Journal of Heredity 2008:99(4):407–416
doi:10.1093/jhered/esn013
Advance Access publication February 28, 2008

Expression and Nucleotide Diversity of the Maize *RIK* Gene

Brent Buckner, Kayleigh A. Swaggart, Cheryl C. Wong, Heath A. Smith, Kelsey M. Aurand, Michael J. Scanlon, Patrick S. Schnable, and Diane Janick-Buckner

From the Department of Biology, Truman State University, 100 East Normal Street, Kirksville, MO 63501 (Buckner, Swaggart, Wong, Smith, Aurand, and Janick-Buckner); the Department of Plant Biology, Cornell University, Ithaca, NY 14850 (Scanlon); the Department of Agronomy, Iowa State University, Ames, IA 50011 (Schnable); and the Department of Genetics, Developmental Biology and Cell Biology, Iowa State University, Ames, IA 50011 (Schnable).

Address correspondence to Dr B. Buckner at the address above, or e-mail: bbuckner@truman.edu.

Abstract

The K homology (KH) domain is a conserved sequence present in a wide variety of RNA-binding proteins. The rough sheath2–interacting KH domain (RIK) protein of maize has been implicated in the maintenance of the repressed chromatin state of *knox* genes during leaf primordia initiation. The amino acid sequences of the publicly available plant RIK proteins contain a splicing factor 1 (SF1)–like KH domain core sequence motif that distinguishes them from all other SF1-like KH domain–containing proteins. We demonstrate that the maize *RIK* gene exhibits surprisingly little nucleotide sequence diversity among *Zea* species and subspecies. Microarray hybridization experiments demonstrate that *RIK* has a higher level of expression in the shoot apical meristem as compared with 14-day seedling. Reverse transcriptase–polymerase chain reaction analysis of *RIK* indicates that the gene is expressed in many tissues, albeit at lower levels in older leaf samples. Taken together, these data suggest that the RIK protein may be involved in the maintenance of an inactive chromatin state of *knox* and possibly other genes in nonmeristematic tissues.

The K homology (KH) motif is a common RNA-binding domain that was first described in the human hnRNP K protein (Siomi et al. 1993). KH domains are approximately 60 amino acids long and contain the highly conserved consensus sequence V/IIG*xx*G*xx*I/V, where *x* can be any amino acid, although positively charged amino acids are preferred (Burd and Dreyfuss 1994; Grishin 2001). This motif is in a variety of proteins ranging from ribosomal proteins to transcriptional modifiers and has been described in Bacteria, Archaea, and Eukaryotes (Burd and Dreyfuss 1994).

The type I KH domains can be divided into 4 distinct subfamilies: vigilin like (Dodson and Shapiro 1997), polynucleotide phosphorylase (PNPase) like (Stickney et al. 2005), poly(C)-binding protein (PCBP) like (Makeyev and Liebhaber 2002), and splicing factor 1 (SF1) like (Liu et al. 2001). KH domains can be found in one to multiple copies per protein. SF1 specifically recognizes the intron branch point sequence UACUAAC in the pre-mRNA transcripts during spliceosome assembly (Liu et al. 2001). Proteins that have the SF1-like KH domain contain a single KH domain. In a bioinformatics analysis of the *Arabidopsis* genome, Lorković and Barta (2002) identified 26 genes that encode KH domain proteins. All these contained either PCBP-like or SF1-like KH domains, with At5g51300 being identified as the *Arabidopsis* SF1 homologue. No vigilin-like or PNPase KH domain proteins were described in this study; however, Walter et al. (2002) cloned and characterized an *Arabidopsis* chloroplast PNPase.

Phelps-Durr et al. (2005) identified and characterized homologues of an SF1-like KH domain protein found in both *Arabidopsis* and maize. These proteins were named rough sheath2–interacting KH domain (RIK) protein because the maize protein was found to physically interact with the maize rough sheath2 (RS2) protein. Likewise, the *Arabidopsis* RIK protein physically interacts with the asymmetric leaf1 (AS1) protein, the *Arabidopsis* homologue of maize RS2. Furthermore, RS2 physically interacts with histone cell cycle regulation defective homolog A (HIRA), a chromatin-remodeling protein. Repression of *knox* genes within the peripheral zone of the shoot apical meristem (SAM) is required for proper leaf development to proceed and occurs through the activity of AS1 in *Arabidopsis* (Byrne et al. 2000) and RS2 in maize (Timmermans et al. 1999; Tsiantis et al. 1999). This repression has been proposed to occur through a DNA replication independent chromatin-remodeling complex that includes the proteins RS2, RIK, HIRA, and AS2 (Phelps-Durr et al. 2005). Based, in part, on

the fact that RIK contains an RNA-helicase domain and a KH domain, which both bind RNA, Phelps-Durr et al. (2005) hypothesized that RIK might bind a silencing RNA and thus contribute to the recruitment of the DNA replication independent chromatin-remodeling complex to the *knox* genes, resulting in maintenance of their repression.

In this study, we describe the genomic sequence of the maize *RIK* gene and the deduced amino acid sequence of the RIK protein. We demonstrate that the amino acid sequences of plant RIK proteins contain an SF1-like KH domain sequence motif that distinguishes them from other SF1-like KH domain proteins. We demonstrate that the 3′-end of the *RIK* gene exhibits surprisingly little nucleotide sequence diversity among *Zea* species and subspecies. Lastly, our qualitative reverse transcriptase–polymerase chain reaction (RT-PCR) analysis of *RIK* indicates that the gene is expressed in many maize tissues. This suggests that the RIK protein may be involved in the maintenance of an inactive chromatin state of *knox* and possibly other genes in nonmeristematic tissues.

Materials and Methods

Plant Material

The inbred line B73 or the offspring of a self-pollinated genetic stock that segregated for the *rs2* mutant reference allele (Timmermans et al. 1999; Tsiantis et al. 1999) and had been extensively backcrossed into a B73 line were used as the sources of all tissues for mRNA isolations. In all, 7 North American inbred lines, 5 New Mexican, and 5 Mexican open-pollinated landraces, obtained from the North Central Regional Plant Introduction Station in Ames, IA, were evaluated in the nucleotide diversity study (Table 1). In addition, several *Zea* species and subspecies (collectively referred to as teosinte) were also analyzed in this study (Table 1).

Primer Design

All primers were designed using the PRIMER3 program (http://workbench.sdsc.edu/). The sequence, target, and thermal cycling parameters of all primers used in this study are included in Supplementary Table 1.

Diversity Study Experimental Design

DNA was isolated from 2 or more plants from each accession of *Zea* species and subspecies (Table 1) and amplified using Extract-N-Amp (Sigma-Aldrich, St Louis, MO). Amplifications were performed as indicated in Supplementary Table 1. Amplified products were cloned, and 3 or more clones from each independent PCR reaction were sequenced (described below). This experimental design was adopted to detect multiple alleles potentially present in the open-pollinated landraces and teosintes and to confirm sequence polymorphisms via independent PCR reactions.

Table 1. Genotypes analyzed in nucleotide diversity study

Species	Accession	Name	Plant type	Location
Zea mays subsp. *mays*	PI 550473	B73	IB	Iowa
	PI 558532	Mo17	IB	Missouri
	PI 587144	Mo24W	IB	Missouri
	PI 531084	NC258	IB	North Carolina
	Ames 19288	Oh43	IB	Ohio
	PI 550518	T8	IB	Tennessee
	PI 587154	W153R	IB	Wisconsin
	PI 218167	Acoma Pueblo	LR	New Mexico
	PI 476867	Isleta	LR	New Mexico
	PI 218133	Mesita Pueblo	LR	New Mexico
	PI 218153	San Felipe Pueblo	LR	New Mexico
	PI 476868	Blue corn (Taos)	LR	New Mexico
	Ames 19485	Bolita	LR	Guanajuato, Mexico
	PI 620859	Chapalote	LR	Sonora, Mexico
	Ames 19539	Conico	LR	Jalisco, Mexico
	Ames 6089	Blando en Gordo	LR	Chihuahua, Mexico
	PI 628502	Gigante Gordo	LR	San Luis Potosi, Mexico
Zea mays subsp. *parviglumis*	Ames 21800		Teo	Guerrero, Mexico
	Ames 21889		Teo	Jalisco, Mexico
Zea mays subsp. *mexicana*	PI 566683		Teo	Mexico, Mexico
	PI 566686		Teo	Guerrero, Mexico
Zea luxurians	PI 306617		Teo	Jutiapa, Guatemala
	PI 441933		Teo	Jutiapa, Guatemala
	Ames 21866		Teo	Guatemala
	Ames 21877		Teo	Jutiapa, Guatemala
	Ames 21879		Teo	Chiquimula, Guatemala

IB, inbred line; LR, Mexican or New Mexican landrace; Teo, teosinte.

Singletons that were not replicated within the entire allelic data set were excluded from the study.

Cloning

All RT-PCR and PCR products that were to be cloned and sequenced were electrophoresed on agarose gels, then excised and purified using Ultrafree®-DA (Millipore, Billerica, MA) centrifugal filter units. The purified DNA was concentrated by alcohol precipitation (Sambrook et al. 1989), ligated into the pGEM®-T Easy Vector, and transformed into JM109 competent cells (Promega, Madison, WI).

DNA Sequencing

Plasmid DNA was isolated from overnight cultures using the QIAprep Spin Miniprep protocol (Qiagen, Valencia, CA). Plasmid DNA was sequenced at a concentration of 0.25 μg/μL at the DNA Facility of the Iowa State University Office of Biotechnology, Ames, IA, using an ABI 3730 DNA Analyzer (Applied Biosystems Inc., Foster City, CA).

Protein Sequence Analysis

Multisequence alignment analyses were performed using CLUSTALX (Thompson et al. 1997). Pairwise alignment parameters were set to a gap penalty of 35 and a gap extension of 0.75. Multiple alignment parameters were set to a gap opening of 15, a gap extension of 0.3, and a delay divergent sequences to 25%. Boxshade diagrams were produced using BOXSHADE version 3.21 available at http://www.ch.embnet.org/software/BOX_form.html. Trees were either constructed in CLUSTALX and visualized using TREEVIEW (Page 1996) or constructed and visualized using PAUP* (Swofford 2002). Amino acid composition and formula weight were determined using the statistical analysis of protein sequences method of Brendel et al. (1992).

Nucleotide Diversity

The statistics π and θ were used to estimate nucleotide diversity. π is a function of the average pairwise difference among sequences in a sample (Nei 1987), whereas θ is a function of both the number of polymorphic sites and the number of sequences in a sample (Watterson 1975). Because these values were very similar and the gene exhibited a very low level of diversity, only π values are discussed. Tajima's *D* (Tajima 1989) test, total π, silent π, and synonymous and nonsynonymous π were calculated using DNASP (Rozas et al. 2003).

Microarray Data

Differentially regulated genes identified via microarray experiments (Ohtsu et al. 2007; Zhang et al. 2007) were annotated as previously described (Buckner et al. 2007) at http://sam.truman.edu.

The RT-PCR

To evaluate *RIK* expression, a representative group of tissues was selected and characterized by RT-PCR from the inbred line B73. In addition, we also compared the expression of *RIK* in 2- and 6-week-old *rs2* mutant and wild-type sibling plants. In all cases, maize tissues were excised, quickly frozen using liquid nitrogen, ground into a fine powder, and stored at −80 °C until analysis. For all 7-day tissues, kernels were germinated on germination paper in a dark incubator set at 30 °C. Fourteen-day seedlings were grown in a growth chamber with a 17-h light and 7-h dark cycle at 27 °C. Total RNA was isolated from frozen ground tissues using the TRIzol™ protocol (Invitrogen, Carlsbad, CA). RNA was then quantified based on absorption at 260 nm. Quantified RNA was reverse transcribed into cDNA utilizing the SuperScript™ First-Strand Synthesis system (Invitrogen) with the supplied polythymine primers. cDNA was amplified via PCR using the primers shown in Supplementary Table 1. The PCR products were separated by electrophoresis on a 1.0% agarose gel and photographed.

Results

Genomic Sequence of *RIK*

Previously, Phelps-Durr et al. (2005) described the RIK protein sequence. However, the full-length genomic sequence of maize *RIK* has not been previously described or deposited as an annotated sequence in GenBank. Three noncontiguous Maize Assembled Genomic Islands (MAGIs; Fu et al. 2005; http://magi.plantgenomics.iastate.edu/) that contain nearly perfect sequence identity to the beginning (MAGI4_148256), middle (MAGI4_106073), and end (MAGI4_114303) of the *RIK* cDNA were identified. These MAGIs had been assembled from DNA sequencing of inbred line B73 (Fu et al. 2005). Primers were used to amplify the inbred line B73 genomic sequences linking these MAGI sequences as indicated in Supplementary Table 1. The amplified DNA was cloned, sequenced, and assembled with the MAGI sequences into one contiguous B73 *RIK* genomic sequence (Supplementary Figure 1). We also used the maize *RIK* cDNA sequence (AY940679) to identify a partially sequenced B73 maize bacterial artificial chromosome (BAC) (AC194252) that contains the *RIK* gene. The maize *RIK* cDNA (AY940679) was compared with the *RIK* genomic sequence both manually and using SPIDEY (http://www.ncbi.nlm.nih.gov/).

Here we report only the genomic sequence that was verified by 2 or more studies (e.g., BAC, MAGI, or our own sequencing experiments). Because only 125 bp upstream of the transcriptional initiation site is available, no bioinformatics analysis of the upstream regulatory elements of this gene is presented. The *RIK* gene of maize is composed of 12 exons, which is consistent with exon number in the *Arabidopsis* (At3g29390) and *Oryza sativa RIK* gene sequences (*japonica* AC134241; *indica* CT828823). Phelps-Durr et al. (2005) reported a 7-bp repeat of A's at positions 1850 to 1856 in their deposited maize cDNA sequence (AY940679). However, BAC AC194252, MAGI4_114303, various maize expressed sequence tags (ESTs) available in the National Center for Biotechnology Information (NCBI) (e.g., AY108824) and

MAGI databases (e.g., MEC_75115_P95-Mar06), and all the diversity study sequence presented below indicate that this is, in fact, a 6-bp repeat of A's (Supplementary Figure 1). This sequence discrepancy is in exon 11 of the *RIK* sequence and causes a frameshift in the deduced C-terminal portion of the RIK protein. The sequence presented in Supplementary Figure 1 represents the true C-terminal end of the maize RIK protein. The full-length RIK protein contains 657 amino acids and has a predicted molecular weight of 69.9 KDa.

Plant RIK Proteins

The maize RIK protein and cDNA sequences were used in BLASTP and TBLASTN searches, respectively, to identify RIK protein sequences in the NCBI database. A single *RIK* gene was identified in both the *Oryza sativa indica* and *japonica* genomic sequences and is located on chromosome 3. The deduced *indica* and *japonica* RIK amino acid sequences differ in only a single amino acid; R 228 (relative to the rice amino acid sequences), a polar, positively charged R-group is replaced in *japonica* by W, a nonpolar, aromatic R-group. Interestingly, this amino acid is in the KH domain and is the amino acid that precedes the KH domain conserved core sequence (Figure 1). For simplicity, only the *indica* sequence is presented in Figures 1 and 2. Additionally, a cDNA sequence (BT013455) that encodes the entire RIK protein of tomato was identified. A *Vitis vinifera* RIK-like protein sequence (AM467023) was identified, but it appeared to be N-terminally truncated and was not included in our final analyses. The plant RIK proteins (Figure 1) contain significant similarity beyond their SF1-like KH domains. The C-terminal ends of the RIK proteins are well conserved and rich in P and D. Interestingly, the human SF1 protein contains a conserved proline-rich C-terminal motif that has been shown to physically interact with the human transcription factor CA150 (Goldstrohm et al. 2001).

Phylogenetic Analysis of the SF1-like KH Domain

For each of the KH domain maize genes that were differentially regulated in the microarray studies (described below; Table 2) a maize EST contig (MEC) was identified from the MAGI database. The deduced amino acid sequences representing the KH domains of the maize MECs were aligned to all the *Arabidopsis* KH domains described by Lorković and Barta (2002) and Phelps-Durr et al. (2005). In the phylogenetic tree derived from this alignment, the maize and *Arabidopsis* SF1-like KH domains form a distinct clade supported by bootstrap analysis (Supplementary Figure 2). Figure 2A shows the multi-sequence alignment of the SF1 RNA-binding domain, including the highly conserved KH domain core consensus sequence V/IIG*xx*G*xx*I/V, when comparing SF1-like KH domains from the *Arabidopsis* genome, rice, and tomato RIK and the maize SF1-like KH domains from the genes that were differentially regulated in the microarray studies (Table 2). A phylogenetic tree representing the relationship of these SF1-like KH domains is shown in Figure 2B. The SF1-like KH domains of plant RIK proteins form a distinct clade from the other SF1-like KH domains of maize and *Arabidopsis* that is well supported by bootstrap values.

Nucleotide Diversity at the 3′-end of the *RIK* Gene

An approximately 500-bp region of *RIK*-containing sequence from exon 11, intron 11, and exon 12 was amplified and sequenced from several *Zea* species and subspecies (Table 1). This region of the gene encodes the C-terminal proline-rich domain of RIK that is conserved in plants (Figure 1). Fourteen different *RIK* haplotypes were observed; however, 8 haplotypes were represented by a single individual (Figure 3 and Supplementary Table 2). Only 6 single-nucleotide polymorphisms (SNPs) were detected in this region. In all, 5 of these SNPs were found in exons and 4 of these resulted in nonsynonymous changes to the amino acid sequence (Figure 3). The substitutions at positions 5946 and 5967 replace T with I and P with L, respectively, both of which are nonconservative changes. However, the substitution at position 5922 replaces a D with G, which is a weak group conservation, whereas the substitution at position 5987 replaces F with L, which is a strong group conservation. The impact of these nonsynonymous changes on the function of the RIK protein is unknown. Only one indel was observed, a 1-bp indel in the intronic sequence of haplotype 11. The frequency of SNPs per base pair in the *RIK* sequence examined was calculated by dividing the total number of SNPs by the length of DNA sequence evaluated. We identified 1 SNP every 77.7 bp. Tenaillon et al. (2001) reported an average of 1 SNP every 27.6 bp when considering 21 loci from chromosome 1 (*RIK* is also on chromosome 1). Thus, by comparison, this portion of the *RIK* gene displays a strikingly low frequency of SNPs.

Three different microsatellite repeats, one 3 bp (ACC) and two 6 bp (TGCCAC and CTAAGG), were observed in exon 11 (Figure 3). The 3-bp microsatellite is found either as a 3- or 4-unit repeat encoding LPP or LPPP, respectively. The 6-bp microsatellite TGCCAC is found as a 2-, 3-, or 4-unit repeat encoding LPLPP, LPLPLPP, or LPLPLPLPP, respectively. The additional amino acids encoded by these repeats are L and P, both of which are already abundant in the C-terminal end of the RIK protein (Figure 1). Thus, these microsatellite repeats may not significantly alter the functional properties of the RIK protein. The 6-bp microsatellite CTAAGG, which is only repeated in haplotype 7, results in the addition of 2 amino acids, A and K (KEE becomes KAKEE). It is interesting to note that inbred lines Mo17, Mo24W, and NC258 each contain 2 different haplotypes of this region, which can be distinguished by microsatellite repeat variation within the ACC and TGCCAC repeats (Figure 3). It is possible that near-identical paralogues of *RIK* exist in these lines; however, it is also possible that this microsatellite repeat is unstable. Instability of microsatellite repeats has been previously described in maize (Matsuoka, et al. 2002, Vigouroux et al. 2002).

Various π values were used to evaluate the nucleotide sequence diversity of *RIK*; all π values calculated for this

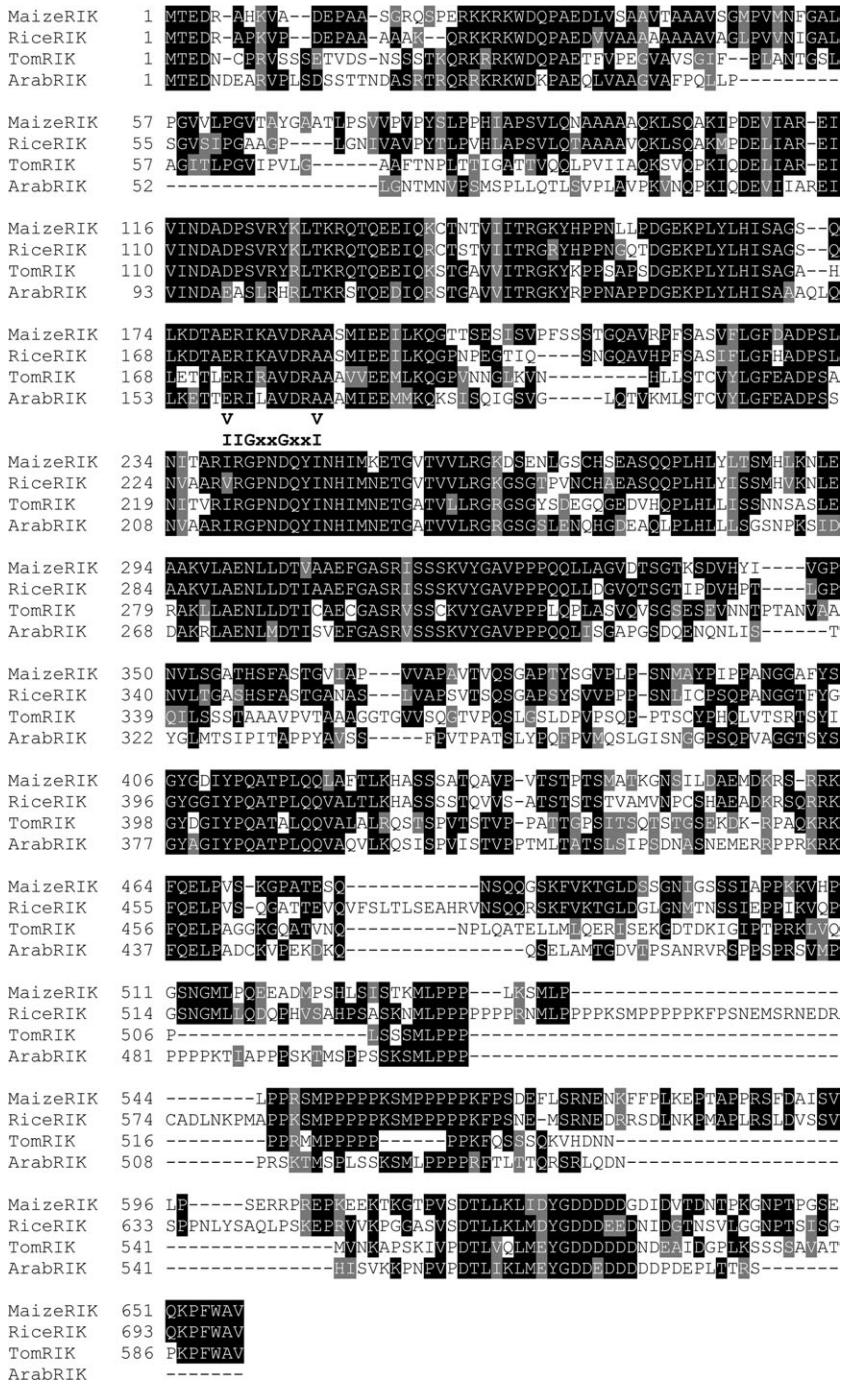

Figure 1. Multisequence alignment of the RIK proteins from maize, rice (EAY89032), *Arabidopsis* (AAY24687), and tomato (BT013455). Alignment was performed using CLUSTALX and BOXSHADE version 3.21 as described in Materials and Methods. The KH domain core consensus sequence is designated above the multisequence alignment. Dark shading indicates identical amino acids, and light shading indicates similar amino acids.

portion of *RIK* were remarkably low in both maize and teosinte (Table 3). However, *RIK* did exhibit a modest decrease in total π and silent π when comparing teosinte to all maize sequences. An analysis of the nucleotide diversity of the KH domain–encoding region among the inbred lines used in this study exhibited a similar paucity of diversity (data not shown). Maize genes that are thought to have been under selection during the domestication and improvement of maize exhibit similarly low π values, but typically these genes show a more dramatic difference when compared

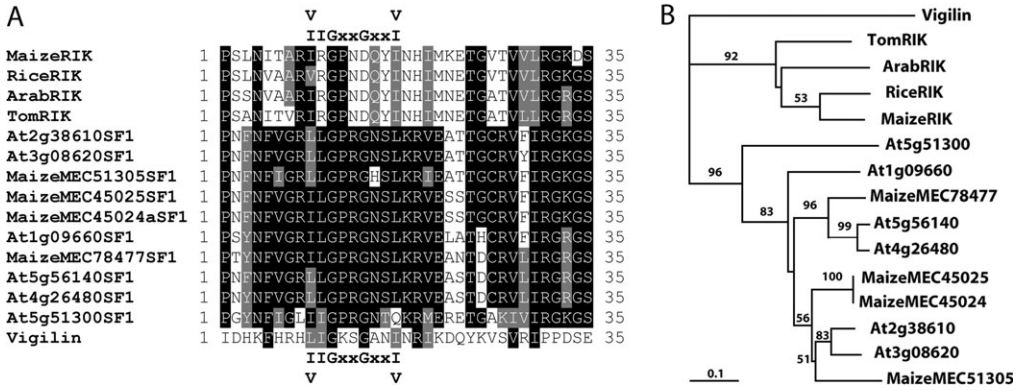

Figure 2. Multisequence alignment and phylogenetic analyses of the RIK SF1-like KH domain. (**A**) Multisequence alignment of the portion of the KH domain containing the SF1-like RNA-binding domain and the highly conserved KH domain core consensus sequence V/IIG*xx*G*xx*I/V. *Arabidopsis* sequences were those identified by Lorkovic and Barta (2002) and Phelps-Durr et al. (2005), whereas maize sequences were those differentially regulated in the microarray studies (Table 2). The SF1-like KH domains of maize, rice (EAY89032), *Arabidopsis* (AAY24687), and tomato (BT013455) RIK amino acid sequences were derived from sequences described in Results and aligned as described in Materials and Methods. (**B**) A phylogenetic tree representing the relationship of the SF1-like KH domains is aligned in Figure 2A. The tree was constructed as described in Materials and Methods. Bootstrap values for nodes supported in >50% of 10 000 bootstrap replicates are given above the branch leading to those nodes. Human vigilin (GI:1431721) was used as an out-group and root. Sequences are designated with the *Arabidopsis* locus number or maize MEC number. RIK sequences used the following abbreviations: Tom, tomato; Arab, *Arabidopsis*.

with teosinte. The North American inbred lines and Mexican open-pollinated landraces studied here were chosen to compare favorably with a subset of those studied by Tenaillon et al. (2001, 2002) and to represent a good cross section of the diversity represented in the North American inbred lines and Mexican germplasms. Tenaillon et al. (2001) characterized nucleotide diversity at 21 loci from chromosome 1. Although the nucleotide diversity of these 21 loci ranged more than 13-fold, the loci with the lowest diversities, including loci thought to be under selection, were comparable to that found in *RIK*. In our study, π values calculated separately for an approximately 469-bp region of the maize *zeaxanthin epoxidase* gene, which was amplified, cloned, and sequenced from the same accessions of maize, and usually the same plant, exhibit high nucleotide diversity characteristic of many maize genes (e.g., total π 14.65 and silent π 20.94). Thus, the low π values calculated for *RIK* are characteristic of the gene, not the population history of the plants studied. When compared with the majority of similarly analyzed maize genes (reviewed by Wright and Gaut 2005), *RIK* showed low nucleotide diversity, which is consistent with purifying selection at this locus.

The Tajima's *D* statistic (Tajima 1989) can help to estimate if selection is taking place at a gene sequence or if a population has experienced a recent change in size. The Tajima's *D* values for *RIK* in this study were not significantly different from zero, showing no evidence for change in the population size or any particular pattern of selection. However, the Tajima's *D* statistic is higher in maize populations relative to teosintes, a trend that has been observed for other maize genes (Tenaillon et al. 2004; Hufford et al. 2007) and is anticipated for genes that have undergone a domestication bottleneck causing a loss of low frequency variants. Thus, based on Tajima's *D* value, we would conclude that *RIK* is experiencing drift-mutation equilibrium.

Table 2. KH domain protein-encoding genes differentially regulated in the SAM of maize

EST accession	MEC[a]	KH type[b]	SAM/seedling[c]
CA829446	78477	SF1	0.4
DV942395	51305	SF1	3.4
BM340627[d]	**75115**	**SF1**	**3.4**
DY400798	45024[e]	SF1	3.7
CB351654	45024[e]	SF1	3.3
CB329430	45025	SF1	4.9
DN207585	15353	PCBP (2)	2.0
CD661999	22818	KHI (2)	3.0
DN218585	40246	KHI	0.5
DV491122	29843[e]	KHI	1.7
DV942176	29843[e]	KHI	1.8
DN214431	56444	KHI	1.9

[a] MEC from http://magi.plantgenomics.iastate.edu/.

[b] KH domain type designation used in NCBI conserved domain feature of BLASTP.

[c] Microarray hybridization experiment comparing laser-dissected SAM tissue to above ground parts of 14-day seedling (Buckner et al. 2007; Ohtsu et al. 2007). Fold change values are indicated.

[d] Maize *RIK* is highlighted in bold.

[e] ESTs that shared significant sequence identity to the same MEC were considered a single gene in Results.

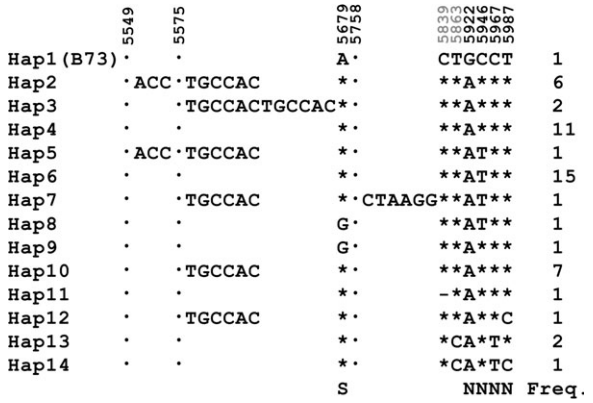

Figure 3. Nucleotide polymorphism analysis of the 3′-end of the maize *RIK* gene. Sequence chart showing all polymorphisms found in the 3′-region of *RIK* analyzed in 26 accessions of *Zea* and *Zea* subspecies. Inbred line B73 (haplotype 1) is used as reference and its sequence is shown at the top. An asterisk under the sequence indicates that the base at this position is the same as in the reference sequence. The numbers above this sequence indicate the position of polymorphisms relative to the genomic sequence of *RIK* (Supplementary Figure 1). Black and gray numbers are exonic and intronic polymorphism positions, respectively. The numbers above a dot indicate the last base of the previous microsatellite unit repeat. The presence of any additional microsatellite unit repeats is shown after the dot for relevant haplotypes. Below the polymorphisms, S and N indicate synonymous and nonsynonymous SNPs, respectively. Numbers to the right of the sequences indicate the number of different plants that exhibit that particular haplotype.

SAM Microarray Studies

Microarray hybridization experiments aimed at understanding the complex genetic circuitry that regulates the maintenance of the SAM and the initiation and development of leaf primordia have been described (Ohtsu et al. 2007; Zhang et al. 2007). The annotation of the differentially regulated genes from these and other studies (Buckner et al. 2007) are available at http://sam.truman.edu. These studies confirmed that the *RIK* gene, as well as several other genes encoding KH domain proteins, is differentially expressed in the SAM when compared with whole seedlings (Table 2). In all, 8 of the 10 genes that were annotated as KH domain proteins were upregulated in the SAM. Five of these KH domain proteins contain an SF1-like KH domain, 4 of these being upregulated. The *RIK* gene was upregulated 3.4-fold in laser-captured SAM tissue when compared with whole seedling. It is not entirely surprising that these KH domain proteins are differentially regulated during SAM development. In recent years, a number of KH domain proteins have been demonstrated to play a role in floral development (HEN4, Cheng et al. 2003), vegetative growth and gynoecium formation (PEPPER, Ripoll et al. 2006), and flowering time (FLK, Lim et al. 2004; Mockler et al. 2004).

RT-PCR Expression Analysis

Phelps-Durr et al. (2005) reported that the *RIK* gene was abundantly expressed in SAM-enriched tissue (SAM and leaf primordia P0 through P4) and young leaves (leaf primordial P5 through P8), whereas it was expressed to a significantly lesser extent in the leaf blade and sheath from 2-week old leaf tissue. In contrast, they also observed that expression of the *RIK* gene was modestly increased in sheath tissue and slightly decreased in blade tissue of *rs2* mutants when compared with wild type. The expression studies of Phelps-Durr et al. (2005) were carried out in inbred line B73 and an *rs2* mutant introgressed into a B73 background (Phelps-Durr TL, personal communication). Our RT-PCR experiments support and extend the expression differences described by Phelps-Durr et al. (2005). The SAM-enriched tissue in our study contains the SAM and several leaf primordia (most likely P0 through P8). Thus, this sample is likely to represent a composite of the SAM-enriched tissue and young leaf tissue studied by Phelps-Durr et al. (2005). In our studies and those of Phelps-Durr et al. (2005), the *RIK* gene is expressed to a slightly higher degree in SAM-enriched tissue as compared with leaf blade and sheath (Figure 4A). The microarray studies also indicate a somewhat (3.4-fold) higher expression of *RIK* in SAM versus whole seedling (Table 2). Our evaluation of 6-week-old leaf blade and sheath expression indicates that the expression of *RIK* decreases as compared with 2-week-old tissue (Figure 4A).

Table 3. Nucleotide diversity statistics for *RIK*

	n	π_{total}[a]	π_{silent}[a]	$\pi_{synonymous}$[a]	$\pi_{nonsynonymous}$[a]	θ_{total}[a]	θ_{silent}[a]	*D*[b]
All maize	30	1.53	0.74	1.49	2.07	1.63	1.45	−0.13865
Inbred	10	1.58	0.00	0.00	2.60	1.52	0.00	0.12030
New Mexican	10	1.19	0.00	0.00	1.97	0.76	0.00	1.46364
Mexican	10	1.91	2.04	4.11	1.89	1.52	2.03	0.83017
All teosinte	21	2.05	1.48	0.00	2.46	2.40	1.60	−0.40863
Zea mays subsp. *parviglumis*	5	0.86	0.00	0.00	1.42	1.03	0.00	−0.81650
Zea mays subsp. *mexicana*	8	1.46	0.00	0.00	2.40	1.66	0.00	−0.44794
Zea luxurians	8	2.84	3.07	0.00	2.78	2.49	2.21	0.58467

[a] Values reported × 10^3.

[b] No values were statistically significant.

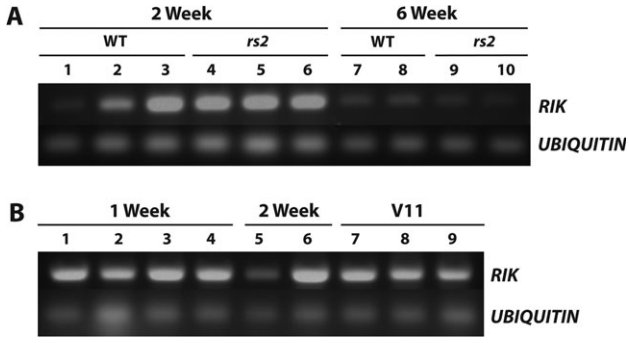

Figure 4. RT-PCR expression analyses of *RIK*. (**A**) *RIK* and *ubiquitin* expression in 2- and 6-week-old wild-type (B73) and *rs2* (B73 background) seedlings. RT-PCR was performed using *RIK* and *ubiquitin* primers from the following tissues of wild-type and *rs2* plants: lanes 1 and 4, sheath from leaves 3 and 4 of 2-week-old seedlings; lanes 2 and 5, blade from leaves 3 and 4 of 2-week-old seedlings; lanes 3 and 6, SAM-enriched tissue of 2-week-old seedlings; lanes 7 and 9, sheath from leaf 5 of 6-week-old seedlings; lanes 8 and 10, blade from leaf 5 of 6-week-old seedlings. (**B**) *RIK* and *ubiquitin* expression in 1- and 2-week-old B73 seedlings and V11 stage B73 plants. RT-PCR was performed using *RIK* and *ubiquitin* primers on the following tissues: lane 1, 1-week-old shoot; lane 2, 1-week-old SAM-enriched tissue; lane 3, 1-week-old root; lane 4, 1-week-old root apical meristem enriched tissue; lane 5, 2-week-old leaf; lane 6, 2-week-old SAM-enriched tissue; lane 7, V11 stage leaf; lane 8, V11 stage developing ear; and lane 9, V11 stage developing tassel.

To further evaluate *RIK* expression, a representative group of tissues was characterized by RT-PCR from inbred B73 plants (Figure 4B). These studies indicate that the *RIK* gene is expressed in all tissues evaluated but that transcripts of *RIK* are found in lower abundance in older leaf tissues (Figure 4). In addition, 454 ESTs from MAGI4_114303 (which contains a portion of the *RIK* gene) were obtained from laser capture microdissected SAM and tapetal cells (http://magi.plantgenomics.iastate.edu/).

Discussion

KH domain proteins from *Arabidopsis* have been described using a bioinformatics approach (Lorković and Barta 2002). However, the study of Lorković and Barta (2002) did not identify the *RIK* gene, which was isolated during yeast two-hybrid protein interaction studies using RS2 as a bait protein (Phelps-Durr et al. 2005). It is likely that the *RIK* gene was not included in the bioinformatics analyses of Lorković and Barta (2002) because the KH domain of *RIK* is distinctly different from all other SF1-like KH core motifs (Figure 2 and Supplementary Figure 2). The KH domain core sequence is highly conserved in the plant RIK SF1-like KH domains (V/IRGPNDQYI), although there are differences compared with the canonical V/IIG*xx*G*xx*I/V core sequence. The canonical I, a nonpolar R-group, in the second amino acid position is replaced by an R, which contains a polar, positive R-group. This is not a conserved change. In addition, the canonical second G, which contains a tiny R-group is replaced by a D, which contains a small polar negative R-group. This is at best a weak conservation. Thus, the KH domain core consensus sequence of the RIK proteins, which is part of the RNA-binding domain, is distinct and contributes to assigning them to a separate clade from all other plant SF1-like KH domains that have been described. Structural and functional characterizations of the RIK proteins and their interaction with target RNAs will be necessary to determine if these core sequence amino acid substitutions have a biologically significant impact on the specificity and function of the RIK proteins. However, we suggest that the SF1-like KH domain of RIK proteins is sufficiently distinct that they might be better described as RIK-like KH domains.

It is interesting to note that a single *RIK* gene has been identified in the complete *Arabidopsis* and rice genomic sequences. Both these plants behave genetically as diploids but are thought to be diploidized paleopolyploids (Blanc and Wolfe 2004b; Paterson et al. 2004; Maere et al. 2005). However, many genes remain duplicated in both genomes and likely represent ancient homeologous paralogues. Maize also behaves as a true diploid, yet evidence from the study of the evolution of maize paralogues supports that maize is an ancient segmental allotetraploid (Gaut and Doebley 1997; Gaut et al. 2000). Studies in *Arabidopsis* have demonstrated that genes of certain functional classes exhibit high retention after a genome duplication event (e.g., transcription, signal transduction, and protein modification), whereas genes in other functional categories (e.g., RNA binding, DNA metabolism, and nuclease activity) exhibit low retention rates (Blanc and Wolfe 2004a; Seoighe and Gehring 2004; Maere et al. 2005). The finding that the *RIK* gene, which encodes a presumptive RNA-binding protein, is present as a single copy in the *Arabidopsis* and rice genomes is consistent with the loss of its homeologous paralogue. Although it is speculative, it is interesting to suggest that there might be a selective advantage to retaining only a single copy of *RIK* during a diploidization event. It will be interesting to determine the *RIK* gene copy number in the soon to be completed maize genomic sequence as well as in true polyploids, including both natural and synthetic polyploids and their progenitors.

Domestication and improvement population bottlenecks initially significantly reduced the genetic diversity of maize (Eyre-Walker et al. 1998; Tenaillon et al. 2004). The nucleotide diversity of maize genes has been further reduced by the positive selection of genes involved in the traits associated with the domestication (White and Doebley 1999; Clark et al. 2004; Tenaillon et al. 2004) and improvement events. Many maize genes have been found to exhibit a paucity of genetic diversity with a genetic signature characteristic of loci under selection (Wright et al. 2005; Hufford et al. 2007).

The *RIK* gene displayed a low frequency of SNPs per base pair and a low nucleotide diversity (Table 3 and Figure 3),

which are consistent with purifying selection at this locus. However, the Tajima's *D* statistic was consistent with neutral equilibrium at the *RIK* locus. It is reasonable to expect that our small sample size and the paucity of genetic diversity that exists at *RIK* would limit the likelihood that this test would detect selection at this locus.

Despite these observations, it is intriguing to note that the *RIK* gene is remarkably conserved between maize and teosinte. The loss of sequence diversity in maize compared with its wild relatives varies depending on the locus of study (Whitt et al. 2002; Zhang et al. 2002; Tenaillon et al. 2004). However, the portion of *RIK* gene analyzed in this study is also highly conserved even between maize and *Tripsicum dactyloides*. In addition, the portion of the RIK protein encoded by this sequence is highly conserved when comparing maize, rice, *Arabidopsis*, and tomato (Figure 1). Taken together, these observations suggest strong functional constraints (i.e., purifying selection) on the 3′-region of the *RIK* gene rather than more recent selection during the domestication or improvement of maize. RIK is known to interact with the non-myb domain region of RS2; however, it has not been determined which portion of RIK is involved in this interaction (Phelps-Durr et al. 2005). The proline-rich C-terminal motif of the human SF1 protein interacts with transcription factor CA150 (Goldstrohm et al. 2001). Thus, we suggest that the conserved proline-rich C-terminal region of RIK is a strong candidate for being involved in functionally significant protein–protein interactions.

Class I *knox* genes are essential for establishing and maintaining the indeterminance of the SAM; repression of *knox* genes within the peripheral zone of the SAM is an early event in the commitment of the leaf founder cells to give rise to leaf primordia (Timmermanns et al. 1999; Tsiantis et al. 1999; Byrne et al. 2000). Although the function of the RIK protein remains to be experimentally determined, it has been hypothesized that the RIK protein binds to a silencing RNA and contributes to the recruitment of the DNA replication independent chromatin-remodeling complex to the *knox* genes, whereon the *knox* genes are maintained in a repressed chromatin state (Phelps-Durr et al. 2005). Laser capture microarray studies of Ohtsu et al. (2007) indicate that the *RIK* gene expression is 3.4-fold upregulated in the SAM when compared with whole seedling (Table 2). These studies do not preclude that *RIK* is expressed in both these tissues (i.e., SAM and seedling), and indeed, our RT-PCR analyses indicate that *RIK* was expressed in all tissues investigated (Figure 4). In situ hybridization studies would help to illuminate the spatial expression pattern of the *RIK* gene, whereas immunohistochemical staining studies would identify the tissues in which the RIK protein accumulates. Studies such as these may help to clarify if differential expression of this gene is required for repressing the *knox* genes. The ubiquitous expression pattern for the *RIK* gene could suggest that either a silencing RNA or some other RIK-interacting RNA or protein is differentially expressed and limits the involvement of the RIK protein in the *knox* repression mechanism. In addition, *RIK* is expressed in a wide variety of tissues where class I *knox* genes are not expressed. Thus, the potential exists that the RIK protein could be involved in the maintenance of an inactive chromatin state of *knox* and possibly other genes in nonmeristematic tissues.

Supplementary Material

Supplementary Tables 1 and 2 and Figures 1 and 2 can be found at http://www.jhered.oxfordjournals.org/.

Funding

National Science Foundation (DBI-0321711 to P.S.S. and DBI-0321515 to M.J.S.).

Acknowledgments

The authors wish to thank Lisa Grantham, Ashley Lough, and B.B.'s Truman State University 2007 Genetics class for their help in characterizing this gene. The authors are also grateful to Dr Anton Weisstein for his discussions of phylogenetic trees and diversity statistics, as well as his advice constructing phylogenetic trees and their interpretation.

References

Blanc G, Wolfe KH. 2004a. Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell. 16:1679–1691.

Blanc G, Wolfe KH. 2004b. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell. 16:1667–1678.

Brendel V, Bucher P, Nourbakhsh IR, Blaisdell BE, Karlin S. 1992. Methods and algorithms for statistical analysis of protein sequences. Proc Natl Acad Sci USA. 89:2002–2006.

Buckner B, Beck B, Browning K, Fritz A, Grantham L, Hoxha E, Kamvar Z, Lough A, Nikolova O, Schnable PS, et al. 2007. Involving undergraduates in the annotation and analysis of global gene expression studies: creation of a maize shoot apical meristem expression database. Genetics. 176:741–747.

Burd CG, Dreyfuss G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. Science. 265:615–621.

Byrne ME, Barley R, Curtis R, Arroyo JM, Dunham M, Hudson A, Martienssen RA. 2000. *Asymmetric leaves1* mediates leaf patterning and stem cell function in Arabidopsis. Nature. 408:967–971.

Cheng Y, Kato N, Wang W, Li J, Chen X. 2003. Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in Arabidopsis thaliana. Dev Cell. 4:53–66.

Clark RM, Linton E, Messing J, Doebley JF. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proc Natl Acad Sci USA. 101:700–707.

Dodson RE, Shapiro DJ. 1997. Vigilin, a ubiquitous protein with 14 K homology domains, is the estrogen-inducible vitellogenin mRNA 3′-untranslated region-binding protein. J Biol Chem. 272:12249–12252.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci USA. 95:4441–4446.

Fu Y, Emrich SJ, Guo L, Wen T-J, Ashlock DA, Aluru S, Schnable PS. 2005. Quality assessment of maize assembled genomic islands (MAGIs) and large-scale experimental verification of predicted genes. Proc Natl Acad Sci USA. 102:12282–12287.

Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA. 94:6809–6814.

Gaut BS, Le Thierry d'Ennequin M, Peek AS, Sawkins MC. 2000. Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci USA. 97:7008–7015.

Goldstrohm AC, Albrecht TR, Suñé C, Bedford MT, Garcia-Blanco MA. 2001. The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. Mol Cell Biol. 21:7617–7628.

Grishin NV. 2001. KH domain: one motif, two folds. Nucleic Acids Res. 29:638–643.

Hufford KM, Canaran P, Ware DH, McMullen MD, Gaut BS. 2007. Patterns of selection and tissue-specific expression among maize domestication and crop improvement loci. Plant Physiol. 144:1642–1653.

Lim MH, Kim J, Kim YS, Chung KS, Seo YH, Lee I, Kim J, Hong CB, Kim HJ, Park CM. 2004. A new Arabidopsis gene, FLK, encodes an RNA binding protein with K homology motifs and regulates flowering time via FLOWERING LOCUS C. Plant Cell. 16:731–740.

Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, Zanier K, Krämer A, Sattler M. 2001. Structural basis for recognition of the intron branch site RNA by splicing factor 1. Science. 294:1098–1102.

Lorković ZJ, Barta A. 2002. Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. Nucleic Acids Res. 30:623–635.

Maere S, De Bodt S, Raes J, Casneuf J, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA. 102:5454–5459.

Makeyev AV, Liebhaber SA. 2002. The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. RNA. 8:265–278.

Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J. 2002. Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet. 104:436–450.

Mockler TC, Yu X, Shalitin D, Parikh D, Michael TP, Liou J, Huang J, Smith Z, Alonso JM, Ecker JR, et al. 2004. Regulation of flowering time in Arabidopsis by K homology domain proteins. Proc Natl Acad Sci USA. 101:12759–12764.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Ohtsu K, Smith M, Emrich S, Borsuk L, Zhou R, Chen T, Zhang X, Timmermans MCP, Beck J, Buckner B, et al. 2007. Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). Plant J. 52:391–404.

Page RDM. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosciences. 12:357–358.

Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA. 101:9903–9908.

Phelps-Durr TL, Thomas J, Vahab P, Timmermans MCP. 2005. Maize rough sheath2 and its *Arabidopsis* orthologue ASYMMETRIC LEAVES1 interact with HIRA, a predicted histone chaperone, to maintain *knox* gene silencing and determinacy during organogenesis. Plant Cell. 17:2886–2898.

Ripoll JJ, Ferrándiz C, Martínez-Laborda A, Vera A. 2006. PEPPER, a novel K-homology domain gene, regulates vegetative and gynoecium development in Arabidopsis. Dev Biol. 289:346–359.

Rozas J, Sanchez-Delvarrio JC, Messegyer X, Rozas X. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. 2nd ed.. New York: Cold Spring Harbor Laboratory Press.

Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. Trends Genet. 20:461–464.

Siomi H, Matunis MJ, Michael WM, Dreyfuss G. 1993. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. Nucleic Acids Res. 21:1193–1198.

Stickney LM, Hankins JS, Miao X, Mackie GA. 2005. Function of the conserved S1 and KH domains in polynucleotide phosphorylase. J Bacteriol. 187:7214–7221.

Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS. 2002. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Genetics. 162:1401–1413.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA. 98:9161–9166.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: A multilocus investigation of the domestication process in maize. Mol Biol Evol. 21:1214–1225.

Timmermans MCP, Hudson A, Becraft PW, Nelson T. 1999. ROUGH SHEATH2: a Myb protein that represses knox homeobox genes in maize lateral organ primordia. Science. 284:151–153.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.

Tsiantis M, Schneeberger R, Golz JF, Freeling M, Langdale JA. 1999. The maize *rough sheath2* gene and leaf development programs in monocot and dicot plants. Science. 284:154–156.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J. 2002. Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol. 19:1251–1260.

Walter M, Kilian J, Kudla J. 2002. PNPase activity determines the efficiency of mRNA3′-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. EMBO J. 21:6905–6914.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:188–193.

White SE, Doebley JF. 1999. The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. Genetics. 153:1455–1462.

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. 2002. Genetic diversity and selection in the maize starch pathway. Proc Natl Acad Sci USA. 99:12959–12962.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. Science. 308:1310–1314.

Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol. 22:506–519.

Zhang L, Peek AS, Dunams D, Gaut BS. 2002. Population genetics of duplicated disease-defense genes, *hm1* and *hm2*, in maize (*Zea mays* spp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*). Genetics. 162:851–860.

Zhang X, Madi S, Borsuk L, Nettleton D, Elshire RJ, Buckner B, Janick-Buckner D, Beck J, Timmermans M, Schnable PS, et al. 2007. Laser microdissection of narrow sheath mutant maize uncovers novel gene expression in the shoot apical meristem. PLoS Genetics. 3:1040–1050.

Received September 28, 2007
Accepted January 11, 2008

Corresponding Editor: Susan Gabay-Laughnan