Journal: Theoretical and Applied Genetics

Title: A haplotype dictionary for the *Rht-1* loci in wheat

Edward P. Wilhelm, Ian J. Mackay, Robert J. Saville, Andrey V. Korolev, Francois Balfourier, Andy J. Greenland, Margaret I. Boulton, and Wayne Powell

National Institute of Agricultural Botany, Huntingdon Rd., Cambridge CB3 0LE, United Kingdom (E.P.W., I.J.M., A.J.G., W.P.)
John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom (E.P.W., R.J.S., A.V.K., M.I.B.)
INRA, UMR 1095, Genetics, Diversity and Ecophysiology of Cereals, F-63100 Clermont-Ferrand, France (F.B.)
UBP, UMR 1095, Genetics, Diversity and Ecophysiology of Cereals, F-63100 Clermont-Ferrand, France (F.B.)

Corresponding Author:
Edward P. Wilhelm
E-mail: edwilhelm@yahoo.com
Phone: 618-530-9691

Present address for Wayne Powell
Institute of Biological, Environmental, and Rural Sciences, Aberystwyth SY23 3DA, United Kingdom

Present address for Robert J. Saville
East Malling Research, New Road, East Malling, Kent ME19 6BJ, United Kingdom

Abstract

The introduction of *Reduced height* (*Rht*)-*B1b* and *Rht-D1b* into bread wheat (*Triticum aestivum*) varieties was a key component of the 'green revolution' and today these alleles are the primary sources of semi-dwarfism in wheat. The *Rht-1* loci encode DELLA proteins, which are transcription factors that affect plant growth and stress tolerance. In bread wheat, *Rht-D1b* and *Rht-B1b* influence resistance to the disease Fusarium Head Blight. To identify *Rht-1* variants, locus specific primers were developed and used to sequence the entire open reading frame (ORF) and 1.7 kb of the 5' and 0.5 kb of the 3' flanking regions of *Rht-A1* (*Rht-A1*+f), *Rht-B1* (*Rht-B1*+f), and *Rht-D1* (*Rht-D1*+f) in bread wheat (36 sequences from each genome) and tetraploid and diploid wheat (one to three sequences from each genome). Among the bread wheat accessions, the *Rht-A1*+f and *Rht-D1*+f sequences contained relatively low genetic diversity and few haplotypes relative to the *Rht-B1*+f sequences. The tetraploid and diploid wheat accessions were relatively rich in genetic diversity and contained the majority of the polymorphic sites. Novel polymorphisms, relative to 'Chinese Spring', discovered among the accessions include 160 bp and 197 bp insertions 5' of *Rht-B1* and a frameshift in the *Rht-B1* ORF. Quantitative real-time PCR using shoot and leaf tissue from five-day old seedlings of genotypes lacking or containing the 5' insertions revealed no major affect on *Rht-B1* transcript accumulation. This research provides insights into the genetic diversity present at the *Rht-1* loci in modern bread wheat and in relation to ancestral wheat accessions.

Keywords:
Reduced height, nucleotide diversity, dwarf, wheat, DELLA, haplotype diversity

Author Contribution Statement:
E.P.W., M.I.B., I.J.M., A.J.G., and W.P. designed the research; E.P.W, R.J.S., A.V.K., and M.I.B. performed the research; F.B. provided 372CC germplasm and marker data; E.P.W. analyzed data; and E.P.W. wrote the paper.

INTRODUCTION

The incorporation of the semi-dwarf alleles *Rht-B1b* & *Rht-D1b* at the *Reduced height* (*Rht)-1* loci in bread wheat (*Triticum aestivum*) was an important component of the 'green revolution', which abated a major worldwide food shortage (Hedden 2003). The *Rht-1* loci encode DELLA proteins, which in plants serve a key biological function by integrating hormonal and environmental signals that affect overall growth (Alvey and Harberd 2005; Achard *et al.* 2006; Alvey and Boulton 2008) and are associated with abiotic and biotic stress tolerance (Achard and Genschik 2009). The DELLA protein functions to restrict plant growth, but in the presence of gibberellin (GA), DELLA is degraded, removing this restriction. The C-terminus of the protein acts to repress growth while the N-terminus, which includes the DELLA domain, is involved in GA sensitivity. In hexaploid bread wheat (AABBDD genomes), the DELLA proteins are encoded by three homoeoloci: *Rht-A1*, *Rht-B1*, and *Rht-D1* located on the group 4 chromosomes of the A, B, and D genomes, respectively.

In wheat, *Rht-1a* alleles encode for wild type (tall) plants with DELLA proteins that are GA sensitive. The open reading frame (ORF) sequences are available for *Rht-A1a* (Genbank acc no. JF930277; Pearce *et al*. 2011), *Rht-B1a* (Genbank acc no. JF930278; Pearce *et al*. 2011), and *Rht-D1a* (Genbank acc no. AJ242531; Peng *et al.* 1999). The sequences of the three *Rht-1a* homoeologs are well conserved, having 96.8% of the amino acid (AA) identities in common (Pearce *et al*. 2011; Wilhelm *et al.* 2013). Rht-A1a, Rht-B1a, and Rht-D1a are each expressed in wheat stem tissue and have similar expression patterns (Pearce *et al.* 2011). Several *Rht-1* alleles that encode GA insensitive proteins and have reduced height have been sequenced, including the *Rht-B1b* and *Rht-D1b* 'green revolution' genes. The *Rht-B1b* and *Rht-D1b* ORFs each contain a single nucleotide polymorphism (SNP) in the DELLA domain, which introduces a premature stop codon (Peng *et al.* 1999). The resulting protein is less responsive to GA due to an apparent N-terminal truncation, which thereby serves to repress plant growth even in the presence of GA (Peng *et al.* 1999). Additional GA insensitive alleles that have been

sequenced include *Rht-B1c*, *Rht-B1d*, *Rht-B1e*, *Rht-D1c*, and *Rht-D1d*. Polymorphisms that cause height differences have been identified in *Rht-B1c* (an insertion in the DELLA domain of up to 2,026 bp that following translation and splicing results in a predicted 30 AA insertion; Pearce *et al.* 2011; Wu *et al.* 2011), *Rht-B1e* (a stop codon in the DELLA domain; Pearce *et al.* 2011; Li *et al.* 2012b), and *Rht-D1c* (increased copy number of the *Rht-D1b* allele; Pearce *et al.* 2011; Li *et al.* 2012a).

Genetic diversity has been examined in only a few homoeologous gene series in hexaploid wheat, which include *Storage protein activator* (*Spa*; Ravel *et al.* 2009), *GA-dependent MYB transcription factor* (*GAMYB*; Haseneyer *et al.* 2008), and *Glutamine synthetase* (*GS*; Li *et al.* 2011). At the *Rht-1* homoeoloci, the prevalence of *Rht-B1b* and *Rht-D1b* have been measured in several germplasm sets using PCR markers based on the semi-dwarf causative SNPs (Chrpova *et al.* 2003; Zhang *et al.* 2006; Knopf *et al.* 2008; Tosovic-Maric *et al.* 2008; Dan *et al.* 2009; Guedira *et al.* 2010; Gulyas *et al.* 2011). These studies have provided valuable information regarding the distribution of these alleles, but do not reveal sequence variation outside of the *Rht-1b* SNPs. Recently, three full-length BAC clones containing *Rht-1* (representing *Rht-A1*, *Rht-B1*, and *Rht-D1*) and 160 kb or more of flanking sequence were isolated and sequenced (Genbank acc. nos. JX978692-JX978694; Febrer *et al*. 2009; Wilhelm *et al*. 2013). Full-length *Rht-1*-containing BAC sequences are also available for the A genome ancestor *Triticum urartu* (Genbank acc. no. JX978695; Wilhelm *et al.* 2013), the D genome ancestor *Aegilops tauschii* (ecotype AL8/78, Genbank acc. no. HQ435330.1; Duan *et al.* 2012), and the D genome of the near isogenic line 'Aibai/CS' (Genbank acc. no. HQ435325.1; Duan *et al.* 2012). In the 5' and 3' regions flanking *Rht-1*, five conserved non-coding sequences (CNSs) located within 2 kb of the ORF were identified on each wheat homoeolog and in nine *Poaceae* members (Duan *et al.* 2012; Wilhelm *et al.* 2013). In plants, CNSs have been identified that are involved in gene regulation (Uchida *et al.* 2007), suggesting the five CNSs could be important *Rht-1* regulatory regions. The *Rht-1*-containing BAC sequences enable the creation of genome-specific primers for examining genetic diversity of the *Rht-1* ORFs and flanking 5' and

3’ regions among wheat accessions, which could provide insight into function and regulation of *Rht-1* and may lead to the discovery of agronomically beneficial alleles.

Genetic diversity of the *Rht-1* loci and the surrounding regions is of particular interest not only due to the important biological and adaptive roles of *Rht-1*, but also because this could aid in identification of useful variants of linked loci. One tightly linked gene is wheat *Teosinte branched 1 (TaTb1;* Duan *et al.* 2012). *Tb1* is a key domestication gene in maize (*Zea mays*) that controls branch number (Doebley 2004) and the barley (*Hordeum vulgare*) ortholog *Intermedium-C* has been associated with changes in tillering and lateral spikelet fertility (Ramsay *et al.* 2011). Another example demonstrating the need for increased knowledge of *Rht-1* and the flanking region is the association of *Rht-B1b* and *Rht-D1b* with changes in susceptibility to the destructive wheat disease Fusarium Head Blight (FHB; Hilton *et al.* 1999; Draeger *et al.* 2007; Srinivasachary *et al.* 2008, 2009; Buerstmayr *et al.* 2012). FHB can reduce yield in wheat and can contaminate the grains with mycotoxins harmful to animals and humans (Gilbert and Tekauz 2000). The association of *Rht-B1b* and *Rht-D1b* with changes in FHB susceptibility may result from an effect of height *per se* (Yan *et al.* 2011). Alternatively, this association may result from a pleiotropic effect; such as an altered cell death response related to changes in the DELLA protein (Saville *et al*. 2011) or, in the case of *Rht-D1b*, may be caused by linkage to a deleterious gene (Draeger *et al.* 2007; Srinivasachary *et al.* 2009).

In this study we determined haplotype diversity of the three *Rht-1* homoeoloci by developing locus-specific primers that were used to sequence the ORFs, 1.7 kb of the 5’ flanking region, and 0.5 kb of the flanking 3’ region (flanking regions referred hereafter as “+f”) of *Rht-A1*, *Rht-B1*, and *Rht-D1* in bread wheat lines and in lines that originate from tetraploid and diploid wheat. We show that among the three homoeoloci, the *Rht-B1*+f sequences contained the greatest nucleotide and haplotype diversity. Quantitative real-time PCR is used to assess whether two insertions upstream of *Rht-B1* affect gene

expression. *Rht-1*+f diversity is discussed in relation to other genes in wheat and in relation to *Rht-1* orthologs from the *Poaceae* family.

MATERIALS AND METHODS

Plant Materials

The bread wheat 1 set (BW1) is composed of 21 natural hexaploid wheat varieties and in each variety the three homoeologous *Rht-1*+f regions were sequenced. Table 1 shows sequenced accessions and seed sources. The BW1 accessions include a subset of twelve varieties that have been grown widely in the United Kingdom (BW1-UK), seven varieties ('Fultz', 'Gaines', 'Kanred', 'Norin 10', 'Norin 10/Brevor-14', 'Siete Cerros', and 'Sonora 64') associated with the origin of 'Norin 10' and early spread of *Rht-B1b* and *Rht-D1b*, one variety containing *Rht-B1e* ('Krasnodari 1'), and 'CS'.

Accessions that composed the Bread Wheat 2 set (Table 1; set BW2) were chosen from the INRA worldwide bread wheat core collection of 372 accessions (372CC; Balfourier *et al.* 2007) using chromosome 4 simple sequence repeat and Diversity Arrays Technology marker scores (F. Balfourier, unpublished) and *Rht-B1* and *Rht-D1* genotype scores (described below). To determine which accessions to select for sequencing the *Rht-A1*+f region, the 16 accessions that contained the greatest combined number of 4A alleles were chosen using the line selection feature, simulated annealing method (N = 1000) of Powermarker v2.5 (Liu and Muse 2005). This procedure was also used to choose accessions for sequencing of the *Rht-B1*+f and *Rht-D1*+f regions using 4B and 4D markers, respectively, with priority placed on markers linked to *Rht-1*. Using these criteria, generally only a single *Rht-1*+f region was selected and sequenced per 372CC accession, with the exceptions of INRA-23996 (*Rht-A1*+f and *Rht-B1*+f sequenced) and INRA-13812 (*Rht-A1*+f, *Rht-B1*+f, and *Rht-D1*+f sequenced). The BW2 accessions are natural *T. aestivum* hexaploids with the following exceptions: INRA-13812 is a synthetic line ('W7984') composed of *T. durum* (AABB genomes) variety 'Altar 84' and *A. tauschii* (DD genomes); INRA-03485 has a

pedigree of *T. turgidum* (AABB genomes) / *A. ventricosa* (DDNN genomes) // *T. aestivum* (AABBDD genomes), hence the sequenced *Rht-A1*+f region may originate from *T. aestivum* or *T. turgidum*.

The tetraploid and diploid wheat set (Table 1; set TDW) consisted of four accessions: *T. urartu* (AA genomes), *T. dicoccoides* 57 (AABB genomes), *T. dicoccoides* 65 (AABB genomes), and the synthetic wheat 'SS7010073' (AABBDD genomes). 'SS7010073' was developed from a cross between the tetraploid *T. dicoccum* (John Innes Centre (JIC) acc. no. 1070026) and the diploid *A. tauschii* (JIC acc. no. 2220053) performed by E. Sears, University of Missouri (Columbia, Missouri, USA) (S. Reader, JIC, pers. comm.). We sequenced *Rht-1* nucleotide coordinates (NCs) -500 to 2300 (negative NCs refer to sequence 5' of the *Rht-1* start nucleotide, NC 1) in the A, B, and D genomes of these accessions and confirmed that the 1070026 and 2220053 sequences perfectly matched 'SS7010073' in the respective genomes.

BW1 and TDW accessions were grown individually in 1 l pots in the glasshouse at the JIC, Norwich, UK. BW2 accessions were grown outdoors at the National Institute of Agricultural Botany (NIAB), Cambridge, UK in 4 l pots with four plants per pot as part of a larger study that examined the entire 372CC. Growth conditions for all plants are described in Online Resource 1. Mature plant height was measured as the distance from the soil surface to the tip of the longest tiller present in each pot, excluding awns.

DNA extraction

For all three sets, leaf tissue was collected from two to three week old plants and DNA extracted using a modification of the method described by Fulton *et al*. (1995). For the BW1 and TDW sets, a single bulk of one to four plants was used in DNA extractions. For the BW2 set, two replicates of four plants each were extracted with one extraction used for forward sequencing reads and the other for reverse reads to further minimize the risk of sequencing errors.

Multiplex PCR assay to detect the *Rht-B1* 160 bp and 197 bp insertions

A multiplex PCR-based assay was developed to determine the presence or absence of the 160 bp and 197 bp *Rht-B1* insertions using the forward primer Rht-F11 and reverse primers 160-R1, 197-R1 and Rht-ABD-R9. Primer sequences are shown in Online Resource 2. PCR reactions comprised a 10 µl reaction mix of 1 × Green GoTaq Reaction Buffer (Promega), 3% glycerol, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 µM forward primer, 0.33 µM of each reverse primer, 0.125 µl Taq Polymerase, and 20 ng of DNA template. The PCR profile consisted of 95°C for 5 min, followed by 40 cycles of [95°C for 30 s; an annealing of 60°C for 30 s, and extension at 72°C for 1 min], followed by 5 min at 72°C. Amplified products were separated in a 1.5% agarose gel in 1 × TBE buffer and visualized under UV light with ethidium bromide. Primers Rht-F11 and Rht-ABD-R9 flank the insertions and in lines without an insertion amplify a 1050 bp product (Online Resource 3). Primer 160-R1 lies within the 160 bp insertion and amplifies a 449 bp product in lines with the 160 bp insertion. Primer 197-R1 lies within the 197 bp insertion and amplifies a 361 bp product in lines with this insertion. The *Rht-B1* insertion assay was able to detect the three heterozygous classes.

Assays to distinguish the *Rht-B1a*, *Rht-B1b*, *Rht-D1a*, and *Rht-D1b* alleles are described in Online Resource 1.

PCR amplification and sequencing of *Rht-1* and the flanking region

Sequences from each accession were generated using locus-specific products amplified from genomic DNA (gDNA) of seedling tissue with the exceptions of the three homoeologous 'CS' *Rht-1*+f sequences (all with 8 × coverage) and the *T. urartu Rht-A1*+f sequence, which were derived from BAC sequences that are complete across the *Rht-1*+f regions (Genbank acc. nos. JX978692 to JX978695; Wilhelm *et al.* 2013). Primer pairs used to amplify the *Rht-1*+f regions were designed using primer3 software (http://frodo.wi.mit.edu/primer3). PCR conditions were as described for the *Rht-B1* insertion assay using the primer concentrations, extension times, and annealing temperatures shown in Online Resource 4. Prior to sequencing,

primers and dNTPs were removed from PCR products by adding 1 × Exonuclease I buffer, 0.75 U Exonuclease I (NEB Biolabs) and 0.25 U shrimp alkaline phosphatase (Promega) to 7.5 µl of PCR product and incubating at 37°C for 30 min and then deactivating enzymes by heating to 80°C for 15 min. Sequencing reactions were performed in 10 µl Big Dye (Applied Biosystems; ABI) sequencing mixes that contained 1 × BigDye Sequence Buffer, 1 µl BigDye (ver. 3.1), 5% dimethyl sulfoxide, 1 µM of a sequencing primer, and 2 µl of a PCR product. The sequencing primer and PCR product used in each reaction are shown in Online Resource 4. PCR profiles consisted of 98°C for 1 min, followed by 25 cycles of 98°C for 10 s, 50°C for 5 s, and 60°C for 4 min. Sequenced products were precipitated using the ABI BigDye v3.1 Cycle Sequencing Kit ethanol/EDTA/sodium acetate method according to manufacturer specifications and suspended in a final volume of 10 µl Hi-Di (ABI) formamide. Sequencing was performed using an ABI 3730 DNA Analyzer and bases called with ABI Sequencing Analysis Software, v5.1. Each of the sequences generated from gDNA has 2 × coverage (one forward and one reverse read) with the exceptions of the following *Rht-D1*+f regions that have 1 × coverage: 'CS'-D NCs -1416 to -1063 & -1033 to -968 in accession 'SS7010073' and 'CS'-D NCs -1128 to -1063 & -1033 to -968 from the remaining 36 *Rht-D1*+f sequences arising from gDNA.

Diversity analysis and bioinformatics

Nucleotide contigs were assembled using the ContigExpress package of Vector NTI (Invitrogen) and assembled sequences aligned using ClustalX (Larkin *et al.* 2007). AA alignments were performed using GeneDoc v2.6.002 software (Nicholas and Nicholas, 1997). Nucleotide diversity per site (π) (Tajima 1983) and Watterson's theta per site (θ) (Nei 1987) were calculated using DnaSP v5 software (Librado and Rozas 2009) on aligned sequences. Haplotype diversity (HD) was calculated using the following formula: $HD = n(1-\Sigma f^2)(n-1)^{-1}$, where n is the sample size and f is the frequency of each haplotype (Nei 1987).

The nucleotide sequences of the 197 and 160 bp insertions served as query in BLASTn searches of the National Center for Biotechnology Information (NCBI) nucleotide collection (nr/nt) (http://blast.ncbi.nlm.nih.gov) and the Triticeae Repeat, nonredundant database (TREP) (http://wheat.pw.usda.gov/GG2/blast).

Rht-1 transcript analysis

Seeds of each accession were surface sterilized and stratified (4°C, two days) on wetted filter paper before being transferred to a controlled environment room (22°C 16/8 hr light/dark) for five days. Three biological replicates were utilized for each accession except 'Cadenza' and 'SS7010073', which had two replicates. For each replicate, shoot and leaf tissues of ten seedlings were collected, pooled, and frozen in liquid nitrogen and ground to a fine powder with a TissueLyser LT (Qiagen) for 30 s. RNA was extracted using the RNeasy Plant Mini Kit (Qiagen). A total of 1 µg of RNA was used to synthesize cDNA using SuperScript III (Invitrogen) and random nonamers (50 µM; Invitrogen) according to the manufacturer's instructions. RNA was removed from the RNA-cDNA duplex using RNase-H (Invitrogen). The resulting cDNA was diluted 1:20 with nuclease-free water prior to quantitative real time PCR (qRT-PCR) performed using a DNA engine Opticon2 Contiguous Fluorescence Detector (MJ Research Inc., Alameda, CA, USA). Primer sequences (Online Resource 2) and reaction conditions are as described in Pearce *et al*. (2011). The internal controls for normalization of expression were Glyceraldehyde 3-phosphate dehydrogenase (GAPDH; McGrann *et al.* 2009) and Elongation factor 1α (EF1α; Coram *et al.* 2008), (Online Resource 2). Target gene expression was calculated relative to that of the normalization factor (a value derived from the geometric mean of the internal control genes using the GeNORM software described by Vandesompele *et al*. (2002)) using the ΔCT method (Pfaffl *et al*. 2001) and corrected for primer efficiencies. Normalized data is presented as the mean of the biological replicates. Statistically significant differences among Rht-B1 transcript levels and among Rht-D1 transcript levels were determined using ANOVA (Genstat, 12th edition) by inputting normalized replicate values.

Genbank accession numbers for the *Rht-1*+f haplotype sequences are shown in Online Resource 5.

RESULTS

Sequenced regions and polymorphic sites

Using three diverse wheat germplasm sets, the *Rht-A1*+f, *Rht-B1*+f, and *Rht-D1*+f regions were sequenced from PCR products amplified from gDNA in 39, 39, and 37 accessions, respectively (Table 1). *Rht-A1*+f sequences consisted of 4122 bp ('CS'-A NCs -1760 to 2362), *Rht-B1*+f sequences consisted of 4494 bp ('CS'-B NCs -1815 to 2322, plus the 160 and 197 bp insertions present in some lines, but not in 'CS'), and *Rht-D1*+f sequences consisted of 4091 bp ('CS'-D NCs -1809 to 2311, minus a sequencing gap from 'CS'-D NCs -1062 to -1034 that is present in all accessions except 'CS') (Fig. 1). The *Rht-D1*+f sequencing gap corresponds to a 29 bp region in 'CS' that contains a 19 bp long poly-C chain. PCR products spanning the *Rht-D1*+f sequence gap that visually matched the 'CS' product length were amplified in each accession, indicating that no large insertions or deletions were present in the amplified region relative to 'CS'. When sequenced, each product terminated in a poly-C chain.

Rht-1+f diversity was first determined in a set of 21 bread wheat accessions (BW1), which included twelve lines that were widely grown in the UK (BW1-UK). In each BW1 accession, the A, B, and D genome *Rht-1*+f regions were sequenced. Mean height measurements were taken from a minimum of five greenhouse-grown plants per accession with the exceptions of 'Alchemy' (no plants) and 'Paragon' (three plants); heights ranged from 46 to 115 cm (Table 1). To explore *Rht-1* diversity in a worldwide collection of bread wheat accessions, the BW2 set was selected from the 372CC using group 4 markers in an attempt to maximize diversity. The BW2 set contained 16 *Rht-A1*+f, 16 *Rht-B1*+f, and 16 *Rht-D1*+f sequences, which were derived from 45 accessions (generally, a single *Rht-1*+f homoeolog was sequenced per

accession). Mean plant heights ranged from 54 to 189 cm. *Rht-1*+f sequence diversity was also examined in a set of tetraploid and diploid wheat (TDW) accessions. Mean heights taken from two to six plants per accession (with the exception of *T. urartu*, which was not grown) ranged from 69 to 87 cm. Over all three sets, there were 74 polymorphic sites (PSs) identified among the *Rht-A1*+f sequences, 49 PSs among the *Rht-B1*+f sequences, and 34 PSs among the *Rht-D1*+f sequences (Table 2; Fig. 1). The percentage of sites polymorphic in the *Rht-1* ORF was 0.48% (27 PSs in 5601 bp) whereas 2.00% of the 5' sites were polymorphic (114 PSs in 5712 bp) and 1.15% of the 3' sites were polymorphic (16 PSs in 1394 bp).

Rht-A1+f diversity

Among the 37 *Rht-A1*+f bread wheat sequences (BW1+BW2), there were nine haplotypes (*Rht-A1a*_1 to *Rht-A1a*_9) and 25 PSs with π of 0.49×10^{-3} and HD of 0.43 (Table 2; Online Resource 6). The BW2 set had π and HD values approximately double that of the BW1 set. There were no PSs present among the 12 UK accessions, which all belonged to haplotype (hap) *Rht-A1a*_2. Hap *Rht-A1a*_2 was present in 28 of the 37 accessions while the remaining eight haplotypes were represented by only one or two accessions (Fig. 2a). 'CS' contained hap *Rht-A1a*_1, which contains a T nucleotide deletion at 'CS'-A NC -1046 relative to the other haplotypes. To rule out the possibility that the -1046 T indel was due to an error in the 'CS' BAC sequence, DNA from leaf tissue of 'CS' seedlings was amplified from this region and sequenced. A forward and reverse sequencing read of the amplified DNA confirmed the absence of the T base in 'CS'. 'CS' differed from the majority of the bread wheat haplotypes by three or fewer PSs, but differed from hap *Rht-A1a*_4 by 17 PSs and hap *Rht-A1a*_9 by 20 PSs (Online Resource 6). The hap *Rht-A1a*_4 and hap *Rht-A1a*_9 PSs were all upstream of the *Rht-1* ORF and these two haplotypes were closely related, differing by only three PSs. Among the bread wheat PSs, there were three predicted AA changes, which all occur in conserved protein domains. Hap *Rht-A1a*_3 contains an S477Y substitution (AA changes indicate the 'CS'

residue followed by the AA number and the substitution) in the PFYRE domain (Online Resource 7, AA alignment position (pos.) 483). Hap *Rht-A1a*_5 contains an S189T substitution in the middle position of a nine-residue poly S string in the Poly S/T/V domain (Online Resource 7, pos. 194). Hap *Rht-A1a*_7 contains a G332S substitution in the VHIID domain (Online Resource 7, pos. 338).

Among the four TDW *Rht-A1*+f sequences, there were four haplotypes, 64 PSs, π was 5.94×10^{-3}, and HD was 1.00 (Table 2; Online Resource 6). Of the PSs, 49 were not present in the BW accessions. Five of the 64 PSs found among the TDW sequences are located in the *Rht-A1* ORF and none result in a predicted AA change. The TDW *Rht-A1*+f sequence most similar to 'CS' is 'SS7010073' (hap *Rht-A1a*_10; A genome from *T. dicoccum*), which differs from 'CS' by 8 PSs. *T. dicoccoides* 57 (hap *Rht-A1a*_11) and *T. dicoccoides* 65 (hap *Rht-A1a*_12) differ from 'CS' by 19 and 16 PSs, respectively. Hap *Rht-A1a*_11 is closely related to bread wheat haps *Rht-A1a*_4 and *Rht-A1a*_9, differing by four PSs and one PS, respectively. Of all the A genome sequences, *T. urartu* (hap *Rht-A1a*_13) contained the largest number of PSs (53) relative to 'CS' and 39 of these are unique to *T. urartu*.

Rht-B1+f diversity

The 37 *Rht-B1*+f bread wheat sequences (BW1+BW2) contained 23 PSs, 13 haplotypes (*Rht-B1a*_1 to *Rht-B1a*_11; *Rht-B1b*_1; *Rht-B1e*_1) with π of 0.86×10^{-3} and HD of 0.89 (Table 2; Online Resource 8). Hap *Rht-B1a*_1, which included 'CS', was the most frequent haplotype occurring in nine accessions (Fig. 2b). Four other haplotypes were present in four or more accessions. The polymorphisms included insertions of 160 bp and 197 bp, a frameshift mutation, two nonsense changes, and three missense substitutions. The two nonsense changes were associated with the *Rht-B1b* and *Rht-B1e* semi-dwarf alleles, as previously reported (Peng *et al.* 1999; Pearce *et al.* 2011; Li *et al*. 2012b). The six accessions with the *Rht-B1b* allele contain hap *Rht-B1b*_1, which is identical to hap *Rht-B1a*_1 except for the 'CS'-B NC 190 SNP that results in the premature stop codon. The *Rht-B1e* allele was only present

in ‘Krasnodari 1’, and this haplotype (*Rht-B1e*_1) contained two PSs in addition to the ‘CS’-B NC 181 SNP that results in the premature stop codon. The 160 bp insertion occurred at ‘CS’-B NC -356 and was present in three haplotypes (*Rht-B1a*_2, *Rht-B1a*_3, and *Rht-B1a*_4) that also contained an E205G substitution in the poly S/T/V domain (Online Resource 7, pos. 209) and zero, one, or two additional PSs. The 197 bp insertion was present at ‘CS’-B NC -591 in haps *Rht-B1a*_5 and *Rht-B1a*_6. Both haplotypes also contained G15R and M25I substitutions that occur outside of the conserved protein domains (Online Resource 7, pos. 15 and 25). The G15R and M25I substitutions were previously identified in the cultivar ‘Tom Thumb’, which contains the *Rht-B1c* allele (Pearce *et al.* 2011; Wu *et al.* 2011). Haps *Rht-B1a*_5 and *Rht-B1a*_6 also contained the largest number of PSs (14 and 15, respectively) relative to ‘CS’ whereas the other bread wheat haplotypes contained one to four PSs. The frameshift mutation results from a T insertion that occurs before ‘CS’ NC 984 (AA 328) at the beginning of the C-terminus (Online Resource 7, pos. 333). The resulting protein is predicted to contain 633 AAs, ending at ‘CS’-B NC 1898. The frameshift is present in INRA-23995 (hap *Rht-B1a*_10), which contains no additional PSs relative to ‘CS’. Relative to the BW2 set, the BW1 *Rht-B1*+f sequences contained more haplotypes and PSs and had greater π and HD values. The BW1-UK subset contained six haplotypes and 18 of the 23 PSs present in the combined BW1 and BW2 sets, including both of the large indels and four AA changes. The π and HD values of the BW1-UK subset were also similar to the values of the combined BW1 and BW2 sets.

The three TDW accessions each represented new haplotypes (*Rht-B1a*_12; *Rht-B1a*_13; *Rht-B1a*_14) and contained 31 PSs with π of 4.15×10^{-3} and HD of 1.00. Of the 31 PSs, 27 were not present in the bread wheat accessions. Relative to ‘CS’, *T. dicoccoides* 57 (hap *Rht-B1a*_14) and *T. dicoccoides* 65 (hap *Rht-B1a*_13) contained 13 and 15 PSs, respectively. ‘SS7010073’ (hap *Rht-B1a*_12), a synthetic line with B genome derived from *T. dicoccum*, contained 23 PSs relative to ‘CS’. Each TDW haplotype contained the 197 bp insertion and there were four SNPs in the insertion that differentiated the haplotypes. The 160 bp insertion was not present in any TDW accession.

Hap *Rht-B1a*_12 also contains a 16 bp deletion at 'CS'-B NC -694. Six SNPs occurred in the *Rht-1* ORF of the TDW accessions with one resulting in an amino acid change (M25I), which was also present in haps *Rht-B1a*_5 and *Rht-B1a*_6.

Rht-D1+f diversity

The 37 *Rht-D1*+f sequences from the bread wheat (BW1+BW2) accessions contained seven PSs, seven haplotypes, π of 0.24, and HD of 0.76 (Table 2; Online Resource 9). Hap *Rht-D1a*_1 is the most common haplotype, occurring in 14 accessions including 'CS' (Fig. 2c). The other haplotypes differ from 'CS' by only one or two PSs. The 11 accessions containing *Rht-D1b* belong to hap *Rht-D1b*_1, which differs from hap *Rht-D1a*_1 by only the 'CS'-D NC 181 SNP that results in a premature stop codon, previously reported by Peng *et al.* (1999). Hap *Rht-D1a*_5, present in INRA-13812 (a synthetic line with D genome derived from *A. tauschii*), contains a G334S substitution relative to 'CS'. The residue change occurs in the VHIID domain and at the equivalent residue as the G332S substitution in hap *Rht-A1a*_7 (Online Resource 7, pos. 338). Comparisons of the *Rht-D1*+f sequences among the bread wheat sets and subsets revealed that BW1 and BW1-UK had slightly higher π and HD than BW2 and contained similar numbers of haplotypes and PSs.

An additional *Rht-D1*+f region derived from an *A. tauschii* genome was sequenced from the synthetic hexaploid 'SS7010073', the lone accession in the D genome TDW set. 'SS7010073' contains hap *Rht-D1a*_7, which differs from 'CS' by 28 PSs. Relative to 'CS', four consecutive nucleotides ('CS'-D NCs 483 to 486) are altered in hap *Rht-D1a*_7 with the latter three resulting in a T162V substitution (Online Resource 7, pos. 165). A G622A substitution (Online Resource 7, pos. 626) is also present in hap *Rht-D1a*_7. Both AA substitutions occur outside of conserved domains. The two substituted AAs in hap *Rht-D1a*_7 match the AAs present in corresponding positions in *Rht-A1* and *Rht-B1* in all of the wheat accessions examined herein.

Effect of the 160 bp and 197 bp insertions on *Rht-B1* expression

To determine if the 160 and 197 bp *Rht-B1* insertions affected Rht-B1 expression, Rht-B1 and Rht-D1 transcript abundance was measured with qRT-PCR using seedling tissue collected from the following genotypes and varieties: no insertion ('CS' and 'Cadenza'); 160 bp insertion ('Mercia' and 'Paragon'); 197 bp insertion ('Kanred', 'Cappelle Desprez', and 'SS7010073'). 'SS7010073' also carries a 16 bp deletion 103 bp upstream of the 197 bp insertion. Transcript abundance was normalized relative to the geometric mean of GAPDH and EF1α. Of the five lines with the *Rht-B1* insertions, only 'Kanred' shows a statistically significant difference ($p < 0.05$) in normalized Rht-B1 transcript amount relative to 'CS' or 'Cadenza' (Fig. 3). Normalized 'Kanred' Rht-B1 transcript levels were 19% that of the mean of 'CS' and 'Cadenza'. The normalized Rht-B1 transcript levels of the remaining four lines were slightly reduced (ranging from 85% to 94% of the mean of 'CS' and 'Cadenza'), but were not significantly different from either line at a probability threshold of 0.05. To determine if the *Rht-B1* insertions differentially affected the expression levels of Rht-B1 relative to other Rht-1 homoeologs, Rht-D1 transcript levels were measured. Similar to the Rht-B1 results, 'Kanred' showed a significant reduction ($p < 0.05$) in transcript level after normalization relative to 'CS' or 'Cadenza'. Normalized 'Kanred' transcript levels were 15% the mean of 'CS' and 'Cadenza'. The normalized 'Cappelle-Desprez' Rht-D1 transcript level was also significantly reduced ($p < 0.05$), having a 69% reduction relative to the mean of 'Cadenza' and 'CS'. No other lines showed a significant reduction in normalized Rht-D1 transcript abundance at $p \leq 0.05$. Although 'Kanred' has statistically significant reductions in Rht-B1 and Rht-D1 transcript amounts following normalization, 'Kanred' transcript amounts were the highest among the lines prior to normalization. Relatively high transcript levels of both normalization genes in 'Kanred' relative to the other accessions resulted in 'Kanred' normalized Rht-B1 and Rht-D1 transcript levels that were reduced relative to the other accessions.

DISCUSSION

Rht-1+f genetic diversity in bread wheat

The two bread wheat sets comprised 37 sequences from each *Rht-1*+f region, and contained a total of 55 PSs (42 SNPs and 13 indels) with π of 0.53 x 10^{-3} when averaged across the three genomes. The homoeologous *Rht-1*+f regions total 12,707 bp for an average of one PS per 231 bp and one SNP per 303 bp. The *Rht-1*+f SNP frequency is in the range previously reported for bread wheat genes, which were as follows: one SNP per 91 bp for the *Spa* homoeologs from 42 accessions (Ravel *et al.* 2009), one SNP per 212 bp for 21 genes among 42 accessions (Ravel *et al.* 2006), one SNP per 441 bp for the *GAMYB* homoeologs from 42 accessions (Haseneyer *et al.* 2008). In a study of 21 loci from 41 bread wheat accessions, one PS was present for every 362 bp (Haudry *et al.* 2007), which is similar to *Rht-1*+f. π of the *Rht-1*+f regions was slightly reduced relative to other bread wheat genes, which had the following π values: 1.83 × 10^{-3} (Ravel *et al.* 2009); 0.9 × 10^{-3} (Ravel *et al.* 2006); 0.83 × 10^{-3} (Haudry *et al.* 2007). The *Rht-1* ORF had a reduced frequency of PSs relative to non-coding sequence, a pattern previously reported in bread wheat (Ravel *et al.* 2006; Haseneyer *et al.* 2008; Ravel *et al.* 2009).

Nucleotide diversity of the bread wheat *Rht-1* ORF (0.41 × 10^{-3}) is greatly reduced relative to orthologs in the *Poaceae* family. Among 92 diverse maize inbreds, π of the *Rht-1* ortholog *Dwarf8* was 1.8 x 10^{-3} (Thornsberry *et al.* 2001). Partial length sequences of *Rht-1* orthologs from 26 *Sorghum bicolor* inbreds and 20 *Pennisetum glaucum* inbreds from West and Central Africa had π values of 1.63 × 10^{-3} and 7.04 × 10^{-3}, respectively (Li *et al.* 2010). In *Eragrostis tef*, π values of the *rht1-1* and *rht1-2* homologs were 3.82 x 10^{-3} and 5.75 x 10^{-3}, respectively (Smith *et al.* 2012). Although the bread wheat gene pool is narrow (Feuillet *et al.* 2008), the reduced diversity of *Rht-1* relative to other bread wheat genes indicates there has been selection at or near *Rht-1*. Intense selection for the semi-dwarf *Rht-B1b* and *Rht-D1b* alleles over the last fifty years may be partially responsible for decreasing diversity at these loci. Selection may also be occurring at tightly linked loci such as

TaTb1 (Duan *et al.* 2012) or a tightly linked gene affecting FHB resistance (Srinivasachary *et al.* 2009).

Among the *Rht-1*+f bread wheat sequences, the B genome contained the greatest π (0.86×10^{-3}), the greatest HD (0.89), the most haplotypes (13), and an intermediate number of PSs (23). *Rht-B1*+f PSs included the two largest indels and the only frameshift mutation. High diversity of the bread wheat B genome relative to the A and D genomes has been reported in several studies (Huang *et al.* 2002; Ravel *et al.* 2006; Wang *et al.* 2007; Haseneyer *et al.* 2008; Li *et al.* 2011). The bread wheat *Rht-D1*+f sequences contained the lowest π (0.24×10^{-3}), an intermediate HD (0.76), the fewest haplotypes (7), and the lowest number of PSs (7). No bread wheat *Rht-D1*+f haplotypes differed by more than four polymorphisms and two haplotypes represent 68% of the sequences. The lack of diversity on the D genome relative to the other bread wheat genomes agrees with the consensus of other studies (Bryan *et al.* 1997; Huang *et al.* 2002; Wang *et al.* 2007; White *et al.* 2008; Chao *et al.* 2009). The bread wheat *Rht-A1*+f sequences have intermediate π (0.49×10^{-3}), an intermediate number of haplotypes (9), the lowest HD (0.43), and the highest number of PSs (25). Low HD results from the predominance of hap *Rht-A1a*_2, which is present in 76% of the bread wheat accessions.

There was little difference in π or HD among the BW1, BW1-UK, and BW2 sets in regards to the *Rht-B1*+f sequences and the *Rht-D1*+f sequences. This is surprising considering the geographically narrow range of germplasm selected for the BW1 and BW1-UK sets relative to the BW2 worldwide diversity set. In contrast, the BW2 *Rht-A1*+f region had π and HD values nearly double those of BW1 and BW1-UK. A previous study of UK bread wheats using genome-wide markers reported significantly higher genetic diversity on the A genome relative to the B and D genomes (White *et al.* 2008), suggesting that the lack of *Rht-A1*+f diversity in the UK wheats may be specific to this locus. One possibility, among many, is that *Rht-A1* may be linked to additional traits that have been under strong selection pressure in the UK. The recent availability of the *Rht-A1* genetic map position (Wilhelm *et al.* 2013) should aid in identifying any linked traits.

Rht-1+f genetic diversity in wild relatives of wheat

The TDW set had several fold greater π and contained more PSs than the bread wheat sets. Between *T. dicoccoides* ($\pi = 3.20 \times 10^{-3}$) and the bread wheat *Rht-A1*+f and *Rht-B1*+f sequences ($\pi = 0.7 \times 10^{-3}$) there is a 78% loss in nucleotide diversity, which is similar to the 69% loss reported between these two species by Haudry *et al*. (2007). *T. urartu* contains the greatest number of PSs relative to the bread wheat A genome, which agrees with prior work showing *T. urartu* is more distantly related to bread wheat than *T. dicoccoides* (Dvorak and Akhunov 2005). Only two *Rht-D1*+f sequences derived from *A. tauschii* were examined (those present in 'SS7010073' and INRA-13812) and these differed by 29 PSs, five-fold more PSs than were present among the 36 bread wheat *Rht-D1*+f sequences. Similarly, Caldwell *et al*. (2004) reported a 30-fold reduction in bread wheat genetic diversity relative to *A. tauschii* in *granule-bound starch synthase*. Relative to 'CS', the *Rht-D1*+f sequence from 'SS7010073' contained 28 PSs and the INRA-13812 *Rht-D1*+f sequence contained just two PSs. A high degree of similarity between the D genomes of INRA-13812 and bread wheat was previously reported (Ravel *et al.* 2006; Haseneyer *et al.* 2008). *A. tauschii* ecotype AL8/78 (Duan *et al.* 2012) differed from 'CS' by eight PSs in the *Rht-D1*+f region and contained six PSs (relative to 'CS') in common with 'SS7010073'. All of the TDW haplotypes and the bread wheat haplotypes *Rht-A1a*_4, *Rht-A1a*_9, *Rht-B1a*_5, and *Rht-B1a*_6 differed markedly from the majority of the bread wheat haplotypes, and represent candidates that can be explored for useful variation at *Rht-1* and nearby loci.

Rht-1+f polymorphisms

Among the PSs were indels of 160 bp and 197 bp that were 5' of *Rht-B1* and within 600 bp of the start codon. The 197 bp insertion was present in the *Rht-B1*+f region of five bread wheat lines and all three TDW accessions. Sequence homologous to the 197 bp insertion was also present in collinear regions of *Rht-A1* and *Rht-D1* in all wheat sequences examined (83% and

85% of the 197 bases were shared with the collinear 'CS' *Rht-A1*+f and 'CS' *Rht-D1*+f sequences, respectively). In addition, the 197 bases closely matched (73% identity over 180 bp) the collinear region in barley, which is approximately 1.5 kb upstream of the *Rht-1* ortholog *Slender1*. These results indicate the presence of the *Rht-B1* 197 bp insertion is the ancestral condition and that a deletion occurred in a progenitor of the bread wheat accessions lacking this sequence. No sequence similar to the 197 bases is present in the TREP database indicating the sequence is likely not a repetitive element. The *Rht-B1* 160 bp insertion is present in eight bread wheat lines, but is not present among the TDW *Rht-B1* sequences or in collinear regions of *Rht-A1* or *Rht-D1* in any accession. These results indicate the 160 bases are an insertion relative to the ancestral condition. No sequence similar to the 160 bases was present in the NCBI nucleotide collection or TREP database, indicating the insertion is not well conserved outside of wheat and is not likely to be a repetitive element. Interestingly, the 160 bases are inserted in the middle of a CNS that is highly conserved among several *Poaceae* members (Duan *et al.* 2012; Wilhelm *et al.* 2013), suggesting the possibility that a *cis*-regulatory region may be disrupted by the insertion.

Changes in normalized Rht-B1 transcript abundance were not clearly associated with the 160 or 197 bp insertion. Significant ($p < 0.05$) reductions in normalized Rht-B1 transcript levels relative to lines without an insertion were present in 'Kanred', which contains the 197 bp insertion; however, significant reductions in Rht-B1 transcript amount did not occur in the remaining accessions with the 197 bp insertion. Normalized 'Kanred' Rht-D1 transcript levels were also significantly ($p < 0.05$) reduced although no Rht-D1 insertion was present. Together, these results indicate that the 197 bp insertion is not likely to be the primary cause of the reduced normalized Rht-B1 expression levels in 'Kanred'. Slight reductions (6-15%) in Rht-B1 transcript abundance in each of the four remaining cultivars containing an *Rht-B1* insertion (two with the 197 bp insertion and two with the 160 bp insertion) relative to lines without an insertion, although non-significant, leave open the possibility that the insertions may have a minor effect on Rht-B1 transcript abundance in seedling tissue. The absence of a signficant effect of

the Rht-B1 insertions could also relate to the age of tissue (five days) and tissue type (leaf and shoot) examined. Rht-1 expression patterns in bread wheat have previously been shown to differ based on tissue type and developmental stage (Pearce *et al.* 2011). Expression levels of the barley and rice (*Oryza sativa*) Rht-1 orthologs (both termed Slender1) were also found to differ based on tissue type (Chandler *et al.* 2002; Kaneko *et al.* 2003). To more fully determine whether the *Rht-B1* insertions affect Rht-B1 expression, it will be necessary to analyze more tissues and to sample at multiple developmental stages.

In the *Rht-1* ORFs, thirteen mutations (three nonsense, a frameshift, and nine missense) were present that are predicted to alter the AA sequence. The nonsense mutations were associated with *Rht-B1b*, *Rht-D1b*, and *Rht-B1e* (Peng *et al.* 1999; Pearce *et al.* 2011; Li *et al.* 2012b). The absence of genetic variation among *Rht-B1b* haplotypes and among *Rht-D1b* haplotypes is likely due to the recent introduction of these alleles (beginning in the 1960s) into Western wheat varieties and the predominant use of a single donor, 'Norin 10' (Gale and Youssefian 1985; Dalrymple 1986). The semi-dwarf causative SNP is the only PS that differentiates the *Rht-B1b* and *Rht-D1b* haplotypes from 'CS'. The close genetic similarity and similar origins of 'CS' (a Chinese landrace) and 'Norin 10' (a Japanese line that may be from Korea; Cho *et al.* 1980) suggest *Rht-B1b* and *Rht-D1b* arose in germplasm closely related to 'CS'.

The *Rht-B1* frameshift mutation in the Russian landrace INRA-23995 is near the beginning of the DELLA protein C-terminus. Mutations that disrupt the C-terminus often lead to a loss of function, which is characterized by a GA-constitutive growth response and a phenotype of elongated and slender stems and leaves (Ikeda *et al.* 2001; Chandler *et al.* 2002). However, due to the buffering effects of *Rht-A1* and *Rht-D1*, only a dominant or semi-dominant mutation is likely to produce an observable phenotype. Plant height of INRA-23995 was similar to accessions that were not semi-dwarf (Table 1), suggesting any height effect is likely weaker than that of *Rht-B1b* or *Rht-D1b*. The frameshift mutation in combination with *Rht-D1b* may lead to an

additional decrease in height by reducing the number of functional wild type copies of DELLA from two to one. Further testing and development of suitable germplasm is required to determine the effect of this allele.

Conclusions

This study provides insight into the nucleotide and haplotype diversity at the homoeologous *Rht-1* loci and flanking sequences, which represents an important first step in the search for useful variation. A lack of diversity was associated with the *Rht-1*+f regions in bread wheat relative to previously examined wheat genes and relative to *Rht-1* orthologs in the *Poaceae* family. Diversity in the *Rht-1*+f regions of the bread wheat A and D genomes was particularly reduced, including the presence of only a single *Rht-A1*+f haplotype in the twelve UK accessions examined. The few sequences derived from wheat ancestral lines contained most of the polymorphisms, and, as previously reported (Feuillet *et al.* 2008), are a rich source of diversity that could be mined for useful variation.

ACKNOWLEDGMENTS

We thank the NIAB trust for funding Ed Wilhelm's research as part of a PhD, with assistance from the Biotechnology and Biological Sciences Research Council. We also thank the NIAB pre-breeding team, the NIAB horticultural staff, and the JIC horticultural staff for their assistance in growing plants, Jizeng Jia (Chinese Academy of Agricultural Sciences) for use of *T. urartu* sequence prior to publication, and Huw Jones (NIAB) for technical assistance with sequencing and critical review of the manuscript.

REFERENCES

Achard P, Cheng H, De Grauwe L, Decat J, Schoutteten H, Moritz T, Van Der Straeten D, Peng JR, Harberd, NP (2006) Integration of plant responses to environmentally activated phytohormonal signals. Science 311:91-94

Achard P, Genschik P (2009) Releasing the brakes of plant growth: how GAs shutdown DELLA proteins. J Exp Bot 60:1085-1092

Alvey L, Boulton MI (2008) DELLA proteins in signalling. Encyclopedia of Life Sciences. John Wiley & Sons, Ltd: Chichester

Alvey L, Harberd NP (2005) DELLA proteins: integrators of multiple plant growth regulatory inputs? Physiologia Plantarum 123:153-160

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M, Charmet G (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. Theor App Gen 114:1265–1275.

Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor Appl Gen 94:557-563

Buerstmayr M, Huber K, Heckmann J, Steiner B, Nelson JC, Buerstmayr H (2012) Mapping of QTL for Fusarium head blight resistance and morphological and developmental traits in three backcross populations derived from *Triticum dicoccum* × *Triticum durum*. Theor Appl Gen 125:1751-1765

Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo M-C, Wolters P, Powell W (2004) Sequence polymorphism in polyploidy wheat and their D-genome diploid ancestor. Genetics 167:941-947

Chandler PM, Marion-Poll A, Ellis M, Gubler F (2002) Mutants at the *Slender1* locus of barley cv Himalaya. Molecular and Physiological Characterization. Plant Physiol 129:181-190

Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo M-C, Dubcovsky J (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. Mol Breeding 23:23-33

Cho CH, Hong BH, Park MW, Shim JW, and Kim BK (1980) Origin, dissemination, and utilization of wheat semi-dwarf genes in Korea. Annual Wheat Newsletter 27:67

Chrpova JS, Skorpik M, Prasilova P, Sip V (2003) Detection of Norin 10 dwarfing genes in winter wheat varieties registered in the Czech Republic. Czech J Genet Plant Breed 39:89–92

Coram TE, Wang MN, Chen XM (2008) Transcriptome analysis of the wheat - *Puccinia striiformis* f. sp *tritici* interaction. Molec Plant Pathol 9:157-169

Dalrymple DG (1986) Development and spread of high-yielding wheat varieties in developing countries. Agency for International Development, Washington, D.C., U.S.A.

Dan L, FangPing Y, ZhongHu H, DaNian Y, XianChun X (2009) Characterization of *Lr34/Yr18*, *Rht-B1b*, *Rht-D1b* genes in CIMMYT wheat cultivars and advanced lines using STS markers. Sci Agric Sin 42:17-27

Doebley JF (2004) The genetics of maize evolution. Ann Rev Gen 38:37-59

Draeger R, Gosman N, Steed A, Chandler E, Thomsett M, Srinivasachary, Schondelmaier J, Buerstmayr H, Lemmens M, Schmolke M, Mesterhazy A, Nicholson P (2007) Identification of QTLs for resistance to Fusarium head blight, DON accumulation and associated traits in the winter wheat variety Arina. Theor Appl Genet 115:617-625

Duan J, Wu J, Liu Y, Xiao J, Zhao G, Gu Y, Jia J, Kong X (2012) New cis-regulatory elements in the *Rht-D1b* locus region of wheat. Funct Integr Genomics 12:489-500

Dvorak J, Akhunov ED (2005) Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploidy evolution in the *Aegilops-Triticum* alliance. Genetics 171:323-332

Ellis MH, Spielmeyer W, Gale KR, Rebetzke GJ, and Richards RA (2002) "Perfect" markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. Theor Appl Gen 105:1038-1042

Febrer M, Wilhelm E, Al-Kaff N, Wright J, Powell W, Bevan, MW, Boulton, MI (2009) Rapid identification of the three homoeologues of the wheat dwarfing gene Rht using a novel PCR-based screen of three-dimensional BAC pools. Genome 52:993-1000

Feuillet C, Langridge P, Waugh R (2008) Cereal breeding takes a walk on the wild side. Trends in Genetics 24:24-32

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207-209

Gale MD, Youssefian S (1985) Dwarfing genes in wheat. Progress in Plant Breeding, Vol. 1, ed. GE Russell, pp 1-35

Gilbert J, Tekauz A (2000) Review: recent developments in research on Fusarium head blight of wheat in Canada. Can J Plant Pathol 22:1-8

Guedira M, Brown-Guedira G, Van Sanford D, Sneller C, Souza E, Marshall D (2010) Distribution of Rht genes in modern and historic winter wheat cultivars from the Eastern and Central USA. Crop Sci 50:1811-1822.

Gulyas, G, Bognar Z, Lang L, Rakszegi M, Bedo Z (2011) Distribution of dwarfing genes (*Rht-B1b* and *Rht-D1b*) in Martonvasar wheat breeding materials. Acta Agronomica Hungarica 59:249-254

Haseneyer G, Ravel C, Dardevet M, Balfourier F, Sourdille P, Charmet G, Brunel D, Sauer S, Geiger HH, Graner A, Stracke S (2008) High level of conservation between genes coding for the GAMYB transcription factor in barley (*Hordeum vulgare* L.) and bread wheat (*Triticum aestivum* L.) collections. Theor Appl Genet 117:321-331

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S, David J (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol Biol Evol 24:1506-1517

Hedden P (2003) The genes of the Green Revolution. Trends in Genet 19:5-9

Hilton AJ, Jenkinson P, Hollins TW, Parry DW (1999) Relationship between cultivar height and severity of Fusarium ear blight in wheat. Plant Path 48:202-208

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. Proc Nat Acad Sci USA 99:8133-8138

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, Kitano H, Koshioka M, Futsuhara Y, Matsuoka M, Yamaguchi J (2001) *Slender rice*, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI*/*RGA*/*RHT*/*D8*. Plant Cell 13:999-1010

Itoh H, Ueguchi-Tanaka M, Sato Y, Ashikari M, Matsuoka M (2002) The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. Plant Cell 14: 57-70

Kaneko M, Itoh H, Inukai Y, Sakamoto T, Ueguchi-Tanaka M, Ashikari M, Matsuoka M (2003) Where do gibberellin biosynthesis and gibberellin signaling occur in rice plants? Plant J 35:104-115

Knopf C, Becker H, Ebmeyer E, Korzun V (2008) Occurrence of three dwarfing *Rht* genes in German winter wheat varieties. Cer Res Comm 36:553-560

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm, A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948

Li Y, Xiao J, Wu J, Duan J, Liu Y, Ye X, Zhang X, Guo X, Gu Y, Zhang L, Jia J, Kong X (2012a) A tandem segmental duplication (TSD) in green revolution gene *Rht-D1b* region underlies plant height variation. New Phytol 196:282-291

Li A, Yang W, Guo X, Liu D, Sun J, Zhang A (2012b) Isolation of a gibberellin-insensitive dwarfing gene, *Rht-B1e*, and development of an allele-specific PCR marker. Mol Breeding 30:1443-1451

Li XP, Zhao, XQ, He X, Zhao GY, Li B, Liu DC, Zhang AM, Zhang XY, Tong YP, Li ZS (2011) Haplotype analysis of the genes encoding glutamine synthetase plastic isoforms and their association with nitrogen-use- and yield-related traits in bread wheat. New Phytologist 189:449-458

Li Y, Bhosale S, Haussmann BIG, Stich B, Melchinger AE, Parzies HK (2010) Genetic diversity and linkage disequilibrium of two homologous genes to maize *D8*: Sorghum *SbD8* and pearl millet *PgD8*. J Plant Breed and Crop Sci 2:117-128

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452

Liu K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21:2128-2129

McGrann GRD, Townsend BJ, Antoniw JF, Asher MJC, Mutasa-Goettgens ES (2009) Barley elicits a similar early basal defence response during host and non-host interactions with Polymyxa root parasites. European J of Plant Pathol 123:5-15

Nei, M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York

Nicholas KB, Nicholas HB (1997) GeneDoc: a tool for editing and annotation multiple sequence alignments. Distributed by the author (www.psc.edu/biomed/genedoc)

Pearce S, Saville R, Vaughan SP, Chandler PM, Wilhelm EP, Sparks CA, Al-Kaff N, Korolev A, Boulton MI, Phillips AL, Hedden P, Nicholson P, Thomas SG (2011) Molecular characterisation of *Rht-1* dwarfing genes in hexaploid wheat. Plant Physiol 157:1820-1831

Peng JR, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256-261

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29(9):e45

Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WT, Macaulay M, MacKenzie K, Simpson C, Fuller J, Bonar N, Hayes PM, Lundqvist U, Franckowiak JD, Close TJ, Muehlbauer GJ, Waugh R (2011) *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. Nature Genet 43:169-173

Ravel C, Martre P, Romeuf I, Dardevet M, El-Malki R, Bordes J, Duchateau N, Brunel D, Balfourier F, Charmet G (2009) Polymorphism in the wheat

transcriptional activator *Spa* influences its pattern of expression and has pleiotropic effects on grain protein composition, dough viscoelasticity, and grain hardness. Plant Physiol 151:2133-2144

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour F, Chalhoub B, Brunel D, Béckert M, Charmet G (2006) Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.) Genome 49:1131-1139

Smith SM, Yuan Y, Doust AN, Bennetzen JL (2012) Haplotype analysis and linkage disequilibrium at five loci in *Eragrostis tef*. Genes Genomes Genet 2:407-419

Srinivasachary, Gosman N, Steed A, Hollins TW, Bayles R, Jennings P, Nicholson P. (2009) Semi-dwarfing *Rht-B1* and *Rht-D1* loci of wheat differ significantly in their influence on resistance to Fusarium head blight. Theor Appl Genet 118:695-702

Srinivasachary, Gosman N, Steed A, Simmonds J, Leverington-Waite M, Wang Y, Snape J, Nicholson P (2008) Susceptibility to Fusarium head blight is associated with the *Rht-D1b* semi-dwarfing allele in wheat. Theor Appl Gen 116:1145-1153

Tajima, F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286-289

Tian C, Wan P, Sun S, Li J, Chen M (2004) Genome-wide analysis of the GRAS gene family in rice and Arabidopsis. Plant Molec Biol 54:519-532

Tosovic-Maric B, Kobiljski B, Obreht D, Vapa L (2008) Evaluation of wheat *Rht* genes using molecular markers. Genetika 40:31-38

Uchida N, Townsley B, Chung K-H, Sinha N (2007) Regulation of *SHOOT MERISTEMLESS* genes via an upstream-conserved noncoding sequence coordinates leaf development. Proc Natl Acad Sci USA 104:15953-15958

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: RESEARCH0034

Wang H, Wang X, Chen P, Liu D (2007) Assessment of genetic diversity of Yunnan, Tibetan, and Xinjiang wheat using SSR markers. J of Genetics and Genomics 34:623-633

White J, Law JR, MacKay I, Chalmers KJ, Smith JSC, Kilian A, Powell, W (2008) The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. Theor Appl Gen 116:439-453

Wilhelm EP, Howells RM, Al-Kaff N, Jia J, Baker C, Leverington-Waite MA, Griffiths S, Greenland AJ, Boulton MI, Powell W (2013) Genetic characterization and mapping of the *Rht-1* homoeologs and flanking sequences in wheat. Theor Appl Gen DOI: 10.1007/s00122-013-2055-3

Wu J, Kong X, Wan J, Liu X, Zhang X, Guo X, Zhou R, Zhao G, Jing R, Fu X, Jia J (2011) Dominant and pleiotropic effects of a GAI gene in wheat results from a lack of interaction between DELLA and GID1. Plant Physiol 157: 2120-2130

Yan W, Li HB, Cai SB, Ma HX, Rebetzke GJ, Liu CJ (2011) Effects of plant height on type I and type II resistance to fusarium head blight in wheat. Plant Path 60:516-512

Ye S, Dhillon S, Ke X, Collins AR, Day INM (2001) An efficient procedure for genotyping single nucleotide polymorphisms. Nucl Acids Res 29(17):e88

Zhang X, Yang S, Zhou Y, He Z, Xia X (2006) Distribution of the *Rht-B1b*, *Rht-D1b* and *Rht8* reduced height genes in autumn-sown Chinese wheats detected by molecular markers. Euphytica 152:109-116

Figure and Table headings

Fig. 1 Diagrammatic representation of polymorphic sites (PSs) identified in the *Rht-1*+f regions. Sequenced regions are represented by rectangles. Gray rectangles represent the *Rht-1* open reading frames (ORFs) with dark gray regions corresponding to conserved protein domains (Itoh *et al.* 2002; Tian *et al.* 2004). Unfilled rectangles represent 5' and 3' sequences. The start and end nucleotide coordinates (NCs) bracket each sequence. NCs bracketing each sequence and in the scale are relative to the start nucleotide of the *Rht-1* ORF of 'Chinese Spring' ('CS' NC 1), with negative numbers indicating 5' sequence. The area between closed unfilled rectangles on *Rht-D1* represents a 29 bp ('CS' NCs -1062 to -1034) un-sequenced region in all *Rht-D1* sequences except 'CS'. Vertical lines immediately above rectangles represent PSs present in one or more bread wheat variety relative to 'CS', vertical lines immediately below rectangles represent PSs present in one or more tetraploid/diploid wheat accession relative to 'CS', and vertical lines that span the rectangles represent PSs present in both bread wheat and tetraploid/diploid wheat relative to 'CS'. Arrows indicate the positions and sizes of insertions (ins.) and deletions (del.) greater than 10 bp in length and indicate predicted AA changes, including a one bp ins. in the *Rht-B1* ORF of accession INRA-23995 leading to a predicted frameshift. For AA changes, the number of the affected residue is shown, preceded by the one-letter code of the 'CS' residue and followed by the one-letter code of the substituted residue. Asterisks represent premature stop codons. The four PSs contained within the 197 bp insertion are not shown.

Fig. 2 Distribution of haplotypes among the *Rht-1*+f sequences from the (a) A genome (b) B genome, and (c) D genomes of the bread wheat sets. Haplotype designations are shown followed by the number of accessions (out of 37 per genome) containing that haplotype.

Fig. 3 Expression of Rht-B1 and Rht-D1 in accessions with the *Rht-B1* 160 bp insertion (ins.), *Rht-B1* 197 bp ins., or no ins. Normalized data is presented as the mean of two or three biological replicates and error bars denote the 95% confidence interval (2 × the standard error of the mean) of each sample. GAPDH = Glyceraldehyde 3-phosphate dehydrogenase, EF1α = Elongation factor 1α

Table 1 *Rht-1* and flanking region (*Rht-1*+f) haplotypes and plant heights of sequenced accessions

[a] BW1 is bread wheat set 1. BW1-UK denotes a subset of BW1 accessions widely grown in the UK. Bread wheat 2 (BW2) contains accessions from the INRA worldwide bread wheat core collection of 372 accessions (372CC; Balfourier *et al.* 2007). TDW accessions originate from tetraploid or diploid wheat.

[b] 'Chinese Spring' and *T. urartu* sequences are from BAC clones, Genbank acc. nos. JX978692-JX978695 (Wilhelm *et al*. 2013). The remaining sequences originate from genomic DNA of plants grown from seed from the following collections: JIC, the John Innes Centre Germplasm Resources Unit Triticum collection (http://www.jic.ac.uk/germplasm); US, the USDA-ARS National Small Grains Collection (http://www.ars-grin.gov/npgs); NIAB, the National Institute of Agricultural Botany Triticum collection; INRA, the INRA 372CC. Accession numbers follow the dash.

[c] na = not available.

[d] Geographic (Geog.) origins are: AFG Afghanistan; ARM Armenia; AUT Austria; BEL Belgium; CAN Canada; CHE Switzerland; CHN China; DEU Germany; ESP Spain; FIN Finland; FRA France; GBR Great Britain; ISR Israel; JPN Japan; MAR Morocco; MEX Mexico; NPL Nepal; NZL New Zealand; PAK Pakistan; PSE Palestine; POL Poland; PRT Portugal; RUS

Russia; SYR Syria; TKM Turkmenistan; TUN Tunisia; TUR Turkey; USA United States; YUG Yugoslavia; ZAF South Africa; ZWE Zimbabwe.

[e] Growth habit (GH): S = spring; W = winter; Fac. = facultative.

[f] Reg. yr = Year of varietal registration.

[g] L = landrace; F = fixed line

[h] For the BW1 and TDW sets, *Rht-1*+f haplotypes were determined for the three *Rht-1* homoeologs present in each accession. For BW2, *Rht-1*+f haplotypes were determined for the most diverse 4A, 4B, and 4D chromosomes in the 372CC; therefore, generally only a single homoeolog was sequenced per BW2 accession. Polymorphisms associated with each haplotype are shown in Online Resources 6, 8, and 9. Genbank accession numbers associated with each sequence are shown in Online Resource 5.

[i] Values are means ± standard deviation

Table 2 Summary of diversity measurements within wheat sets

[a] π = nucleotide diversity per site; θ = Watterson's theta per site; na = not applicable.

[b] Indel = insertion-deletions; bp = base pairs; SNP = single nucleotide polymorphisms; amino acid (AA) changes indicate missense, nonsense, and frameshift mutations.

[c] +f refers to the *Rht-1* flanking 5' and 3' regions. The number of bps sequenced in each region is shown in parentheses. *Rht-B1* sequence length includes the 160 and 197 bp insertions.

[d] BW1 = bread wheat 1; BW1-UK = subset of BW1 lines widely grown in the UK; BW2 = bread wheat 2; TDW = tetraploid/diploid wheat accessions.

[e] Sequences were derived from genomic DNA in this study, with the exceptions of 'CS' *Rht-A1*+f, *T. urartu Rht-A1*+f , 'CS' *Rht-B1*+f, and 'CS' *Rht-D1*+f, which are from BAC clone sequences (Wilhelm *et al*. 2013).

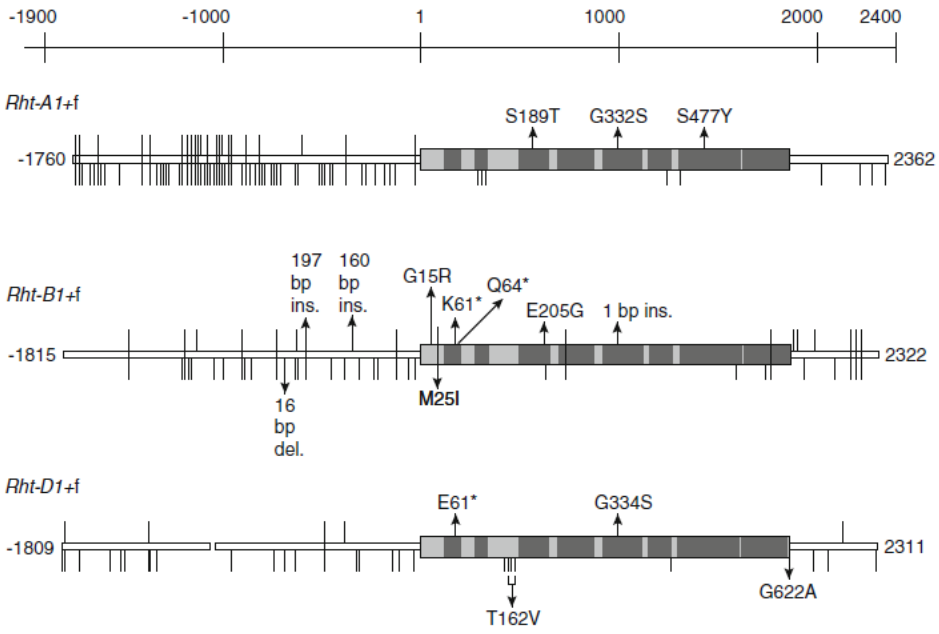
-1900
-1000
1
1000
2000
2400
Rht-A1+f
S189T
G332S
S477Y
-1760
2362
Rht-B1+f
197 bp ins.
160 bp ins.
G15R
Q64*
K61*
E205G
1 bp ins.
-1815
2322
16 bp del.
M25I
Rht-D1+f
E61*
G334S
-1809
2311
G622A
T162V

Figure 1.

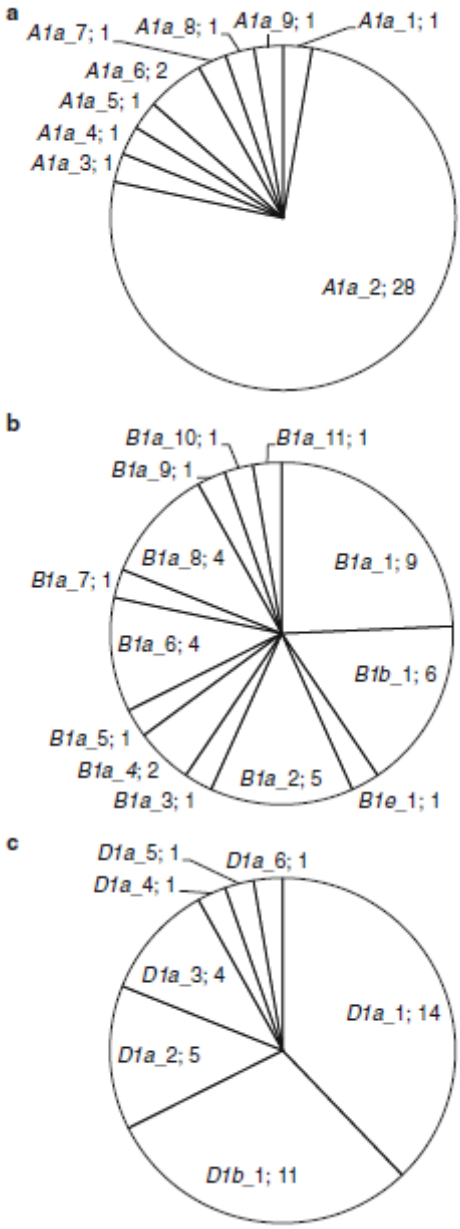
a
A1a_7; 1
A1a_8; 1
A1a_9; 1
A1a_1; 1
A1a_6; 2
A1a_5; 1
A1a_4; 1
A1a_3; 1
A1a_2; 28
b
B1a_10; 1
B1a_11; 1
B1a_9; 1
B1a_8; 4
B1a_1; 9
B1a_7; 1
B1a_6; 4
B1b_1; 6
B1a_5; 1
B1a_4; 2
B1a_2; 5
B1a_3; 1
B1e_1; 1
c
D1a_5; 1
D1a_6; 1
D1a_4; 1
D1a_3; 4
D1a_1; 14
D1a_2; 5
D1b_1; 11

Figure 2.

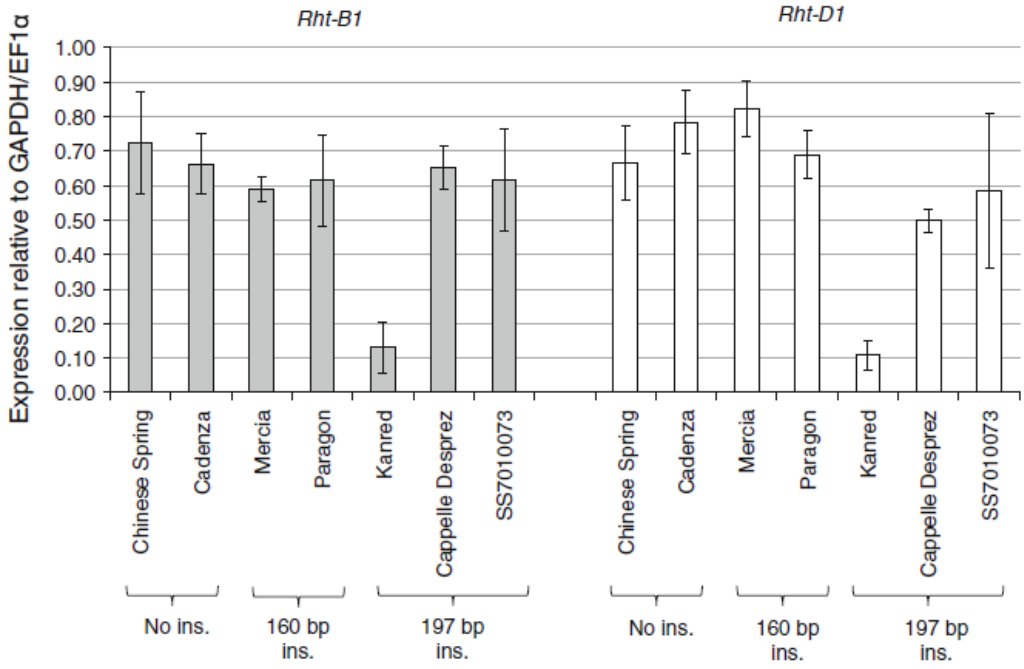

Figure 3.

Table 1 *Rht-1* and flanking region (*Rht-1*+f) haplotypes and plant heights of sequenced accessions

Set[a]	Accession source[b]	Accession name[c]	Geog. origin[d]	GH[e]	Reg. year[f]	L/ F[g]	*Rht-1*+f haplotype[h]			Ht (cm)[i]
BW1	BAC clone	Chinese Spring	CHN	S	na	L	*A1a*_1	*B1a*_1	*D1a*_1	82 ± 4
BW1	JIC-748	Fultz	USA	W	1871	F	*A1a*_2	*B1a*_7	*D1a*_2	115 ± 9
BW1	JIC-814	Gaines	USA	W	1961	F	*A1a*_2	*B1a*_8	*D1b*_1	68 ± 3
BW1	JIC-741	Kanred	USA	W	1917	F	*A1a*_2	*B1a*_5	*D1a*_1	95 ± 10
BW1	JIC-7208	Krasnodari 1	RUS	W	na	F	*A1a*_2	*B1e*_1	*D1a*_1	47 ± 5
BW1	US-PI156641	Norin 10	JPN	W	1935	F	*A1a*_2	*B1b*_1	*D1b*_1	46 ± 1
BW1	US-CItr13253	Norin 10/Brevor-14	USA	W	1949	F	*A1a*_2	*B1b*_1	*D1b*_1	51 ± 3
BW1	JIC-614	Siete Cerros	MEX	S	1966	F	*A1a*_4	*B1b*_1	*D1a*_2	55 ± 5
BW1	JIC-613	Sonora 64	MEX	S	1964	F	*A1a*_3	*B1a*_9	*D1b*_1	60 ± 3
BW1-UK	NIAB	Alchemy	GBR	W	2005	F	*A1a*_2	*B1a*_3	*D1b*_1	na
BW1-UK	JIC-2	April Bearded	GBR	S	na	L	*A1a*_2	*B1a*_2	*D1a*_3	107 ± 8
BW1-UK	NIAB-EW2	Avalon	GBR	W	1979	F	*A1a*_2	*B1a*_2	*D1b*_1	70 ± 2
BW1-UK	NIAB-EW68-3	Cadenza	GBR	S	1992	F	*A1a*_2	*B1a*_8	*D1a*_1	69 ± 6
BW1-UK	NIAB-EW3	Cappelle Desprez	FRA	W	1946	F	*A1a*_2	*B1a*_6	*D1a*_3	90 ± 5
BW1-UK	NIAB-EW7	Hobbit 'Sib'	GBR	W	1975	F	*A1a*_2	*B1a*_6	*D1b*_1	74 ± 3
BW1-UK	JIC-9333	Mercia	GBR	W	1984	F	*A1a*_2	*B1a*_2	*D1a*_3	77 ± 3
BW1-UK	NIAB-EW10	Paragon	GBR	S	1998	F	*A1a*_2	*B1a*_4	*D1a*_3	80 ± 8
BW1-UK	NIAB-EW66-3	Robigus	GBR	W	2003	F	*A1a*_2	*B1b*_1	*D1a*_2	64 ± 2
BW1-UK	NIAB-EW9	Soissons	FRA	W	1987	F	*A1a*_2	*B1b*_1	*D1a*_4	71 ± 4
BW1-UK	JIC-8551	Squarehead's Master	GBR	W	1911	F	*A1a*_2	*B1a*_8	*D1a*_1	86 ± 7
BW1-UK	NIAB-EW11-3	Xi19	GBR	W	2002	F	*A1a*_2	*B1a*_8	*D1b*_1	57 ± 2
BW2	INRA-00537	CH62022	CHE	W	na	F	*A1a*_2			114 ± 6
BW2	INRA-00748	A.4	AFG	W	na	F	*A1a*_5			160 ± 1
BW2	INRA-00822	Aifeng-4	CHN	W	1971	F	*A1a*_2			86 ± 1
BW2	INRA-00957	Arawa	NZL	W	1955	F		*B1a*_6		137 ± 7
BW2	INRA-01192	Balkan	YUG	W	1979	F			*D1a*_1	107 ± 1
BW2	INRA-01697	Bung Epi Blanc	NPL	W	na	L			*D1a*_1	157 ± 8
BW2	INRA-01974	CF4563-1-5-3-2-5	FRA	W	na	F	*A1a*_2			94 ± 1

BW2	INRA-02411	Daeraad	ZAF	S	1958	F	*A1a*_6			130 ± 3
BW2	INRA-03170	Fronthatch	USA	S	1963	F		*B1a*_1		140 ± 5
BW2	INRA-03220	G72300	GRC	S	na	F			*D1a*_2	146 ± 4
BW2	INRA-03485	H93-70	ESP	W	na	F	*A1a*_2			172 ± 2
BW2	INRA-03942	JO3045	FIN	S	na	F			*D1a*_1	146 ± 5
BW2	INRA-03970	Jufy II	BEL	S	1954	F		*B1a*_2		127 ± 3
BW2	INRA-04645	Mars Suede Rouge Barbu	FRA	S	1922	F			*D1a*_1	183 ± 3
BW2	INRA-04796	Miche	FRA	W	1954	F	*A1a*_2			121 ± 8
BW2	INRA-04901	Mocho de Espiga Bianca	PRT	S	1928	F	*A1a*_2			135 ± 0
BW2	INRA-05096	N67M2	ISR	S	na	F			*D1b*_1	54 ± 4
BW2	INRA-05260	Norin 60	JPN	S	1965	F	*A1a*_2			100 ± 3
BW2	INRA-05816	Precoce a Barbe Blanche	PRT	S	1955	F	*A1a*_2			153 ± 13
BW2	INRA-06047	Redman	CAN	S	1946	F			*D1a*_1	137 ± 6
BW2	INRA-06318	Rouge de Marchissy	FRA	W	1929	F		*B1a*_1		189 ± 1
BW2	INRA-06396	S975-A4-A1	ZWE	S	na	F			*D1a*_1	89 ± 9
BW2	INRA-06740	Strubes Dickkopf	DEU	W	1880	F	*A1a*_2			146 ± 8
BW2	INRA-07040	Tremesino Meira	ESP	W	na	L		*B1a*_11		155 ± 11
BW2	INRA-08194	Neelkant	SYR	W	1980	F			*D1a*_2	121 ± 1
BW2	INRA-08287	DC147U	FRA	W	na	F			*D1b*_1	118 ± 4
BW2	INRA-09077	Non Plus Extra	AUT	W	1919	F		*B1a*_2		149 ± 21
BW2	INRA-13310	Fruh Weizen	DEU	W	na	F		*B1a*_4		167 ± 3
BW2	INRA-13436	Fondard Crespin	FRA	W	1948	F			*D1a*_1	165 ± 2
BW2	INRA-13445	Volt	FRA	W	1994	F			*D1a*_1	91 ± 3
BW2	INRA-13471	Ornicar	FRA	W	1998	F	*A1a*_7			90 ± 4
BW2	INRA-13812	W7984 (Synthetic)	MEX	S	na	F	*A1a*_6	*B1b*_1	*D1a*_5	112 ± 6
BW2	INRA-13861	Auguste	FRA	W	1998	F		*B1a*_6		80 ± 0
BW2	INRA-15950	AS68VM4-3-2/TJB636 13	FRA	W	na	F			*D1b*_1	106 ± 1
BW2	INRA-23891	na	ARM	S	na	L		*B1a*_1		142 ± 5
BW2	INRA-23896	na	TUR	S	na	L			*D1a*_6	161 ± 9
BW2	INRA-23909	na	MAR	S	na	L	*A1a*_8			144 ± 6
BW2	INRA-23964	Thori 212-Var.8/1	PAK	S	1934	F	*A1a*_9			129 ± 4
BW2	INRA-23989	na	GEO	S	1931	L			*D1a*_1	171 ± 4
BW2	INRA-23995	na	RUS	S	1950	L		*B1a*_10		150 ± 6
BW2	INRA-23996	Guisuiskaya Syao-Bai-Mai	CHN	S	1953	F	*A1a*_2	*B1a*_1		126 ± 8
BW2	INRA-24056	na	TUR	Fac.	na	L		*B1a*_1		176 ± 1
BW2	INRA-24180	Palestinskaya	PSE	S	1927	F		*B1a*_1		139 ± 6
BW2	INRA-24184	na	PSE	S	1927	L		*B1a*_1		184 ± 5
BW2	INRA-24185	na	TKM	S	na	L		*B1a*_1		152 ± 8
TDW	JIC-7010073	SS7010073 (Synthetic)	na	W	na	F	*A1a*_10	*B1a*_12	*D1a*_7	82 ± 8
TDW	US-PI428054	*T. dicoccoides* 57	TUR	W	na	L	*A1a*_11	*B1a*_14		69 ± 7
TDW	US-PI428097	*T. dicoccoides* 65	ISR	W/S	na	L	*A1a*_12	*B1a*_13		87 ± 6
TDW	BAC clone	*T. urartu*	na	na	na	L	*A1a*_13			na

Table 2 Summary of diversity measurements within wheat sets

			Diversity[a]		Polymorphic sites[b]					Haplotypes	
Re-gion[c]	Set[d]	Seq. (no.)[e]	π x 10^{-3}	θ x 10^{-3}	Indel (no.)	Indel (bp)	SNP (no.)	Total (no.)	AA changes	no.	Di-versity
	BW1	21	0.35	1.01	3	3	15	18	1	4	0.27
Rht-A1+f	BW1-UK	12	0.00	0.00	0	0	0	0	0	1	0.00
	BW2	16	0.68	1.46	3	3	20	23	2	6	0.62
(4122	BW1+BW2	37	0.49	1.22	4	4	21	25	3	9	0.43
bp)	TDW	4	5.94	6.35	16	39	48	64	0	4	1.00
	Overall	41	1.24	3.17	18	41	56	74	3	13	0.54
	BW1	21	0.96	0.99	6	368	16	22	5	11	0.90
Rht-B1+f	BW1-UK	12	1.13	1.11	4	364	15	18	4	6	0.88
	BW2	16	0.76	0.94	5	365	14	19	5	7	0.75
(4494	BW1+BW2	37	0.86	0.85	7	369	16	23	6	13	0.89
bp)	TDW	3	4.15	4.15	4	28	27	31	1	3	1.00
	Overall	40	1.43	2.17	10	394	39	49	6	16	0.91

Rht-D1+f (4091 bp)	BW1	21	0.26	0.20	1	3	3	4	1	5	0.78
	BW1-UK	12	0.29	0.24	1	3	3	4	1	5	0.80
	BW2	16	0.20	0.29	1	1	4	5	2	5	0.67
	BW1+BW2	37	0.24	0.29	2	4	5	7	2	7	0.76
	TDW	1	na	na	na	na	na	na	na	1	na
	Overall	38	0.52	1.56	9	17	27	34	4	8	0.77