# Genome-Wide Characterization of the HD-ZIP IV Transcription Factor Family in Maize: Preferential Expression in the Epidermis[1][C][W]

**Marie Javelle, Catherine Klein-Cosson, Vanessa Vernoud, Véronique Boltz, Chris Maher, Marja Timmermans, Nathalie Depège-Fargeix, and Peter M. Rogowsky***

Université de Lyon, Ecole Normale Supérieure de Lyon, Université Lyon 1, Unité Reproduction et Développement des Plantes, F–69364 Lyon, France (M.J., C.-K.C., V.V., V.B., N.D.-F., P.M.R.); INRA, UMR879 Reproduction et Développement des Plantes, F–69364 Lyon, France (M.J., C.-K.C., V.V., V.B., N.D.-F., P.M.R.); CNRS, UMR5667 Reproduction et Développement des Plantes, F–69364 Lyon, France (M.J., C.-K.C., V.V., V.B., N.D.-F., P.M.R.); and Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 (M.J., C.M., M.T.)

Transcription factors of the plant-specific homeodomain leucine zipper IV (HD-ZIP IV) family have been found from moss to higher plants, and several family members have been associated with epidermis-related expression and/or function. In maize (*Zea mays*), four of the five characterized HD-ZIP IV family members are expressed specifically in the epidermis, one contributes to trichome development, and target genes of another one are involved in cuticle biosynthesis. Assessing the phylogeny, synteny, gene structure, expression, and regulation of the entire family in maize, 12 novel *ZmHDZIV* genes were identified in the recently sequenced maize genome. Among the 17 genes, eight form homeologous pairs duplicated after the split of maize and sorghum (*Sorghum bicolor*), whereas a fifth duplication is shared with sorghum. All 17 *ZmHDZIV* genes appear to be derived from a basic module containing seven introns in the coding region. With one possible exception, all 17 *ZmHDZIV* genes are expressed and show preferential expression in immature reproductive organs. Fourteen of 15 *ZmHDZIV* genes with detectable expression in laser-dissected tissues exhibit a moderate to very strong expression preference for the epidermis, suggesting that at least in maize, the majority of HD-ZIP IV family members may have epidermis-related functions. Thirteen *ZmHDZIV* genes carry conserved motifs of 19 and 21 nucleotides in their 3′ untranslated region. The strong evolutionary conservation and the size of the conserved motifs in the 3′ untranslated region suggest that the expression of *HD-ZIP IV* genes may be regulated by small RNAs.

Are transcription factor families expected to have conserved functions? At first sight, the answer seems to be no, because the classification of transcription factor families is based on DNA-binding motifs (Pabo and Sauer, 1992), and there is no obvious reason to postulate a link between a particular type of DNA-protein interaction and a biological function. On the other hand, the enormous expansion of particular transcription factor families in certain eukaryotic taxa suggests that at least some families evolved from a common ancestor, providing a basis for a possible functional conservation among extant family members. In the case of the homeodomain (HD), a 60-amino acid DNA-binding domain found in all eukaryotic organisms, the different families containing this domain in plants regulate very diverse biological processes, but there seems to be at least some functional conservation within each of these families (Ariel et al., 2007).

The 14 distinct HD-containing families in plants are defined by the presence of additional functional protein domains (Mukherjee et al., 2009). Four of them are regrouped in the HD-ZIP superfamily, defined by the association of the HD with a Leu zipper (ZIP) domain, a combination found exclusively in the plant kingdom (Schena and Davis, 1992). The HD-ZIP families III and IV are defined by the presence of two additional domains, the steroidogenic acute regulatory protein-related lipid transfer (START) domain and the START adjacent domain (SAD). These two families can be distinguished by a fifth domain, the C-terminal MEKHLA motif, which is present in the HD-ZIP III family and absent in the HD-ZIP IV family (Mukherjee and Bürglin, 2006). The latter family might present an example for an evolutionarily conserved transcription factor family with shared contributions to epidermal development (Javelle et al., 2011).

HD-ZIP IV transcription factors are likely present in all land plants. They have been described specifi-

[1] This work was supported by the National Science Foundation (grant no. 0820610 to M.T.) and by Ph.D. fellowships of the French Ministry of Higher Education (to M.J. and C.K.-C.).

* Corresponding author; e-mail peter.rogowsky@ens-lyon.fr.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Peter M. Rogowsky (peter.rogowsky@ens-lyon.fr).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[W] The online version of this article contains Web-only data.

www.plantphysiol.org/cgi/doi/10.1104/pp.111.182147

 

cally in the dicots Arabidopsis (*Arabidopsis thaliana*; Nakamura et al., 2006), sunflower (*Helianthus annuus*; Valle et al., 1997), cotton (*Gossypium hirsutum*; Guan et al., 2008), tomato (*Solanum lycopersicum*; Isaacson et al., 2009), and apple (*Malus domestica*; Dong et al., 1999), in the monocots maize (*Zea mays*; Ingram et al., 2000), rice (*Oryza sativa*; Ito et al., 2003), sorghum (*Sorghum bicolor*; Swigonová et al., 2004), and *Phalaenopsis* (Nadeau et al., 1996), in the gymnosperm Norway spruce (*Picea abies*; Ingouff et al., 2003), the lycophyte *Selaginella*, and the bryophyte *Physcomitrella* (Prigge and Clark, 2006). A comprehensive analysis of 10 plant genomes established the number of genes belonging to the HD-ZIP IV family in the genomes of Arabidopsis, poplar (*Populus* spp.), rice, maize, moss, and spike moss but failed to detect any *HD-ZIP IV* genes in the genomes of three green algae and one red alga (Mukherjee et al., 2009).

Interestingly, expression of nearly all *HD-ZIP IV* genes studied so far is associated with the epidermis in general or with specific epidermal structures. In Arabidopsis, the founding member of the HD-ZIP IV family, *GLABRA2* (*GL2*), is expressed specifically in leaf trichomes and in root atrichoblast cells (Rerie et al., 1994; Masucci et al., 1996), whereas Arabidopsis *MERISTEM LAYER1* (*AtML1*) and *PROTODERMAL FACTOR2* (*PDF2*) are both expressed throughout the epidermal layer of numerous organs (Lu et al., 1996; Abe et al., 2003). In contrast, *ANTHOCYANINLESS2* (*ANL2*) is mainly expressed in the subepidermal layer of the vegetative meristem, rosette leaves, and seedling roots (Kubo et al., 2008). An exhaustive characterization of the 12 remaining Arabidopsis *HD-ZIP IV* genes by promoter::*GUS* fusions revealed that the promoters of *HOMEODOMAIN GLABRA1* (*HDG1*), *HDG2*, *HDG5*, *HDG7*, *HDG11*, and *HDG12* direct preferential expression in the epidermal layer of shoot organs (Nakamura et al., 2006). In rice, the five *Rice outermost cell-specific* genes for which detailed expression data are available are all specifically expressed in the epidermis (Ito et al., 2002, 2003). Similarly, in maize, four of the five characterized *Outer cell layer* (*OCL*) genes (*OCL1*, *OCL3*, *OCL4*, and *OCL5*) are almost exclusively expressed in the epidermal layer, whereas *OCL2* expression is restricted to the subepidermal layer of floral meristems (Ingram et al., 1999, 2000). Expression data from other species are sparser, but *PaHB1* (for homeobox 1) in Norway spruce (Ingouff et al., 2001) provides yet another example of an HD-ZIP IV family member with epidermis-specific expression, whereas *Ovule39* of *Phalaenopsis*, with broad expression in all ovule cell types, presents a counter example (Nadeau et al., 1996).

Functional data based on the phenotypic characterization of *HD-ZIP IV* mutants or transgenic knockouts have been obtained in Arabidopsis, maize, and tomato, and the great majority substantiate a link between HD-ZIP IV transcription factor activity and epidermal development and function. In Arabidopsis, the *gl2* mutant shows abnormal trichome expansion and ectopic root hair differentiation (Rerie et al., 1994; Di Cristina et al., 1996), and the *hdg11* mutant is characterized by excessive trichome branching (Nakamura et al., 2006). Consistent with the subepidermal layer expression of *ANL2*, the *anl2* mutant does not produce subepidermal anthocyanin pigments and has an aberrant cellular organization of the primary root (Kubo et al., 1999). Single mutants in the remaining 13 *HD-ZIP IV* genes lack detectable phenotypes, but the *atml1pdf2* double mutant has a very strong epidermal phenotype; it never forms an organized epidermal layer in the apical part of the proembryo and dies at the seedling stage under standard conditions (Abe et al., 2003). In maize, the *ocl4* mutant is characterized by ectopic macrohairs (a class of maize trichomes) on the leaf margins and the formation of an additional, subepidermal cell layer in the anther wall (Vernoud et al., 2009), whereas a dominant negative form of *OCL1* causes a strong delay in kernel development (Khaled et al., 2005). In tomato, a point mutation in *Cutin deficient2* is likely responsible for the dramatically reduced cutin content in the tomato fruit (Isaacson et al., 2009). In addition, numerous direct or indirect targets of the maize HD-ZIP IV transcription factors are predicted to function in epidermal processes, which further supports the proposed link between this transcription factor family and epidermis-related functions (Abe et al., 2003; Javelle et al., 2010). In particular, OCL1 activates genes involved in lipid metabolism and cuticle biosynthesis in maize (Javelle et al., 2010).

The sequencing of the maize genome (Schnable et al., 2009) offered us the unique opportunity to expand our initial work on five *OCL* genes to all members of the HD-ZIP IV family in maize. We present here a detailed analysis of the phylogeny, synteny, gene structure, expression, and conserved regulatory motifs of the family. In particular, a comparison of the expression in the epidermal and subepidermal layers of two different organs provides further evidence for a specialization of HD-ZIP IV transcription factors in epidermis-related functions.

## RESULTS

### Origin and Structure of 17 *ZmHDZIV* Genes in the Maize Genome

To detect all members of the HD-ZIP IV family present in the maize genome, TBLASTN searches of all maize bacterial artificial chromosome sequences of genotype B73 present in GenBank were performed separately with the HD-ZIP domain and the START domain sequences of all HD-ZIP IV family members in Arabidopsis (16) and rice (11) and of the published family members in maize (five). Based on a value of $e < 10^{-5}$ and the absence of the MEKHLA domain characteristic of the HD-ZIP III family, 17 genes encoding HD-ZIP IV transcription factors were identi-

fied in release 4a.53 of the maize genome (Table I), including the five published genes *OCL1* to *OCL5* (Ingram et al., 1999, 2000). The additional genes were named *ZmHDZIV6* to *ZmHDZIV17* based on nomenclature recommendations for grass transcription factors (Gray et al., 2009). The gene model(s) predicted for these loci were manually verified and improved. The major modifications and the resulting amino acid sequences are presented in Supplemental Figure S1. The annotated nucleotide sequence data are available in the Third Party Annotation section of the GenBank/EMBL/DDBJ databases under the accession numbers BK008026 to BK008042.

In comparison with a previous analysis based on contigs of EST and Genomic Survey Sequence data (Mukherjee et al., 2009), we found only 17 rather than 21 genes in maize. This difference was likely due to duplicate entries for *ZmHDZIV1_OCL1* (*OCL1* and AZM459893) and *ZmHDZIV2_OCL2* (*OCL2* and AZM469787) and a triplicate entry for *OCL9* (PUBHC67TD, AZM4101813, and TF1) in the previous study. While the respective sequences were not identical, they clearly converged to a single gene model in the present B73 genome sequence and represented either variations between genotypes or alternatives in splicing. Similarly, we excluded one rice *HD-ZIP IV* entry (P0409D09.32 = Os07g24350.1) from our study because the corresponding gene model contained only an HD domain but no START or SAD domain within 10 kb of the downstream sequence, reducing the number of rice *HD-ZIP IV* genes to 11.

Analysis of the positions of the 17 *ZmHDZIV* genes in the maize genome did not reveal a strong clustering on particular chromosomes (Table I). The only exception was *ZmHDZIV11* and *ZmHDZIV12*, which were located within 153 kb of each other on chromosome 2. Since none of the neighboring genes were duplicated, the two genes likely arose from a local rather than a whole genome duplication event.

One-quarter of the genes in the maize genome are located in duplicated chromosomal segments, which are remnants of the allotetraploid origin of maize (Swigonová et al., 2004; Schnable et al., 2009). Since the whole genome duplication leading to tetraploidization occurred after the split from sorghum and rice, these duplicate regions are defined by cosynteny of the respective homeologous maize genes to common reference genes in sorghum or rice. Using the CoGe Web site (http://synteny.cnr.berkeley.edu/CoGe/), we checked whether *ZmHDZIV* genes were located in duplicate regions of the maize genome and/or in syntenous regions with rice (Supplemental Fig. S2). We established that 13 of the 17 *ZmHDZIV* genes were present in duplicate regions and that *ZmHDZIV5_OCL5*/*ZmHDZIV10*, *ZmHDZIV6*/*ZmHDZIV8*, *ZmHDZIV13*/*ZmHDZIV14*, and *ZmHDZIV15*/*ZmHDZIV17* likely formed homeologous gene pairs in four such duplicate chromosomal segments (Table I). With the exception of *ZmHDZIV1_OCL1*, the same *ZmHDZIV* genes as well

**Table I.** *ZmHDZIV genes in the maize genome*

| Name | GenBank Accession No. | Identifier in the Maize Genome (Release 4a.53) | Chromosome | Position (Start in Release 4a.53) | Position (End in Release 4a.53) | Orientation | Protein Size | Maize Chromosome with Syntenous Region | *ZmHDZIV* Gene in Syntenous Region[a] | Presence in Sorghum and Rice[b] | Rice Chromosome with Syntenous Region | *Roc* Gene in Syntenous Region[a] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | *amino acids* | | | | | |
| *ZmHDZIV1_OCL1* | BK008026 | GRMZM2G026643_T02 | Zm3 | 27,197,008 | 27,204,687 | Forward | 803 | Zm8 | No | Sb, (Os) | No | na |
| *ZmHDZIV2_OCL2* | BK008027 | AC235534.1_FGT007_mod | Zm10 | 136,522,908 | 136,529,482 | Forward | 834 | Zm2 | No | Sb, Os | Os4 | *Roc4* |
| *ZmHDZIV3_OCL3* | BK008028 | GRMZM2G116658_T01 | Zm7 | 136,100,609 | 136,106,796 | Forward | 863 | Zm2 | No | Sb, Os | Os9 | *Roc6* |
| *ZmHDZIV4_OCL4* | BK008029 | GRMZM2G123140_T03 | Zm1 | 99,763,128 | 99,769,093 | Reverse | 884 | No | na | Sb, Os | Os10 | *Roc3* |
| *ZmHDZIV5_OCL5* | BK008030 | GRMZM2G130442_T01 | Zm4 | 28,833,054 | 28,839,237 | Reverse | 796 | Zm10 | *ZmHDZIV10* | Sb, Os | Os8 | *Roc1* |
| *ZmHDZIV6* | BK008031 | GRMZM2G001289_T02 | Zm2 | 8,237,590 | 8,243,890 | Forward | 795 | Zm10 | *ZmHDZIV8* | Sb, Os | Os4 | *Roc2* |
| *ZmHDZIV7* | BK008032 | GRMZM2G004957_T01 | Zm10 | 84,135,045 | 84,138,429 | Forward | 769 | No | No | Sb, Os | Os8 | *Roc7* |
| *ZmHDZIV8* | BK008033 | GRMZM2G122897_T01 | Zm10 | 142,981,287 | 142,986,049 | Reverse | 742 | Zm2 | *ZmHDZIV6* | Sb, Os | Os4 | *Roc2* |
| *ZmHDZIV9* | BK008034 | GRMZM2G386276_T01 | Zm3 | 187,939,546 | 187,944,608 | Reverse | 701 | Zm6 | No | Sb, Os | Os1 | *OsTF1* |
| *ZmHDZIV10* | BK008035 | GRMZM2G118063_T01 | Zm10 | 42,115,584 | 42,124,124 | Forward | 802 | Zm4 | *ZmHDZIV5_OCL5* | Sb, Os | Os8 | *Roc1* |
| *ZmHDZIV11* | BK008036 | GRMZM2G438260_T01 | Zm2 | 87,878,884 | 87,882,865 | Forward | 750 | No | na | (Sb) | No | na |
| *ZmHDZIV12* | BK008037 | GRMZM2G060444_T01_mod | Zm2 | 88,035,597 | 88,038,630 | Reverse | 747 | No | na | (Sb) | No | na |
| *ZmHDZIV13* | BK008038 | GRMZM2G126646_T01 | Zm4 | 176,668,879 | 176,673,174 | Reverse | 698 | Zm5 | *ZmHDZIV14* | Sb | Os6 | *Roc8* |
| *ZmHDZIV14* | BK008039 | GRMZM2G145690_T01 | Zm5 | 210,323,721 | 210,327,636 | Reverse | 692 | Zm4 | *ZmHDZIV13* | Sb | Os6 | *Roc8* |
| *ZmHDZIV15* | BK008040 | GRMZM2G004334_T01 | Zm6 | 111,001,775 | 111,004,906 | Reverse | 732 | Zm9 | *ZmHDZIV17* | Sb, Os | Os6 | *Roc8* |
| *ZmHDZIV16* | BK008041 | GRMZM2G109252_T01 | Zm7 | 36,835,558 | 36,838,693 | Forward | 611 | No | na | Sb | No | na |
| *ZmHDZIV17* | BK008042 | AC201766.3_FGT002_mod | Zm9 | 8,971,612 | 8,975,297 | Forward | 699 | Zm6 | *ZmHDZIV15* | Sb, Os | Os6 | *Roc8* |

[a]na, Not applicable. [b]Os, rice; Sb, sorghum; (), not in syntenous position.

as *ZmHDZIV4_OCL4* localized to chromosomal regions with syntenous regions in rice, which in all cases contained a rice *HD-ZIP IV* gene. The four duplicate gene pairs all showed cosynteny with a single locus in rice, the gene pairs *ZmHDZIV13*/*ZmHDZIV14* and *ZmHDZIV15*/*ZmHDZIV17* sharing cosynteny with *Roc8* (Table I).

The phylogenetic relationship between all maize, sorghum, rice, Arabidopsis, *Selaginella*, and *Physcomitrella HD-ZIP IV* genes was established by the Bayesian method (Fig. 1). The corresponding tree divided the family into three clades (I–III), one of which contained three subclades of equal rank (Ia–Ic). As expected, each of the previously identified gene pairs *ZmHDZIV5_OCL5*/*ZmHDZIV10*, *ZmHDZIV6*/*ZmHDZIV8*, *ZmHDZIV13*/*ZmHDZIV14*, and *ZmHDZIV15*/*ZmHDZIV17* clustered with a single sorghum gene. However, only in the last case were the two maize genes closer to each other than to the sorghum gene, confirming that the gene pair *ZmHDZIV15*/*ZmHDZIV17* arose during the whole genome duplication after the split of maize and sorghum. Each of the remaining maize genes had a sorghum ortholog, including the gene pair *ZmHDZIV11*/*ZmHDZIV12*, situating this local duplication event prior to the split of maize and sorghum.

Three branches of the tree had values below 75%. They all concerned the position of Arabidopsis sequences, which tended to cluster among each other, whereas sorghum and rice sequences were interspersed between maize sequences. And if the use of alternative settings or methods (maximum likelihood) yielded slightly different trees (data not shown), it did not change the clustering of monocot or dicot sequences. While the clustering can readily be explained with independent duplication events in dicots, it makes it very difficult to infer the function of particular cereal *HD-ZIP IV* genes from mutant data in the model species Arabidopsis. For example, the maize genes *ZmHDZIV1_OCL1* and *ZmHDZIV2_OCL2* are close neighbors of the functionally characterized Arabidopsis genes *ANL2* (Kubo et al., 2008) and *FLOWERING LATE* (*FWA;* Kinoshita et al., 2004) in clade Ia, yet it is impossible to determine from the phylogenetic tree whether *ANL2*, *FWA*, *HDG1*, or *HDG7* is the Arabidopsis ortholog of *ZmHDZIV2_OCL2* or *ZmHDZIV1_OCL1*. More generally, the numerous duplications in cereals in general and maize in particular may hamper future functional analyses due to functional redundancy, as documented in Arabidopsis (Nakamura et al., 2006).

Analysis of the intron-exon structure of *ZmHDZIV* genes revealed substantial variation, with the intron number varying between six and 11. When focusing on the coding sequence, a basic module of eight exons and seven introns emerged, from which all gene structures can be derived by the insertion of introns and/or the fusion of exons (Fig. 2). *ZmHDZIV* genes that shared a particular novel intron tended to be clustered in the phylogenetic tree. For instance, the division of exon 3 into two exons, 3a and 3b, was found in all members of clade Ib but not in clades Ia, II, or III; the additional intron in *HDZIV9* was actually located at a different position in exon 3. Similarly, the division of exon 7 into exons 7a and 7b was not found in any member of clade Ia but was shared by all members of clades Ib and II with the exception of *ZmHDZIV7*, which might be explained by a secondary fusion event concerning not only exon 7 but also exons 5 and 6. Finally, the division of exon 8 and the presence of an intron in the 3′ untranslated region (UTR) were shared by all genes in clades II and III but were not found in any other clade (Fig. 2). It is further noteworthy that none of the new introns inserted within the highly conserved homeodomain and START domain.

On the other hand, the apparent intron deletions were specific to individual *ZmHDZIV* genes. Whenever an intron at the same position was lost in more than one *ZmHDZIV* gene, the resulting intron-exon structures were clearly distinct. For example, the two events involving intron 2 (*ZmHDZIV4_OCL4* and *ZmHDZIV15*) did not resemble each other. In the case of *ZmHDZIV7* and *ZmHDZIV16*, two introns appeared to have been lost, leading to the most compact genes, with only six introns.

### Preferential Expression of *ZmHDZIV* Genes in Immature Reproductive Organs

To complement the existing expression data of *ZmHDZIV1_OCL1* to *ZmHDZIV5_OCL5* (Ingram et al., 2000) and to establish the expression patterns of the remaining *ZmHDZIV* genes, quantitative reverse transcription (qRT)-PCR experiments were carried out on RNA extracted from major organs of the maize plant. Both vegetative organs such as young seedlings, juvenile and adult leaves, stems, or roots and reproductive organs comprising male (tassel) and female (ear) inflorescences, silk (maize styles), and kernels at 12 or 35 d after pollination were examined.

Most *ZmHDZIV* genes exhibited a rather broad expression profile. With the exception of roots and stems, in which the transcript levels of most *ZmHDZIV* genes were low or undetectable, expression of several to many *ZmHDZIV* genes was observed in the other organs tested (Fig. 3; Supplemental Table S2). All the *ZmHDZIV* genes, except *ZmHDZIV12*, had a relatively high level of expression in reproductive compared with vegetative organs. Indeed, most showed a preferential expression either in immature tassels (*ZmHDZIV4_OCL4*, *ZmHDZIV5_OCL5*, *ZmHDZIV6*, *ZmHDZIV8*, *ZmHDZIV10*, *ZmHDZIV13*, *ZmHDZIV14*, *ZmHDZIV15*, and *ZmHDZIV17*) or immature ears (*ZmHDZIV1_OCL1*, *ZmHDZIV2_OCL2*, and *ZmHDZIV7*). Only the transcript levels for *ZmHDZIV16* and *ZmHDZIV9* were most abundant in young developing kernels and for *ZmHDZIV3_OCL3* in silk (Fig. 3). Interestingly, *ZmHDZIV* expression levels were consistently higher in immature, developing organs compared with mature, differentiated ones, both in vegetative and reproductive structures. Consequently, the relative richness of reproductive organs in young

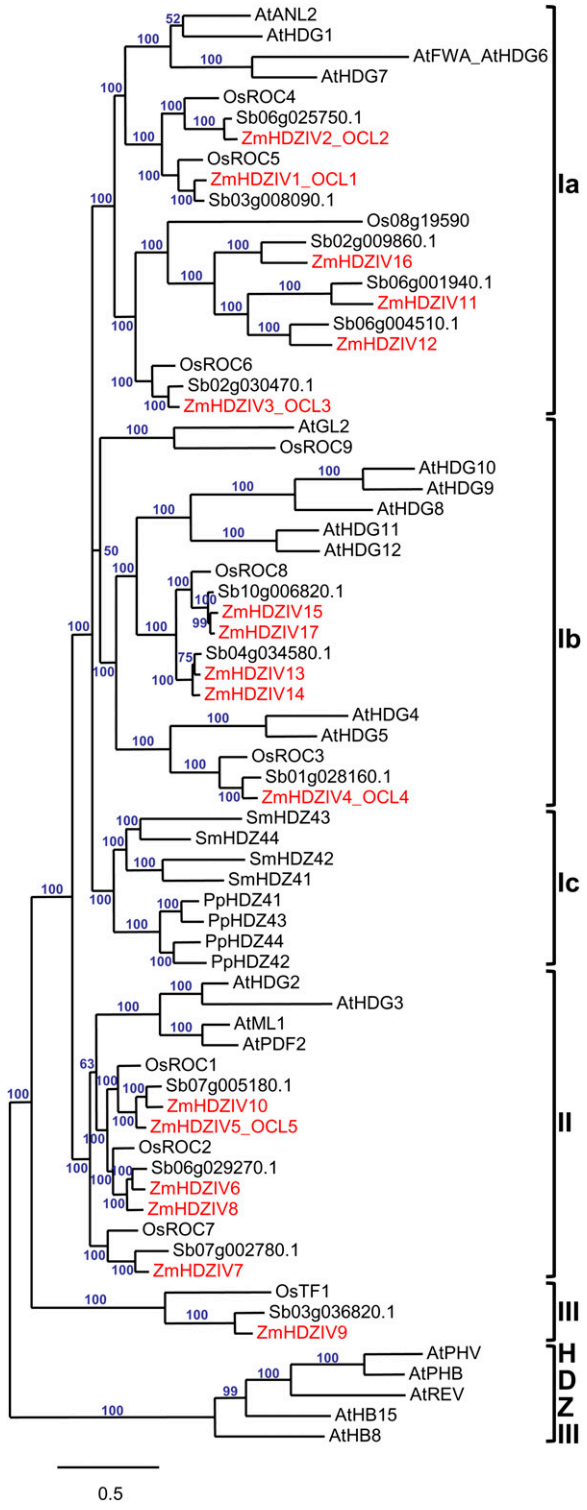


**Figure 1.** Phylogenetic tree of *HD-ZIP IV* genes from maize, sorghum, rice, Arabidopsis, *Physcomitrella*, and *Selaginella*. A Bayesian phylogenetic tree was generated using all Arabidopsis (At), maize (Zm; red), rice (Os), sorghum (Sb), *Physcomitrella* (Pp), and *Selaginella* (Sm) *HD-ZIP IV* coding sequences with MrBayes software (2,000,000 generations; average SD of split frequencies = 0.0028). Coding sequences of *HD-ZIP III* genes from Arabidopsis behaved as an outgroup (HDZIII). Different clades (roman numerals) were annotated manually. [See online article for color version of this figure.]

dividing tissues may have contributed to their generally high overall expression levels. One notable exception was *ZmHDZIV11*, which was almost exclusively expressed in mature tassels that included pollen. Expression of its paralog *ZmHDZIV12* was not detected in any of the 13 organs tested, despite the use of several distinct primer pairs that all amplified genomic DNA. This result is substantiated by the absence of ESTs for this gene in GenBank. In conclusion, 16 of the 17 *ZmHDZIV* genes were expressed, and the majority of them showed preferential expression in immature aerial organs and more precisely young male and female inflorescences.

### Preferential Expression of *ZmHDZIV* Genes in the Epidermal/L1 Cell Layer

Previous in situ hybridization experiments had shown that *ZmHDZIV1_OCL1*, *ZmHDZIV3_OCL3*, *ZmHDZIV4_OCL4*, and *ZmHDZIV5_OCL5* exhibited epidermis-specific expression notably in embryo, meristems, and young organ primordia (Ingram et al., 1999, 2000), whereas *ZmHDZIV2_OCL2* expression was restricted to the subepidermal layer (L2) in floral meristems. To evaluate whether this cell layer specificity was shared by the remaining *ZmHDZIV* genes, we performed laser microdissection (LMD) followed by RT-PCR experiments on RNA extracted from epidermal versus mesophyll cells of the central part of fully expanded juvenile leaves and on L1 versus L2 cells of the shoot apical meristem (SAM; Fig. 4, A and C). Of the 17 *ZmHDZIV* genes tested on leaf epidermal and mesophyll cells, amplification products were detected for nine of them (Fig. 4B). As observed previously (Javelle et al., 2010), *ZmHDZIV1_OCL1* transcripts accumulated preferentially but not exclusively in epidermal cells of juvenile leaves. Similar results were obtained for *ZmHDZIV3_OCL3* and *ZmHDZIV10*. On the other hand, highly enriched epidermal expression was detected for *ZmHDZIV5_OCL5* and epidermis-specific expression for *ZmHDZIV6*, *ZmHDZIV7*, *ZmHDZIV8*, *ZmHDZIV14*, and *ZmHDZIV16*. Experiments on RNA extracted from SAM L1 and L2 cells confirmed these results and gave us information for five additional genes. *ZmHDZIV4_OCL4*, *ZmHDZIV13*, *ZmHDZIV15*, and *ZmHDZIV17* were also extremely enriched in the L1 layer of the meristem, while *ZmHDZIV9* was very weakly expressed in L2 cells but not detected in L1. As in juvenile leaves, no expression of *ZmHDZIV11* or *ZmHDZIV12* was detected in the SAM samples.

The previous in situ hybridization data in embryos, shoot apices, inflorescences, and flowers (Ingram et al., 1999, 2000), the leaf LMD data, and the SAM LMD data were generally in good agreement. However, some quantitative differences were observed between leaf and SAM, such as the increased outer cell layer specificity

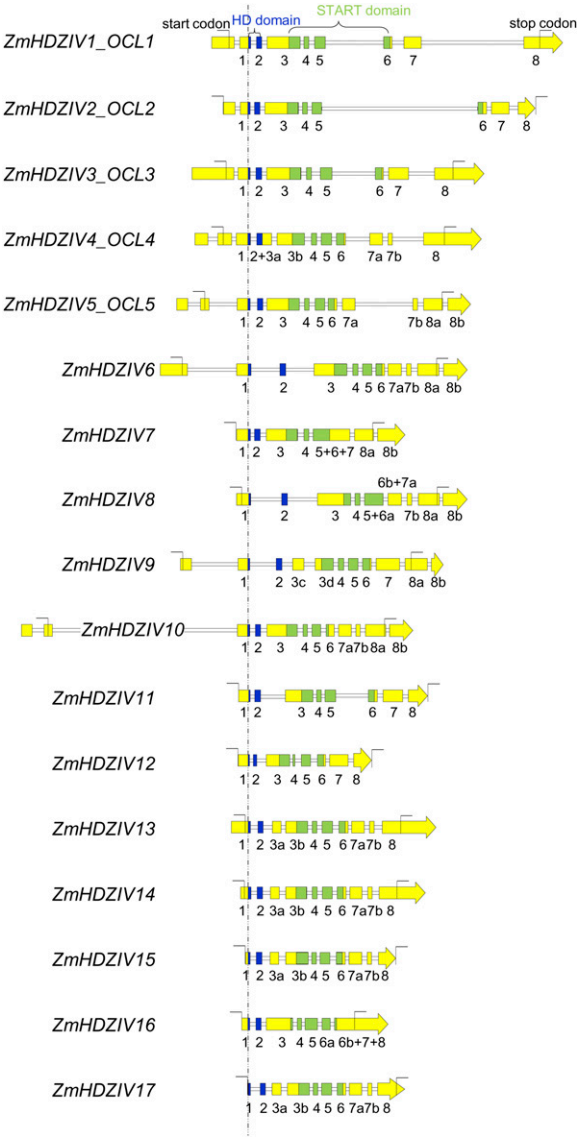


**Figure 2.** Genomic structure of maize *ZmHDZIV* genes. Exons are indicated by thick boxes and introns by thin boxes. The regions coding for the HD and the START domain are colored in blue and green, respectively. Vertical bars indicate the start and stop codons, respectively. Conserved exons are given the same number in all genes, the numbering starting arbitrarily with the exon containing the beginning of the highly conserved homeodomain. The accession numbers of the sequences used are listed in Table I. [See online article for color version of this figure.]

of *ZmHDZIV3_OCL3* or the decreased specificity of *ZmHDZIV16*. In addition, notable qualitative differences were observed for *ZmHDZIV2_OCL2*. While in situ hybridization had shown that *ZmHDZIV2_OCL2* mRNA was excluded from the L1 in floral meristems (Ingram et al., 2000), the gene was preferentially expressed in this cell layer in the SAM. This exception demonstrated that the preferential or near-specific expression in one of the cell layers could not be automatically extrapolated from the leaf or SAM data to the entire plant, although the conservation of layer preference throughout the plant remained the most likely scenario.

### Conservation of Two Short Sequence Motifs in the 3′ UTR of *HD-ZIP IV* Genes

Previously, the 3′ UTR of *HD-ZIP IV* genes had attracted attention due to the conservation of a 17-nucleotide motif between the two *Picea* genes *PaHB1* and *PaHB2* and several *HD-ZIP IV* genes of Arabidopsis, sunflower, and cotton (Ingouff et al., 2003). Analysis of the *ZmHDZIV* 3′ UTR sequences detected this motif in 13 of the 17 *ZmHDZIV* genes and allowed us to extend the consensus sequence to 21 nucleotides, 5′-GGTGGTTCGGGTATTGACTTC-3′ (Supplemental Table S3). Further analysis of *HD-ZIP IV* genes from moss, lycopods, gymnosperms, monocots, and eudicots demonstrated that nine of 13 sorghum, eight of 11 rice, 10 of 16 Arabidopsis, four of four *Selaginella*, and four of four *Physcomitrella HD-ZIP-IV* genes contained this conserved motif in their 3′ UTR. Two maize genes, *ZmHDZIV4_OCL4* and *ZmHDZIV7*, carried a mismatch in the strongly conserved core region; in the case of *ZmHDZIV4_OCL4*, this mismatch was conserved in the orthologs in sorghum and rice, hinting at functional implications of the primary sequence of the motif.

Moreover, detailed analysis of the 3′ UTR sequences revealed a second conserved motif of 19 nucleotides (consensus sequence 5′-GGAGTCAAGAACGCACC-TC-3′) that was located upstream of the 21-nucleotide motif in all the species under investigation (Fig. 5; Supplemental Table S3). Interestingly, the two conserved sequences were either both present or both absent in a given 3′ UTR, suggesting a functional link between the two motifs. The presence of the conserved motifs in all four *HD-ZIP IV* genes of *Physcomitrella patens*, which holds a basal position in the phylogeny of land plants, suggested that the ancestral *HD-ZIP IV* gene(s) possessed the two motifs.

Since the two motifs not only coevolved but also were partially complementary, we investigated the possibility that the two sites formed a specific secondary RNA structure within the 3′ UTR using the secondary structure prediction software RNAfold (Gruber et al., 2008). Of the 13 *ZmHDZIV* genes containing the two motifs, *ZmHDZIV1_OCL1*, *ZmHDZIV2_OCL2*, *ZmHDZIV3_OCL3*, *ZmHDZIV6*, *ZmHDZIV7*, *ZmHDZIV8*, *ZmHDZIV10*, and *ZmHDZIV17* were predicted to form a stem-loop structure with considerable base-pairing probability between the two motifs (Supplemental Fig. S3). The predicted secondary structures were also well conserved in Arabidopsis and rice, despite the fact that the two motifs were not always located at the same distance from the stop codon or separated by the same number of nucleotides (data not shown). In conclusion, the presence of evolutionarily conserved 19- and 21-nucleotide motifs suggested that a posttranscriptional regulatory mechanism controlled *ZmHDZIV* gene expression, possibly via base pairing of the two motifs.

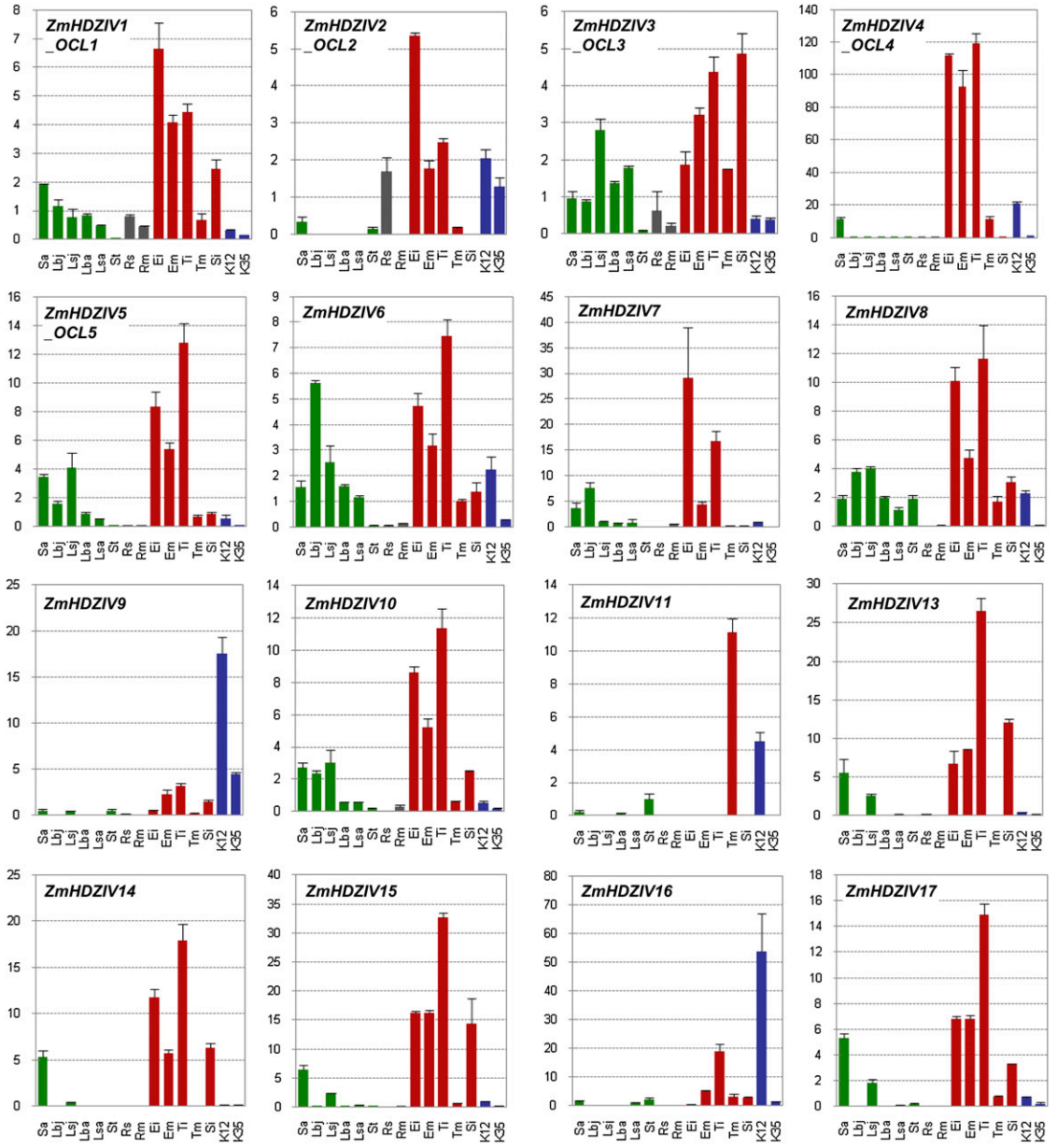


**Figure 3.** qRT-PCR analysis of *ZmHDZIV* transcript levels in different maize organs. Total RNA was extracted from the aerial part of seedlings (Sa), juvenile leaf blade (Lbj), juvenile leaf sheath (Lsj), adult leaf blade (Lba), adult leaf sheath (Lsa), stem (St), seedling root (Rs), mature root (Rm), immature ear (Ei), mature ear (Em), immature tassel (Ti), mature tassel (Tm), silk (Si), and kernel at 12 d after pollination (K12) and 35 d after pollination (K35). Organs are color coded: vegetative in green, root in gray, reproductive in red, and kernel in blue. For each gene, the expression levels obtained by normalization with the *18S* RNA gene are presented on relative scales that differ from gene to gene. Data are average values ± SD from two experiments, each carried out in duplicate. For each sample, organs from two to six independent plants were pooled. No data are presented for *ZmHDIV12*, since this gene was not expressed in any of the organs tested.

## DISCUSSION

### Duplications, Losses, and Transpositions of Cereal *HD-ZIP IV* Genes

Cereal genome evolution has been marked by two whole genome duplications. The first concerns all cereals and took place approximately 90 million years ago in the common ancestor of rice, sorghum, and maize, while the second occurred approximately 12 million years ago specifically in the lineage leading to maize (Salse et al., 2009). As a consequence of the second duplication event, approximately 25% of the genes in the maize genome possess closely related paralogs (Schnable et al., 2009). Thus, it is not surprising that the maize genome with 17 members contains more *HD-ZIP IV* genes than the sorghum and rice genomes with 13 and 11 members,

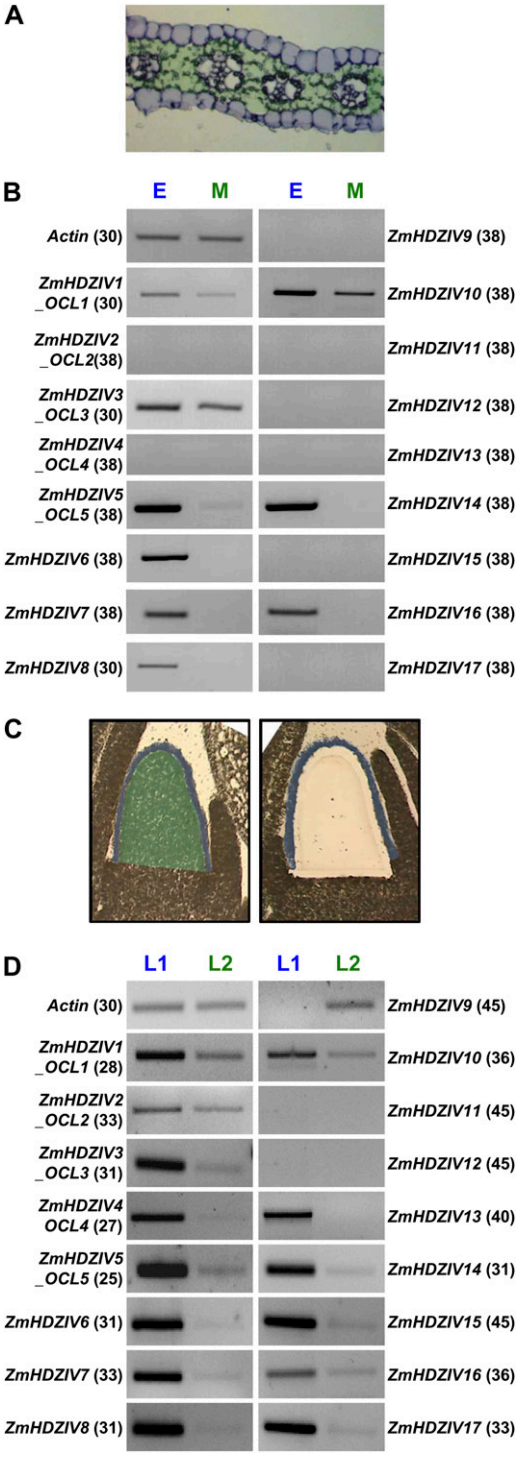


**Figure 4.** Expression of *ZmHDZIV* genes in outer and inner cell layers of juvenile leaf and SAM. A, Leaf section showing epidermal cells (blue) and mesophyll cells (green) isolated using the infrared laser-capture microdissection system. B, RT-PCR experiments assessing the expression of *ZmHDZIV* genes in microdissected epidermal (E) and mesophyll (M) cells using gene-specific primers. C, SAM section showing L1 cells (blue) and L2 cells (green) isolated using the UV laser-capture microdissection system before (left) and after (right) capture of the L2 cells. D, RT-PCR experiments assessing the expression of *ZmHDZIV* genes in microdissected L1 and L2 cells using gene-specific primers. In B and D, the concentration of the cDNA templates was normalized according to the abundance of the *Actin* RT-PCR product; the number of PCR cycles is indicated in parentheses.

respectively. Unexpectedly in three cases, one of the maize genes was slightly more similar to its sorghum ortholog than to its maize homeolog (Fig. 1), raising the possibility of a more ancient duplication in the common ancestor of maize and sorghum after its split from rice and subsequent gene loss in sorghum. While this possibility cannot be excluded, several arguments favor the hypothesis that the homeologous pairs *ZmHDZIV5_OCL5/ZmHDZIV10*, *ZmHDZIV6/ZmHDZIV8*, and *ZmHDZIV13/ZmHDZIV14* arose during the second whole genome duplication event approximately 12 million years ago, just like the pair *ZmHDZIV15/ZmHDZIV17*, for which the two maize genes are closer to each other than to their sorghum ortholog. First, all six maize genes concerned are located in the 25% of the genome duplicated during the maize-specific whole genome duplication. Second, the six maize regions show synteny only to the three sorghum regions carrying *HD-ZIP IV* genes and not to any other part of the genome. Third, there is no documentation of a major duplication event concerning the common ancestor of maize and sorghum but not rice. Finally, a strong bias in gene loss and expression suggests that maize homeologs evolve at different rates, providing a possible explanation for the stronger conservation of one of them (Schnable et al., 2011).

With regard to rice, the gene pairs *ZmHDZIV5_OCL5/ZmHDZIV10* and *ZmHDZIV6/ZmHDZIV8* cluster as expected with single genes from rice. However, the phylogenetic clade encompassing the two maize gene pairs *ZmHDZIV13/ZmHDZIV14* and *ZmHDZIV15/ZmHDZIV17* contains two sorghum genes but only a single rice gene, *Roc8*, suggesting either gene loss in rice or a specific duplication in the branch leading to sorghum and maize. We favor the first hypothesis, and more precisely a gene loss on rice chromosome 2, since there is not only strong synteny between rice chromosome 2 (position 32,600,000 containing no *Roc*) and rice chromosome 6 (the region around *Roc8*) but also between this region of rice chromosome 2 and maize chromosome 4 (the region around *ZmHDZIV13*), maize chromosome 5 (the region around *ZmHDZIV14*), maize chromosome 6 (the region around *ZmHDZIV15*), and maize chromosome 9 (the region around *ZmHDZIV17*). In fact, the data converge to a model in which a single ancestor gene was duplicated during the first whole genome duplication approximately 90 million years ago in the ancestor of rice, sorghum, and maize. During the second whole genome duplication approximately 12 million years ago, both copies were duplicated in maize, whereas the copy on chromosome 2 was lost in rice.

Duplication of the maize *ZmHDZIV11* and *ZmHDZIV12* genes likely involved a different evolutionary path.

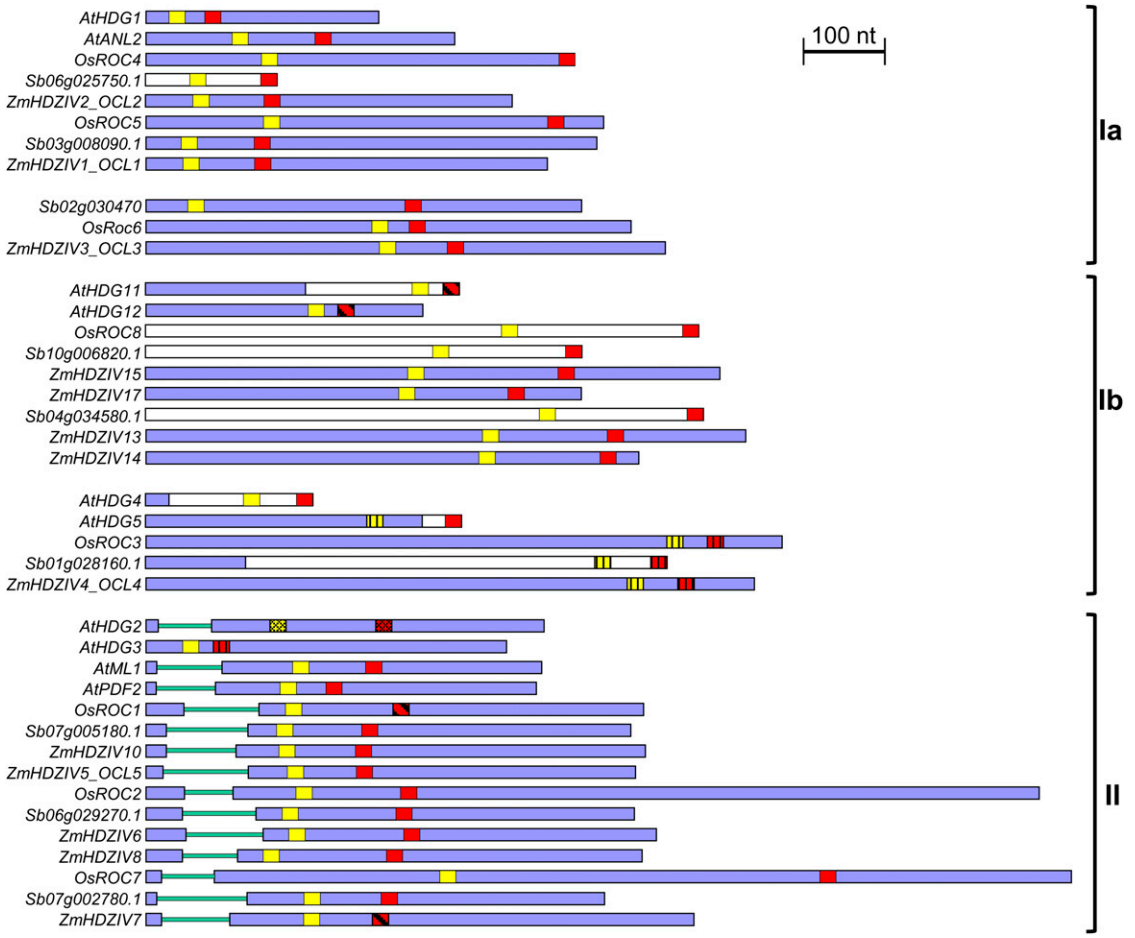


**Figure 5.** Positions of conserved 21- and 19-nucleotide (nt) motifs in the 3′ UTR of maize, sorghum, rice, and Arabidopsis *HD-ZIP IV* genes. The 3′ UTRs of maize (Zm), sorghum (Sb), rice (Os), and Arabidopsis (At) *HD-ZIP IV* genes are drawn to scale starting with the first nucleotide after the stop codon. 3′ UTRs with corresponding EST data are represented in blue and 3′ UTRs without EST data are in white. Thin green boxes depict introns. Yellow and red boxes indicate the positions of the 19- and 21-nucleotide conserved motifs, respectively. Motifs with nonconserved nucleotides in their core region are depicted with the same motif. Sequences are grouped according to the phylogenetic clades in Figure 1.

*ZmHDZIV11* and *ZmHDZIV12* are located near each other on maize chromosome 2, share less extensive sequence similarity than the paralogs described previously, and each gene has a sorghum counterpart. While they are phylogenetically related to *ZmHDZIV16*, neither the pair *ZmHDZIV11*/*ZmHDZIV12* nor *ZmHDZIV16* is located in duplicate regions of the maize genome. These data led to a hypothesis in which a copy of *ZmHDZIV16* was inserted elsewhere in the genome of the common ancestor to maize and sorghum, associated with a local duplication either during this insertion event or afterward. One might speculate that this insertion was possibly mediated by transposable elements, helitrons being prime candidates (Yang and Bennetzen, 2009).

The remaining six *ZmHDZIV* genes all possess a single sorghum and rice ortholog. While the high sequence similarity between these maize and rice genes is generally reflected by synteny between their chromosomal locations, there is one exception. *ZmHDZIV1_OCL1* and *Roc5* share strong sequence similarity and are closest neighbors in the phylogenetic tree, yet *ZmHDZIV1_OCL1* is located in a region of maize chromosome 3 that has no synteny to the rice genome; in fact, this region corresponds to a gap of 2.92 Mb in the otherwise strong synteny between maize chromosome 3 and rice chromosome 1. On the other hand, the region of rice chromosome 2 containing *Roc5* shows extensive synteny to regions of maize chromosome 4 and maize chromosome 5, which do not contain *ZmHDZIV* genes. Consequently, one of the two genes must have been displaced in the genome, *ZmHDZIV1_OCL1* being the better candidate due to the well-documented plasticity of the maize genome (Wang and Dooner, 2006), which seems to be higher than that of rice (Sweredoski et al., 2008).

As reported previously, whole genome duplications affecting *HD-ZIP IV* genes also occurred indepen-

dently in the eudicot branch leading to Arabidopsis, where 14 of the 16 family members have been defined as paralogous gene pairs by Nakamura et al. (2006): *HDG1/ANL2*, *HDG2/HDG3*, *HDG4/HDG5*, *FWA/HDG7*, *HDG8/HDG9*, *HDG11/HDG12*, and *ATML1/PDF2*. Consequently, it is not surprising to find more Arabidopsis than rice, maize, or sorghum genes in some clades. For example, the *ZmHDZIV4_OCL4* clade includes the Arabidopsis *HDG4/HDG5* pair and only one gene from rice (*Roc3*), maize (*ZmHDZIV4_OCL4*), and sorghum (*Sb01g028160*). More unexpectedly, even the Arabidopsis gene pairs tend to cluster together in the phylogenetic tree rather than being interspersed among the cereal genes. This suggests the presence of a rather limited number of *HD-ZIP IV* genes in the common ancestor to monocots and dicots and numerous duplication events after the separation of the two lineages. Taking into account the clustering of the eight extant *HD-ZIP IV* genes from *Physcomitrella* and *Selaginella* in a single clade, one may want to postulate gene loss in these two species and the presence of three *HD-ZIP IV* genes corresponding to clades I, II, and III (Fig. 1) in the ancestor of land plants. These findings raise the question of why the majority of duplicated genes have been retained in the respective genomes. Previous findings in eudicots suggest that this is a general feature of transcription factor genes, which show significantly higher retention than genes fulfilling basic cellular functions (Tang et al., 2008). In addition, one may speculate that the specialization of epidermal functions in land plants necessitated additional control mechanisms, and it may not be a coincidence that all the *HD-ZIP IV* genes involved in trichome development, which are *GL2*, *HDG11*, and *HDG12* from Arabidopsis (Nadeau et al., 1996), *ZmHDZIV4_OCL4* from maize (Vernoud et al., 2009), and *GaHOX1* from cotton (Guan et al., 2008), are members of clade Ib (data not shown for *GaHOX1*).

### Intron Insertions and Losses in Cereal *HD-ZIP IV* Genes

In parallel to gene number, the structure of *HD-ZIP IV* genes underwent evolutionary changes. Applying the rules of parsimony and taking into account both the conservation of certain modifications of the intron-exon structure and the phylogenetic positions of the corresponding genes, an ancestral *HD-ZIP IV* gene with seven introns in the coding region can be deduced. This basic unit was modified by the insertion of introns into exons 3, 7, and/or 8, which in each case are shared between closely related family members. Based on the published intron-exon structure of Arabidopsis (Nakamura et al., 2006), the complete absence of additional introns in exons 3, 7, and 8 of all genes belonging to clade Ia, the presence of an additional intron in both exons 3 and 7 of all members of clade Ib (with the exception of *AtGL2* lacking the insertion in exon 3), and the restriction of an additional intron in exon 8 to clades II and III were evolutionarily conserved between maize and Arabidopsis. In the case of exon 8, the new intron is positioned just downstream of the coding region in the 3′ UTR. The splicing of this intron may play a role in transcript stability, as suggested by the precedent of the Arabidopsis *THIAMINE C SYNTHASE A* gene (Bocobza et al., 2007).

On the other hand, the predicted intron losses seem to have occurred more recently, and as independent events, since these occurred in nonconserved positions and in distantly related *ZmHDZIV* genes. Beyond the coding sequence, introns of varying number and size are also found in the 5′ UTR, possibly provoking an intron-mediated enhancement phenomenon, which enhances the level of translation. For example, in the *AtEFIα-A3* gene, the presence of a long intron in the 5′ UTR is sufficient to enhance gene expression in plants in a size-dependent manner (Chung et al., 2006). Neither the insertions nor the losses of introns can readily be linked to modifications in gene function, since even in the model species Arabidopsis, very few of the 16 *hdg* single or double mutants have been attributed to detectable developmental phenotypes, despite rather exhaustive study (Nakamura et al., 2006).

### Widespread Epidermal Expression among *ZmHDZIV* Genes

The expression profiles of the 17 *ZmHDZIV* genes in 15 organs of the maize plant represent, to our knowledge, the first comprehensive analysis of the HD-ZIP IV family in monocots and complement an earlier study in the dicot model species Arabidopsis (Nakamura et al., 2006). No expression was detected for *ZmHDZIV12* in any of the organs tested, suggesting that the gene is either not expressed or exhibits a highly restricted spatiotemporal expression pattern. All of the remaining 16 *ZmHDZIV* genes show strong expression in reproductive organs (ear, tassel), and half show moderate expression in leaves. This roughly reflects the situation in Arabidopsis, where only two of the 16 Arabidopsis genes, *HDG3* and *HDG8*, are not expressed in flowers. Contrary to Arabidopsis, where approximately half of the 16 *ZmHDZIV* genes have substantial expression in root and stem, only *ZmHDZIV1_OCL1*, *ZmHDZIV2_OCL2*, and *ZmHDZIV3_OCL3* are expressed above basal levels in these organs. Another difference with Arabidopsis resides in the fact that the four paralogous pairs in maize have very similar expression patterns, whereas only two of the six paralogous pairs in Arabidopsis behave in a similar fashion (Nakamura et al., 2006). This may indicate that the duplication events in Arabidopsis are generally more ancient than the ones in maize. All these observations lead to the conclusion that individual genes have lost part of the ancestral expression pattern and/or gained new expression territories.

The use of two different developmental stages for five of the maize organs allowed a comparison of *ZmHDZIV* expression between young, developing and old, differentiated organs. With the notable exception of *ZmHDZIV11*, which is expressed in mature but not in immature tassels, the expression of most *ZmHDZIV*

genes is more abundant in immature compared with mature organs. This observation is consistent with the functional analysis of *ZmHDZIV4_OCL4*, which demonstrated that morphological defects in *ocl4* mutant leaves and anthers become microscopically detectable at stages when *ZmHDZIV4_OCL4* is no longer expressed, but these occur precisely in those tissues that express *ZmHDZIV4_OCL4* earlier in development (Vernoud et al., 2009). Extrapolating these observations, *ZmHDZIV* genes may set the stage in immature organs for developmental events manifesting themselves in mature organs.

Within a given organ, *HD-ZIP IV* genes often show predominant expression in a single tissue layer, typically the epidermis or, occasionally, the subepidermal cell layer. In Arabidopsis, in situ hybridization and/or gene-specific promoter fusions to the *GUS* reporter gene revealed that nine of the 16 *HD-ZIP IV* genes show strongly enriched expression in the epidermis, whereas expression of one family member is limited to the subepidermal layer (Rerie et al., 1994; Lu et al., 1996; Masucci et al., 1996; Nakamura et al., 2006; Kubo et al., 2008). Similar results were obtained for all those genes characterized by in situ hybridization in rice (five genes) and maize (five genes; Ingram et al., 1999, 2000; Ito et al., 2002, 2003). Using laser microdissected outer and inner cell layers of juvenile leaves and of the SAM, we obtained layer-specific data for all *ZmHDZIV* genes but *ZmHDZIV11* and *ZmHDZIV12*. The expression of all the genes was moderately to very strongly enriched in the outer cell layer of both structures, lending further evidence to the hypothesis that most members of the HD-ZIP IV family have epidermis-related functions. As in Arabidopsis, there was one exception, *ZmHDZIV9* being weakly but specifically expressed in L2 of the vegetative meristem.

Expression of almost all family members in the SAM and of only a limited number in juvenile or adult leaves further substantiates the previously mentioned preference for *ZmHDZIV* genes to be expressed in immature organs and suggests a role of the family very early in the development of the maize leaf in general and its epidermis in particular.

### Regulation of *HD-ZIP IV* Genes via a Small RNA?

Earlier observations had identified a highly conserved 17-nucleotide sequence motif in the 3′ UTR of several *HD-ZIP IV* genes from *Picea*, Arabidopsis, sunflower, and cotton (Ingouff et al., 2003). Through an exhaustive analysis of 3′ UTR sequences from all maize, rice, sorghum, Arabidopsis, *Selaginella*, and *Physcomitrella HD-ZIP IV* genes, we established a 21-nucleotide consensus sequence for this motif. In addition, we identified a second conserved motif of 19 nucleotides located between the stop codon and the 21-nucleotide motif. The two motifs are frequently predicted to base pair in the stem of hairpins in the minimum free-energy structure of the 3′ UTR.

Given the size and the high evolutionary conservation of the motifs, their primary sequence and/or the formation of a stem-loop structure through their base pairing may have regulatory implications. Considering the lengths of the motifs, it is interesting to postulate that one or both may be binding sites for a small RNA and consequently play a role in the regulation of *HD-ZIP IV* genes. Indeed, the majority of plant microRNAs (miRNAs) are known to regulate gene expression through transcript cleavage and/or translation repression by near-perfect base pairing with their mRNA targets at conserved sites in the coding sequence or the 3′ UTR (Voinnet, 2009). One of many examples, genes of the HD-ZIP III family are regulated posttranscriptionally by *miR165/166* binding to a conserved site in the coding sequence corresponding to the beginning of the START domain (Bowman, 2004). It seems unlikely that both motifs are miRNA target sites. Although some plant genes are known to be targets for multiple distinct small RNAs, such examples are not conserved throughout evolution (Axtell et al., 2007; Howell et al., 2007). However, it remains possible that one of the conserved motifs corresponds to a small RNA target site, with the second motif regulating its accessibility. Such a situation exists in the *p27* 3′ UTR in *Caenorhabditis elegans*, where the binding of Pumilio protein to the first element induces a local change in RNA structure that facilitates the recognition of the second element by *miR221/miR222* and leads to efficient down-regulation of *p27* expression (Kedde et al., 2010). More generally, tight secondary structures surrounding a miRNA-binding site, such as the paired stem structures predicted to form in the 3′ UTR of *ZmHDZIV* genes, may influence the accessibility of a miRNA to its target, and modulation of these secondary structures by binding of cofactors may control miRNA-binding site accessibility (Long et al., 2007). Since the base pairing between plant miRNAs and their targets is generally very strict, the sequence of a hypothetical miRNA binding to the 19-nucleotide or the 21-nucleotide motif can be deduced. However, no corresponding miRNA has been found in the present releases of miRNA databases (data not shown).

Alternatively, one may assume that the secondary structure on its own plays a role without the involvement of a miRNA or other small RNA and that the role of the two motifs is limited to the formation of this hairpin. Indeed, such hairpins serve as binding sites for a variety of proteins. For instance, TAR RNA-binding protein is involved in the regulation of mRNAs via binding to a hairpin (Svoboda and Di Cara, 2006). Independently of the underlying mechanism, the biological function of the hypothetical regulation of *HD-ZIP IV* genes by the conserved motifs in their 3′ UTR remains open. A link with the preferential expression of *HD-ZIP IV* genes in the outer cell layer seems to be compromised by the example of *HDZIV2_OCL2*, which is not L1 preferential but contains the conserved 19- and 21-bp elements in its 3′ UTR. Other

potential roles include the fine-tuning of *HD-ZIP IV* mRNA or HD-ZIP IV protein stability, which may also contribute to the maintenance of layer-specific cell fates by HD-ZIP IV transcription factors.

## CONCLUSION

With 17 family members, the HD-ZIP IV family in maize is larger than in sorghum, rice, or Arabidopsis. *ZmHDZIV* genes are expressed in most organs and show highest mRNA levels in immature reproductive organs. Highly preferential expression in the epidermal layer is widespread, possibly suggesting a specialization of this transcription factor family in epidermis-related functions. Finally, the regulation of numerous family members may be linked to the presence of two conserved motifs in their 3′ UTR.

## MATERIALS AND METHODS

### Sequence Analysis

Gene models were downloaded from release 4a.53 of the B73 maize (*Zea mays*) genome assembly (http://www.maizesequence.org) and manually annotated by comparisons with maize EST data (http://www.maizesequence.org/blast) and alignments with Vector NTI software (Invitrogen). The resulting nucleotide sequence data are available in the Third Party Annotation section of the GenBank/EMBL/DDBJ databases under accession numbers BK008026 to BK008042. Partial cDNA sequences for *ZmHDZIV6* to *ZmHDZIV17* (excluding *ZmHDZIV12*) have been assigned the accession numbers JN003608 to JN003618. Duplications in the maize genome and synteny between maize and rice (*Oryza sativa*) were established with the SynMap tool and confirmed with the GEvo tool at the CoGe Web site (http://synteny.cnr.berkeley.edu/CoGe/). Secondary structures of 3′ UTR sequences were predicted using RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

### Phylogenetic Analysis

The cDNA coding sequences were aligned with MUSCLE version 3.5 (Edgar, 2004), and phylogenetic trees were established with the MrBayes version 3.1 software for the Bayesian inference of phylogeny (Ronquist and Huelsenbeck, 2003) using standard settings (number of substitution types = 6, rates = invgamma). After an initial analysis with 100,000 generations, additional generations were added until the SD of split frequencies fell below 0.01. The trees were formatted with TreeDyn (Chevenet et al., 2006).

### qRT-PCR

Total RNA from different maize organs (described by Vernoud et al., 2009) was extracted using TRIzol reagent according to the manufacturer's instructions (Invitrogen). Genomic DNA contaminants were removed from RNA samples using TURBO DNase from Ambion. DNase treatment and inactivation were performed according to the manufacturer's protocol (Ambion). Approximately 5 $\mu$g of total RNA was reverse transcribed using random hexamers (Amersham Biosciences) and reverse transcriptase without RNaseH activity (Fermentas) in a final volume of 20 $\mu$L.

qPCR was conducted on an ABI StepOne Plus cycler (Applied Biosystems) using the FastStart Universal SYBRGreen Master Mix (Roche). Reactions were prepared in a total volume of 20 $\mu$L with 5 $\mu$L of 50 times diluted cDNA, 0.6 $\mu$L of each primer to a final concentration of 300 nM, and 10 $\mu$L of 2x SYBRGreen mix. Amplification was achieved by a 10-min initial enzyme activation at 95°C followed by 40 cycles composed of 30 s of denaturation at 95°C and 1 min of primer annealing/extension at 60°C. To verify the specificity of the amplification of each primer pair, melting-curve analysis was performed ranging from 60°C to 95°C with temperature increasing in steps of 0.1°C $s^{-1}$. Data were analyzed with version 2.1 of the StepOne software (Applied Biosystems), and gene expression levels relative to the *18S* rRNA reference gene were calculated by the comparative cycle threshold method (Schmittgen and Livak, 2008).

Gene-specific primers (Supplemental Table S1) were designed at the end of the coding sequence and/or in the 3′ UTR. In the case of homeologous gene pairs, PCR product identity was verified by cloning and sequencing. Whenever possible, intron-spanning primers were chosen to ascertain the absence of genomic DNA contamination. For primers not fulfilling this criterion, controls without reverse transcriptase were performed.

### Laser-Capture Microdissection and RT-PCR

From the region of maximum width of a fully expanded leaf 4, 1-$cm^2$ sections were fixed in acetone and paraffin embedded as described (Ohtsu et al., 2007). Epidermal and mesophyll subepidermal cells were microdissected from 10-$\mu$m sections using the Arcturus XT infrared laser-capture microdissection system with the following settings for epidermal/mesophyll cells, respectively: laser spot size, 10/20 $\mu$m; laser pulse duration, 20/30 ms; and laser power, 50/70 mW. About 5,000 epidermal cells (predominantly adaxial) and 2,500 mesophyll cells were collected. For SAM microdissection, maize shoot apices were dissected from 2-week-old plants in a block of 3 × 2 mm and fixed using the same procedures. Sections of 8 $\mu$m were spread on PEN 1.0 membrane slides (Zeiss). About 250,000 $\mu m^2$ of L1 and 500,000 $\mu m^2$ of L2 cells were captured into an AdhesiveCap 500 tube (Zeiss) using the PALM MicroBeam system with the following parameters: cut at 36 to 50 UV-Energy/65 UV-Focus and laser pressured catapulting at 61 to 75 UV-Energy/63 UV-Focus. In both cases, RNA was extracted with the PicoPure RNA isolation kit (Arcturus). RNA samples were treated with DNase I (Qiagen) and amplified (two rounds) with the TargetAmp 2-Round aRNA Amplification kit 2.0 (Epicentre Biotechnologies). Amplified RNA was reverse transcribed and amplified by PCR as described above.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers BK008026 to BK008042 and JN003608 and JN003618.

### Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Deduced amino acid sequences of HD-ZIP IV proteins in maize, Arabidopsis, rice, sorghum, *Selaginella*, and *Physcomitrella*.

**Supplemental Figure S2.** Synteny between maize, rice, and sorghum *HD-ZIP IV* genes.

**Supplemental Figure S3.** Prediction of base pairing between the conserved 19- and 21-nucleotide motifs in the 3′ UTR of *ZmHDZIV* genes.

**Supplemental Table S1.** Primers used in this study.

**Supplemental Table S2.** Expression of 17 *ZmHDZIV* genes in major organs of the maize plant as determined by qRT-PCR.

**Supplemental Table S3.** Conservation of 21- and 19-nucleotide sequences in the 3′ UTR of *HD-ZIP IV* genes from different species.

## ACKNOWLEDGMENTS

We thank Christine Arnould and the members of the Microscopy Centre at INRA Dijon for essential help in the microdissection experiments and Doreen Ware for advice on the bioinformatic analysis of putative small RNA precursors. Isabelle Desbouchages, Alexis Lacroix, and Priscilla Angelot are acknowledged for maize culture and Hervé Leyral and Claudia Bardoux for the preparation of buffers and media. qRT-PCR experiments were carried out at the Genetic Analysis Platform (responsible: Bariza Blanquier) of the Structure Fédérative de Recherche Biosciences Gerland-Lyon Sud.

Received June 21, 2011; accepted August 6, 2011; published August 8, 2011.

## LITERATURE CITED

**Abe M, Katsumata H, Komeda Y, Takahashi T** (2003) Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development **130:** 635–643

**Ariel FD, Manavella PA, Dezar CA, Chan RL** (2007) The true story of the HD-Zip family. Trends Plant Sci **12:** 419–426

**Axtell MJ, Snyder JA, Bartel DP** (2007) Common functions for diverse small RNAs of land plants. Plant Cell **19:** 1750–1769

**Bocobza S, Adato A, Mandel T, Shapira M, Nudler E, Aharoni A** (2007) Riboswitch-dependent gene regulation and its evolution in the plant kingdom. Genes Dev **21:** 2874–2879

**Bowman JL** (2004) Class III HD-Zip gene regulation, the golden fleece of ARGONAUTE activity? Bioessays **26:** 938–942

**Chevenet F, Brun C, Bañuls AL, Jacq B, Christen R** (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics **7:** 439

**Chung BY, Simons C, Firth AE, Brown CM, Hellens RP** (2006) Effect of 5′UTR introns on gene expression in Arabidopsis thaliana. BMC Genomics **7:** 120

**Di Cristina M, Sessa G, Dolan L, Linstead P, Baima S, Ruberti I, Morelli G** (1996) The Arabidopsis Athb-10 (GLABRA2) is an HD-Zip protein required for regulation of root hair development. Plant J **10:** 393–402

**Dong YH, Yao JL, Atkinson RG, Morris BA, Gardner RC** (1999) Mdh3 encoding a Phalaenopsis O39-like homeodomain protein expressed in ovules of Malus domestica. J Exp Bot **50:** 141–142

**Edgar RC** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:** 1792–1797

**Gray J, Bevan M, Brutnell T, Buell CR, Cone K, Hake S, Jackson D, Kellogg E, Lawrence C, McCouch S, et al** (2009) A recommendation for naming transcription factor proteins in the grasses. Plant Physiol **149:** 4–6

**Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL** (2008) The Vienna RNA websuite. Nucleic Acids Res **36:** W70–W74

**Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ, Chen XY** (2008) The HD-Zip IV gene GaHOX1 from cotton is a functional homologue of the Arabidopsis GLABRA2. Physiol Plant **134:** 174–182

**Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC** (2007) Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. Plant Cell **19:** 926–942

**Ingouff M, Farbos I, Lagercrantz U, von Arnold S** (2001) PaHB1 is an evolutionary conserved HD-GL2 homeobox gene expressed in the protoderm during Norway spruce embryo development. Genesis **30:** 220–230

**Ingouff M, Farbos I, Wiweger M, von Arnold S** (2003) The molecular characterization of PaHB2, a homeobox gene of the HD-GL2 family expressed during embryo development in Norway spruce. J Exp Bot **54:** 1343–1350

**Ingram GC, Boisnard-Lorig C, Dumas C, Rogowsky PM** (2000) Expression patterns of genes encoding HD-ZipIV homeo domain proteins define specific domains in maize embryos and meristems. Plant J **22:** 401–414

**Ingram GC, Magnard JL, Vergne P, Dumas C, Rogowsky PM** (1999) ZmOCL1, an HDGL2 family homeobox gene, is expressed in the outer cell layer throughout maize development. Plant Mol Biol **40:** 343–354

**Isaacson T, Kosma DK, Matas AJ, Buda GJ, He Y, Yu B, Pravitasari A, Batteas JD, Stark RE, Jenks MA, et al** (2009) Cutin deficiency in the tomato fruit cuticle consistently affects resistance to microbial infection and biomechanical properties, but not transpirational water loss. Plant J **60:** 363–377

**Ito M, Sentoku N, Nishimura A, Hong SK, Sato Y, Matsuoka M** (2002) Position dependent expression of GL2-type homeobox gene, Roc1: significance for protoderm differentiation and radial pattern formation in early rice embryogenesis. Plant J **29:** 497–507

**Ito M, Sentoku N, Nishimura A, Hong SK, Sato Y, Matsuoka M** (2003) Roles of rice GL2-type homeobox genes in epidermis differentiation. Breed Sci **53:** 245–253

**Javelle M, Vernoud V, Depège-Fargeix N, Arnould C, Oursel D, Domergue F, Sarda X, Rogowsky PM** (2010) Overexpression of the epidermis-specific homeodomain-leucine zipper IV transcription factor Outer Cell Layer1 in maize identifies target genes involved in lipid metabolism and cuticle biosynthesis. Plant Physiol **154:** 273–286

**Javelle M, Vernoud V, Rogowsky PM, Ingram GC** (2011) Epidermis: the formation and functions of a fundamental plant tissue. New Phytol **189:** 17–39

**Kedde M, van Kouwenhove M, Zwart W, Oude Vrielink JA, Elkon R, Agami R** (2010) A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility. Nat Cell Biol **12:** 1014–1020

**Khaled AS, Vernoud V, Ingram GC, Perez P, Sarda X, Rogowsky PM** (2005) Engrailed-ZmOCL1 fusions cause a transient reduction of kernel size in maize. Plant Mol Biol **58:** 123–139

**Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T** (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science **303:** 521–523

**Kubo H, Kishi M, Goto K** (2008) Expression analysis of ANTHOCYANINLESS2 gene in Arabidopsis. Plant Sci **175:** 853–857

**Kubo H, Peeters AJ, Aarts MG, Pereira A, Koornneef M** (1999) ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis*. Plant Cell **11:** 1217–1226

**Long D, Lee R, Williams P, Chan CY, Ambros V, Ding Y** (2007) Potent effect of target structure on microRNA function. Nat Struct Mol Biol **14:** 287–294

**Lu P, Porat R, Nadeau JA, O'Neill SD** (1996) Identification of a meristem L1 layer-specific gene in *Arabidopsis* that is expressed during embryonic pattern formation and defines a new class of homeobox genes. Plant Cell **8:** 2155–2168

**Masucci JD, Rerie WG, Foreman DR, Zhang M, Galway ME, Marks MD, Schiefelbein JW** (1996) The homeobox gene GLABRA2 is required for position-dependent cell differentiation in the root epidermis of Arabidopsis thaliana. Development **122:** 1253–1260

**Mukherjee K, Brocchieri L, Bürglin TR** (2009) A comprehensive classification and evolutionary analysis of plant homeobox genes. Mol Biol Evol **26:** 2775–2794

**Mukherjee K, Bürglin TR** (2006) MEKHLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins. Plant Physiol **140:** 1142–1150

**Nadeau JA, Zhang XS, Li J, O'Neill SD** (1996) Ovule development: identification of stage-specific and tissue-specific cDNAs. Plant Cell **8:** 213–239

**Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T** (2006) Characterization of the class IV homeodomain-leucine zipper gene family in Arabidopsis. Plant Physiol **141:** 1363–1375

**Ohtsu K, Takahashi H, Schnable PS, Nakazono M** (2007) Cell type-specific gene expression profiling in plants by using a combination of laser microdissection and high-throughput technologies. Plant Cell Physiol **48:** 3–7

**Pabo CO, Sauer RT** (1992) Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem **61:** 1053–1095

**Prigge MJ, Clark SE** (2006) Evolution of the class III HD-Zip gene family in land plants. Evol Dev **8:** 350–361

**Rerie WG, Feldmann KA, Marks MD** (1994) The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis. Genes Dev **8:** 1388–1399

**Ronquist F, Huelsenbeck JP** (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:** 1572–1574

**Salse J, Abrouk M, Bolot S, Guilhot N, Courcelle E, Faraut T, Waugh R, Close TJ, Messing J, Feuillet C** (2009) Reconstruction of monocotyledonous proto-chromosomes reveals faster evolution in plants than in animals. Proc Natl Acad Sci USA **106:** 14908–14913

**Schena M, Davis RW** (1992) HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily. Proc Natl Acad Sci USA **89:** 3894–3898

**Schmittgen TD, Livak KJ** (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc **3:** 1101–1108

**Schnable JC, Springer NM, Freeling M** (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA **108:** 4069–4074

**Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al** (2009) The B73 maize genome: complexity, diversity, and dynamics. Science **326:** 1112–1115

**Svoboda P, Di Cara A** (2006) Hairpin RNA: a secondary structure of primary importance. Cell Mol Life Sci **63:** 901–908

**Sweredoski M, DeRose-Wilson L, Gaut BS** (2008) A comparative computational analysis of nonautonomous helitron elements between maize and rice. BMC Genomics **9:** 467

**Swigonová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J** (2004) Close split of sorghum and maize genome progenitors. Genome Res **14:** 1916–1923

**Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH** (2008)

Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res **18:** 1944–1954

**Valle EM, Gonzalez DH, Gago G, Chan RL** (1997) Isolation and expression pattern of hahr1, a homeobox-containing cDNA from Helianthus annuus. Gene **196:** 61–68

**Vernoud V, Laigle G, Rozier F, Meeley RB, Perez P, Rogowsky PM** (2009) The HD-ZIP IV transcription factor OCL4 is necessary for trichome patterning and anther development in maize. Plant J **59:** 883–894

**Voinnet O** (2009) Origin, biogenesis, and activity of plant microRNAs. Cell **136:** 669–687

**Wang Q, Dooner HK** (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proc Natl Acad Sci USA **103:** 17644–17649

**Yang L, Bennetzen JL** (2009) Distribution, diversity, evolution, and survival of helitrons in the maize genome. Proc Natl Acad Sci USA **106:** 19922–19927