The EMBO Journal vol.3 no.12 pp.2745-2752, 1984

Structure and chromosomal arrangement of leghemoglobin genes in kidney bean suggest divergence in soybean leghemoglobin gene loci following tetraploidization

Jong Seob Lee and Desh Pal S.Verma

Plant Molecular Biology Laboratory, Department of Biology, McGill University, 1205 Docteur Penfield Avenue, Montreal, Quebec, Canada H3A 1B1

Communicated by B.Hohn

We have determined the structure of one of the leghemoglobin (Lb) genes of *Phaseolus vulgaris* (kidney bean) and deduced the chromosomal arrangement of leghemoglobin genes by genomic hybridizations with Lb and two other sequences, each specific to the 5′ or 3′ region of the soybean leghemoglobin loci. By comparing this organization with two other species of legumes, *Glycine max* (soybean) and *G. soja* (wild soybean), a phylogeny of leghemoglobin gene loci was traced. The intragenic structure of the kidney bean leghemoglobin gene shows the same intron/exon arrangement as that of soybean leghemoglobin genes and extensive sequence homologies in both coding as well as 5′ and 3′ non-coding regions. The presence in the kidney bean genome of four leghemoglobin genes suggests that tandem duplications of a single primordial plant globin gene had occurred to generate four leghemoglobin genes in an 'Lb-locus' before *Glycine* and *Phaseolus* species diverged. Chromosome duplication by tetraploidization in *Glycine* generated two loci containing four genes each. A large deletion in one of the two four-gene loci in soybean resulted in the generation of the Lbc_2 locus containing two leghemoglobin genes, one functional and another pseudo ($Lb\psi_2$). This pseudogene, unlike that present on the main locus, is represented by only two and a half exons and appears to be truncated. The two other truncated genes (LbT_1 and LbT_2) were probably generated similarly in the genome of *Glycine* spp. following tetraploidization before the divergence of *G. max* and *G. soja*.

Key words: leghemoglobin/gene organization/*Phaseolus vulgaris*/tetraploidization

Introduction

Leghemoglobins are encoded by a family of closely related sequences in soybean (*Glycine max*) (Sullivan *et al.*, 1981; Brisson *et al.*, 1982), which exist as functional, pseudo and truncated genes (Brisson and Verma, 1982; Hyldig-Nielsen *et al.*, 1982; Wiborg *et al.*, 1982). The primary structure of leghemoglobin (Lb) genes was found to be very similar to those of mammalian globin genes with respect to the position of two introns common to all globin genes as well as the presence of several regulatory sequences on the 5′ end of these genes (Lee and Verma, 1984; Brown *et al.*, 1984). The presence of the third intron in the Lb genes suggested that plant globin genes are primitive globin genes (Lewin, 1981; Lee and Verma, 1984). Furthermore, the arrangement of Lb genes on the chromosome is very similar to animal globin loci (Efstratiadis *et al.*, 1980; Lee *et al.*, 1983). Although it has been suggested that Lb genes may have been transferred horizontally from animals to legumes *via* a retrovirus-like vector (Jeffreys, 1982), their presence in non-leguminous plants that are able to fix nitrogen symbiotically (Appleby *et al.*, 1983) indicate a much older origin of these genes.

Since more information on the structure and chromosomal arrangements of Lb genes in other legumes is essential to fully understand the evolution of these genes, we have carried out experiments with two other closely related legume species, *Glycine soja* (wild soybean) and *Phaseolus vulgaris* (kidney bean). We isolated and characterized, at the nucleotide level, a Lb gene from a kidney bean genomic library. Also, we searched their genomes for the presence of two sequences which were found to be at the 5′ and 3′ regions of the main Lb locus of soybean. The latter appears to be a characteristic of all loci containing Lb sequences. These results made it possible to trace the chromosomal arrangement of Lb genes and the evolutionary relationship among these three related legumes.

Results

Leghemoglobin gene loci in soybean

The chromosomal arrangement of Lb genes in soybean (Figure 1) consists of four loci containing Lb sequences (Lee *et al.*, 1983). Two loci carry functional genes while the other two have truncated Lb sequences. To investigate how these loci may have been generated, we characterized the chromosomal arrangements of Lb genes in two other related species, *G. soja* and *P. vulgaris*. We searched the genomes of these two species for sequences that surround the main Lb locus in soybean carrying four Lb genes, Lba, Lbc_1, $Lb\psi_1$ and Lbc_3. The first sequence is the 3′ sequence (designated as an 's' in Figure 1) which was found to be specific to the 3′ ends of all four loci containing Lb sequences in soybean. The other sequence which is present at the 5′ end, has been found to be expressed in root and leaf tissue (designated as R/L in Figure 1). We deduced from these results (see below), possible chromosomal arrangements of Lb genes in these two species and compared them with that of soybean to trace the mode of evolution of Lb gene arrangement.

Structure of the second leghemoglobin gene on the soybean Lbc_2 locus

The Lbc_2 locus contains two leghemoglobin genes, one of which was identified as the Lbc_2 gene (Wiborg *et al.*, 1982; see also Lee *et al.*, 1983), but the other upstream gene was not yet characterized. We analyzed the nucleotide sequence of the gene upstream of Lbc_2 called ψ_2 in Figure 1. This gene carries only the first two exons and a half of the third exon of leghemoglobin (Figure 2). The nucleotide sequence revealed the presence of several copies of a short repeated sequence, CCA/TCCC at the end of the gene. Also, two important point mutations were found, one at the second exon (TTG to TAG) and the other at the splicing junction between the second exon and intron (GT to CT). The first would result in an in-phase termination while the second might affect RNA splicing. Otherwise, this DNA sequence shows a high hom-

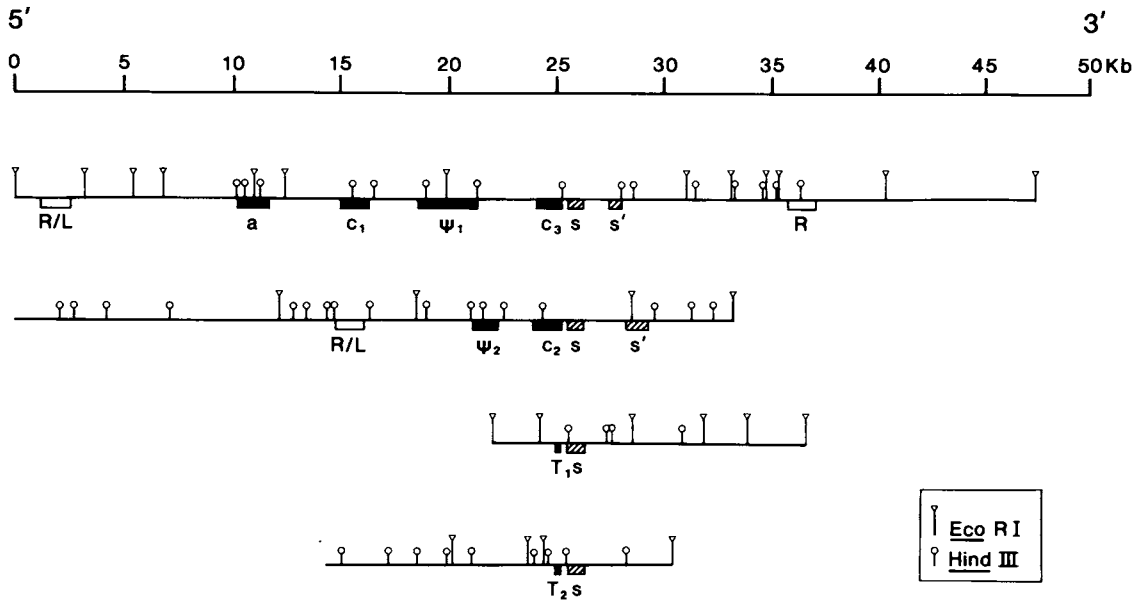

Fig. 1. Chromosomal arrangement of leghemoglobin genes in soybean. The *Eco*RI and *Hind*III restriction maps of the regions carrying leghemoglobin sequences were derived from the detailed analysis of lambda clones of these regions of the chromosome. R and R/L are sequences expressed in root and root/leaf, respectively, s and s' are two repeat elements [see Lee *et al.* (1983) for more details].

ology with that of the Lba gene (Wiborg *et al.*, 1982) in the main (Lba) locus.

Structure of a kidney bean leghemoglobin gene

We constructed a partial cDNA library of kidney bean nodule mRNA and isolated a plasmid, pJSLb15, which contains an insert of ~720 bp long, representing almost the complete Lb sequence of *P. vulgaris* (Lee, 1984). The nick-translated insert of pJSLb15 was used as probe for screening a *Phaseolus* genomic library and for Southern hybridization of *Phaseolus* genomic DNA to determine the number and possible arrangement of Lb genes in the kidney bean genome.

A kidney bean *Mbo*I partial genomic library constructed in lambda 1059 was screened with the insert of pJSLb15. The positive clones were mapped with the restriction enzymes, *Eco*RI, *Bam*HI and *Hind*III. A Lb cDNA-hybridizing region was localized on one of the clones, PvLb1. The nucleotide sequence of the Lb-containing region (Figure 3) indicated that this Lb gene is complete. It is interrupted by three intervening sequences which are located between codons 32 and 33, 68 and 69 and 103 and 104, the same positions as those of soybean Lb genes (see Brisson and Verma, 1982). The corresponding intervening sequences are 78, 95 and 88 nucleotides long.

The amino acid sequence deduced from the nucleotide sequence matched the known amino acid sequence of kidney bean leghemoglobin consisting of 145 amino acids (Lehtovaara and Ellfolk, 1975b), except at the positions 98 (Ser-Asn), 99 (Asn-Asp), 100 (Asp-Ser) and 125 (Gln-Glu). The changes at positions 98, 99 and 100 could be due to errors in the protein sequence whereas the change at position 124 is due to a transition (CAA to GAA).

The region upstream from the initiation codon revealed the conservation of sequences with those of soybean Lb genes, including two sequences, CCAAT and TATAAA that are common to other eukaryotic genes. However, the position of the CCAAT box is closer to the TATAAA box in the *Phaseolus* Lb gene as compared with that in soybean. A sequence (~30 bp) identified in the region surrounding the 'cap' site of soybean Lb and animal globin genes (Lee and Verma, 1984; Brown *et al.*, 1984) also shows extensive homology with the same region of the kidney bean gene, suggesting that it may be essential for the expression of these genes. Analysis of the 3' non-coding region revealed a consensus sequence (AATAAA) for the poly(A) addition signal. Thus, these results suggest that this Lb gene has all the attributes of a normal eukaryotic gene. Since the bulk of the leghemoglobin in *Phaseolus* nodules is represented by the deduced sequence, this gene may be a functional gene.

Four leghemoglobin genes in kidney bean genome

Genomic DNA from kidney bean embryonic axes was digested with various restriction enzymes, separated on agarose gel, transferred onto GeneScreen Paper and hybridized with the nick-translated insert of pJSLb15. As shown in Figure 4, it revealed several distinct bands, suggesting a family of related sequences. The presence of four hybridizing bands in most lanes, whether from single or double digestions, indicated that there are four Lb sequences on the chromosome of kidney bean.

In the kidney bean, there is only one major Lb component, PhLba, which is post-translationally modified into one minor component, PhLbb (Lehtovaara and Ellfolk, 1975a). It is unclear whether one gene is expressed and the others are silent or all four genes code for one identical component.

Presence of the common 3' repeat sequence of soybean in the genomes of kidney bean and G. soja

As shown in Figure 1, four Lb gene loci in soybean share a common sequence specific to the 3' end (Lee *et al.*, 1983). We searched genomic DNAs of two other species for the presence of this sequence, using as a probe a 2.7-kb *Hind*III fragment adjacent to the Lbc_3 gene which had been subcloned into pBR322. Figure 5(A) shows the hybridization pattern of kidney bean genomic DNA digested with various restriction enzymes. It revealed a single band on each lane, suggesting the presence of only one sequence in the kidney bean genome. However, there are as many as four hybridizing bands in the

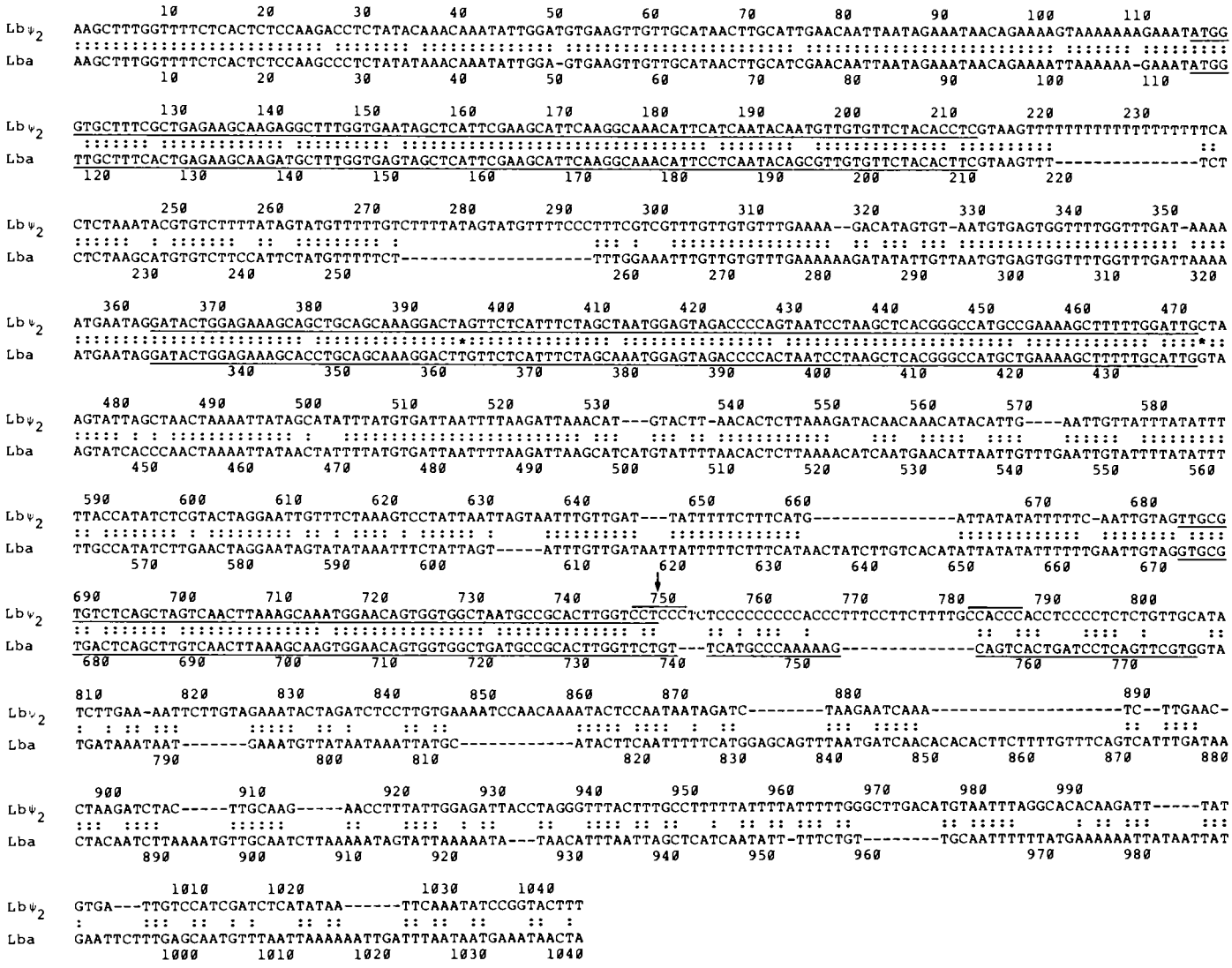

Fig. 2. Nucleotide sequence of the second gene on the Lbc_2 locus. The nucleotide sequence of a pseudo (truncated) gene on the Lbc_2 locus was determined and compared with the corresponding region of the Lba gene on the main (Lba) locus (Hyldig-Nielsen *et al.*, 1982). Homologous sequences between two genes are indicated by colons, two point mutations are shown by stars and exons are underlined. The deletion point is indicated by an arrow. Dashes (-) represent the sequences that might have been deleted or inserted. A short repeated sequence, CCA/TCCC is upperlined at the end of this gene.

case of *G. soja* [Figure 5(B)]. Since it was found that the 2.7-kb *Hind*III fragment contains two sequences, only one of which is common to all four loci in soybean (Lee *et al.*, 1983), the presence of almost the same number of hybridizing bands as that in soybean may suggest that Lb genes in *G. soja* are arranged similarly to those in soybean.

Search for the 5′ sequence expressed in root and leaf

In soybean, the Lba and Lbc_2 loci were found to be flanked by a sequence 5′ to them which is expressed preferentially in root and leaf tissue (Lee *et al.*, 1983). To determine whether this gene is always linked to the Lb locus, we searched for this sequence in the genomes of two other legume species. Figure 6 shows the hybridization pattern of genomic DNAs digested with *Bam*HI or *Eco*RI using as a probe a 1.5-kb *Hind*III fragment from the Lbc_2 locus which had been subcloned into pBR322. It revealed that *G. soja*, like soybean, carries two copies of this sequence while only one copy of this sequence was found in the kidney bean genome.

Discussion

The second Lb gene on the soybean Lbc_2 locus is truncated

The nucleotide sequence of the second Lb gene on the Lbc_2 locus indicates that the Lb gene is truncated and is different from two other truncated genes (LbT_1 and LbT_2) carrying only the last exon (Brisson and Verma, 1982; Lee, 1984). How this truncated gene was generated on the Lbc_2 locus is unclear. However, a comparison of the 5′ non-coding region of the $Lb\psi_2$ gene on the Lbc_2 locus with that of the Lba gene revealed 96% homology between two genes. Also, a similarity was found between the Lbc_2 and Lbc_3 genes with respect to the number of amino acids. They encode one extra amino acid as compared with the Lba and Lbc_1 genes. Thus, a deletion in the main locus (Lba) may have given rise to the Lbc_2 locus. The mechanism by which the deletion occurred is not apparent. The presence of a few copies of a short repeat sequence at the breakpoint (see arrow in Figure 2) may suggest the possibility of the involvement of these sequences in

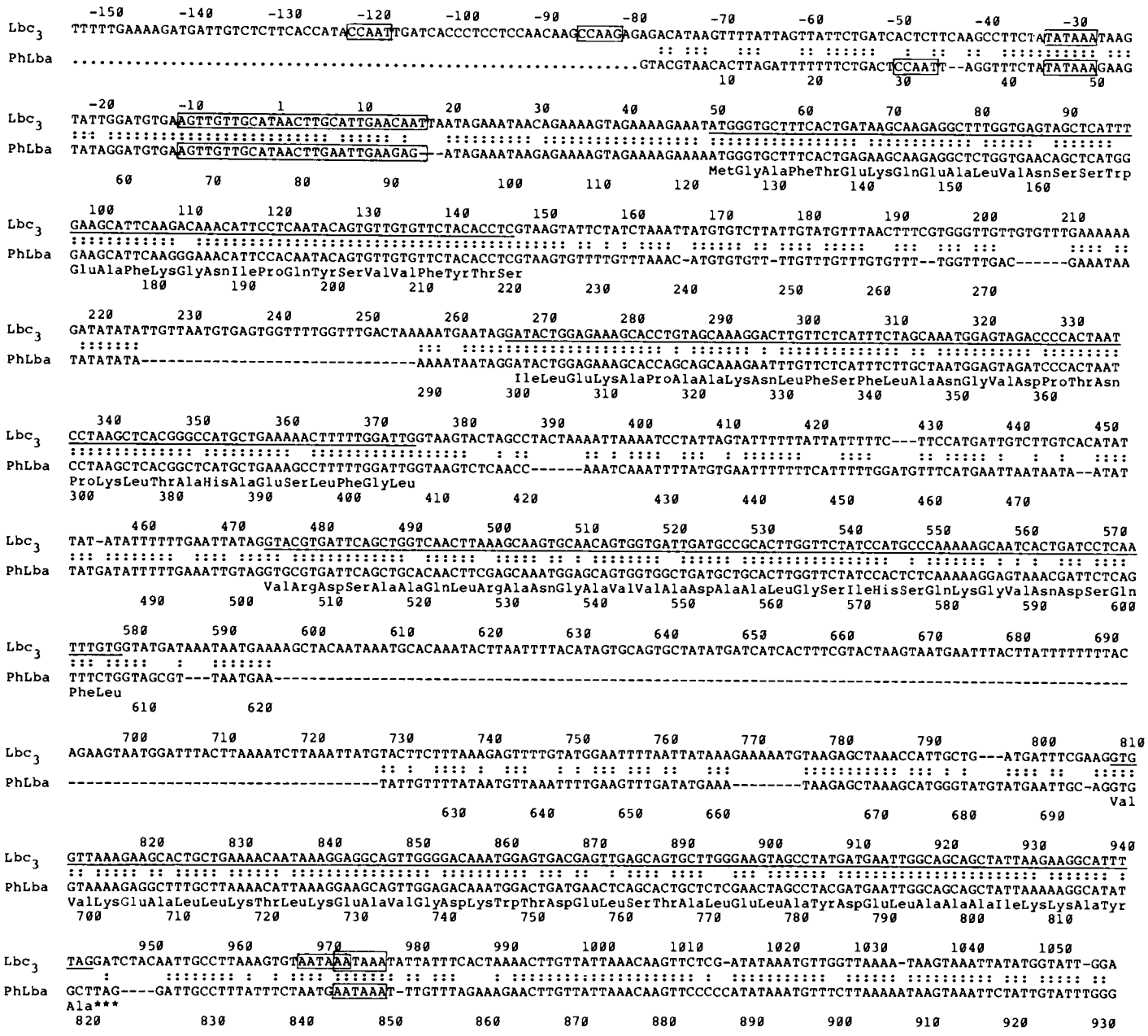

Fig. 3. Nucleotide sequence of a kidney bean leghemoglobin gene and its comparison with that of the soybean Lbc_3 gene. The nucleotide sequence of a kidney bean Lb gene present on the clone PvLb1 was compared with that of the Lbc_3 gene of soybean (Brisson and Verma, 1982). Dashes (-) represent the sequences which might have been deleted or inserted. Consensus sequences found at the 5′ and 3′ regions (Brown *et al.*, 1984; Lee and Verma, 1984) are boxed. Homologous sequences between two genes are indicated by colons. The exons in the Lbc_3 gene are underlined and the amino acids corresponding to exons in the *Phaseolus* gene are indicated.

homologous recombination. It is probable that the deletion is the result of a non-homologous exchange during replication as recently observed in the human β-globin gene cluster (Vanin *et al.*, 1983).

Kidney bean leghemoglobin gene structure is identical to that of soybean leghemoglobin genes

The nucleotide sequence of a kidney bean leghemoglobin gene indicated that this complete gene is interrupted by three intervening sequences which are located at the same positions as those in soybean leghemoglobin genes (Brisson and Verma, 1982) but are shorter. The difference in the number of introns also found in the *Phaseolus* Lb gene is consistent with the idea that these genes are ancestral to the animal globin genes (Lewin, 1981; see Lee and Verma, 1984) and that introns could have been removed during evolution (Blake, 1983) as with the actin genes and the preproinsulin gene.

Sequence comparison of this gene with one of the soybean Lb genes showed a significant structural similarity, indicating their close evolutionary relationship. Also, it suggests that they have been under certain constraints that prevent them from diverging. Similar comparison between seed storage proteins of soybean and kidney bean (Schuler *et al.*, 1983) shows that although conglycinin and phaseolin genes may have diverged from a single ancestral gene, their level of divergence is much greater than that of Lb genes of two species.

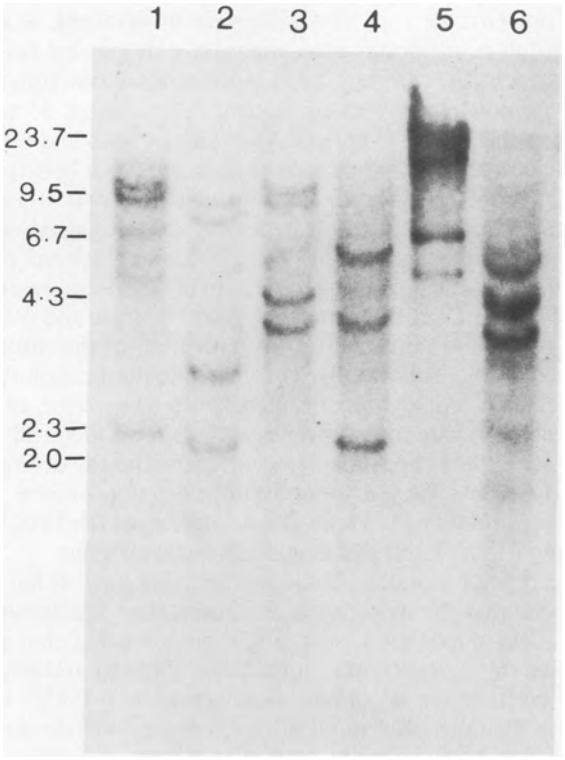

Fig. 4. Identification of restriction fragments containing leghemoglobin sequences in kidney bean genomic DNA. DNA isolated from kidney bean embryonic axes and digested with various restriction enzymes was subjected to Southern hybridization with nick-translated inserts of pJSLb15. **Lanes:** **1**, *Eco*RI; **2**, *Eco*RI and *Hind*III; **3**, *Hind*III; **4**, *Eco*RI and *Bam*HI; **5**, *Bam*HI; and **6**, *Hind*III and *Bam*HI. The positions and sizes (in kb) of the *Hind*III-digested lambda DNA are shown.

Chromosomal arrangements of leghemoglobin genes in kidney bean and G. soja

The genomic hybridizations indicate that two sequences, each specific to the 5′ or 3′ region of the soybean Lb loci are also present in the *Phaseolus* genome. As identified in the Lba and Lbc_2 loci of soybean (Lee *et al.*, 1983), the presence of only one copy of each sequence in the kidney bean genome suggests that the four Lb genes may be arranged on one location similarly to the arrangement on the Lba locus of soybean. However, it seems likely that the intergenic regions of the kidney bean 'Lb locus' are quite long as compared with those of soybean loci. Also, the 3′ sequence specific to Lb loci does not appear to be as closely associated with any kidney bean Lb sequence as that in Lb loci of soybean. This could be explained by the genome sizes of the two species. Although soybean genome was duplicated by tetraploidization ($2n = 40$) (Hadley and Hymowitz, 1973), its size (1.8×10^9 bp) (Walbot and Goldberg, 1978) is almost identical to that (1.9×10^9 bp) of kidney bean genome ($2n = 22$) (Sun *et al.*, 1981; Evans 1980).

G. soja is the closest ancestral relative of the modern soybean (*Glycine max*). *G. soja* and *G. max* both have 40 chromosomes and the two species can be intercrossed (Hymowitz and Newell, 1980). Therefore, we also searched the genomic DNA of *G. soja* for the presence of two 5′ and 3′ sequences in order to predict the chromosomal arrangement of Lb genes in *G. soja*. The presence of the two copies of the 5′ gene and at least four copies of the 3′ repeat sequence indicates that the Lb genes in *G. soja* are arranged on the chromosome, similarly to those of soybean, probably in four loci.

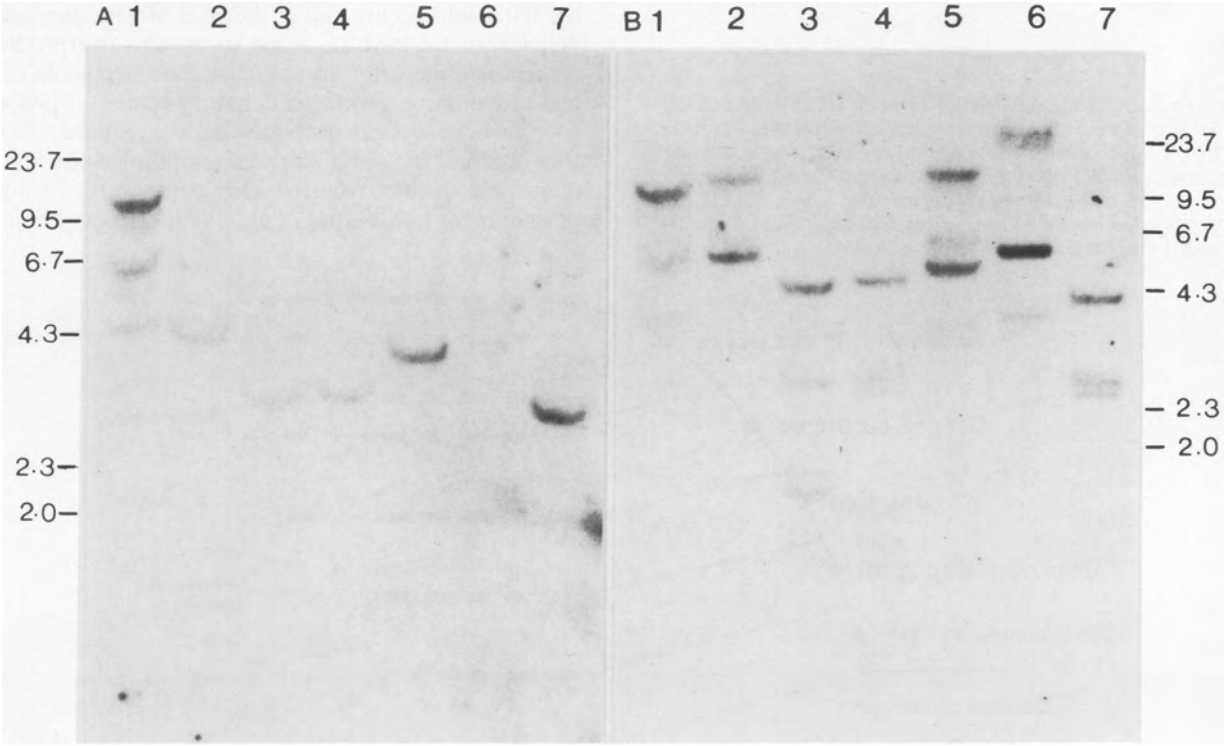

Fig. 5. Southern blots of kidney bean genomic DNA **(A)** and *G. soja* DNA **(B)** for the search of the 3′ repeat sequence. Genomic DNAs from these species were digested with combinations of *Eco*RI, *Hind*III and *Bam*HI and hybridized with the nick-translated 2.7 kb *Hind*III fragment adjacent to the Lbc_3 gene of soybean (Lee *et al.*, 1983). The positions and sizes (in kb) of the markers (*Hind*III-digested lambda DNA) are shown. **Lanes: 1**, soybean DNA cut with *Eco*RI; **2**, *Eco*RI; **3**, *Eco*RI and *Hind*III; **4**, *Hind*III; **5**, *Eco*RI and *Bam*HI; **6**, *Bam*HI; and **7**, *Hind*III and *Bam*HI.

Evolutionary relationship among leghemoglobin genes in different species

We have traced the possible evolutionary divergence in the arrangement of Lb genes in three species (Figure 7). On the basis of our findings we suggest three major evolutionary steps leading to soybean leghemoglobin genes: (i) gene duplication, (ii) genome duplication and finally (iii) deletion in one of the Lb loci.

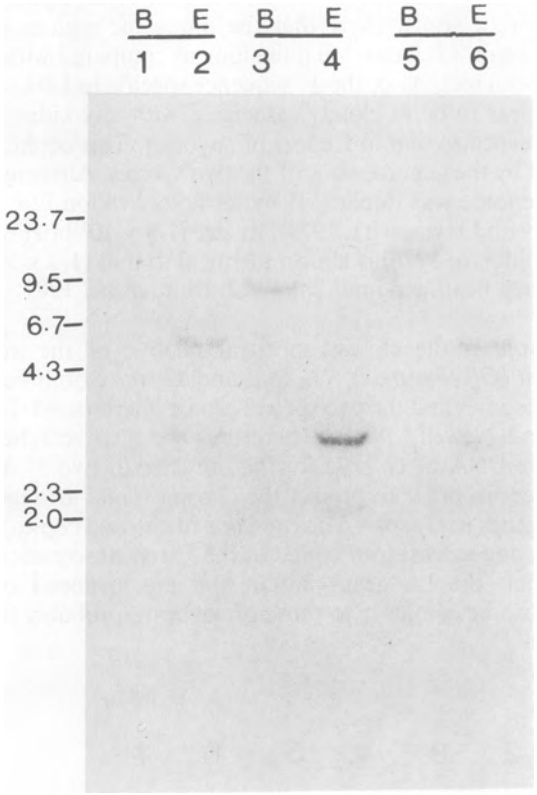

Fig. 6. Southern blot of genomic DNAs of soybean, kidney bean and *G. soja* for the search of the 5′ sequence expressed in the root and leaf tissue. Genomic DNAs of three species were subjected to Southern hybridization after digestion with *Bam*HI (B) or *Eco*RI (E). **Lanes: 1** and **2**, soybean; **3** and **4**, kidney bean; **5** and **6**, *G. soja*. Nick-translated 1.5-kb *Hind*III fragments were used as a probe. The positions and sizes (in kb) of the markers (*Hind*III-digested lambda DNA) are indicated.

Before *Glycine* spp. and kidney bean diverged, a single primordial plant globin gene might have duplicated twice to generate a locus, carrying four Lb genes which was surrounded by two different types of sequences, one at the 5′ region and the other at the 3′ region. After kidney bean and *Glycine* spp. had diverged, a truncated gene might have been generated from the last gene of the four gene locus. It is unclear how the truncated gene was generated and dispersed on the same or different chromosome. A recombinational event, probably unequal crossing over, involving a short repeat sequence found in the truncated gene (Brisson and Verma, 1982) may be responsible for the generation of the truncated sequence. This, followed by chromosome duplication (tetraploidization), could have resulted in the generation of two loci, each carrying four Lb genes and also two identical truncated genes. This presumably occurred before the divergence of two *Glycine* species. In soybean, a deletion on one four-gene locus removed ~12 kb DNA to generate the Lbc_2 locus carrying a functional and pseudo (truncated) gene.

This type of evolution is not unique to legumes. It has been suggested that the *Xenopus laevis* globin gene family evolved by tandem duplication of a single primordial globin gene, followed by chromosome duplication through tetraploidization (Jeffreys *et al.*, 1980; Hosbach *et al.*, 1983). If we assume that the gene duplications occurred before genome duplication, the extensive homologies among soybean Lb genes could have been maintained by concerted evolution (Zimmer *et al.*, 1980; Brown *et al.*, 1984).

Polyploidy creates an abundance of genetic material, which can be exploited by subsequent mutation and selection (Ohno, 1970). It has been suggested that after polyploidization there is a strong tendency to evolve into a diploid state by chromosomal rearrangement and by sequence diversification (Leipoldt and Schmidtke, 1982). This diploidization of the tetraploid occurs also at the level of gene expression. The deletion might have occurred to prevent overproduction of leghemoglobins after tetraploidization. Since only one major leghemoglobin is produced in kidney bean, it is probable that four genes in soybean are expressed at a reduced rate although their multiplicity, variability and temporal expression indicate a possible distinct role for each component during nodule development (see Verma *et al.*, 1979; Verma, 1982).

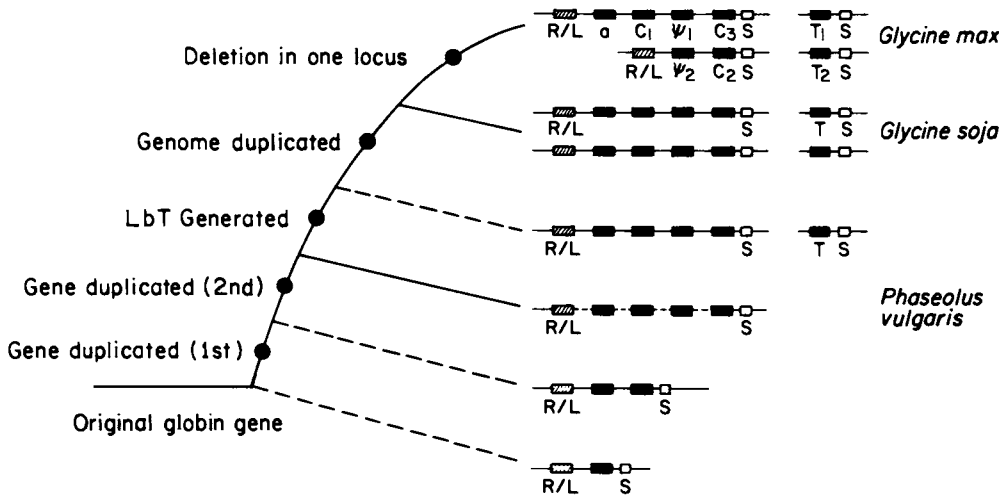

Fig. 7. Phylogeny of the leghemoglobin loci. The chromosomal arrangements deduced from this study revealed possible events which could have occurred through the evolutionary time to lead to the leghemoglobin organization in soybean. LbT, an ancestral truncated gene.

Plant versus animal globin genes

Globins are wide-spread in nature. They include the tetrameric hemoglobins of higher vertebrates, monomeric hemoglobins of protochordates, a variety of invertebrate globins, monomeric myoglobins, the monomeric leghemoglobins in legumes (Hunt *et al.*, 1978) and the dimeric globins found in the root nodules of nitrogen-fixing non-leguminous plants (Appleby *et al.*, 1983). Although plant and animal globin genes might have diverged between 900 million and 1.4 billion years ago (the estimated time for the plant-animal divergence) (Brown *et al.*, 1984), their similarities in structure as well as in chromosomal arrangement (Lee *et al.*, 1983) suggest that they have been under certain evolutionary constraints.

It is generally believed that duplication of hemoglobin and myoglobin genes occurred ~700 million years ago before the divergence of the α- and β-globin genes ~500 million years ago. In contrast with hemoglobin, which functions as a tetrameric protein, myoglobin is monomeric and encoded by a single gene from a family of related sequences in humans. The chromosomal arrangement of these genes is not yet known (Weller *et al.*, 1984). It may be similar to the kidney bean leghemoglobin. However, the presence of only two introns on the myoglobin gene rules out the possibility of a close evolutionary relationship to leghemoglobins.

Although this study made it possible to trace the mode of evolution for three related species for a short time period, the recent findings of dimeric hemoproteins in non-leguminous plants (Appleby *et al.*, 1983) raise an important question about the origin and evolution of plant globin genes. Jeffreys (1982) has suggested that the leghemoglobins arose as a result of a horizontal gene transfer from an animal to an ancestral legume plant. However, this does not seem likely. The X-ray crystallographic studies on lupin leghemoglobins (Vainshtein *et al.*, 1975) suggested that animal globin and leghemoglobin had been generated ~1.2 billion years ago (the estimated time for the plant-animal divergence). Leghemoglobin does not appear to be more closely related to certain globins which may have been transferred. The extent of corrected sequence divergence for amino acid replacements causing base substitutions is >100% between the plant and animal globin sequences (Brown *et al.*, 1984). This may indicate that plant and animal globin gene families diverged over one billion years ago. It is possible that ancestral hemoglobin genes existed in many (if not all) higher and lower plant families, but these genes may have become silent in plants which have not acquired the ability to associate with nitrogen-fixing organisms. The selective advantage which biological nitrogen-fixation provided to the legume plants may have fixed these genes in these plants. Also, since there are other nodule-specific genes in legumes, they may have co-evolved with leghemoglobin genes to optimize this unique association in nature. More data about Lb genes in other legumes, as well as the globin genes in non-leguminous plants, will shed light on the origin and evolution of this unique group of plant genes.

Materials and methods

Growth of plant tissue

Twenty-one day nodules of kidney bean (*P. vulgaris*) were obtained following inoculation of the seedlings with *Rhizobium phaseoli* (strain RCR3610). Plants were grown as described (Verma and Bal, 1976; Verma *et al.*, 1974) and the tissue was stored under liquid nitrogen.

cDNA synthesis and cloning

Poly(A)$^+$ RNA was isolated from total polysomes of nodules as described (Verma *et al.*, 1974; Auger *et al.*, 1979). Synthesis and cloning of double-stranded cDNA of nodule RNA was essentially carried out as described (Fuller *et al.*, 1983). *Pst*I-cut pBR322 tailed with dGTP was obtained from Bethesda Research Laboratories.

Molecular cloning and isolation of DNAs

Various restriction fragments from genomic clones were subcloned into pBR322. Plasmid DNAs were purified on CsCl/ethidium bromide gradients. Genomic DNA from soybean and kidney bean embryonic axes and DNA from leaves of *G. soja* were isolated as described (Varsanyi-Breiner *et al.*, 1979). Phage DNA was isolated by the methods of Blattner *et al.* (1977) and Maniatis *et al.* (1978).

Isolation of a kidney bean genomic clone from a bacteriophage library

About 5 x 10^5 recombinant bacteriophage were screened as described by Woo (1979) from a *Mbo*I partial genomic library of kidney bean embryonic axes DNA constructed in lambda 1059. *Eco*RI strain K802 was used as the host.

Southern hybridization

DNA digested with restriction endonucleases (Boehringer Mannheim) were electrophoresed through agarose gels and transferred to GeneScreen Paper (New England Nuclear) by the method of Southern (1975). Pre-treatment, hybridization and washing of filters were performed as described by Wahl *et al.* (1979) and used previously (Lee *et al.*, 1983).

DNA sequencing

Appropriate DNA fragments were subcloned in M13 mp8 or mp9, propagated in the host JM101, and DNA was purified from phages as suggested by Amersham. DNA sequencing was performed by the dideoxy termination method described by Sanger *et al.* (1977). Sequencing reaction products were electrophoresed on 6% 0.3 mm polyacrylamide-urea gels. Sequences were obtained in both orientations. Comparison of sequences was made using the NucAln computer program of Wilbur and Lipman (1983).

Acknowledgements

This study was supported by the research grants from NSERC, Ottawa and FCAC, Quebec. We wish to thank Drs. J.Slightom and M.Murray for providing the *Phaseolus* genomic library, and Drs. G.Brown and E.Olson for their critical comments on this manuscript. The secretarial assistance of Miss Y. Mark is highly appreciated.

References

Appleby,C.A., Tjepkema,J.D. and Trinick,M.J. (1983) *Science (Wash.)*, **220**, 951-953.

Auger,S. , Baulcombe,D. and Verma,D.P.S. (1979) *Biochim. Biophys. Acta*, **563**, 496-507.

Blake,C. (1983) *Nature*, **306**, 535-537.

Blattner,F.R., Williams,B.G., Blechl,A.E., Denniston-Thompson,K., Faber, H.E., Furlong,L.-A., Grunwald,D.J., Kiefer,D.O., Moore,D.D., Shumm, J.W., Sheldon,E.L. and Smithies,O. (1977) *Science (Wash.)*, **196**, 161-169.

Brisson,N. and Verma,D.P.S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4055-4059.

Brisson,N., Pombo-Gentile,A. and Verma,D.P.S. (1982) *Can. J. Biochem.*, **60**, 272-278.

Brown,G.G., Lee,J.S., Brisson,N. and Verma,D.P.S. (1984) *J. Mol. Evol.*, in press.

Efstratiadis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connell,C., Spritz,R.A., DeRiel,J.K., Forget,B.G., Wiessman,S.M., Slightom,J.L., Blechl,A.E., Smithies,O., Baralle,F.E., Shoulders,C.C. and Proudfoot, N.J. (1980) *Cell*, **21**, 653-668.

Evans,A.M. (1980) in Summerfield,R.J. and Bunting,A.H. (eds.), *Advances in Legume Science*, Royal Botanic Gardens, Kew, pp. 337-347.

Fuller,F., Künstner,P.W., Nguyen,T. and Verma,D.P.S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2594-2598.

Hadley,H.H. and Hymowitz,T. (1983) in Caldwell,B.E. (ed.), *Soybeans: Improvement, Production and Uses*, Madison, Wisconsin: American Society of Agronomy, pp. 97-114.

Hosbach,H.A., Wyler,T. and Weber,R. (1982) *Cell*, **32**, 45-53.

Hunt,L.T., Hurst-Calderone,S. and Dayhoff,M.O. (1978) in Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington, D.C., pp. 229-251.

Hyldig-Nielsen,J.J., Jensen,E.O., Paludan,K., Wiborg,O., Garrett,R., Jorgensen,O.P. and Marcker,K.A. (1982) *Nucleic Acids Res.*, **10**, 689-701.

Hymowitz,T. and Newell,C.A. (1980) in Summerfield,R.J. and Bunting, A.H. (eds.), *Advances in Legume Science*, Royal Botanic Gardens, Kew, pp. 251-264.

Jeffreys,A.J. (1982) in Dover,G.A. and Flavell,R.B. (eds.), *Genome Evolution*, Academic Press, London, pp. 157-176.

Jeffreys,A.J., Wilson,V., Wood,D. and Simons,J.P. (1980) *Cell*, **21**, 555-564.

Lee,J.S. (1984) Ph.D. Thesis, McGill University, Montreal, Canada.
Lee,J.S. and Verma,D.P.S. (1984) in Setlow,J.K. and Hollaender,A. (eds.), *Genetic Engineering: Principles and Methods,* Vol. **6**, Plenum Press, in press.
Lee,J.S., Brown,G.G. and Verma,D.P.S. (1983) *Nucleic Acids Res.,* **11**, 5541-5553.
Lehtovaara,P. and Ellfolk,N. (1975a) *Acta Chem. Scand.,* **B29**, 56-60.
Lehtovaara,P. and Ellfolk,N. (1975b) *Eur. J. Biochem.,* **54**, 577-584.
Leipoldt,M. and Schmidtke,J. (1982) in Dover,G.A. and Flavell,R.B. (eds.), *Genome Evolution,* Academic Press, London, pp. 219-236.
Lewin,R. (1981) *Science (Wash.),* **214**, 426-429.
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,D.K. and Efstratiadis,A. (1978) *Cell,* **15**, 687-701.
Ohno,S. (1970) *Evolution by Gene Duplication,* published by Springer, Berlin, Heidelberg, NY.
Sanger,F., Nicklen,S. and Coulsor,A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5463-5467.
Schuler,M.A., Doyle,J. and Beachy,R.N. (1983) *Plant Mol. Biol.,* **2**, 119-127.
Southern,E.M. (1975) *J. Mol. Biol.,* **98**, 503-517.
Sullivan,D., Brisson,N., Goodchild,B., Verma,D.P.S. and Thomas,D.Y. (1981) *Nature,* **289**, 516-518.
Sun,S.M., Slightom,J.L. and Hall,T.C. (1981) *Nature,* **289**, 37-41.
Vainshtein,B.K., Harutyunyan,E.H., Kuranova,I.P., Borisov,V.V., Sosfenov,N.I., Pavovsky,A.G., Grebenko,A.I. and Konareva,N. (1975) *Nature,* **254**, 163-164.
Vanin,E.F., Henthorn,P.S., Kioussis,D., Grosveld,F. and Smithies,O. (1983) *Cell,* **35**, 701-709.
Varsanyi-Breiner,A., Gusella,J.F., Keys,C., Housman,D.E., Sullivan,D., Brisson,N. and Verma,D.P.S. (1979) *Gene,* **7**, 317-334.
Verma,D.P.S. (1982) in Smith,H. and Grierson,D. (eds.), *The Molecular Biology of Plant Development,* Blackwell Pub. Co., pp. 437-466.
Verma,D.P.S. and Bal,A.K. (1976) *Proc. Natl. Acad. Sci. USA,* **73**, 3843-3847.
Verma,D.P.S., Nash,D.T. and Schulman,H.M. (1974) *Nature,* **251**, 74-77.
Verma,D.P.S., Ball,S., Guérin,C. and Wanamaker,L. (1979) *Biochemistry (Wash.),* **18**, 476-483.
Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA,* **76**, 3683-3687.
Walbot,V. and Goldberg,R. (1978) in Davis,J. and Hall,T. (eds.), *Nucleic Acids in Plants,* CRC Press, Boca Raton, FL, pp. 3-40.
Weller,P., Jeffreys,A.J., Wilson,V. and Blanchetot,A. (1984) *EMBO J.,* **3**, 439-446.
Wiborg,O., Hyldig-Nielsen,J.J., Jensen,E., Paludan,K. and Marcker,K.A. (1982) *Nucleic Acids Res.,* **10**, 3487-3494.
Wilbur,W.J. and Lipman,D.J. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 726-730.
Woo,S.L.C. (1979) *Methods Enzymol.,* **68**, 389-395.
Zimmer,E.A., Martin,S.L., Beverley,S.M., Kan,Y.W. and Wilson,A.C. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 2158-2162.

Received on 23 August 1984