Intragenic diversity and functional conservation of the three homoeologous loci of the KN1-type homeobox gene Wknox1 in common wheat

Morimoto, Ryoko
Kosugi, Takayuki
Nakamura, Chiharu
Takumi, Shigeo

(Citation)
Plant Molecular Biology,57(6):907-924

(Issue Date)
2005-04

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000347

Introduction

Common wheat (*Triticum aestivum* L.) has evolved through allohexaploidization between cultivated tetraploid or emmer wheat (*T. dicoccum*) with A and B genomes and wild *Ae. squarrosa* with D genome (Kihara, 1944; McFadden and Sears, 1944). A and B genomes of cultivated emmer wheat have presumably been derived from A genome of *T. urartu* (Chapman *et al*., 1976; Dvorák *et al*., 1993; Takumi *et al*., 1993) and S genome of *Ae. speltoides* (Sarkar and Stebbins, 1956; Sasanuma *et al*., 1996), respectively, after domestication of wild emmer wheat (*T. dicoccoides*). The well-investigated phylogenetic relationships among a series of *Triticum* and *Aegilops* species should facilitate our understanding of the wheat genome divergence and evolution. An important feature associated with allopolyploidization is the dramatic structural alterations in the constituent genomes (Sasakuma *et al*., 1993; Belyayev *et al*., 2000). Several studies showed that low-copy DNA sequences were rapidly and specifically eliminated at the early phases of allopolyploidization (Liu *et al*., 1998a; 1998b; Ozkan *et al*., 2001). Allopolyploidization also affects gene expression via genetic and epigenetic changes (Kashkush *et al*., 2002; 2003). Recent comparative studies of physical contigs covering orthologous loci have revealed a large number of transposon and retroposon insertions in contiguous intergenic regions (Wicker *et al*., 2001; Sanmiguel *et al*., 2002), and a significant part of the increase in wheat genome size has been attributed to these mobile elements (Wicker *et al*., 2003a). These studies indicated that the magnitude of structural modifications in intergenic regions became large enough to significantly reduce homology among them during the wheat genome evolution. On the other hand, information on the structural changes in intragenic regions is limited, particularly in genes containing large introns. Duplications and deletions of large segments in orthologous loci encoding low molecular weight glutenins increased their differentiation (Wicker

et al., 2003b). A comparative study of nucleotide sequences of exon/intron regions among three homoeologous *Waxy* genes of common wheat also revealed intragenic differentiation, but their introns are all short with only 69- to 140-bp in length (Murai *et al*., 1999). Long introns such as the second intron of *Arabidopsis AGAMOUS* gene sometimes possess regulatory elements to control the functional mRNA levels (Deyholos and Sieburth, 2000). Therefore, a comparative analysis of intragenic nucleotide diversity in genes with long introns is important for studying genome differentiation and evolution.

Homeobox genes encode transcription factors, which play important roles in developmental processes in both animals and plants. A maize *knotted1* (*kn1*) gene is the first homeobox gene identified in plants and cloned by a transposon tagging strategy (Vollbrecht *et al*., 1991). The *kn1*-type homeobox (*KNOX*) genes isolated later from many other plant species are classified into two subgroups. The class I *KNOX* genes include maize *kn1* and *rough sheath1* (Schneeberger *et al*., 1995), rice *OSH1* (Matsuoka *et al*., 1993) and *OSH15* (Sato *et al*., 1998), barley *HvKnox3* (Müller *et al*., 1995) and *Arabidopsis STM1* (Long *et al*., 1996) and *KNAT1* (*BREVIPEDICELLUS*) (Lincoln *et al*., 1994; Douglas *et al*., 2002; Venglat *et al*., 2002). Their spatial expression was confined primarily in the meristematic tissues and has a profound effect on vegetative development, in contrast to the constitutive expression of the class II *KNOX* genes. The class I *KNOX* genes control the maintenance of shoot apical meristem (SAM), determination of cell fates and differentiation of vegetative tissues (Kerstetter *et al*., 1997). In *Arabidopsis* and tobacco, overexpression of these genes results in the development of lobed leaves, distally displaced stipules and occasional ectopic meristems on the adaxial leaf surface (Sinha *et al*., 1993; Sato *et al*., 1996). A number of dominant mutations were identified in the class I *KNOX* gene loci of various plant species. Such gain-of-function mutations induce dramatic morphological changes, but the patterns of alteration assume species-specific features (Williams-

Carrier *et al*., 1997). The mutant maize *Kn1* alleles disrupt normal leaf development and induce knot-formation on the leaf vein. This *Kn1* dominant phenotype is due to altered gene expression patterns by structural mutations at the *kn1* locus. Most of the mutations result from insertion of *Ds* or *Mutator* transposons into the longest intron (3rd intron) of the *kn1* gene, and one allele (*Kn1-0*) is caused by a tandem duplication of the 17-kb genomic region including the entire coding region (Vollbrecht *et al*., 1993). A tomato mutation *Mouse ears* dramatically changes the leaf architecture forming compound leaves, which is caused by a fusion event involving a class I *KNOX* gene *LeT6* (Chen *et al*., 1997). In barley *Hooded* (*K*) mutants possessing a dominant allele of *HvKnox3*, inflorescence-like structures differentiate on the upper lemmas and awns (Müller *et al*., 1995). This hooded phenotype is caused by a 305-bp duplication in the longest intron (4th intron) at the *HvKnox3* locus (Müller *et al*., 1995). A similar phenotype was recovered in transgenic barley plants overexpressing the maize *kn1* cDNA (Williams-Carrier *et al*., 1997).

The class I *KNOX* gene family commonly has a large intron of several kilobase pairs in length and additional three or four short introns. This large intron contains some regulatory elements controlling the level of transcription and tissue-specificity (Santi *et al*., 2003). Therefore, structural mutations in the large intron often induce ectopic expression of the *KNOX* genes and exhibit dominant mutant phenotypes. These features render the *KNOX* genes a suitable candidate for studying structural and functional divergence associated with allopolyploidization in common wheat. We previously isolated three highly homologous cDNA clones encoding the KN1-type homeobox protein from young spikes of Japanese common wheat cultivar ‘Norin 26’ based on the homology to *kn1* and *OSH1* (Takumi *et al*., 2000). The three cDNA copies were assigned to the homoeologous group 4 chromosomes in the three component genomes of common wheat and designated *Wknox1a*, *Wknox1b* and *Wknox1d*. Their predicted

amino acid sequences were highly conserved, and most of the polymorphic sites among the *Wknox1* cDNA sequences were synonymous substitution. The high degree of conservation of the *Wknox1* gene likely reflects their important functional roles in the monocot development. Despite the conserved feature among the *kn1* homologs, structural mutations with phenotypic alterations were frequently identified in long introns of the maize *kn1* and barley *HvKnox3*. Therefore, objectives of our study were to identify structural mutations and evolutionary changes through allopolyploid evolution in the *Wknox1* loci and their consequences on the *Wknox1* expression. Here, we present our results of a comparative analysis of the three genomic sequences of the *Wknox1* homoeologs in a common wheat cultivar 'Chinese Spring' (CS). We found a large number of structural mutations in the 4th intron and the 5'-upstream region among them and allocated some of the mutation events in the predicted lineage of the wheat genome evolution. Despite such mutations, the function of the three *Wknox1* homoeologs was conserved in the wheat genome evolution.

Materials and methods

Plant materials

Genomic sequences of the three *Wknox1* loci and their expression were determined using a common wheat (*Triticum aestivum* L.) cultivar 'Chinese Spring' (CS). Evaluation of the *Wknox1* function was made in a heterologous system using transformants of *Nicotiana tabacum* cv. 'Petite Havana'. To study the distribution of seven large mutations at the *Wknox1* loci, a total of 48 *Triticum* and *Aegilops* accessions were used (Table 1). They included 10 accessions of einkorn

wheat (wild *T. urartu* and *T. boeoticum* and cultivated *T. monococcum*, genome constitution AA), 4 of *Aegilops speltoides* (SS), 8 of four other *Aegilops* species in the section Sitopsis (SS) and 5 of *Ae. squarrosa* (DD). Two common wheat (AABBDD) cultivars, 'CS' and 'S-615', 10 accessions of wild tetraploid *T. dicoccoides* (AABB) and 9 accessions of cultivated tetraploid *T. dicoccum* (AABB) were also used to analyze the distribution of insertions and deletions (indels). These tetraploid wheat accessions were selected to cover the major areas of their distribution. Single plants from all accessions were grown in greenhouse for verification of the species classification and for DNA extraction.

mRNA in situ *hybridization of the* Wknox1 *gene and production of transgenic tobacco plants with the* 35S::Wknox1 *chimeric gene*

Wheat tissues for mRNA *in situ* hybridization were fixed in 0.1 M sodium phosphate buffer containing 4% (v/v) paraformaldehyde and 0.25% (v/v) glutaraldehyde for more than 12 hr at 4 ˚C. The fixed tissues were embedded in Paraplast Plus (Sherwood Medical) after dehydration with ethanol and infiltration with xylene. The embedded samples were sliced into 10 µm sections by rotary microtome HM325 (Zeiss). Hybridization of *Wknox1* mRNA with digoxigenin-labeled sense and antisense probes produced from the *Wknox1a* coding region was performed overnight at 50 ˚C. The RNA probes were directly synthesized with T3 and T7 RNA polymerase (TOYOBO, Japan) from a cDNA clone encoding *Wknox1a* (Takumi et al., 2000). After hybridization, the sections were washed in 4x SSPE at 65 ˚C. Immunological detection of the hybridized probe was according to Sentoku *et al*. (1999).

A cDNA clone encoding *Wknox1a* was introduced into the *Xba*I/*Sac*I site of pBI121 (Clontech) to produce the *CaMV35S::Wknox1a* construct. The *35S::Wknox1a* construct was

introduced into leaf discs of 'Petite Havana' using *Agrobacterium tumefaciens* LBA4404 to study phenotypic changes induced by over-expressed *Wknox1a*. Transformants were selected on the Murashige-Skoog (MS) medium (Murashige and Skoog, 1962) containing 0.1 mg/l NAA, 1.0 mg/l BA and 250 mg/l kanamycin, and regenerated on hormone-free MS medium containing 50 mg/l kanamycin. Southern and northern blot analyses were conducted respectively to confirm integration and expression of the transformed *Wknox1a* gene as previously reported (Ohno *et al*., 2001).

Cloning and sequencing of the homoeologous Wknox1 *loci*

Genomic clones containing the *Wknox1* sequences were screened from a wheat transformation-competent artificial chromosome (TAC) library constructed from CS (Liu *et al*., 2000). The 1st and 2nd screenings were performed according to the procedure of Liu *et al*. (2000) and five positive clones were isolated. For characterization of the isolated TAC clones, DNA inserts were digested with a restriction enzyme *Hin*dIII, resolved by electrophoresis through 0.8% agarose gel and blotted onto nylon membranes. Southern hybridization with the *Wknox1a* cDNA as a probe was carried out using an enhanced chamiluminescence (ECL) direct nucleic acid labeling/detection system (Amersham, UK). The *Wknox1* genomic regions were subcloned into pBluescriptII SK^- plasmid vector (Stratagene, USA) and the three loci, all with ca. 10-kb in length, were sequenced by the automated fluorescent dye deoxy terminator cycle sequencing system using ABI PRISM^{TM} 310 Genetic Analyzer (PE Applied Biosystems, USA). The DNA sequences were analyzed by a DNASIS (Hitachi, Japan) and a multiple alignment was calculated by the Waterman's algorithm (Waterman, 1986).

Detection of polymorphisms among the three Wknox1 *loci*

Based on the multiple alignment data, a large number of single nucleotide polymorphisms (SNPs) and indel mutations were found among the three *Wknox1* loci. To exclude cultivar differences, sequence data of all the three homoeologous loci were obtained from a single cultivar of common wheat, CS. Base-change mutations were classified into the following two groups. In one group nucleotides at corresponding sites were all different among the three homoeologs, and in another identical nucleotides were found in two homoeologs. Hence, we defined the former group as 'unclassified base changes' and the latter as 'genome-specific SNPs'. Nucleotide changes unique in one homoeolog were tentatively classified into insertion mutations, and sequences uniquely lacking in one homoeolog were classified into deletion mutations.

Based on the number of the indel mutations and the genome-specific SNPs, we calculated mutation rates in the *Wknox1* loci. The unclassified base changes were omitted from the calculation because their genome specificity could not determine from the SNP data. The calculated mutation rates among the three homoeologous loci were considered to reflect the levels of differentiation of the A, B and D genomes in the hexaploid genome.

PCR analysis of insertion/deletion mutations at the Wknox1 *loci*

A comparison of genomic sequences of the three *Wknox1* loci revealed a number of indel mutations with longer than 6-bp in length (see Results). Six indel mutations and one recombination event were selected to study the distribution of mutations in *Triticum* and *Aegilops* species. Specific primer sets for PCR analysis were designed for individual indel mutation sites (Table 2). The primer positions were set in the flanking regions of the mutation sites. All the

primer sets precisely recognized and amplified the objective regions from the ancestral species of common wheat. The template DNA (50 ng) and 10 nM each of the primers were mixed with 2 µl of 10x DNA polymerase buffer, 100 µM of a dNTP mixture (equimolar dATP, dCTP, dGTP and dTTP), 10% (v/v) glycerol and 1 unit of Taq DNA polymerase in a total volume of 20 µl. PCR consisted of the following steps: denaturation at 94˚C for 1 min, annealing at the indicated temperature for 1 min, and DNA synthesis at 72˚C for 1 min. Thirty-five cycles of PCR were performed, and the amplified DNA fragments were analyzed by ethidium-bromide staining after 1.5% agarose gel electrophoresis.

RT-PCR analysis of the homoeologous Wknox1 *transcripts and functional analysis of the 5'-upstream sequences of the* Wknox1 *loci*

Total RNA was extracted from the following tissues; young spikes, shoots containing SAM, pistils at two different developmental stages, lemmas and paleas of 'S-615' according to Takumi et al. (2000). cDNA was synthesized from DNaseI-treated RNA with Rever-Tra Ace α (TOYOBO, Japan) and subjected to RT-PCR (reverse transcriptase PCR) analysis. In RT-PCR of the three homoeologous *Wknox1* transcripts, a common forward primer, 5'-GGAGGGTGGAGACGCAACTCAACT-3', and the following three reverse primers were used, which were designed based on the homoeolog-specific sequences of the 3' non-translated region; 5'-CACCGACCAAGGTCACCAGT-3', 5'-CGCCGACCAAGGTGACCGGC-3' and 5'-CCAAGGTCACCGGTAACGGT-3' for *Wknox1a*, *Wknox1b* and *Wknox1d*, respectively. For amplification of the ubiquitin gene as a control, the following gene-specific primer set was used; 5'-GCATGCAGATATTTGTGAA-3' and 5'-GGAGCTTACTGGCCAC-3'. All of the amplified

DNA fragments were separated by electrophoresis through 2% agarose gel and stained with ethidium-bromide.

The 5'-upstream sequences, all about 2-kb in length, of the three *Wknox1* homoeologs were individually fused to the *gusA* reporter gene to analyze their tissue-specific expression pattern. An additional construct was produced to contain the *gusA* gene under control of the 680-bp 5'-upstream region of the *Wknox1d* homoeolog. Transformation was conducted as already described. GUS activity was assessed histochemically as described previously (Takumi and Shimada, 1996). The chlorophyll of histochemically-stained leaves was extracted with ethanol.

Results

Functional orthology of the Wknox1 *gene to the* kn1*-type homeobox gene*

In the previous RT-PCR study (Takumi *et al*., 2000), we showed that the *Wknox1* genes were specifically expressed in SAM-containing shoot and young spikes and that the tissue-specific expression pattern of the *Wknox1* gene was similar to that of the *kn1*-type homeobox gene family including maize *kn1* and rice *OSH1* (Kerstetter *et al*., 1994; Sentoku *et al*., 1999). To confirm this, mRNA *in situ* hybridization using the antisense probe was performed in the longitudinal sections of the shoot apex from one-day-old wheat seedlings. The *Wknox1* expression was confined in the interior (L2) cells of SAM and developing tissues in the stem (Fig. 1A). There was no detectable *Wknox1* mRNA in young leaf primordia and old leaves. Control experiment using the sense probe showed no hybridization signals (data not shown). The result suggested that the characteristic tissue specificity of the *kn1*-type genes was conserved in the wheat *Wknox1*

gene. To further analyze biological effects of its ectopic expression, *Wknox1a* cDNA was overexpressed under control of the CaMV35S promoter in transgenic tobacco plants. More than thirty transgenic plants were recovered. Integration into the tobacco genome and expression in leaves were confirmed by Southern and northern blot analyses in all transgenic plants (data not shown). A majority of the transformants showed abnormal morphologies including aberrant leaves. The altered morphology was inherited to the next generations (data not shown). Flower morphology in the transformants was slightly altered, showing more deeply split petals than the wild-type (Fig. 1B vs 1C). In the transformants with mild effects, the leaf edges were slightly wrinkled (Fig. 1D vs 1E). Some transformants showed round leaves. Severely altered transformants developed compound leaves (Fig. 1F), similar to those observed in transgenic tobacco plants overexpressing rice *OSH1* (Sato *et al*., 1996) and *OSH15* (Sato *et al*., 1998). In such transformants, ectopic leaves and/or shoots were occasionally formed on the adaxial side of the leaves (Fig. 1G, 1H). This phenotype is similar to that of transgenic tobacco overexpressing *kn1*, *OSH1* and *HvKnox3* (Sinha *et al*., 1993; Müller *et al*., 1995; Sato *et al*., 1996). To confirm ectopic expression of the introduced *Wknox1a* gene, *in situ* hybridization with the antisense probe was conducted using the sections of leaves forming ectopic shoots. The *Wknox1* mRNA was detected in the ectopic shoots on the leaves of the transgenic tobacco plants (Fig. 1I). The ectopic expression of the *Wknox1a* gene triggering the typical ectopic tissue-formation indicated that the *Wknox1*-derived protein played a functionally similar role to that of other KN1-type homeobox proteins.

Genomic sequence diversity in the homoeologous Wknox1 *loci*

The maize *kn1* and rice *OSH1* have five exons and four introns (Vollbrecht *et al*., 1991; Matsuoka *et al*., 1993), while the barley *HvKnox3* (*Bkn3*) and wheat *Wknox1* have an additional exon and intron at the 5' non-coding region (Müller *et al*., 1995; Takumi *et al*., 2000). This exon/intron structure is conserved among the three *Wknox1* homoeologs (Takumi *et al*., 2000). To further study structural divergence among them, we isolated five genomic clones from the TAC library of CS. Three clones contained full-length *Wknox1* genomic sequences with different *Hin*dIII restriction fragment length polymorphism patterns. Nucleotide sequences of the *Wknox1* regions in these clones were identical to each one of the three *Wknox1* cDNA sequences. Thus, we concluded that the TAC clones contained the all three homoeologous loci, i.e. *Wknox1a*, *Wknox1b* and *Wknox1d*. The *Wknox1* genomic sequences including 5'-upstream regions were subcloned and their nucleotide sequences were determined. Total lengths covering the 5'-non-coding region, the open reading frame (ORF) and the 3'-downstream to the polyadenylation site were 7,243-, 7,256- and 7,444-bp in *Wknox1a*, *Wknox1b* and *Wknox1d* loci, respectively (Table 3). The *Wknox1* homoeologs all possessed GC-AG nucleotides instead of GT-AG at the exon/intron borders of the second intron. Nucleotide sequences of the 5'-upstream regions were 2,512-, 3,592- and 2,581-bp in length, respectively.

High homology was observed throughout the exon/intron regions of the three *Wknox1* homoeologous loci, but a large number of indel-type polymorphisms were found among them (Fig. 2A; Table 3). A total number of indel mutations in the studied region was 114, among which 29 indels were longer than 6-bp in length and 15 of them were localized in the *Wknox1b* locus. A total number of short indel mutations less than 5-bp in length was 85, and most of them were found in the fourth introns (Table 3). One 3-bp deletion was observed in the ORF of *Wknox1a*, and 4 indels occurred from the stop codon to polyadenylation site as previously reported (Takumi et al., 2000). Four insertion sequences (ISs) showed high homology to the

miniature inverted repeat transposable element (MITE) classified into the *Stowaway* family in barley and *Heteranthelium piliferum* (Bureau and Wessler, 1994). The *Stowaway* family is an abundant transposable element in the grass genomes. Two of these four ISs were 123- and 160-bp in length, respectively contained 36- and 41-bp terminal inverted repeats (TIRs), and located in the fourth intron of *Wknox1b*. Sequence identities of the TIRs were 94.4% and 90.2% in the 123- and 160-bp ISs, respectively. An IS in the fourth intron of *Wknox1a* was 49-bp and that of *Wknox1d* was 122-bp in length. An identity of TIRs was lower in the 122-bp IS (86.0%) than in the 123- and 160-bp ISs, and only 12-bp TIR could be recognized in the 49-bp IS. A short TA-dinucleotide repeat (a target site duplication), which was characteristic to the *Stowaway* family, was found at the ends of all these ISs. The results indicated that these four ISs belonging to the *Stowaway* family of MITE was active at least after differentiation of the A, B and D genomes. One tandem duplication was found in the fourth intron of *Wknox1d* and one in the 3' non-coding region of *Wknox1b*. These duplications were respectively 50- and 22-bp in length.

A large number (538) of SNPs were found among the three *Wknox1* loci. Most of the SNPs were in the intron regions, especially in the fourth introns (Table 3). The observed SNPs were considered to reflect the degree of differentiation of the A, B and D genomes derived from a common ancestral genome. All base changes within the indels belonged to the unclassified base changes, whereas a total of 444 genome-specific SNPs were recognized among the three *Wknox1* loci (Table 3). A majority (231) of genome-specific SNPs was found in the *Wknox1b* locus, and the number of SNPs in *Wknox1b* was more than the sum of SNPs in *Wknox1a* and *Wknox1d*.

A total of 558 mutation events excluding the unclassified base changes were detected in the total length of the three *Wknox1* sequences (22,643-bp). Thirty-eight mutations were in the exon regions (4,501-bp) with an average mutation rate of 0.0084. In the intron regions (17,426-bp), 521 mutations were found with an average mutation rate of 0.030. The mutation rate in the

intron regions thus was more than 3.5-times higher than that in the exon regions. The mutation rates in the 3' non-coding region were higher than those in the open reading frame at all three loci. The average mutation rate of *Wknox1b* (0.041) was the highest among the three homoeologs in both exon and intron regions. The mutation rate of *Wknox1a* (0.025) was higher than that of *Wknox1d* (0.013). Most of the mutation events were concentrated in the fourth introns, but the magnitude of mutation rate in the fourth introns was similar to those in the fifth introns and the total intron regions (Table 3). This result indicated that the number of mutation events was largely dependent on the intron length.

The upstream-half regions of the fourth introns of the three *Wknox1* loci showed high homology to the corresponding region of the barley *HvKnox3* (accession number X83518). On the other hand, the downstream-half regions showed much less homology. In the upstream-half region of the fourth intron, PHO-binding sites (ATGGC) and $(GA/TC)_n$ repeats present in *HvKnox3* (Santi *et al*., 2003) were conserved in the fourth introns of all three *Wknox1* loci (Fig. 2B). The repeat number of the $(GA/TC)_n$ sequence varied among the three *Wknox1* homoeologs and *HvKnox3*. Insertion sites of a *Mutator* transposable element were assigned to this intron region in the maize *kn1* (Vollbrecht *et al*., 1993), and a 305-bp tandem duplication occurred in the same region of the barley *HvKnox3* (Müller *et al*., 1995). These insertion and duplication events resulted in the dominant *Kn1* and *K* mutations of maize and barley, respectively.

In the 5'-upstream regions of the *Wknox1* loci, sequence alignment could be constructed within ca. 1.5-kb regions between *Wknox1a* and *Wknox1b*, while homology between these and *Wknox1d* was found only within a 680-bp upstream region. The loss of homology in the upstream region of *Wknox1d* indicated a recombination at this boarder in this locus. In the 680-bp upstream region, two ISs longer than 6-bp were found in *Wknox1d*. A 30-bp tandem

duplication and a 144-bp IS were recognized in the upstream region of *Wknox1a* compared with the *Wknox1b* sequence.

Allocations of the structural mutations in the evolutional lineage of the polyploid wheat genome

To allocate the major structural mutations in the evolutional lineage of the polyploid wheat genome, PCR analyses were conducted using gene-specific primer sets to detect polymorphisms among the three *Wknox1* homoeologs. In this study, accessions from putative diploid donor species with A or S (B) genomes and tetraploid and hexaploid wheat were used. We expected that polymorphisms in the PCR-amplified fragments among species and/or accessions should enable us to assign structural mutations to the postulated lineage in the wheat genome evolution. Fig. 3A shows, as an example, the result of PCR amplification of the 123-bp MITE insertion in the fourth intron.

Six insertion events including one tandem duplication and one recombination event were mapped into the history of wheat genome evolution (Fig. 3B). Two MITE insertion events (123-bp and 160-bp) were observed in all accessions of *T. dicoccoides* but not in einkorn wheat (A genome) or Sitopsis species (S and S-related genomes), inferring that the MITE insertion occurred in the fourth intron of *Wknox1b* after allotetraploidization between *T. urartu* and *Ae. speltoides*. The 157-bp insertion in the fifth intron of *Wknox1b* was found in all accessions of the cultivated tetraploid (*T. dicoccum* and *T. durum*) and common wheat but not in accessions of the wild tetraploid wheat (*T. dicoccoides*). This result suggested that the 157-bp insertion occurred in the process of domestication of emmer wheat, although more accessions of tetraploid wheat should be studied. A 144-bp insertion and a tandem duplication in the 5'-upstream region of *Wknox1a* were found in *T. urartu*, and a 122-bp MITE insertion in the fourth intron and a

recombination in the 5'-upstream region of *Wknox1d* was found in *Ae. squarrosa*. These mutation events likely occurred in the differentiation process of A and D genomes from an ancestral diploid genome. It was notable that the tandem duplication in *T. urartu* accessions was not found in other A genome diploid (Einkorn) species, *T. boeoticum* and *T. monococcum.* The tandem duplication event therefore likely occurred after the divergence of *T. urartu* from the other Einkorn species. All polyploid wheat accessions examined in this study possessed this tandem duplication, supporting the view that the A genome of polyploid wheat was derived from *T. urartu* (Dvorák *et al*., 1993; Takumi *et al*., 1993).

Comparison of Wknox1 *expression patterns among the three homoeologs*

Because of the significant number of mutations in the three *Wknox1* homoeologs, we compared their expression patterns in different wheat organs. Based on the genome-specific polymorphisms in the 3' non-coding regions (Table 3), we designed three reverse primers to distinguish the three *Wknox1* homoeologs-derived transcripts (Fig. 4A). The specificity of the primers was confirmed by PCR amplification of the three homoeologs with the subcloned plasmid DNAs (data not shown). All three homoeologous transcripts showed quite similar accumulation patterns after RT-PCR (Fig. 4B). Young spikes and SAM-containing shoots seemed to show higher expression levels of the three homoeologs than in paleas and pistils. No transcripts were detected in lemmas. Intensities of the amplified fragments did not differ after 25 and 30 cycles of the PCR among the three homoeologs. The result indicated that the three *Knox1* homoeologs retained the similar tissue-specific expression patterns.

To study in more detail tissue-specific expression of the three homoeologous *Wknox1* loci, we constructed three chimeric GUS reporter genes under control of the 5'-upstream regions of

Wknox1a, *Wknox1b* and *Wknox1d* (Fig. 5A), and introduced them into the tobacco genome. More than ten transgenic plants were recovered for each construct. GUS assay was performed in the kanamycin-resistant plants at T_1 generation. Strong GUS expression was observed in the shoot apices and weak expression in the vascular systems (Fig. 5B). This expression pattern of the *Wknox1::GUS* constructs was consistent with the results obtained from *in situ* hybridization of *Wknox1* mRNA (Fig. 1A) and RT-PCR of the three homoeologous transcripts (Fig. 4). Identical patterns of the GUS expression in the three homoeologous constructs indicated that no structural mutations influenced the SAM-specific expression pattern of the *Wknox1* homoeologs.

The upstream of *Wknox1d* excluding the 680-bp region showed no homology with the two other homoeologs or with any other sequences deposited in database, implying that a recombination occurred at this border. To examine the promoter function of the 680-bp upstream sequence, it was fused with the *GUS* gene (Fig. 5A) and the *Wknox1d680::GUS* construct was introduced into tobacco. Sixteen transgenic plants were recovered, and the GUS assay was performed in the kanamycin-resistant plants at T_1 generation. The GUS expression pattern of the *Wknox1d680::GUS* construct was identical to that of the *Wknox1::GUS* constructs (Fig. 5B). The result indicated that the recombination at the 680-bp border in *Wknox1d* had no detectable influence on its expression patterns, and that this region contained essential *cis*-elements required for the SAM-specific expression of *Wknox1*.

Discussion

The maize *kn1*-type homeobox genes play essential roles in the maintenance of SAM, determination of cell fates and differentiation of vegetative tissues (Kerstetter *et al*., 1997). They

show a highly conserved structure in their coding regions (exons) but possess large introns (more than 5-kbp in length) with many structural changes such as tandem duplications and transposon/retroposon insertions (Vollbrecht *et al*., 1993; Müller *et al*., 1995). Our previous study of three homoeologous *Wknox1* cDNA sequences isolated from a Japanese common wheat cv. 'Norin 26' suggested that they were orthologs of the maize *kn1*, rice *OSH1* and barley *HvKnox3* (*Bkn3*) genes because of their sequence homology and chromosomal location (Takumi *et al*., 2000). We now proven the functional orthology of these *Wknox1* homoeologs to the *kn1*-type homeobox genes by their SAM-specific expression pattern in wheat and the morphological alterations induced by overexpression in transgenic tobacco plants (Fig. 1). Abundant accumulation of the *Wknox1* transcripts in the meristematic tissues agreed well with the characteristic expression pattern of *kn1* and *OSH1* (Kerstetter *et al*., 1994; Sentoku *et al*., 1999). Dramatic morphological alterations in transgenic tobacco plants by ectopic overexpression of *Wknox1* were similar to those reported in transgenic tobacco overexpressing maize *kn1*, barley *HvKnox3* and rice *OSH1* and *OSH15* (Sinha *et al*., 1993; Müller *et al*., 1995; Sato *et al*., 1996; Sato *et al*., 1998). These morphological changes were consistent with the reported ectopic meristem formation in maize *Kn1* mutant leaves (Kerstetter *et al*., 1994). We therefore considered that the *Wknox1* homoeologs should fulfil a requirement as target loci for studying the intragenic diversity that is associated with allopolyploid evolution of the wheat genome.

We isolated genomic sequences representing a set of three homoeologous *Wknox1* loci locating on the three homoeologous chromosomes of a common wheat cv. CS. A large number of structural mutations including indels, tandem duplications and SNPs were found in these loci (Table 3 and Fig. 2A). Most of the mutations were mapped within the longest fourth intron, of which a majority were in its downstream-half region. The result suggests that the upstream-half region of the fourth intron is more conserved than the downstream-half region. In fact, the

downstream-half region of the fourth intron lacks homology to that in the barley *HvKnox3* locus (Fig. 2B). Several large indel mutations including four MITE insertions occurred in this region. Using these polymorphic sites, a series of diploid species with A, S (B) or D genomes and polyploid wheat with AB and ABD genomes were studied to determine when such structural mutations occurred in the evolutionary lineages of the wheat genomes (Fig. 3). Seven mutation events were mapped in the history of wheat genome evolution and gave the following inferences. Two MITE-insertion events occurred after allotetraploidization between *T. urartu* and *Ae. speltoides* in the fourth intron of *Wknox1b*. A 157-bp insertion occurred in the process of domestication of emmer wheat. A 144-bp insertion and a tandem duplication in the 5'-upstream region of *Wknox1a* and a 122-bp MITE insertion in the 4th intron and a recombination in the 5'-upstream region of *Wknox1d* occurred in the differentiation processes of A and D genomes from an ancestral diploid genome. The unique presence of the tandem duplication in *T. urartu* among Einkorn species and the common presence of this mutation in the A genome of tetraploid and hexaploid wheat supported that the A genome of polyploid wheat was derived from *T. urartu* (Dvorák *et al*., 1993; Takumi *et al*., 1993). Our results obtained using the *Wknox1* homoeologs as a target gene demonstrate the usefulness of the intragenic diversity within the homoeologous loci containing long introns in the evolutionary studies of allopolyploid wheat. Effects of amphidiloidization and domestication on the accumulation of mutations should be further investigated using more homoeologous gene pairs possessing desirable characteristics.

We also addressed a question if the three *Wknox1* homoeologs went through different modes of diversification. The mutation rate observed in each one of the *Wknox1* homoeologs gave some inferences on this aspect. The mutation rate in the *Wknox1b* locus was the highest among the three *Wknox1* homoeologous loci. The number of indel events in the *Wknox1b* locus was also larger than in the *Wknox1a* and *Wknox1d* loci. The observed mutation rates among the three

homoeologous loci should reflect the levels of differentiation of the A, B and D genomes in the hexaploid genome evolved before and after allopolyploidization of ancestral diploid genomes. The highest mutation rate observed in the *Wknox1b* locus might be ascribed to the heterogeneity due to a high outcrossing frequency in *Ae. speltoides*, and/or to significant changes in the course of evolution of the S genome of *Ae. speltoides* to the B genome of polyploid wheat. An exact reason for such a high level of differentiation in the S and B genomes is unknown, although some hypotheses were proposed earlier (Zohary and Feldman, 1962; Furuta and Tanaka, 1970). The mutation rate in the *Wknox1a* locus was higher than that in the *Wknox1d* locus. *Ae. squarrosa* is widely distributed from Turkey to China, and its intraspecific DNA variation revealed by amplified fragment length polymorphism analysis is twice larger than that of *T. urartu*, which distributes in a restricted area from Lebanon to Armenia (Mizumoto *et al*., 2002). Because the interspecific genetic distance between *T. urartu* and *Ae. speltoides* is nearly equal to that between *T. urartu* and *Ae. squarrosa*, some mechanisms other than the genome-wide levels of differentiation in these ancestral diploid species are required for explaining this higher mutation rate in the *Wknox1a* locus than in the *Wknox1d* locus. Molecular analysis of the *Ae. squarrosa* population suggests that the modern forms of *T. aestivum* share a common D-genome genepool (Dvorák *et al.*, 1998). The bottleneck effect at allopolyploidization between tetraploid wheat and *Ae. squarrosa* seems to result in the lower mutation rate in the *Wknox1d*. Moreover, whole-genome duplication often allows accelerated gene evolution, which has recently been proven in yeast (Kellis *et al*., 2004). One plausible explanation might be that processes of allopolyploidization and/or domestication of polyploid wheat caused such differential alterations in the *Wknox1* homoeologs. Although we could not specify any mutation events in the *Wknox1a* locus after tetraploidization, we found three insertion events in the *Wknox1b* locus that likely occurred in the speciation process of *T. dicoccoides* and the domestication of cultivated emmer

wheat. In case of tobacco *NtGA* gene encoding an α subunit of heterotrimeric GTP-binding protein, one of the two homoeologous pairs accumulated more mutations in the evolutional process of this allotetraploid species (Takumi *et al*., 2002). To clarify the mode of genome differentiation after alloplolyploidization, more information is required on the differential mutation rates among homoeologous gene pairs containing long introns.

Another question we addressed in this study was related to the functional differentiation of the three *Wknox1* homoeologs. We compared their expression patterns by RT-PCR using the homoeolog-specific primer set and by using transgenic tobacco plants harboring the *Wknox1*-promoter::*GUS* fusion genes. All three *Wknox1* homoeologs showed the well conserved expression pattern characteristics to the cereal *KNOX* genes in different wheat organs (Fig. 4), suggesting that they are expressed under the same transcriptional regulation. The 5'-upstream region of the three *Wknox1* homoeologs also caused the conserved expression pattern in the transgenic tobacco plants (Fig. 5). The tissue-specificity of the *Wknox1* expression was controlled by *cis*-elements located within ca. 700-bp upstream region, which agreed with the previous report that the 655-bp promoter of *HvKnox3* controls the tissue-specific expression in barley (Santi *et al*., 2003). The highly conserved expression pattern indicates the functional identity among the three *Wknox1* homoeologs. In cases of maize *Kn1* and barley *HvKnox3*, structural mutations in the upstream-half region of the longest intron result in dominant mutations. The structural mutations induce their ectopic expression resulting in the formation of ectopic meristems on maize leaves and ectopic spikelets instead of awns in barley (Vollbrecht *et al*., 1993; Müller *et al*., 1995). The presence of some important regulatory elements was reported in the intron region (Santi *et al*., 2003). In the wheat *Wknox1* loci, however, no structural mutations could change their SAM-specific expression pattern and function of the homoeologs. The

putative amino acid sequences are either identical or highly homologous among the three WKNOX1 homoeologs (Takumi *et al*., 2000).

A wheat *Hooded* (*Hd*) mutation, one of the dominant awnless mutations, reduces the length of the awns and generally causes curved and twisted awns near the base (Watkins and Ellerton, 1940). The *Hd* locus is located on chromosome 4A of CS (Sears, 1954), which carries the *Wknox1a* homoeolog. The barley *Hooded* (*K*) mutants form additional spikelets at the top of lemmas or in the middle of awns by ectopic expression of the *HvKnox3* gene (Müller *et al*., 1995). In the *K* mutants, two additional and oppositely oriented spikelets are formed (Williams-Carrier *et al*., 1997). Wheat *Hd* allele on the short arm of chromosome 4A is considered to be orthologous to the *K* allele on the barley chromosome 4H (Linde-Laursen *et al*., 1997). Therefore, a question arises if the *Wknox1a* locus in CS represents the mutant allele responsible for the *Hd* mutation. No polymorphic sites that are potentially associated with the ectopic expression of the *Wknox1* loci were found in the exon and intron regions by comparison of the genomic sequences of the three *Wknox1* homoeologs, although a number of mutations were identified in the upstream-half region of the *Wknox1a* 4th intron. The three *Wknox1* homoeologs showed the well-conserved expression patterns in different wheat organs (Fig. 4) and their 5'-upstream regions induced the SAM-abundant expression in transgenic tobacco plants (Fig. 5). In spite of the chromosomal synteny between the barley *K* and wheat *Hd* mutations (Linde-Laursen *et al*., 1997), the *Wknox1a* gene is not likely a candidate gene for the *Hd* mutation. Unlike the barley *K* mutants, clear formation of ectopic meristems cannot be recognized readily in the wheat *Hd* mutants (Takumi *et al*., 2000), but an ectopic membranous tissue is occasionally formed at the lower region of the hooked awn in the *Hd* accessions (S. Takumi, R. Morimoto and C. Nakamura, unpublished observation). The wheat *Hd* phenotype resembles to the barley *K* phenotype, but the phenotypic effects observed in the ectopic tissues formed at the hooked awn in the *Hd* mutants

are much less pronounced than those in the *K* phenotype. Two explanations can be postulated for the reduced phenotypic effect of the *Hd* mutation if the *Hd* phenotype results from altered expression of some wheat class I type *KNOX* gene. It was reported that wheat plants carrying a single barley *K* allele on the monosomically added barley chromosome 4H developed a weak *K* phenotype, similar to the barley lines having a weaker allele of the *K* locus, K^e (Taketa and Takeda, 1997). This observation in barley together with the weak phenotype in wheat indicates that the severity of the mutant phenotype is affected not only by different mutant alleles but also by ploidy levels in the genetic background. Another explanation can be that the wheat *Hd* mutation is caused by altered expression of yet unknown class I type *KNOX* gene, which plays less essential roles in the SAM-formation and maintenance than the *Wknox1* loci. In maize and rice, the additional class I type KNOX genes, *Knox3* and *OSH3*, are present at the closely linked chromosomal regions of *kn1* and *OSH1*, respectively (Kerstetter *et al*., 1994; Sentoku *et al*., 1999). The rice *OSH3* locus is only 15-kb apart from that of *OSH1* (Sentoku *et al*., 1999). A number of mutations accumulated in the *OSH3* intron regions and two amino acid substitutions were found at invariant positions of its homeodomain region (Sato *et al*., 2001). Structural mutations within and/or near the *Wknox3a* locus, which is located at the vicinity of the *Wknox1a* locus (S. Takumi, R. Morimoto and C. Nakamura, unpublished result), should be further studied to clarify the relationship between the wheat *KNOX* genes and the *Hd* mutation.

Acknowledgements

We thank Dr. Y. Ogihara for the first screening of the TAC clones. We are grateful to Dr. K. Tsunewaki, Dr. M. Matsuoka, Dr. N. Mori, Dr. K. Murai and Dr. Y. Matsuoka for their kind

advices. We also thank the Plant Germ-Plasm Institute, Faculty of Agriculture, Kyoto University, Japan, the National Small Grains Research Facility, USDA-ARS, USA and Dr. E. Nevo, the Institute of Evolution, Haifa University, Israel, for providing us with the seed stocks used in this study. The work was supported in part by a 'grant-in-aid' from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to ST, no. 13760006). Sequence data from this article have been deposited in the DDBJ/EMBL/GenBank Data Libraries under accession nos. AB182943, AB182944 and AB182945 for *Wknox1a*, *Wknox1b* and *Wknox1d*, respectively. Contribution No. 168 from the Laboratory of Plant Genetics, Faculty of Agriculture, Kobe University.

References

Belyayev, A., Raskina, O., Korol, A. and Nevo, E. 2000. Coevolution of A and B genomes in allotetraploid *Triticum dicoccoides*. Genome 43: 1021-1026.

Bureau, T.E. and Wessler, S.R. 1994. *Stowaway*: A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6: 907-916.

Chapman, V., Miller, T.E. and Riley, R. 1976. Equivalence of the A genome of bread wheat with that of *Triticum urartu*. Genet. Res. 27: 69-76.

Chen, J.J., Janssen, B.J., Williams, A. and Sinha, N. 1997. A gene fusion at a homeobox locus: alterations in leaf shape and implications for morphological evolution. Plant Cell 9: 1289-1304.

Deyholos, M.K. and Sieburth, L.E. 2000. Separable whorl-specific expression and negative

regulation by enhancer elements within the *AGAMOUS* second intron. Plant Cell 12: 1799-1810.

Douglas, S.J., Chuck, G., Dengler, R.E., Pelecanda, L. and Riggs, C.D. 2002. *KNAT1* and *ERECTA* regulate inflorescence architecture in Arabidopsis. Plant Cell 14: 547-558.

Dvorák, J., Terlizzi, P.D., Zhang, H.B. and Resta, P. 1993. The evolution of polyploid wheats: identification of the A genome donor species. Genome 36: 21-31.

Furuta, Y. and Tanaka, M. 1970. Experimental introgression in natural tetraploid *Aegilops* species. Jpn J Genet. 45: 129-145.

Kashkush, K., Feldman, M. and Levy, A.A. 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160: 1651-1659.

Kashkush, K., Feldman, M. and Levy, A.A. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nature Genet. 33: 102-106.

Kawahara, T. 1997. *Catalogue of Aegilops-Triticum Germ-Plasm Preserved in Kyoto University.* Plant Germ-Plasm Institute, Kyoto University, Kyoto, Japan.

Kellis, M., Birren, B.W. and Lander, E.S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428: 617-624.

Kerstetter, R.A., Vollbrecht, E., Lowe, B., Yamaguchi, J. and Hake, S. 1994. Sequence analysis and expression patterns divide the maize *knotted1*-like homeobox genes into two classes. Plant Cell 6: 1877-1887.

Kerstetter, R.A., Laudencia-Chingcuanco, D., Smith, L.G. and Hake, S. 1997. Loss-of-function mutations in the homeobox gene, *knotted1*, are defective in shoot meristem maintenance. Development 124: 3045-3054.

Kihara, H. 1944. Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agric. Hort. Tokyo 19: 889-890 (in Japanese).

Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K. and Hake, S. 1994. A *knotted1*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. Plant Cell 6: 1859-1876.

Linde-Laursen, I., Heslop-Harrison, J.S., Shepherd, K.W. and Taketa, S. 1997. The barley genome and its relationship with the wheat genomes. A survey with an internationally agreed recommendation for barley chromosome nomenclature. Hereditas 126: 1-16.

Liu, B., Bega, J.M., Segal, G., Abbo, S., Rodoya, M. and Feldman, M. 1998a. Rapid genomic changes in newly synthesized amphidiploids of *Triticum* and *Aegilops*. I. Changes in low-copy non-coding DNA sequences. Genome 41: 272-277.

Liu, B., Bega, J.M. and Feldman, M. 1998b. Rapid genomic changes in newly synthesized amphidiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. Genome 41: 535-542.

Liu, Y.-G., Nagaki, K., Fujita, M., Kawaura, K., Uozumi, M. and Ogihara, Y. 2000. Development of an efficient maintenance and screening system for large-inserted genomic DNA libraries of hexaploid wheat in a transformation-competent artificial chromosome (TAC) vector. Plant J. 23: 687-695.

Long, J.A., Moan, E.I., Medford, J.I. and Barton, M.K. 1996. A member of the *KNOTTED* class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. Nature 379: 66-69.

McFadden, E.S. and Sears, E.R. 1944. The artificial synthesis of *Triticum spelta*. Rec. Genet. Soc. Am. 13: 26-27.

Matsuoka, M., Ichikawa, H., Saito, A., Tada, Y., Fujimura, T. and Kano-Murakami, Y. 1993. Expression of a rice homeobox gene causes altered morphology of transgenic plants. Plant Cell 5: 1039-1048.

Mizumoto, K., Hirosawa, S., Nakamura, C. and Takumi, S. 2002. Nuclear and chloroplast genome genetic diversity in the wild einkorn wheat, *Triticum urartu*, revealed by AFLP and SSLP analyses. Hereditas 137: 208-214.

Müller, K., Romano, N., Gerstner, O., Garcia-Maroto, F., Pozzi, C., Salamini, F. and Rohde, W. 1995. The barley *Hooded* mutation caused by a duplication in a homeobox gene intron. Nature 374: 727-730.

Murai, J., Taira, T. and Ohta, D. 1999. Isolation and characterization of the three Waxy genes encoding the granule-bound starch synthase in hexaploid wheat. Gene 234: 71-79.

Murashige, T. and Skoog, F. 1962 A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant. 15: 473-497.

Nevo, E., Golenberg, E. and Beiles, A. 1982. Genetic diversity and environmental accessions of wild wheat, *Triticum dicoccoides*, in Israel. Theor. Appl. Genet. 76: 321-332.

Ohno, R., Takumi, S. and Nakamura, C. 2001. Expression of a cold-responsive *Lt-Cor* gene and development of freezing tolerance during cold acclimation in wheat (*Triticum aestivum* L.). J. Exp. Bot. 52: 1-8.

Ozkan, H., Levy, A.A. and Feldman, M. 2001. Allopolyploidy-induced rapid genome evolution in the (*Aegilops-Triticum*) group. Plant Cell 13: 1735-1747.

Sanmiguel, P.J., Ramakrishna, W., Bennetzen, J.L., Busso, C.S. and Dubcovsky, J. 2002. Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome $5A^m$. Funct. Integr. Genomics 2: 70-80.

Santi, L., Wang, Y., Stile, M.R., Berendzen, K., Wanke, D., Roig, C., Pozzi, C., Müller, K., Müller, J., Rohde, W. and Salamini, F. 2003. The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene *Bkn3*. Plant J. 34: 813-826.

Sarkar, P. and Stebbins, G.L. 1956. Morphological evidence concerning the origin of the B genome in wheat. Am. J. Bot. 43: 297-304.

Sasakuma, T., Ogihara, Y. and Tsujimoto, H. 1993. Genome rearrangement of repetitive sequences in the processes of hybridization and amphidiploidization in Triticenae. *In*: Li, ZS. and Yin, Z.Y. (eds) Proc. 8th Int. Wheat Genet. Symp., Beijing. China Agricultural Scientech Press, Beijing, pp563-566.

Sasanuma, T., Miyashita, N.T. and Tsunewaki, K. 1996. Wheat phylogeny determined by RFLP analysis of nuclear DNA. 3. Intra- and interspecific variations of five *Aegilops* Sitopsis species. Theor. Appl. Genet. 92: 928-934.

Sato, Y., Tamaoki, M., Murakami, T., Yamamoto, N., Kano-Murakami, Y. and Matsuoka, M. 1996. Abnormal cell divisions in leaf primordia caused by the expression of the rice homeobox gene *OSH1* lead to altered morphology of leaves in transgenic tobacco. Mol. Gen. Genet. 251: 13-22.

Sato, Y., Sentoku, N. Nagato, Y. and Matsuoka, M. 1998. Isolation and characterization of a rice homeobox gene, *OSH15*. Plant Mol. Biol. 38: 983-998.

Sato, Y., Fukuda, Y. and Hirano, H. 2001. Mutations that cause amino acid substitutions at the invariant positions in homeodomain of OSH3 KNOX protein suggest artificial selection during rice domestication. Genes Genetic. Syst. 76: 381-392.

Schneeberger, R.G., Becraft, P.W., Hake, S. and Freeling, M. 1995. Ectopic expression of the *knox* homeo box gene *rough sheath1* alters cell fate in the maize leaf. Genes Dev. 9: 2292-2304.

Sears, E.R. 1954. *The aneuploids of common wheat.* Missouri Agricultural Experiment Station Research Bulletin 572: 59pp.

Sentoku, N., Sato, Y., Kurata, N., Ito, Y., Kitano, H. and Matsuoka, M. 1999. Regional expression of the rice KN1-type homeobox gene family during embryo, shoot, and flower development. Plant Cell 11: 1651-1663.

Sinha, N.R., Williams, R.E. and Hake, S. 1993. Overexpression of the maize homeo box gene, *Knotted-1*, causes a switch from determinate to indeterminate cell fates. Genes Dev. 7: 787-795.

Taketa, S. and Takeda, K. 1997. Expression of dominant marker genes of barley in wheat-barley hybrids. Genes Genet. Syst. 72: 101-106.

Takumi, S., Nasuda, S., Liu, Y.-G. and Tsunewaki, K. 1993. Wheat phylogeny determined by RFLP analysis of nuclear DNA. 1. Einkorn wheat. Jpn. J. Genet. 68: 73-79.

Takumi, S. and Shimada, T. 1996. Production of transgenic wheat through particle bombardment of scutellar tissues: frequency is influenced by culture duration. J. Plant Physiol. 149: 418-423.

Takumi, S., Kosugi, T., Murai, K., Mori, N. and Nakamura, C. 2000. Molecular cloning of three homoeologous cDNAs encoding orthologs of the maize KNOTTED1 homeobox protein from young spikes of hexaploid wheat. Gene 249: 171-181.

Takumi, S., M. Ida, Y. Haisa, S. Ando, and C. Nakamura, 2002. Genomic structure and homoeologous relationship of the two α subunit genes of a heterotrimeric GTP-binding protein in tobacco. Genome 45: 626-633.

Venglat, S.P., Dumonceaux, T., Rozwadowski, K., Parnell, L., Babic, V., Keller, W. and Martienssen, R. 2002. The homeobox gene *BREVIPEDICELLUS* is a key regulator of inflorescence architecture in *Arabidopsis*. Proc. Natl. Acad. Sci. USA. 99: 4730-4735.

Vollbrecht, E., Veit, B., Sinha, N. and Hake, S. 1991. The develeopmental gene *Knotted-1* is a member of a maize homeobox gene family. Nature 350: 241-243.

Vollbrecht, E., Kerstetter, R., Lowe, B., Veit, B. and Hake, S. 1993. Homeobox genes in plant development: mutational and molecular analysis. In: Sprading, A.C. (ed.), *Evolutionary Conservation of Developmental Mechanisms.* Wiley-Liss, New York, pp. 111-123.

Waterman, M.S., 1986. Multiple sequence alignment by consensus. Nucleic Acids Res. 14: 9905-9102.

Watkins, A.E. and Ellerton, S. 1940. Variation and genetics of the awn in *Triticum*. J. Genet. 40: 243-270.

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E. and Keller, B. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J. 26: 307-316.

Wicker, T., Guyot, R., Yahiaoui, N. and Keller, B. 2003a. CACTA transposons in Triticeae: A diverse family of high-copy repetitive elements. Plant Physiol. 132: 52-63.

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.-D., Dubcovsky, J. and Keller, B. 2003b. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15: 1186-1197.

Williams-Carrier, R.E., Lie, Y.S., Hake, S. and Lemaux, P.G., 1997. Ectopic expression of the maize *kn1* gene phenocopies the *Hooded* mutant of barley. Development 124: 3737-3745.

Zohary, D. and Feldman, M. 1962. Hybridization between amphidiploids and the evolution of polyploids in wheat (*Aegilops-Triticum*) group. Evolution 16: 44-61.

Legends of figures

Figure 1. Tissue-specific expression of the *Wknox1* gene in wheat and morphological alterations induced by *Wknox1* over-expression in transgenic tobacco plants. (A) *In situ* hybridization of the *Wknox1* mRNA in the longitudinal section of a shoot apex of a one-day-old wheat seedling. The *Wknox1* mRNA was detected in the wild-type SAM. Sections were hybridized to a digoxigenin-labeled *Wknox1* antisense probe. C, coleoptile; P0, leaf primordia. (B to I) Transgenic tobacco plants with the *35S::Wknox1* chimeric gene. (B, D) Flower and leaf morphology of a wild-type tobacco plant. (C, E) Flower and leaf morphology of a transformant with a mild phenotype. (F) A lobbed leaf from a transformant with a severe phenotype. (G, H) Ectopic leaves and shoots formed on the adaxial leaf surface of a transformant with a severe phenotype. (I) *In situ* hybridization of the *Wknox1* mRNA in a leaf section of a transformant with a severe phenotype. An arrow head indicates an ectopic shoot.

Figure 2. Major structural mutations in the three *Wknox1* homoeologous loci. (A) The mutation sites of indels with more than 6-bp in length. The filled triangles represent insertion sites with their length shown above. Four MITE sequences and three tandem duplications are indicated above the filled triangles. The deletion sites are indicated by open triangles with their deletion lengths. The site between bold and thin lines in the 5'-upstream region of *Wknox1d* indicates a recombination site. (B) Comparison of putative regulatory elements in the upstream-half regions of the 4th intron among the *HvKnox3* and *Wknox1* loci. The PHO-binding sites and $(GA/TC)_n$ repeats are shown. A shaded box indicates the 305-bp tandem duplication in the barley *K* mutation.

Figure 3. Allocations of insertion, tandem duplication and recombination events in the three *Wknox1* homoeologous loci through the evolutionary lineage of the wheat genome. (A) PCR amplification of the 123-bp MITE insertion in the 4th intron of *Wknox1b*. Total DNA was extracted from leaves of diploid, tetraploid and hexaploid species of *Triticum* and *Aegilops*. Arrows indicate PCR-amplified fragments. M; 100-bp ladder size marker, nc; negative control, Ldn; *T. durum* cv. 'Langdon', CS; *T. aestivum* cv. 'Chinese Spring'. (B) Postulated lineages of six insertion events and one recombination event in the 4th introns and 5'-upstream regions of the *Wknox1* loci allocated in the wheat genome evolution.

Figure 4. RT-PCR analysis of the *Wknox1* transcripts from the three homoeologous loci. (A) Alignment of the primer-containing regions in the three *Wknox1* homoeologs. Primer locations are shown by boldface letters. (B) Comparison of the *Wknox1a, Wknox1b* and *Wknox1d* transcript levels in several wheat organs. The amplified RT-PCR products were 548-bp in length. Total RNAs were extracted from lemmas, paleas, young spikes, SAM-enriched shoots, and small (S) and large (L) pistils. The ubiquitin gene was used as an internal control. Numbers of PCR cycles are shown at the right side of the electrophoresis gel.

Figure 5. Functional comparison of the 5'-upstream sequences of the three *Wknox1* homoeologs in transgenic tobacco plants. (A) Schemes of the four GUS gene constructs under control of the 5'-upstream regions of the *Wknox1* homoeologs. The *Wknox1a::GUS, Wknox1b::GUS, Wknox1d::GUS* constructs contained about 1.5 to 2.0-kbp 5'-upstream

sequences, the first exons and introns, and a part of the second exons. The 5'-upstream sequence of the *Wknox1d680::GUS* contained a 680-bp upstream region showing high homology among the three homoeologs. (B) Histochemical localization of the GUS activity near the shoot apices of the transgenic tobacco plants (T_1 generation).

Table 1. Accession numbers and sources of plant materials used in this study

T. urartu (3 accessions): KU199-1, KU199-6, KU199-11	
T. monococcum (2 accessions): KU105, KU3636	
T. boeoticum (3 accessions): KU101-2, KU101-3, KU103	
Ae. speltoides (4 accessions): KU1-2, KU1-3, KU2-4, KU2-5	
Ae. bicornis (2 accessions): KU3-2, KU3-3	
Ae. longissima (1 accession): KU4-5	
Ae. searsii (2 accessions): KU-4-6, KU4-7	
Ae. sharonensis (2 accessions): KU5-2, KU5-3	
Ae. squarossa (5 accessions): KU20-1, KU20-3, KU20-6, KU20-9, KU20-10	
T. dicoccoides (10 accessions): KU1945, KU1952, KU1959B, KU1978B, KU8935, KU8736A, KU8817, N (pop2), N (pop3), N (pop4)	
T. dicoccum (9 accessions): KU7309, KU1582, KU10497, KU10501, Citr4013, PI94628, PI27531, KU1063A, KU3371	
T. durum cv. ‘Langdon’	
T. aestivum cv. ‘Chinese Spring’, ‘S-615’	

KU: Plant Germ-Plasm Institute, Faculty of Agriculture, Kyoto University, Japan

PI: National Small Grains Research Facility, USDA-ARS, USA

Citr: National Small Grains Research Facility, USDA-ARS, USA

N: E. Nevo, Institute of Evolution, University of Haifa 31999, Israel.

For details of collection localities, see Kawahara (1997) and Nevo *et al*. (1982).

Table 2. Nucleotide sequences of primer sets used in the investigation of structural mutations in the *Wknox1* loci of *Triticum* and *Aegilops* species.

Mutation	Region	Locus	Forward (F) and reverse (R) primer sequences	length of the expected product	annealing temp.
tandem	5' up	*Wknox1a*	F; 5'-CATGTTGAGTGGACCTTTGA-3'	411-bp	53˚C
			R; 5'-CTGCAAGTACTGGGTTCAAG-3'		
144bp-in	5' up	*Wknox1a*	F; 5'-CTTGAACCCAGTACTTGCAG-3'	333-bp	55˚C
			R; 5'-TTATCGTATTGTCGCATCGT-3'		
recombi	5' up	*Wknox1d*	F; 5'-GCACCCCACCAATAGACCCA-3'	306-bp	55˚C
			R; 5'-TCCACACAGCCCTGCAAGAA-3'		
123bp-in	4th int	*Wknox1b*	F; 5'-ACAGATATCACGAGTTTCCG-3'	520-bp	58˚C
			R; 5'-TGGACATGGTATTGTATGCA-3'		
160bp-in	4th int	*Wknox1b*	F; 5'-AATCCATAATCATTTGCTTC-3'	840-bp	55˚C
			R; 5'-ATACAAGACTTGGAGAAAGC-3'		
122bp-in	4th int	*Wknox1d*	F; 5'-AAAAAAAAGGTTAAATGGAC-3'	400-bp	55˚C
			R; 5'-ACCTTATACATGATTGGGAA-3'		
157bp-in	5th int	*Wknox1b*	F; 5'-GCTGAAGCACCATCTCCTGA-3'	589-bp	63˚C
			R; 5'-CATGTAGAAGGCGGCGTTAG-3'		

in; insertion, tandem; tandem duplication, recombi; recombination, int; intron

Table 3. Summary of the comparative analysis of genomic sequences and the mutation rates in the three homoeologous *Wknox1* loci of common wheat.

Region	Locus	Nucleotide length (bp)	Genome specific SNPs	Insertion*	Deletion*	Total number of mutation events	Mutation rate
1st exon	*Wknox1a*	26	-	-	-	0	0
	Wknox1b	26	-	-	-	0	0
	Wknox1d	26	1	-	-	1	0.038
1st intron	*Wknox1a*	92	-	1	-	1	0.011
	Wknox1b	91	-	-	-	0	0
	Wknox1d	91	-	-	-	0	0
2nd exon	*Wknox1a*	458	5	-	1	6	0.013
	Wknox1b	461	-	-	-	0	0
	Wknox1d	461	-	-	-	0	0
2nd intron	*Wknox1a*	199	2	-	-	2	0.010
	Wknox1b	199	3	-	-	3	0.015
	Wknox1d	202	5	1	-	6	0.030
3rd exon	*Wknox1a*	123	1	-	-	1	0.008
	Wknox1b	123	4	-	-	4	0.033
	Wknox1d	123	-	-	-	0	0
3rd intron	*Wknox1a*	104	-	-	-	0	0
	Wknox1b	100	3	-	1	4	0.04
	Wknox1d	104	-	-	-	0	0
4th exon	*Wknox1a*	145	-	-	-	0	0
	Wknox1b	145	-	-	-	0	0
	Wknox1d	145	-	-	-	0	0
4th intron	*Wknox1a*	5227	114	12 (3)	15 (4)	141	0.027
	Wknox1b	5104	200	21 (4)	38 (8)	259	0.051
	Wknox1d	5418	70	4 (3)	11 (3)	85	0.016
5th exon	*Wknox1a*	233	1	-	-	1	0.004
	Wknox1b	233	5	-	-	5	0.021
	Wknox1d	233	1	-	-	1	0.004
5th intron	*Wknox1a*	117	3	-	1	4	0.034
	Wknox1b	276	11	3 (1)	1	15	0.054
	Wknox1d	118	1	-	-	1	0.008
	Wknox1a	204	-	-	-	0	0

6th exon	*Wknox1b*	204	2	-	-	2	0.010
(to stop codon)	*Wknox1d*	204	1	-	-	1	0.005
6th exon	*Wknox1a*	315	7	-	-	7	0.022
(from stop codon	*Wknox1b*	294	3	1 (1)	1 (1)	5	0.017
to poly(A) site)	*Wknox1d*	319	2	1 (1)	1	4	0.013
	Wknox1a	1189	7	-	1	8	0.007
total exon	*Wknox1b*	1192	11	-	-	11	0.009
(to stop codon)	*Wknox1d*	1192	3	-	-	3	0.003
	Wknox1a	1504	14	-	1	15	0.010
total exon	*Wknox1b*	1486	14	1 (1)	1 (1) **	16	0.011
(to poly(A) site)	*Wknox1d*	1511	5	1 (1)	1	7	0.005
	Wknox1a	5723	119	13 (3)	16 (4)	148	0.026
total intron	*Wknox1b*	5770	217	24 (5)	40 (8)	281	0.049
	Wknox1d	5933	76	5 (3)	11 (3)	92	0.016
	Wknox1a	7243	133	13 (3)	17 (4)	163	0.023
total length	*Wknox1b*	7256	231	25 (6)	41 (9)	297	0.041
	Wknox1d	7444	80	6 (4)	12 (3)	98	0.013

*: The numbers of indel events of more than 6-bp length sequences are shown in parentheses.

**: This mutation was a change of poly(A) binding site.

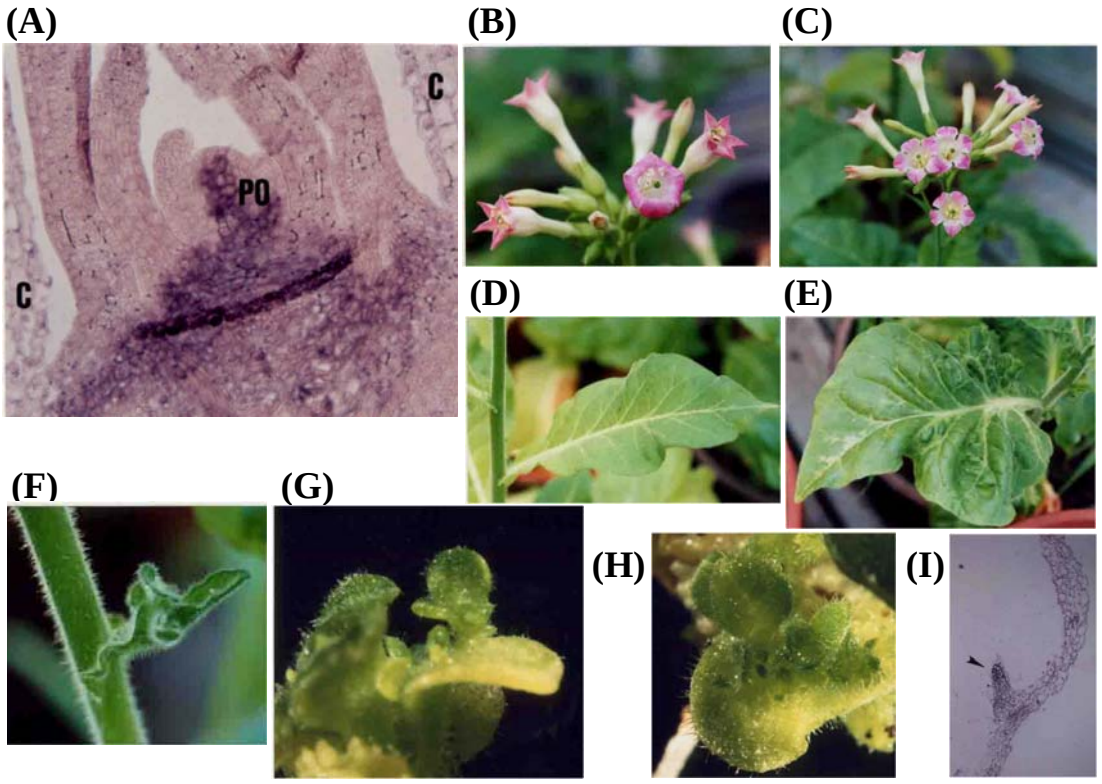


Fig. 1 (Morimoto et al.)

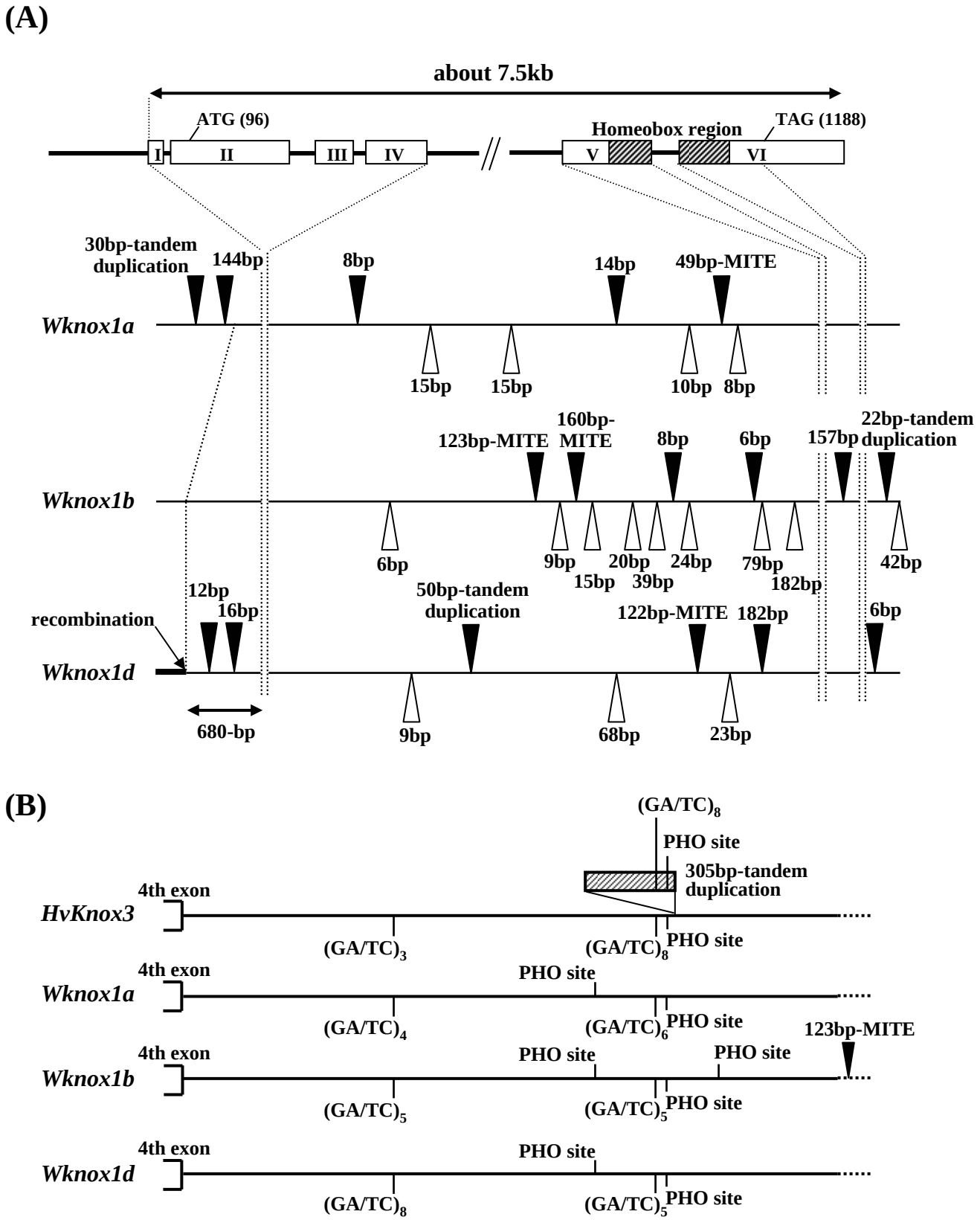
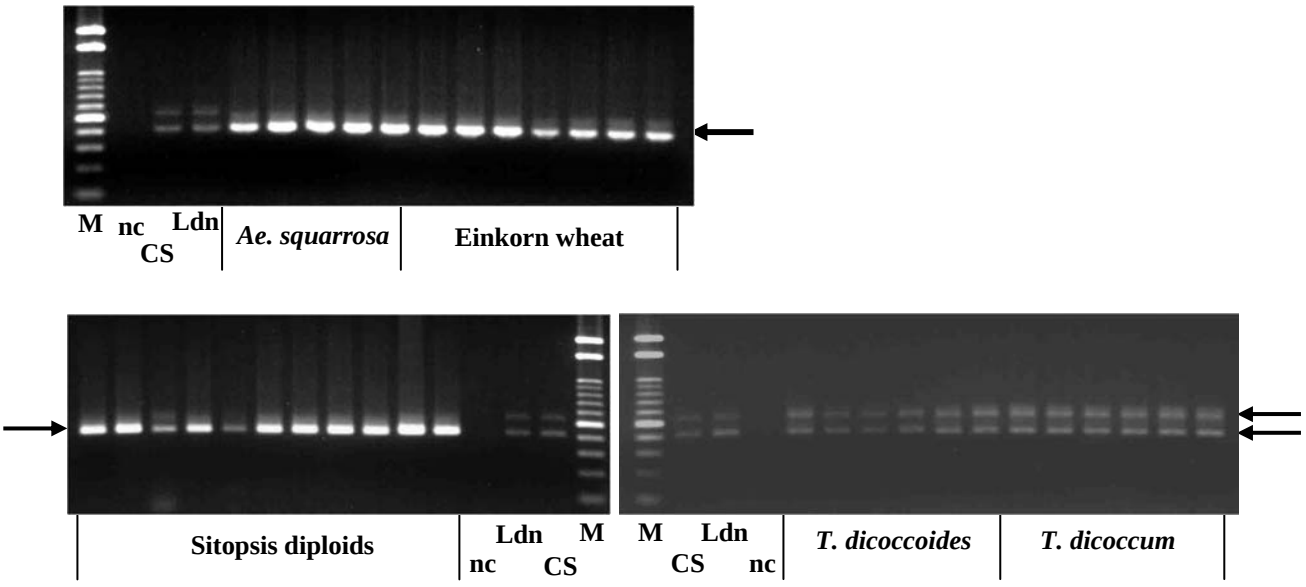


Fig. 2 (Morimoto et al.)

(A)



(B)

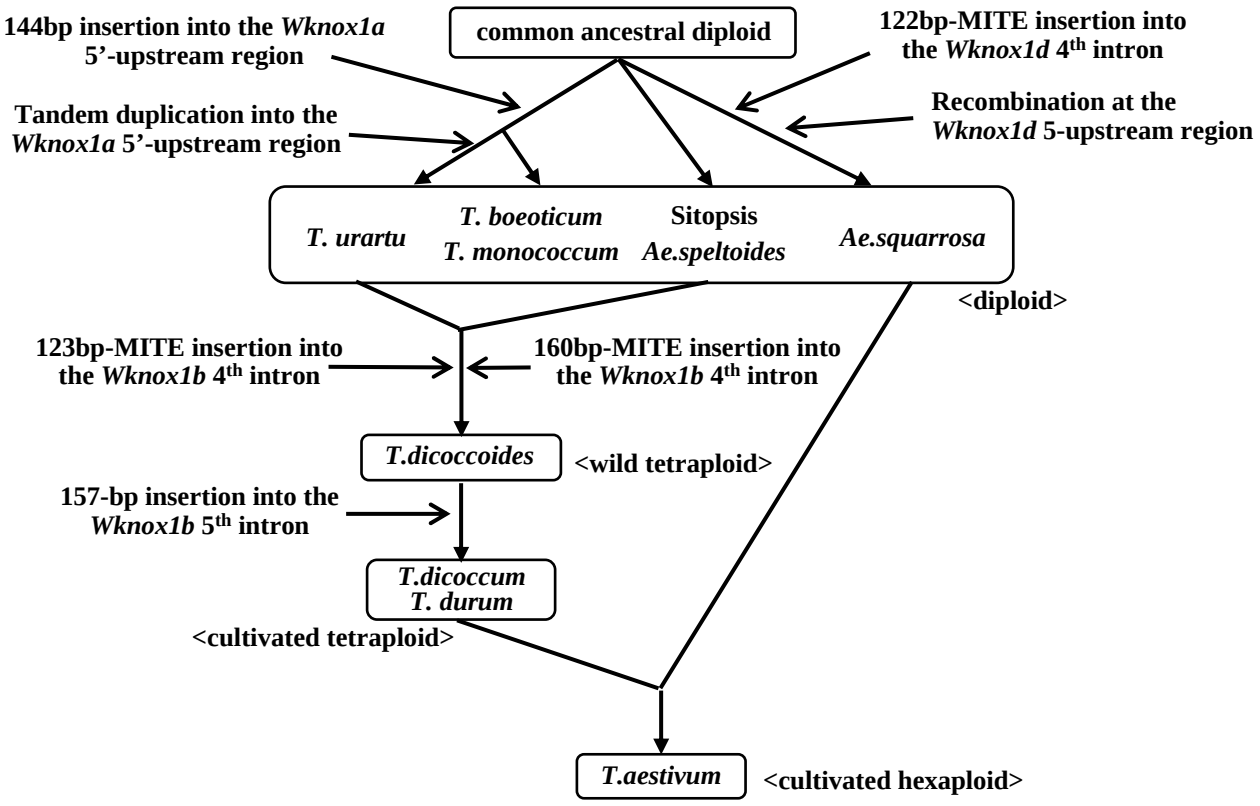
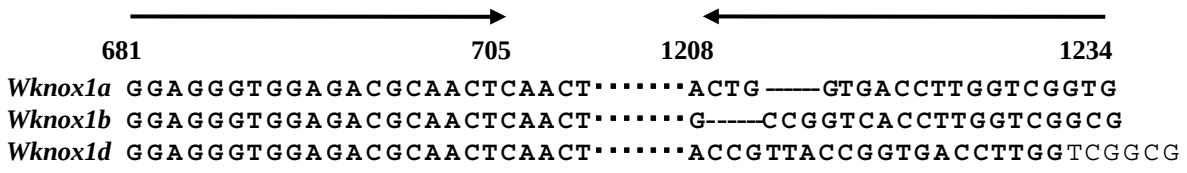


Fig. 3 (Morimoto et al.)

(A)



(B)

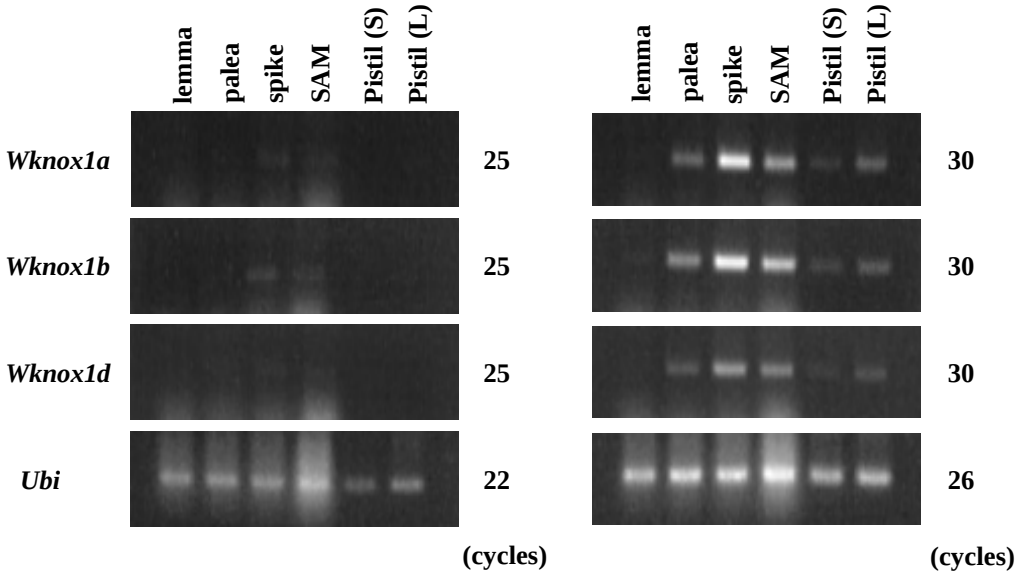


Fig. 4 (Morimoto et al.)

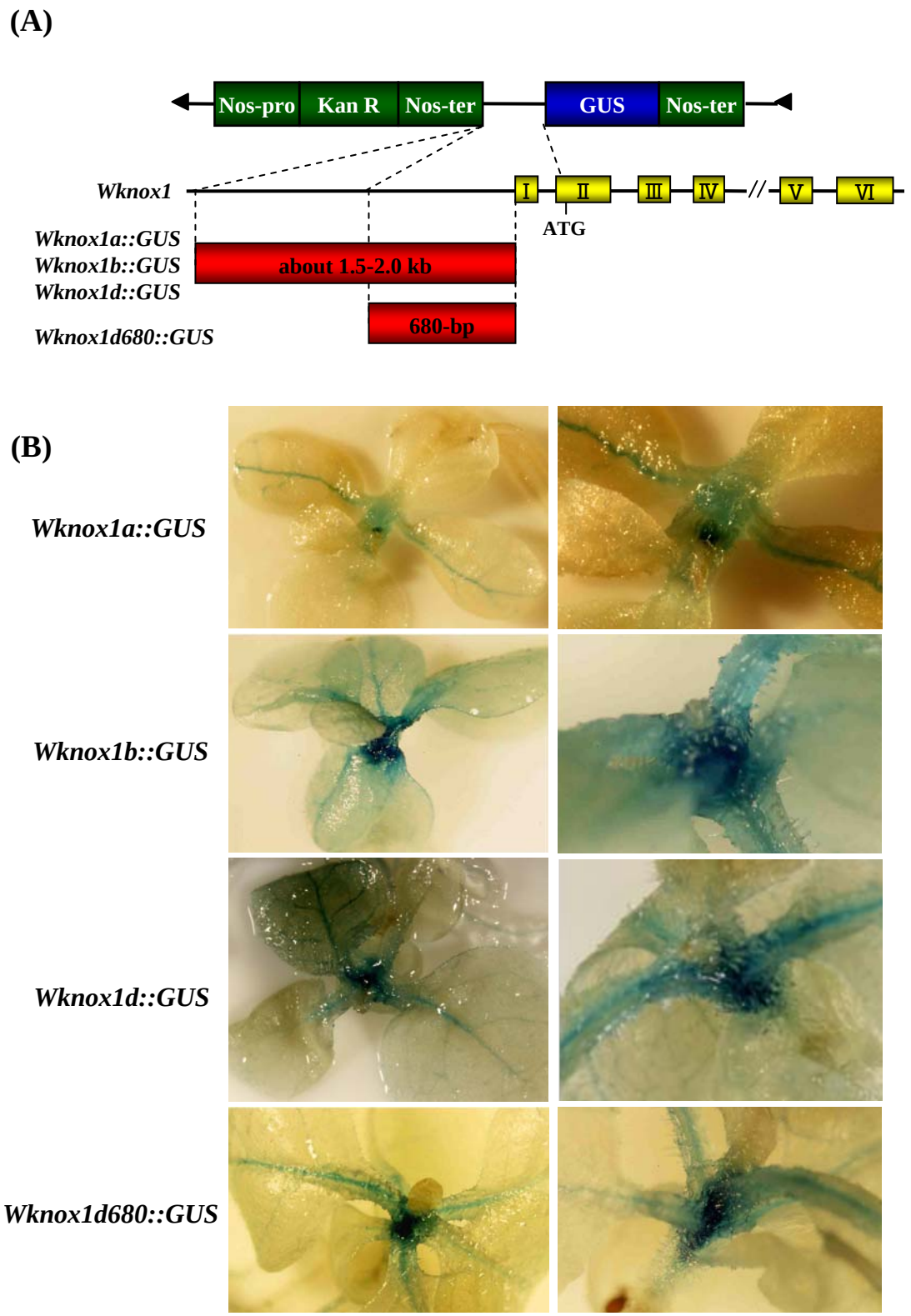


Fig. 5 (Morimoto et al.)