Resequencing Vrs1 gene in Spanish barley landraces revealed reversion of six-rowed to two-rowed spike

Ana M. Casas1*, Bruno Contreras-Moreira1,2, Carlos P. Cantalapiedra1, Shun Sakuma3,4, María Pilar Gracia1, Marian Moralejo5, José Luis Molina-Cano6, Takao Komatsuda7, Ernesto Igartua1

Concise title: Natural variation in Vrs1

1Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas (CSIC), Zaragoza, Spain
2Fundación ARAID, Zaragoza, Spain
3Faculty of Agriculture, Tottori University, Tottori, Japan
4Current address: Leibniz Institute of Plant Genetics and Crop Plant Research, IPK, Gatersleben, Germany
5Universitat de Lleida, Lleida, Spain
6Institut de Recerca i Tecnología Agroalimentàries, (IRTA), Lleida, Spain (retired)
7Institute of Crop Science, National Agriculture and Food Research Organisation, Tsukuba, Japan

* Correspondence: Ana M. Casas, acasas@eead.csic.es
Phone: +34 976716085
Abstract

Six-rowed spike 1 (Vrs1) is a gene of major importance for barley breeding and germplasm management as it is the main gene determining spike row-type (2-rowed vs 6-rowed). This is a widely used DUS trait, and has been often associated to phenotypic traits beyond spike type. Comprehensive re-sequencing Vrs1 revealed three two-rowed alleles (Vrs1.b2; Vrs1.b3; Vrs1.t1) and four six-rowed (vrs1.a1; vrs1.a2; vrs1.a3; vrs1.a4) in the natural population. However, the current knowledge about Vrs1 alleles and its distribution among Spanish barley subpopulations is still underexploited. We analyzed the gene in a panel of 215 genotypes, made of Spanish landraces and European cultivars. Among 143 six-rowed accessions, 57 had the vrs1.a1 allele, 83 were vrs1.a2 and three showed the vrs1.a3 allele. Vrs1.b3 was found in most two-rowed accessions, and a new allele was observed in 7 out of 50 two-rowed Spanish landraces. This allele, named Vrs1.b5, contains a ‘T’ insertion in exon 2, originally proposed as the causal mutation giving rise to the six-row vrs1.a2 allele, but has an additional upstream deletion that results in the change of 15 amino acids and a potentially functional protein. We conclude that eight Vrs1 alleles (Vrs1.b2, Vrs1.b3, Vrs1.b5, Vrs1.t1, vrs1.a1, vrs1.a2, vrs1.a3, vrs1.a4) discriminate two and six-rowed barleys. The markers described will be useful for DUS identification, plant breeders, and other crop scientists.

Keywords: barley, landraces, Vrs1, SNP
A feature relevant to describe the history of the barley crop *Hordeum vulgare* subsp. *vulgare* is the spike row-type, two- and six-rowed, according to the fertility of the lateral spikelets of each triplet sitting at each rachis node. *H. vulgare* subsp. *spontaneum*, the wild ancestor of barley, is two-rowed, as explained at length in Komatsuda et al. (2007). So far, five genes determining row-type (*vrs1*, *vrs2*, *vrs3*, *vrs4* and *vrs5*) have been cloned. *Vrs1* (syn. *HvHoX1*) encodes a homeodomain-leucine zipper class I (HD-ZIP I) transcription factor that inhibits the development of lateral spikelets (Komatsuda et al. 2007). *Vrs2* encodes a homolog of *SHORT INTERNODES* (Youssef et al. 2017a). *Vrs3* encodes a putative Jumonji histone demethylase (Bull et al. 2017; van Esse et al. 2017). *Vrs4* encodes an orthologue of the maize *RAMOSA 2* gene (Koppolu et al. 2013). The last gene, *Vrs5* (syn. *Int-c*) encodes a homologue of maize *TEOSINTE BRANCHED 1* (Ramsay et al. 2011). However, only two of them, *Vrs1* and *Vrs5*, affect spike row-type in natural populations (Komatsuda et al. 2007; Saisho et al. 2009; Ramsay et al. 2011; Youssef et al. 2017b).

Several possible mutations at *Vrs1* convert the sterile lateral spikelets into fertile ones. These mutations occur naturally and may have been favored by farmers that interpreted increased fertility as a yield-increasing trait, even though biomass productivity differs little between the two forms (Evans and Wardlaw 1976). The six-rowed trait has appeared independently in several occasions during the history of the crop, acting as a driver of germplasm differentiation (von Bothmer et al. 2003; Komatsuda et al. 2007). Actually, the distinction between two-rowed and six-rowed types is one of the main divides in barley germplasm, as breeders tend to maintain their stocks apart, to avoid the cumbersome process of recovery of pure spike types.

In addition to two two-rowed alleles (*Vrs1.b2* and *Vrs1.b3*), at least three independent mutations in *Vrs1* (*vrs1.a1*, *vrs1.a2* and *vrs1.a3*, all six-rowed) are found among current barleys (Komatsuda et al. 2007). These last alleles are caused by a deletion (*vrs1.a1*) or insertion (*vrs1.a2*) in the coding sequence, resulting in frame shifts; whereas *vrs1.a3* is due to an amino acid change in the homeodomain region. Different surveys resequencing the gene across wild and cultivated barleys (Saisho et al. 2009; Cuesta-Marcos et al. 2010; Ramsay et al. 2011; Youssef et al. 2012, 2017b), ancient DNA from historic landraces (Leino and Hagenblad 2010), or through the use of specific KASP markers in herbarium specimens (Lister et al. 2013), also found the previous alleles and identified another two, *Vrs1.t* (*deficiens*), which is caused by a single amino acid change in the C-terminal region of the protein (Sakuma et al. 2017) and *vrs1.a4* (six-rowed, without any apparent change in the open reading frame (ORF) of *Vrs1.b* alleles).
The Vrs1 region is associated not just to row-type. In fact, it turns up in many barley association scans, for a wide variety of agronomic traits (Cuesta-Marcos et al. 2010; Muñoz-Amatriain et al. 2014; Alqudah et al. 2016). Recently, some studies have provided functional proof of the gene involvement in several phenotypic traits. Liller et al. (2015) found an effect of Vrs1 on tillering, while Thirolugachandar et al. (2017) reported increased leaf width, vein number, leaf nitrogen content, and grain number associated to the six-rowed allele.

Spain is at the end of the routes of distribution that brought barley and other crops to Europe, starting from the Neolithic and ending probably in the Middle Ages (Fischbeck 2003; Komatsuda et al. 2007), receiving crops with adaptations to environmental factors encountered along different routes (Banks et al. 2013; Müller 2015). Therefore, Spanish landraces actually summarize the evolution of the crop in, at least, Southern Europe. Additionally, Spain is one of the few European countries in which cereal landraces were collected and kept in germplasm banks before they disappeared from cultivation, because they were cultivated up to the second half of the 20th century (Igartua et al. 1998; Pujol-Andreu 2011).

Before cloning Vrs1, polymorphisms in the linked Chloroplast Elongation Factor G gene (cMWG699), closely linked to the Vrs1 locus, were used as a surrogate for its row-type characterization (Komatsuda et al. 1998; Tanno et al. 1999, 2002). Cuesta-Marcos et al. (2010) developed four SNP markers within the Vrs1 gene that have been widely used by the scientific and plant breeding community as diagnostic for row-type. However, the current knowledge about Vrs1 alleles and its distribution among barley subpopulations is still underexploited. In this work, we carry out a comprehensive survey of haplotypes at this locus by analyzing Spanish barley landraces, which are complementary of barley accessions analyzed in other studies (Saisho et al. 2009; Cuesta-Marcos et al. 2010; Ramsay et al. 2011; Sakuma et al. 2017; Youssef et al. 2017b), with a predominance of six-rowed over two-rowed accessions.

Materials and Methods

Plant materials

This study involved 176 Spanish barley landraces (50 two-rowed and 126 six-rowed), most of them collected before 1954 (Supplementary Table 1). Of these landraces, 137 (126 six-rowed and 11 two-rowed) belonged to the Spanish Barley Core Collection (SBCC, Igartua et al. 1998)
and 39 come from the set assembled by Moralejo et al. (1994). Thirty-six additional cultivars originated from other countries were studied for comparative purposes: 8 landraces from Morocco (4 two-rowed and 4 six-rowed), 7 of them obtained from the USDA World collection, and 28 widely diverse cultivars (15 two-rowed and 13 six-rowed) that represented the cultivated gene pool from the world. Finally, we included 3 wild barley accessions from Morocco (Molina-Cano et al. 1982) donated by J.L. Molina-Cano (Table 1).

**SNPs genotyping**

Four SNPs within the Vrs1 gene (HORVU2Hr1G092290 in the new barley genome sequence (Mascher et al. 2017)) were genotyped as a part of the 9k Infinium iSelect SNP chip (Comadran et al. 2012). Two of those SNPs (12_30897 and 12_30901) were included in the Barley SNP Panel from Eureka Genomics Corporation (Hercules, CA). The CAPs marker cMWG699/TaqI (Komatsuda et al. 1998; Tanno et al. 1999), corresponding to HORVU2Hr1G092180, was evaluated in all the samples as previously described (Casas et al. 2005). The plant materials tested with each system are detailed in Table 1.

**Sanger sequencing**

Sanger-sequencing of Vrs1 and Int-c (HORVU4Hr1G007040) of selected accessions were carried out as described by Ramsay et al. (2011).

**Exome sequencing**

Exome capture was performed according to the methods described by Mascher et al. (2013). DNA sequencing, made at CNAG (Centro Nacional de Análisis Genómico, Barcelona), and data analysis were performed as described in Cantalapiedra et al. (2016). Briefly, mapping of paired-end reads (2x101 bp) to the Morex WGS assembly was carried out with BWA MEM (Li and Durbin 2009). Variant calling was done by combining SAMtools (Li et al. 2009) and GATK (McKenna et al. 2010). In addition, snpEff (Cingolani et al. 2012) was used to estimate the effect of polymorphisms on coding sequences. Data for Vrs1 and Int-c were retrieved by inspecting the corresponding Morex WGS contigs (contig_135757, Vrs1, and contig_5747, Int-c), as identified by BLASTN alignment at http://webblast.ipk-gatersleben.de/barley_ibsc/viroblast.php.

**RNA extraction and reverse-transcription PCR**
Total RNA was extracted from immature spikes, leaf blades, leaf sheaths, nodes and internodes at awn primordium stage using TRIzol (Invitrogen). RNA was quantified using a NanoDrop 2000 (Thermo Fisher Scientific). To remove genomic DNA contamination, RNA was treated with RNase-free DNase I (Roche). First-strand cDNA was synthesized with SuperScript III (Invitrogen) and first-strand cDNA derived from 20 ng RNA was used as PCR template. Primers used for RT-PCR are listed in Supplementary Table 2. Barley Actin gene was used as positive control.

Phylogenetic analysis

Over two hundred Vrs1 nucleotide sequences were retrieved with BLASTN from the NCBI nt database, using the sequence of SBCC153 (1,133 bp) as query. Hits with low similarity or query coverage, as well as redundant ones, were removed. The surviving sequences were trimmed to the length of SBCC153 and renamed using the allele and haplotype names defined in the work of Saisho et al. (2009). Haplotypes “hap13” and “hap3.2” were also filtered out for bearing large deletions. Sequences for barley genotypes widely used as standards for genomic studies, Barke, Bowman, Haruna Nijo and Morex Vrs1 were included. A multiple alignment was computed with clustal-omega-1.2.1 (Sievers et al. 2011). A parsimony haplotype network was generated with software TCS v1.21 (Clement et al. 2000) and default parameters, which consider gaps as a fifth character. The resulting phylogenetic tree was optimally plotted with http://cibio.up.pt/software/tcsBU (dos Santos et al. 2016).

Results

SNPs genotypes of Vrs1

This study investigates 215 accessions that offer a good perspective of the diversity of the crop, with particular focus on the western Mediterranean region (Table 1). Vrs1 alleles were initially inferred combining spike row-type and the four SNPs within Vrs1 presented by the 9k Illumina Infinium assay (Cuesta-Marcos et al. 2010). The SNP markers provided information to discriminate among genotypes (Table 2). BOPA marker 12_30897 (G/A) differentiates Vrs1.b3 (two-rowed) and vrs1.a3 (six-rowed) from the other alleles. Marker 12_30900 (C/G) is specific to the vrs1.a3 allele. Marker 12_30896 (G/A), although it does not contribute to function of Vrs1, can be used to identify six-rowed barleys with the vrs1.a1 allele from the rest. The last marker 12_30901 (G/A) separates one of the major branches of the phylogenetic tree of Vrs1.
(with alleles Vrs1.b2, vrs1.a2 and Vrs1.t) from the other two major branches of cultivated barley (with alleles vsr1.a1, Vrs1.b3 and vsr1.a3). The BOPA scores of the 215 accessions are shown in Supplementary Table 1.

Sequence validation of Vrs1 alleles

Sanger sequencing of 28 accessions were also carried out (Supplementary Table 1). The sequence data discriminated between some non-committal genotypes and concluded all the allele calls by SNP genotyping. As a whole, nine polymorphisms differentiated seven Vrs1 alleles (Table 2). The sequences of lines identified as vsr1.a1, vsr1.a2, Vrs1.b2, Vrs1.b3, and Vrs1.t were identical to the sequences downloaded from NCBI. A new allele, named Vrs1.b5, was found in seven two-rowed Spanish accessions. Exome sequencing of 73 genotypes (mostly Spanish landraces) independently confirmed the polymorphisms identified (Supplementary Table 3), providing full allelic discrimination. Thus, 26 lines were correctly classified as vsr1.a1, 37 lines were vsr1.a2, 2 lines had the vsr1.a3 allele, and 7 carried the Vrs1.b3 allele and 1 as Vrs1.b5.

Discovery of a novel allele Vrs1.b5

The new Vrs1 allele was identified in seven two-rowed Spanish landraces (Fig. 1A). This allele has been named Vrs1.b5, the number being the next in sequence available for this gene. The Vrs1.b5 allele, presents the distinctive thymine ('T') insertion in exon 2, typical of vsr1.a2 but, on top of it, there was a single base deletion G/_ in the same exon, 45 bp upstream of this insertion. The coupled deletion/insertion results in a frameshift of a stretch of 15 amino acids (Fig. 1B) and restored two-rowed spike from the six-rowed spike (Fig. 1C). The frameshift was located outside of the homeodomain, in a region apparently not relevant for the function of the DNA-binding domain (Fig. 1B). The spike phenotype of the lines carrying the new allele is definitely two-rowed (Fig. 1C), indicating that the 15 substituted amino acids were not essential for the function of VRS1 in terms of suppressing the development of lateral florets.

This change, a single base deletion, has not been observed in any other sequence reported for this gene. A molecular phylogeny analysis with 47 unique sequences from both domesticated and wild barleys, positioned the new Vrs1.b5 allele only one-step apart from the six-row vrs1.a2 allele, both sharing Vrs1.b2 as common ancestor (Fig. 2).
Expression of Vrs1.b5

Transcript of Vrs1 was detected in the two accessions (SBCC153 and SBCC155) carrying Vrs1.b5 (Fig. 3). Vrs1 was predominantly expressed in the immature spikes, as previously reported (Sakuma et al. 2010, 2013). A six-rowed accession (SBCC039) carrying vrs1.a2, the immediate ancestor of Vrs1.b5 allele, showed the same gene expression pattern with the Vrs1.b5 carriers. Implication was that the gain of Vrs1 function in Vrs1.b5 was effected by the frameshift, not by any change of transcription.

Diversity and geographical distribution of Vrs1 alleles

Using the four SNPs from Cuesta-Marcos et al. 2010, together with the morphological identification of the row number, allows the precise identification of most, but not all, Vrs1 alleles. To illustrate this, we predicted the Vrs1 allele for another study involving 138 European winter cultivars (Digel et al. 2016, Supplementary Table 4). Marker-based assignment of Vrs1 alleles allowed unequivocal identification of 60 Vrs1.b3 alleles, 4 vrs1.a1, and 30 vrs1.a3. For the rest, marker information complemented with spike row-type allowed assigning allele vrs1.a2 to 39 cultivars, and five cultivars (all two-rowed) were still inconclusive. Scoring the rest of polymorphisms in Table 2 would allow further differentiation of the Vrs1.b2, Vrs1.b5 and Vrs1.t alleles.

We could also predict correctly row type and specific allele for 109 of the 126 landraces studied by Russell et al. (2016). We retrieved the data for SNPs and indels within Vrs1 (Morex WGS contig_135757) from that study, identifying four polymorphic sites of the nine presented in Table 2 (Supplementary Table 5). Most of the 72 six-rowed accessions carried the vrs1.a1 allele (52), originating from Asia and in Africa; vrs1.a3 was present in 13 accessions, mainly from Eastern Europe, and vrs1.a2 was only found in 4 landraces from Spain and the French Pyrenees. Regarding two-rowed accessions, 39 were identified as having the Vrs1.b3 allele. Further 13 accessions were identified as carrying one of the Vrs1.b2 alleles, and four could not be determined due to missing data.

Association of Vrs1 and EF-G alleles
Previously, variation in the *EF-G* locus, closely linked to *Vrs1*, was used to infer different origins within cultivated barley (Tanno et al. 1999, 2002; Casas et al. 2005), and allows further differentiating the *Vrs1.t* allele (A-type) from the *Vrs1.b2* and *vrs1.a2* lineage (D-type). Most two-rowed accessions analyzed in this study (57 out of 72) had the *Vrs1.b3* allele, associated with the K-type in *EF-G*. Twelve two-rowed lines, (3 wild, 2 landraces from Morocco and 7 Spanish landraces), however, showed a D-type, which is typical of six-rowed lines.

The wild barleys and cultivated landraces from Morocco all carried the *Vrs1.b2* allele, as expected, as a similar finding was reported previously by Komatsuda et al. (2007).

Among 143 six-rowed accessions analyzed, 83 carried *vrs1.a2* allele, and 78 of the *vrs1.a2* accessions carried the D-type of *EF-G* (*cMWG699*) confirming their tight linkage (Komatsuda et al. 1999) and association (Tanno et al. 2002). The *vrs1.a2* was derived from *Vrs1.b2* (Fig. 3) as described earlier (Komatsuda et al. 2007) and all the *Vrs1.b2* carriers were D-type carriers (Table 2, Supplementary Table 1).

All the accessions carrying *Vrs1.b5* and most of the six-rowed accessions (78 out of 83) with *vrs1.a2* also carried the D-type at the *EF-G* locus, indicating that both alleles belong to the same lineage, an implication of the restoration of gene function in *Vrs1.b5* from *vrs1.a2* due to the deletion. The *Int-c* gene that intervenes in the size of the lateral spikelets was sequenced in several lines (Supplementary Table 1). Using Sanger and exome sequence data available for 75 accessions, six-rowed lines carried the *Int-c.a* allele, as expected, whereas seven two-rowed lines, including two with the *Vrs1.b5* allele, were *Int-c.b1*, typical of two-rowed cultivated lines (Ramsay et al. 2011).

**Discussion**

The present study adds a new two-rowed allele *Vrs1.b5* to the catalogue of *Vrs1* diversity. The new allele was created by a restoration of gene function in the ancestral recessive allele by a single nucleotide deletion. This sort of mutation, a kind of gain-of-function, is unusual in nature. The direction of mutation from recessive *vrs1* to dominant *Vrs1* was opposite to the normal direction so far discovered (Komatsuda et al. 2007). In the present study, we identified 7 two-rowed lines with the *Vrs1.b5* allele, out of 50 two-rowed Spanish lines analyzed, i.e., 14% of two-rowed Spanish barleys carry the new allele. The phenotype of these lines is definitely two-rowed, and two accessions surveyed for *int.c* carry the allele typically found in
two-rowed lines (Ramsay et al. 2011). A maximum parsimony phylogenetic analysis suggests
that Vrs1.b5 was a reversion of vrs1.a2 to the two-rowed state through a new mutation. The
reversion of the six-rowed to two-rowed seems far more uncommon than the opposite in the
history of the crop (Komatsuda et al. 2007). Phylogenetically, loss-of-function allele vrs1.a2
seems to derive from Vrs1.b2. Later, a deletion in vrs1.a2 likely gave rise to Vrs1.b5, which
restored the ORF and reverted to the two-rowed phenotype. Thirulogachandar et al. (2017)
performed a phylogenetic analysis for plants HD-ZIP I proteins, identifying putative motifs
evolutionary conserved. Motif 16, 14 amino acids long starting from Lys29, which corresponds
to part of the 15 amino acid region changed in the Vrs1.b5 allele, separated the monocot from
the dicot proteins, as reported by those authors. The amino acid change in Vrs1.b5 allele did
not affect functionality of the protein although the motif was predicted to have a nuclear
localization signal ('RRRRRRSAR').

Origin of Vrs1.b5

The presence of the Vrs1.b2 to vrs1.a2 lineage in the western Mediterranean was previously
reported by Komatsuda et al. (2007), who proposed that the vrs1.a2 allele could be native to
the region. This view was supported by studies carried out with the EF-G locus (marker
cMWG699). Several surveys done with this marker concluded that the D-type (associated to
vrs1.a2 and Vrs1.b2) was found preferentially in the Mediterranean region (Tanno et al. 1999,
2002; Casas et al. 2005; Baba et al. 2011), but it was also present in winter six-rowed cultivars
from Germany, France, and other western European countries (Casas et al., 2005). Since both
two- and six-rowed barleys carrying the D allele are present in North Africa, Baba et al. (2011)
proposed that the origin of the D allele was in Morocco. Our results illustrate that the D allele
is profusely present in Spanish six-rowed landraces, most of them with the vrs1.a2 allele, and
in a small group of two-rowed landraces featuring the new Vrs1.b5 allele. Moreover, even
today, a large proportion of six-rowed vrs1.a2 genotypes can still be found among European
modern winter cultivars, as derived from data provided in Digel et al. (2016). Therefore, its
geographic origin cannot be indicated with certainty. Vrs1.b2 probably appeared in the Middle
East. Recent sequencing of a 6,000-year-old barley from a cave in Israel revealed that it carried
a putative two-rowed Vrs1.b2 genotype (Mascher et al. 2016). Vrs1.b2 has also been found in
the old landrace Palmella Blue (Komatsuda et al. 2007), collected in Egypt early in the 20th
century (https://npgsweb.ars-grin.gov/gringlobal/accessiondetail.aspx?id=1025310). In this
study, we have found this allele in two landraces from Morocco which, could represent
remains of the same genetic stock, found at the two ends of its geographical distribution after
westwards expansion though the Mediterranean during the Neolithic (Zilhão 2011).

Considering together the presence of the precursor allele of Vrs1.b5 (vrs1.a2) exclusively in
Western Europe and, in lower frequencies, in Morocco (Casas et al. 2005, Baba et al. 2011, and
data derived from Digel et al. 2016, and Russell et al. 2016), and the discovery of Vrs1.b5 in
Spanish landraces, we hypothesize that this new allele is native to the western part of the Old
World.

Application of the SNP markers for germplasm characterization

An updated classification of the Vrs1 alleles can facilitate the analysis and differentiation of
genotypes. Earlier efforts by Cockram et al. (2012), who developed a set of KASP markers for
morphological traits assessed to determine distinctiveness, uniformity and stability (DUS) of
new plant breeding varieties, including Vrs1, did not differentiate among all possible alleles.
Similarly, the four SNP markers developed by Cuesta-Marcos et al. (2010) are not fully
diagnostic, but at least allow an easy first discriminating step that could be implemented in all
datasets based on the 9K and 50K (Bayer et al. 2017) barley chips. We illustrated its application
by correctly estimating the Vrs1 allele for 133 of the 138 cultivars evaluated by Digel et al.
(2016), with five inconclusive (Supplementary Table 4). Four of the remaining genotypes
shared a common parent with the Vrs1.t allele (Intro) in their pedigree, and two of them have
recently been identified as deficiens (Sakuma et al. 2017).

Different studies carried out genome wide association analysis for type of spike or yield related
traits and found QTL in the region of the Vrs1 locus. Even though Cuesta-Marcos et al. (2010)
developed the 4 SNP BOPA markers described in Table 2, these authors reported more highly
significant associations by the use of ‘synthetic markers’ which summarized the
dominant/recessive nature of the Vrs1 allele than with any SNP within the gene. In another
genome-wide study with six-rowed cultivars and advanced breeding lines, Berger et al. (2013)
identified markers in the vrs1 region associated with QTL for test weight but were not able to
differentiate the lines. The authors concluded that either there was a closely linked gene for
test weight, or there were two or more vrs1 alleles segregating in the breeding materials with
one of them contributing directly to increased test weight. An examination of the specific Vrs1
alleles present in their dataset would have shed further light on these hypotheses. In the work
by Muñoz-Amatriain et al. (2014), with the USDA barley core collection, the largest panel
tested up to date, the top hit for spike row number corresponded to marker 12_30896. This
marker separates six-rowed \textit{vrs1.a1} accessions from the rest, which includes not only 2-rowed
accessions, but also other 6-rowed ones with different alleles. Therefore, this marker does not
discriminate row type. It was captured by GWAS probably because there was an imbalance of
allelic frequencies in the genotypes, with a majority of 6-rowed presenting \textit{vrs1.a1}. In the last
study published by Thirulogachandar et al. (2017), the two SNPs associated with variation in
leaf area are 12_30896 and 12_30900 (both within \textit{Vrs1} and differentiating six-rowed lines
with allele \textit{vrs1.a1} or \textit{vrs1.a3}, respectively). Overall, a comprehensive evaluation of \textit{Vrs1} for
the markers differentiating all alleles, as haplotypes instead of considering them
independently, would allow an accurate characterization of the alleles present in barley
materials as the ones reported in those studies. A definitive allele characterization could
provide new insights on association of specific \textit{Vrs1} alleles to the relevant agronomic and
morphological traits reportedly related to this gene.

This study completes the allelic catalogue of gene \textit{Vrs1}, offers new insights on explanations for
their geographic distribution, and provides a full list of SNP markers useful for breeders and
germplasm banks to better analyse genetic variation associated to this gene, and facilitate
germplasm classification.

\section*{Data accessibility}

Newly reported sequences for \textit{Int-c} and \textit{Vrs1} are accessible at European Nucleotide Archive
under references LT727691-LT727723.

\section*{Authors’ contributions}

AMC, EI and TK conceived this work. PG, MM and JMC selected and provided the plant
accessions. AMC and SS performed laboratory work. AMC, CPC and BCM analyzed the DNA
sequence data. BCM was responsible for the phylogenetic analysis. AMC, BCM, EI and TK
drafted the document. All the authors read and approved the manuscript.

\section*{Acknowledgments}

The authors would like to thank Dr. Alessandro Tondelli, who provided data for marker
12_30896 on the barley genotypes tested by Digel et al. (2016). This work was supported by
Spanish Ministry of Economy, Industry and Competitiveness grants AGL2010-21929, AGL2013-48756-R, RFP2012-00015-00-00, RTA2012-00033-C03-02 and EUI2009-04075 (national code for Plant-KBBE project ExpResBar). CPC was funded by Spanish Ministry of Economy, Industry and Competitiveness grant no. BES-2011-045905 (linked to project AGL2010-21929). TK and SS were supported by a research fund by the Ministry of Agriculture, Forestry, and Fisheries of Japan (Genomics for Agricultural Innovation grants no. TRS1002). SS was supported by a Grant-in-Aid from the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellow for Research Abroad and a Grant-in-Aid for Young Scientists (B) (no. 16K18635).

Reference List


Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D’Ascenzo M, Akhunov ED, Hedley PE, Gonzales AM, Morrell PL,


Table 1 Plant materials and genotyping platforms used in this study.

<table>
<thead>
<tr>
<th>Country/Region</th>
<th>Type of material</th>
<th>Spike row-type</th>
<th>No.</th>
<th>Genotyping</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morocco</td>
<td>Wild</td>
<td>Two-rowed</td>
<td>3</td>
<td>Eureka Genomics, EF-G, Sanger</td>
<td>Molina-Cano et al. 1982</td>
</tr>
<tr>
<td>Spain</td>
<td>Landraces</td>
<td>Two- and six-rowed</td>
<td>137</td>
<td>9k iSelect, EF-G, exome capture (66), Sanger (14)</td>
<td>Igartua et al. 1998</td>
</tr>
<tr>
<td>Europe/USA</td>
<td>Cultivars</td>
<td>Two- and six-rowed</td>
<td>28</td>
<td>9k iSelect, EF-G, exome capture (7), Sanger (3)</td>
<td>This study</td>
</tr>
</tbody>
</table>
Table 2  Vrs1 alleles detected in this study.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Exon 1</th>
<th>Exon 2</th>
<th>Intron 2</th>
<th>Exon 3</th>
<th>3’ UTR</th>
<th>Row</th>
<th>EF-G</th>
<th>No. of accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vrs1.b2</td>
<td>1067 G</td>
<td>1240 G</td>
<td>1246 G</td>
<td>1288 C</td>
<td>1393 T</td>
<td>1608 G</td>
<td>1725 G</td>
<td>1818 A</td>
</tr>
<tr>
<td>vrs1.a2</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>Vrs1.b5</td>
<td>G</td>
<td>–</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>vrs1.a1</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>–</td>
<td>C</td>
<td>C</td>
<td>–</td>
<td>A</td>
</tr>
<tr>
<td>Vrs1.b3</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>–</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>vrs1.a3</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>–</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>Vrs1.t1</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>–</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>G</td>
</tr>
</tbody>
</table>

Effect: Gly8>Asp, Glu26>F.S, Glu26>Asp, Ala40>F.S, Phe75>Leu, Glu152>F.S, Ser184>Gly

BOPA markers 12_30901 and 12_30896 interrogate the complementary strand, therefore are usually reported as A/G and G/A, respectively.
Figure legends

Fig. 1 Discovery of Vrs1.b5. (A) Alignment of Vrs1.b2, vrs1.a2 and Vrs1.b5 DNA partial sequences of exon 2 indicating the creation of Vrs1.b5 by 1-bp deletion in vrs1.a2. (B) Multiple alignment of protein sequences of different alleles of the Vrs1 gene. Secondary structure elements of the homeodomain (HD) and the leucine zipper dimerization domain (Leu zipper) are shown as color-filled boxes. Predicted protein-DNA interfaces residues are marked with circles. Boxed sequences highlight differences among alleles and asterisks mark premature stop codons. Three wild barley lines ("OUH" identifiers) used as outgroups in a previous phylogeny (Komatsuda et al. 2007) are also shown. Alignment was computed with clustal-omega-1.2.1. (C) Barley spikes from Spanish landraces SBCC039 (vrs1.a2) six-rowed (left) and SBCC155 (Vrs1.b5) two-rowed (right).

Fig. 2 Sequence analysis of 47 aligned Vrs1 alleles. Parsimony network where edges represent single-mutation transitions, and nodes correspond to haplotypes. Tiny circles represent intermediate states (with no associated genotypes). Six-rowed barley accessions are displayed as thick circles. The central haplotype corresponds to OUH630. The labeled circles match the haplotypes shown in Figure 1B.

Fig. 3 Reverse transcription PCR (RT-PCR) analysis of Vrs1
Vrs1 was predominantly expressed in the immature spikes in both six-rowed (SBCC039) and two-rowed barley (SBCC153 and SBCC155). All organs were collected from main tiller at the awn primordium stage. The PCR product of Vrs1 was directly sequenced to confirm the specificity. Actin was used as a control.
<table>
<thead>
<tr>
<th></th>
<th>Spike</th>
<th>Leaf blade</th>
<th>Leaf sheath</th>
<th>Node</th>
<th>Internode</th>
<th>gDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vrs1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBCC039 (vrs1.a2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBCC153 (Vrs1.b5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBCC155 (Vrs1.b5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>