# Unexpected Complexity of Poly(A)-Binding Protein Gene Families in Flowering Plants: Three Conserved Lineages That Are at Least 200 Million Years Old and Possible Auto- and Cross-Regulation

**Dmitry A. Belostotsky**[1]

*Department of Biological Sciences, State University of New York, Albany, New York 12222*

Manuscript received August 16, 2002
Accepted for publication October 9, 2002

ABSTRACT

Eukaryotic poly(A)-binding protein (PABP) is a ubiquitous, essential factor involved in mRNA biogenesis, translation, and turnover. Most eukaryotes examined have only one or a few PABPs. In contrast, eight expressed PABP genes are present in *Arabidopsis thaliana*. These genes fall into three distinct classes, based on highly concordant results of (i) phylogenetic analysis of the amino acid sequences of the encoded proteins, (ii) analysis of the intron number and placement, and (iii) surveys of gene expression patterns. Representatives of each of the three classes also exist in the rice genome, suggesting that the diversification of the plant PABP genes has occurred prior to the split of monocots and dicots ≥200 MYA. Experiments with the recombinant PAB3 protein suggest the possibility of a negative feedback regulation, as well as of cross-regulation between the Arabidopsis PABPs that belong to different classes but are simultaneously expressed in the same cell type. Such a high complexity of the plant PABPs might enable a very fine regulation of organismal growth and development at the post-transcriptional level, compared with PABPs of other eukaryotes.

POLY(A)-binding protein (PABP) is ubiquitous in eukaryotes, and its function is essential in yeast (Sachs *et al.* 1987), Aspergillus (Marhoul and Adams 1996), Drosophila (Sigrist *et al.* 2000), and *Caenorhabditis elegans* (A. Petcherski and J. Kimble, personal communication). PABP participates in at least three major post-transcriptional processes: initiation of protein synthesis, mRNA turnover, and mRNA biogenesis.

The ability of PABP to stimulate translation is largely due to its interaction with the translation initiation factor eIF4G. Simultaneous interactions of eIF4G with cap-binding protein eIF4E, on the one hand, and PABP, on the other hand, bring about circularization of the mRNA, which could facilitate ribosome recycling. However, the first initiation event is also stimulated by PABP *in vitro* (Tarun and Sachs 1995). Yeast PABP-dependent mRNA circularization has been visualized by direct methods (Wells *et al.* 1998). Its functional consequences have been studied by measuring translation efficiency of reporter mRNAs that either contain or lack the 5′ cap or 3′ poly(A) in extracts containing or lacking PABP or containing variants of PABP that are unable to interact with eIF4G (Tarun *et al.* 1997). The important role of the PABP/eIF4G interaction is also supported by *in vivo* experiments in Xenopus oocytes (Wakiyama *et al.* 2000). However, the mechanism of translational stimulation may be more complex than just an interaction between PABP and eIF4G, since certain mutations uncouple these two phenomena (Kessler and Sachs 1998). Moreover, observations made in yeast strains conditionally defective in poly(A) synthesis suggest the possibility that PABP interacts directly with ribosomes (Proweller and Butler 1996). Finally, Chlamydomonas has a form of PABP (RB47) that is imported into chloroplasts and acts as a part of a message-specific translational activator complex (Yohn *et al.* 1998). Thus, PABP can stimulate translation in multiple ways.

PABP also plays a complex role in mRNA degradation. On the one hand, PABP inhibits mRNA deadenylation, as well as decapping (Bernstein *et al.* 1989; Caponigro and Parker 1995; Wilusz *et al.* 2001). According to the deadenylation-dependent decapping model (Caponigro and Parker 1996), dissociation of the last molecule of PABP disrupts the interaction between the mRNA 5′ and 3′ ends, thus enabling the decapping enzyme to attack the 5′ cap. However, circularization of the mRNA via the eIF4G/PABP interaction accounts for only a part of the inhibitory effect of PABP upon decapping, as partial inhibition can still be observed when eIF4E is prevented from interacting with the 5′ cap (Wilusz *et al.* 2001). In addition, deletion of the eIF4G-interacting domain from the yeast PABP that was tethered to the mRNA in a poly(A)-independent manner did not affect its ability to block decapping (Coller *et al.* 1998). Therefore, additional mechanisms of the inhibition of mRNA decapping by PABP must exist.

Whereas PABP inhibits deadenylation (Bernstein *et al.* 1989), it is also required for the proper rate of dead-

[1]*Address for correspondence:* Department of Biological Sciences, 1400 Washington Ave., State University of New York, Albany, NY 12222. E-mail: dab@albany.edu

enylation *in vivo* (Caponigro and Parker 1995). A possible resolution of this paradox can be envisioned if PABP actually promotes the entry of the mRNA into the decay pathway rather than accelerates deadenylation *per se*. Indeed, yeast strains lacking PABP (but viable due to bypass suppressor mutations) exhibit a temporal lag before mRNA decay commences (Caponigro and Parker 1995). This lag likely reflects a role of PABP in efficient mRNA biogenesis.

The multiplicity of the cellular functions of PABP raises the question as to which of them is essential for viability. Using cross-species complementation of the yeast *pab1* null mutant by the Arabidopsis *PAB3* cDNA, it was shown that rescue of viability required neither the restoration of poly(A)-dependent translation nor the protection of the 5′ cap from premature removal (Chekanova *et al.* 2001). However, plant PABP eliminated or at least significantly reduced the lag prior to mRNA decay in yeast cells (Chekanova *et al.* 2001). These data show that the function of PABP in mRNA biogenesis is conserved between yeast and plants and that this function alone can be sufficient for viability in yeast. However, it is also possible that PABP's functions in translation and in the control of mRNA decapping, alone or in combination, could be sufficient as well.

With the exception of these cross-species complementation studies, the function of plant PABPs *in vivo* has not been demonstrated. While the Arabidopsis PAB3 protein could protect polyadenylated RNA from 3′ → 5′ exonuclease activity *in vitro* (Chekanova *et al.* 2000), the *in vivo* relevance of this finding could be evaluated only after the relative contributions of the 5′ → 3′ and 3′ → 5′ pathways to overall mRNA decay in plants are better understood. Early observations that poly(A) tails can enhance expression of reporter mRNAs electroporated into plant protoplasts were interpreted as evidence for the role of poly(A) tails in translation in plants (Gallie 1991). However, it has recently become clear that electroporation experiments may not faithfully reflect translational stimulation (Brown and Johnson 2001). Moreover, yeast *pab1* null mutant strains exhibit poly (A)-dependent stimulation of expression of the electroporated reporter mRNAs similar in magnitude to that of the *PAB1* strains (Brown and Johnson 2001).

Most eukaryotes examined appear to have only one (*Saccharomyces cerevisiae*, *Drosophila melanogaster*), two (*C. elegans*, *Xenopus laevis*), or three (*Homo sapiens*) functional PABP genes. In contrast, eight expressed PABP genes are found in Arabidopsis. Moreover, individual members of the Arabidopsis PABP gene family exhibit a degree of sequence divergence that is unusually high for this generally well-conserved protein. Furthermore, various Arabidopsis PABPs are differentially expressed. This multiplicity, high sequence divergence, and differential expression present a broader functional potential to affect organismal growth and development than that apparent for PABPs in other eukaryotes.

## MATERIALS AND METHODS

**Sources of data:** Arabidopsis DNA sequences were from MIPS (http://mips.gsf.de/proj/thal/db/index.html). Exon boundaries in *PAB1*, *-2*, *-3*, and *-5* were experimentally verified (Belostotsky and Meagher 1993, 1996; Chekanova *et al.* 2001; D. A. Belostotsky, unpublished data). *PAB4* and *-8* cDNA sequences (generated by SPP Consortium and RIKEN) were retrieved from GenBank via SIGnAL database links (http://signal.salk.edu). Exon predictions in the *PAB6* and *-7* genes are supported by colinearity of amino acid sequences. Exon 1 and a portion of exon 2 of *PAB1* are missing in the MIPS database, whereas *PAB8* sequence is erroneously extended. Corrected *PAB1* and *PAB8* gene models were constructed using alignments with respective 5′ and 3′ end expressed sequence tags (ESTs). EST frequency data were compiled from TAIR Locus Reports and cross-checked with TIGR, SIGnAL, and GenBank databases. mRNA decay data (Gutierrez *et al.* 2002) were obtained from Stanford microarray database (http://afgc.stanford.edu/afgc_html/expression.html).

**Phylogenetic analyses:** Amino acid sequence alignments were produced with PileUp and edited with LineUp (Genetics Computer Group, Madison, WI). Phylogenetic analysis was performed using the PAUP package (Swofford 1993) v. 4.0b8 for Macintosh. Gaps were treated as "missing data." A PROTPARS matrix was used to produce a maximum parsimony phylogenetic tree from the alignable amino acid sequences. Bootstrap analysis (10,000 replicates) was run in branch-and-bound mode.

**Expression analyses:** Arabidopsis plants (cv. Columbia) were transformed (Clough and Bent 1998) with pDB414, made by subcloning the *Bcl*I fragment that includes ~2000 bp of the 5′ flanking sequence and the first 16 codons of *PAB3* open reading frame (ORF), into pBI101.2 (Jefferson *et al.* 1987). Histochemical assays (An *et al.* 1996) were performed on 10 independent transgenic lines. *In situ* hybridization was done according to the protocol developed by G. Drews and J. Okamuro (http://godot.ncgr.org/cshl-course/5-in_situ.html). Antibodies against the unique C-terminal fragment of PAB3 and immunoblotting conditions were described previously (Chekanova *et al.* 2001). This antibody is specific for PAB3. Reverse transcriptase (RT)-PCR assays were conducted as in Chekanova *et al.* (2000).

**Electrophoretic mobility shift assays:** The C-terminally $His_6$-tagged recombinant PAB3 protein, lacking its first 41 amino acids, was described previously (Chekanova *et al.* 2000). Gel mobility shift assays (Kuhn and Pieler 1996) were conducted with 2 μg/ml tRNA and 0.25 mg/ml heparin included in all binding reactions. RNA probes were generated by PCR of portions of 5′-untranslated regions (5′-UTRs) that included the A-rich segments indicated in the text, using T7 promoter sequence encoded in the sense primer, followed by *in vitro* transcription and gel purification. Bands were quantitated on a PhosphoImager. $K_d$ values were obtained by fitting the data to the equation $y = 1/(1 + K_d/x)$ with KaleidaGraph, where $y$ is the fraction of the RNA probe bound and $x$ is total protein concentration. For the *PAB5* probe, the $K_d$ was assumed to be equal to the protein concentration that gave 50% binding.

## RESULTS

**Evolutionary relationships among the Arabidopsis and rice PABP amino acid sequences suggest the existence of three ancient plant PABP gene lineages that are at least 200 million years old:** A distinctive feature of PABP is four highly conserved, tandemly arranged

**TABLE 1**

**Arabidopsis PABP gene expression data summary**

| Gene | MIPS entry | Chr. no. | Expression | Distribution of ESTs | | | | | | | | | | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AG | R | RL | L | SQ | SD | W | O | F | Σ | |
| *PAB1* | At1g34140 | 1 | Low, tissue sp. | | 1[a] | | | | | | | | 1 | Orphan |
| *PAB2* | At4g34110 | 4 | Broad, high | 3 | 16 | | | | 2 | 22 | 3[b] | 3 | 49 | II |
| *PAB3* | At1g22760 | 1 | Reprod. | | | | | | | | | 1[a] | 1 | I |
| *PAB4* | At2g23350 | 2 | Broad, high | 3 | 10 | 1 | 2 | 1 | | 4 | | | 21 | II |
| *PAB5* | At1g71770 | 1 | Reprod. | | | | | | | 1 | | 1[a] | 2 | I |
| *PAB6* | At3g16380 | 3 | Low, tissue sp. | | | | | | | | 1[c] | | 1 | III |
| *PAB7* | At2g36660 | 2 | Low, tissue sp. | | | | | | | | 1[c] | | 1 | III |
| *PAB8* | At1g49760 | 1 | Broad, high | 2 | 6 | | | 3 | 3 | 3 | | 1 | 18 | II |

Chr., chromosome; AG, above ground organs (pooled); R, roots; RL, rosette leaves; L, leaves; SQ, siliques; SD, seeds; W, whole plant; O, other tissues; F, flowers; Σ, total ESTs; sp., specific.

[a] (Belostotsky and Meagher 1996).

[b] Two ESTs from etiolated seedlings, one from suspension culture cells.

[c] This work.

RNA recognition motifs (RRMs) in the N-terminal part of the protein. The RRMs have been individually conserved during evolution; that is, each is more similar to the corresponding RRM in a PABP from a distant species than to another RRM within the same protein. By applying these criteria, eight *bona fide* PABP genes were identified in Arabidopsis using BLAST (Altschul *et al.* 1997) searches (Table 1). All of the Arabidopsis PABP genes are widely dispersed in the genome. All of them, except *PAB1* and *PAB6*, also contain a conserved C-terminal non-RRM motif. Even though they lack this motif, PAB1 and PAB6 are most likely functional PABPs. The conserved C-terminal motif of PABP is involved in numerous interactions, *e.g.*, with eRF3 (Cosson *et al.* 2001), Paip1 (Roy *et al.* 2002), and Paip2 in humans (Khaleghpour *et al.* 2001) and eRF3 and Pbp1p in yeast (Mangus *et al.* 1998; Hoshino *et al.* 1999). Nevertheless, it is dispensable *in vivo* (Sachs *et al.* 1987). Furthermore, just the first two RRMs of PABP are sufficient for high-affinity binding to oligo(A) RNA (Burd *et al.* 1991; Kuhn and Pieler 1996; Deo *et al.* 1999), as well as for reconstitution of protein synthesis *in vitro* (Kessler and Sachs 1998), and support the interactions with eIF4G (Imataka *et al.* 1998; Kessler and Sachs 1998). As shown in Figure 1, the RNP II and RNP I sequence motifs of RRM1 and RRM2 of all Arabidopsis PABPs, including PAB1 and PAB6, contain all of the conserved residues (or conservative replacements thereof) that were shown by X-ray crystallographic analysis to directly contact oligo(A) in human PABP (Deo *et al.* 1999). The only exception is a nonconservative change from histidine in human PABP to glutamine in Arabidopsis PABPs in the RNP I of RRM2. Importantly, however, this glutamine is conserved in all plant PABPs sequenced to date. Furthermore, the *in vitro* translated PAB1 protein was able to bind to poly(A) Sepharose with specificity (not shown).

A high amount of sequence divergence suggested that amino acid sequences rather than DNA sequences should be used in phylogenetic analysis of this protein family. To minimize noise, all nonhomologous regions, as well as the C-terminal domain (which is absent from *PAB1* and *PAB6*), were excluded from the sequence alignment (see supporting information at http://www.genetics.org/supplemental/). This alignment was used in a maximum-parsimony analysis. The branching order of the resulting unrooted tree (Figure 2A) allows the placement of the eight Arabidopsis PABP genes into three classes: class I composed of *PAB3* and *PAB5*; class II containing *PAB2*, *PAB4*, and *PAB8*; and class III containing *PAB6* and *PAB7*. Trees with identical topology

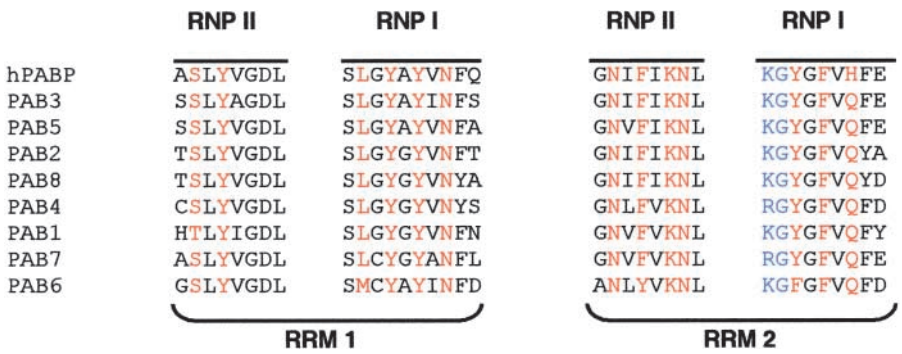


Figure 1.—Comparison of the RNP II and RNP I sequence motifs of RRMs 1 and 2 in Arabidopsis and human PABPs. Amino acids that contact RNA via side chains in the human PABP (Deo *et al.* 1999) are highlighted in red, and those that contact RNA via the main chain are highlighted in blue.

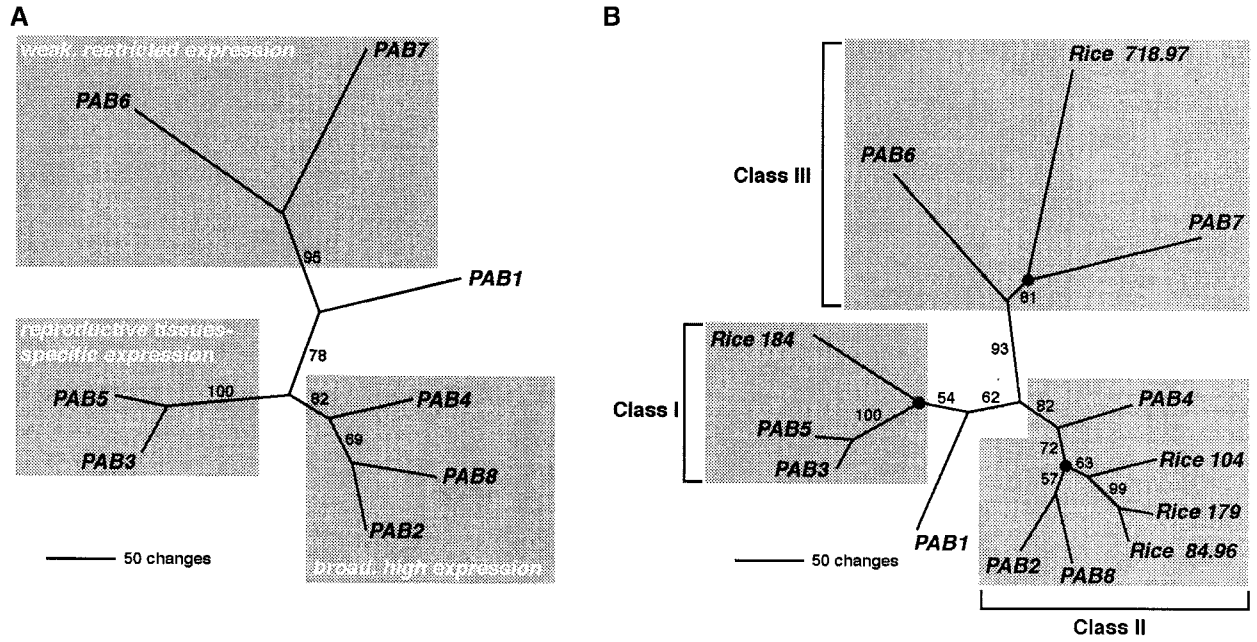


FIGURE 2.—Evolutionary relationships of the plant PABP amino acid sequences. Maximum-parsimony trees are based on the alignment of the central portion of the Arabidopsis (A) and Arabidopsis and rice (B) PABP amino acid sequences. Bootstrap values are shown next to the respective branches (10,000 replicates). Circled branch points in B represent speciation events rather than gene duplication events.

were obtained using UPGMA and neighbor-joining methods (data not shown). Moreover, bootstrap analysis (10,000 replicates) lends strong support to many aspects of this branching order. While rooting the tree by midpoint suggests that class III is basal to the class I and II sister groups (data not shown), attempts to root the tree using either metazoan or fungal PABPs as outgroups proved fruitless, since the degree of sequence divergence between most of the Arabidopsis PABPs was comparable or even exceeded the degree of divergence between a given Arabidopsis PABP and other PABPs used as outgroups. As a consequence, the relationship of *PAB1* to the rest of the Arabidopsis PABP genes remains uncertain. Deep branching suggests that it should be classified as an orphan gene.

To gain insight into when these plant PABP gene classes arose in evolution, the genome of rice, a monocot, was also examined. A total of five rice PABP genes were identified through BLAST searches (at http://portal.tmri.org/rice/) and the most conserved 317-amino-acid segment of their encoded products (see supporting information at http://www.genetics.org/supplemental/) was subjected to PAUP analysis as above. The only change in the topology of the Arabidopsis PABP gene tree upon adding the rice sequences to the data set concerns the *PAB1* gene, which moved over by one node (Figure 2B). More importantly, the resulting tree reveals that the rice genome has representatives of each of the three PABP gene classes identified in Arabidopsis: The *OsPAB184* gene is a member of class I, *OsPAB718.97* is a member of class III, and *OsPAB179*, *OsPAB104*, and *OsPAB84.96* are members of class II. Thus, the duplication events that gave rise to the three classes of the PABP genes in flowering plants must have occurred prior to the divergence of monocots and dicots ≥200 MYA (WOLFE *et al.* 1989).

**Plant PABP gene structures and their evolution:** Introns were observed in a total of 19 positions in the Arabidopsis PABP genes (Figure 3). Introns 14 (in *PAB3* and *PAB5*), 15 (in *PAB7*), and 16 (in *PAB2*, *PAB4*, and *PAB8*) occur in close, but not identical, positions relative to the coding sequence. The amino acid sequence alignment is less certain in this segment than in the RRM region. Furthermore, all of these introns occur in phase zero. Intron phase refers to its position within a codon, and phase zero introns are those that occur between codons. Phase zero introns are about twice as likely to be ancient than to result from recent insertion events (DE SOUZA *et al.* 1998). Thus, introns 14, 15, and 16 could be related by common descent and therefore are referred to as 14–16 in the ancestral gene model (Figure 3). On the other hand, introns 6 and 7, as well as introns 8 and 9, which also occur in close positions, differ in phase and therefore are considered distinct.

Although many of the introns are conserved, the differences between the intron numbers and locations allow several inferences about the evolutionary history of the Arabidopsis PABP gene family. First, the structures of the class I PABP genes (*PAB3* and *PAB5*) are identical, as are the class II PABP gene structures (*PAB2*, *PAB4*, and *PAB8*), and these two groups differ from one another by the absence of introns 2 and 12 from the

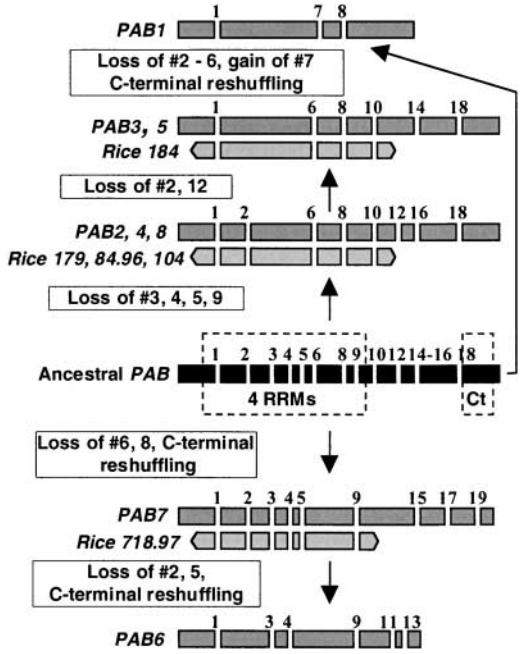


FIGURE 3.—Arabidopsis and rice PABP gene structures and a proposed model of their evolution. Exons are represented by light gray (rice), dark gray (Arabidopsis), or black (ancestral gene model) boxes. Only the central portions of the rice genes that could be unambiguously reconstructed are shown. The model assumes that introns in positions 14, 15, and 16 are related by ancestry. Regions corresponding to the four RRMs and the conserved C-terminal domain are indicated by dashed boxes in the ancestral gene model.

former. Second, class III genes, *PAB6* and *PAB7*, share in common introns 3, 4, and 9, which are absent from all other PABP genes. Intron 5 is unique to *PAB7*. Introns 11, 13, 15, 17, and 19 in the C-terminal portion of the *PAB6* and *PAB7* ORFs are not conserved between these two genes, and their amino acid sequences differ significantly as well. The orphan *PAB1* gene lacks all but introns 1, 7 (which is unique to *PAB1*), and 8, and it also lacks the C-terminal domain.

The following minimum-evolution model can be proposed for the Arabidopsis PABP gene structures. The ancestral PABP gene (Figure 3) contained introns 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 14–16, and 18. The hypothetical common progenitor of the class I and class II PABP genes was derived from this ancestral gene via a loss of introns 3, 4, 5, and 9. Subsequent loss of introns 2 and 12 marked the separation of the class I lineage. The *PAB6* and *PAB7* genes (class III) were derived from the ancestral gene independently from the above lineages, by a loss of introns 6 and 8 and a reshuffling of the distal portion of the gene. The reshufflings must have occurred independently in *PAB6* and *PAB7*, resulting in unique intron positioning (introns 11 and 13 in *PAB6* and introns 17 and 19 in *PAB7*) and a loss of the conserved C-terminal domain in *PAB6*, but not in *PAB7*. The loss of introns 2 and 5 also marked the separation of *PAB6* from *PAB7*. The *PAB1* gene arose from the ancestral gene independently of others, via a loss of all but introns 1 and 8, gain of intron 7, and an additional rearrangement that resulted in a loss of the conserved C-terminal domain. While assuming an independent evolution of the *PAB1* gene increases the total number of events in the model, it agrees best with the deep branching order obtained for *PAB1* in PAUP analysis (Figure 2).

The primordial status of introns 1, 3, 5, and 6 is supported by their presence in the same location and phase in the human PABP gene (HORNSTEIN *et al.* 1999). Introns 2, 8, 10, 14–16, and 18 are also considered primordial because they occur in more than one PABP gene class and are all in phase zero. Of the remaining introns presumed to be ancestral, intron 12 is also in phase zero, while introns 4 and 9 are not. Assuming the presence of intron 12 in the ancestral gene allows the model to be constructed with fewer events. On the other hand, introns 4 and 9 could have equally likely resulted from recent intron gains in the class III lineage.

Gene models for the central regions of the five rice PABP genes were also reconstructed and compared with the Arabidopsis PABP gene models (poor sequence conservation beyond the four RRMs precluded full reconstructions). This analysis revealed that the structure of the central segment of the *OsPAB184* gene is identical to those of the class I Arabidopsis PABP genes; the structures of the *OsPAB179*, *OsPAB104*, and *OsPAB84.96* rice PABP genes are identical to those of the class II Arabidopsis PABP genes; and the structure of *OsPAB718.97* is identical to that of the Arabidopsis PAB7, a member of class III. These findings further support the notion of the three ancient classes of the PABP genes in flowering plants.

**Expression of the plant PABP genes:** Expression of the Arabidopsis PABP genes was analyzed by examining the EST databases, as well as experimentally in cases where no evidence for expression existed. Numerous ESTs were found for *PAB2*, *PAB4*, and *PAB8*, suggesting that these genes are expressed highly and in a broad range of cell types (Table 1; frequency distributions may not proportionally reflect relative expression in tissues, since EST data were compiled from more than one cDNA library). In contrast, expression of *PAB5* and *PAB3* is restricted to reproductive tissues. *PAB5* is expressed in tapetum, pollen, ovules, and developing seeds (BELOSTOTSKY and MEAGHER 1996), whereas *PAB3* is restricted to tapetum and pollen (Figure 4, C–G). *PAB1* is expressed in roots, most likely at a very low level and/or in restricted cell types since the transcript could be detected only by RT-PCR, and expression of *PAB1* is even weaker in flowers (BELOSTOTSKY and MEAGHER 1993). No ESTs were found for *PAB6* and *PAB7*. However, both transcripts were detectable by RT-PCR, although not by Northern blotting, which probably reflects their low expression and/or restricted ex-

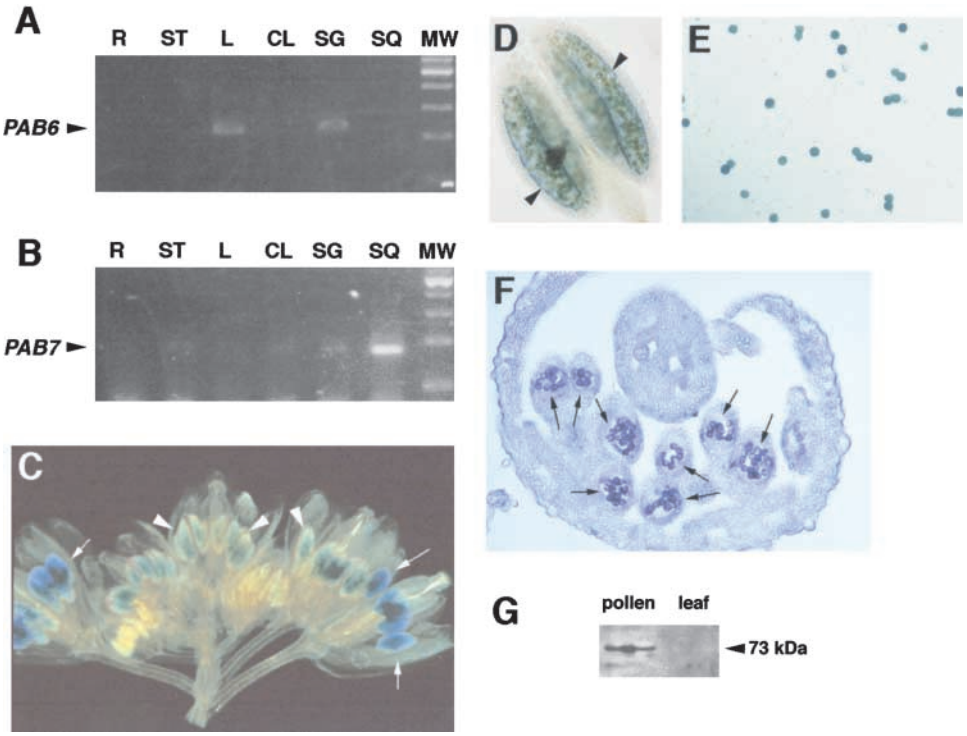


Figure 4.—Expression of the Arabidopsis PABP genes. (A and B) Results of the RT-PCR analyses of the expression of *PABP6* (A) and *PAB7* (B) in roots, stems (ST), leaves (L), cauline leaves (CL), whole 10-day-old seedlings (SG), and siliques (SQ). MW, molecular weight marker (100-bp ladder). (C–G) Analysis of the *PAB3* gene expression pattern using (C–E) the *PAB3* upstream control region translational fusion to β-glucuronidase in transgenic Arabidopsis (tapetum expression is highlighted in D; E shows mature pollen that was first isolated from anther and then stained); (F) *in situ* hybridization to the transverse section through the wild-type flower; and (G) Western blotting of the pollen and leaf total extracts with antibody specific for PAB3.

pression domains. *PAB6* mRNA was detected in leaves and young seedlings. *PAB7* was found predominantly in siliques, although trace amounts of RT-PCR signal could be seen in all tissues tested except roots (Figure 4, A and B).

Thus, all eight Arabidopsis PABP genes are transcriptionally active and can be grouped into three classes on the basis of similarity in their expression patterns. The broadly and highly expressed class is composed of *PAB2*, *PAB4*, and *PAB8*. The reproductive tissue-specific class is represented by *PAB3* and *PAB5*. The third class, whose expression is weak and/or restricted to a small subset of cell types, includes the *PAB6* and *PAB7* genes. *PAB1* appears to be an orphan gene, whose expression is weak and spatially restricted and may include reproductive tissues. Remarkably, this expression-based classification is in complete agreement with the ones derived from the analyses of Arabidopsis PABP amino acid sequences (Figure 2A) and their gene structures (Figure 3). Moreover, BLAST searches for the rice PABP ESTs produced no matches for *OsPAB718.97*, multiple matches for *OsPAB179*, *OsPAB104*, and *OsPAB84.96* that were derived from a broad range of tissues, and a single match for *OsPAB184*, derived from the endosperm cDNA. This distribution is fully consistent with the expression patterns found for the Arabidopsis PABP gene classes III, II, and I, respectively.

**Possible autoregulation and cross-regulation of plant PABP genes:** A notable feature of fungal and metazoan PABP genes is the presence of the A-rich segments in their 5′-untranslated regions that could serve as PABP-binding sites (Sachs *et al.* 1986; Nietfeld *et al.* 1990; Burd *et al.* 1991; de Melo Neto *et al.* 1995). PABP can bind to oligo(A) stretches as short as 12 residues (Sachs *et al.* 1987), as well as to nonhomopolymeric A-rich sequences (Gorlach *et al.* 1994). The interaction of PABP with its own 5′-UTR downregulates its own translation *in vivo* (Bag 2001). Comparably A-rich stretches are present in the 5′-UTRs of several Arabidopsis PABP genes. For instance, sequences $A_6GGA_{11}$, $A_{11}GGA_6$, and $A_7GTCA_6GTTTCGA_4TCCA_7$ occur in the 5′-UTRs of *PAB3*, *PAB5*, and *PAB2*, respectively. Thus, a similar autoregulatory control may operate in plant cells. Furthermore, since any given 5′-UTR can be bound by more than one PABP and any given PABP could bind to the 5′-UTR of more than one PABP mRNA, yet another level of complexity might exist in those cell types in which several PABPs are expressed simultaneously. For example, *PAB2* and *PAB5* are coexpressed in pollen (Belostotsky and Meagher 1996; Palanivelu *et al.* 2000b). In addition, since *PAB3* is closely related to *PAB5* and its expression in flowers was detected by RT-PCR (Belostotsky and Meagher 1993), a possibility that it is also expressed in pollen was also examined. Results of the analyses employing promoter-reporter fusions in transgenic plants, *in situ* hybridization, as well as immunoblotting of extracts from different Arabidopsis organs showed that the *PAB3* gene is expressed in pollen and tapetum, but not in any other tissue or cell type (Figure 4, C–G).

A purified recombinant PAB3 was then tested for its ability to interact with the 5′-UTRs of *PAB5*, *PAB2*, and its own, using gel mobility shift assays (Figure 5). Recombinant PAB3 interacted with its own 5′-UTRs and with

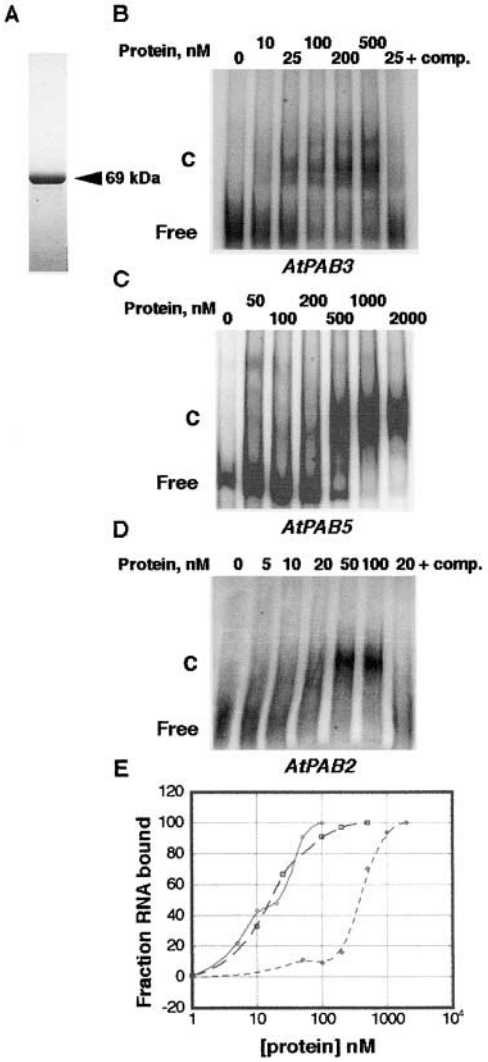


FIGURE 5.—Arabidopsis PAB3 protein interacts with the 5′-UTRs of *PAB2*, *PAB3*, and *PAB5* mRNAs *in vitro*. (A) An SDS PAGE gel was loaded with 2 μg of the purified recombinant His-tagged PAB3. (B–D) Gel mobility shift assays were carried out using the amounts of PAB3 protein indicated at the top, with *PAB3*, *PAB5*, and *PAB2* RNA probes. Oligo(A) was included as a competitor (150-fold excess) in the rightmost lanes of B and D. Free probe and shifted complex are indicated on the left. (E) Quantitation of the binding data: *PAB2* probe, circled symbols and solid line; *PAB3* probe, squared symbols and dashed line; *PAB5* probe, diamond-shaped symbols and dashed line. All data points are averages of two independent experiments that varied by <20%.

the 5′-UTR of *PAB2* with high and comparable affinity ($K_d \sim 20$ nM) and with lower affinity ($K_d \sim 200$ nM) with the 5′-UTR of *PAB5*. The lack of perfect correlation between the length of uninterrupted oligo(A) stretch and the apparent binding affinity may suggest that non-A residues of the respective 5′-UTRs make significant contributions to binding. These interactions were specific, since they were observed in the presence of 2 μg/ml tRNA as a nonspecific competitor, but were abolished by an excess of unlabeled oligo(A). Moreover, PABP affinity for nonspecific RNA is known to be considerably lower [$K_d \geq 0.5$ μM for mammalian PABP (GORLACH *et al.* 1994); $K_d = 1.5$ μM for yeast PABP (DEARDORFF and SACHS 1997)]. No binding to any of these probes was observed with up to 2 μM of nonspecific protein (glutathione S-transferase, not shown). These results suggest the possibility of negative feedback regulation, as well as cross-regulation, among the multiple plant PABPs that are coexpressed in the same cell.

## DISCUSSION

In this article, evidence is provided that genomes of monocot and dicot plants contain at least three ancient lineages of PABP genes. These findings suggest that orthologous PABP genes should exist in most monocots and dicots and possibly even in the clades of the flowering plants basal to the monocot/dicot split. Arabidopsis has eight PABP genes, all of which are expressed and potentially functional, and at least three of them (*PAB2*, *PAB3*, and *PAB5*) are able to complement the *pab1* null mutant of *S. cerevisiaie* (BELOSTOTSKY and MEAGHER 1996; PALANIVELU *et al.* 2000a; CHEKANOVA *et al.* 2001). This represents the largest known PABP multigene family in any species so far, which is particularly striking considering the relatively small size of the Arabidopsis genome. Phylogenetic analysis of the Arabidopsis PABP amino acid sequences, comparative analysis of their gene structures, and surveys of the available gene expression data demonstrate that class I is composed of the reproductive tissue-specific PABPs, *PAB3* and *PAB5* genes, class II of the broadly and strongly expressed *PAB2*, *PAB4*, and *PAB8* genes, and class III of the tissue-specific, weakly expressed *PAB6* and *PAB7* genes. In rice, the *OsPAB184* gene has been identified as a member of class I; *OsPAB179*, *OsPAB104* and *OsPAB84.96* as members of class II; and *OsPAB718.97* as a member of class III. While detailed expression data for the rice PABP genes are not yet available, the tissue distribution and abundance of the reported ESTs are fully consistent with the prediction that the expression of class I, II, and III PABP genes in rice follows the same pattern as it does in Arabidopsis. Experiments showing that the developmental regulation of the *PAB2* promoter in transgenic tobacco is virtually identical to the one seen in Arabidopsis also support the view that the expression patterns of the plant PABP genes are evolutionarily conserved (PALANIVELU *et al.* 2000b). Placement of the Arabidopsis *PAB1* gene is less certain, principal possibilities being that it either belongs to class I or is a sole member of a separate class that is present in Arabidopsis but absent from rice. Alternatively, *PAB1* might be a gene that is no longer under selective pressure and is on its way to becoming a pseudogene. Sequencing of other dicot genomes should help resolve this issue.

The potential for interactions of the plant PABPs with their own 5′-UTR and the 5′-UTRs of the other PABP genes expressed in the same cell type represents a level of complexity not seen in other eukaryotes to date. Results of binding studies suggest that the PAB3 protein may regulate the function of its own mRNA, as well as of mRNAs of the other PABP genes expressed in pollen. Curiously, PAB3 binds with lower affinity to the 5′-UTR of the *PAB5* transcript than to the 5′-UTR of *PAB2* and to its own 5′-UTR. This could serve to limit the expression of the class II PAB2 protein in pollen, thus allowing the reproductive-specific PABP (PAB5) to predominate. In addition, while the binding to the *PAB2* and *PAB3* probes fit well to a standard equation describing noncooperative single-site interaction, the binding curve for the *PAB5* probe was steeper; *i.e.*, it deviated toward positive cooperativity (Figure 5). The possible significance of this remains to be investigated.

The outcomes of these multiple interactions could be complex and dependent on the *in vivo* concentrations of the respective PABPs, their $K_d$'s for the various 5′-UTRs, concentrations and secondary structures of the 5′-UTRs themselves, and competing RNA-binding proteins. It should also be noted that PABPs might bind to the 5′-UTRs of other transcripts containing similarly A-rich elements, such as the pectinesterase gene (At2g47050), whose 5′-UTR contains the sequence $A_6CA_4CCA_{19}GACA_9$, where the last A is the first base of the initiator codon.

The high degree of amino acid sequence divergence and vastly different expression patterns of the class I, II, and III PABPs argue for the existence of functional differences between the classes and possibly even between members of the same class. For instance, microarray-based mRNA decay experiments (Gutierrez *et al.* 2002; http://afgc.stanford.edu/afgc_html/expression.html) have revealed that, while the *PAB8* mRNA is stable, the two other members of class II, *PAB2* and *PAB4*, are encoded by unstable mRNAs ($T_{1/2} < 2$ hr). This may indicate that the plant needs to adjust levels of some, but not all, class II PABPs relatively rapidly, depending on changes in the environment. However, the potential for functional redundancy must also be taken into account. Selection for the enhanced quantity of the encoded gene product, for higher fidelity of the process, for the emergent function(s), or for the divergent function(s) have all been considered as possible reasons for maintaining redundant genes (Thomas 1993). Divergence in gene function could also be driven by acquisition of novelties in *cis*-regulatory regions controlling the expression patterns of otherwise interchangeable ORFs (reviewed in Pickett and Meeks-Wagner 1995). Knockout mutants in all eight Arabidopsis PABP genes have been recently isolated in this laboratory, enabling rigorous experimental testing of their functions and possible mechanisms of redundancy and/or functional specialization.

I gratefully acknowledge J. Chekanova for stimulating discussions, C.-B. Stewart for sharing her expertise in phylogenetic analysis, J. Chekanova, R. Shaw, and R. Lartey for help with some of the experiments, and C.-B. Stewart and R. Meagher for critical reading of the manuscript. I also thank H. Tedeschi for his support and R. Meagher for stimulating my interest in evolution of multigene families. This project was supported by USDA NRICGP and the Basic Biosciences Minigrant Program.

## LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

An, Y. Q., S. Huang, J. M. McDowell, E. C. McKinney and R. B. Meagher, 1996 Conserved expression of the Arabidopsis ACT1 and ACT 3 actin subclass in organ primordia and mature pollen. Plant Cell **8:** 15–30.

Bag, J., 2001 Feedback inhibition of poly(A)-binding protein mRNA translation. A possible mechanism of translation arrest by stalled 40 S ribosomal subunits. J. Biol. Chem. **276:** 47352–47360.

Belostotsky, D., and R. B. Meagher, 1993 Differential organ-specific expression of three poly(A) binding protein genes from *Arabidopsis*. Proc. Natl. Acad. Sci. USA **90:** 6686–6690.

Belostotsky, D. A., and R. B. Meagher, 1996 Pollen-, ovule-, and early embryo-specific poly(A) binding protein from *Arabidopsis* complements essential functions in yeast. Plant Cell **8:** 1261–1275.

Bernstein, P., S. W. Peltz and J. Ross, 1989 The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability *in vitro*. Mol. Cell. Biol. **9:** 659–670.

Brown, J. T., and A. W. Johnson, 2001 A cis-acting element known to block 3′ mRNA degradation enhances expression of polyA-minus mRNA in wild-type yeast cells and phenocopies a ski mutant. RNA **7:** 1566–1577.

Burd, C. G., E. L. Matunis and G. Dreyfuss, 1991 The multiple RNA binding domains of the mRNA poly(A)-binding protein have different RNA-binding activities. Mol. Cell. Biol. **11:** 3419–3424.

Caponigro, G., and R. Parker, 1995 Multiple functions for the poly(A) binding protein in mRNA decapping and deadenylation in yeast. Genes Dev. **9:** 2421–2432.

Caponigro, G., and R. Parker, 1996 Mechanisms and control of mRNA turnover in Saccharomyces cerevisiae. Microbiol. Rev. **60:** 233–249.

Chekanova, J. A., R. J. Shaw, M. A. Wills and D. A. Belostotsky, 2000 Poly(A) tail-dependent exonuclease AtRrp41p from Arabidopsis thaliana rescues 5.8 S rRNA processing and mRNA decay defects of the yeast ski6 mutant and is found in an exosome-sized complex in plant and yeast cells. J. Biol. Chem. **275:** 33158–33166.

Chekanova, J. A., R. J. Shaw and D. A. Belostotsky, 2001 Analysis of an essential requirement for the poly(A) binding protein function using cross-species complementation. Curr. Biol. **11:** 1207–1214.

Clough, S. J., and A. F. Bent, 1998 Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. **16:** 735–743.

Coller, J. M., N. K. Gray and M. P. Wickens, 1998 mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. Genes Dev. **12:** 3226–3235.

Cosson, B., A. Couturier, S. Chabelskaya, D. Kiktev, S. Inge-Vechtomov *et al.*, 2001 Poly(A)-binding protein acts in translation termination via eukaryotic release factor 3 interaction and does not influence [PSI(+)] propagation. Mol. Cell. Biol. **22:** 3301–3315.

Deardorff, J. A., and A. B. Sachs, 1997 Differential effects of aromatic and charged residue substitutions in the RNA binding domains of the yeast poly(A)-binding protein. J. Mol. Biol. **269:** 67–81.

de Melo Neto, O. P., N. Standart and M. Martins de Sa, 1995 Autoregulation of poly(A)-binding protein synthesis *in vitro*. Nucleic Acids Res. **23:** 2198–2205.

Deo, R. C., J. B. Bonanno, N. Sonenberg and S. K. Burley, 1999

Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell **98:** 835–845.

de Souza, S. J., M. Long, R. J. Klein, S. Roy, S. Lin *et al.*, 1998 Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. Proc. Natl. Acad. Sci. USA **95:** 5094–5099.

Gallie, D. R., 1991 The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev. **5:** 2108–2116.

Gorlach, M., C. G. Burd and G. Dreyfuss, 1994 The mRNA poly (A)-binding protein: localization, abundance, and RNA-binding specificity. Exp. Cell Res. **211:** 400–407.

Gutierrez, R. A., R. M. Ewing, J. M. Cherry and P. J. Green, 2002 Identification of unstable transcripts in Arabidopsis by cDNA microarray analysis: rapid decay is associated with a group of touch- and specific clock-controlled genes. Proc. Natl. Acad. Sci. USA **99:** 11513–11518.

Hornstein, E., A. Git, I. Braunstein, D. Avni and O. Meyuhas, 1999 The expression of poly(A)-binding protein gene is translationally regulated in a growth-dependent fashion through a 5′-terminal oligopyrimidine tract motif. J. Biol. Chem. **274:** 1708–1714.

Hoshino, S., M. Imai, T. Kobayashi, N. Uchida and T. Katada, 1999 The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3′-poly (A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. J. Biol. Chem. **274:** 16677–16680.

Imataka, H., A. Gradi and N. Sonenberg, 1998 A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. EMBO J. **17:** 7480–7489.

Jefferson, R. A., T. A. Kavanagh and M. W. Bevan, 1987 GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6:** 3901–3907.

Kessler, S. H., and A. B. Sachs, 1998 RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. Mol. Cell. Biol. **18:** 51–57.

Khaleghpour, K., A. Kahvejian, G. De Crescenzo, G. Roy, Y. V. Svitkin *et al.*, 2001 Dual interactions of the translational repressor Paip2 with poly(A) binding protein. Mol. Cell. Biol. **21:** 5200–5213.

Kuhn, U., and T. Pieler, 1996 *Xenopus* poly(A) binding protein: functional domains in RNA binding and protein-protein interaction. J. Mol. Biol. **256:** 20–30.

Mangus, D. A., N. Amrani and A. Jacobson, 1998 Pbp1p, a factor interacting with Saccharomyces cerevisiae poly(A)-binding protein, regulates polyadenylation. Mol. Cell. Biol. **18:** 7383–7396.

Marhoul, J. F., and T. H. Adams, 1996 *Aspergillus* fabM encodes an essential product that is related to poly(A)-binding proteins and activates development when overexpressed. Genetics **144:** 1463–1470.

Nietfeld, W., H. Mentzel and T. Pieler, 1990 The *Xenopus laevis* poly(A) binding protein is composed of multiple functionally independent RNA binding domains. EMBO J. **9:** 3699–3705.

Palanivelu, R., D. A. Belostotsky and R. B. Meagher, 2000a Arabidopsis thaliana poly(A) binding protein 2 (PAB2) functions in yeast translational and mRNA decay processes. Plant J. **22:** 187–198.

Palanivelu, R., D. A. Belostotsky and R. B. Meagher, 2000b Conserved expression of Arabidopsis thaliana poly(A) binding protein 2 (PAB2), in distinct vegetative and reproductive tissues. Plant J. **22:** 199–210.

Pickett, F. B., and D. R. Meeks-Wagner, 1995 Seeing double: appreciating genetic redundancy. Plant Cell **7:** 1347–1356.

Proweller, A., and J. S. Butler, 1996 Ribosomal association of poly(A)-binding protein in poly(A)-deficient *Saccharomyces cerevisiae.* J. Biol. Chem. **271:** 10859–10865.

Roy, G., G. De Crescenzo, K. Khaleghpour, A. Kahvejian, M. O'Connor-McCourt *et al.*, 2002 Paip1 interacts with poly(A) binding protein through two independent binding motifs. Mol. Cell. Biol. **22:** 3769–3782.

Sachs, A. B., M. W. Bond and R. D. Kornberg, 1986 A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. Cell **45:** 827–835.

Sachs, A. B., R. W. Davis and R. D. Kornberg, 1987 A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. Mol. Cell. Biol. **7:** 3268–3276.

Sigrist, S. J., P. R. Thiel, D. F. Reiff, P. E. Lachance, P. Lasko *et al.*, 2000 Postsynaptic translation affects the efficacy and morphology of neuromuscular junctions. Nature **405:** 1062–1065.

Swofford, D. L., 1993 *PAUP: Phylogenetic Analysis Using Parsimony.* Illinois Natural History Survey, Champaign, IL.

Tarun, S., and A. B. Sachs, 1995 A common function for mRNA 5′ and 3′ ends in translation initiation in yeast. Genes Dev. **9:** 2997–3007.

Tarun, S. Z., S. E. Wells, J. A. Deardorff and A. B. Sachs, 1997 Translation initiation factor eIF4G mediates *in vitro* poly(A) tail-dependent translation. Proc. Natl. Acad. Sci. USA **94:** 9046–9051.

Thomas, J. H., 1993 Thinking about genetic redundancy. Trends Genet. **9:** 380–384.

Wakiyama, M., H. Imataka and N. Sonenberg, 2000 Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. Curr. Biol. **10:** 1147–1150.

Wells, S. E., P. E. Hillner, R. D. Vale and A. B. Sachs, 1998 Circularization of mRNA by eukaryotic translation initiation factors. Mol. Cell **2:** 135–140.

Wilusz, C. J., M. Gao, C. L. Jones, J. Wilusz and S. W. Peltz, 2001 Poly(A)-binding proteins regulate both mRNA deadenylation and decapping in yeast cytoplasmic extracts. RNA **7:** 1416–1424.

Wolfe, K. H., M. Gouy, Y.-W. Yang, P. M. Sharp and W.-H. Li, 1989 Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86:** 6201–6205.

Yohn, C. B., A. Cohen, A. Danon and S. P. Mayfield, 1998 A poly(A) binding protein functions in the chloroplast as a message-specific translation factor. Proc. Natl. Acad. Sci. USA **95:** 2238–2243.

Communicating editor: M. S. Sachs