

Duplication of a well-conserved homeodomain-leucine zipper transcription factor gene in barley generates a copy with more specific functions

Shun Sakuma · Mohammad Pourkheirandish · Takashi Matsumoto · Takato Koba · Takao Komatsuda

Received: 16 March 2009 / Revised: 7 July 2009 / Accepted: 3 August 2009 / Published online: 26 August 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Three spikelets are formed at each rachis node of the cultivated barley (*Hordeum vulgare* ssp. *vulgare*) spike. In two-rowed barley, the central one is fertile and the two lateral ones are sterile, whereas in the six-rowed type, all three are fertile. This characteristic is determined by the allelic constitution at the *six-rowed spike 1* (*vrs1*) locus on the long arm of chromosome 2H, with the recessive allele (*vrs1*) being responsible for the six-rowed phenotype. The *Vrs1* (*HvHox1*) gene encodes a homeodomain-leucine zipper (HD-Zip) transcription factor. Here, we show that the *Vrs1* gene evolved in the Poaceae via a duplication, with a second copy of the gene, *HvHox2*, present on the short arm of chromosome 2H. Micro-collinearity and polypeptide sequences were both well conserved between *HvHox2* and its Poaceae orthologs, but *Vrs1* is unique to the barley tribe. The *Vrs1* gene product lacks a motif which is conserved among the *HvHox2* orthologs. A phylogenetic analysis demonstrated that *Vrs1* and *HvHox2* must have diverged after the separation of *Brachypodium distachyon* from the Pooideae and suggests that *Vrs1* arose following

the duplication of *HvHox2*, and acquired its new function during the evolution of the barley tribe. *HvHox2* was expressed in all organs examined but *Vrs1* was predominantly expressed in immature inflorescence.

Keywords Barley · Poaceae · Micro-collinearity · Gene duplication

Abbreviations

DNA	Deoxyribonucleic acid
cv	Cultivar
cDNA	Complementary DNA
BLAST	Basic local alignment search tool
NCBI	National Center for Biological Information
ESTs	Expressed sequence tags
PCR	Polymerase chain reaction
CAPS	Cleaved amplified polymorphic sequence
dCAPS	Derived CAPS
RNA	Ribonucleic acid
RT-PCR	Reverse transcription PCR

Electronic supplementary material The online version of this article (doi:10.1007/s10142-009-0134-y) contains supplementary material, which is available to authorized users.

S. Sakuma · M. Pourkheirandish · T. Matsumoto · T. Komatsuda (✉)
National Institute of Agrobiological Sciences (NIAS),
Plant Genome Research Unit,
2-1-2 Kan-non-dai,
Tsukuba, Ibaraki 305-8602, Japan
e-mail: takao@affrc.go.jp

S. Sakuma · T. Koba
Graduate school of Horticulture, Chiba University,
648 Matsudo,
Matsudo, Chiba 271-8510, Japan

Introduction

The grasses (Poaceae) form a monophyletic family of monocotyledonous plants which includes all the cereal crops, notably rice (*Oryza sativa* L.), maize (*Zea mays* L.), wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.), and sorghum (*Sorghum bicolor* L.). These cereals share a common ancestor from which they have diverged over a period of some 60 million years ago (Devos 2005); nevertheless, some synteny has been retained between them (Devos 2005; Gale and Devos 1998; Lu and Faris 2006). For example, rice chromosome 4 and 7 align well with

chromosome 2 of barley and wheat (Chen et al. 2009; Devos 2005; Moore et al. 1995). With the complete rice genomic sequence to hand (International Rice Genome Sequencing Project 2005), it has become possible to demonstrate both where collinearity has been retained at the fine-scale level (Bennetzen and Ma 2003; Bossolini et al. 2007; Faris et al. 2008; Srinivasachary et al. 2007; Yan et al. 2003), and where it has collapsed as a result of inversions, deletions, duplications, and other intrachromosomal rearrangements (Ilic et al. 2003; La Rota and Sorrells 2004; Li and Gill 2002; Liu et al. 2006; Tarchini et al. 2000). Other full grass species genome sequencing project either completed or underway include those for sorghum (Paterson et al. 2009; Sasaki and Antonio 2009) and *Brachypodium distachyon*, a small genome, short growth cycle, self-fertile, model temperate grass (Ozdemir et al. 2008).

Inflorescence structure is one of the main determinants of grain yield in the cereals. The inflorescence can take the form of a panicle (rice, sorghum, and maize) or a spike (wheat, barley, and *Brachypodium*). Some evidence supports the notion that the spike has evolved from the panicle (Vegetti and Anton 1995). The barley spike carries a set of three spikelets at each rachis node. In “two-rowed” barley, the two lateral spikelets are reduced in size and sterile, but in the “six-rowed” type, all three spikelets are fertile. The six-rowed phenotype is genetically determined by homozygosity for the recessive allele (previously referred to as *vrs1*) at the *vrs1* locus, which has been identified as a homeobox gene (*HvHox1*) encoding a transcription factor containing a homeodomain (HD) with a leucine zipper motif (Zip; Komatsuda et al. 2007). HD-Zip proteins have been grouped into four families (Ariel et al. 2007), with the *Vrs1* gene product (VRS1) belonging to the family I. Although HD can be found in all eukaryotic genomes, the HD-Zip family is restricted to the plant kingdom. The HD-Zip protein is dimerized by the Zip domain, and uses the HD to bind specifically to dyad-symmetrical DNA recognition sequences, based on the strict spatial relationship between HD and Zip (Sessa et al. 1993). VRS1 is thought to suppress the development of the lateral spikelets, since its expression was restricted to the lateral-spikelet primordia in the immature spikes (Komatsuda et al. 2007). The loss of *Vrs1* function resulted in the complete conversion of the rudimentary lateral spikelets of a two-rowed barley into fully developed fertile spikelets, just as in the six-rowed type.

Phylogenetic analysis demonstrated that the origin of the six-rowed phenotype was probably polyphyletic, both temporally and spatially, and occurred via a series of independent mutations at the *Vrs1* (Komatsuda et al. 2007). The higher seed set of the six-rowed type would have been readily selected during the domestication process (Harlan et

al. 1973). Micro-collinearity between rice and barley is disrupted in the *Vrs1* region, but a *Vrs1* ortholog has been identified on rice chromosome 7 (Pourkheirandish et al. 2007). The barley EST (scsnp06322), mapping to the centromere region of chromosome 2H, is homologous to rice Os07g0581000 (LOC_Os07g39280), which co-locates with the rice *Vrs1* ortholog Os07g0581700 (LOC_Os07g39320), (Pourkheirandish et al. 2007; Rostoks et al. 2005). This genomic location suggests the original site of *Vrs1* to be the centromere region of chromosome 2H prior to the chromosomal rearrangement, which has been responsible for the local loss of synteny between rice and barley, but it is plausible that *Vrs1* evolved as a ‘copy’ of an indispensable ‘master’ gene, which is still present in its ancestral location on chromosome 2H (Pourkheirandish et al. 2007). Neither the structure nor the function of *Vrs1* orthologs in any of the other Poaceae members has been elucidated. The objective of this study was to compare the genomic organization of the regions containing a *Vrs1* ortholog in a set of Poaceae species, as a means of inferring the refinement of the function of *Vrs1* by gene duplication in the speciation of barley.

Materials and methods

Plant materials

The two-rowed barley cv. Kanto Nakate Gold (KNG, NIAS accession number JP 15436) and the six-rowed barley cv. Azumamugi (AZ, JP 17209; maintained in the Gene Bank, NIAS, Tsukuba, Japan) were intercrossed to allow the development of a population of 99 F₁₂ recombinant inbred lines (RILs). The wild barley (*H. vulgare* ssp. *spontaneum*) strain OUH602 was obtained from Research Institute for Bioresources, Okayama University, Kurashiki, and used to generate a population of 186 F₇ RILs from the cross OUH602 × KNG. A set of chromosome addition lines (CALs), in which six of barley cv. Betzes chromosomes (2H–7H) are present, in turn, in a background of bread wheat cv. Chinese Spring (Shepherd and Islam 1981) were kindly provided by Dr. A. K. M. R. Islam, Department of Plant Science, Waite Institute, University of Adelaide, Australia.

Barley full-length cDNA library

Seedling shoots and roots of cv. Haruna Nijo were used as a source of mRNAs to construct full-length cDNA libraries, following the methods described by Carninci et al. (1996). A sample of clones was end-sequenced (both 5′ and 3′). A detailed description of this library and its construction will appear elsewhere.

Searching for *Vrs1* orthologs in Poaceae

Nucleotide-BLAST (BLASTN), protein–protein BLAST (BLASTP), and translated nucleotide–protein BLAST (TBLASTN) searches were made against the following sequence databases: barley, Barley Full-Length cDNA End Sequence Database of NIAS (unpublished); rice, Rice Annotation Project Database (<http://rapdb.dna.affrc.go.jp/>) and The Institute for Genomic Research (TIGR) Rice genome annotation (<http://rice.plantbiology.msu.edu/>); maize, MaizeSequence.org (<http://www.maizesequence.org/index.html>); sorghum, Department of Energy Joint Genome Institute (JGI) *S. bicolor* (<http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.html>); *B. distachyon*, BRACHYPODIUM.ORG (<http://www.brachypodium.org/>); wheat, TIGR Wheat Genome database (<http://www.tigr.org/tdb/e2k1/tae1/>), and NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Phylogenetic and peptide motif analysis

Sequence data were aligned using ClustalW2 software (<http://www.ebi.ac.uk/Tools/clustalw2/>). A phylogenetic tree was constructed by the neighbor-joining method, using PAUP 4.0b10 software (Sinauer, Sunderland, Massachusetts) employing 100 bootstrap replicates. Insertion/deletion characters were not included. Peptide motifs were analyzed using the Surveyed conserved motif ALignment diagram and the Associating Dendrogram (SALAD) database (<http://salad.dna.affrc.go.jp/salad/en/>; Mihara et al. 2008). A graphical display of motif composition was constructed using Interactive SALAD analysis.

Barley ESTs data

Barley ESTs giving the best match to rice genes on chromosome 7 were selected from the Gramene database (http://www.gramene.org/Oryza_sativa_japonica/index.html). The copy number of each EST was investigated using Plant Repeat Database at Michigan State University (<http://plantrepeats.plantbiology.msu.edu/index.html>). Exon–intron junctions were assumed to be conserved between rice and barley, and were extracted from NCBI BLAST 2 SEQUENCES (<http://blast.ncbi.nlm.nih.gov/bl2seq/wblast2.cgi>).

PCR amplification

Plant genomic DNA was extracted as described by Komatsuda et al. (1998). PCR primers were designed from the predicted exon regions with Oligo5 software (W. Rychlick, National Bioscience, Plymouth, MN, USA) and synthesized commercially (BEX, Tokyo, Japan) (Supple-

mentary Table 2). PCR amplification was carried out in 10 µl reactions containing 0.25 U ExTaq polymerase (Takara, Tokyo, Japan), 1× ExTaq polymerase buffer, 0.3 µM of each primer, 200 µM dNTP, 2 mM MgCl₂, 0–5% *v/v* dimethyl sulphoxide (DMSO), and 20 ng genomic DNA. Each PCR was cycled through a denaturation step (94°C/5 min), followed by 30 cycles of 94°C/30 s, 55–65°C (primer-dependent)/30 s, 72°C/30–90 s with a final incubation of 72°C/7 min. Amplicons were electrophoresed through either agarose (Agarose ME, Iwai Kagaku, Tokyo, Japan) or a MetaPhor agarose (Cambrex Bio Science Rockland Inc., Rockland, MA, USA) gels, depending on their size, and were visualized by ethidium bromide staining.

Development of CAPS and dCAPS markers

PCR products were purified using the QIAquick PCR purification Kit (Qiagen, Germantown, MD, USA) and subjected to cycle sequencing using a Big Dye Terminator Kit (Applied Biosystem, Foster, CA, USA). Sequencing reactions were purified by Agencourt CleanSEQ (Beckman, Beverly, MA, USA), and analyzed with an ABI prism 3130 genetic analyzer (Applied Biosystem). Sequence data were aligned by Sequencher DNA Sequencing Software (Hitachi-Soft, Yokohama, Japan). Polymorphic restriction sites were identified via the Restriction Maps option of Molecular Toolkit (<http://arbl.cvmbs.colostate.edu/molkit/mapper/>) or with dCAPS Finder 2.0 (<http://helix.wustl.edu/dcaps/dcaps.html>). PCR products were digested at the recommended temperature for 2–3 h in reactions of 15 µl of reaction mixture containing 10 µl PCR products, 1× reaction buffer, and 1 U restriction enzyme.

Genetic mapping

The AZ × KNG and OUH602 × KNG RILs population were genotyped using polymorphic CAPS and dCAPS markers, and MAPMAKER/EXP ver.3.0 (Lander et al. 1987) was used to integrate the resulting loci into the linkage map constructed by Komatsuda and Tanno (2004).

RNA extraction and RT-PCR assay

Total RNA was extracted from various tissues by using TRIzol (Invitrogen, Carlsbad, CA). A first-strand cDNA was synthesized with SuperScript II (Invitrogen). The first-strand cDNA was used as a template, and amplification was performed for 30 or 40 PCR cycles (1 min at 94°C, 30 s at 60/65°C, 30 s at 72°C) followed by 7 min at 72°C. The primers for *HvHox2* were 5'-GCGTGGTTCGAGTGGTTAGCCTGT-3' (sense) and 5'-GAGAGCTACCGGTACTACTTGC-3' (anti-

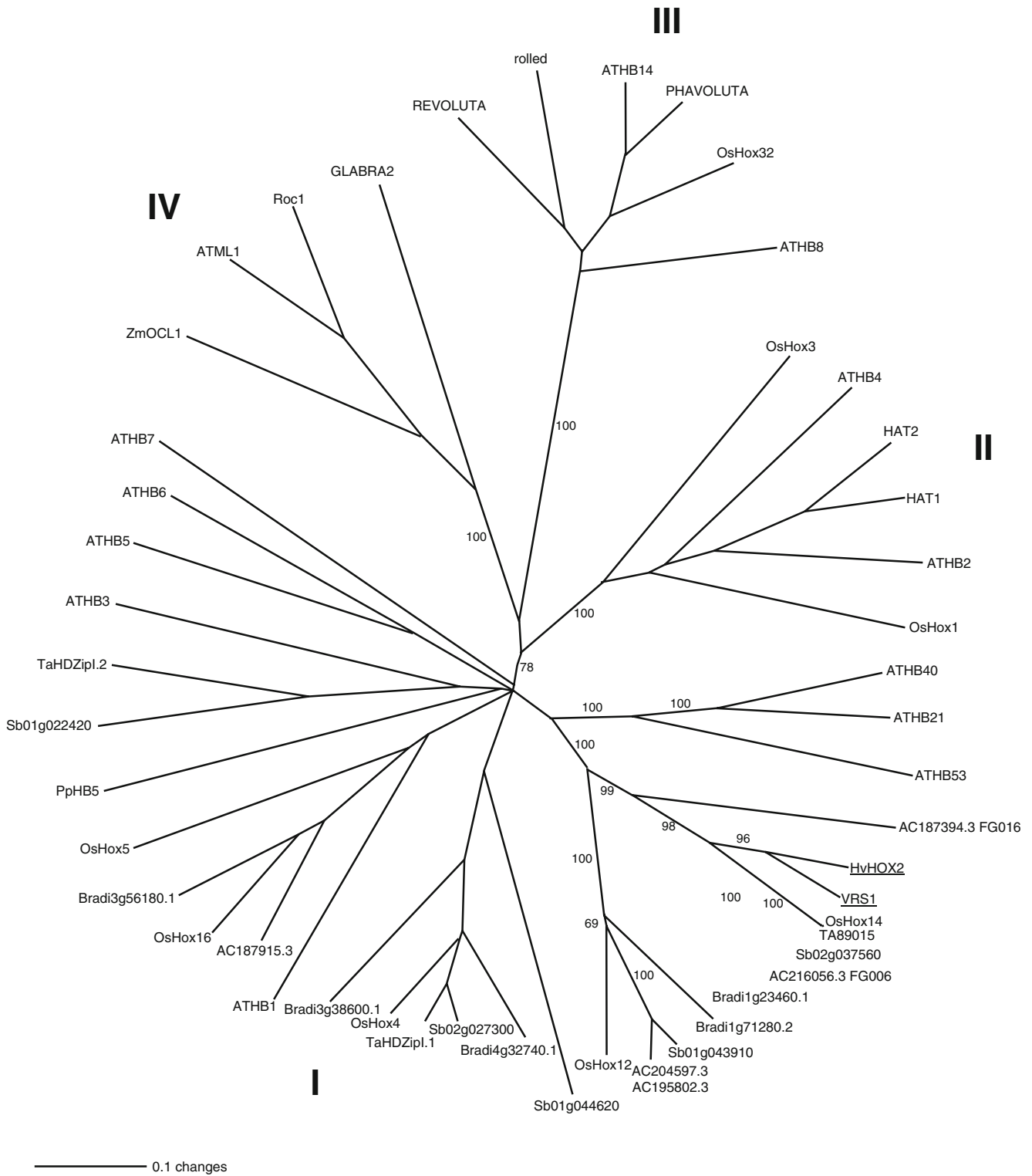


Fig. 1 Phylogenetic analysis of the predicted peptide sequences of the *Vrs1* homologues among the Poaceae and HD-Zip family genes in *A. thaliana*. The phylogenetic tree was constructed by the neighbor-

joining method using PAUP 4.0b10. Local bootstrap values after 100 replicates are indicated near the branching points

sense). *Vrs1* primers were 5'-GGTTTTTAGCATGAATTA GAGTTTA-3' (sense) and 5'-TATACAGGCTAAAAACCA AAGATTA-3' (antisense). *Actin* primers were 5'-GTCCTTTTCCAGCCATCTTTC-3' (sense) and 5'-CAA GAATCGACCCTCCAATCC-3' (antisense). RT-PCR assay was performed at least twice for each sample.

Results

Identification of a *Vrs1* paralog in barley

The best hit (*E*-value, e-110) from a BLASTN search based on the *Vrs1* cDNA sequence (accession no. AB259783) as a

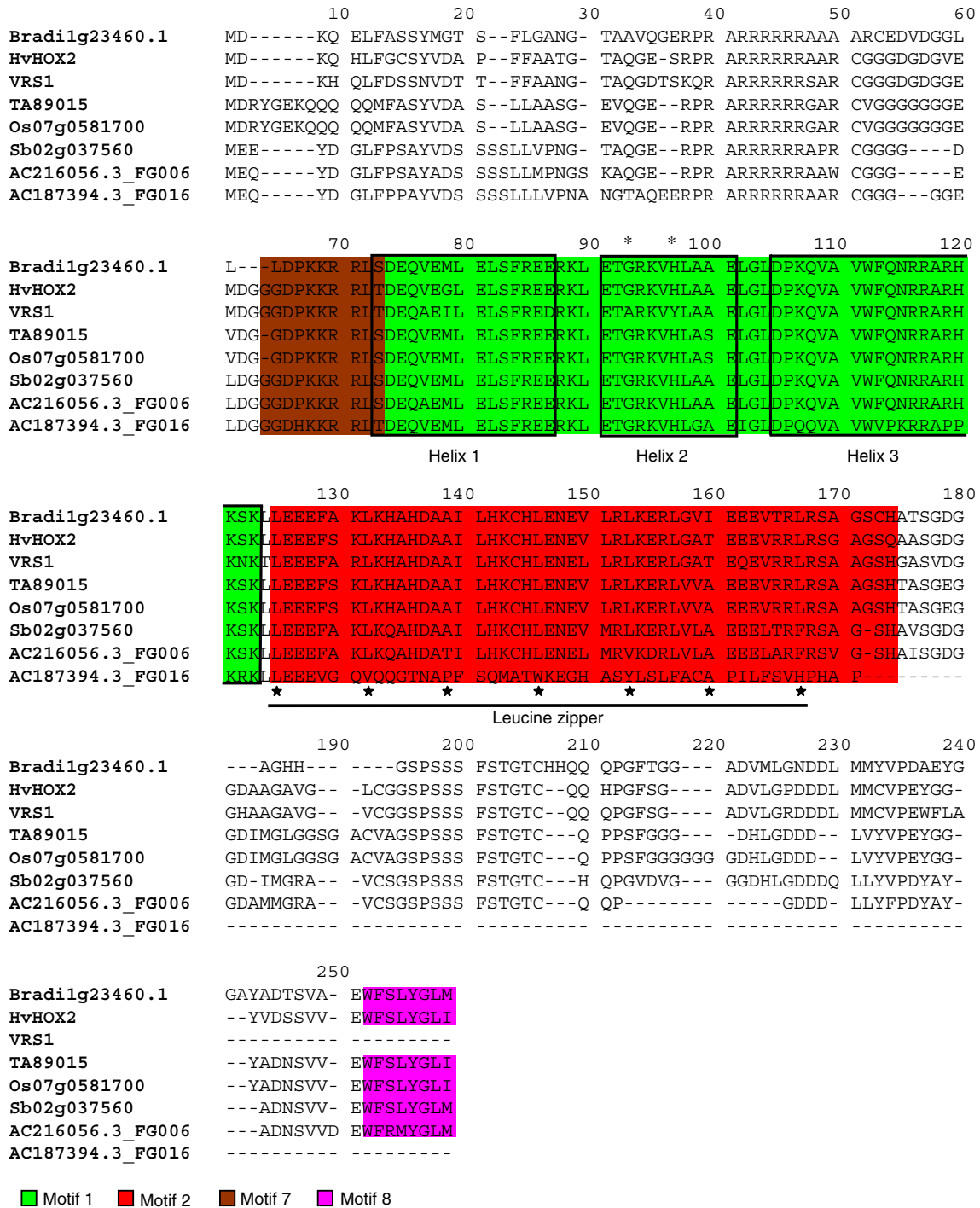


Fig. 2 Comparison between the deduced peptide sequences of VRS1, HvHOX2 and its Poaceae orthologs. The three homeodomain helical sequences are shown boxed, and a leucine zipper is indicated by

staring and underlining. Hyphens indicate a gap introduced to facilitate alignment. Colored boxes show the motifs. Asterisks show the position of Ala 84 (A) and Tyr 88 (Y) in VRS1

query was the barley full-length cDNA clone Hv2074A10 (accession no. AB490233) extracted from the seedling shoots and root of cv. Haruna Nijo (NIAS barley database). Four lesser hits (E -values $< e^{-10}$) were also identified from the same library. The search using *Vrs1* HD-Zip coding region as the query sequence hit three entries (E -values $< e^{-10}$) in the same database. *Vrs1* and Hv2074A10 shared an identical exon/intron structure (Supplementary Fig. 1). The second and third exons of both genes contained homeodomain-leucine zipper motif. As a result, the Hv2074A10 was named “*HvHox2*”. Alignment of the *Vrs1* and *HvHox2* cDNA sequences identified a 300-bp insertion in *Vrs1* and a 44 bp insertion in the third exon of *HvHox2* (Supplementary Fig. 2). Neither the 300 nor the 44 bp sequences were homologous to any accession in the public domain. The *Vrs1* insertion resulted in the generation of a stop codon, producing a polypeptide 14 residues shorter than the *HvHox2* product (Supplementary Fig. 3). The deduced polypeptide sequences of the *HvHox2* (HvHOX2, 236 aa) and *Vrs1* (VRS1, 222 aa) shared 88% identity in the HD-Zip domain and 84% identity in the whole protein.

Identification of *Vrs1* homologues in the Poaceae

The *Vrs1* cDNA sequence was used to query each species database in turn. *Vrs1* homologues were present in rice (Os07g0581700, *Oshox14*), *B. distachyon* (Bradi1g23460.1), sorghum (Sb02g03750), and maize (AC216056.2_FG006 and AC187394.3_FG016). The presence of two (rather than just one) *Vrs1* homologues in maize probably reflects the cryptic tetraploidy of the maize genome (Paterson et al. 2004; Swigonova et al. 2004; Wei et al. 2007). BLASTN search using *HvHox2* as the query sequence produced the same hits. All the *Vrs1* homologues contained an HD-Zip motif (Supplementary Fig. 3).

Phylogenetic analysis of VRS1 homologues

Polypeptide sequences matching that of VRS1 were selected by BLASTP or TBLASTN search against the wheat, *B. distachyon*, rice, sorghum, and maize databases (Supplementary Table 1). The top five sequences with significant similarity to VRS1 were selected from each taxon to form a phylogenetic tree. *Arabidopsis thaliana* HD-Zip family (classes I–IV) proteins were included to infer the class to which the Poaceae proteins belonged. The resulting neighbor-joining tree showed a clear separation of the four families of HD-Zip families, with each clad being strongly supported by bootstrap analysis (Fig. 1). All the VRS1 homologues clustered within the HD-Zip I family (along with *A. thaliana* ATHB40, ATHB21, and ATHB53), suggesting that these sequences are orthologous. The high degree of similarity between VRS1 and HvHOX2 (boot-

strap support level 88%) suggests that they became separated from one another after the separation of barley (Fig. 1). The sister clad, containing the rice, *B. distachyon*, wheat, sorghum and maize sequences, was also highly supported (100%). The peptide motifs 1 and 7 within VRS1 form the homeodomain and motif 2 forms the leucine zipper (Fig. 2). Motif 3, 4, and 6 were highly conserved across all the orthologs (Supplementary Fig. 4)—their function is unknown. Motif 8, in the C-terminal region of the gene, was highly conserved across all the non-barley orthologs and HvHOX2, but not in VRS1 (Fig. 2). Thus, we concluded that *HvHox2*, rather than *Vrs1*, is probably the original representative of the gene in barley.

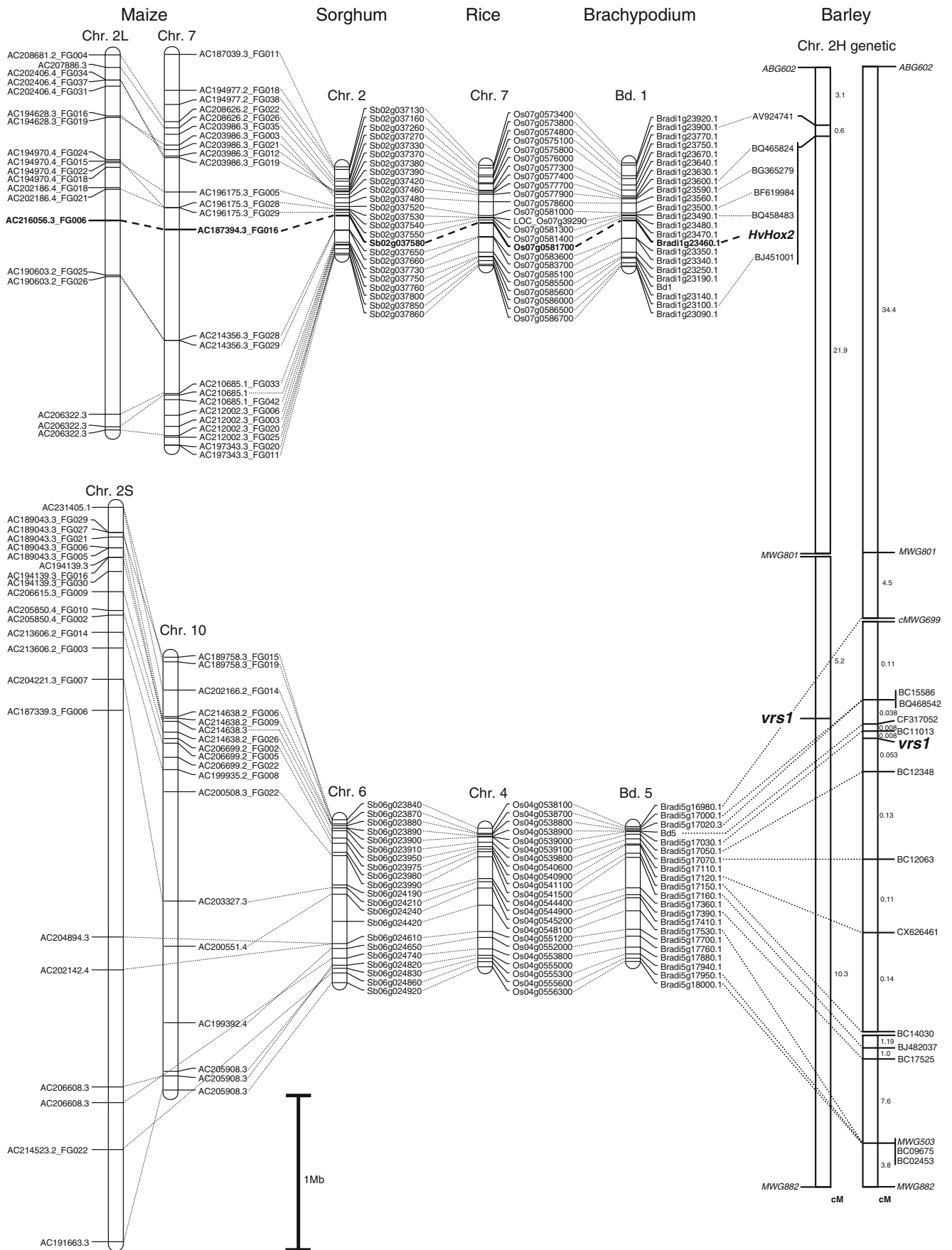
Genetic mapping of *HvHox2* in barley

HvHox2 mapped 27.1 cM from *vrs1* on chromosome 2H (OUH602 \times KNG; Fig. 3). A more saturated map of the *HvHox2* region was obtained by including the loci corresponding to 24 barley ESTs with homology to rice genes in the chromosome 7 containing the rice *HvHox2* ortholog *Oshox14* (syn. Os07g0581700). PCR primers targeting the predicted exon regions generated a single-PCR product for 20 of these 24 ESTs (Supplementary Table 2). CALS analysis assigned eight of 20 to chromosome 2H, with the remaining 12 non-assigned, because the wheat and barley amplicons co-migrated on the gels. The OUH602 and KNG amplicon sequence differed for ten of the EST sequences, but in the contrast between AZ and KNG, only one amplicon sequence was polymorphic. Based on the OUH602 \times KNG RILs, five of the loci co-segregated with *HvHox2*, mapping to a location 0.6 cM proximal to the AV924741 and 3.7 cM proximal to *ABG602* on the short arm of chromosome 2H (Fig. 3).

Micro-collinearity in the Poaceae

The outcome of a study of fine-scale micro-collinearity in the *HvHox2* homologue region between rice chromosome 7 and *B. distachyon* Bd. 1 is shown in Supplementary Table 3.

Fig. 3 Micro-collinearity among the Poaceae in the *Vrs1* and *HvHox2* regions. The maps of *Brachypodium*, rice, sorghum, and maize are physical, and barley genetical. Dotted lines connect barley markers with their Poaceae orthologs. Broken lines connect *HvHox2* with their Poaceae orthologs. In the barley genetic maps, OUH602 \times KNG (left), AZ \times KNG (right), markers shown in *italics* are PCR marked loci derived from RFLPs and AFLPs, while other markers are EST-derived. Markers with ‘BC’ denote barley EST clusters: BC15586 (CB872532 and CA025026), BC11013 (BF064588 and BJ453556), BC12348 (AJ468022 and CB881790), BC12063 (BU973565 and BU987455), BC14030 (CA005338 and CA019164), BC17525 (AJ486478 and BQ664265), BC09675 (AV920420 and BM100120) and BC02453 (BM816591; after Pourkheirandish et al. 2007). Scale bar shows 1 Mb



A series of BLASTN searches based on rice cDNA queries showed that both gene contents and orthologs order was well conserved between these two species (Fig. 3). Similarly, micro-collinearity in the *HvHox2* region was conserved between rice chromosome 7, sorghum chromosome 6, and maize chromosomes 2 S and 10 (Fig. 3). Micro-collinearity was also conserved in the *Vrs1* region (rice chromosome 4, *B. distachyon* Bd. 5, sorghum chromosome 2 and maize chromosome 2 L and 7—see Fig. 3). However, there were no *Vrs1* orthologs present in this region in any of these species.

Expression analysis of *HvHox2* and *Vrs1*

RT-PCR analysis showed that *HvHox2* was expressed in all four organs tested (leaves, shoots and roots of the seedlings, and immature inflorescences in 5 mm long; Fig. 4). *Vrs1* was strongly expressed in the immature inflorescences but slightly or not expressed in the other organs. There was no difference of gene expression pattern between AZ and KNG.

Discussion

A duplication of *HvHox2* gave rise to *Vrs1*

Micro-collinearity can be disrupted by inversion, tandem duplication, indel formation, or transposition (Devos 2005). Two alternative hypotheses have been proposed to account for the loss of micro-collinearity between barley and rice in the *Vrs1* region: (1) the chromosomal segment containing *Vrs1* was transposed from the short arm to the long arm within chromosome 2H; or (2) *Vrs1* evolved from a copy of an indispensable master gene, which is still present in its ancestral location on chromosome 2H (Pourkheirandish et al. 2007). The results of the present study are supportive of the latter hypothesis, since the phylogenetic analysis suggested that *HvHox2* and *Vrs1* are paralog (Fig. 1), while the comparative genetic mapping showed that the ancestral copy is *HvHox2*, not *Vrs1* (Fig. 3). The duplication of *HvHox2* must have occurred after the separation of *B. distachyon* from the Pooideae (Fig. 1), an event which has been timed at ~35–40 million years ago (Catalan and Olmstead 2000). It seems probable that *HvHox2* and *Vrs1* diverged after the split between *Triticum* and *Hordeum*, because no *Vrs1* homologue is represented in the wheat EST database, whereas a wheat *HvHox2* ortholog has been identified. *Vrs1* is present in *H. vulgare* ssp. *spontaneum* (Komatsuda et al. 2007; this study) and *H. bulbosum* (M. Pourkheirandish et al. unpublished) but until its distribution among *Hordeum* spp. more generally has been established, it is not possible to estimate how early during the evolution

of this tribe the duplication event occurred. Although it is not possible to exclude the possibility of the ‘duplicated’ *Vrs1* pseudogene formation in process of the speciation in the other taxa of Poaceae, a BLAST search did not detect any pseudogenes located in the *Vrs1* region in other species (Fig. 3, Supplementary Table 4).

The *HvHox2* sequence is highly conserved among the Poaceae

The HD-Zip I genes share intron and exon distribution (Ariel et al. 2007), and *HvHox2* and its Poaceae orthologs have a similar genomic structure, including HD-Zip motif and peptide motifs (Fig. 2). The expression of HD-Zip I genes is regulated by drought, cold, and osmotic stresses, and by various hormones (Dezar et al. 2005; Himmelbach et al. 2002; Olsson et al. 2004; Rueda et al. 2005). Treatment of *A. thaliana* plants with either abscisic acid or salt increased the transcription of *ATHB21*, *-40*, *-53* genes which are phylogenetically closest to the *HvHox2* (Fig. 1). Most HD-Zip genes (including the three above) are expressed (in seedlings, roots, leaves, stems, and flowers), but a few show organ-specific patterns of expression (Agalou et al. 2008; Henriksson et al. 2005). *HvHox2* was expressed in the leaves, shoots, roots, and immature inflorescences, while *Vrs1* was expressed only in the immature inflorescences (Fig. 4). *In situ* hybridization showed that *Vrs1* expression was very much localized to the lateral-spikelet primordia in spikes at the triple-mound stage and glum-primordium stages (Komatsuda et al. 2007), with zero detectable expression in young leaves (Fig. 4). While *HvHox2* and *Vrs1* may regulate a wholly or partially overlapping set of downstream genes, their biological functions may have been differentiated by their expression profile. On the basis of polypeptide sequences, it seems likely that the molecular role of *HvHox2* and its Poaceae orthologs is identical. Within barley itself, the genome sequence of *HvHox2* (accession no. AB490234) in the two-rowed type (KNG and Haruna Nijo) is the same as that in

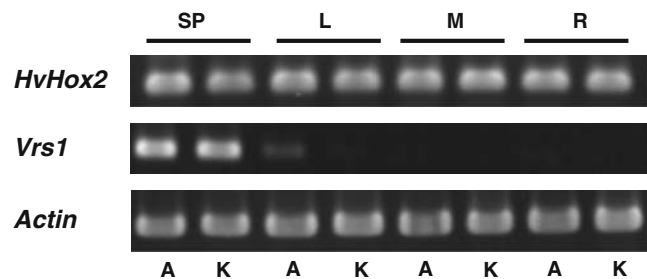


Fig. 4 Expression analysis of *HvHox2* and *Vrs1* by RT-PCR. SP, immature inflorescence (5 mm in length); L leaves of seedling, M coleoptile node segment of seedling, R roots of seedling. A six-rowed barley cv. Azumamugi, K two-rowed barley cv. Kanto Nakate Gold

the six-rowed type (AZ and Morex) (Supplementary Fig. 1), whereas the *Vrs1* sequence is remarkably variable among these same cultivars (Komatsuda et al. 2007). The lack of a *Oshox14* mutant in the *Tos17* insertion mutant collection (Hirochika 2001) does not contradict with the notion that *HvHox2* is essential for plant growth and development in the cereals. In contrast, the loss of *Vrs1* function both in natural variants, and in many induced mutants (including the total deletion of the gene) indicates that *Vrs1* is non-essential for plant growth and development (Komatsuda et al. 2007).

Neo-functionalization of *Vrs1*

Gene duplication is closely associated with creation of novel gene functions (Force et al. 1999). Most paralogs are lost after few million years, but those which either gain a new function (“neo-functionalization”) or cover old function (“sub-functionalization”) survive. In rice, the two C-class MADS box genes *OSMADS3* and *OSMADS58*, which were generated by a gene duplication, have been partially sub-functionalized (Yamaguchi et al. 2006). While *HvHox2* probably has retained its ancestral functions, *Vrs1* may have acquired the role of suppressing the lateral-spikelet development (Fig. 4). *Vrs1* expression was localized to the lateral-spikelet primordia (Komatsuda et al. 2007). Presumably, this phenotype is associated with a selective advantage in nature (Pourkheirandish and Komatsuda 2007). It is unclear, however, whether the recruitment of *Vrs1* to its present function occurred before or after the gain of the triplet in *Hordeum*.

HD-Zip proteins form dimers that recognize the pseudo-palindromic DNA sequence (Chan et al. 1998; Palena et al. 1999, 2001). Two residues in helix II, and one in the loop between helix I and helix II make contact with the target DNA (Tron et al. 2004), and these contacts are critical for aligning the recognition helix correctly. In the VRS1 protein, the glycine 84 and histidine 88 residues present in the helix II region of HvHOX2 and its Poaceae orthologs have been mutated to alanine 84 and tyrosine 88, respectively (Fig. 2). One, or both of these alterations may change either binding affinity or the target DNA sequence. However, if HvHOX2 and VRS1 still share the same target DNA sequence and retain the same level of affinity, it is possible that VRS1 competes with HvHOX2 to bind to *cis*-elements of downstream genes. Since *Vrs1* is expressed only in the lateral-spikelet primordia, VRS1 would tend to out-compete HvHOX2 for binding in the lateral-spikelet primordia. The 300-bp insertion into the *Vrs1* sequence both introduced a new stop codon (Supplementary Fig. 3) and removed a conserved peptide in motif 8 in the C-terminal region (Fig. 2). The role of the motif 8 is not annotated in the protein family (Pfam) database (<http://pfam.sanger.ac.uk/>),

but its loss may be sufficient to differentiate the functions of the *Vrs1* and *HvHox2* gene products. Transcriptional co-activators enhance transcription by interacting with both general and gene-specific transcription factors (Zanetti et al. 2004). Thus, motif 8 in HvHOX2 could interact with certain classes of transcriptional co-activators, and then act as an activator for the transcription of downstream genes, while VRS1 would not be able to replicate this interaction as it lacks motif 8. The formation of HvHOX2/VRS1 heterodimers may reduce the concentration of HvHOX2 homodimer in the cell. None of the above three hypotheses are mutually exclusive, so may be combined to explain the function of *Vrs1*.

Acknowledgments We thank N. Wang, M. Sameri, G. Chen, M. Mihara, H. Sassa and S. Kikuchi for their help and advice. We also thank R. Koebner for linguistic assistance in the preparation of this manuscript. This research was financially supported by the Ministry of Agriculture, Forestry, and Fisheries of Japan (Genomics for Agricultural Innovation grant no. TRC1004).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Reference

- Agalou A, Purwantomo S, Overnas E, Johannesson H, Zhu X, Estiati A, de Kam RJ, Engstrom P, Slamet-Loedin IH, Zhu Z, Wang M, Xiong L, Meijer AH, Ouwerkerk PB (2008) A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. *Plant Mol Biol* 66:87–103
- Ariel FD, Manavella PA, Dezar CA, Chan RL (2007) The true story of the HD-Zip family. *Trends Plant Sci* 12:419–426
- Bennetzen JL, Ma J (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. *Curr Opin Plant Biol* 6:128–133
- Bossolini E, Wicker T, Knobel PA, Keller B (2007) Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. *Plant J* 49:704–717
- Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M, Muramatsu M, Hayashizaki Y, Schneider C (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* 37:327–336
- Catalan P, Olmstead RG (2000) Phylogenetic reconstruction of the genus *Brachypodium* P. Beauv. (Poaceae) from combined sequences of chloroplast *ndhF* gene and nuclear ITS. *Plant Syst Evol* 220:1–19
- Chan RL, Gago GM, Palena CM, Gonzalez DH (1998) Homeoboxes in plant development. *Biochim Biophys Acta* 1442:1–19
- Chen A, Brule-Babel A, Baumann U, Collins NC (2009) Structure-function analysis of the barley genome: the gene-rich region of chromosome 2HL. *Funct Integr Genomics* 9:67–79
- Devos KM (2005) Updating the ‘crop circle’. *Curr Opin Plant Biol* 8:155–162
- Dezar CA, Gago GM, Gonzalez DH, Chan RL (2005) *Habb-4*, a sunflower homeobox-leucine zipper gene, is a developmental

- regulator and confers drought tolerance to *Arabidopsis thaliana* plants. *Transgenic Res* 14:429–440
- Faris JD, Zhang Z, Fellers JP, Gill BS (2008) Micro-colinearity between rice, *Brachypodium*, and *Triticum monococcum* at the wheat domestication locus *Q*. *Funct Integr Genomics* 8:149–164
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545
- Gale MD, Devos KM (1998) Comparative genetics in the grasses. *Proc Natl Acad Sci USA* 95:1971–1974
- Harlan JR, JMJD W, Price EG (1973) Comparative Evolution of Cereals. *Evolution* 27:311–325
- Henriksson E, Olsson AS, Johannesson H, Johansson H, Hanson J, Engstrom P, Soderman E (2005) Homeodomain leucine zipper class I genes in *Arabidopsis*. Expression patterns and phylogenetic relationships. *Plant Physiol* 139:509–518
- Himmelbach A, Hoffmann T, Leube M, Hohener B, Grill E (2002) Homeodomain protein ATHB6 is a target of the protein phosphatase ABI1 and regulates hormone responses in *Arabidopsis*. *EMBO J* 21:3029–3038
- Hirochika H (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr Opin Plant Biol* 4:118–122
- Ilic K, SanMiguel PJ, Bennetzen JL (2003) A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. *Proc Natl Acad Sci USA* 100:12265–12270
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436:793–800
- Komatsuda T, Tanno K (2004) Comparative high resolution map of the six-rowed spike locus 1 (*vsr1*) in several populations of barley, *Hordeum vulgare* L. *Hereditas* 141:68–73
- Komatsuda T, Nakamura I, Takaiwa F, Oka S (1998) Development of STS markers closely linked to the *vsr1* locus in barley, *Hordeum vulgare*. *Genome* 41:680–685
- Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci USA* 104:1424–1429
- La Rota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. *Funct Integr Genomics* 4:34–46
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–181
- Li W, Gill BS (2002) The colinearity of the *Sh2/Al* orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the triticeae. *Genetics* 160:1153–1162
- Liu S, Zhang X, Pumphrey MO, Stack RW, Gill BS, Anderson JA (2006) Complex microcolinearity among wheat, rice, and barley revealed by fine mapping of the genomic region harboring a major QTL for resistance to Fusarium head blight in wheat. *Funct Integr Genomics* 6:83–89
- Lu H, Faris JD (2006) Macro- and microcolinearity between the genomic region of wheat chromosome 5B containing the *Tsn1* gene and the rice genome. *Funct Integr Genomics* 6:90–103
- Mihara M, Itoh T, Izawa T (2008) In silico identification of short nucleotide sequences associated with gene expression of pollen development in rice. *Plant Cell Physiol* 49:1451–1464
- Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution. Grasses, line up and form a circle. *Curr Biol* 5:737–739
- Olsson AS, Engstrom P, Soderman E (2004) The homeobox genes *ATHB12* and *ATHB7* encode potential regulators of growth in response to water deficit in *Arabidopsis*. *Plant Mol Biol* 55:663–677
- Ozdemir BS, Hernandez P, Filiz E, Budak H (2008) *Brachypodium* genomics. *Int J Plant Genomics* 2008:536104
- Palena CM, Gonzalez DH, Chan RL (1999) A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein *Habb-4* with DNA. *Biochem J* 341 (Pt 1):81–87
- Palena CM, Tron AE, Bertocini CW, Gonzalez DH, Chan RL (2001) Positively charged residues at the N-terminal arm of the homeodomain are required for efficient DNA binding by homeodomain-leucine zipper proteins. *J Mol Biol* 308:39–47
- Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA* 101:9903–9908
- Paterson AH, Bowers JE, Rm B, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Ollilar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman WD, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551–556
- Pourkheirandish M, Komatsuda T (2007) The importance of barley genetics and domestication in a global perspective. *Ann Bot (Lond)* 100:999–1008
- Pourkheirandish M, Wicker T, Stein N, Fujimura T, Komatsuda T (2007) Analysis of the barley chromosome 2 region containing the six-rowed spike gene *vsr1* reveals a breakdown of rice-barley micro colinearity by a transposition. *Theor Appl Genet* 114:1357–1365
- Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol Genet Genomics* 274:515–527
- Rueda EC, Dezar CA, Gonzalez DH, Chan RL (2005) *Habb-10*, a sunflower homeobox-leucine zipper gene, is regulated by light quality and quantity, and promotes early flowering when expressed in *Arabidopsis*. *Plant Cell Physiol* 46:1954–1963
- Sasaki T, Antonio BA (2009) Plant genomics: *Sorghum* in sequence. *Nature* 457:547–548
- Sessa G, Morelli G, Ruberti I (1993) The *Athb-1* and *-2* HD-Zip domains homodimerize forming complexes of different DNA binding specificities. *EMBO J* 12:3507–3517
- Shepherd K, Islam A (1981) Wheat: barley hybrids-the first eighty years. In: Evans R, Peacock K (eds) *Wheat science today and tomorrow*. Cambridge University Press, Cambridge, pp 107–128
- Srinivasachary DMM, Gale MD, Devos KM (2007) Comparative analyses reveal high levels of conserved colinearity between the finger millet and rice genomes. *Theor Appl Genet* 115:489–499
- Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaça V, Bennetzen JL, Messing J (2004) On the tetraploid origin of the maize genome. *Comp Funct Genomics* 5:281–284
- Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *Adh1-adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* 12:381–391
- Tron AE, Welchen E, Gonzalez DH (2004) Engineering the loop region of a homeodomain-leucine zipper protein promotes

- efficient binding to a monomeric DNA binding site. *Biochemistry* 43:15845–15851
- Vegetti A, Anton AM (1995) Some evolution trends in the inflorescence of Poaceae. *FLORA* 190:225–228
- Wei F, Coe E, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S, Fuks G, Sanchez-Villeda H, Schroeder S, Fang Z, McMullen M, Davis G, Bowers JE, Paterson AH, Schaeffer M, Gardiner J, Cone K, Messing J, Soderlund C, Wing RA (2007) Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS Genet* 3:e123
- Yamaguchi T, Lee DY, Miyao A, Hirochika H, An G, Hirano HY (2006) Functional diversification of the two C-class MADS box genes *OSMADS3* and *OSMADS58* in *Oryza sativa*. *Plant Cell* 18:15–28
- Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. *Proc Natl Acad Sci USA* 100:6263–6268
- Zanetti ME, Chan RL, Godoy AV, Gonzalez DH, Casalongue CA (2004) Homeodomain-leucine zipper proteins interact with a plant homologue of the transcriptional co-activator multiprotein bridging factor 1. *J Biochem Mol Biol* 37:320–324