A High-Density, Sequence-Enriched Genetic Map of *Hordeum bulbosum* and Its Collinearity to *H. vulgare*

**Neele Wendler, Martin Mascher, Axel Himmelbach, Federica Bini, Jochen Kumlehn, and Nils Stein***

**Abstract**

*Hordeum bulbosum* L., a wild grass and close relative of cultivated barley (*Hordeum vulgare* L.), gained importance in plant breeding as inducer of haploid plants in crosses with barley and also as a genetic resource for introgression of disease resistance/tolerance genes into cultivated barley. Genetic mapping of genes introgressed from *H. bulbosum* is a prerequisite for their efficient utilization in barley breeding, but often hindered due to repressed recombination. The mechanism underlying the reduced frequency or lack of meiotic recombination between *H. bulbosum* and *H. vulgare* chromatin in introgressed segments is not understood. It may be explained by lack of genome collinearity or other structural differences between both genomes. In the present study, two F1 mapping populations of *H. bulbosum* were analyzed by genotyping-by-sequencing (GBS) and four dense *H. bulbosum* genetic maps containing 1449, 996, 720, and 943 SNP markers, respectively, revealed overall a high degree of collinearity for all seven homeologous linkage groups of *H. vulgare* and *H. bulbosum*. The patterns of distribution of recombination along chromosomes differed between barley and *H. bulbosum*, indicating organizational differences between both genomes.

**Core Ideas**

- Genotyping-by-sequencing was used to create high-density genetic maps of the out-crossing species *Hordeum bulbosum*
- Collinearity was compared between the created *H. bulbosum* genetic maps and genetic maps of barley
- A de novo *H. bulbosum* assembly was generated using a doubled haploid *H. bulbosum* plant and gene models were predicted on the assembly
- *H. bulbosum* de novo contigs were integrated into the genetic maps to construct an anchored sequence reference of *H. bulbosum*

**Among all cereal species**, barley (*Hordeum vulgare* subsp. *vulgare*) ranks fourth globally for harvested area and produced tons (http://faostat.fao.org, verified 8 Sept. 2017). It is utilized mainly as animal feed, and to a lesser extent as a resource for malting and various food products (Blake et al., 2011). Barley crop production is threatened by more than 80 different pathogen-caused diseases (Mathre, 1997), of which some may even cause complete yield loss (Paulitz and Steffenson, 2010). For several diseases, genetic control can be achieved by crossing naturally occurring resistance or tolerance genes from genetic resources into elite germplasm (Hajjar and
Hodgkin, 2007). These genetic resources are either represented by genotypes from the primary *H. vulgare* gene pool, which includes the fully compatible wild subspecies *H. vulgare* subsp. *spontaneum*, but also by accessions of further wild relatives.

*Hordeum bulbosum* L. is a close relative of barley, and both species have diverged about six million years ago (Blattner, 2006). *Hordeum bulbosum* occurs either as a diploid (2n = 2x = 14) or tetraploid (4n = 4x = 28) perennial outbreeding species with a self-incompatibility system (Kakeda et al., 2008). Due to crossing barriers to *H. vulgare*, *H. bulbosum* has been defined as a single species representing the secondary gene pool of cultivated barley (von Bothmer et al., 1995). Nonetheless, it has been recognized as a valuable source of genetic diversity for barley improvement, in particular regarding disease resistance or tolerance (Jie and Snape, 1988; Pickering et al., 1997; Ruge et al., 2003; Szigat and Szigat, 1991; Walther et al., 2000). Fertile hybrids of intercrosses between both species may be obtained through embryo rescue via in vitro culture circumventing seed abortion due to crossing barriers. Introggression lines carrying only fragments of *H. bulbosum* chromosomes in the background of the *H. vulgare* genome can be generated by either selfing or backcrossing of tetraploid or triploid *H. vulgare*/*H. bulbosum* hybrids, respectively. Such introgression lines have been shown to carry disease resistance against pathogens like leaf rust (*Puccinia hordei*), stem rust (*Puccinia graminis*), bunt virus, barley yellow dwarf virus (BYDV), scald (*Rhynchosporium commune*), powdery mildew, and Septoria speckled leaf blotch (*Septoria passarini*; Fetch et al., 2009; Pickering et al., 2004; Ruge et al., 2003; Scholz et al., 2009; Shtaya et al., 2007; Singh et al., 2004; Toubia-Rahme et al., 2003). To make introgressed resistances more accessible to plant breeding, the underlying genes need to be precisely mapped and closely linked markers have to be established for efficient marker-assisted selection. Furthermore, negative linkage drag, resulting from *H. bulbosum* alleles that are lying on introgressed segments, has to be removed through marker-assisted selection of segments that were reduced in size through recombination.

Recently, molecular methods based on next-generation sequencing and an extensive pool of molecular markers, sequence resources and single-nucleotide polymorphism (SNP) information was produced for the majority of the published *H. vulgare*/*H. bulbosum* introgression lines (Wendler et al., 2015, 2014). These genetic and genomic resources were generated by using genetic and physical maps of *H. vulgare* as genomic reference model for *H. bulbosum* (International Barley Sequencing Consortium, 2012; Mascher et al., 2013a) assuming high genome collinearity between both closely related species (Wendler et al., 2015, 2014). Genetic mapping of introgressed loci from *H. bulbosum* often revealed highly reduced frequency of recombination in introgressed intervals. This phenomenon may be caused by any kind of structural variation or lack of micro-collinearity between both *Hordeum* genomes. So far, collinearity between barley and *H. bulbosum* was analyzed based on 136 restriction fragment length polymorphism markers, revealing conserved order in most cases (Jaffe et al., 2000; Salvo-Garrido et al., 2001). Rearranged markers were found in centromeric regions only (Jaffe et al., 2000; Salvo-Garrido et al., 2001). In a recent study, 12 gene-containing probes located in the centromeric region of barley chromosome 3H were analyzed for studying collinearity to *H. bulbosum* using fluorescent in situ hybridization (Aliyeva-Schnorr et al., 2016) revealing 100% conserved order of these probes. This supports a high degree of synteny of genes between barley and its wild relative (Aliyeva-Schnorr et al., 2016).

In the present study, we established four high-density genetic linkage maps of *H. bulbosum* by using state-of-the-art GBS. To underpin the accuracy and confidence in SNP identification in *H. bulbosum*, a de novo draft sequence assembly of the *H. bulbosum* genome was produced from whole-genome shotgun (WGS) sequencing of a single doubled haploid *H. bulbosum* plant. GBS marker map positions could be directly associated to the draft reference genome of barley (International Barley Sequencing Consortium, 2012; Mascher et al., 2013a) revealing in general a high level of genome collinearity but also different patterns of distribution of recombination between both species.

**Material and Methods**

**Plant Material for Genome Sequencing in *H. bulbosum***

*Hordeum bulbosum* is an outbreeding species, hence wild accessions are expected to be heterozygous at all/most loci of the genome. Heterozygosity complicates whole-genome sequencing and assembly; therefore, we aimed at identifying a highly homozygous accession for shotgun sequencing. A plant of the diploid (2n) *H. bulbosum* accession FB2012-011-2 (FBB), which was collected south of Ubique, Andalusia, Spain (altitude: 730 m) was used to produce a doubled haploid plant (FBB_170) through anther culture as described below. Grains of this accession were kindly provided by Frank Blattner (IPK Gatersleben, Germany). Flow cytometric analysis (Achigan-Dako et al., 2008) of the plant FBB_170 confirmed its spontaneously doubled genome constitution. Genomic DNA of this plant was used for WGS sequencing. Due to lack of fertility and overall weakness, this plant could not be maintained. The lack of fertility and weakness of the doubled haploid plant may be due to genomic instabilities that arise when the normally outcrossing species gets homozygous. Inbreeding depression is a common phenomenon in animals and outbreeding plants and has been discussed before (Charlesworth and Willis, 2009). However, the donor plant is maintained clonally.

**Plant Material for Genetic Mapping in *H. bulbosum***

To produce *H. bulbosum* F₁ mapping populations, seeds of the diploid *H. bulbosum* accessions FB2012-010 (FBC), collected in Cortes de la Frontera (Sierra Bermecha),
Spain (altitude: 548 m) and FB2012-007A (FBD), collected in northwestern Terifa, Spain, were grown together with FBB in the greenhouse in the autumn of 2013. Grains of these accessions were kindly provided by Frank Blattner (IPK Gatersleben, Germany). Starting at four weeks after germination, young plants were maintained at 4°C for 8 wk for vernalization.

Clonal bulbs of the diploid *H. bulbosum* accession PB1 (Salvo-Garrido et al., 2001) were kindly provided by Pierre Devaux (Florimond Desprez, Cappelle-en-Pévèle, France). After recovery of the shoots and roots, the PB1 plants were kept at 4°C for 8 wk to be vernalized.

To produce F1 progeny, flowering ears of a single plant of each accession as well as the clonal plants of PB1 were bagged together in pairs as described earlier (Jaffe et al., 2000; Salvo-Garrido et al., 2001). This resulted in two F1, mapping populations of 190 (FBB1 × PB1) and 272 (FBC_3 × FBD_5) individuals, respectively. Thus, each mapping population goes back to a single individual of the accessions FBB, FBC, FBD, and PB1, respectively. The F1 populations as well as the parental genotypes were subjected to GBS. Two technical replicates for each parental genotype were included into the GBS analyses. The parents and the progeny are maintained at the IPK Gatersleben. DNA of the doubled haploid plant as well as fresh leaf material as described earlier (Stein et al., 2001).

**Production of Doubled Haploid**

*H. bulbosum* Plants

Doubled haploids of *H. bulbosum* accession FBB were produced largely following a protocol previously published for *H. vulgare* (Lippmann et al., 2015). Briefly, anthers containing highly vacuolated, premiotic microspores were dissected from surface-sterilized spikes and subjected to starvation in 0.4 M mannitol solution for 4 d so as to induce pollen embryogenic competence. Viable pollen grains were extracted from pretreated anthers by tissue homogenization, sieving and discontinuous density gradient centrifugation and were cultivated at a density of 100,000/mL in liquid KPB medium as described by Coronado et al. (2005). Pollen-derived embryogenic calli and embryo-like structures were then grown on fresh KPB medium solidified by supplementation with 0.4% Phytagel (P8169, Sigma-Aldrich, St. Louis, MO). Plantlets which emerged after further subcultivation on K4NB regeneration medium (Kumlehn et al., 2006) were subsequently established in soil. More than 70% of the plants analyzed by flow-cytometry (Daghma et al., 2014) had spontaneously undergone whole-genome duplication.

**Next-Generation Sequencing Library Preparation**

Genomic DNA of the doubled haploid *H. bulbosum* plant FBB_170 (accession FB2012-011-2, FBB) was used for WGS sequencing. Whole-genome shotgun library preparation and sequencing followed previously established procedures (Mascher et al., 2013b). In brief, Illumina TruSeq paired-end libraries (Illumina Part No. 15026486) were prepared according to the manufacturer’s instructions (Illumina, Inc., San Diego, CA). Genomic DNA (1 μg) was fragmented (size range: 300–400 bp) using Covaris microTUBES and the Covaris S220 Instrument (175 W ultrasonic power, 10% duty factor, 200 cycles per burst, 60 s treatment time, Covaris, Inc., Woburn, MA).

After adaptor ligation, the DNA library was size selected (size range: 320–420 bp) by excision from a SYBR-Gold stained agarose gel (Life Technologies GmbH, Invitrogen, Darmstadt, Germany) and purified using AmPure XP beads (Beckman Coulter GmbH, Krefeld, Germany) and the library size profile was monitored on a DNA 7500 chip by using an Agilent 2100 Bioanalyzer (Agilent Technology, Santa Clara, Part No. 5067-1506). After quantification by qPCR as described elsewhere (Mascher et al., 2013b) the DNA library was sequenced on an Illumina HiSeq2000 device (Illumina, Inc., San Diego, CA).

For genotyping the F1 populations, GBS libraries were prepared from genomic DNA by using the two restriction enzymes *MspI* and *PstI* High-Fidelity (NEB, Ipswich, UK), and sequenced as described earlier (Wendler et al., 2013, 2014). Up to 100 samples with individual indices (Supplementary File S1) were multiplexed and sequenced together on a single HiSeq2000 lane (Illumina, San Diego, CA). The formation of clusters on the cBot instrument and 1 × 100 bp single-end sequencing (HiSeq2000) were performed as recommended by the manufacturer.

**Next-Generation Sequencing Data Analysis**

Sequencing raw reads of the doubled haploid *H. bulbosum* accession FBB_170 were quality-trimmed and assembled using the applications quality_trim with default parameters and clc_novo_assemble (parameters: fbs 150520) of the software package CLC assembly cell 3.2.2 (http://www.clcbio.com/, verified 8 Sept. 2017).

Barley high-confidence (HC) and low-confidence (LC) gene sequences (International Barley Sequencing Consortium, 2012) were mapped against the *H. bulbosum* assembly using the blastall command of BLAST (Altschul et al., 1990) with default parameters. Only BLAST hits with a maximum e-value of 0, at least 90% sequence identity and 300 bp length were kept for further analysis. In cases of multiple hits, best hits were chosen considering e-value and bit-scores. The genetic map positions of the barley genes were taken from The International Barley Sequencing Consortium (2012) and the population sequencing (POPSEQ) barley map (Mascher et al., 2013a). Furthermore, barley HC and LC genes were mapped against the *H. bulbosum* assembly using the GMAP program (Wu and Watanabe, 2005; parameters: -K 50000 and -L 1000000) to observe gene model coverage by the assembly.

Demultiplexing of raw GBS sequence reads was achieved with the CASAVA pipeline 1.8 (Illumina, San Diego, CA). Adaptor sequences were trimmed from GBS sequence reads using cutadapt (Martin, 2011) and reads shorter than 30 bp after adaptor removal were discarded. Trimmed GBS reads were mapped against the de novo assembled *H. bulbosum* reference sequence using BWA
version 0.7.5 (Li and Durbin, 2009). The BWA command “mem” was called with parameter “-k 10” to set the minimum seed length and “-M” to mark shorter split hits as secondary, otherwise default settings were used.

SNP calling on GBS sequence data was performed with the SAMtools pipeline (v.0.1.19) using the commands “samtools mpileup” and “bcftools” with default settings (Li et al., 2009). The SAMtools command mpileup was called with the parameter “-C 50” for downgrading mapping quality for reads containing excessive mismatches. The additional parameter “-D” was used for SAMTools mpileup to obtain per sample read depth. Subsequently, GBS genotype calls were filtered with a custom AWK script (Mascher et al., 2013c) and the R statistical environment (R Core Team, 2012; http://www.r-project.org/, verified 8 Sept. 2017), selecting only SNPs with a minimum 10-fold sequence read depth and genotype quality of at least 10 for homozygous and heterozygous calls.

Overlaps between called SNPs and GMAP-predicted barley genes on the *H. bulbosum* assembly were identified using the command “bedtools intersect” of the BEDTools suite with default parameters (Quinlan and Hall, 2010).

### High-Density Genetic Map Calculation

A high-density genetic map was calculated separately for each individual parental line, using markers that are heterozygous in one parent while being homozygous in the other (i.e., AB × AA). Seven progeny plants that turned out to have arisen from self-pollination were removed from the analysis. Thus, at SNP positions where the two crossing parents had alternative homozygous SNP calls, these progenies appeared to have fixed homozygous markers of the selfed parent only. GBS SNPs were only considered if lying on a *H. bulbosum* assembly contig that could be anchored to a barley gene as described above. Map calculation was performed using the Kosambi function and Maximum Likelihood Mapping of the program JoinMap v.4.1 (Van Ooijen, 2011). The JoinMap “CP” function for outcrossing species was applied (Van Ooijen, 2011). Markers and individuals were excluded from the analysis if they exhibited more than 10 and 20% missing data, respectively. A minimum LOD score of 14 was considered to separate the seven linkage groups. For individual chromosome maps of the four parental lines, initially a first map was calculated and then subjected to manual correction. Nonfitting markers were excluded and the maps were recalculated until no uncertain marker positions remained. Removing nonfitting markers sufficed to reach a stable map. The resulting marker order of each map was compared with high-density genetic maps of barley (International Barley Sequencing Consortium, 2012; Mascher et al., 2013a).

### Anchoring of Additional Assembly Contigs into the Framework Maps

Genotyping-by-sequencing SNPs lying on *H. bulbosum* WGS contigs, which were not included in the high-density genetic maps at first instance could be included into these frameworks subsequently by applying an approach that anchors markers with high numbers of missing values into genetic bins of a framework map as described earlier (Mascher et al., 2013a). This comprised also GBS SNPs that were lying on *H. bulbosum* assembly contigs, which could not be assigned to barley genes. In brief, this method uses segregation patterns in a biparental population to linearly arrange sequence contigs by assigning them to recombination bins of a framework linkage map of the genome (Mascher et al., 2013a). To do so, a nearest-neighbor search was performed to find the set of framework markers with minimal Hamming distance to a given SNP (i.e., the smallest number of alternative SNP alleles that are required to change an observed segregation pattern into the reference allele, and vice versa). If multiple framework markers showed identical minimal distances, the following consistency checks were applied: (i) ≥95% of the framework markers had to be positioned on the same chromosome, and (ii) the median absolute deviation of their genetic positions was required to be <3 cm. SNPs were placed into the framework map using the program anchoring as published (Mascher et al., 2013a). Subsequently, SNP markers were filtered using the R statistical environment (R Core Team, 2012) to keep only *H. bulbosum* contigs where the major chromosome is supported by ≥95% of framework map markers and the standard deviation of centiMorgan positions was at maximum 3.

### Results

#### A Draft de novo Genome Assembly of *H. bulbosum*

Stringent read mapping is a most crucial prerequisite for using GBS for genetic mapping purposes. Highest stringency can only be obtained if a sequence assembly of the exact species instead of a close relative is used as has been employed in previous studies using GBS analysis of *H. vulgare* vs. *H. bulbosum* (Wendler et al., 2014). A draft genome assembly of a doubled haploid *H. bulbosum* plant (FBB_170) was assembled from WGS sequencing data corresponding to about 15-fold haploid genome coverage. The automated de novo sequence assembly resulted in a total of 634,982 contigs (Supplementary File S2) with an L50 (50% of all nucleotides of the assembly are contained in contigs of at least the length of L50; International Human Genome Sequencing Consortium, 2001) of 3376 bp and a total assembly length of 0.94 Gbp (Table 1), which equals about 20% of the estimated 4.5 Gbp *H. bulbosum* genome size (Eilam et al., 2009; Jakob et al., 2004).

For subsequent steps of analysis, like comparative GBS marker mapping in *H. vulgare* and *H. bulbosum*, it is important to have a comprehensive understanding of putatively orthologous sequence loci between both *Hordeum* species. We determined uniquely alignable homologous regions between the draft assemblies of the barley and the *H. bulbosum* genomes by aligning *H. vulgare* gene sequences (International Barley Sequencing Consortium, 2012) against the de novo assembly of *H. bulbosum* generated in the present study. Only HC matches of a
Table 1. *Hordeum bulbosum* assembly statistics.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum contig length</td>
<td>36,023 bp</td>
</tr>
<tr>
<td>Minimum contig length</td>
<td>200 bp</td>
</tr>
<tr>
<td>Average contig length</td>
<td>1487 bp</td>
</tr>
<tr>
<td>L50†</td>
<td>3376 bp</td>
</tr>
<tr>
<td>Total assembly length</td>
<td>0.94 Gbp</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>634,982</td>
</tr>
<tr>
<td>Contigs &gt; 1000 bp in size</td>
<td>229,986</td>
</tr>
<tr>
<td>Contigs &gt; 1500 bp in size</td>
<td>171,048</td>
</tr>
<tr>
<td>Contigs &gt; 2000 bp in size</td>
<td>133,531</td>
</tr>
</tbody>
</table>

† 50% of all nucleotides of the assembly are contained in contigs of at least the length of L50.

The International Human Genome Sequencing Consortium (2001).

minimum length of 300 bp and 90% sequence identity were considered. Under these criteria, 18,342 barley gene models could be connected to 25,304 *H. bulbosum* de novo sequence contigs (Supplementary File S3), allowing us to connect genetic positions of barley contigs (International Barley Sequencing Consortium, 2012; Mascher et al., 2013a) to *H. bulbosum* assembly contigs. This analysis provided the basis for the subsequent global analysis of genome collinearity. By taking a closer look at the representation of *H. vulgare* gene models in the de novo *H. bulbosum* assembly, it was found that 16,227 barley HC and 31,428 LC genes (confidence definition according to the International Barley Sequencing Consortium, 2012) were covered to at least 90% in the *H. bulbosum* sequences (Supplementary File S4).

High-Density Genetic Maps

To calculate genetic maps of *H. bulbosum*, GBS was performed with the two F₁ mating populations of *H. bulbosum* (FBB₁ × PB1 and FBC₃ × FBD₅). The GBS sequencing reads were mapped against the *H. bulbosum* de novo assembly for variant detection. High quality SNP calls supported by at least 10-fold sequence read coverage in all individuals of the mapping population were used for map calculation. An initial visual check of the graphical GBS SNP genotypes revealed one and six self-pollinated samples in the FBB₁ and FBC₃ populations, respectively. Thus, SNP positions with alternative homozygous allele calls in the putative crossing parents were fixed for the allele of the selfed parent in these progenies. These samples were removed from further analysis. GBS SNPs lying on *H. bulbosum* contigs connected to *H. vulgare* gene models were used for the calculation of high-density genetic maps. As described earlier (Ward et al., 2013) only SNP markers segregating in only one parental genotype were used for the analysis. Thus, individual maps were calculated basically for each parental line using SNP markers, which were heterozygous in one parent and homozygous in the other parent (i.e., AA × AB or AB × BB; Ward et al., 2013). Genetic maps for each parent were calculated using only SNPs with a maximum of 10% and individuals with at most 20% missing data points, respectively. Seven linkage groups corresponding to the seven *H. bulbosum* chromosomes were obtained at a LOD score threshold of 14. Map calculation for individual linkage groups was iterated several times by repeatedly excluding nonfitting SNP markers and SNP markers with uncertain map positions (i.e., SNP markers with multiple double crossovers, segregations distortions that did not fit the overall segregation, and missing data points at recombination points). The genetic map of *H. bulbosum* accession PB1 is shown in Fig. 1.

High-density genetic maps and their anchoring information to the *H. vulgare* genetic maps (International Barley Sequencing Consortium, 2012; Mascher et al., 2013a) can be found in Supplementary File S5. The final maps contained 1449, 996, 720, and 943 markers for individual parental maps of PB1, FBB₁, FBC₃, and FBD₅, respectively. Numbers of markers per chromosome and number of genetic bins per chromosome are listed in Table 2. The cumulative map lengths were 764, 621, 658, and 716 cM for the parental maps of PB1, FBB₁, FBC₃, and FBD₅, respectively.

In total, 2412 unique *H. bulbosum* assembly contigs were anchored to at least one of the four genetic maps. The cumulative length of these contigs exhibited 18,435,588 bp (18.4 Mbp).

Five-hundred and eighty-six contigs harbored markers that were placed in at least two maps, enabling the analysis of the collinearity between pairs of maps (Fig. 2). Overall, maps were collinear, but the low number of shared markers precludes an in-depth comparison.

Collinearity Analysis between Barley and *H. bulbosum*

Collinearity could be compared between genetic maps of barley and *H. bulbosum* because 1325, 921, 672, and 877 markers of the individual parental maps of PB1, FBB₁, FBC₃, and FBD₅ contained homology information to anchored barley genes (Fig. 3). Genetic positions between the barley POPSEQ map (Mascher et al., 2013a) were compared with the genetic positions of their putative orthologs in our *H. bulbosum* maps of FBB₁, FBC₃, FBD₄, and PB1, respectively. Overall, collinearity and synteny between the presently most dense barley genetic map (Mascher et al., 2013a) and the calculated *H. bulbosum* maps was very high. Thus, from 3757 GBS markers, 3549 (95%) were in collinear order between *H. bulbosum* and *H. vulgare*.

Minor rearrangements could be detected for markers being positioned in a different order within a chromosome or even between chromosomes. From a total of 208 rearranged markers, the positions of 25 were consistent when compared with the International Barley Sequencing Consortium (2012) map. Thus, these rearrangements are rather due to a misordering in the barley POPSEQ map (Mascher et al., 2013a) than representing true rearrangements in the *H. bulbosum* vs. the *H. vulgare* genome. Furthermore, 73 markers showed rearrangements below a 3 cM resolution. Since the barley genetic map does not resolve below a 3 cM resolution (Mascher et al., 2013a), these rearrangements cannot be assigned as certain.
Fig. 1. Genetic map of *Hordeum bulbosum* accession PB1. Genetic map of the *H. bulbosum* accession PB1 was calculated using JoinMap v.4.1 (Van Ooijen, 2011). Markers that are heterozygous in one parent while being homozygous in the other (i.e., ABxAA) were used for map calculation. Map calculation was performed using the Kosambi function and Maximum Likelihood Mapping of the program JoinMap. The JoinMap “CP” function for out-crossing species was applied (Van Ooijen, 2011). Markers and individuals were excluded from the analysis if they exhibited >10 and 20% missing data, respectively. A minimum LOD score of 14 was considered to separate the seven linkage groups (1Hb to 7Hb). GBS markers were labeled as “M1” to “Mx” for simplification. The figure was generated using the MapChart program (Voorrips, 2002).
Table 2. Genetic bins and markers per *H. bulbosum* chromosome (B/Chr and M/Chr).

<table>
<thead>
<tr>
<th>Chr</th>
<th>B/Chr</th>
<th>M/Chr</th>
<th>B/Chr</th>
<th>M/Chr</th>
<th>B/Chr</th>
<th>M/Chr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1Hb</td>
<td>19</td>
<td>30</td>
<td>15</td>
<td>7</td>
<td>150</td>
<td>10</td>
</tr>
<tr>
<td>2Hb</td>
<td>38</td>
<td>230</td>
<td>19</td>
<td>79</td>
<td>27</td>
<td>137</td>
</tr>
<tr>
<td>3Hb</td>
<td>40</td>
<td>128</td>
<td>24</td>
<td>136</td>
<td>25</td>
<td>110</td>
</tr>
<tr>
<td>4Hb</td>
<td>19</td>
<td>148</td>
<td>9</td>
<td>87</td>
<td>21</td>
<td>108</td>
</tr>
<tr>
<td>5Hb</td>
<td>44</td>
<td>216</td>
<td>46</td>
<td>178</td>
<td>42</td>
<td>129</td>
</tr>
<tr>
<td>6Hb</td>
<td>38</td>
<td>230</td>
<td>19</td>
<td>79</td>
<td>27</td>
<td>137</td>
</tr>
<tr>
<td>7Hb</td>
<td>34</td>
<td>188</td>
<td>25</td>
<td>158</td>
<td>26</td>
<td>153</td>
</tr>
</tbody>
</table>

repositioning in the *H. bulbosum* genome. Of the remaining inconsistent markers, 92 markers were swapped between chromosomes and 18 markers were reordered within a chromosome, but above 3 cM resolution.

**Anchoring WGS Contigs into High-Density Framework Maps**

To enrich the HC genetic map of *H. bulbosum* by more sequences from the de novo WGS assembly, a method that integrates SNPs into a framework map based on their segregation patterns was applied (Mascher et al., 2013a). As described above, additional GBS SNP markers were genetically placed by assigning them to their nearest neighbors in the high-density genetic map of each parental plant. Additional SNPs that were to be anchored were filtered to have <90% missing data points. Using this approach, 9611 additional *H. bulbosum* contigs were integrated into the *H. bulbosum* maps (Supplementary File S6). Overall, a total of 12,023 contigs were anchored representing a cumulative length of more than 71 Mbp (71,158,787 bp), which equals about 1.6% of the *H. bulbosum* genome (Eilam et al., 2009; Jakob et al., 2004).

**Discussion**

**Sequence-Based High-Density Genetic Maps of *H. bulbosum***

Using GBS, four high-density genetic maps of *H. bulbosum* were generated. Compared with its relative *H. vulgare*, linkage mapping in *H. bulbosum* is considerably more complicated, owing to the out-breeding mating system of this species. While in self-pollinating species such as barley most loci will be expected to be homozygous, individual plants of *H. bulbosum* are likely to be heterozygous at many loci. Furthermore, *H. bulbosum* is self-incompatible (Kakeda et al., 2008), so mapping has to be performed in F₁ generations that are derived from highly heterozygous crossing parents. The latter may result in up to four different alleles segregating at a given locus. Nonetheless, F₁ populations allow the creation of individual parental maps based on markers that are heterozygous in one parent while being homozygous in the other parent (i.e., AA × AB or AB × BB).

The four *H. bulbosum* maps generated in the present study have cumulative map lengths of 764, 621, 658, and 716 cM for PBI, FBB1, FBC_3, and FBD_5, respectively. Therefore, the maps seem to represent nearly the entire *H. bulbosum* genome, which was estimated to be about 700 cM in size (Salvo-Garrido et al., 2001). This is about 30% shorter than genetic maps of barley, whose lengths generally comprise about 1000 cM (Close et al., 2009; Mascher et al., 2013a; Poland et al., 2012; Stein et al., 2007).

The final maps comprised 3757 unique GBS markers on 2412 unique *H. bulbosum* assembly contigs. Furthermore, with 166, 167, 251, and 252 individuals the used populations were at least two times larger than those used in previous studies (Jaffe et al., 2000; Salvo-Garrido et al., 2001). Thus, these *H. bulbosum* genetic maps have a much higher resolution and density than those constructed in previous mapping efforts, where 137 molecular markers were mapped on the basis of 111 individuals (Jaffe et al., 2000; Salvo-Garrido et al., 2001).

When comparing marker order to the barley POPSEQ genetic map (Mascher et al., 2013a), some rearrangements were detected below 3 cM resolution. However, the POPSEQ map loses accuracy at a threshold of 3 cM (Mascher et al., 2013a). Moreover, 92 markers were positioned on different chromosomes between barley and *H. bulbosum*, and 18 markers were reordered within a chromosome but above 3 cM resolution. These markers may represent truly different genomic noncollinear positions. Reasons could include translocation, inversion, duplication, and differential deletion of paralogs. However, it cannot be excluded that they are a result of (i) mis-assemblies of either the barley or *H. bulbosum* map, (ii) unspecific sequence alignments of *H. bulbosum* GBS reads caused by repetitive sequences or errors in the *H. bulbosum* WGS assembly. Overall, *H. bulbosum* maps exhibited highly conserved marker order when compared to *H. vulgare*. This is in line with previous studies, which suggested up to 100% conserved synteny between the relatives (Aliyeva-Schnorr et al., 2016; Jaffe et al., 2000; Salvo-Garrido et al., 2001).

A different distribution of recombination between barley and *H. bulbosum* was observed: in *H. bulbosum* the recombination rate in centromeric regions was lower, while it was higher at telomeres compared with *H. vulgare* (Fig. 3). The latter phenomenon has been observed before and was related to the observation that introgressions of *H. bulbosum* in barley have strong bias toward the telomeres (Salvo-Garrido et al., 2001).

Interestingly, a total of seven plants in two populations appeared to be derived from self-pollination. Self-fertility (at least partial) has not been reported before in *H. bulbosum*. However, the two loci-based self-incompatibility systems (Kakeda et al., 2008) may allow a certain degree of self-fertilization as is known from rye (Secale cereale L.) were self-seeding is expected at a rate of about 1 to 10% (Lundqvist, 1958).

**An Anchored *H. bulbosum* Sequence Assembly**

Anchored and ordered sequence assemblies of species provide a valuable resource as a reference for read mapping and SNP detection in resequencing studies (Jost et
Fig. 2. Comparison between different genetic maps of *H. bulbosum* and barley. The genetic positions of markers that are physically co-located on the same sequence contigs were compared across maps. Each row corresponds to a pair of two genetic maps as indicated to the left. Columns correspond to chromosomes.
al., 2016; Mascher et al., 2014). In *H. bulbosum*, no such anchored assembly has been available yet.

The present study implemented an approach (Mascher et al., 2013a). This allowed integration of *H. bulbosum* WGS contigs into the high-density framework maps and to establish a reference of anchored *H. bulbosum* sequence contigs. This allowed to integrate 9611 additional *H. bulbosum* contigs into the framework of *H. bulbosum* maps, resulting in a total of 12,023 anchored contigs. These contigs exhibited a cumulative length of more than 71 Mbp and thereby represented approximately 7.5% of the total *H. bulbosum* assembly (0.94 Gbp). In total, 16,227 HC and 31,428 LC *H. vulgare* genes are represented by these 12,023 contigs to at least 90%. Thus, this anchored WGS sequence assembly will be a valuable resource, enabling future resequencing experiments (e.g., as a reference for variant detection). In principal, the reference assemblies of the close relative *H. vulgare* can be used for read mapping of *H. bulbosum* as well (Wendler et al., 2014). However, it was shown that due to sequence diversity between barley and *H. bulbosum*, only 40 and 70% of *H. bulbosum* GBS and exome capture resequencing reads could be mapped to the barley genome sequence, respectively (Wendler et al., 2014). By contrast, 95% of the GBS reads could be mapped to the current *H. bulbosum* assembly. Thus, the efficiency of detecting polymorphisms will be accordingly higher when using the reference of *H. bulbosum* itself. Furthermore, predicted gene models may be used for candidate gene prediction in map-based cloning approaches involving introgression lines carrying chromosomal segments from *H. bulbosum* in a barley background (Fetch et al., 2009; Johnston et al., 2009; Pickering et al., 2004; Ruge et al., 2003; Scholz et al., 2009).

**Conclusion**

The present study provided the first high-density genetic maps of *H. bulbosum* and a comprehensive collinearity comparison to barley using the power of GBS. Furthermore, de novo assembled *H. bulbosum* contigs were integrated into the framework maps. The four genetic maps comprised in total 3757 markers. 12,023 WGS *H. bulbosum* contigs of a cumulative length of 71 Mbp could be anchored into this framework. The comparison of genetic marker/gene order revealed a very high degree of genome collinearity between *H. vulgare* and *H. bulbosum*, confirming earlier findings. The genetic maps and the knowledge of synteny will provide an important tool for a more efficient utilization of previously established genomic and genetic resources of *H. bulbosum* (e.g., SNP data and introgression lines). Anchored sequence contigs of *H. bulbosum* will provide a reference for read mapping in future resequencing approaches. And finally, predicted gene models on the *H. bulbosum* assembly may be used to identify candidate genes.

**Accession Numbers**

Sequence data from this article can be found in the EMBL/GenBank data libraries under accession number PRJEB20657 (GBS) and PRJEB20646 (WGS assembly).
Supplementary Information Available

Supplementary File S1. List of sample IDs and genotyping-by-sequencing indices. This file contains a list of sample IDs and indices that were used for GBS experiments. File can be downloaded from eDAL (Arend et al., 2014) at http://dx.doi.org/10.5447/IPK/2017/14.

Supplementary File S2: De novo assembly of *H. bulbosum*. De novo assembled contigs of the *H. bulbosum* doubled haploid line FBB_170. File can be downloaded from eDAL (Arend et al., 2014) at http://dx.doi.org/10.5447/IPK/2017/12.

Supplementary File S3. Blast alignments of barley gene sequences on *H. bulbosum* assembly. barley high-confidence (HC) and low-confidence (LC) gene sequences (International Barley Sequencing Consortium, 2012) were mapped against the *H. bulbosum* assembly using BLAST (Altschul et al., 1990). Thresholds were set to a maximum e-value of 0, at least 90% sequence identity and at minimum 300 bp alignment length. In cases of multiple hits, best hits were chosen considering e-value and bit-scores. The genetic map positions of the barley genes were taken from The International Barley Sequencing Consortium (2012) and the population sequencing (POPSEQ) barley map (Mascher et al., 2013a). File can be downloaded from eDAL (Arend et al., 2014) at http://dx.doi.org/10.5447/IPK/2017/11.

Supplementary File S4. Gmap alignments of barley gene sequences on *H. bulbosum* assembly. barley high-confidence (HC) and low-confidence (LC) gene sequences (International Barley Sequencing Consortium, 2012) were mapped against the *H. bulbosum* assembly using the GMAP program (Wu and Watanabe, 2005). File can be downloaded from eDAL (Arend et al., 2014) at http://dx.doi.org/10.5447/IPK/2017/12.

Supplementary File S5. High-density genetic maps of *H. bulbosum* and anchoring information to barley. Parental genetic maps were calculated for *H. bulbosum* using the program JoinMap v4.1 (Van Ooijen, 2011) for each individual parental line. Markers were used that are heterozygous in one parent while being homozygous in the other (i.e. ABKAA). GBS SNPs were only considered if lying on a *H. bulbosum* assembly contig that could be anchored to a barley gene as described above. Markers and individuals were excluded from the analysis if they exhibited more than 10% and 20% missing data, respectively. A minimum LOD score of 14 was considered to separate the seven linkage groups. The resulting marker order of each map was compared to high-density genetic maps of barley (International Barley Sequencing Consortium, 2012; Mascher et al., 2013a). File can be downloaded from eDAL (Arend et al., 2014) at http://dx.doi.org/10.5447/IPK/2017/16.

Supplementary File S6. Integration of de novo assembled *H. bulbosum* contigs into framework maps. *H. bulbosum* WGS contigs, which were not included in the high-density genetic maps at first instance were included into these frameworks by applying an approach that anchors markers into genetic bins of a framework map as described earlier (Mascher et al., 2013a). To do so, a nearest-neighbor search was performed to find the set of framework markers with minimal Hamming distance to a given SNP. SNPs were placed into the framework map using the program canchor as published (Mascher et al., 2013a). Subsequently, SNP markers were filtered using the R statistical environment (R Core Team, 2012) to keep only *H. bulbosum* contigs where the major chromosome is supported by > = 95% of framework map markers and the standard deviation of cM positions was at maximum 3. File can be downloaded from eDAL (Arend et al., 2014) at http://dx.doi.org/10.5447/IPK/2017/15.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

Acknowledgments

We gratefully acknowledge the excellent technical support by Ines Walde, Manuela Knauf, Susanne König, Sandra Driesslein and Enk Geyer (IPK). We like to thank Anne Siebold for raw data submission and Uwe Scholz and the research group Bioinformatics and Information Technology at IPK for computational resources. We also give our best thanks to Uwe Scholz for uploading additional files at eDAL (Arend et al., 2014). Furthermore, thanks to Frank Blattner and Pierre Devaux for providing us with material of the *H. bulbosum* accessions. The work was financially supported by a grant (‘TRANSBULB’, FKZ 0315966) from the German Federal Ministry of Education and Research (BMBF) to NS.

References


