Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Regulation of Plant Gene Expression

R. B. Goldberg

Phil. Trans. R. Soc. Lond. B 1986 **314**, 343-353
doi: 10.1098/rstb.1986.0057

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1986 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

Phil. Trans. R. Soc. Lond. B **314**, 343–353 (1986)
Printed in Great Britain

Regulation of plant gene expression

By R. B. Goldberg

Department of Biology, University of California, Los Angeles, California 90024, *U.S.A.*

Differential gene expression is required to establish and maintain specific developmental states in higher plants. For example, an anther has at least 11000 diverse mRNAs that are absent from the polysomes of other organ systems, and the root has at least 7000 organ-specific mRNAs. Both transcriptional and post-transcriptional processes regulate the sequence composition and prevalence distribution of each developmental-specific mRNA set. Soybean seed protein genes represent an excellent example of a highly regulated gene set. These genes are temporally and spatially regulated during embryogenesis, and are either inactive or expressed at low levels in mature plant organ systems. Gene transfer experiments indicate that soybean seed protein genes retain their developmental-specific expression programme in transformed tobacco plants. In addition, large polygenic clusters can be transferred from soybean to tobacco, and the expression pattern of each gene within the cluster is maintained in the foreign cell environment. Although the DNA sequences and cellular factors required to control seed protein gene expression are not yet known, gene transfer studies and emerging DNA binding protein technology should facilitate their identification in the near future.

Introduction

Higher plants have many unique developmental processes. First, as shown in figure 1, plants have an alternation of spore-forming and gamete-forming generations that have distinct anlagen and are separated spatially and temporally during the life cycle. Unlike animals, higher plants do not have a germline that is set aside during early embryogeny. In contrast, the dominant, sporophytic generation undergoes a series of ontological events that lead from a single-celled zygote to a mature, spore-producing plant. The haploid spores divide mitotically and differentiate into a three-celled male gametophyte or a seven-celled female gametophyte. The male and female gametophytes reside in distinct sporophytic organ systems and contain the sperm and egg cells, respectively. Gametophytic cells are functionally distinct from one another and probably arise as a consequence of differential expression of the haploid, gametophyte genome. Second, higher plants undergo a continuous programme of differentiation and development. Unlike animals that complete their major morphogenetic events during embryogeny, plants contain meristematic cells capable of differentiating into the sporophytic organ systems throughout the life cycle. The root and shoot meristems are formed during embryogenesis and retain the ability to divide, differentiate, and establish the morphological pattern for the root, stem, leaf and flower of the post-embryonic plant. Third, higher plant morphogenesis occurs in the absence of cell movement and results from asymmetric divisions in different cell planes. The lack of cell movement is due to the presence of a semi-rigid wall that surrounds the plant cell and glues it to other cells in the immediate vicinity. Fourth, environmental factors play a major role in plant development and are able to trigger dormancy periods or periods of rapid development. Proper light, temperature and water conditions are

22-2

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences.* www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

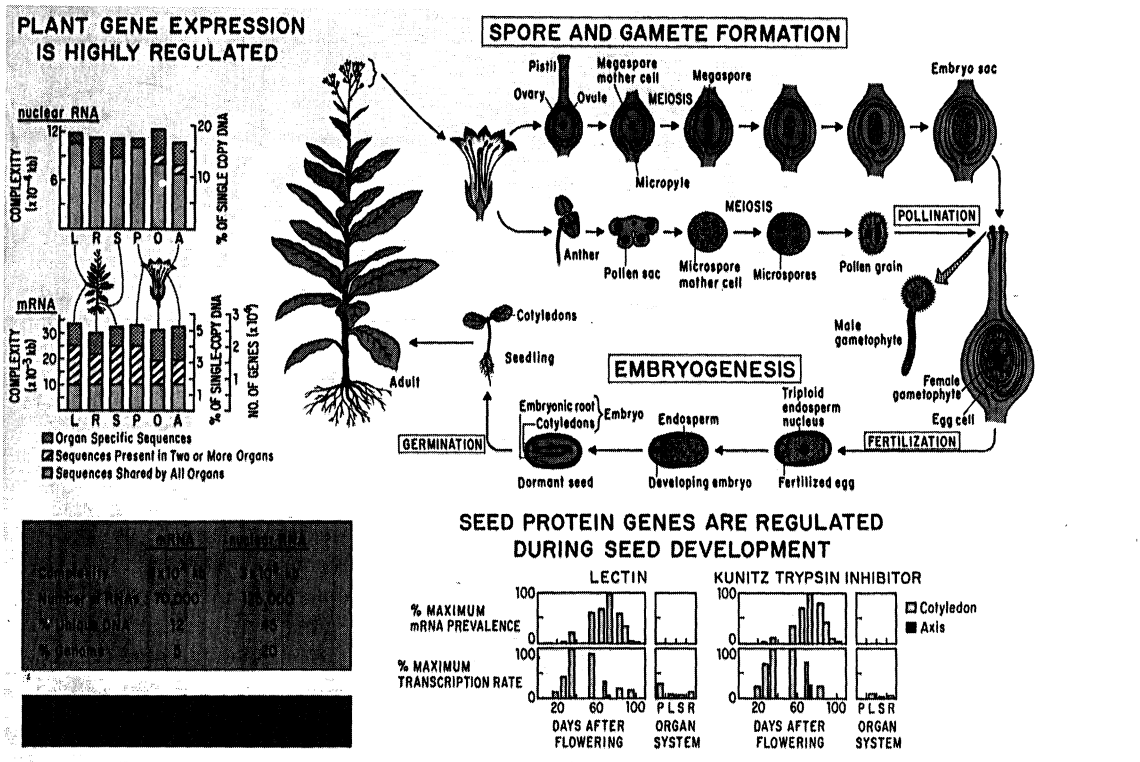


FIGURE 1. Gene expression during the plant life cycle. Tobacco single-copy DNA sequence representation in organ system nuclear and polysomal RNAs was taken from the data of Kamalay & Goldberg (1980, 1984). L, R, S, P, O, A refer to leaf, root, stem, petal, ovary and anther, respectively. Developmental accumulation of soybean seed protein mRNAs and relative rates of seed protein gene transcription were taken from the data of Walling *et al.* (1986). P, L, S, R refer to post-germination cotyledons, leaf, stem and root, respectively. The plant life cycle depicts events that occur in tobacco development. Mature soybean seeds do not contain endosperm tissue.

required for normal growth and development and are able to interact with cellular factors to control ontological events. Finally, many plant cells are totipotent and retain their ability to regenerate into fully differentiated, fertile plants. Totipotency manifests itself through somatic embryogensis, or morphogenesis from meristematic foci, and raises the question of whether irreversible determination plays a major role in plant development.

The underlying cellular and molecular processes that control plant development are not yet understood. For example, studies with the electron microscope indicate that the egg cell is polar and distributes its cytoplasm asymmetrically to embryonic cells upon first zygotic division (Raghavan 1976, 1986). The large basal cell and small terminal cell of the two-celled zygote have different fates during embryogenesis and ultimately give rise to different organs of the mature plant (Raghavan 1976, 1986). It is not yet known, however, whether the egg cell contains morphogenetic determinants that are formed during female gametophyte development and influence early embryological events, nor is it known the extent to which maternal factors and embryonic inductive processes programme ontological events during early embryogenesis. Thus the roles that the maternal and zygotic genomes play in plant embryogeny, and the molecular events that occur during gamete formation and early embryogenesis, are not yet understood. Similarly, the processes responsible for pattern formation and differentiation of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

meristematic cells into the mature sporophytic organ systems are unknown. Nor is it known what cellular processes are responsible for enabling a somatic cell to differentiate into a mature plant. Although I have listed only a few examples, it is clear that new ideas and approaches must be used before we understand in a mechanistic way the nature of plant development.

Gene expression in tobacco organ systems

Most of what is known about plant gene expression comes from RNA-excess DNA–RNA hybridization experiments with sporophytic organ system and embryonic RNAs (Goldberg *et al.* 1978; Kamalay & Goldberg 1980, 1984; Goldberg *et al.* 1981*b*; Galau & Dure 1981). Recently, however, hybridization studies have been done with mature male gametophyte (pollen grain) mRNA (Willing & Mascarenhas 1984). As shown in figure 1, there are approximately 25000 diverse genes expressed in each tobacco organ system or 5% of the single-copy DNA. A similar number of genes are expressed at several stages of soybean embryogenesis (Goldberg *et al.* 1981*b*) and cotton embryogenesis (Galau & Dure 1981) as well as in the *Tradescantia* pollen grain (Willing & Mascarenhas 1984). These findings indicate that a large number of genes are expressed at different periods of the plant life cycle and in both the sporophytic and gametophytic generations.

Figure 1 shows that gene expression is highly regulated in tobacco organ systems. Each organ system contains a developmental-specific mRNA set that is undetectable in the *polysomes* of heterologous organ systems (Kamalay & Goldberg 1980). Most organ-specific mRNAs react kinetically as rare class sequences that are present in an average only 10 molecules per cell. For example, there are approximately 10000 diverse ovary-specific mRNAs and 6000 stem-specific mRNAs. These messages constitute about 40% and 25% of all diverse ovary and stem mRNAs, respectively (figure 1). In addition to the developmental-specific mRNA set, each organ system has a mRNA set that is shared by all organ systems (figure 1). This set contains approximately 8000 diverse messages and probably encodes proteins required by all cells. As shown in figure 1, at least 70000 diverse genes are expressed collectively in the mature sporophytic plant. This is probably a minimum estimate but indicates that a large number of genes is required to programme and maintain the sporophytic phase of the plant life cycle.

Figure 2 compares the representation of individual organ-specific mRNAs in different tobacco organ systems. These messages were identified by differential screening of anther, ovary, petal, stem and root copy DNA (cDNA) libraries with a variety of cDNA probes (K. Cox, G. N. Drews, J. Truettner, T. L. Sims & R. B. Goldberg, unpublished results). As shown in figure 2, mRNAs can be identified that are present exclusively, or at very high levels, in each organ system. These messages are relatively abundant and constitute 0.1 to 1% of the polysomal poly(A) mRNA mass, or approximately 500 to 5000 molecules per cell, depending upon the sequence (Goldberg *et al.* 1978). Thus results from population hybridization experiments and studies with cDNA clones indicate that there are both prevalent and rare mRNAs that are specific to each organ system.

Recently the spatial expression patterns of several tobacco organ-specific genes were investigated by hybridization *in situ* (K. Cox, G. N. Drews & R. B. Goldberg, unpublished results). These experiments showed that organ-specific mRNAs can be restricted to specific regions, tissues and/or cell-types of a given organ system. For example, one root-specific mRNA (TR-2) was present preferentially in elongation region ground meristem cells. In addition,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

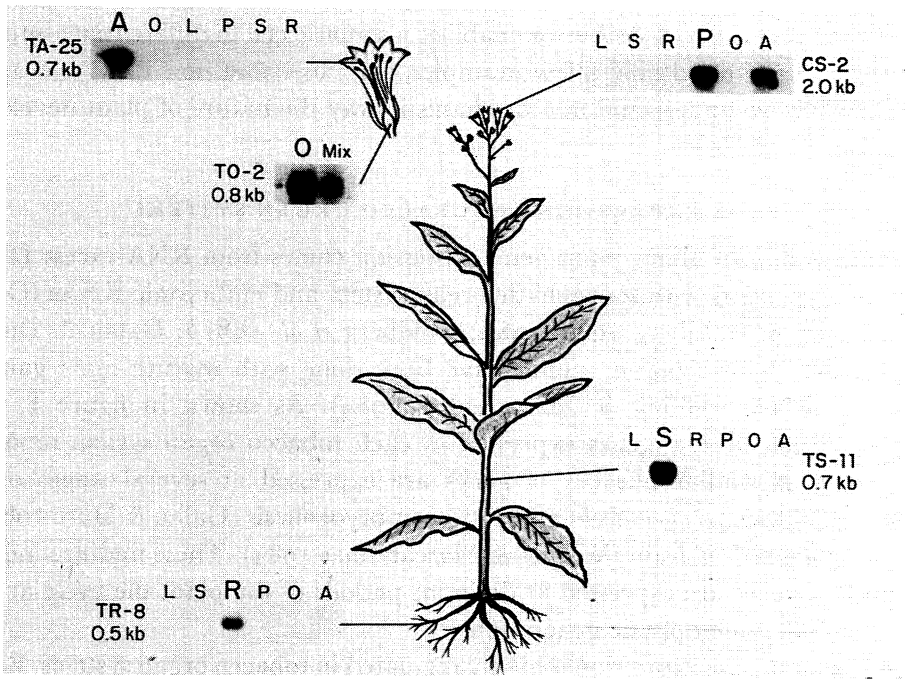


FIGURE 2. Tobacco organ-specific gene expression. Tobacco anther (A), ovary (O), petal (P), stem (S) and root (R) cDNA libraries were constructed according to the procedure of Gubler & Hoffman (1983). The developmental characteristics of tobacco organ systems used to construct the libraries have been presented elsewhere (Kamalay & Goldberg 1980). Differential colony hybridization (St John & Davis 1979) was used to identify cDNA clones that represented mRNAs present exclusively, or at maximum prevalences, in each organ system. Organ-specific mRNA prevalences ranged from 0.1 to 1.0% of the relevant polysomal poly(A) mRNA. Representative cDNA clones were hybridized with gel blots containing anther, ovary, leaf (L), petal, stem and root mRNAs. The cDNA clone designation and corresponding mRNA size is listed next to each mRNA gel blot. CS-2 refers to a chalcone synthetase genomic clone (G. N. Drews & R. B. Goldberg, unpublished results). Mix refers to an equimolar mixture of leaf, stem, root, petal and anther mRNAs.

chalcone synthetase mRNA (CS-2, figure 2) was represented at high levels in upper epidermal cells of the petal pigment zone. These results indicate that organ-specific mRNA sets probably represent a complex collage of genes expressed in various cell and tissue types, and that an individual gene can be expressed at a high level in one cell type and be inactive in another cell type of the same organ system. This situation can result in an *apparent* low expression level for an individual gene when averaged over the entire organ system. Thus many rare class, organ-specific mRNAs identified in RNA-excess DNA–RNA hybridization experiments could be abundant messages in specific cell types.

Both transcriptional and post-transcriptional processes contribute to the production of each organ-specific mRNA set (Kamalay & Goldberg 1980, 1984). Figure 1 shows that each tobacco organ system has a nuclear RNA complexity of approximately 1.1×10^5 kilobase pairs (kb), or about 18% of the single-copy DNA. The nuclear RNA sequence complexity is 3.5 times greater than that found in polysomal RNA indicating that only 30% of diverse nuclear transcripts are exported to the cytoplasm. The aggregate organ system nuclear RNA complexity is approximately 3×10^5 kb, indicating that at least 45% of the single-copy DNA is transcribed and represented collectively in vegetative and floral nuclei during the sporophytic phase of the life cycle (figure 1).

As shown in figure 1, each organ system has a developmental-specific nuclear RNA set. For

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

example, 35% of diverse ovary nuclear RNA sequences (4×10^4 kb) are not detectable in the nuclei of heterologous organ systems (Kamalay & Goldberg 1984). Within detection limits, the ovary-specific nuclear RNA set contains all ovary-specific mRNAs, implying that ovary-specific genes are regulated in part at the transcriptional level. In contrast, 20% of the stem nuclear complexity (2.4×10^4 kb) is not detectable in heterologous nuclear RNAs. However, stem-specific mRNAs are *not* represented in the stem-specific nuclear set. Rather, they are represented in the nuclear RNAs of other organ systems even though they are undetectable at the polysomal level (Kamalay & Goldberg 1980, 1984). These results indicate that post-transcriptional processes are also important in regulating tobacco organ-specific gene expression.

Contrasting soybean seed protein gene expression patterns

A unique aspect of plant development is the formation of the dormant seed. Figure 1 shows that seed development is a consequence of fertilization and marks the beginning of the sporophytic phase of the life cycle. The seed contains the fully differentiated embryo that is programmed to continue growth and morphogenesis in the post-germination sporophyte. As pointed out above, approximately 20000 genes are expressed at any given stage of seed development (Goldberg *et al.* 1981*b*; Galau & Dure 1981). Seed protein genes, however, represent the most striking example of a highly regulated embryonic gene set. In the soybean plant this set includes several unique gene families that encode highly prevalent polypeptides stored in the dormant seed (Goldberg *et al.* 1981*a*, *b*, 1983; Walling *et al.* 1986). These families include genes that encode storage proteins, trypsin inhibitors and lectins. Tobacco, as well as all other plants, contains seed protein genes (Higgins 1984). In fact my laboratory recently identified tobacco mRNAs related to soybean glycinin and β-conglycinin storage protein gene families (S. Barker & R. B. Goldberg, unpublished results).

Figure 1 shows that lectin and Kunitz trypsin inhibitor genes are highly regulated during seed development (Goldberg *et al.* 1981*a*, 1983). Lectin and Kunitz trypsin inhibitor mRNAs accumulate and decay during specific embryonic periods. This temporal pattern of gene expression is similar for all seed protein genes, although the specific timing may be different for each gene class (Walling *et al.* 1986). In addition, Kunitz trypsin inhibitor mRNA and other seed protein messages are more prevalent in the embryonic cotyledons than in the axis (Goldberg *et al.* 1981*a*; Walling *et al.* 1986). Thus there is both temporal and spatial regulation of seed protein gene expression. The relative rates of transcription of lectin and Kunitz trypsin inhibitor genes correlate well with corresponding mRNA fluctations during development, indicating that seed protein genes are regulated in part at the transcriptional level (figure 1) (Walling *et al.* 1986). However, the relative transcription rates of seed protein and non-seed protein genes encoding mRNAs differing by 10000-fold in prevalence are similar, suggesting that post-transcriptional events also play a role in regulating seed protein gene expression (Walling *et al.* 1986).

Table 1 shows that there is no detectable storage protein mRNAs in mature plant organ systems at a detection limit of approximately 0.01 molecules per cell (Goldberg *et al.* 1981*a*; J. J. Harada, R. L. Fischer & R. B. Goldberg, unpublished results). The absence of storage protein mRNAs correlates with highly repressed gene transcription rates (figure 1) (Walling *et al.* 1986). In contrast, lectin mRNA is detected at a low level in the root (Okamuro *et al.*

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

1986) and Kunitz trypsin inhibitor sequences are represented as rare class messages in leaf, stem, and root (K. D. Jofuku and R. B. Goldberg, unpublished results). These findings indicate that all seed protein genes are expressed at high levels during embryogenesis but that they exhibit differential expression patterns in the mature sporophytic plant.

I have proposed a tentative model to explain the contrasting seed protein gene expression patterns. As shown in figure 3, each seed protein gene has a *cis*-acting control element (E) responsible for regulating transcription during embryogenesis. The embryo control element interacts with an embryo-specific, *trans*-acting nuclear factor (E_{T1}) to activate transcription at appropriate developmental periods. In addition, seed protein genes can contain other *cis*-acting control elements (L,S,R) that programme gene expression in the mature plant. Each element interacts with an organ-specific, *trans*-acting factor to trigger transcription in the relevant organ-system (e.g. factor R_T interacts with the root control element). Thus each seed protein gene contains a hierarchy of *cis*-control sequences that allows it to respond to unique

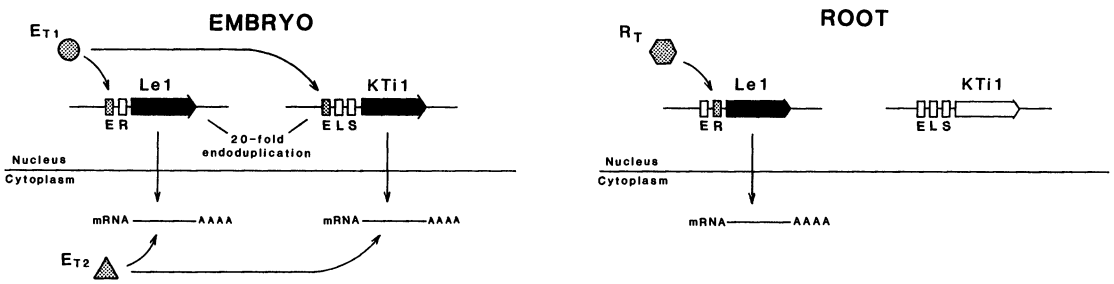


FIGURE 3. Regulation of soybean seed protein gene expression. Model is based on the mRNA prevalence data presented in table 1, and the mRNA accumulation and gene transcription experiments of Walling *et al.* (1986). Le1 and KTi1 refer to specific lectin and Kunitz trypsin inhibitor gene family members, respectively (Goldberg *et al.* 1981*a*; 1983; Walling *et al.* 1986; K. D. Jofuku & R. B. Goldberg, unpublished results). E, R, L, S refer to embryo, root, leaf and stem *cis*-control elements, respectively. E_{T1} and R_T represent embryo and root *trans*-acting factors that interact with E and R *cis*-control elements to activate gene transcription. E_{T2} represents a cytoplasmic factor that interacts with seed protein mRNAs to increase their stabilities.

TABLE 1. SEED PROTEIN mRNA REPRESENTATION (PERCENTAGES) DURING THE SOYBEAN LIFE CYCLE[a]

		mature plant		
mRNA	embryogenesis[b]	leaf	stem	root
glycinin	10[c]	ND	ND	ND
β-conglycinin	11	ND	ND	ND
lectin	0.75	ND	ND	4×10^{-5}
Kunitz trypsin inhibitor[d]	3	5×10^{-4}	10^{-4}	10^{-3}

[a] Taken from the data of Goldberg *et al.* (1981*a*), Walling *et al.* (1986), Okamuro *et al.* (1986), and unpublished results from my laboratory. Prevalences represent minimum estimates. ND indicates not detectable at a sensitivity of 0.01 molecules per cell.

[b] Mid-maturation stage embryos, or about 70 days after flowering (Walling *et al.* 1986).

[c] Percentage of polysomal poly(A) mRNA; 10% of embryo mRNA represents approximately 25000 molecules per cell (Goldberg *et al.* 1981*a*) whereas 4×10^{-5}% of root mRNA represents about 0.2 molecules per cell (Okamuro *et al.* 1986).

[d] Represents expression of several Kunitz trypsin inhibitor family members. Expression of individual Kunitz trypsin inhibitor genes can be restricted to specific mature plant organ systems. For example, the *Kti*1 and *Kti*2 genes encoding the A-37 Kunitz trypsin inhibitor mRNA (Goldberg *et al.* 1981*a*; Walling *et al.* 1986) are not expressed in the root but are expressed during seed development and in the leaf and stem (K. D. Jofuku & R. B. Goldberg, unpublished results).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

developmental situations. For example, the model predicts that storage protein genes have only the embryo element whereas the lectin gene has embryo and root elements (table 1). To account for post-transcriptional processes, I have proposed an additional *trans*-acting factor (E_{T_2}) that is present in the embryo cytoplasm and stablizes seed protein mRNAs during specific embryonic phases. Alternatively, factor E_{T_2} could be nuclear and serve to facilitate the selective export of transcripts to the cytoplasm. In either case, factor E_{T_2} acts on gene transcripts rather than on the gene and/or its flanking regions. Although speculative, this model accounts for all available gene transcription and mRNA accumulation data (Walling *et al.* 1986) and leads to testable predictions.

Is there a candidate for an embryo-specific *cis*-control element? Figure 4 presents a schematic representation of the soybean lectin gene (Vodkin *et al.* 1983). This gene does not contain introns and possesses the relevant eukaryotic consensus sequences implicated in RNA polymerase binding and poly(A) addition. A comparison of the DNA sequences in the 5′ flanking regions

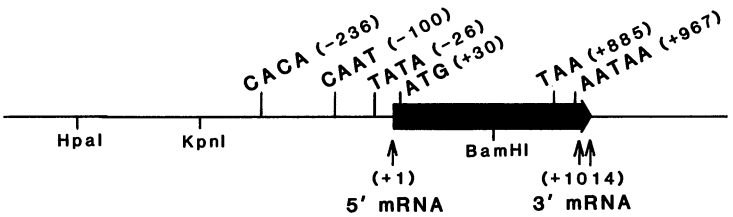


FIGURE 4. Schematic representation of the soybean lectin gene. Taken from the data of Vodkin *et al.* (1983). Nucleotides +1 and +1014 designate the 5′ and 3′ gene ends, respectively. The sequence CACA is part of a *cis* nucleotide consensus sequence present in the 5′ flanking region of all soybean seed protein genes sequenced so far (see figure 5).

GENE		SEQUENCE	POSITION
Lectin	Le 1	T [ ] C T	−236
Glycinin	G2	A [ ] C T	−194
	G2	T [ ] C A	−124
	G1	C [ ] A T	−420
Kunitz Trypsin Inhibitor	KTi1	A [ ] C T	−691
	KTi1	T [ ] A A	−502
	KTi2	A [ ] A C	−339
	KTi2	A [ ] A A	−222

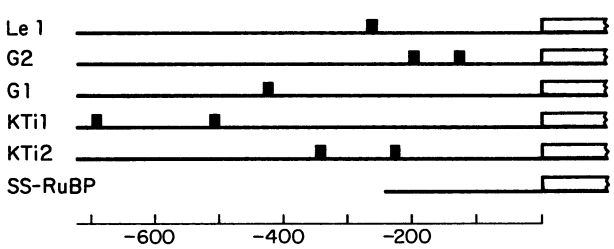


FIGURE 5. A consensus sequence in soybean seed protein gene 5′ regions. The VAX computer was used to search for sequence homologies in the upstream regions of several soybean seed protein genes (Staden 1979). Le1, G1 and G2 sequences were taken from the data of Vodkin *et al.* (1983), T. L. Sims & R. B. Goldberg, unpublished results, and N. C. Nielsen, unpublished results, respectively. The Kunitz trypsin inhibitor sequences were those of K. D. Jofuku & R. B. Goldberg, unpublished results. The sequence CAACACAGT is present at nucleotide −104 in the 5′ CG-4 β-conclycinin gene region (S. Barker & R. B. Goldberg, unpublished results).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

of four different seed protein genes is presented in figure 5 (T. L. Sims & R. B. Goldberg, unpublished results). This analysis revealed the presence of a consensus sequence

$$5' \quad \begin{matrix} A \\ T \\ C \end{matrix} AACACA \begin{matrix} A \\ C \end{matrix} \begin{matrix} A \\ T \\ C \end{matrix} \quad 3'$$

that is present at least once in the 5′ flanking region of all soybean seed protein genes sequenced so far. Using the equations of Davidson *et al.* (1983), I calculate that the probability of this sequence's being present within a 400 nucleotide 5′ flanking region of six different genes is less than 3×10^{-6}. Although the functional relevance, if any, of this sequence is not yet known, it is possible that it could play some role in regulating seed protein gene expression.

Soybean seed protein genes are expressed in transformed tobacco plants

As a first step in identifying the *cis*-control elements that programme seed protein gene expression (figure 3), soybean β-congylcinin, Kunitz trypsin inhibitor and lectin genes were transferred individually to tobacco plants by means of a Ti-plasmid vector (Zambryski *et al.* 1983). In each tobacco transformant there was one unrearranged copy of the introduced soybean gene region, and each seed protein gene retained its expression programme during the tobacco life cycle (Okamuro *et al.* 1986; K. D. Jofuku, S. Barker & R. B. Goldberg, unpublished results).

The lectin gene was transferred to tobacco plants as part of a 17.1 kb DNA fragment containing four non-seed protein genes (Okamuro *et al.* 1986). The non-seed protein genes encode low-prevalence messages present on soybean embryo, leaf, root and stem polysomes (Okamuro *et al.* 1986). As shown in table 1 and figure 3, the lectin gene is expressed at a high level in soybean embryos and at a low level in the mature plant root. All genes present within the 17.1 kb soybean region retained their expression programme during tobacco development (Okamuro *et al.* 1986). Although slight quantitative differences occurred, each gene was expressed at the predicted times of the tobacco life cycle. For example, the lectin gene was expressed during seed development and in the mature plant root whereas the non-seed protein genes were expressed constitutively in the tobacco seed and in mature sporophytic organ systems. Figure 6 shows that the soybean lectin gene is regulated temporally during tobacco seed development as well. Analogously to events that occur during soybean embryogeny (figure 6*b*), lectin mRNA increased and then decreased during tobacco seed development (figure 6*a*, *b*). The exact timing of these mRNA prevalence changes, however, reflected physiological events inherent in tobacco seed development (i.e. 35 days compared with 110 days; figure 6*b*). These data indicate that the lectin gene is expressed in the correct differentiated states and is regulated temporally in transformed tobacco plants. Thus the DNA sequence elements and cellular factors necessary for both developmental-specific and constitutive gene expression are conserved between soybean and tobacco plants. Clearly the tobacco transformation system should permit the identification of DNA sequences required to control seed protein gene expression.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

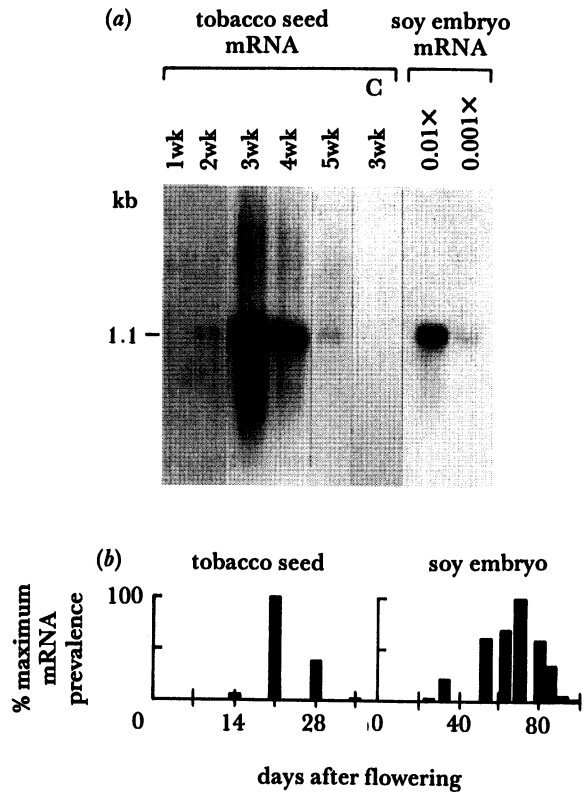


FIGURE 6. Soybean lectin gene expression in transformed tobacco plants. The soybean Le1 lectin gene (Goldberg *et al.* 1983) was transferred to tobacco plants by means of Ti-plasmid vector (Okamuro *et al.* 1986). The lectin gene was present on a 17.1 kb *Eco*RI DNA fragment that contains at least four non-seed protein genes (Goldberg *et al.* 1983; Okamuro *et al.* 1986). A lectin cDNA plasmid, designated L-9 (Goldberg *et al.* 1983), was hybridized with polysomal poly(A) mRNAs isolated from different stages of tobacco seed development. Wk refers to weeks after flowering. 0.01 × and 0.001 × refer to dilutions of soybean mid-maturation stage embryo mRNA, and represent lectin mRNA prevalences of 7.5×10^{-3} and 7.5×10^{-4} %, respectively (Okamuro *et al.* 1986).

THE IDENTIFICATION OF *TRANS*-ACTING FACTORS THAT REGULATE SEED PROTEIN GENE EXPRESSION

The model presented in figure 3 predicts that each seed protein gene is controlled by one or more regulatory genes that encode *trans*-acting factors. These factors interact with the seed protein gene control elements to activate transcription, or interact with seed protein gene transcripts to regulate post-transcriptional processes. To begin to identify *trans*-acting factors required to activate seed protein genes during development, a series of studies of DNA-binding proteins were carried out in my laboratory (K. D. Jofuku & R. B. Goldberg, unpublished results). Figure 7 presents the experimental approach that was utilized. In brief, embryo nuclear protein extracts were prepared and then incubated with labelled lectin gene fragments in the presence of unlabelled poly(dI:dC) competitor DNA (Singh *et al.* 1986). A specific protein–DNA complex was detected by mobility retardation of the labelled DNA fragment during electrophoresis (Singh *et al.* 1986).

The soybean lectin gene formed a specific complex with embryo nuclear protein. No detectable complex was observed with leaf, stem or root nuclear proteins. In addition, neither

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

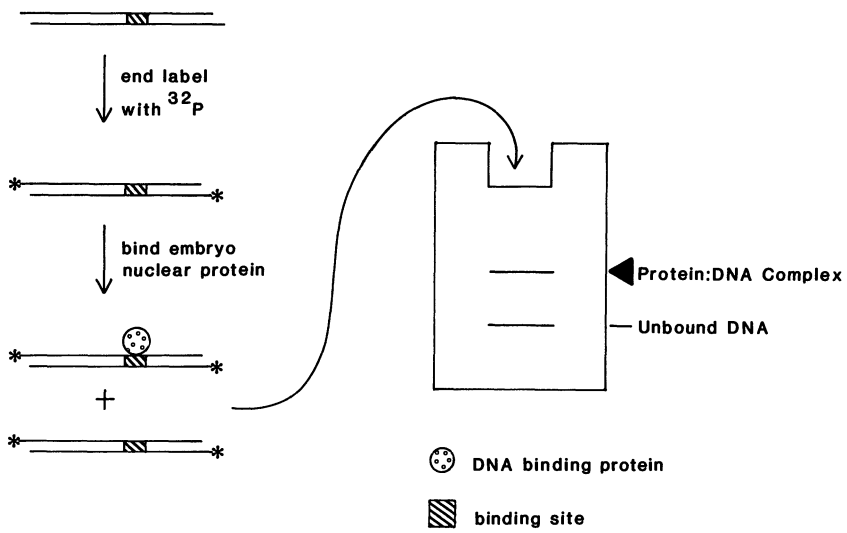


FIGURE 7. Detection of DNA binding proteins in nuclear extracts.

pBR322 DNA nor a soybean leghaemoglobin gene fragment competed with labelled lectin DNA for embryo nuclear protein binding. Thus the DNA–protein complex was specific for the lectin gene. A series of deletions were used to localize the protein-binding region. These experiments showed that the DNA-binding protein interacts with the 5′ lectin gene flanking region between −100 and −400 nucleotides from the transcription start site (figure 3). Although the physiological relevance of the embryo DNA-binding protein activity is not yet known, it is possible that we have identified a *trans*-acting factor necessary for the control of lectin gene expression. Identification of the specific DNA-binding sequence should permit the purification and characterization of the embryo nuclear protein(s) that bind to the lectin gene flanking region (Diffley & Stillman 1986). Clearly, a mechanistic understanding of the cellular processes that regulate plant expression should emerge in the not too distant future.

I express my deep appreciation to Diane Jofuku, Jack Okamuro, Joe Kamalay, Linda Walling, Gary Drews, Kathleen Cox, Jessie Truettner, Gisela Hoschek, Tom Sims, Bob Fischer, Susan Barker and John Harada for their dedication, thoughtfulness and tireless efforts in carrying out the experiments reported in this paper. These studies were supported by grants from the National Science Foundation and the United States Department of Agriculture.

REFERENCES

Davidson, E. H., Jacobs, H. T. & Britten, R. J. 1983 Very short repeats and coordinate induction of genes. *Nature, Lond.* **301**, 468–470.

Diffley, J. F. X. and Stillman, B. 1986 Purification of a cellular, double-stranded DNA binding protein required for initiation of Adenovirus replication by using a rapid filter-binding assay. *Molec. Cell Biol.* **6**, 1363–1373.

Galau, G. A. & Dure, L. S. 1981 Developmental biochemistry of cottonseed embryogenesis and germination: changing mRNA populations as shown by reciprocal heterologous cDNA/mRNA hybridization. *Biochemistry, Wash.* **20**, 4169–4178.

Goldberg, R. B., Hoschek, G., Kamalay, J. C. & Timberlake, W. E. 1978 Sequence complexity of nuclear and polysomal RNA in tissues of the tobacco plant. *Cell* **14**, 123–131.

Goldberg, R. B., Hoschek, G., Ditta, G. S. & Breidenbach, R. W. 1981*a* Developmental regulation of cloned superabundant embryo mRNA in soybean. *Devl Biol.* **83**, 218–231.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

Goldberg, R. B., Hoschek, G., Tam, S. H., Ditta, G. S. & Breidenbach, R. W. 1981*b* Abundance, diversity and regulation of mRNA sequence sets in soybean embryogenesis. *Devl Biol.* **83**, 201–217.

Goldberg, R. B., Hoschek, G. & Vodkin, L. O. 1983 An insertion sequence blocks the expression of a soybean lectin gene. *Cell* **33**, 465–475.

Gubler, U. & Hoffman, B. J. 1983 A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

Higgins, T. J. V. 1984 Synthesis and regulation of major proteins in seeds. *A. Rev. Pl. Physiol.* **35**, 191–221.

Kamalay, J. C. & Goldberg, R. B. 1980 Regulation of structural gene expression in tobacco. *Cell* **19**, 935–946.

Kamalay, J. C. & Goldberg, R. B. 1984 Organ-specific nuclear RNAs in tobacco. *Proc. Natn. Acad. Sci. U.S.A.* **81**, 2801–2805.

Okamuro, J. K., Jofuku, K. D. & Goldberg, R. B. 1986 The soybean seed lectin gene and flanking nonseed protein genes are developmentally regulated in transformed tobacco plants. *Proc. natn. Acad. Sci. U.S.A.* (In the press.)

Raghavan, V. 1976 *Experimental embryogenesis in vascular plants.* (603 pages.) New York: Academic Press.

Raghavan, V. 1986 *Embryogenesis in angiosperms.* (303 pages.) Cambridge University Press.

Singh, H., Sen, R., Baltimore, D. & Sharp, P. A. 1986 A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature, Lond.* **319**, 154–158.

St John, T. P. & Davis, R. 1979 Isolation of galactose-inducible DNA sequences from *Saccharomyces cervisiae* by differential plaque hydridization. *Cell* **16**, 443–452.

Staden, R. 1979 A strategy of DNA sequencing employing computer programs. *Nucl. Acids Res.* **6**, 2601–2610.

Vodkin, L. O., Rhodes, P. R. & Goldberg, R. B. 1983 A lectin gene insertion has the structural features of a transposable element. *Cell* **34**, 1023–1031.

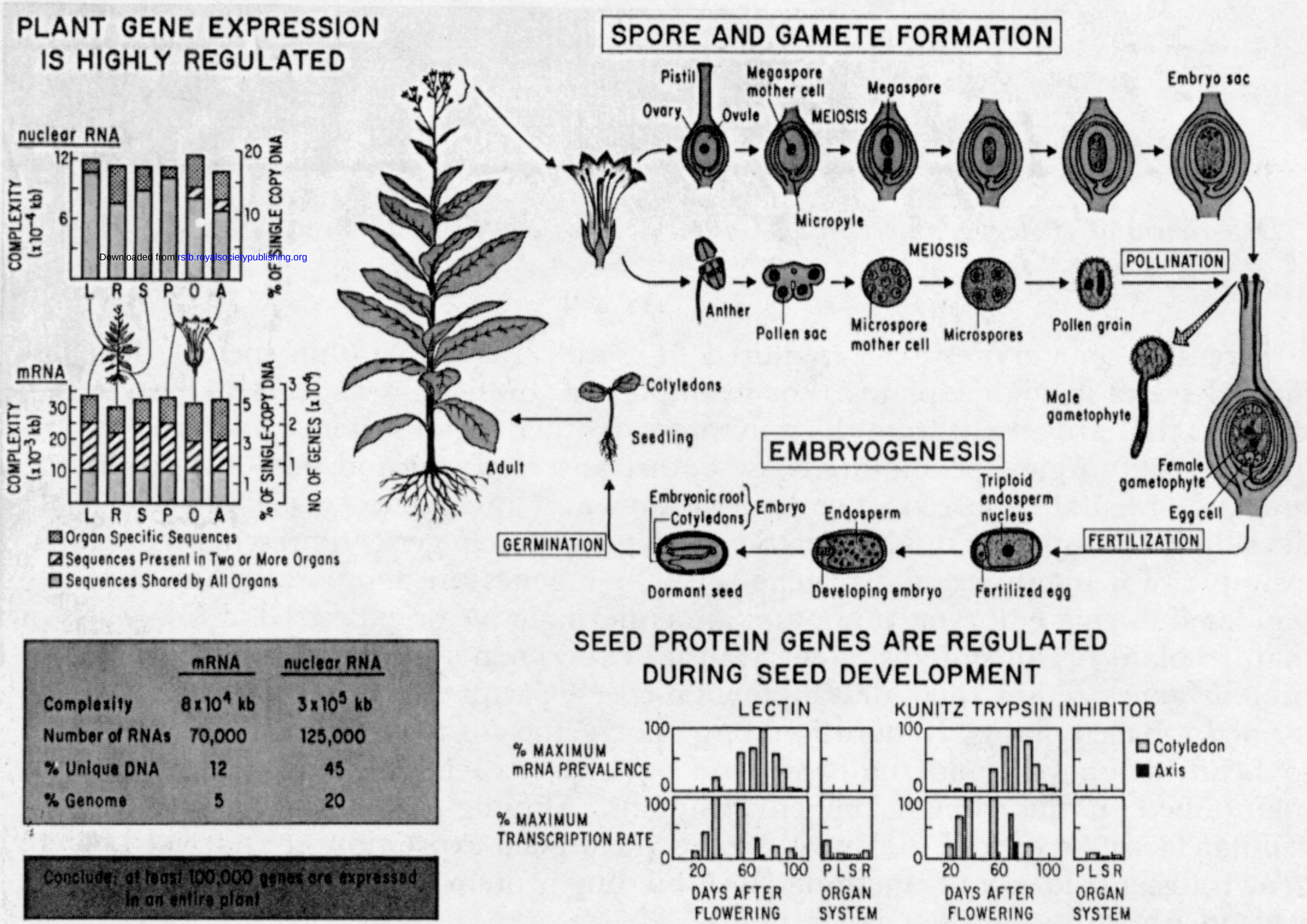
Walling, L., Drews, G. N, & Goldberg, R. B. 1986 Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2123–2127.

Willing, R. P. & Mascarenhas, J. P. 1984 Analysis of the complexity and diversity of mRNAs from pollen and shoots of *Tradescantia. Pl. Physiol.* **75**, 865–868.

Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M. & Schell, J. 1983 Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *EMBO J.* **2**, 2143–2150.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



Downloaded from rstb.royalsocietypublishing.org

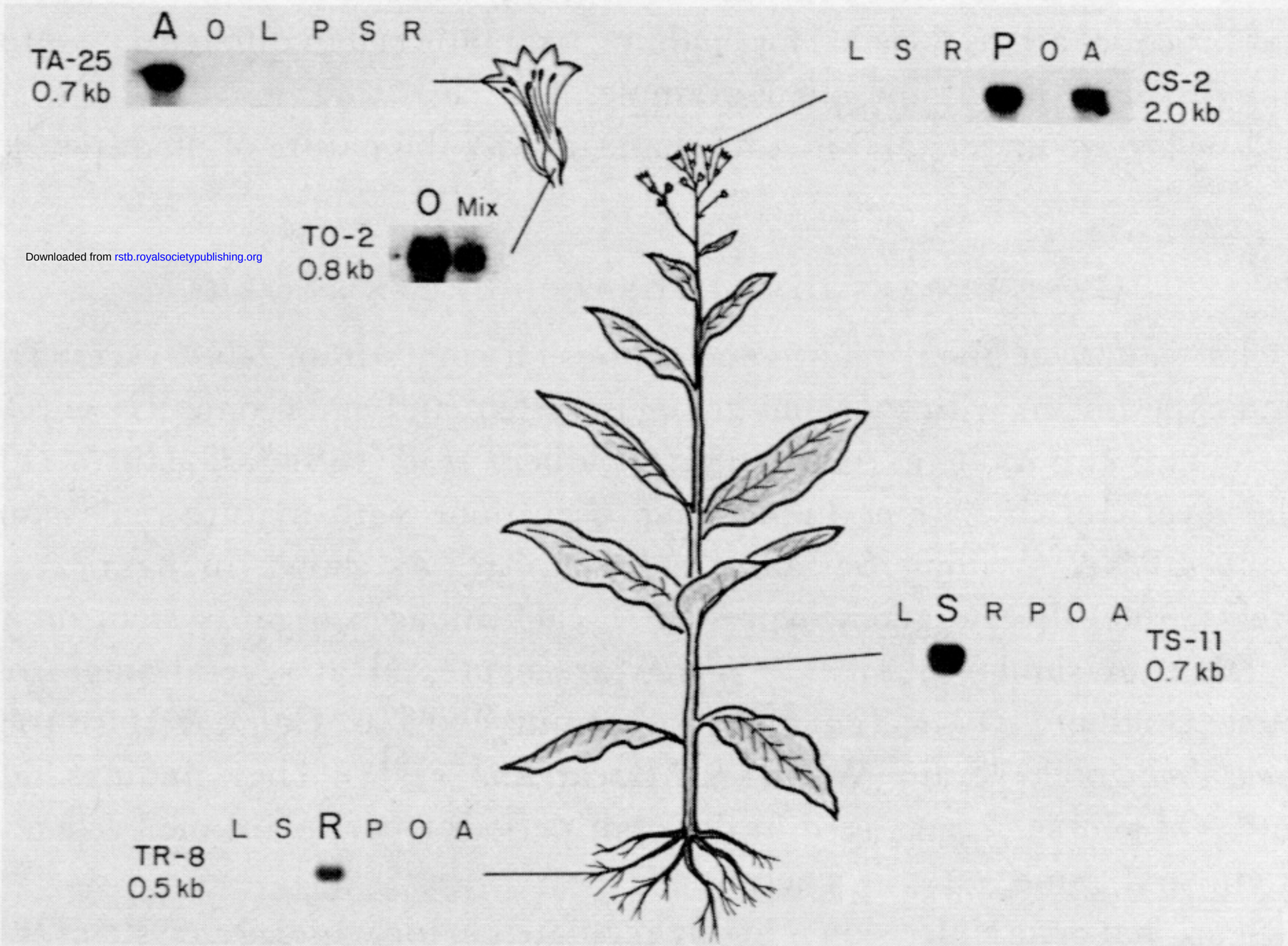
	mRNA	nuclear RNA
Complexity	8×10^4 kb	3×10^5 kb
Number of RNAs	70,000	125,000
% Unique DNA	12	45
% Genome	5	20

Conclude: at least 100,000 genes are expressed in an entire plant

FIGURE 1. Gene expression during the plant life cycle. Tobacco single-copy DNA sequence representation in organ system nuclear and polysomal RNAs was taken from the data of Kamalay & Goldberg (1980, 1984). L, R, S, P, O, A refer to leaf, root, stem, petal, ovary and anther, respectively. Developmental accumulation of soybean seed protein mRNAs and relative rates of seed protein gene transcription were taken from the data of Walling *et al.* (1986). P, L, S, R refer to post-germination cotyledons, leaf, stem and root, respectively. The plant life cycle depicts events that occur in tobacco development. Mature soybean seeds do not contain endosperm tissue.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



IGURE 2. Tobacco organ-specific gene expression. Tobacco anther (A), ovary (O), petal (P), stem (S) and root (R) cDNA libraries were constructed according to the procedure of Gubler & Hoffman (1983). The developmental characteristics of tobacco organ systems used to construct the libraries have been presented elsewhere (Kamalay & Goldberg 1980). Differential colony hybridization (St John & Davis 1979) was used to identify cDNA clones that represented mRNAs present exclusively, or at maximum prevalences, in each organ system. Organ-specific mRNA prevalences ranged from 0.1 to 1.0% of the relevant polysomal poly(A) mRNA. Representative cDNA clones were hybridized with gel blots containing anther, ovary, leaf (L), petal, stem and root mRNAs. The cDNA clone designation and corresponding mRNA size is listed next to each mRNA gel blot. CS-2 refers to a chalcone synthetase genomic clone (G. N. Drews & R. B. Goldberg, unpublished results). Mix refers to an equimolar mixture of leaf, stem, root, petal and anther mRNAs.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

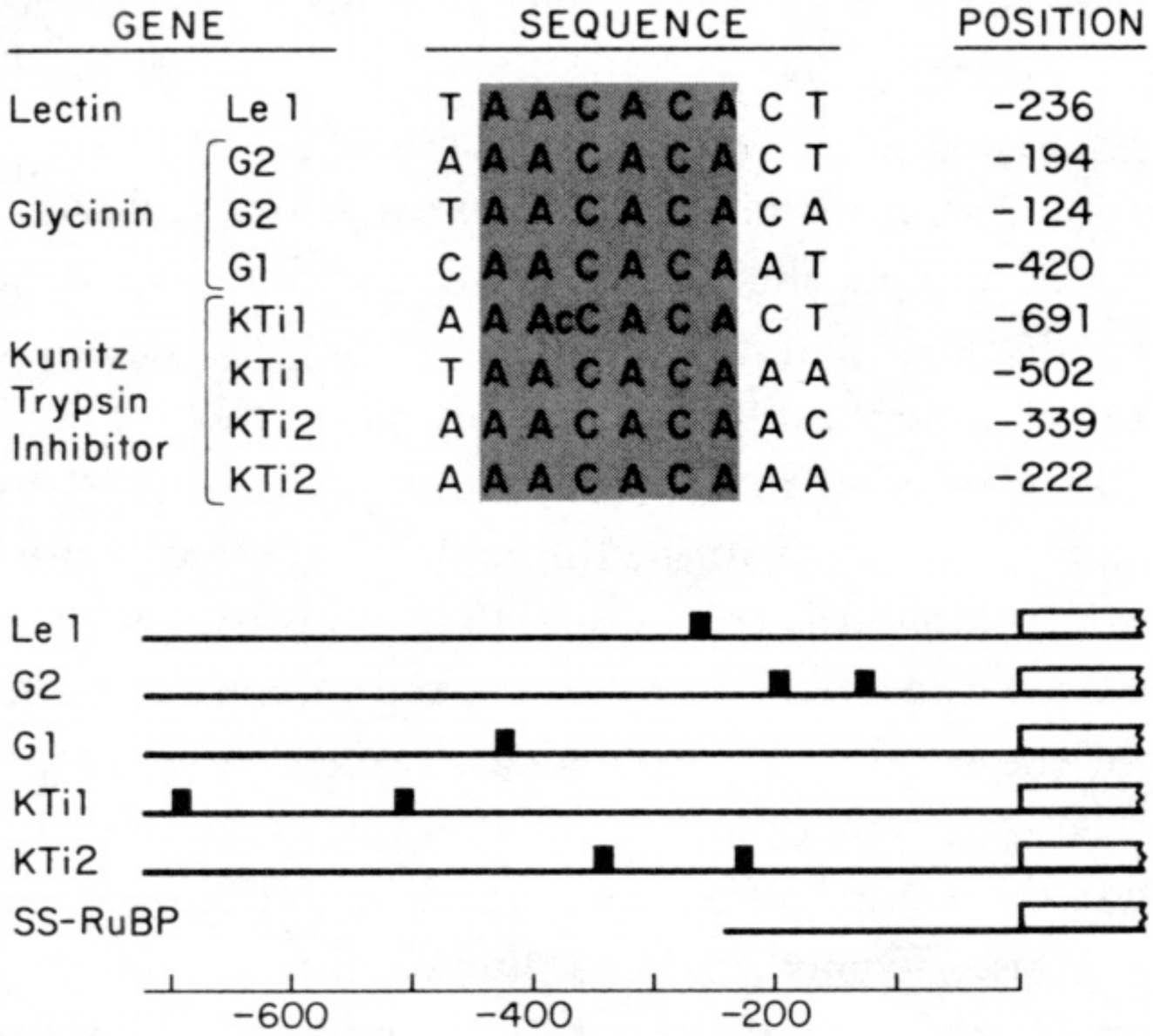


FIGURE 5. A consensus sequence in soybean seed protein gene 5′ regions. The VAX computer was used to search for sequence homologies in the upstream regions of several soybean seed protein genes (Staden 1979). Le1, G1 and G2 sequences were taken from the data of Vodkin *et al.* (1983), T. L. Sims & R. B. Goldberg, unpublished results, and N. C. Nielsen, unpublished results, respectively. The Kunitz trypsin inhibitor sequences were those of K. D. Jofuku & R. B. Goldberg, unpublished results. The sequence CAACACAGT is present at nucleotide −104 in the 5′ CG-4 β-conclycinin gene region (S. Barker & R. B. Goldberg, unpublished results).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

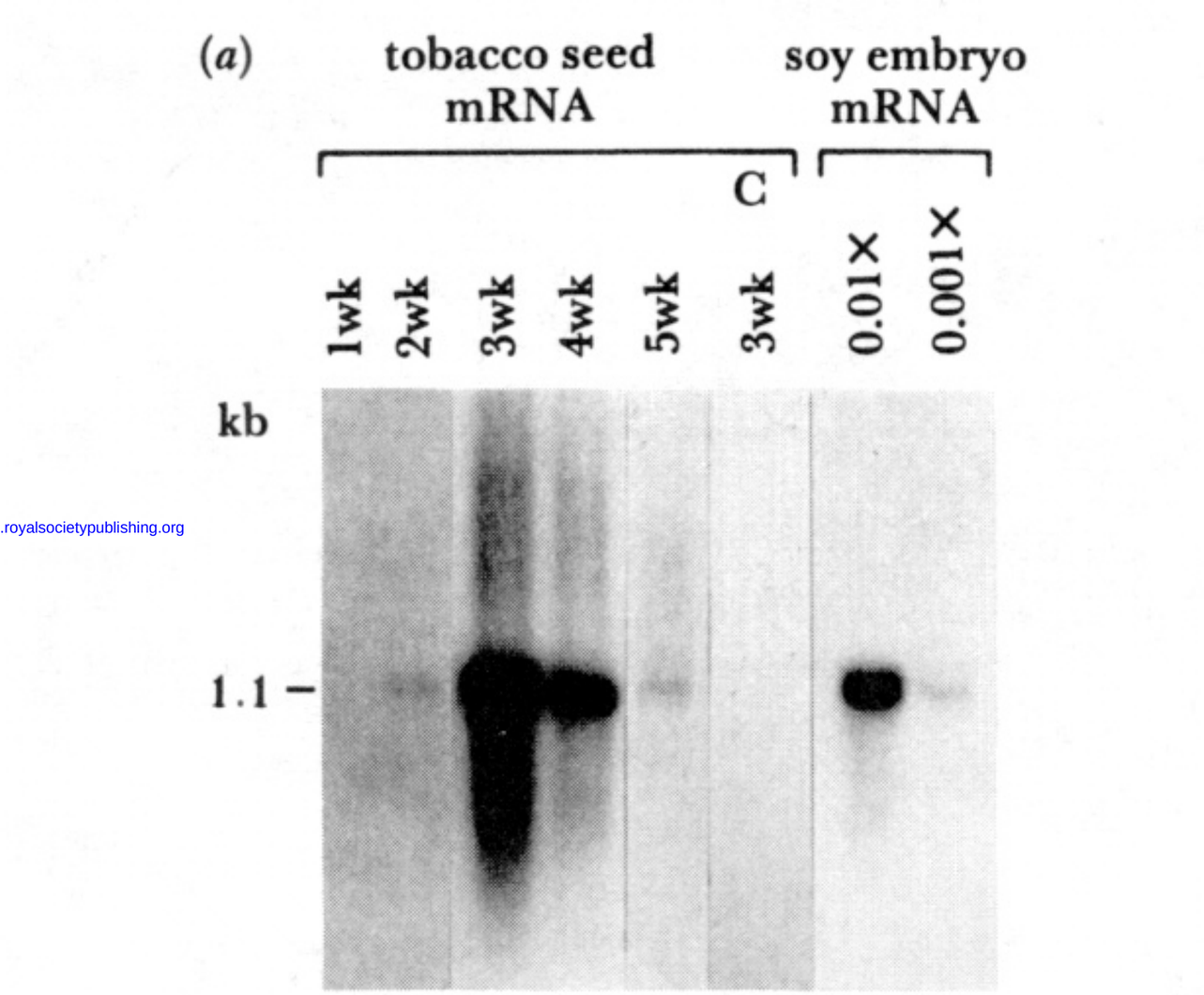


FIGURE 6. Soybean lectin gene expression in transformed tobacco plants. The soybean Le1 lectin gene (Goldberg *et al.* 1983) was transferred to tobacco plants by means of Ti-plasmid vector (Okamuro *et al.* 1986). The lectin gene was present on a 17.1 kb *Eco*RI DNA fragment that contains at least four non-seed protein genes (Goldberg *et al.* 1983; Okamuro *et al.* 1986). A lectin cDNA plasmid, designated L-9 (Goldberg *et al.* 1983), was hybridized with polysomal poly(A) mRNAs isolated from different stages of tobacco seed development. Wk refers to weeks after flowering. 0.01 × and 0.001 × refer to dilutions of soybean mid-maturation stage embryo mRNA, and represent lectin mRNA prevalences of 7.5×10^{-3} and 7.5×10^{-4} %, respectively (Okamuro *et al.* 1986).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES