*Plant Molecular Biology* **32:** 673–683, 1996.



# Three *knotted1*-like homeobox genes in *Arabidopsis*

Kyle A. Serikawa[2], Antonio Martinez-Laborda[1] and Patricia Zambryski*
*Department of Plant Biology, 111 Koshland Hall, University of California, Berkeley, CA 94720, USA (*author for correspondence);[1]Present addresses: Department Genetica y Microbiologica, Universidad de Alicante, Campus de San Juan, Aptd 374, 03080 Alicante, Spain; [2]Dept. Botany, Box 355325, University of Washington, Seattle, WA 98195, USA*

Received 26 October 1995; accepted in revised form 10 July 1996

*Key words: Arabidopsis thaliana*, homeobox, homeodomain, *kn1*, *KNAT*, light regulation

## Abstract

Five *arabidopsis kn1*-like homeobox genes were cloned through low-stringency screening of *Arabidopsis* cDNA libraries with the *kn1* homeobox from maize. These five genes were named *KNAT1-5* (for *kn1*-like *Arabidopsis thaliana*). An analysis of *KNAT1* and *2* has been presented previously [19]. Here we present an analysis of the genes *KNAT3*, *4* and *5*. On the basis of sequence and expression patterns, these three genes belong to the class II subfamily of *kn1*-like homeobox genes [16]. Low-stringency Southern analysis suggests several additional members of the class II genes exist in the *Arabidopsis* genome. The predicted amino acid sequences of the three genes share extensive homology outside of the homeodomain, including 84% between KNAT3 and KNAT4. Northern analysis shows that although all three genes are expressed in all tissues examined, the level of *KNAT3* RNA is highest in young siliques, inflorescences and roots, *KNAT4* RNA level is strongest in leaves and young siliques, and *KNAT5* RNA level is highest in roots. The specificity of these patterns was confirmed by RNA fingerprint analysis. *KNAT3* and *4* are light-regulated as they show reduced expression in etiolated seedlings and also in *hy3*, *cop1* and *det1* mutant backgrounds.

## Introduction

The study of development often and naturally leads to the study of evolutionary conservation and homology between species. Finding underlying principles of development becomes simpler when different systems can be compared for their common features. This was true for experimental embryologists at the turn of the century and is true today, the main difference being the tools and the characteristics used in studying development.

As researchers have begun describing the molecular mechanisms underlying development in a wide variety of systems, many similarities have been revealed. In some cases homology is limited to short stretches of protein sequences that appear in proteins with very different roles. In other cases similarities extend to entire coding sequences, the arrangements of genes within genomes, interactions between proteins, and general pattern-forming mechanisms. It has been suggested that a least some of these more complex levels of homology might serve as major if not predominant criteria for assigning phylogenetic groupings [1, 11]. For example, having a Hox cluster, a linear complex of homeobox genes found in mammals, *Drosophila*, nematodes and other organisms, has been proposed as a defining characteristic for metazoans [35].

Our work has focused on members of a conserved gene family, the homeobox genes of *Arabidopsis*. Homeobox genes are among the most studied of the genes known to be involved in development. Hundreds of homeoboxes have been discovered and characterized in diverse animal, fungal and plant systems [12, 18, 39]. Some homeobox genes have been closely main-

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X92392 (*KNAT3*), X92393 (*KNAT4*) and X92394 (*KNAT5*).

tained in function and sequence over millions of years of evolutionary divergence. Others appear to have been recruited and modified for new roles. The protein motif encoded by the homeobox binds DNA. Homeodomain proteins act as transcription factors [24]. Mutational analyses have demonstrated that many homeodomain proteins act as switches during development, turning cascades of genes on and off.

The plant homeobox genes present a particular challenge to study. Of the identified genes, only *KN1* and *RS1* in maize, *Hooded* in barley, and *g12*, *stm* and *be11* in *Arabidopsis* have an associated mutant phenotype [2, 22, 30, 32, 33, 43]. Furthermore, an interpretation of the maize and barley phenotypes is complex, as *KN1*, *RS1*, and *Hooded* are dominant, neomorphic mutations. *KN1* and *RS1* cause aberrant leaf morphology and *Hooded* causes the development of a second flower in each floret. *g12*, a recessive mutation, causes plants to lack normal trichomes. Plants lacking *stm* gene expression do not form vegetative meristems. *be11* results in ovules lacking outer integuments.

The plant homeobox genes without associated mutant phenotypes have been isolated in a variety of ways. Some, such as the Homeodomain-Zip class, were obtained through low stringency screening or PCR with degenerate oligonucleotides [5, 10, 27, 34, 37]. Others were obtained through cross-hybridization to *kn1* [16, 19, 23, 26], and still others such as *ZmHox1a* through binding to promoter elements [3, 17, 38].

We have focused our analysis on three *kn1*-like homeobox genes in *Arabidopsis*. As none have an associated mutant phenotype, we have chosen a comparative approach to discern their function. Expression patterns and sequence data from our studies of *Arabidopsis* are compared with similar information from maize and other species. Examining the similarities and the differences should stimulate hypotheses about how different members of the gene family may be involved in developmental patterns. Here we describe *KNAT3*, *4* and *5* (*kn1*-like *Arabidopsis thaliana*), three members of the *kn1*-like class of homeobox genes.

## Materials and methods

### *cDNA library screening, clone isolation and characterization*

We screened ca. $1\times 10^5$ plaques of a young plant (ecotype Wassilewskija) cDNA library at low stringency with the homeobox-containing fragment of the maize *kn1* gene. Hybridization was carried out in 9× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate), 2% SDS, 2 mM EDTA, 10 mM Tris-HCl pH 7.5, 200 μg/ml salmon sperm DNA, 5× Denhardt's solution, and 20 mM sodium phosphate buffer at 58 °C. Filters were washed at 58 °C in 1× SSC, 0.2% SDS for 1 h. From this screen we obtained partial cDNAs for *KNAT2* and *KNAT3*. An additional young plant cDNA library (ecotype Wassilewskija) and two flower cDNA libraries (ecotype Landsberg erecta) were screened at low stringency with the homeoboxes of *kn1*, *KNAT2* and *KNAT3*. Both seedling libraries were constructed in λZAPII and obtained from K. Feldmann. One of the flower libraries was constructed in λgt10 and the other in λZAPII. Both flower libraries were obtained from E. Meyerowitz. From these screens we obtained several additional clones of *KNAT2* and *KNAT3*, as well as clones of *KNAT1*, *KNAT4* and *KNAT5*. Clones ranged in size from 0.5 to 1.7 kb.

Both strands of cDNAs were sequenced using the dideoxy-chain termination method. Sequences presented correspond to Landsberg erecta genes. Sequencing was performed with the USB Sequenase 2.0 kit on double-stranded plasmid templates. To obtain complete sequences, 5′ and 3′ deletions were generated using the Erase-A-Base kit (Promega). Sequences were aligned and analysed with the Intelligenetics software or by hand.

### *Southern analysis and mapping*

Genomic DNA isolated from seedlings of ecotype Wassilewskija was digested with *Hin*dIII and separated by electrophoresis through agarose TBE gels at a concentration of 1–3 ug of DNA per lane. High-stringency Southerns were performed as described [36]. Low-stringency Southerns were hybridized and washed under the same conditions as the cDNA library screens.

All three genes were tested for polymorphisms between La-O and Col-O, the two parent ecotypes for the recombinant inbred lines generated by Caroline Dean and Claire Lister [21]. Polymorphisms were found with *Cla*I for KNAT3 and *Hin*dIII for *KNAT5*. DNA was made from the 100 recombinant inbred lines, digested with the appropriate enzymes and probed with *KNAT3* and *KNAT5*. The resulting segregation patterns were analysed using the Map Manager Software [25].

*RNA isolation and northern analysis*

Tissue for RNA isolations was collected from plants grown under long-day conditions (16 h light/8 h dark) at 22 °C in a Percival growth chamber, except for young plants and roots and etiolated seedlings, all of which were grown on plates of germination media (Gambourg's B5 salts with minimal organics, 1% sucrose, 0.1 mg/l biotin, 0.5 mg/l 2-[N-morpholino]ethanesulfonic acid, pH 5.8). Etiolated seedlings were grown for ten days in the dark and then collected. As tissue was collected it was quick-frozen in liquid nitrogen. The mutations *det1-1*, *cop1-4* and *hy3cno-02* were used for light expression analysis.

RNA was isolated by the method of Condit *et al.* [8]. 10 $\mu$g total RNA per lane was electrophoresed through a formaldehyde-agarose gel, transferred onto Hybond N nylon membranes (Amersham), and UV cross-linked. Membranes were hybridized with either DNA or RNA probes. DNA probes were labelled with [$\alpha$-$^{32}$P]-dCTP (3000 Ci/mmol, ICN Pharmaceuticals) using random hexamer labelling. RNA probes were generated by transcription of linearized cDNA-containing plasmids with either T7 or T3 RNA polymerase. RNA probes were labelled with [$\alpha$-$^{32}$P]-dUTP (>600 Ci/mmol, ICN Pharmaceuticals).

Hybridization conditions for DNA probes were 5× SSC, 3× Denhardt's solution, 0.5% SDS and 50 $\mu$g/ml tRNA at 65 °C. Hybridization conditions for RNA probes were as above plus 50% formamide. In both cases blots were washed in 1× SSC, 0.1% SDS for 20 min at room temperature followed by 3× 20 min in 0.2× SSC, 0.1% SDS at 65 °C. As a loading control, blots were hybridized with either a random hexamer-labelled 18S ribosomal cDNA fragment or cyclophilin cDNA [15, 20].

*RNA fingerprinting*

RNA fingerprinting analysis followed the general procedure of Bohl and Apel [4] with the following modifications. 1 pmol of the primer 5′RNAamp was added to 5 $\mu$g of total RNA in 10 $\mu$l of hybridization buffer (60 mM NaCl, 50 mM Tris-HCl pH 8.3, 10mM DTT). Samples were placed in a 90 °C water bath which was allowed to cool to ca. 30 °C over an hour's time. 2 $\mu$l of 36 mM $MgCl_2$ in hybridization buffer was added. A third of each sample was mixed with 6 $\mu$l of reverse-transcription mix (60 mM NaCl, 6 mM $MgCl_2$, 50 mM Tris-HCl pH 8.3, 10 mM DTT, 375 $\mu$M dATP, dGTP, dCTP and dTTP, and 1 unit AMV reverse transcriptase) and placed at 50 °C for 30 min.

PCR was performed in 25 $\mu$l reaction volumes using an MJ Research Programmable Thermal Cycler. Reactions contained primers at a final concentration of 1 $\mu$M each, nucleotides at a final concentration of 0.2 mM each and 1× Boehringer Mannheim *Taq* Polymerase buffer. Reactions were heated to 94 °C before 2.5 units of *Taq* Polymerase were added and the reaction overlaid with mineral oil. The cycling program was: (1) 3 min at 94 °C; (2) 20 × 1 min at 94 °C, 2 min at 54 °C, 2 min at 72 °C; (3) 5 min at 72 °C. After PCR, samples were extracted once with chloroform to remove mineral oil and precipitated with 1 volume of 4 M ammonium acetate and 2 volumes of ethanol.

Samples were resuspended in 40 $\mu$l of distilled water and a tenth of this used for labelling through one additional PCR cycle. Reaction conditions for the labelling cycle included 0.2 $\mu$M of each primer, 20 $\mu$M for dATP, dGTP and dTTP, 10 $\mu$Ci [$\alpha$-$^{32}$P]-dCTP (3000 Ci/mmol, ICN Pharmaceuticals) and 1× *Taq* Polymerase Buffer. Samples were overlaid with mineral oil and heated to 94 °C at which point 0.2 units of *Taq* Polymerase were added. The final cycle ran 3 min at 94 °C, 2 min at 45 °C and 10 min at 72 °C. Samples were again extracted and precipitated. Each sample was resuspended and digested with *Hin*fI. After digestion, 10 000 cpm were loaded and electrophoresed through a 6% non-denaturing acrylamide gel.

The two primers used were synthesized by Operon Technologies, Inc and had the following sequences. 3′RNA amp: 5′-TYGCCTTATCTTCCTCAG-3′; 5′RNAamp: 5′-TYGCCTTATCTTCCTCAG-3′. Control fingerprinting experiments were performed on plasmids containing cDNAs for *KNAT3*, *KNAT4* and *KNAT5*. Otherwise all procedures were the same.

## Results

*Sequence analysis of* Arabidopsis KNAT *genes*

The maize *kn1* homeobox was used as a probe at low stringency to isolate 5 different cross-hybridizing cDNAs from *Arabidopsis*. *KNAT1* and *2* were most similar to *kn1* and have been analysed further elsewhere [19]. The homeoboxes of *KNAT3*, *4* and *5* were sequenced and their amino acid translations compared to other known maize and *Arabidopsis* KN1-like homeodomains in Fig. 1A. The KN1-like homeodomains can be subdivided into two main classes: Class I

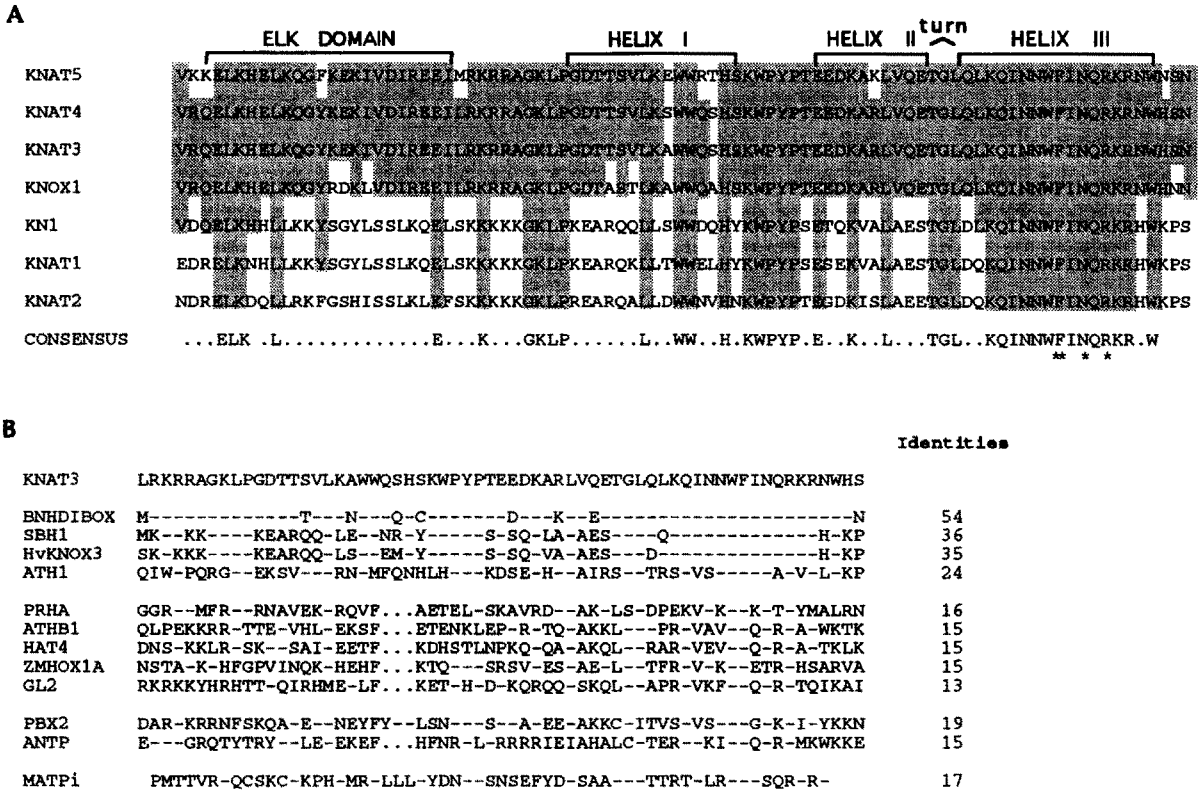


*Figure 1.* Homeodomain sequence comparisons. A. The homeodomains and ELK domains of KNAT3, 4 and 5 are shown in comparison with KNOX1 and kn1 of maize, and KNAT1 and KNAT2 of *Arabidopsis*. A consensus sequence is shown below. The four amino acids invariant in nearly all known homeodomains are marked with asterisks (*). Homologous amino acids are shaded when shared between at least three sequences. B. The homeodomain of KNAT3 is shown in alignment with several plant and animal homeodomains [3, 5, 12, 17, 23, 29–31, 33, 34, 37]. BNHDIBOX is from *Brassica napus*; SHB1 is from soybean; HvKNOX3 is from barley; ATH1, PRHA, ATHB1, HAT4 and GL2 are from *Arabidopsis*; ZMHOX1A is from maize; PBX2 is a paired-related homeodomain from humans; ANTP is from *Drosophila*; and MATPi is from yeast. A three amino acid insertion has been added to several sequences to maximize alignments. Amino acids homologous to those in KNAT3 are represented by a dash (–).

genes share greatest similarity to *kn1*, and class II genes have greatest homology to *knox1* (for *kn1*-like maize homeobox [16]). These classes predate the establishment of monocots and dicots as separate lineages since representatives of both classes are found in *Arabidopsis* (this report and [19]). The three genes we have studied belong to class II and share both sequence and expression pattern similarities with their maize counterparts. The similarities between the maize and *Arabidopsis* amino acid sequences extend N-terminal to the homeodomain into the ELK domain, a conserved 20 amino acid stretch adjacent to the homeodomain. The function of the ELK motif is unknown but has been postulated to play a role in protein-protein interactions [42]. The homeodomain of KNAT3, a representative member of the class II genes, is shown in alignment with several other plant, animal and fungal homeodomains in Fig. 1B. The divergence of the *KNAT* genes from these other homeobox genes serves to contrast how strongly the *kn1*-like genes have been conserved as a group.

A comparison of the full-length sequences of KNAT3 and KNAT4 in Fig. 2 shows remarkable conservation N-terminal to the homeodomain. An unusual feature of both polypeptides are the amino acid stretches including strings of prolines, glutamines, alanines and asparagines. Poly-proline and -glutamine stretches have been described in other transcription factors [13] and can aid in activating transcription, but the function of the other repeated stretches is unclear.

Overall, KNAT3 and KNAT4 share 84% homology (98% in the homeodomain). The level of identity between *KNAT3* and *4* is only slightly less (80%, data not shown). An interesting observation is that many


*Figure 2.* Comparison of KNAT3 and KNAT4 ORFs. The KNAT3 and KNAT4 amino acid sequences were aligned using the GENALIGN program of the Intelligenetics suite. Several regions were anchored to improve overall alignment. The strings of poly-amino acids have been underlined.

stretches of nucleotide homology are invariant, showing no third position substitutions. The lack of conservative changes may indicate that the exact nucleotide sequences of these stretches have a functional importance, such as recognition sites for nucleic acid binding proteins. We have not yet obtained a complete *KNAT5* cDNA with its entire open reading frame but the available amino acid sequence (consisting of approximately the C-terminal two-thirds of the protein) shows KNAT5 also shares many of the conserved N-terminal regions (data not shown).

*KNAT3* and *KNAT5* have been mapped by RFLP analysis with the recombinant inbred lines generated by Lister and Dean [21]. 100 lines were used in the analysis. *KNAT3* maps to the top half of chromosome 5 ca. 9.3 cM distal to marker m291 and 9.5 cM proximal to marker g4560. We have not mapped *KNAT4* as we have not yet found an RFLP associated with this gene. *KNAT5* maps to the bottom of chromosome 4 approximately 4.2 cM distal to marker g8300 and 5.7 cM proximal to marker g3088. Confidence intervals from these experiments can be found elsewhere [21].

*KNAT3*, *4* and *5* hybridize with several bands when used as probes on genomic Southern blots at low stringency as shown in Fig. 3. The probes used for low-stringency screening consisted of the homeobox of each gene and some flanking sequences. The hybridization patterns seen at low stringency overlap partially between genes: *KNAT3*, *4* and *5* cross-hybridize with each other, and each gene also detects novel fragments, suggesting several other class II *KNAT* genes exist in the genome of *Arabidopsis*. In addition to the bands detected by each probe at high stringency (arrows, Fig. 3), each lane contains several other weakly hybridizing bands. Hybridization of genomic DNA after independent digestion with a second enzyme (*Eco*RI) also revealed several weakly hybridizing bands, although not as many as seen with *Hin*dIII. After taking into account the overlap in hybridization patterns we estimate that perhaps at most ten to eleven additional class II genes exist in the *Arabidopsis* genome although the number could be substantially lower if some of the bands seen represent the products of restriction sites within a single gene rather than hybridization to independent genes.

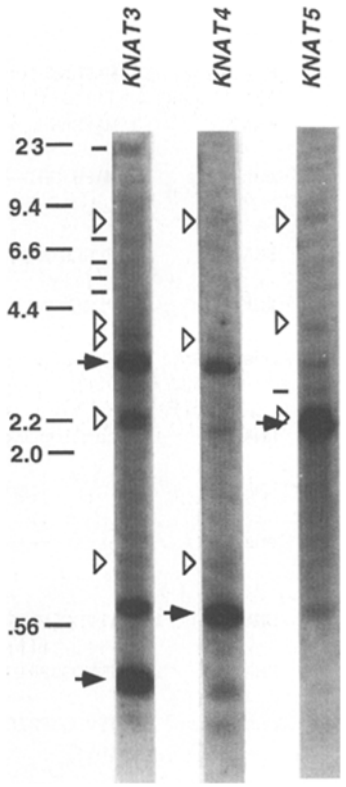


*Figure 3.* Low-stringency southern analysis. The homeobox containing regions of *KNAT3*, *4* and *5* were used as probes in low stringency Southern analysis. Genomic DNA from ecotype Wassilewskija was cut with *Hin*dIII and separated on an agarose gel and blotted. Each lane represents the hybridization patterns seen with one of the three genes. Several similarly migrating and unique bands are visible for each. In each lane, the band(s) that a probe detects at high stringency (data not shown) is marked by an arrow. These bands correspond to the gene used as the probe for that lane. In other lanes, similarly migrating bands are not marked and represent cross-hybridization among *KNAT3*, *4* and *5* under low stringency. Where non-gene-specific, similarly migrating fragments are detected by two or three probes, the fragments are labelled by open arrowheads in all lanes in which those fragments appear. Fragments only detected by one probe are labelled with bars.

*RNA expression patterns of* KNAT *genes*

We probed blots containing total RNA extracted from several plant organs. The blots were hybridized at high stringency with both RNA probes made from 3′-untranslated sequences and DNA probes made from our longest cDNAs. Both probes gave the same results and representative autoradiographs are shown in Fig. 4. Each probe detected a message of ca. 1.5–1.7 kb. Like many of the maize class II genes, *KNAT3*, *4* and *5*, are expressed in almost all organs [16] and are expressed preferentially in one or more organ types. *KNAT3* has its highest steady state message level in young siliques, inflorescences and roots. *KNAT4* signal appears strongest in leaves and young siliques. *KNAT5* RNA is found mainly in roots.

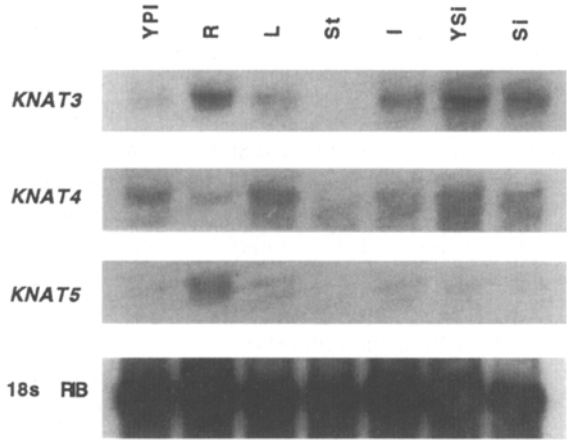


*Figure 4.* Northern analysis. RNA was isolated from different tissues and used in a northern analysis. RNA from 3-week old plants (YP1), roots (R), leaves (L), inflorescence axes (St), inflorescences (I), young (<5 mm) siliques (YSi) and older (>5 mm) siliques (Si) was isolated. Blots were probed with both RNA and DNA probes. Shown are autoradiographs from DNA probes. The bottom set of lanes represents a blot stripped and reprobed with an 18S ribosomal probe as a loading control.

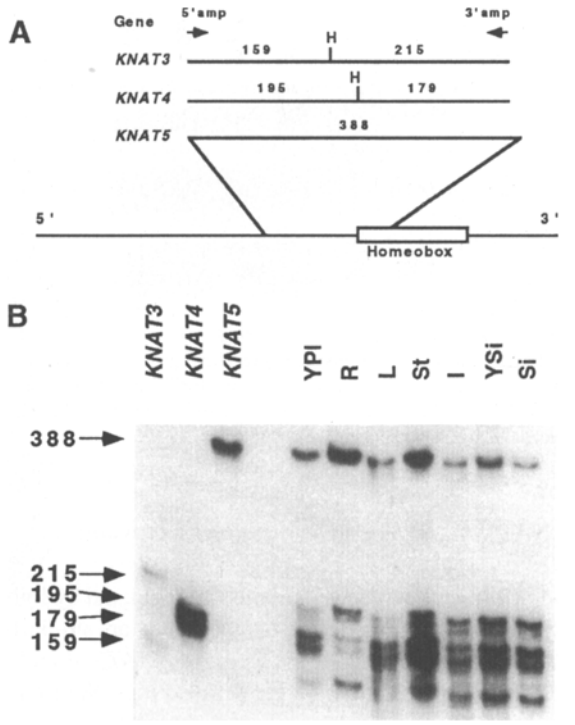


*Figure 5.* RNA fingerprint analysis. A. This diagram presents a schematic of the fingerprint analysis. Shown are the regions of the three homeobox genes to be amplified by the technique of Boh1 and Apel [4] using the primers 3′RNAamp and 5′RNAamp, and the expected sizes of the fragments produced upon digestion with *Hin*fI (H). In this technique the relative levels of expression among closely related genes within an RNA sample can be determined. B. The first three lanes are control loadings of amplified plasmid DNA labelled and cut with *Hin*fI to show their respective sizes. The next seven lanes represent amplified RNA from each of the tissue types probed in Fig. 4. Comparisons of intensity can only be made between bands in a single lane and not between separate lanes.

The high homology between the three *KNAT* genes raises the question of whether our northern analysis accurately reflects expression patterns. To rule out the possibility of cross-hybridization and confirm the above analysis, we used a PCR-based RNA fingerprinting technique [4]. This technique can be used to monitor relative expression levels between different members of a gene family within a particular RNA sample, as long as all members of a gene family are amplified equally. In this technique, universal PCR primers are designed from conserved regions in a multigene family. These primers are used to reverse transcribe and amplify total RNA. The resulting PCR fragments are amplified for one more round in the presence of radiolabelled nucleotides, and subsequently cut with a restriction enzyme giving a diagnostic restriction pattern for each member of the gene family. This strategy when applied to *KNAT3*, *4* and *5* is diagrammed in Fig. 5A. The fragments are separated on an acrylamide gel and the intensities among bands within a sample are compared. In this way, the relative levels of expression among the *KNAT* genes in a particular RNA sample can be compared.

Controls for amplification of *KNAT3*, *4* and *5* were performed using plasmid DNAs of the three genes. Each pairwise combination of genes was amplified and examined for signs of preferential amplification. *KNAT4* was found to amplify preferentially to *KNAT3* and *5* but this discrepancy could be alleviated if a small ($\leq$20) number of amplification cycles was used (data not shown). *KNAT3* and *5* amplified equally well in control reactions.

As shown in Fig. 5B, the RNA fingerprint data agree with that from our northern analysis (Fig. 4). The levels of *KNAT3* and *5* RNAs are greater in the roots than *KNAT4* RNA as expected since *KNAT3* and *5* are high in roots and *KNAT4* is not; in leaves, *KNAT4* RNA is present at a higher level; and in inflorescences, young siliques and siliques, *KNAT3* and *4* RNAs are found at higher levels than that of *KNAT5*.

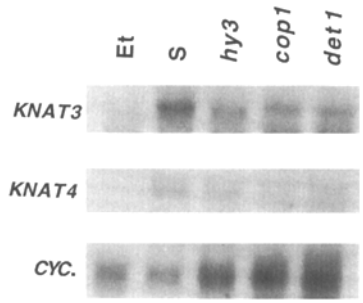


*Figure 6.* Light regulation of *KNAT3* and *4*. RNAs from etiolated (Et) and light-grown (S) wild-type seedlings, light-grown *hy3*, and dark-grown *det1* and *cop1* mutant plants grown for 10 days were probed with *KNAT3* and *4*, and cyclophilin control. The autoradiograph shown is a six day exposure.

### KNAT3 *and* 4 *show light-regulated expression*

Some homeobox genes in plants are regulated by light [5, 31]. To see if any of the genes we study are similarly regulated, we examined the levels of *KNAT3*, *4* and *5* RNA in etiolated and light grown seedlings. We found both *KNAT3* and *4* RNAs were greatly reduced in etiolated tissues, as shown in Fig. 6. Levels of *KNAT5* remained unchanged (data not shown). To determine what aspect of light-regulated growth is necessary for normal expression of *KNAT3* and *KNAT4*, we performed a northern analysis using RNA isolated from light-grown seedlings of the mutant *hy3*, and dark-grown seedlings of the mutants *det1* and *cop1*. During early seedling development, light-grown *hy3* plants resemble etiolated seedlings but still undergo chloroplast development and chlorophyll accumulation [41]. Dark-grown *cop1* and *det1* plants have the morphology of wild-type light-grown plants but lack chlorophyll accumulation [7, 9]. *KNAT3* and *4* are expressed at reduced levels relative to wild-type light-grown seedlings in all three mutants, although their expression is not reduced to the extent seen in etiolated seedlings (Fig. 6).

## Discussion

### *Comparisons of* kn1 *homeobox genes between maize and* Arabidopsis *reveal similarities in sequence and general expression patterns*

Here we characterize three *Arabidopsis kn1*-like homeobox genes. The data suggest the *kn1*-like genes have been highly conserved over evolutionary history. Conservative estimates place the divergence between the monocots and dicots at 100 million years ago, but the class II homeodomains of *Arabidopsis* and maize only differ by 8–10 amino acids out of 61. A comparison between available maize and *Arabidopsis* amino acid sequences outside of the homeodomains shows that homology comparable to that seen within the *Arabidopsis* genes exists N-terminal to the homeodomain (data not shown, R. Kerstetter, personal communication.) A similar degree of conservation exists between the maize and *Arabidopsis* class I genes [19].

Southern analysis suggests other class II genes exist in the *Arabidopsis* genome. Further evidence comes from work by Kerstetter *et al.* [16] who assembled a dendrogram based upon the sequence similarities among the homeodomains of all *kn1*-like homeobox genes. In their dendrogram, *KNAT5* falls on the outermost branch of the class II genes while *KNAT3* and *4* are more identical to each other and to a tomato gene, *THB12* than they are to any of the four known class II maize genes. If a general one-to-one correspondence between class II members among species exists (after taking into account the duplication of the maize genome), it is likely that maize genes corresponding to *KNAT3*, *4* and *5* remain to be described. It is also probable that *Arabidopsis* genes corresponding to class II maize genes on other branches of the dendrogram have not yet been isolated. Alternatively, a lack of direct correspondence between *Arabidopsis* and maize genes may reflect the evolutionary divergence between these two species, with gene duplication events having occurred independently in one species or the other.

Do the amino acid similarities reflect conserved functions for the *kn1*-like homeobox gene classes between these different species? Both class I and class II genes display similar expression patterns between maize and *Arabidopsis*. Class I genes are mainly expressed in meristematic tissues while class II genes are expressed in most tissues, with elevated levels in some organ types (present studies as well as [16, 19]). These data suggest a functional separation between the class I and class II genes as groups, and potentially a functional conservation within each group.

In a more direct assay for function, overexpression studies with the class I genes have shown that *KNAT1* of *Arabidopsis* and *kn1* of maize produce similar phenotypes when driven by the CaMV 35S promoter in transgenic tobacco and *Arabidopsis* [19, 40]. When ectopically expressed, *kn1* and *KNAT1* cause production of shoot bearing leaves in tobacco and highly lobed

and serrated leaves in *Arabidopsis*. Given this similarity, are genes of the same class redundant? Four maize class I genes were examined through *in situ* analysis and their expression patterns found to contain both overlapping and unique domains. This finding suggests each gene has a specific and separate role [14]. Unfortunately, as *stm* is the only loss-of-function *kn1*-like homeobox mutation in maize and *Arabidopsis* [22], the question of redundancy is difficult to answer.

### KNAT3, 4 *and* 5 *may play roles in development*

No maize or *Arabidopsis* class II gene has been found to correspond to a known mutant. Despite this lack, it may be possible to clarify the functions of these genes through molecular approaches. Our northern analysis indicates that the steady-state message level of each gene is at a higher level in some organ types, possibly reflecting unique functions for each gene. *KNAT3* RNA is highest in young siliques, inflorescences and roots; *KNAT4* RNA is highest in leaves and young siliques, and *KNAT5* RNA is highest in roots.

Our analysis of the expression of *KNAT3*, *4* and *5* RNAs in light and dark-grown seedlings, and in three mutants in the light signal-transduction pathway also reveals gene specific differences. While message levels of *KNAT5* are the same in both etiolated and light-grown seedlings, both *KNAT3* and *4* are reduced in etiolated tissues. Possibly the RNA expression level of *KNAT5*, a mainly root specific gene, is the same irrespective of which light-regulated developmental pathway the shoot is following, since root development is not as greatly affected as shoot development by the presence or absence of white light.

To better understand which aspects of light-grown development result in increased *KNAT3* and *4* expression, we examined the levels of both genes in the *Arabidopsis* mutants *hy3*, *cop1* and *det1*. Plant lacking *hy3* have aspects of the etiolated phenotype when grown in white light, including elongated hypocotyls and reduced cotyledon expansion, although they still undergo chloroplast development and chlorophyll accumulation [41]. Mutants in *cop1* and *det1* fail to repress photomorphogenesis in the dark [7, 9]. The alleles of *cop1* and *det1* we used were chosen because they undergo photomorphogenesis in the dark but do not have the anthocyanin accumulation, retarded growth and adult lethality characteristic of the strongest alleles [28].

If *KNAT3* and *4* were regulated solely by photomorphogenesis, we would expect wild-type light-grown levels of both genes in the dark-grown *cop1* and *det1* mutant lines, which show photomorphogenesis in the dark, and reduced levels in light-grown *hy3*, which shows dark-grown morphogenesis. If, on the other hand, *KNAT3* and *4* were regulated in some way by chloroplast development and chlorophyll accumulation, we might expect normal RNA levels in light-grown *hy3*, which has normal chloroplast development and chlorophyll accumulation, and reduced levels in dark-grown *cop1* and *det1*, both of which show no chlorophyll accumulation.

Surprisingly, the levels of *KNAT3* and *4* are reduced in all three mutant backgrounds, although not to the extent seen in etiolated wildtype seedlings. This suggests *KNAT3* and *4* may require both photomorphogenesis and greening for normal, light-grown expression. Clearly, since levels of *KNAT3* and *4* are lower than wild-type in the *hy3* background, more than greening is necessary for the full expression of these two genes. Examinations of the expression of a *KNAT3* promoter-GUS fusion in seedlings grown under red or far-red light indicate *KNAT3* is regulated in different ways by PhyA and PhyB (manuscript in press).

### *Homeobox genes in evolution*

Homeobox genes are interesting not only because they regulate key biological pathways but also because they provide evidence for the historical pathways evolution took and constrain the directions evolution may proceed in the future. Examples can be found in studies of the homeobox genes in metazoans. Evidence suggests that some of the patterns of development among the winged insects can be traced to differences in the regulation and targets of homeobox genes like *Ubx* [6].

Nature abhors many things and one of them is originality. This is the case with homeobox genes, which arose early during the evolution of eukaryotes and appear in all major eukaryotic taxa. So far the number of homeoboxes described in plants is relatively small but evidence strongly supports that many more exist. Studying the homeoboxes of plants should provide not just an understanding of the molecular basis of development, but also a phylogenetic framework upon which to hang the diverse morphological patterns seen among plants.

## Acknowledgements

We thank Sarah Hake and Erik Vollbrecht for the initial gift of the *kn1* cDNA as a probe, Cynthia Lincoln for helpful comments and criticisms, Caroline Dean and Claire Lister for seeds of the recombinant inbred lines, Renate Schmidt for help in mapping, Doris Wagner for the *det1*, *cop1* and *hy3* seeds, X.-W. Deng for the 18S ribosomal plasmid, C. Gasser for the cyclophilin plasmids, E. Meyerowitz and K. Feldmann for the cDNA libraries, and Han Kim for help in sequencing and mapping. This research was supported by DOE grant ER13882 to P.Z., by NIH Genetics Training grant 5 T32 GMO7127-20, and by a postdoctoral fellowship of the Spanish Ministerio de Educacion y Ciencia (Ministry of Education and Science).

## References

1. Averof M, Akam M: HOM/Hox genes of *Artemia*: implications for the origin of insect and crustacean body plans. Curr Biol 3: 73–78 (1993).
2. Becraft PW, Freeling M: Genetic analysis of *Rough sheath1* developmental mutants of maize. Genetics 136: 295–311 (1994).
3. Bellman R, Werr W: Zmhoxla, the product of a novel maize homeobox gene, interacts with the *Shrunken* 26 bp *feedback* control element. EMBO J 11: 3367–3374 (1992).
4. Bohl S, Apel K: A novel fingerprint method for analyzing the expression of complex multigene families of very low transcript abundance. Plant J 3: 887–893 (1993).
5. Boivin R, Hamel F, Beauseigle D, Bellemare G: Stage-specific transcription of the homeobox gene *Bnhd1* in young tissues and flowers of *Brassica napus*. Biochim Biophys Acta 1219: 201–204 (1994).
6. Carroll S: Developmental regulatory mechanisms in the evolution of insect diversity. Development 1994 (Supplement): 217–223 (1994).
7. Chory J, Peto C, Feinbaum R, Pratt L, Ausubel F: *Arabidopsis thaliana* mutant that develops as a light-grown plant in the absence of light. Cell 58: 991–999 (1989).
8. Condit C, Hagen TJ, McKnight TD, Meagher RB: Characterization and preliminary mapping of cauliflower mosaic virus transcripts. Gene 25: 101–108 (1983).
9. Deng X-W, Caspar T, Quail P: *cop1*: a regulatory locus involved in light-controlled development and gene expression in *Arabidopsis*. Genes Devel 5: 1172–1182 (1991).
10. Feng X-H, Kung SD: Identification of differentially expressed members of tobacco homeobox families by differential PCR. Biochem Biophys Res Commun 198: 1012–1019 (1994).
11. Garcia-Fernandez J, Holland PWH: Archetypal organization of the amphioxus HOX gene cluster. Nature 370: 563–566 (1994).
12. Gehring WJ, Affolter M, Burglin T: Homeodomain proteins. Annu Rev Biochem 63: 487–526 (1994).
13. Gerber HP, Seipel K, Georgiev O, Hofferer M, Hug M, Rusconi S, Schaffner W: Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 263: 808–811 (1994).
14. Jackson D, Veit B, Hake S: Expression of maize *knotted1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. Development 120: 405–413 (1994).
15. Jorgenson RA, Cuellar RE, Thompson WF, Kavanagh TA: Structure and variation in ribosomal RNA genes of pea: characterization of a cloned rRDNA repeat and chromosomal rDNA variants. Plant Mol Biol 8: 3–12 (1987).
16. Kerstetter R, Vollbrecht E, Lowe B, Veit B, Yamaguchi J, Hake S: Sequence analysis and expression patterns divide the maize *knotted1*-like homeobox genes into two classes. Plant Cell 6: 1877–1887 (1994).
17. Korfhage U, Trezzini GF, Meier I, Hahlbrock K, Somssich IE: Plant homeodomain protein involved in transcriptional regulation of a pathogen defense-related gene. Plant Cell 6: 695–708 (1994).
18. Kornberg TB: Understanding the Homeodomain. J Biol Chem 268: 26813–26816 (1993).
19. Lincoln C, Long J, Yamaguchi J, Serikawa K, Hake S: A *knotted1*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. Plant Cell 6: 1859–1876 (1994).
20. Lippuner V, Chou IT, Scott SV, Ettinger WF, Theg SM, Gasser CS: Cytosolic and chloroplast stromal forms of cyclophilin from Arabidopsis thaliana. J Biol Chem 269: 7863–7868 (1994).
21. Lister C, Dean C: Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4: 745–750 (1993).
22. Long JA, Moan EI, Medford JI, Barton MK: A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. Nature 379: 66–69 (1996).
23. Ma H, McMullen MD, Finer JJ: Identification of a homeobox-containing gene with enhanced expression during soybean (*Glycine max* L.) somatic embryo development. Plant Mol Biol 24: 465–473 (1994).
24. Manak JR, Scott MP: A class act: conservation of the homeodomain protein functions. Development (Supplement) S61–77 (1994).
25. Manly KF: A Macintosh program for storage and analysis of experimental genetic mapping data. Mammal Genome 4: 303–313 (1993).
26. Matsuoka M, Ichikawa H, Saito A, Tada Y, Fujimura T, Kano-Murakami Y: Expression of a rice homeobox gene causes altered morphology of transgenic plants. Plant Cell 5: 1039–1048 (1993).
27. Mattsson J, Soderman E, Svenson M, Borkird C, Engstrom P: A new homeobox-leucine zipper gene from *Arabidopsis thaliana*. Plant Mol Biol 18: 1019–1022 (1992).
28. McNellis TW, von Arnim AG, Araki T, Komeda Y, Misera S, Deng X-W: Genetic and molecular analysis of an allelic series of *cop1* mutants suggests functional roles for the multiple protein domains. Plant Cell 6: 487–500 (1994).
29. Monica K, Galili N, Nourse J, Saltman D, Cleary ML: *PBX2* and *PBX3*, new homeobox genes with extensive homology to the human proto-oncogene *PBX1*. Mol Cell Biol 11: 6149–6157 (1991).
30. Muller KJ, Romano N, Gerstner O, Garcia-Maroto F, Pozzi C, Salamini F, Rohde W: The barley *Hooded* mutation caused by a duplication in a homeobox gene intron. Nature 374: 727–730 (1995).

31. Quaedvlieg N, Dockx J, Rook R, Weisbeek P, Smeekens S: The homeobox gene *ATH1* of *Arabidopsis* is derepressed in the photomorphogenic mutants *cop1* and *det1*. Plant Cell 7: 117–129 (1995).
32. Reiser L, Modrusan Z, Margossian L, Samach A, Ohad N, Haughn GN, Fischer RL: The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the *Arabidopsis* ovule primordium. Cell 83: 735–42 (1995).
33. Rerie WG, Feldmann KA, Marks MD: The *GLABRA2* gene encodes a homeo domain protein required for normal trichome development in *Arabidopsis*. Genes Devel 8: 1388–1399 (1994).
34. Ruberti I, Sessa G, Lucchetti S, Morelli G: A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif. EMBO J 10: 1787–1791 (1991).
35. Ruddle FH, Bartels JL, Bentley KL, Kappen C, Murtha MT, Pendleton JW: Evolution of Hox Genes. Annual Review of Genetics 28: 423–442 (1994).
36. Sambrook J, Fritsch EJ, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Cold Spring Harbor, NY (1989).
37. Schena M, Davis R: HD-Zip proteins: members of an *Arabidopsis* homeodomain protein superfamily. Proc Natl Acad Sci 89: 3894–3898 (1992).
38. Schindler U, Beckmann H, Cashmore AT: HAT3.1, a novel *Arabidopsis* homeodomain protein containing a conserved cysteine-rich region. Plant J 4: 137–150 (1993).
39. Scott MP, Tamkun JW, Hartzell III GW: The structure and function of the homeodomain. Biochim Biophys Acta 898: 25–48 (1989).
40. Sinha NR, Williams RE, Hake S: Overexpression of the maize homeo box gene, *Knotted-1*, causes a switch from determinate to indeterminate cell fates. Genes Devel 7: 787–795 (1993).
41. Somers DE, Sharrock RA, Tepperman JM, Quail PH: The *hy3* long hypocotyl mutant of *Arabidopsis* is deficient in phytochrome B. Plant Cell 3: 1263–1274 (1991).
42. Vollbrecht E, Kerstetter R, Lowe B, Veit B, Hake S: Homeobox genes in plant development: mutational and molecular analysis. In: Spradling AC (ed) Evolutionary conservation of Developmental Mechanisms, pp. 111–123. Wiley-Liss, New York (1993).
43. Vollbrecht E, Veit B, Sinha N, Hake S: The developmental gene *Knotted-1* is a member of a maize homeobox gene family. Nature 350: 241–243 (1991).