Chromosomal Arrangement of Leghemoglobin Genes in Soybean and Kidney Bean

by

Jong Seob Lee

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Biology
McGill University
Montreal, Quebec
Canada

July 1984

To my wife Jung Sun

ABSTRACT

Leghemoglobin (Lb) genes are induced only following infection of the legume plant by Rhizobium. In soybean, there are four major leghemoglobins which are encoded by four separate genes. A chromosomal walk has been carried out to investigate the arrangement of these Lb genes on the soybean chromosome by screening AluI-HaeIII and EcoRI genomic libraries. A cluster of four different Lb genes was isolated from the libraries in a set of overlapping clones which together include 45 kilobases (kb) of contiguous DNA. These four Lb genes, including a pseudogene, are present in the same transcriptional orientation and are arranged in the order: 5'-Lba-Lbc_1-Lbψ_1-Lbc_3-3'. The intergenic regions average 2.5 kb. In adddion to the main locus, there are other Lb genes in three other loci which do not appear to be contiguous to this locus. The second locus contains another leghemoglobin (Lbc_2) gene and a second pseudo (truncated) gene. The two other loci only contain truncated sequences. A sequence which appears to be common to the 3' regions of all the Lb loci was found flanking the Lbc_3 gene. The 3' flanking region of the main locus also contains a sequence that appears to be expressed more abundantly in root tissue. Another sequence which is primarily expressed in root and leaf was found 5' to two Lb loci.

In order to investigate the possible mechanism by which four Lb loci were generated, the structure and chromosomal arrangement of leghemoglobin genes in kidney bean have been determined. The nucleotide sequence of a kidney bean Lb gene showed the same intron/exon arrangement as that of soybean Lb genes, indicating their close evolutionary relationship. The presence in the kidney bean genome of four leghemoglobin genes with two sequences, each specific to the 5' or 3' region of the soybean Lb loci

suggests that tandom duplication of a single primordial plant globin gene had occurred to generate a four leghemoglobin gene locus before Glycine and Phaseolus species diverged. A large deletion in one of the 2 four-gene loci in soybean resulted in the generation of the Lbc_2 locus containing two leghemoglobin genes. The truncated gene appears to have been generated in the genome of Glycine species, prior to the chromosome duplication by tetraploidization before the divergence of G. max and G. soja.

RÉSUMÉ

Les gènes de la leghémoglobine (Lb) sont induits seulement après l'infection de la légumineuse par le Rhizobium. Il y a quatre leghémoglobines majeures chez la fève soya qui sont codées par quatre gènes séparés. En parcourant le chromosome par hybridations successives dans des librairies génomiques AluI-HaeIII et EcoRI l'arrangement des genes de la leghémoglobine sur le chromosome de la fève soya a été analysé. Un groupe de quatre différents gènes de Lb ont été isolé des librairies dans un ensemble de clônes se chevauchant lesquels totalisent 45 kilobases (kb) d'ADN contigu. Ces quatre gènes Lb, incluant un pseudogène, sont présents dans la meme orientation transcriptionnelle et sont placés dans l'ordre: 5'-Lba-Lbc_1-Lb ψ_1-Lbc_3-3'. Les régions intergéniques sont en moyenne de 2.5 kb. En plus du locus principal, on trouve d'autres gènes de la Lb dans trois autres loci qui ne semblent pas liés à ce locus. Le deuxième locus contient un autre gène Lb (Lbc_2) et un second pseudogène (tronqué). Les deux autres loci ne contiennent que des séquences tronqueés. Une séquence qui semble être commune aux régions 3' de tous les loci Lb à été trouveé près du gène Lbc_3. La région avoisinante de l'extremité 3' du locus principal contient aussi une séquence qui semble être exprimeé plus fréquemment dans le tissu racinaire. Une autre séquence qui est exprimeé principalement dans les racines et les feuilles a été trouveé à l'extremité 5' de deux loci de la Lb.

Dans le but d'étudier les mécanismes possibles par lesquels les quatre loci de Lb ont été générés, la structure et l'arrangement chromosomiques des gènes de la Lb dans la fève rouge ont été déterminés. La séquence des nucléotides des gènes de la Lb de la fève rouge montre le même arrangement

intron/exon que les gènes de la Lb chez la fève soya, indiquant leur proche parenté évolutionaire. La présence dans le génome de la fève rouge de quatre gènes de leghémoglobine en plus des deux séquences, chacune spécifique aux régions 5' ou 3' du locus de la Lb de la fève soya, suggère qu'il y a eu une duplication en tandem d'un seul gène primitif de globine végétale avant que les espèces Glycine et Phaseolus ne divergent. Une large délétion dans un des deux loci de quatre gènes dans la fève soya a généré le locus Lbc_2 contenant deux gènes de la leghémoglobine. Le gène tronqué semble avoir été généré dans le génome de l'espèce Glycine avant la duplication des chromosomes par tétraploidisation survenue avant la divergence de G. max de G. soja.

Traduit par Diane Longtin

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr. Desh Pal S. Verma, for his support, guidance and discussions throughout my work in his laboratory. I wish also to thank the members of my supervisory commitee, Drs. Yutaka Nishioka and Gorden Shore, for their guidance and discussions in the course of my study at McGill university.

My thanks are extended to the members of the laboratory, particularly Forrest Fuller, Norman Brisson, Eric Olson, Michael Sadowsky, Claude Lalande, Vincent Mauro, Shri Purohit, Indra Verma, Truyen Nguyen, Yvette Mark, Diane Longtin, Rose-Marie Riha, Victoria Foster and Joel Delisle for their various help.

I also thank Drs. Robert Goldberg of the University of California, Los Angeles and Jerry Slightom of the Agrigenetic Inc., Madison, Wisconsin for providing soybean or kidney bean genomic libraries and Mr. Guy L'Heureux and Robert Lamache for their photographic work.

TABLE OF CONTENTS

Table of contents (cont'd) Page

Table of contents (cont'd) Page

Table of contents (cont'd) page

LIST OF FIGURES

List of figures (cont'd) page

List of figures (cont'd) page

LIST OF TABLE

ABBREVIATIONS

bp	Base pair
c	copies
cDNA	Complementary deoxyribonucleic acid
Ci	Curie
cm, mm	Centimeter, millimeter
cv.	Cultivar
^{o}C	Degrees Celcius
d(T)	Deoxythymidilate
EDTA	Ethylenediaminetetraacetate
g, mg, ug, ng	Gram, milligram, microgram, nonogram, respectively
hr	Hour
IPTG	Isopropyl- β-D-thio-galactopyranoside
kb	Kilobase
l, ml, ul	Liter, milliliter, microliter, respectively
Lb	Leghemoglobin
M, mM, uM	Molar, millimolar, micromolar, respectively
min	Minute
mRNA	Messenger ribonucleic acid
OD_{600}	Optical density at 600 nanometers
PEG	Polyethylene glycol
Pv	*Phaseolus vulgaris*
poly(A)$^{+}$RNA	Polyadenylated RNA

Abbreviations (cont'd)

RNAase	Ribonuclease
rpm	Revolution per minute
SDS	Sodium lauryl sulfate
SSC	Standard saline citrate
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl) aminomethane
x-gal	5-bromo-4-chloro-3-indolyl- β -galactoside

I. INTRODUCTION AND LITERATURE REVIEW

Biological nitrogen fixation, the enzymatic reduction of dinitrogen (N_2) to ammonia, plays a crucial role in maintaining nitrogen resources in the biosphere. While only certain procaryotic organisms have evolved the enzymatic capacity to reduce dinitrogen directly, very few plants and no animals can assimilate atmospheric dinitrogen. Quite distinctive in this respect are certain angiosperms, which have aquired the ability to associate with specific microorganisms (Rhizobium species and actinomycetes). This symbiotic association between plants and microorganisms results in the development of specialized tissues called nodules, where free gaseous nitrogen is converted into fixed nitrogen for eventual assimilation and storage into plant proteins. This plant-microorganism collaboration is so intimate that one may conceive of the association as a new form of life. The complexities of the nodule tissue, and the high degree of specialization for various enzymatic processes and their regulations, indicate a long-standing co-evolution between the two partners. Although certain non-leguminous angiosperms are known to fix nitrogen by the symbiotic association with actinomycetes (Bond, 1967; 1974; Burns and Hardy, 1975), more attention has been paid to the symbiotic association between legumes and Rhizobium species because of their agricultural importance.

The process of symbiotic nitrogen fixation and the steps leading to it are highly complex and are influenced by genetic factors of both the host and the endosymbiont. While the mechanism and biochemistry of nitrogen fixation is relatively well understood, little is known about bacterial and plant gene products involved in the development and maintenance of the

symbiosis. In this review, I have concentrated upon the general molecular biology of the legume-Rhizobium association with an emphasis on the contribution and the involvement of the host genes in this novel phenomenon in nature. The molecular genetics of nitrogen fixation has been reviewed recently (Brill, 1980; Beringer et al., 1980; Kondorosi and Jonston, 1981; Postgate, 1982).

A. Development of the Root Nodule

Successful interaction of a rhizobial strain with its host results in the development of an organized structure, the root nodule, in which Rhizobium lives endosymbiotically within a group of infected host cells. This involves several sequences of events, during which the host plant not only has to provide the total energy for the process of nitrogen fixation and the survival of Rhizobium as well as a mechanism of protecting the nitrogenase system, but also must suppress any pathogenic responses against the invader.

(1). Invasion of root hair cells

The first step in the development of root nodule symbiosis is mutual recognition. Little is known of the nature and functioning of recognition mechanisms in either pathogenic or symbiotic plant/microorganism interactions. The most promising current hypothesis is that host plant lectins (carbohydrate binding-proteins) interact selectively with microbial cell surface carbohydrates and serve as determinants of recognition or host specificity (Bohlool and Schmidt, 1974; Dazzo and

Hubbell, 1975). The outcome of interactions between plants and microorganisms is normally determined by the recognition capacities of the plants. The recognition mechanism senses or detects the presence of the microorganism. In the case of symbiotic associations this involves the suppression of the host defence mechanism(s). Verma and Nadler (1984) suggested that this could be accomplished in two ways: 1) by suppressing the plant response(s) to invasion or 2) the removal from Rhizobium of the appropriate determinants which trigger the plant defence mechanism. It is possible that Rhizobium is able to overcome the host defenses, but cannot avoid recognition and the constitutive defenses so that the host restricts the spread of this infection by compartmentalization and tissue specialization (root nodule formation).

An early structural response of the host to the attachment of Rhizobium is curling of the root hairs (Nutman, 1959; Dart, 1974). This phenomenon is interpreted as the inhibition of the cell growth at the attachment site (Bauer, 1981). The curled root hair forms a 'pocket' entrapping the adhering Rhizobium. The formation of a pocket enclosing bacteria may serve to increase the concentration of nutrients for the microsymbiont as well as the concentration of signals and effectors from the microsymbiont.

The Rhizobium then penetrates the root hair and moves towards the cortical tissue of the root via a tubular structure known as the infection thread formed by the plant (Nutman, 1956). Root hair invasion and infection thread formation which may require a localized hydrolysis of plant pectin are postulated to involve bacterial pectinase and host cellulase enzyme activities (Ljunggren, 1969; Verma et al., 1978a; Dazzo and Hubbell, 1981). The composition of the infection thread wall has been

shown to be similar to the primary plant cell wall (Dart, 1974; Newcomb, 1976). The diameter and content of infection threads vary: in Glycine max, they appear to contain a single file of dividing bacterial cells and little or no matrix material (Goodchild and Bergersen, 1966). In Pisum sativum, the threads are wider and contain masses of bacteria embedded in a finely-granular matrix (Newcomb, 1976). When the infection threads penetrate the cortex cells, they may branch both within and between plant cells and more than one thread penetrate a single cell (Dart and Mercer, 1964).

(2). Infection of cortical cells

Successful infection results from release of bacteria from the infection thread into the host cell. Here they remain bounded by a membrane vesicle derived from the host plasma membrane due to endocytosis at the time of their release (Goodchild and Bergersen, 1966; Dixon, 1967; Newcomb, 1976). After the infection threads enter the cells in the meristematic area, they develop protrusion-like terminal pockets (Goodchild and Bergersen, 1966). Individual bacteria become attached to these pockets in the infection thread membrane, and are budded-off into individual vesicles referred to as peribacteroid membranes (Bergersen and Briggs, 1958; Robertson et al., 1978). Once inside the host cell, the free-living Rhizobium is transformed into a bacteroidal form, which is characterized by a number of structural and metabolic changes (Sutton et al., 1981).

The cortical cells in which bacteria have been released show a number of morphological, cytological and physiological changes. Newly infected cells are characterized by large nuclei with up to three prominant nucleoli, and by a relative dense cytoplasm with numerous free ribosomes (Newcomb et al., 1979). The area of rough endoplasmic reticulum increases. Golgi bodies, endoplasmic reticulum and peribacteroid membranes are observed fusing with small vesicles (Robertson et al., 1978). During the late stages of bacteroid multiplication, the infected cells become greatly enlarged (Bergersen and Goodchild, 1973). Their nuclei also enlarge and become characteristically amaeboid. The nuclei increase several fold in volume and are often polyploid due to endoreduplication which seems to be caused by Rhizobium-produced cytokinin (Libbenga et al., 1973; Libbenga and Torrey, 1973). A single central vacuole may be formed by coalescence of several smaller vacuoles (Kijne, 1975). Mitochondria and amyloplasts become localized at the cell periphery, adjacent to the intercellular spaces. As many as 2.4×10^4 to 3.6×10^4 bacteria are present per infected cells in soybean (Greshoff and Rolfe, 1978).

Up to 50% of the nodule cells remain uninfected. These uninfected cells are referred to as interstitial cells. The DNA content in these cells remains relatively low (2 to 4 copies) (Truchet, 1978) as compared to the infected cells [(4c, 8c or even 16c levels of DNA (Libbenga and Bogers, 1974)]. Ultrastructural studies indicated that some uninfected cells may be specialized in metabolic pathways related to the assimilation of nitrogen (Gunning et al., 1974; Newcomb and Tandan, 1981). The size and shape of effective nodules is controlled by the plant host rather than the Rhizobium (Kidby and Goodchild, 1966).

B. Nodule Maturation

There are undoubtedly many regulatory factors that operate directly or indirectly to influence nodule maturation (Vincent, 1980). Development may be slowed down or terminated by environmental changes or deficiences in essential nutrients. Numerous Rhizobium mutations causing ineffectiveness (non-nitrogen fixing nodules) have been shown to operate at the nodule maturation stage. Variations in plant genotypes may cause similar effects. Therefore, it is clear that the legume and Rhizobium cells must exchange correct signals at each stage of nodule development.

(1). Plant factors

The plant genome plays an important role in the development and the effectiveness of the nodules in nitrogen fixation. The plant determines the size, shape, distribution and morphology of the nodule including the intracellular organization of infected host cells (Vincent, 1974; Nutman, 1981; Kidby and Goodchild, 1966). Host genes may control the selectivity among Rhizobium species and strains and the mode of infection (Dart, 1977). Bacterial differentiation and morphology are greatly influenced by the host (Dart, 1977; Sutton et al., 1982).

So far, a few genes which affect normal development of nodules have been identified genetically. Single plant genes, transmitted as Mendelian recessives, cause Nod^- phenotype in clover (Nutman, 1954), soybean (Weber, 1966) and pea (Lie, 1971). Several genes for ineffectiveness which can be reversed by modifying plant genes or bacterial mutations have been identified in Trifolium (Gibson, 1964),

Medicago (Gibson, 1962; Viands et al., 1979), Pisum (Holl, 1973, 1975; Lie and Timmermans, 1979), and Glycine max (Vest, 1970; Vest and Caldwell, 1973). Three dominant genes conferring ineffectiveness have been identified in soybean (Caldwell and Vest, 1977).

(2). Bacterial factors

The intracellular location of bacteroids give them considerable potential to influence the host cells by depleting or excreting metabolites or controlling factors. Substances likely to be produced in addition to ammonia in bacteroids and excreted into the host cytoplasm include phytohormones, heme, extracellular polysaccharides and CO_2.

Legume nodules contain high concentrations of all three groups of plant growth-promoting hormones. The major growth promoter is IAA (Pate, 1958). Rhizobium cultures are able to convert exogeneous L-tryptophan into IAA (auxins) (Dullaart, 1970). It was suggested that rhizobial auxins may stimulate nodule growth (Thimann, 1936) or induce a localized re-orientation of plant cell wall synthesis to form infection threads (Kefford et al., 1960). In addition to IAA, Rhizobium may produce other substances. It was suggested that rhizobial cytokinins may trigger endoreduplication and mitosis in the infected root cortex (Libbenga and Torrey, 1973) and that nodule gibberellins may function as inhibitors of further nodulation (Radley, 1961).

The heme component of leghemoglobin is shown to be synthesized in the bacteroids. The rate-limiting enzyme for heme synthesis δ-aminolevulinic acid synthetase, was confined to the bacteroid fraction from soybean nodules (Nadler and Avissar, 1977).

The role of Rhizobium bacteroids in mature nodules is very specialized, particularly in relation to nitrogen metabolism. Rather than conserving their combined nitrogen for growth, bacteroid suspensions excrete it as ammonium ions (Bergersen and Turner, 1967). The enzyme capable of catalyzing the reduction of atmospheric nitrogen to ammonia is the nitrogenase enzyme which consists of two protein components (Ljones, 1974). The larger subunit, MoFe protein (component I) binds the reducible substrate; the smaller one, Fe protein (component II), transfers electrons to the MoFe protein. The enzyme requires ATP and Mg^{++} for activity (Burns, 1977). Nitrogenase generally does not function in free-living Rhizobium. Since this enzyme is highly oxygen-sensitive, all diazotrophs have evolved systems to protect the enzyme from oxygen damage. Some organisms only fix dinitrogen under anaerobic conditions, while others have developed ways of protecting the enzyme while fixing in a generally aerobic environment.

C. Assimilation of Symbiotically Fixed Nitrogen

Ammonia is the first stable product of nitrogen fixation. The mechanisms of assimilation utilized by the bacteroid and the host plant are of great importance in maintaining nitrogen flow without affecting reduction of nitrogen. Here, however, the mechanisms utilized by the host plant will be described.

Ammonia is transported into the plant cytoplasm, where it is assimilated as amides or ureides (Miflin and Lea, 1976). The discovery by Tempest et al. (1970) of glutamate synthase (GOGAT) in Enterobactor aerogenes revealed a route involving glutamine synthatase (GS) and

glutamate synthase which is capable of utilizing NH_4^+ in the cytosol fraction of nodules from many legumes (Miflin and Lea, 1976; Boland et al., 1980). Strong positive correlation between GS activity of nodule cytosol from developing nodules and either nitrogenase activity or nodule leghemoglobin content in lupin (Robertson et al., 1975) provides convincing evidence that cytosol GS is of major importance in the assimilation of the product of nitrogen fixation in this species.

Following the incorporation of newly fixed nitrogen into glutamine via the GS-GOGAT pathway of the nodule cytosol, nitrogen may be transported via the xylem to other parts of the plant where it is utilized. However, although glutamine may comprise a substantial fraction of the transported nitrogen, it rarely constitutes the major nitrogenous solute of xylem in legumes fixing nitrogen. Secondary reactions, involving transfer of the amide- or amino-nitrogen of glutamine to other products must therefore comprise major metabolic routings within the nodule. It has been reported that many 'temperate' legumes, including Pisum, Lupinus and Medicago export nitrogen from the nodule mainly as asparagine (Pate et al., 1969; Streeter, 1972; Scott et al., 1976; Atkins et al., 1978). Other legumes, those designated as 'tropical', for example Glycine, Phaseolus and Vigna appear to change from transporting amino acids to the export of ureides (mainly allantoin and allantoic acid) (Atkins et al., 1978; Matsumoto et al., 1978; Herridge et al., 1978; Schubert, 1981). Some of the enzymes involved in these specialized pathways may be encoded by host genes (see below).

(1). Asparagine synthesis

Several radioactive labelling studies in serradella (Kennedy et al., 1966) and Vicia (Lawrie and Wheelerg, 1975) nodules showed that aspartate was quickly labelled later appearing in asparagine, only after glutamate. In soybean, which transports mainly ureides, poor labelling of asparagine occurred following feeding labelled nitrogen to nodules (Ohyama and Kumazawa, 1980). These results indicated the presence of an active glutamine dependent asparagine synthesis system in these nodules.

The pathway of asparagine synthesis in higher plants is considered to involve two enzymes, aspartate aminotransferase and asparagine synthetase (Lea and Miflin, 1980). In lupin, both enzyme activities increase in parallel with the activities of glutamine synthetase, glutamate synthase and nitrogenase (Reynolds and Farnden, 1979; Scott et al., 1976). In soybean, however, both enzyme activities also increased during the initial stages of nodulation but, as ureide biosynthesis proceeded their activities declined (Reynolds et al., 1982).

(2). Ureide synthesis

It is generally believed, that ureide synthesis in plants not capable of fixing nitrogen proceeds through oxidation of purine nucleotides rather than by condensation of urea and a two-carbon compound such as glyoxylate (Reinbothe and Mothes, 1962). The presence in nodules of ureide-producing legumes of significant activities of xanthine dehydrogenase, xanthine oxidase, uric acid oxidase and allantoinase has been taken as evidence

that a similar purine-based pathway of ureide synthesis might operate in the nitrogen-fixing tissue of nodules.

Tajima and Yamamoto (1975) reported high activities of xanthine oxidase, uricase and allantoinase in the root nodules of soybean plants providing evidence of purine catabolism. Matsumoto et al (1977) reported that nodulated soybean plants accumulated high concentrations of allantoic acid and allantoin in the soluble nitrogen fraction of plants. Fugihara and Yamaguchi (1978) have used an inhibitor, allopurinol, to prevent the formation of allantoic acid and urea in seedlings and nodules of soybean. This resulted in the accumulation of xanthine but not hypoxanthine. Triplett et al. (198Ø) have proposed a pathway of ureide synthesis involving the degradation of inosine-5-monophosphate and some related purines including inosine, xanthine-5-monophosphate, xanthosine, hypoxanthine and xanthine. Increasing evidences from labelling and inhibitor studies (Matsumoto et al., 1977; Fujihara and Yamaguchi, 1978; Ohyama and Kumazuwa, 1978; Boland and Schubert, 1982; Atkins et al., 198Ø) support the purine degradative pathway as the primary route of ureide biosynthesis. Localization of enzymes involved in this pathway (Tajima and Yamamoto, 1975; Triplett et al., 198Ø) provides evidence that plant cells in the nodule are the sites of ureide biosynthesis. The peroxisomal enzymes, uricase and catalase, also allantoinase and a number of enzymes involved in the synthesis of purine precursors were found to be present in the uninfected cells of soybean nodules at much higher specific activities (Hanks et al., 1981; 1983). This may suggest that the whole of purine synthesis and catabolism pathway occur predominantly in the uninfected cells of the nodule using amino acid transported from the nitrogen-fixing infected cells.

D. Plant Genes Expressed in the Root Nodule

A number of host genes have been implicated to be involved in the development of a successful association between a legume plant and its nodulating Rhizobium species (Caldwell and Vest, 1977; Nutman, 1981). They have an effect on a number of events in this symbiosis which determine the ability to nodulate, the abundance of nodules, the intracellular organization and the differentiation of bacteroids (Verma, 1982; Verma and Long, 1983).

In an effort to estimate the approximate number of structural genes active in nodules and to examine for the presence of any nodule-specific host sequences, the complexity of total cytoplasmic poly(A)$^+$ RNA of the nodule tissue was compared with that of the uninfected root (Auger and Verma, 1981). The results indicated a substantial sequence homology; however, in addition to leghemoglobin, which is a major component in nodules, a significant increase in the concentration of middle-abundant mRNA was observed. Also, immunological techniques were used to detect the presence of several nodule-specific plant proteins (Legocki and Verma, 1980). The proteins of host origin which are present in nodules but are undetected in the uninfected root were named nodulins (Legocki and Verma, 1980).

(1). Nodulins

Two major groups of host gene products, leghemoglobins and nodulins are induced specifically during symbiotic nitrogen fixation. Nodulins may exist as at least three types: (i) proteins responsible for the

maintenance of nodule structure, (ii) enzymes necessary for the specific assimilation of reduced nitrogen, and (iii) proteins that support bacteroid function and thus facilitate nitrogen reduction (Fuller et al., 1983). One "superabundant" protein of the third type is leghemoglobin, which functions in facilitating oxygen diffusion for bacteroid respiration. Other nodule-specific polypeptides were identified in nodules of several legumes (Legocki and Verma, 1979, 1980; Bisseling et al., 1983). Nodule-specific uricase and glutamine synthetase are proteins of the second type (Legocki and Verma, 1979; Bergmann et al., 1983; Cullimore et al., 1983).

(a). Uricase

Uricase is one of the first purified nodulins in legumes (Bergmann et al., 1983). Uricase has been shown to increase in soybeans upon nodulation (Tajima and Yamamoto, 1975). In tropical legumes such as soybean and cowpea, the major part of fixed nitrogen is incorporated into allantoin and allantoic acid for storage and transport to other parts of the plant (Schubert, 1981) and for both species a new pathway of ureide synthesis has been proposed which involves the uricase enzyme (Schubert, 1981; Shelp et al., 1983).

Analysis of the total soluble proteins from nodules effective in nitrogen fixation revealed a peptide of 35,000 MW termed nodulin-35 (Legocki and Verma, 1979). Purification of this protein in native form and determination of some of its chemical and physical characteristics indicate that the nodulin-35 is a subunit of specific uricase which is induced in the soybean nodules and is shown to be located in uninfected

cells of these tissues (Bergmann et al., 1983). This is consistent with the idea (Hanks et al., 1983) that uninfected cells play an important role in ureide synthesis. The uricase functions in oxidizing uric acid, a product of purine catabolism, to allontoin (Tajima and Yamamoto, 1975). Absence of any cross-reacting molecules in root tissue with the antibody against uricase indicates that the nodulin-35 is encoded by a unique plant gene which is induced in the nodule tissue (Bergmann et al., 1983).

(b). Glutamine synthetase

Glutamine synthetase has been purified from root nodules of Glycine (McParland et al., 1976), Lupinus (McCormack et al., 1982) and Phaseolus (Cullimore et al., 1983). In nodules of Phaseolus two forms of glutamine synthetase (GS1 and GS2) are present and they differ from both leaf forms of the enzyme in this species (Cullimore et al., 1983). While GSn2 (n: nodule-specific) [identical to the root enzyme (GSr)] makes up only about 12% of the activity in nodules, GSn1 accounts for over 85% of total nodule GS activity (Lara et al., 1983), almost all of it being in the plant cytosol (Awonaike et al., 1981). It is not yet certain whether the nodule-specific glutamine synthetase is a product of a separate gene or represents modification of the root enzyme.

(c). Other nodulins

Analysis of RNA:cDNA hybridization kinetics revealed the presence of a moderately abundant nodule-specific RNA class (Auger et al., 1979; Auger and Verma, 1981) which has been postulated to encode nodulins. Recently

four nodulin coding sequences were isolated from a cDNA library prepared from soybean nodule poly(A)$^+$ RNA (Fuller et al., 1983). They encode polypeptides of Mr 44,000, 27,000, 24,000 and 100,000 - 120,000, respectively (Fuller et al., 1983). RNA dot-blot hybridizations (Fuller and Verma, 1984) indicated that these nodulin sequences began to appear and accumulate in Rhizobium infected root tissues as early (day 3 to 5 after infection) as leghemoglobin sequences and reached fully induced levels by day 11 (Fuller and Verma, 1984). Different ineffective Rhizobium strains had a pleiotropic affect on the accumulation of each sequence.

Immunological studies revealed the presence of about 30 nodulins detectable during nodule development of pea plants (Bisseling et al., 1983). Some of them were preferentially synthesized before nitrogen fixation started, whereas the majority were synthesized concomitantly with leghemoglobin (Bisseling et al., 1983). This suggests that there may be proteins which reflect the formation of nodule tissue while other nodulins may be essential for nitrogen fixation.

The studies on nitrogen assimilation revealed the involvement of a number of enzymes which either increase substantially in activity during nodulation and/or are produced as different isoforms in nodules compared to roots. They are a subunit of glutamine synthetase (GSn1) in Phaseolus (Cullimore, 1983), aspartate aminotransferase (AAT-P2) (Reynolds and Farnden, 1979; Boland et al., 1982), uricase (Tajima and Yamamoto, 1975; Christensen and Jochimsen, 1983; Bergmann et al., 1983), xanthine dehydrogenase (Triplett et al., 1982; Christensen and Jochimsen, 1983), asparagine synthetase (Scott et al., 1976), phosphoenolpyruvate carboxylase (Christeller et al., 1977) and xanthine oxidase (Tajima and Yamamoto, 1975). They have been identified as nodule-specific and thus are

good candidates for nodulins. Therefore, further studies on nodulins will provide useful information on the number of nodule-specific host proteins involved in symbiosis and the relationship between their expression during nodule development. Also, they will facilitate our understanding about the nature of their regulatory phenomena.

(2). Leghemoglobin

Kubo (1939) showed for the first time that the red pigment of legumincus root nodules is a hemoprotein similar to hemoglobin. This was confirmed by Kelin and Wang (1945) by showing that it was capable of completely reversible oxygenation and deoxygenation. This hemoprotein, leghemoglobin, is similar to myoglobin of the muscle tissue of vertebrates and is present only in the root nodules of leguminous plants fixing nitrogen. Recent findings of hemoglobins in the nodules of non-leguminous nitrogen-fixing plants (Appleby et al., 1983) suggest that they may play an important role in the nitrogen fixation

(a). Location

Since Smith (1949) reported that leghemoglobin was contained only in the large bacteroid-containing cells of nodules, there has been a contraversy over its intracellular location (see review by Bergersen, 1980). Using electron microscope autoradiography of serradella nodules labelled with ^{59}Fe, Dilworth and Kidby (1968) localized leghemoglobin between the bacteroid and the membrane envelope surrounding the bacteroids. Bergersen and Goodchild (1973) obtained similar results using

a diaminobenzidine staining method. These studies were based on the indirect methods of localization of leghemoglobin as heme or iron. However, using ferritin-conjugated antibodies to leghemoglobin, Verma and Bal (1976) showed that leghemoglobin is restricted to the cytoplasm of the infected cells of the nodules. This was further confirmed by isolating the membrane envelopes enclosing the bacteroids and demonstrating that no antibody reactive material is present in these vehicles (Verma et al., 1978b). Since leghemoglobin is not synthesized as a large precursor (Verma et al., 1979), it suggests that this protein does not cross the membrane envelope enclosing the bacteroids and remains in the host-cell cytoplasm after translation. However, the possibility remains that the minor components that are modified post-translationally may be present in the peribacteroid space (see Bergersen and Appleby, 1981)

(b). Structure

Leghemoglobin comprises a prosthetic group, protoporphyrin, and a protein component, globin (Ellfolk, 1972). The former, an ion porphyrin, consists of four pyrrole rings in a plane around an iron atom, which is coupled to the four nitrogen atoms. The structure of leghemoglobin is similar to that of myoglobin (Vainshtein et al., 1975). Leghemoglobin is a monomeric hemoprotein while hemoglobins found in the non-leguminous plants are dimeric (Appleby et al., 1983).

Leghemoglobin is almost invariably heterogenous. In soybean, there are four major components, Lba, Lbc_1, Lbc_2 and Lbc_3 (Fuchsman and Appleby, 1979), each of which is post-translationally modified into minor components, Lbb, Lbd_1, Lbd_2 and Lbd_3, respectively (Whittaker et

al., 1981). This modification involves N-terminal acetylation in the case of the three types of Lbc, whereas the N-terminal residue of Lba is removed enzymatically before acetylation occurs (Whittaker et al., 1981).

Other legumes also contain multiple components of leghemoglobin. Lupin nodules have been found to contain four distinct leghemoglobins and serradella nodules contain three components (Dilworth, 1969). While snakebean contains from two to six components (Broughton and Dilworth, 1971), kidney bean contains two components, one of which is the major component and the other is the product of post-translational modification of the main component (Lehtovaaro and Ellfolk, 1975a). A recent analysis of the leghemoglobins from alfalfa indicated the presence of five components, Lbl to Lb5 (Jing et al., 1982). The multiplicity and variability of leghemoglobin in various legumes may suggest that they have some functional importance (Fuchsman et al., 1976).

Determination of the complete amino acid sequence has been reported for leghemoglobins from five different plants. They are Lupinus luteus LbI (Jegorov et al., 1976), L. luteus LbII (Jegorov et al., 1978), Vicia faba LbI (Richardson et al., 1975), Glycine max Lba (Ellfolk and Sievers, 1971), G. max Lbc (Sievers et al., 1978), Phaseolus vulgaris Lba (Lehtovaara and Ellfolk, 1975b), and Pisum sativum LbI (Lehtovaara et al., 1980). The alignment of the amino acid sequences of these leghemoglobins using Lupinus leghemoglobin as reference (Lehtovaara et al., 1980) showed that 50 residues which represent one third of the molecule are common to all seven leghemoglobins. As shown in Table I, soybean leghemoglobins are closely related to Phaseolus leghemoglobin a with about 30 amino acid residues variant. In addition, Pisum leghemoglobin I and Vicia leghemoglobin I both share extensive homology, again with about 30 amino

Table I. Difference Matrix for Leghemoglobins[a]

The first sum indicates the number of differences between each two leghemoglobin sequences alighned. The second sum indicates the number of polymorphic residues where both identical and differing variants are found when the two sequences are alighned.

		(b)	(c)	(d)	(e)	(f)	(g)
Pisum LbI (147)	(a)	33+15	57+ 8	54+10	52+ 6	69+ 4	70+ 3
Vicia LbI (143)	(b)	0	53+ 8	51+ 9	53+ 7	75+ 7	75+ 7
Glysine Lba (142)	(c)		0	8+ 6	31+ 1	73+ 1	72+ 1
Glycine Lbc (143)	(d)			0	26+ 4	71+ 3	68+ 3
Phaseolus Lba (145)	(e)				0	75	75
Lupinus LbI (153)	(f)					0	20
Lupinus LbII (153)	(g)						0

[a] from Lehtovaar et al. (1980).

The number of amino acids coded by each leghemoglobin is shown in parentheses.

acids variant. Lupinus leghemoglobins, however, do not show close relationship to any leghemoglobin in other legumes. The high degree of conservation in their primary structure is reflected in their immunological cross-reactivity (Hurrell et al., 1977, 1979). These demonstrate their close evolutionary relationship. Assuming three-dimensional homology for leghemoglobins aligned, it seems that residues forming the heme pocket on the ligand-binding side have been best conserved during evolution, whereas mutations have been incorporated especially on some outer corners of the molecule, which are less critical for its function (Lehtovaara et al., 1980).

The high degree of similarities in the primary and tertiary structures of leghemoglobins and mammalian globins indicates that plant and animal globins have a common evolutionary origin. Alignment of kidney bean leghemoglobin a (PhLba) with different globin sequences showed that PhLba has 28 - 32 residues identical with mammalian hemoglobin beta, delta and gamma chains, 25 with Glycera dibranchiata hemoglobin, 19 - 24 with mammalian alpha chains and 15 - 20 residues identical with myoglobins (Lehtovaara and Ellfolk, 1975b). The molecular geometry of the hemoproteins changed remarkably little in passing from mammals through invertebrates to plants. Molecular conformations of globins of various evolutionary stages showed that globin chains are similar though not identical in size and helix content (Vainshtein et al., 1975; Hurrell et al., 1979).

(c). Functions

The average oxygen pressure in the cells of the central tissues of soybean nodules is estimated to be only about Ø.Ø1 mm Hg and is maintained at this level during vigorous oxygen consumption primarily by the bacteroids (Wittenberg et al., 1972). However, it is not clear how a large inflex of oxygen is sustained, while the oxygen pressure within the bacteroids is kept at a low level.

A correlation has been observed between the ability of nodules to fix nitrogen and the presence of leghemoglobins (Smith, 1949), suggesting that leghemoglobin plays a fundamental role in nitrogen fixation. The development of nitrogenase activity in free-living rhizobia, without induction of leghemoglobin (Pagan et al., 1975), rules out the possibility of direct involvement of leghemoglobin in the nitrogen fixation. Rather, the equibria and kinetics of the reaction of leghemoglobin with oxygen are consistent with the facilitation of oxygen diffusion in an environment of very low oxygen pressure (Wittenberg et al., 1972; Imamura et al., 1972). This suggests that leghemoglobin may facilitate oxygen diffusion to the bacteroids.

In order to define the molecular mechanisms by which leghemoglobin augments the oxygen consumption and nitrogenase activity of bacteroids, Wittenberg et al. (1974) investigated the effect of oxyleghemoglobin on oxygen consumption and nitrogenase activity. They observed that oxyleghemoglobin added to shaken suspensions of bacteroids strongly enhances both the rate of oxygen uptake and the activity of nitrogenase as measured by the rate of reduction of acetylene to ethylene. From these results they proposed that a function of leghemoglobin is to facilitate the diffusion of oxygen across the layer of solution adjacent to the

bacteroid surface and to deliver free (dissolved) oxygen to the bacteroid (Wittenberg et al., 1974).

Although the overall molecular structure of leghemoglobin is similar to that of animal myoglobins and hemoglobins, this molecule has a much higher affinity for oxygen. It releases oxygen for bacteroid respiration only at low partial pressure (Wittenberg et al., 1974; Appleby et al., 1975). By providing a sustained oxygen flux to bacteroids at very low $_pO_2$, it also protects the nitrogenase enzyme from excess oxygen.

(d). Biosynthesis

Leghemoglobin is synthesized only when the plant and Rhizobium species live symbiotically. The leghemoglobin apoprotein is encoded by plant genome (Baulcombe and Verma, 1978), whereas the bacterial partner synthesizes heme which can be inserted into apoleghemoglobin (Cutting and Schulman, 1969, Nadler and Avissar, 1977).

Heme synthesis has been proposed to be the function of Rhizobium. By incubating fractionated nodule extracts with -aminolevulinic acid, Cutting and Schulman (1969) found that the bacteroid fraction of nodule extracts can convert δ-aminolevulinic acid into heme. Avissar and Nadler (1978) observed that the rate of heme secretion from bacteria increased in vitro under low partial pressure of oxygen. Rhizobium meliloti mutant strains with reduced levels of δ-aminolevulinic acid synthase, the first enzyme of heme biosynthesis, form small white nodules which do not fix nitrogen (Leong et al., 1982).

Leghemoglobins are encoded by poly(A)-containing 9S RNA (Verma et al., 1974). This RNA is preferentially translated on free polysomes (Verma and Bal, 1976), which is consistent with the cellular location of leghemoglobin (Verma et al., 1978b). Leghemoglobin is synthesized six to seven days after Rhizobium infection and four days before the appearance of nitrogenase activity, suggesting that the inductions of leghemoglobin and nitrogenase are independent (Verma et al., 1979). Once leghemoglobin synthesis is initiated, it is influenced by the ineffectiveness of the nodules (Verma et al., 1981). Although the expression of leghemoglobin genes appears to be controlled at the transcriptional level by Rhizobium (Verma et al., 1974, 1981), its actual mechanism is yet to be investigated.

Analysis both in vivo and in vitro of relative rate of synthesis of two electrophoretically distinguishable components of leghemoglobin in soybean indicated that the rate of synthesis of the Lbc components is higher than that of the Lba in young nodules (12 days), but that Lba synthesized more rapidly in older nodules (21 days) (Verma et al., 1979). The separation and determination (by isoelectric focusing) of the Lb components from root nodules at the different stages of development (Fuchsman and Appleby, 1979), also indicate that the ratio of Lbc_3 to Lba content decreases dramatically as nodules mature. Therefore, as with animal globin genes (Efstratiadis et al., 1980), leghemoglobin gene switching may occur during development. However, it remains to be determined whether all leghemoglobin components have distinct biochemical and physiological roles in nitrogen fixation.

(e). Leghemoglobin genes

Leghemoglobin sequences have been suggested to be present in soybean genome in about 40 copies per haploid compliment (Baulcombe and Verma, 1978), whereas hybridization of EcoRI-digested genomic DNA of soybean with a nick-translated Lb cDNA clone revealed about seven EcoRI bands (Sullivan et al., 1981). These observations suggest the presence of a family of related genes in the soybean genome. These genes exist in soybean as functional, truncated or pseudo genes (Brisson and Verma, 1982).

Most of the leghemoglobin genes have been isolated from the genomic libraries and analyzed at the nucleotide level. They are the Lba (Hyldig-Nielsen et al., 1982), Lbc_1 (Hyldig-Nielsen et al., 1982), Lbc_2 (Wiborg et al., 1982), Lbc_3 (Brisson and Verma, 1982; Wiborg et al., 1982), Lb ψ_1 (Brisson and Verma, 1982; Wiborg et al., 1983) and LbT_1 (Brisson and Verma, 1982; Brisson, 1982) genes. All the functional leghemoglobin genes are interrupted by three intervening sequences at codons 32, 68-69 and 103-104. The three introns start with the sequence GT and terminate with AG, as is typical of other eucaryotic genes (Lerner et al., 1980). The region upstream from the initiation codon revealed the presence of two sequences, the TATAAA (Goldberg, 1979) and CCAAT (Corden et al., 1981) boxes, which are common to other eucaryotic genes and have been proposed to be involved in transcription. Also, the 3' non-coding region revealed a sequence, AATAAA, a consensus sequence for the poly(A) addition signal (Proudfoot and Brownlee, 1976). The presence of these sequences in the leghemoglobin genes suggests that mechanisms for gene transcription and RNA splicing in plants may be similar to these in other eucaryotes. Indeed, the region upstream from the TATAAA box of plant genes

revealed the presence of a possible regulatory sequence referred to as the AGGA box (Messing et al., 1983), which is also found in the leghemoglobin genes. Since the CCAAT box has been found in some plant genes but is not apparent in others, it seems likely that despite the similarities of animal and plant systems at the molecular level, the latter has its own characteristic features. However, since the induction of leghemoglobin genes occurs only after infection of the plant by Rhizobium, they may also contain other regulatory sequences which are responsible for the induction.

Intriguingly, leghemoglobin gene structure was found to be very similar to that of mammalian globin genes, both with respect to the positions of two introns common to all globin genes as well as to the presence of several regulatory sequences on the 5' region of these genes (Brisson and Verma, 1982; Brown et al., 1984). These suggest their close evolutionary relationship.

At the time this research project was initiated, a few leghemoglobin genes were isolated from a soybean genomic library (Sullivan et al., 1981) and their arrangement on the chromosome was unknown. Since the chromosomal arrangement of leghemoglobin genes is essential to understand the regulation of their expression and their evolutionary relationship to mammalian globin genes, I have carried out chromosomal walks by screening the genomic libraries using various fragments as probes. The results showed that leghemoglobin genes in soybean are arranged in four loci on the chromosome (Lee et al., 1983). In order to investigate how these four loci have been generated in soybean , I have carried out experiments with

two other legumes which are phylogenetically related to soybean; Phaseolus vulgaris (kidney bean) and Glycine soja (wild soybean). I 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. Chromosomal arrangements of leghemoglobin genes of P. vulgaris and G. soja were deduced from the results and were compared with that of soybean. These revealed possible evolutionary events leading to soybean leghemoglobin genes. Also, the structure as well as chromosomal arrangement of leghemoglobin genes will be compared with those of animal globin genes and their evolutionary relationship will be described in this report.

II. MATERIALS AND METHODS

A. Materials

(1). Enzymes and chemicals

Restriction endonucleases and DNA and RNA modifying enzymes were purchased from New England Biolabs, Boehringer Mannheim and Bethesda Research Laboratories (BRL), except for avian myeloblastosis virus reverse transcriptase which was obtained from Dr. J. Beard (Life Sciences Inc.), and S1 nuclease, a gift of Dr. F. Fuller.

The following chemicals were obtained from Sigma Chemical Co.: acrylamide, bis-acrylamide, Trisma base (Tris), urea, sodium lauryl sulfate (SDS), sucrose (RNAase free), N-lauryl sarcosine, ethidium bromide, ammonium persulfate, ampicillin, tetracycline, chloramphenicol, bovine serume albumin (BSA), polyethylene glycol (8,000), isopropyl-β-D-thio-galactopyranoside (IPTG), dimethylformamide, RNAase A, 5-bromo-4-chloro-3-indolyl-β-galactoside (X-gal), agar, Triton X-100, thiamine-HCl, and polyvinylpyrrolidone; BDH chemicals: sodium chloride (NaCl), tri-sodium citrate, sodium hydroxide (NaOH), sodium acetate, magnesium chloride ($MgCl_2$), boric acid, glycerol, di-sodium hydrogen orthophosphate (Na_2HPO_4), potassium dihydrogen orthophosphate (KH_2PO_4), ammonium chloride (NH_4Cl), ethylene diaminetetraacetic acid (disodium salt) (Na_2EDTA), calcium chloride ($CaCl_2$), potassium cloride (KCl), hydrochloric acid (HCl), acetic acid, and boric acid; Fisher Scientific Co.: magnesium sulfate and Paraffin oil; Bio-Rad

Laboratories: acrylamide and agarose; Difco Laboratories: bacto-tryptone, bacto agar and yeast extract; Pharmacia: Ficoll 400, Sephadex G-50, and dextran sulfate; Boeringer Mannheim: Tris; R-plus: urea; Estman Kodak: 2-mercaptoethanol, xylene cyannol FF and bromophenol blue; and BRL: phenol.

The cesium chloride was obtained from Kawecki Berylco Industries; oligo-(dT) was from Collaborative Research; deoxynucleotides and dideoxynucleotides were from P.L Biochemicals; [^{32}-P]-labelled nucleotides were from New England Nuclear and Amersham Corp.; GF/C and DE-81 filters were from Whatman; nitrocellulose paper (0.45 um pore size) was from Schleicher and Schnell (BA85) and from Millipore; GeneScreen paper was from New England Nuclear; Miracloth was from Calbiochem. and N-Z-Amine (type A) was from Humko Sheffield Chem.

(2). Biological materials

Soybean (Glycine max (L.) Merrill, cv. "Prize") seeds were obtained from Strayer Seed Farm, Hudson, Iowa and inoculated with Rhizobium japonicum strain 61A76 (Nitrogen Co., Milwaukee).

Kidney bean (Phaseolus vulgaris) seeds were obtained from W. H. Perron Co., Laval, Quebec and inoculated with R. phaseoli strain RCR3610 from Dr. T. Hall, University of Wisconsin.

Seeds of Glycine soja were obtained from N. Neilson, Purdue University.

Escherichia coli K12 strain K802 (hsr, hsm, galK, suII, lacY, met) was obtained from Dr. F. Blattner, University of Wisconsin and E. coli strain DH1 (recA1, endA1, gyrA96, thi-1, hsdR17, supE44) was obtained from Dr. D. Hanahan, Harvard University. E. coli strain JM101 (Δ(lac, pro), thi, strA, supE, endA, sbcB, hsdR$^-$, F'traD36, proAB, lacI, ZΔ M15).

The plasmid vectors pBR322 and pBR325 were propagated in E. coli strain DH1. Lambda DNA digested with HindIII was obtained from BRL.

(3). Growth media for bacteria and bacteriophage

Lb medium

Per liter:	bacto-tryptone:	10 g
	yeast extract:	5 g
	NaCl:	5 g
(for plates)-	bacto agar:	18 g

2 X YT medium

Per liter:	bacto-tryptone:	16 g
	yeast extract:	10 g
	NaCl:	5 g
(for plates)-	bacto agar:	15 g

H plates

Per liter:	bacto-tryptone:	10 g
	NaCl:	8 g
	bacto agar:	12 g

H top agar

Per liter:	bacto-tryptone:	10 g
	NaCl:	8 g
	agarose:	8 g

M9-glucose medium

Per liter:	Na_2HPO_4:	6 g
	KH_2PO_4:	3 g
	NaCl:	0.5 g
	NH_4Cl:	1 g
	$MgSO_4 7H_2O$:	0.25 g
	$CaCl_2$:	15 mg
	glucose:	4 g
	thiamine-HCl:	2 mg

NZY medium

Per liter:	NZ amine A:	10 g
	NaCl:	5 g
	$MgCl_2$ $7H_2O$:	2 g
	yeast extract:	5 g
(for plates)-	bacto-agar:	11 g
(for top agar)-	agarose:	6.5 g

B. Methods

(1). Growth of plant tissues

Soybean (Glycine max), wild soybean (G. soja) or kidney bean (Phaseolus vulgaris) 3-day old seedlings were inoculated with the appropriate Rhizobium species (R. japonicum or R. phaseoli) and grown in vermiculite in controlled environment chambers with a 16 hr photoperiod at 27°C and a dark period at 21°C. The plants were watered every day with nitrogen-free nutrient solution (Cutting and Schulman, 1969). Root nodules were harvested 3 weeks following infection and immediately stored in liquid nitrogen after harvesting.

Embryonic axes were hand-dissected from seeds presoaked in water and mature leaves were harvested from 3-week old plants. Uninfected root tips were obtained from seeds germinated in the dark in moist vermiculite for three days at 27°C.

(2). Preparation of DNA

(a). DNA from plant tissues

Nuclei were isolated from about 15 g of tissue as described by Beadich et al. (1980) and Varsanyi-Breiner et al. (1979). The tissue was ground in liquid nitrogen to a fine powder and then homogenized in cold buffer containing 50 mM TrisHCl (pH 8.0), 5 mM $MgCl_2$, 0.1 M diethyl-thiocarbonate and 0.3 M sucrose. The cell debri were removed by filtering

through cheesecloth. The nuclei were collected by centrifugation for 10 min at 9,000 rpm at 4°C. The nuclear pellet was resuspended in cold buffer containing 50 mM Tris-HCl (pH 8.0), 20 mM Na_2EDTA, 0.3 M sucrose and 1% sodium lauryl sarcosinate. When the lysate appeared uniformly dispersed, CsCl was added to a final concentration of 60%. This solution was centrifuged for 20 min at 12,000 rpm at 4°C and then filtered through Miracloth. Ethidium bromide was added to a final concentration of 400 ug/ml and the density was adjusted to 1.57 g/ml, as estimated by refractive index ($\underline{n}$=1.3875). Density gradients were performed by centrifuging for 40 hrs at 40,000 rpm in a Beckman 70 Ti rotor at 20°C. The band of DNA which fluoresced when viewed under UV light was removed with a syringe by puncturing the side of the tube. The ethidium bromide was extracted with CsCl-saturated isopropanol. The DNA was dialyzed against TE buffer [10 mM Tris-HCl (pH 7.5) and 1 mM EDTA]. The concentration of DNA was estimated from UV absorption at 260 nm using 1 O.D.= 50 ug/ml.

(b). Plasmid DNA

The plasmids were prepared using the rapid boiling method of Holmes and Quigley (1981). Bacterial cultures (5 ml) were grown overnight at 37°C in 2 X YT broth containing 50 ug/ml of ampicillin or 20 ug/ml of tetracycline. Bacteria were pelleted at 6,000 rpm for 5 min and resuspended in 0.35 ml of 8% sucrose, 5% Triton-X 100, 50 mM EDTA, 50 mM Tris-HCl (pH 8.0) (STET buffer) and 25 ul of a 10 mg/ml stock of freshly prepared lysozyme was added. The solution was placed in a boiling water

bath for 40 s and centrifuged immediately at 18,000 rpm for 10 min at either room temperature or 4°C. The supernatant was drawn off the slightly gelatinous pellet with a pasteur pippet and precipitated at -18°C for 10 min by addition of an equal volume of isopropanol. The precipitate was collected by centrifugation at 15,000 rpm for 10 min and was resuspended in 100 ul of 0.3 M sodium acetate (pH 7.0), followed by centrifugation at 10,000 rpm for 5 min. The supernatant was precipitated at -18°C for 10 min by addition of 250 ul of cold ethanol. The precipitate was collected by centrifugation at 15,000 rpm for 10 min, washed with 70% cold ethanol and resuspended in 50 ul of TE buffer. 5 ul of this solution was sufficient to yield visible DNA bands on a gel after restriction enzyme digestion in the presence of 20 ug/ml of RNAase A.

Large scale preparation of plasmids was performed using the standard precedure scaled up to 100 times. Bacteria pelleted from 500 ml culture were resuspended in 35 ml of STET buffer and processed exactly as described in the standard procedure up to the isopropanol precipitation step except that the solution was brought to boiling over an open flame, followed by incubation in a boiling water bath for 40 s. Precipitated nucleic acids were pelleted at 12,000 rpm for 10 min and resuspended in 8 ml of 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA, followed by the addition of 8 g of CsCl and 0.3 ml of a 10 mg/ml stock of ethidium bromide. The solution was centrifuged in a Beckman Ti 70.1 rotor at 20°C for 48 hrs at 40,000 rpm. The lower plasmid band was removed after visualization under UV light. The ethidium bromide was extracted with CsCl-saturated isopropanol and the plasmid was dialized against TE buffer.

(c). Phage DNA

The DNA from bacteriophage was prepared as suggested by Blattner et al. (1977) and Maniatis et al. (1978). About 10^6 phage were mixed with 3 ml of an overnight culture of E. coli strain K802 grown in NZY or Lb medium containing 10 mM $MgCl_2$. The phage were allowed to adsorb to the bacteria for 10 min at 37°C and mixed with 100 ml of top agar kept molten at 50°C. The top agar was then poured on the bottom agar in an aluminum tray, followed by incubation at 37°C overnight. The top agar, showing confluent lysis, was scraped off and transferred to a 250 ml centrifuge tube and 100 ml of phage buffer [10 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$] and 5 ml of chloroform were added to the tube. The resulting suspension was shaken for 2 - 3 hrs at room temperature and then centrifuged at 6,000 rpm for 10 min at 4°C. The supernant was transferred to another centrifuge tube. NaCl (60 g/l) and polyethylene glycol 8,000 (PEG) (70 g/l) were added to the tube. The solution was then kept at 4°C overnight. The precipitate was centrifuged at 6,000 rpm for 10 min and the pellet was dissolved in 8 ml of phage buffer. The solution was transferred to a 15 ml centrifuge tube, extracted with 3 ml of chloroform and centrifuged at 8,000 rpm for 10 min at 4°C. The supernatant was collected and CsCl was added to the final concentration of 1.5 g/ml. The solution was then centrifuged at 40,000 rpm at 20°C for 20 hrs in a Ti 70.1 rotor. The phage band was collected with a syringe by puncturing the side of the tube and the phage was dialyzed against phage buffer. The DNA was deproteinized 3 times with phenol and then was washed 3 times with chloroform. The DNA was precipitated with ethanol and dissolved in TE at a concentration of 0.1 ug/ul.

(3). Isolation of poly(A)$^+$ RNA

Poly(A)-containing RNA was isolated as described by Auger et al. (1979). Tissue (root, leaf or nodule) frozen in liquid nitrogen were ground in liquid nitrogen with a mortar and pestle and homogenized in 25 ml of polysome extraction buffer [0.15 M Tris-acetate (pH 8.5), 50 mM KCl, 30 mM Mg acetate, 0.3 M surose, 10 mM β-mecarptoethanol, 0.4% Non-Idet P-40]. The solution was passed through a layer of cheesecloth, transferred to a centrifuge tube and centrifuged at 15,000 rpm for 10 min at 4^{o}C. The supernatant was filtered through a layer of Miracloth, transferred to a ultracentrifuge tube over a 5 ml sucrose cushion containing 1.5 M sucrose, 50 mM Tris-acetate (pH 8.5), 50 mM KCl and 10 mM Mg acetate and centrifuged at 55,000 rpm for 90 min at 4^{o}C in a fixed angle rotor. The polysomal pellet was dissolved in a 0.5 ml of extraction buffer [0.1 M Tris-HCl (pH 9.0). 0.1 M NaCl, 0.01 M Na_2EDTA, 1% SDS], extracted with phenol 2 or 3 times, washed with chloroform and precipitated with 70% ethanol at -70^{o}C. The RNA was pelleted and washed with ethanol. The RNA pellet was dissolved in 2 ml H_2O, made up to 0.4 M NaCl and 0.1% SDS and fractionated on an oligo-(dT) cellulose column which had previously been washed with 0.1 M KOH (or NaOH), rinced with water and equilibrated with a solution containing 0.4 M NaCl and 0.1% SDS. The column was sequentially washed with a solution (0.4 M NaCl, 0.1% SDS), followed by a wash with the same solution without SDS. The small amount of RNA which remained bound to the column was eluted with 3 - 4 ml of H_2O in a polycarbonate tube and poly(A)-containing RNA was collected in a pellet by centrifugation at 55,000 rpm for 16 - 18 hrs at 4^{o}C. The pellet was dissolved in H_2O or TE buffer and kept at -70^{o}C. The presence of poly(A)$^+$ RNA in the

eluted fraction was assessed by hybridization to ^{3}H-poly(U) as described by Verma et al. (1974).

(4). DNA digestion and electrophoresis

(a). Restriction enzyme digestions

Restriction enzymes were used as recommended by the suppliers. Usually four units of enzymes were used per ug of DNA. The reactions were stopped by adding loading buffer (75% sucrose, 0.001% bromophenol blue and 0.01% xylene cyanol FF) and the DNA was loaded on the gel for electrophoresis. If the reaction volume was too much to load on the gel, the DNA was precipitated and dissolved in the small volume before adding the loading buffer.

(b). Gel electrophoresis

1. Polyacrylamide gel electrophoresis

Analytical or preparative polyacrylamide gel electrophoresis was carried out using a Bio-Rad gel apparatus with 1.5 mm or 3 mm spacers. The concentration of acrylamide was 5% and the ratio of acrylamide to bisacrylamide was 30:1. The polymerization was initiated with TEMED and ammonium persulfate. Electrophoresis was carried out at 150 volts in TBE buffer [100 mM Tris borate (pH 8.3), 1 mM Na_2EDTA]. The DNA bands were visualized by staining the gels with ethidium bromide. Plasmid pBR322, cut with restriction enzyme HinfI, was used as molecular weight markers.

2. Agarose gel electrophoresis

Agarose gels (Ø.8% to 1.5%) were prepared in TAE buffer [4Ø mM Tris acetate (pH 8.Ø), 1 mM EDTA] with 1 ug/ml of ethidium bromide. Agarose gel electrophoresis was carried out at 7Ø volts in TAE buffer. Lambda DNA digested with HindIII was used as molecular weight markers.

(5). Southern hybridization

(a). Transfer of DNA from gels

DNA was transferred to GeneScreen by the method of Southern (1975). After electrophoresis was completed, the gel with a ruler along its edge was photographed. The gel was transferred to a glass baking dish and the DNA was partially hydrolyzed by acid depurination by soaking the gel twice for 1Ø min in Ø.25 M HCl at room temperature (Wahl et al., 1978). The DNA was then denatured by soaking the gel twice for 15 min in 1.5 M NaCl and Ø.5 NaOH at room temperature with constant shaking. The gel was neutralized by soaking twice for 15 min in 1.5 M NaCl, Ø.5 M Tris-HCl (pH 7.5) at room temperature with constant shaking. Two sheets of Whatman 3 MM paper saturated with the neutralization solution were prepared on the glass plate over a glass tray containing the solution. The gel was then placed on the Whatman paper. Saran wrap was placed on the Whatman paper around the gel to prevent contact between the dry paper to be placed above the gel and the wet paper beneath. The GeneScreen paper was positioned on the top of the gel, followed by the additions of two sheets of dry Whatman

paper, then a 3-inch layer of paper towels, and finally a light weight, to ensure even contact. Transfer of DNA fragments from an agarose gel was carried out for 16 hrs. After transfer was complete, the towels and the Whatman papers above the gel were removed and the GeneScreen paper was taken off the gel and soaked in the neutralization solution at room temperature for 5 min. The dried filter was placed between two sheets of Whatman paper and baked for 2 hrs at 80°C under vacuum.

(b). Hybridization

The hybridization probes were prepared by nick-translation of double stranded DNA using DNA polymerase as described by Rigby et al. (1977). cDNA probes were prepared from polysomal poly(A)-containing RNA as described by Fuller et al. (1983).

1. Hybridization of DNA bound to GeneScreen

The baked filter was prehybridized for 2 hrs at 65°C in a sealed plastic bag containing prehybridization solution {5 X SET [0.5 M NaCl, 0.15 M Tris-HCl (pH 8.0), 5 mM Na_2EDTA], 10 X Denhardt solution (100 X Denhardt: 2% BSA, Ficoll and polyvinyl pyrrolidone), 10% Dextran sulfate, 1% w/v SDS, 50 ug/ml of sheared and denatured calf thymus DNA and 50 ug/ml of poly(A)}. The prehybridization solution was discarded and replaced with an identical solution containing the hybridization probe. Hybridization was performed for 20 hrs at 60 to 65°C. After the unhybridized probe was removed by washing the filter in 2 X SSC (20 X SSC: 3 M NaCl and 0.3 M tri-sodium citrate), 0.1% SDS, the filter was dried and covered with Saran

wrap. The hybridized DNA fragments were revealed by autoradiography at -70°C with Kodak XR-5 X-ray films and a Dupont Cronex Lightening-Plus intensifying screen.

2. Colony hybridization

Colony hybridization was carried out by the method of Grunstein and Hogness (1975). Using sterile toothpicks, the colonies were simultaneously consolidated onto a master agar plate and onto a nitrocellulose filter laid on the surface of a second agar plate containing the selective antibiotics. After a period of growth at 37°C, the colonies on the nitrocellulose filter were lysed for 2 to 3 min on 3 MM paper that had been saturated with 0.5 M NaOH and 1.5 M NaCl. The filter was then neutralized on 3 MM paper that had been saturated with 1.5 M NaCl and 0.5 M Tris-HCl (pH 7.5). After 5 min, the filter was blotted and dried at room temperature and baked for 2 hrs at 80°C in a vacuum oven. Hybridization of the filter was performed as described before.

(6). Screening of genomic libraries

The screening of the genomic library was performed using the precedure of Woo (1981), which included the additional amplification step of Benton and Davis (1977).

Each time, about 0.5 X 10^6 bacteriophage were screened. Bacteriophage were grown up to 10^4 plague forming units per plate using E. coli strain K802 (for soybean libraries) or DH1 (for a kidney bean genomic library) as hosts. For phage amplification, an overnight culture

of K802 grown in Lb broth containing 10 mM $MgCl_2$, was diluted 10-fold with 180 ml of culture medium at room temperature in a 15-cm-diameter petri dish. Nitrocellulose filters, precut to circles 8.6 cm in diameter, were marked with a water-proof pen and submerged in the diluted cell suspension. The filters coated with cells were blotted and dried on Whatman 3 MM papers. The entire precedure was carried out in a laminar flow transfer hood. The nitrocellulose filters were then layered onto the phage-containing agar plates for 5 min to allow the transfer of phage particles. The filters were gently peeled off the phage-containing soft agar and transferred onto a fresh Lb agar plate supplemented with 10 mM $MgCl_2$, with the side that had come into contact with phage plaques facing up. All filters were prepared in duplicate. The fresh petri dishes, containing the nitrocellulose filters, were then incubated at 37°C overnight.

The next day, the nitrocellulose filters were lifted off the agar plates and layered to rest on Whatman paper saturated with 0.5 M NaOH and 1.5 M NaCl for 5 min at room temperature. The bacteria and phage particles were lysed (the denatured DNAs were fixed onto the nitocellulose filters during this treatment). The filters were neutralized by transferring to Whatman paper saturated with 0.5 M Tris-HCl (pH7.5) and 1.5 M NaCl, washed with 20 X SSC and blotted on Whatman papers for drying. The air-dried filters were baked for 2 hrs under vacuum at 80°C.

Hybridization of filters were carried out as described before. Usually strong hybridization signals were obtained after overnight autoradiography. All positive clones were scored when the spot appeared on both filters. The film was placed on a light box and the plates were

oriented above the film. Using a sterile Pasteur pipet, a 1 cm^2 area was removed from the top agar layer above the spot on the film and the top agar was placed in an Eppendorf tube containing 1 ml of phage buffer and Ø.1 ml of chloroform. The tube was then vortexed. The suspension was used for a second screening at lower phage concentrations (1ØØ - 5ØØ plaque forming units per plate) so that individual phage could be isolated. The phages in the plaques produced in the second screening were then individually amplified by removing them from a plate with a sterile Pasteur pipet, suspending them in Ø.5 ml phage buffer and Ø.Ø5 ml chloroform and plating them. The amplified phage stock was used for the preparation of phage DNA.

(7). Molecular cloning

(a). Synthesis and cloning of cDNA

A partial cDNA library of nodule mRNA of kidney bean was constructed according to the precedure of Fuller et al. (1983). Poly(A)$^+$ RNA was isolated from total polysomes of nodules as described before. Double stranded cDNA was synthesized to nodule RNA essentially by the method of Wickens et al. (1978), except that the second strand was synthesized at pH 8.Ø. First strand reaction mixtures contained 5Ø mM Tris-HCl (pH 8.3), 1Ø mM $MgCl_2$, 1ØØ mM KCl, Ø.5 mM dATP/dCTP/dGTP/dTTP, MeHgOH-denatured nodule poly(A)$^+$ RNA at 1ØØ ug/ml, oligo-(dT) at 4Ø ug/ml, and reverse transcriptase at 1Ø units/ug of RNA. The first strand was converted to double strand in 3Ø min by using 1Ø units of DNA polymerase I per ug of input RNA. Double-stranded cDNA was treated with nuclease S1 (Vogt, 1973)

and tailed with dCTP (Villa-Komaroff et al., 1978). PstI-cut pBR322 tailed with dGTP (purchased from BRL) was hybridized to tailed cDNA and transformed into E. coli strain DH1 as described below.

(b). Subcloning of genomic clones

The DNA from genomic clones were digested with a restriction enzyme and inserted into plasmid pBR322 in order to facilitate their analysis.

Phage DNA and pBR322 (each 1 ug) were digested in different Eppendorf tubes with restriction enzyme HindIII or BamHI for 1 hr, extracted with phenol, precipitated with ethanol and dissolved in 20 ul of TE buffer. Ligation was carried out overnight at 14°C in 20 ul reaction mixtures containing 15 ul of phage DNA, 2 ul of plasmid DNA, 2 ul of 10 X ligation buffer [200 mM Tris-HCl (pH 7.6), 10 mM EDTA, 100 mM dithiothreitol, 6 mM ATP, 100 mM $MgCl_2$] and 0.1 unit of T_4 DNA ligase.

Transformation of ligated DNA into competent E. coli strain DH1 was carried out by using a modified calcium chloride treatment method (Mandel and Higa, 1970). One ml of a fresh overnight culture of DH1 was inoculated in 50 ml of Lb broth and grown to $O.D._{600}$ = 0.3 - 0.6. The cells were cooled on the ice and collected by centrifugation at 5,000 rpm for 5 min at 4°C. The cells were resuspended in 25 ml of cold calcium chloride solution [10 mM Tris-HCl (pH 8.0) and 50 mM $CaCl_2$], kept on the ice for 20 min and collected by centrifugation at 3,000 rpm for 5 min at 4°C. The pellet was resuspended in 2.5 ml of the calcium solution. Competent cells (200 ul) were transferred to glass test tubes on ice containing 5 ul of ligated DNA. After 40 min, the tubes were heated at 42°C for 3 min and 1 ml of Lb medium was added to each tube. The cells were incubated at

37°C for 1 hr and plated on Lb plate containing 50 ug/ml of ampicillin to select for the transformed bacteria. Resistant colonies were then replica plated on Lb plates containing 20 ug/ml of tetracycline. Plasmid DNA from the tetracycline-sensitive colonies was isolated and analyzed for the presence of inserts.

(8). DNA sequencing

Nucleotide sequences were determined by the dideoxy chain termination method of Sanger et al. (1977).

(a). Cloning precedures

1. Isolation of DNA fragments

DNA fragments to be sequenced were prepared from subclones of the genomic clones. Plasmid DNA was digested with a restriction enzyme(s) and separated on agarose or polyacrylamide gels. After visualizing the bands, a slice of agarose (or polyacrylamide) containing the band of interest was cut out and the DNA was recovered from the gel slice by electroelution into a dialysis bag (McDonnell et al., 1977). All the buffer in the bag was recovered, extracted with phenol, precipitated with ethanol and dissolved in binding buffer [0.2 M NaCl, Tris-HCl (pH 7.2), 1 mM EDTA for the DNA less than 1000 bp in length, or 1.0 M NaCl, Tris-HCl (pH 7.2), 1 mM EDTA for the DNA greater than 1000 bp in length]. The DNA was further purified using NACS Prepac mini-column using the protocol of BRL. DNA recovered from the mini-column was pure enough to be digested with any

restriction enzyme.

2. Ligation

Replicating form (RF) DNA of M13 phage, mp8 or mp9 (supplied by Amersham) was digested with one of the restriction enzymes having a unique recognition sequence in the cloning region on the vector, extracted with phenol, precipitated with ethanol and dissolved in TE to give a final concentration of approximately 10 ng/ul. About 1 ug of the purified DNA fragment was digested with a restriction enzyme (AluI, RsaI, Sau3AI or TagI), extracted with phenol, precipitated with ethanol and dissolved in TE buffer to 20 ng/ul. Ligation reaction mixture contained 5 ul of DNA insert (100 ng), 2 ul of DNA vector (20 ng), 1 ul of the 10 X ligation buffer, and 1 ul of T_4 DNA ligase (0.1 unit/ul for sticky ends and 0.4 units/ul for blunt ends) and was incubated overnight at 14°C.

3. Transformation

Transformation of ligated DNA into competent cells of E. coli strain JM101 (supplied by Amersham) was carried out as described before. After cells were heat shocked at 42°C for 3 min, 280 ul of fresh cells/X-gal/IPTG mix (40 ul of 100 mM IPTG, 40 ul of 2% X-gal in diethyl formide and 200 ul of fresh E. coli cells) was added to each tube containing 3 ml molten H top agar kept at 50°C, mixed and poured immediately onto a H plate. The plates were inverted and incubated at 37°C overnight.

4. Preparation of single-stranded template

After overnight growth, infected cells of recombinant colorless plaques were grown up to produce single-stranded template for the sequencing reaction. Phage were amplified from single plaques in 1.5 ml cultures and harvested from culture supernatant by precipitating with polyethylene glycol and NaCl. Protein coat was removed by treatment with phenol and DNA was precipitated with ethanol, dissolved in 50 ul of TE buffer (pH 8.0) and stored in -20^{o}C for the subsequent sequencing.

(b). Sequencing procedures

1. Annealing primer to template

The annealing reaction mixture contained 5 ul of the single-stranded DNA template, 1 ul of M13 primer (1.2 ug/ml), 1.5 ul of Klenow reaction buffer [100 mM Tris-HCl (pH 8.5), 100 mM $MgCl_2$] and 2.5 ul of double-distilled water and was incubated in a laboratory oven at 55 - 60^{o}C for 1 - 2 hrs.

2. Sequencing reaction

The annealed primer was extended in 5' to 3' direction with Klenow fragment in presence of dNTPs and ddNTP. After annealing, the tubes were centrifuged briefly. 2 ul of labelled nucleotide ([α-^{32}P] dATP at 800 Ci/mmol) and 1 ul of Klenow fragment (1 unit/ul) were added to each tube and mixed carefully by pipetting in and out. 2.5 ul of the annealed

template/label/enzyme mix were added to mirocentrifuge tubes marked A, C, G, or T. To each tube, 2 ul of the relevant dNTP/ddNTP mix were added and the reaction was started. After 15 min, 2 ul of 0.5 mM dATP were added to each tube so that the reaction was not limited by the concentration of the single labelled dNTP supplied. After further 15 min, the reaction was stopped by adding 10 ul of formamide dye mix (100% formamide, 0.03% xylene cyanol FF, 0.03% bromophenol blue, 20 mM Na_2EDTA).

3. Gel electrophoresis

The sequencing gels contained 6% acrylamide with a ratio of acrylamide to bisacrylamide of 20:1, and were made in TBE buffer [100 mM Tris borate (pH 8.3), 1 mM Na_2EDTA] containg 8 M urea. The gels were pre-run about 1 hr in TBE buffer. The samples were boiled for 5 min and loaded on to two gels. The gels were electrophoresed at about 1,500 volts. One gel was run until the bromophenol blue dye front was just running off the bottom of the gel and the other 15 min after the xylene cyanol FF dye front ran off the bottom of the gel.

4. Autoradiography

After electrophoresis, the gel was transferred onto Whatman 3 MM paper, covered with Saran wrap and dried using a Bio-Rad gel dryer. The dried gel was exposed to Kodak X-ray film at room temperature overnight.

5. Comparison of DNA sequences

Comparison of DNA sequences was made using the NucAln cumputor program of Wilbur and Lipman (1983).

III. RESULTS

A. Chromosomal Arrangement of Leghemoglobin Genes in Soybean

(1). Isolation of genomic fragments containing leghemoglobin sequences

There are several lines of evidence which suggest that leghemoglobins are encoded by a family of closely related plant genes in soybean. The primary amino acid sequences of major leghemoglobin species are different from each other (Ellfolk and Sievers, 1971; Whittaker et al., 1981). Saturation hybridization studies using a kinetically purified leghemoglobin cDNA probe showed that there may be as many as 40 copies of leghemoglobin genes (Baulcombe and Verma, 1978). Southern blot analysis of EcoRI-digested genomic DNA revealed the presence of at least 7 different genomic fragments containing leghemoglobin sequences in soybean genome (Sullivan et al., 1981).

To find out the linkage of leghemoglobin genes on the chromosome, genomic recombinant molecules containing leghemoglobin sequences were isolated from a soybean genomic library which was constructed by R. Goldberg from a partial AluI-HaeIII digest of soybean genomic DNA using Charon 4A as a vector. About 5 X 10^5 recombinant phage were screened with a ^{32}P-labelled Lb cDNA clone called pLb14 constructed by B. Goodchild. The first screening of the genomic library gave rise to 12 independent clones.

These 12 clones were analyzed using restriction enzymes after large scale purification of phage recombinant DNA from each clone. Each phage DNA was digested with EcoRI, subjected to electrophoresis on a 1.2%

agarose gel and transferred on to the GeneScreen paper after visualization by staining with ethidium bromide. Lb-specific DNA sequences were visualized by hybridization with radioactively labelled pLbl4. Figure 1 shows the autoradiogram of the Southern analysis of these clones. The comparison of the EcoRI digestion pattern of each clone as well as its autoradiogram indicated the presence of 6 different clones. Each of these 6 clones was analyzed further using different enzymes HindIII, BglII, BamHI or PstI. The analysis of the 6 clones revealed partial overlapping of one clone to certain other clone: clone 19 to clone 28 (Group A); clone 2 to clone 4 (Group B); and clone 36 to clone 43 (Group C). The structures of inserts in these 6 clones are shown in Figure 2. Group A includes about 17 kb of genomic DNA. Group B contains two leghemoglobin sequences which are separated by an intergenic region of about 2.5 kb and one of these contains only the 5' region of a leghemoglobin sequence. Group C contains only one leghemoglobin sequence on a 1 kb HindIII fragment. The cross-hybridization between clones of these three groups using, as probes, flanking DNA fragments isolated from each clone showed no overlapping between three groups.

(2). Chromosomal walk on a leghemoglobin gene locus

Ch4GmLbl1 (Gml1) is one of the genomic clones isolated from a EcoRI-partial soybean genomic library (Sullivan et al., 1981). This clone carries a 11.5 kb EcoRI fragment which contains the Lbc_3 gene and the last two exons of a pseudogene (Brisson and Verma, 1982). The comparison of the restriction map of Gml1 with that of Group A revealed that they overlap with each other. One 4.0 kb HindIII fragment containing the Lbc_3

Figure 1. Analysis of 12 clones isolated from the AluI-HaeIII partial soybean genomic library. DNAs from 12 clones were cleaved with EcoRI, electrophoresed on a 1.2% agarose gel, transferred on to GeneScreen paper and hybridized with ^{32}P-labelled Lb cDNA clone pLb14. The number on the top indicates the name of each clone. Detailed restriction map of each clone is shown in Figure 2. Clones Gm11 and Gm4 were previously isolated from an EcoRI genomic library (Sullivan et al., 1981).

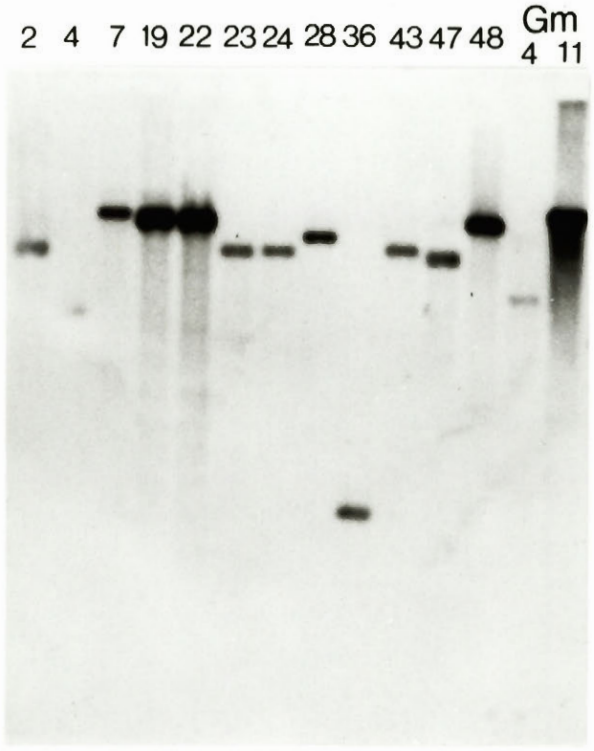
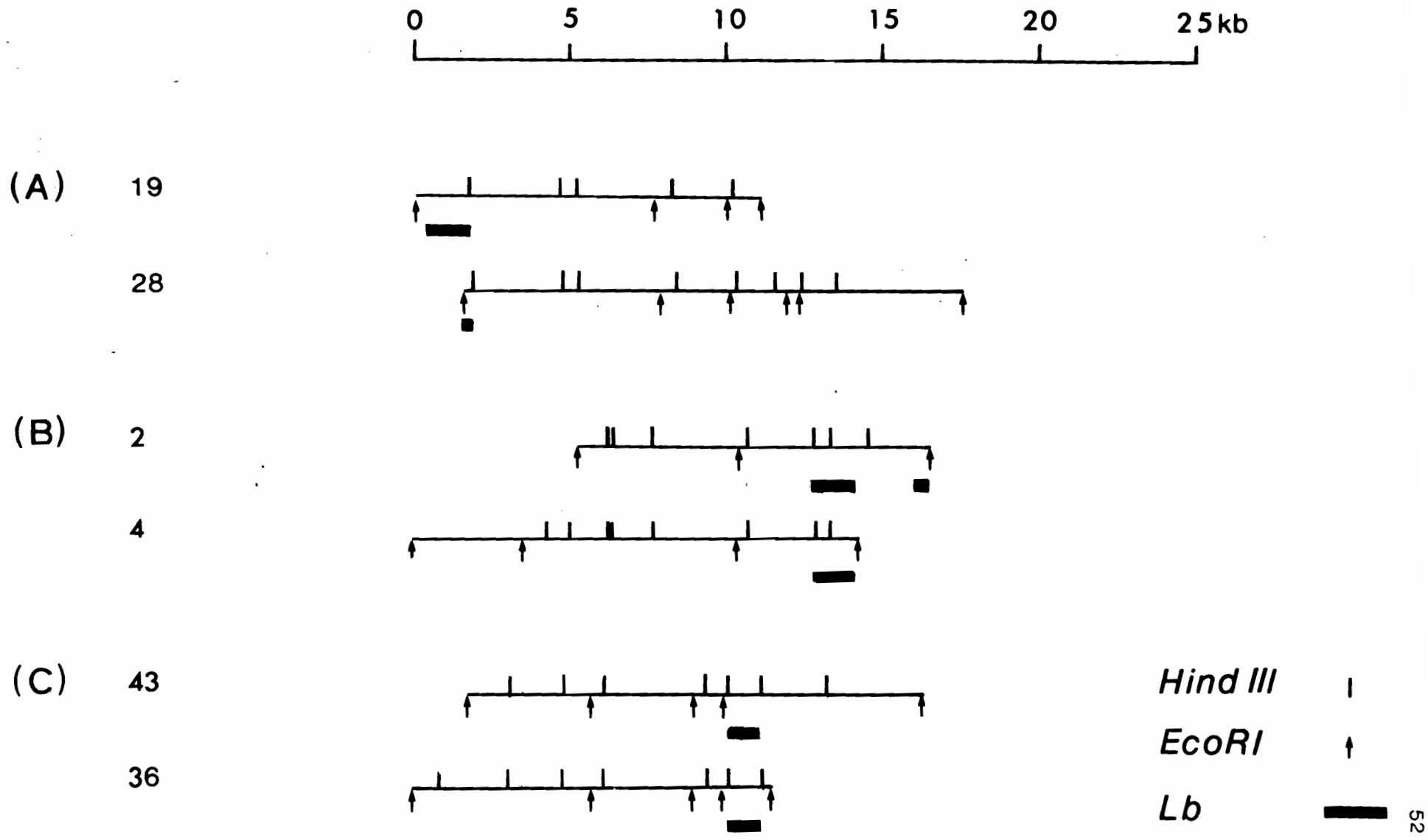

2
4
7
19
22
23
24
28
36
43
47
48
Gm
4
11

Figure 2. Restriction map of 6 genomic clones. The map was constructed with the enzymes EcoRI and HindIII. 6 different clones from original 12 clones were grouped into Group A, B and C, depending on their overlapping. Solid boxes indicate the regions hybridizing with pLbl4.


0
5
10
15
20
25kb
(A)
19
28
(B)
2
4
(C)
43
36
Hind III
EcoRI
Lb

gene was subcloned into pBR322. As shown in Figure 3, a 1.2 kb HaeIII intergenic fragment between a pseudogene and the Lbc_3 gene was then isolated from the subclone, labelled and used as a probe to screen the AluI-HaeIII library to reach the upstream region from the Lbc_3 gene on the chromosome. This yielded two clones, clone 13 and clone 15, whose structures are shown in Figure 4, with those of other clones. As shown in Figure 4, clone 15 carries two EcoRI fragments, one of which is part of the 11.5 kb fragment of Gml1 and the other fragment is upstream of the fragment and contains the first two exons of the pseudogene (Wiborg et al., 1983). Clone 13 carries three EcoRI fragments, all of which hybridize with the Lb cDNA clone. Jensen et al. (1981) isolated a 7.5 kb EcoRI fragment containing a leghemoglobin sequence from the EcoRI-partial library. The restriction map of this 7.5 kb fragment was exactly the same as that of one of three EcoRI fragments on clone 13. Another EcoRI fragment of 5.Ø kb in size turned out to be part of the 11.5 kb fragment. The small EcoRI fragment of 1.Ø kb in size also contains a leghemoglobin sequence. This indicated that clone 13 contains four leghemoglobin sequences.

Further walks across the chromosome were carried out by screening the AluI-HaeIII and EcoRI libraries using various fragments as probes. A screening of the AluI-HaeIII library with pLbl4 yielded clone 6. The screening of the EcoRI library with the 5.Ø kb EcoRI fragment from clone 28 yielded clone 32E. Another screening of the EcoRI library with the pLbl4 yielded clones 6ØE and 7ØE. The structures of clones 6, 32E and 7ØE are shown in Figure 4, while that of clone 7ØE is shown in Figure 5A.

Figure 4 shows the EcoRI and HindIII restriciton maps derived from the detailed analyses of lambda clones of this region of the chromosome.

Figure 3. Restriction map of the 11.5 kb fragment of Gml1 carrying the Lbc_3 gene and part of a pseudogene, $Lb\psi_1$. A 4.Ø kb HindIII fragment from Gml1 was subcloned into pBR322. From this subclone, a 1.2 kb HaeIII fragment was isolated and used as a probe to carry out the chromosomal walk.

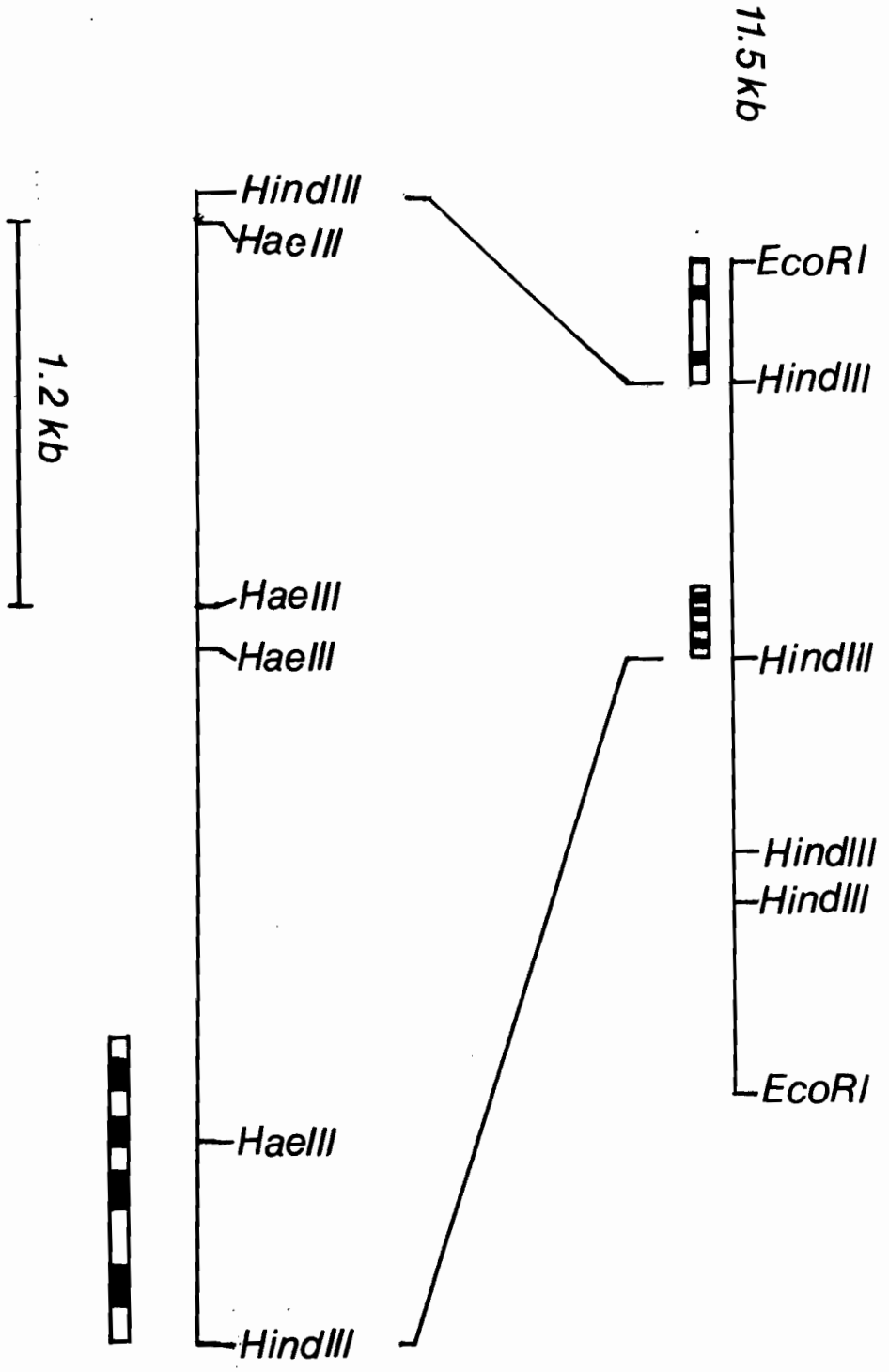
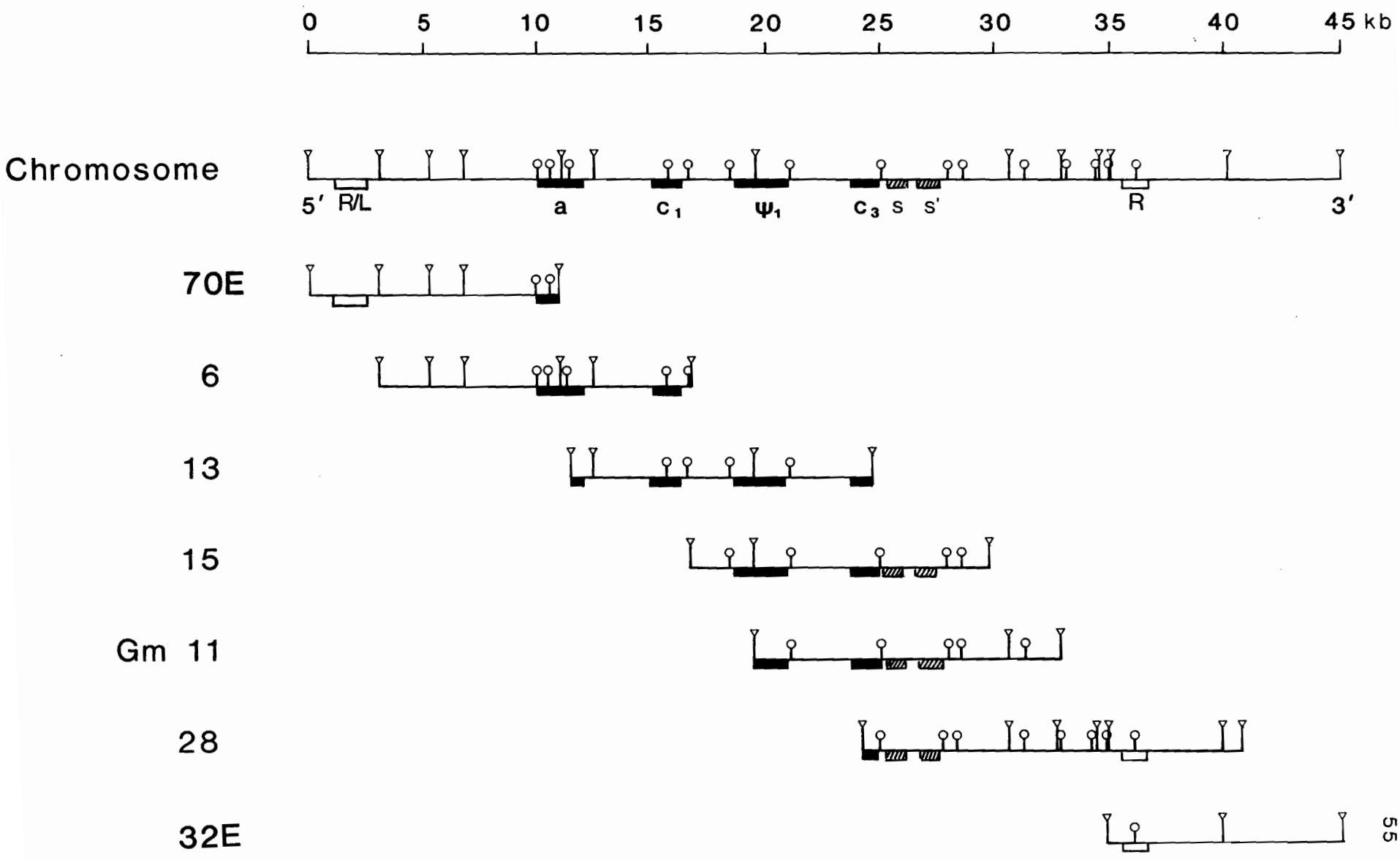

11.5 kb
HindIII
HaeIII
EcoRI
HindIII
1.2 kb
HaeIII
HaeIII
HindIII
HindIII
HindIII
HaeIII
EcoRI
HindIII

Figure 4. Chromosomal organization of a leghemoglobin locus. Clones 70E, 6, 13, 15, and 28 flanking the Gml1 region contain 4 leghemoglobin genes (solid boxes). Open boxes indicate the position of sequences which appear to be expressed in root and leaf tissues (see Figures 13 and 14). The dotted boxes indicate the positions of sequences which are homologous to the 3' region of the three other Lb gene regions (see Figure 5). The E on the clone number indicates that the clone was isolated from the EcoRI genomic library. Arrow heads, EcoRI and circle, HindIII sites. Clone Gml1 was isolated previously (Sullivan et al., 1981).


0
5
10
15
20
25
30
35
40
45 kb
Chromosome
5′ R/L
a
c 1
ψ1
c 3 s s'
R
3′
70E
6
13
15
Gm 11
28
32E

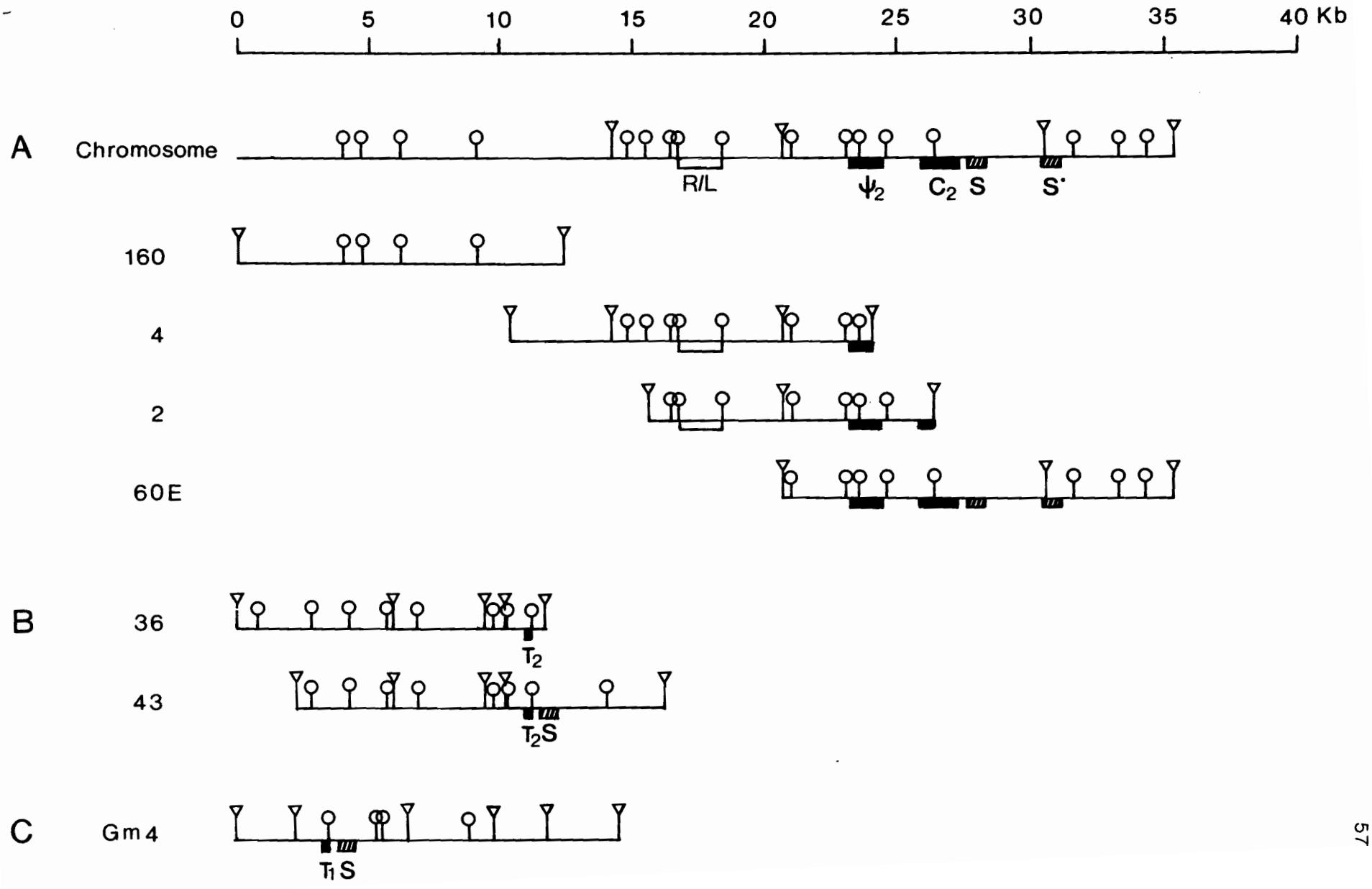
These 7 overlapping clones cover about 45 kb of contiguous DNA on the soybean genome. This region carries four EcoRI fragments of 11.5, 7.5, 4.2 and 1.4 kb in size containing four leghemoglobin genes.

The four leghemoglobin genes identified in this region are arranged in the order: Lba-Lbc_1-$Lb\psi_1$-Lbc_3. The positions of Lba and Lbc_1 genes were deduced by aligning the restriction maps of the cloned DNAs with those inferred from Lba and Lbc_1 sequences (Hyldig-Nielsen et al., 1982). Restriction enzyme analyses (Figure 4) and sequencing data (Brisson and Verma, 1982; Hyldig-Nielsen et al., 1982) suggest that all of these genes including the pseudogene are present in the same transcriptional orientation. The intergenic regions between these leghemoglobin genes average 2.5 kb and are shorter than those in typical mammalian globin loci. Since no other leghemoglobin genes were found within 10 kb of the 5' end of the Lba gene or within 20 kb of the 3' end of the Lbc_3 gene and two other plant genes flank each end (see below), this gene cluster repersents one complete leghemoglobin locus (named the Lba locus).

(3). Other regions containing leghemoglobin sequences

Similarly, a chromosomal walk was carried out in the region covered by clones 2 and 4 and yielded two more clones 160 and 60E. Figure 5A and B show the EcoRI and HindIII restriction maps of two other Lb gene-containing regions isolated from the AluI-HaeIII and EcoRI libraries. Clones 2, 4, 160 and 60E cover about 35 kb of contiguous DNA on the chromosome. This region contains two linked leghemoglobin sequences with a spacer of 2.5 kb. Nucleotide sequence analyses of these two sequences

Figure 5. Other leghemoglobin loci in soybean. Three Lb regions (A, B and C) which are not contiguous to the main Lb locus were isolated from the libraries. The region defined by clones 16Ø, 2, 4 and 6ØE contains two leghemoglobin sequences (solid boxes) and sequences homologous to the 3' region of the Lbc_3 gene (dotted boxes) (see Figure 4). A sequence at the 5' end of the leghemoglobin gene which is expressed in root and leaf tissues (Figure 14) is shown by an open box. Clones 36 and 43 represent a leghemoglobin sequence (solid box) which is flanked by the sequence homologous to the 3' region of the Lbc_3 gene. Gm4, carrying a truncated Lb sequence [see Brisson and Verma (1982)] also contains the repeat sequence found at the main locus as well as A and B loci. Arrow heads, EcoRI and circles, HindIII sites.

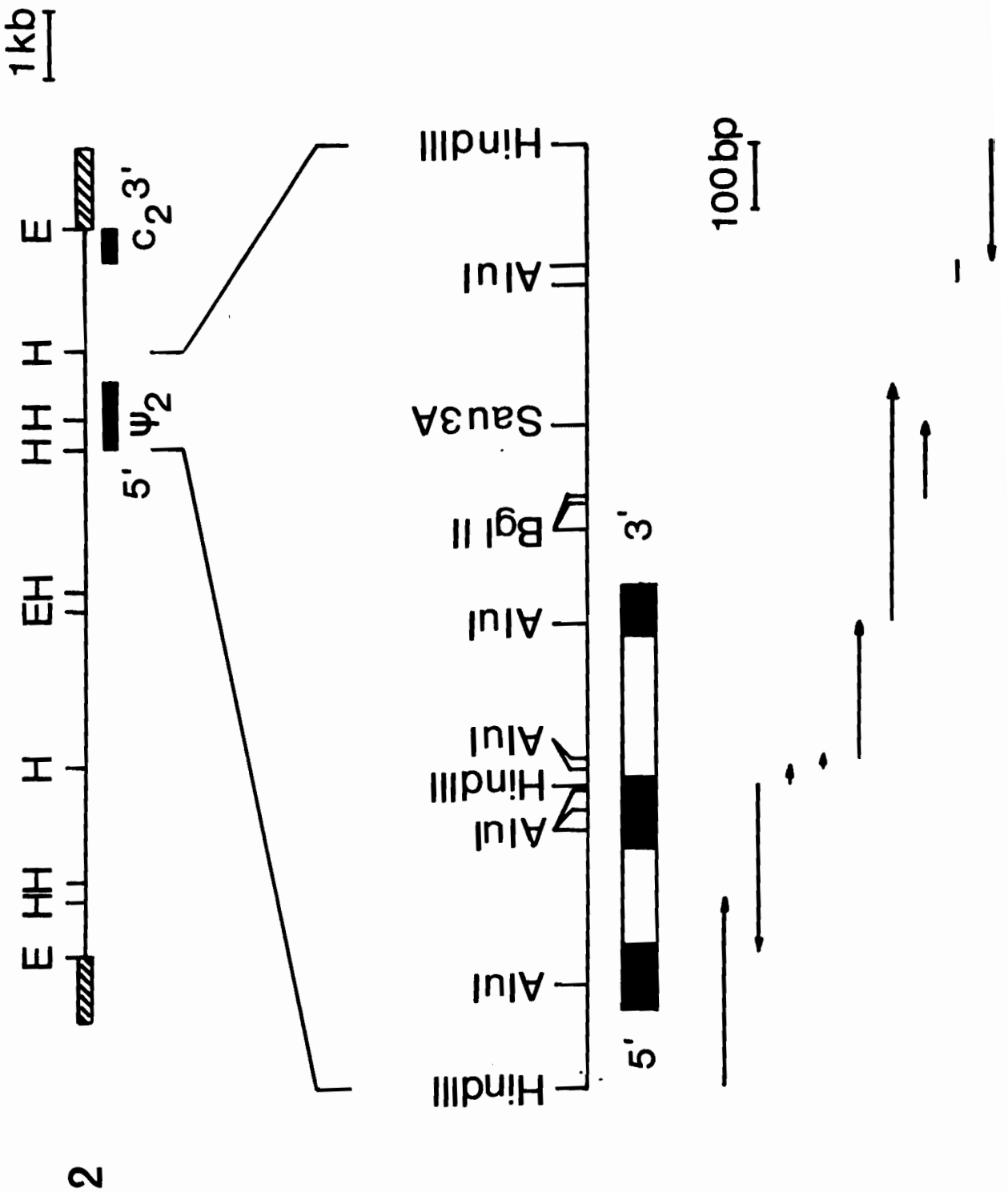

0
5
10
15
20
25
30
35
40 Kb
A
Chromosome
R/L
ψ2
C2
S
S·
160
4
2
60E
B
36
T2
43
T2S
C
Gm4
T1S

revealed that the first gene contains only the first two exons and part of the third exon and the second gene codes for Lbc_2 (Wiborg et al., 1982) (see below). The locus containing two leghemoglobin genes was named the Lbc_2 locus. The other region which includes clones 36 and 43 contains a leghemoglobin sequence. The nucleotide sequence of the Lb-containing region on the clone 43 indicated that only the last exon of a leghemoglobin gene is present in this region (the LbT_2 locus) (For more details, see below). This is similar to the previously identified truncated sequence on clone Gm4 shown in Figure 5C (the LbT_1 locus) (Brisson and Verma, 1982; Brisson, 1982). Hybridization of EcoRI-digested genomic DNA with a BglII fragment flanking the 3' region of the truncated Lb sequence showed that the Lb-containing 5.8 kb EcoRI fragment on clone 43 represents a 6.0 kb genomic fragment (data not shown).

(4). Analysis of leghemoglobin sequences on clone 2

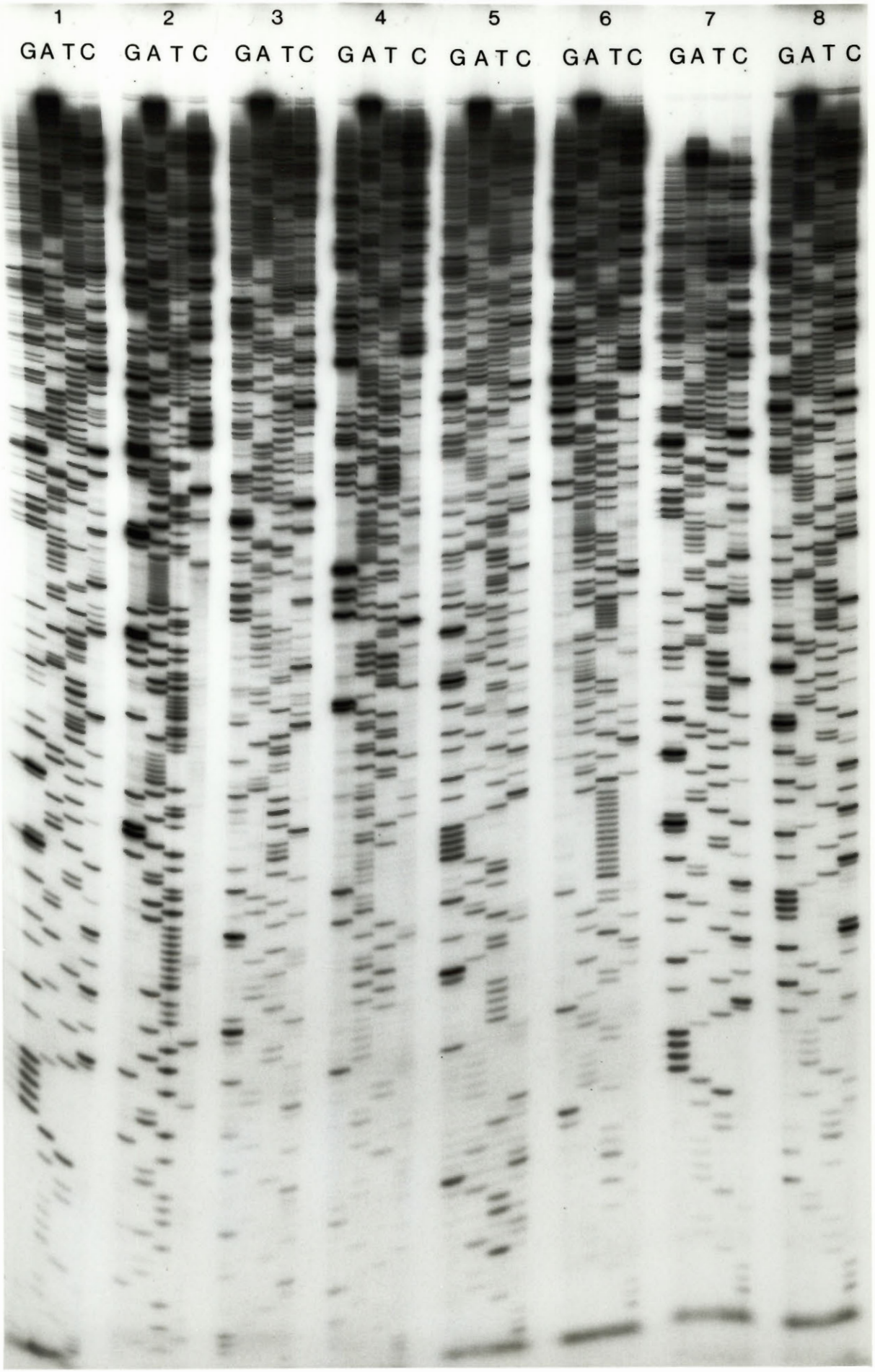
The region covered by overlapping clones 2, 4, 160 and 60E (Figure 5A) contains two leghemoglobin genes. Partial nucleotide sequence of the second gene (V. Mauro, unpublished data) showed that the gene is the Lbc_2 gene whose nucleotide sequence has been known (Wiborg et al., 1982). However, the other gene was not characterized. Therefore, I analyzed the nucleotide sequence of this gene to find out the nature of this sequence. Two HindIII fragments containing this gene were subcloned into pBR322 and the nucleotide sequence of the Lb-containing region was determined according to the strategy illustrated in Figure 6. An example of an autoradiograph of a sequencing gel is shown in Figure 7.

Figure 6. Restriction map of the region of clone 2, showing two linked leghemoglobin genes. Two HindIII fragments containing the Lb ψ_2 gene were subcloned into pBR322 and sequenced. The extended region shows the detailed restriction map. Horrizontal arrows indicate the extent and direction of sequencing. Coding regions are shown by solid boxes and the open boxes indicate the intervening sequences. Hatched boxes on clone 2 indicate lambda arms.


1kb
E
H
HH
EH
H
HH
E
$c_2$3'
ψ_2
5'
HindIII
AluI
Sau3A
BglII
AluI
AluI
HindIII
AluI
AluI
HindIII
3'
5'
100bp

2

Figure 7. Example of an autoradiograph of a sequencing gel. 8 different M13 recombinant clones were subjected to sequencing by the dideoxy chain termination method as described in Methods and electrophoresed on a Ø.3 mm polyacrylamide sequencing gel until the bromophenol blue dye reached the bottom of the gel.


1
2
3
4
5
6
7
8
GATC
GATC
GATC
GATC
GATC
GATC
GATC
GATC

The nucleotide sequence of the first gene on clone 2 is shown in Figure 8, together with the nucleotide sequence of the Lba gene (Hyldig-Nielsen et al., 1982). It revealed that the gene carries only the first two exons and part of the third exon, representing a truncated gene. The deletion appears to have occurred at position 750 (representing codon 90). This truncated gene is different from two other truncated genes carrying only the last exon (Brisson and Verma, 1982; also see below). 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 in the second exon (TTG to TAG) and the other at the splicing junction between the second exon and intron (GT to CT), which are commonly found in the pseudogenes (Treisman et al., 1983; Proudfoot et al., 1982). The first would result in an in-phase termination, while the second would affect RNA splicing. Otherwise, this DNA sequence shows high homology with the Lba gene. These two leghemoglobin sequences are arranged in the order:5'-Lbψ_2-Lbc$_2$-3'. They are present in the same transcriptional orientation within a space of 2.5 kb. Since the overall organization of the Lba locus carrying four leghemoglobin genes is similar to that of the Lbc$_2$ gene locus, it suggests that a large deletion in a locus carrying four leghemoglobin genes might have resulted in the generation of this two-gene Lbc$_2$ locus (For more details, see below).

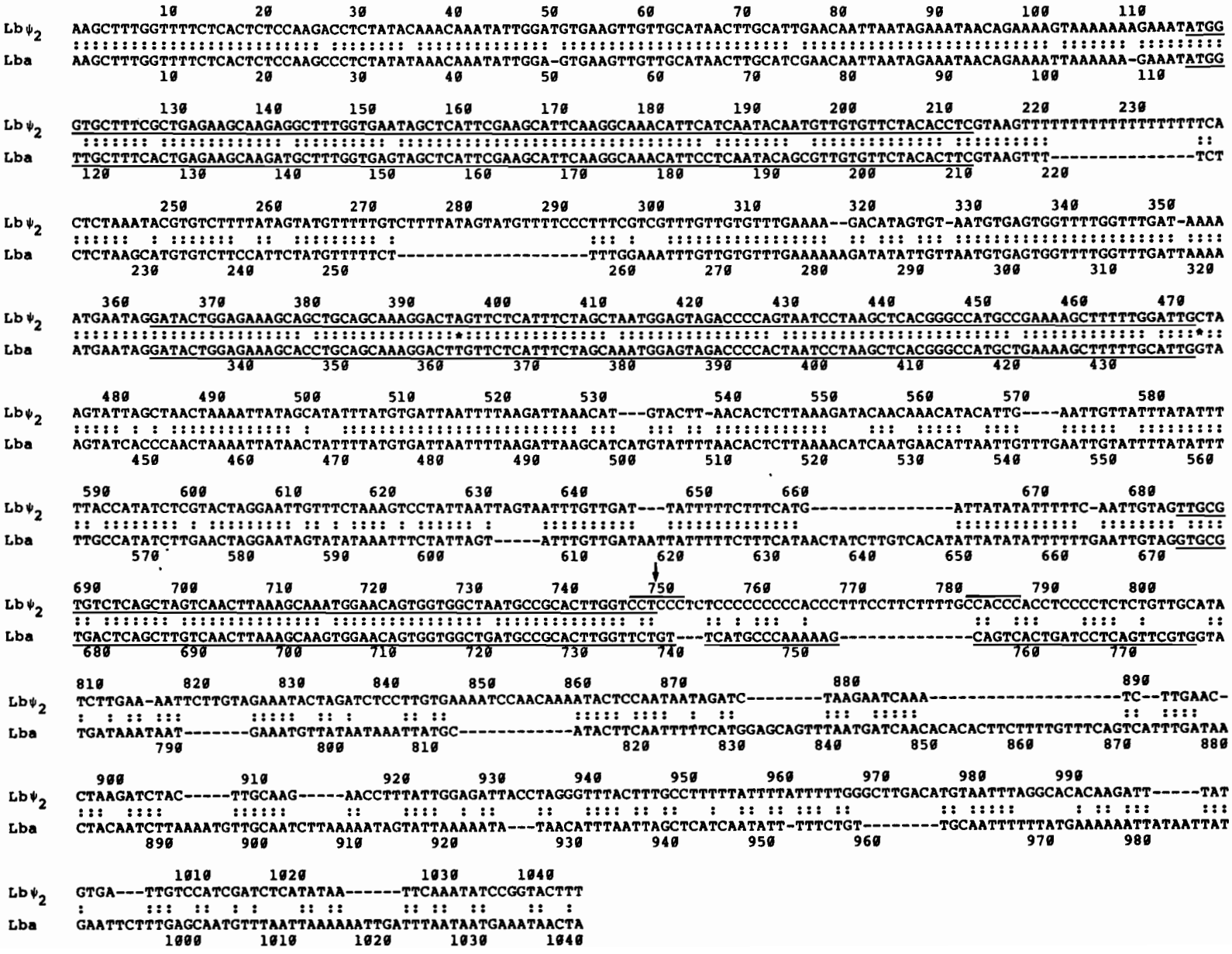

Lbψ2
Lba

Figure 8. Nucleotide sequence of the $Lb\psi_2$ 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 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 underlined at the end of the truncated gene.

(5). Structure of a leghemoglobin sequence on clone 43

In order to facilitate the nucleotide sequence analysis of this sequence, a 1.Ø kb HindIII fragment containing the gene was subcloned into pBR322. The fragment was then isolated from the subclone and subjected to DNA sequencing using the dideoxy chain termination method according to the strategy illustrated in Figure 9. Also, the nucleotide sequence of part of a 2.8 kb HindIII fragment next to the 1.Ø kb fragment on clone 43 was determined (its sequencing strategy is not indicated in Figure 9). The sequence is shown in Figure 1Ø, together with that of a truncated gene on Gm4 (Brisson and Verma, 1982; Brisson, 1982). It revealed that this gene (LbT_2) carries only the last exon as the LbT_1 gene on Gm4. The phenylalanine (coded by TTT) at the C-terminal end indicates that these sequences are derived from either the Lbc_2 gene or the Lbc_3 gene. Extensive sequence homology between the LbT_1 and LbT_2 genes suggests the possibility of tetraploidization being responsible for their origin.

(6). A sequence flanking the leghemoglobin loci

In order to determine if some or all of the leghemoglobin gene loci share any sequences other than those in the coding regions, various probes of 5' and 3' flanking fragments as well as intergenic regions, were hybridized to restricted Lb-containing recombinant phage DNAs. Genomic DNAs restricted with the same enzyme were analyzed in parallel. A 2.7 kb HindIII fragment immediately flanking the 3' end of the Lbc_3 gene (indicated by a dotted bracket in Figure 4) when used as a probe, hybridized to several genomic clones as shown in Figure 11B.

Figure 9. Restriction map of the Lb-containing region of clone 43. The HindIII fragment containing the LbT_2 gene was subcloned into pBR322 and sequenced using the dideoxy chain termination method, as illustrated in this figure. Hatched boxes on clone 43 indicate lambda arms. The extended region shows the detailed restriction map.

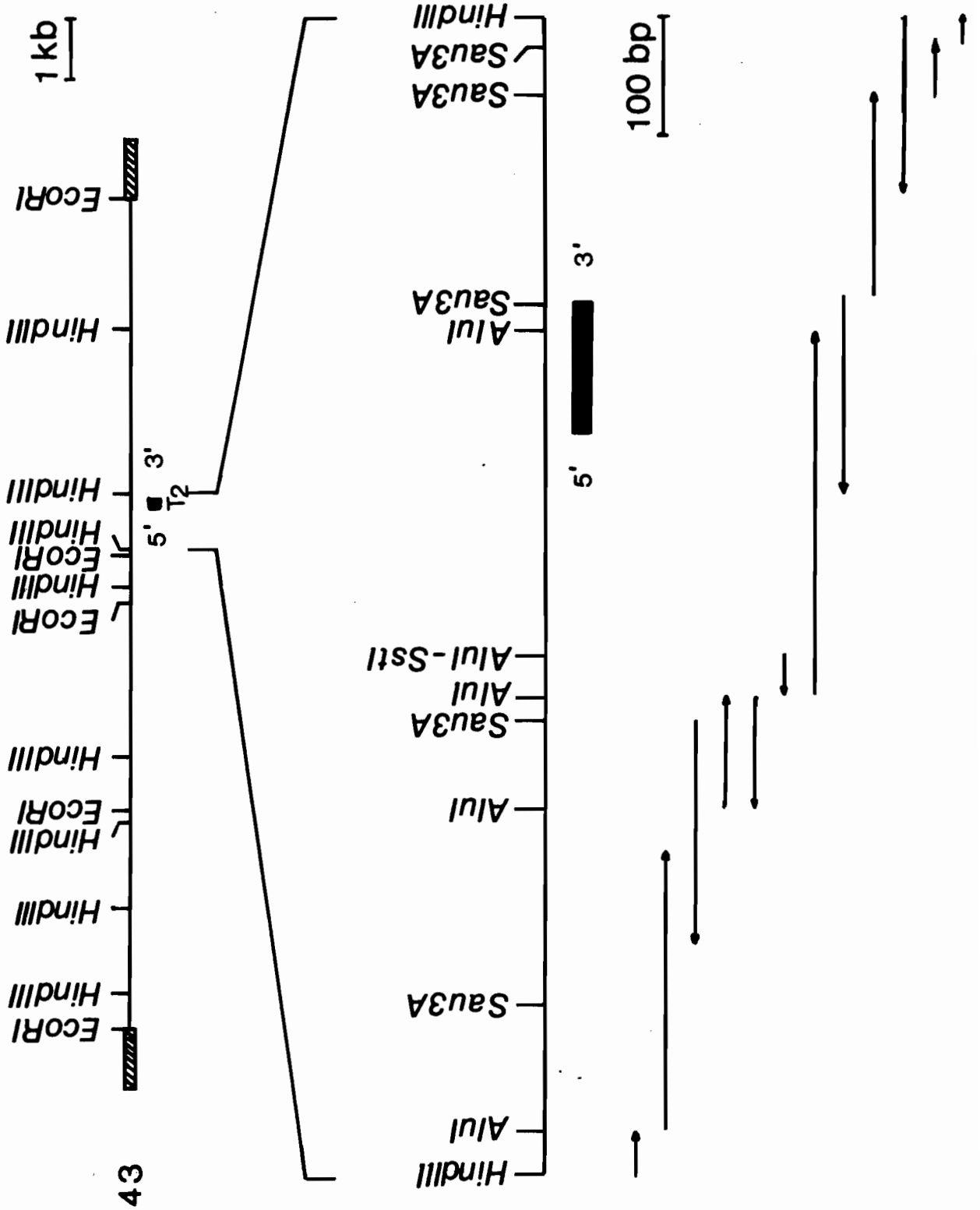

1 kb
EcoRI
HindIII
HindIII
5'
3'
T2
HindIII
EcoRI
HindIII
EcoRI
HindIII
HindIII
EcoRI
HindIII
HindIII
EcoRI
43
HindIII
AluI
Sau3A
Sau3A
AluI
Sau3A
5'
3'
Sau3A
AluI
AluI-SstI
AluI
Sau3A
AluI
Sau3A
AluI
HindIII
100 bp

Figure 10. Comparison of DNA sequence of the LbT_2 gene with that of the LbT_1 gene on Gm4 (Brisson, 1982; Brisson and Verma, 1982). Homologous sequences between two truncated genes are indicated by colons. Dashes (-) represent the sequences that might have been deleted or inserted. The coding region is underlined and the repeated sequence is upperlined.

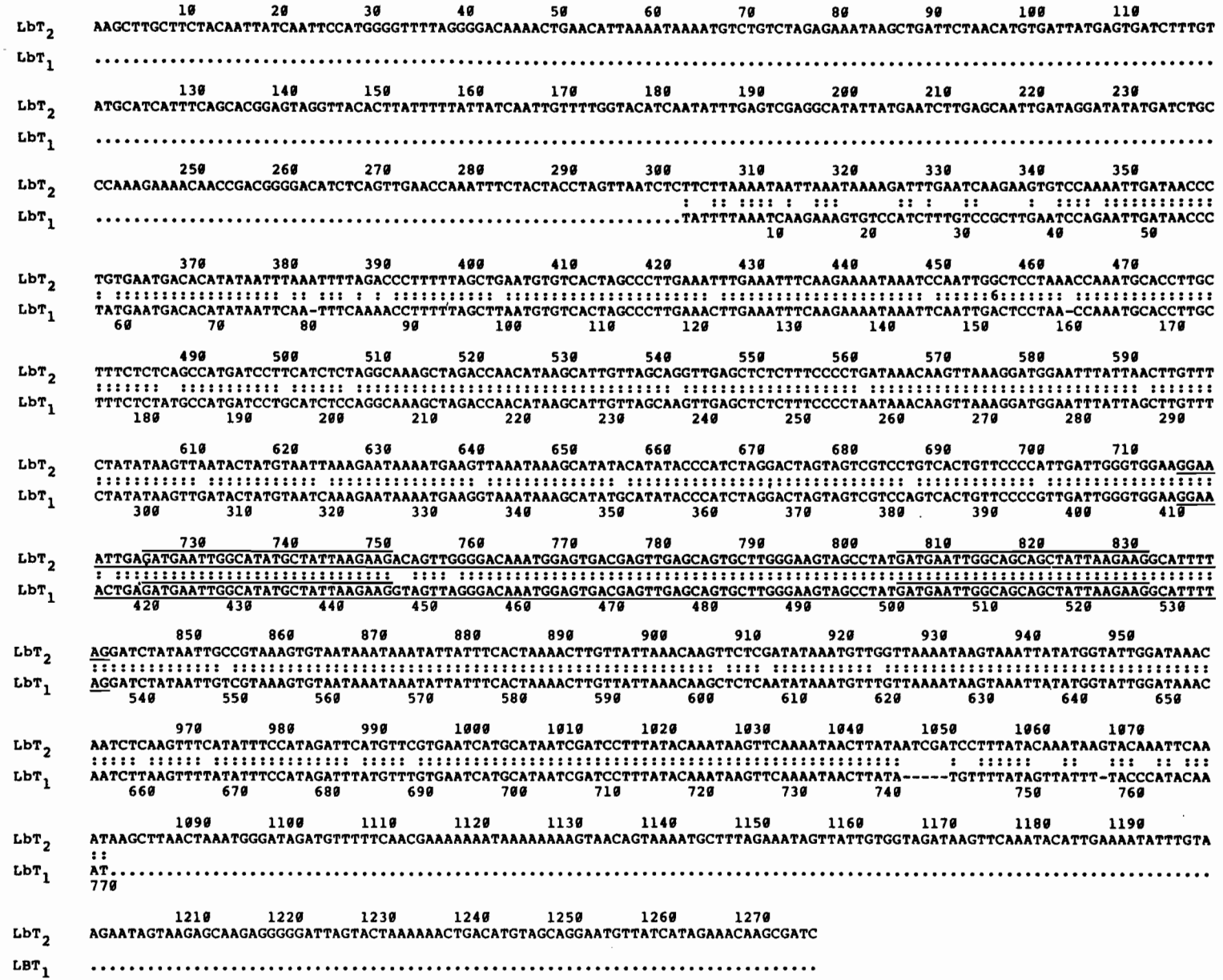
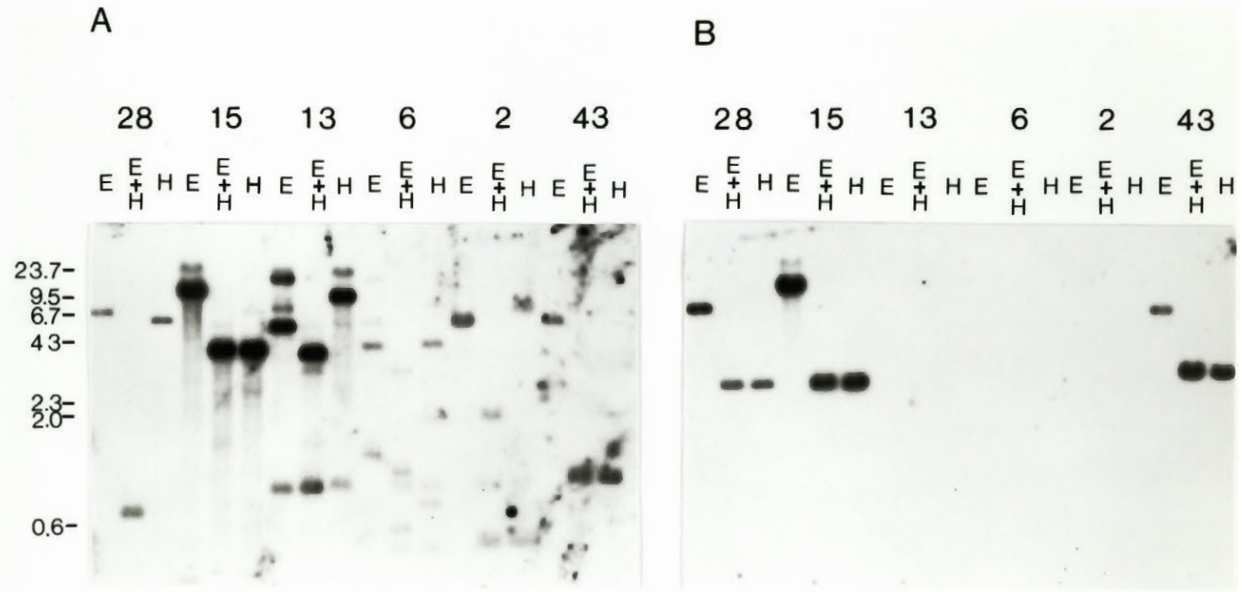

LbT2
LbT1
LBT1

Figure 11. Southern hybridization of genomic clones with a 2.7 kb HindIII fragment from Gml1. DNAs from clones shown on the top were digested with restriction enzyme EcoRI, HindIII or both, separated on the agarose gel, transferred on to GeneScreen and hybridized with nick-translated pLb14 (A) or the 2.7 kb HindIII fragment flanking the 3' side of the main leghemoglobin locus (B) which had been subcloned into pBR322 (indicated by a dotted bracket in Figure 4).


A
B
28
15
13
6
2
43
E
E+H
H
23.7
9.5
6.7
4.3
2.3
2.0
0.6

A homologous region was found 3' to the Lbc_2 locus represented by clones 2, 4, 16Ø and 6ØE (Figure 5A). In addition, hybridization was also detected in the region 3' to the Lb sequence on clone 43 and Gm4, both of which possess a trancated Lb gene (Figure 5B and C). Their locations are marked by an open dotted box (Figure 5A, B and C).

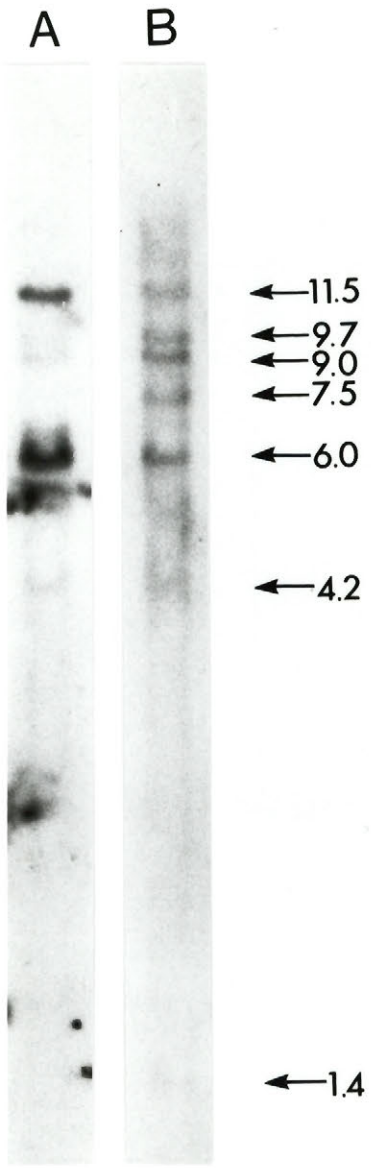
The 2.7 kb HindIII fragment hybridized to all the EcoRI-digested Lb-containing genomic fragments, except to the Lb-containing fragments of 7.5 and 1.4 kb (Figure 12B). These fragments correspond to those containing the Lba, Lbc_1 and $Lb\psi_1$ genes and thus, the results are consistent with the map shown in Figure 4. The hybridization observed in Figure 12A just below the 6.Ø kb band may represent a part of the 2.7 kb HindIII fragment which only hybridizes to a 5.2 kb EcoRI band of clone 6ØE. The 5.2 kb fragment does not hybridize to clones 43 and Gm4 (data not shown), indicating that there are two types of sequences in the 2.7 kb fragment: one is common to all loci, while the other only present on the Lba and Lbc_2 loci.

Only one fragment from the 5' end of the leghemoglobin genes on various recombinant phages was found to cross-hybridize. This is a small HindIII fragment of clones 2 and 4 which hybridized to a 3.4 kb EcoRI fragment 5' to the Lba gene found on clone 7ØE as indicated in Figure 4. This fragment also hybridized to root and leaf cDNAs (see below).

(7). Other genes flanking leghemoglobin loci

The clustering of the genes specifying a functionally related group of proteins is often responsible for coordinate regulation of transcription. Concomitant activation of clustered genes may be achieved in eucaryotes by

Figure 12. Sequences in the soybean genome homologous to the 3' region of the leghemoglobin loci. Southern hybridization of genomic DNA (10 ug) with the 2.7 kb HindIII fragment (A) and leghemoglobin cDNA clone (B). Genomic DNA was digested with EcoRI, separated on the agarose gel, transferred to GeneScreen and hybridized with the radioactively labelled 2.7 kb HindIII fragment (A). The same filter was hybridized with a nick-translated Lb cDNA clone (B). The size shown in kb of each band was determined by comparison of their mobility with HindIII-digested bacteriophage lambda DNA as size marker. The difference in intensity could be due to the difference in homology of sequence.


A
B
11.5
9.7
9.0
7.5
6.0
4.2
1.4

zone-specific modification of chromatin structures. Therefore, in order to see if soybean leghemoglobin genes are located in a unique region of the chromosome which contains genes that are only expressed during nodulation, Southern blots of EcoRI-digested recombinant phage DNAs were sequentially hybridized with root, leaf and nodule cDNAs.

Figure 13 shows that clones 2 and 4 each contain an EcoRI fragment that hybridizes with root and leaf cDNA, and to a less extent with nodule cDNA, indicating that they carry sequences which are expressed in root and leaf tissues. These two clones actually contain the same hybridizing sequence. The difference in the size is due to the fact that one of the EcoRI sites in each clone is formed by an artificial linker (see Figure 5A). Root cDNA hybridized to a restriction fragment from clone 28. This hybridizing region was mapped 3' to the main locus (Figure 4). The root and leaf sequence that hybridizes to clones 2 and 4 maps 5' to the Lbc_2 locus which contains only two leghemoglobin genes (Figure 5A). The 1.5 kb HindIII fragment containing the 5' root/leaf sequence was isolated and hybridized with EcoRI-digested DNA of clones 6 and 70E (data not shown). It showed that this sequence is also present 5' to the main locus (Figure 4). The clone 28 region and the clones 2 and 4 region did not cross-hybridize, indicating the 5' and 3' flanking sequences that are expressed in leaf and root tissues are different.

To determine the uniqueness of these sequences in the genome, EcoRI-digested genomic DNA was hybridized with a sub-fragment from clone 2 or 28 corresponding to the expressed sequences. Two genomic EcoRI fragments appear to contain the 5' root/leaf sequence (Figure 14B). The sizes of these fragments are the same as those of two fragments found in the Lba and Lbc_2 loci. On the other hand, five EcoRI fragments hybridize

Figure 13. Hybridization of EcoRI fragments of clones containing leghemoglobin sequences with cDNA to nodule (A), leaf (B), or root (C). DNA from each recombinant phage numbered at the bottom was digested with EcoRI, separated on the agarose gel, transferred to the GeneScreen and hybridized with cDNA to poly(A)$^+$ RNA. Note the hybridization of fragments in clone 28 (open arrow heads) with root, 2 and 4 (solid arrow heads) with root, leaf and nodule cDNAs. The major hybridizing bands in (A) are due to Lb sequences (8.5 kb as well as other larger fragments which do not correspond to the fragments containing leghemoglobin sequence represent partial digestion as confirmed by the hybridization with Lb cDNA clone, data not shown). The position and size (in kb) of the markers (HindIII-digested bacteriophage lambda DNA) are indicated.

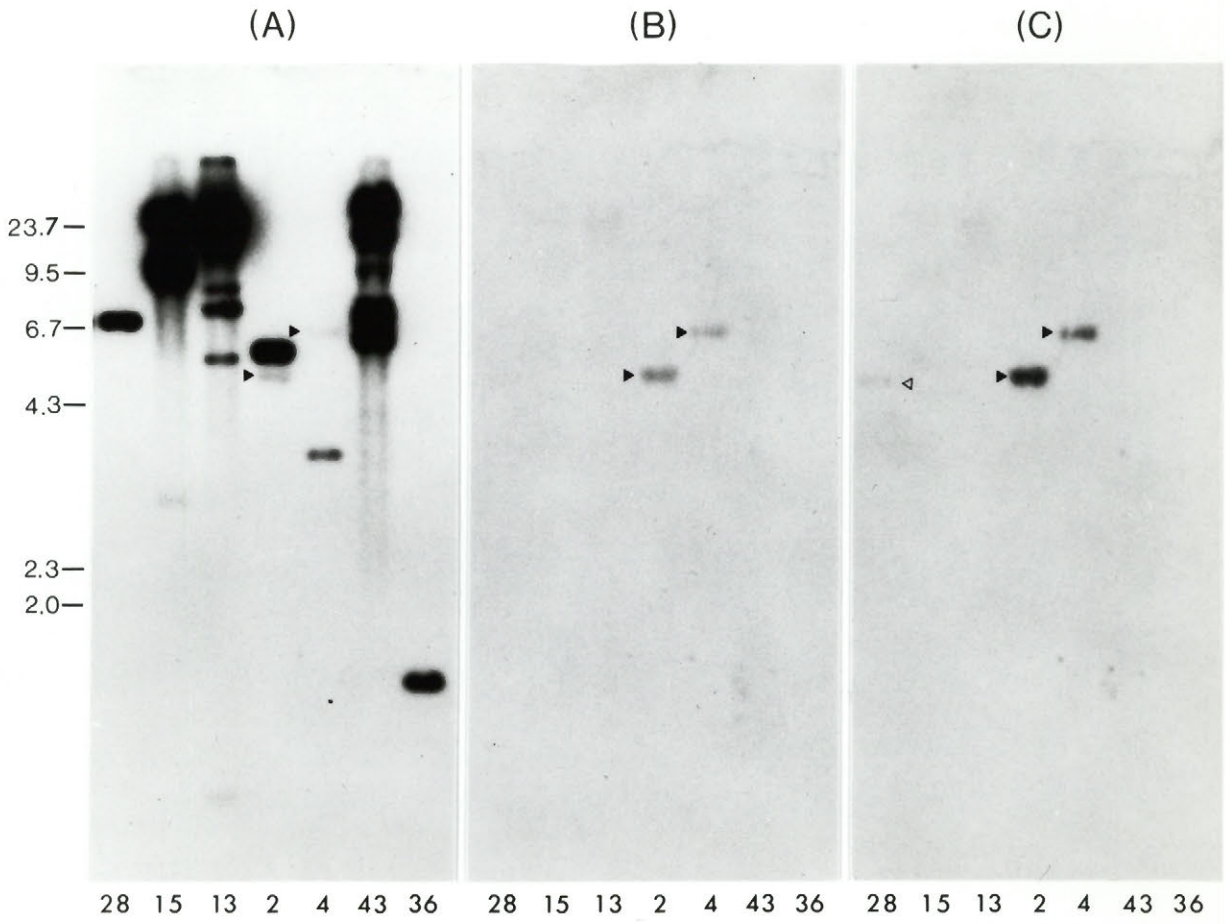
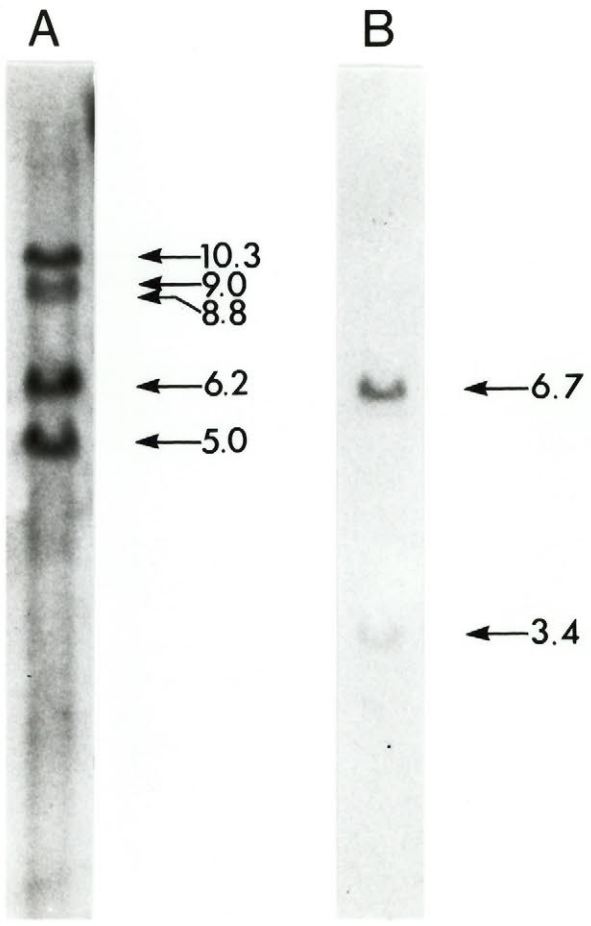

(A)
(B)
(C)
23.7
9.5
6.7
4.3
2.3
2.0
28 15 13 2 4 43 36
28 15 13 2 4 43 36
28 15 13 2 4 43 36

Figure 14. Southern hybridization of genomic DNA with the fragment containing the sequences expressed in root and leaf tissues (see Figure 4). (A) with a 5.0 kb EcoRI fragment from clone 28, (B) with a 1.5 kb HindIII fragment from clone 2. The filter containing EcoRI digested genomic DNA was hybridized with a radioactively-labelled DNA fragment which had been subcloned into pBR325 or pBR322 from clone 28 or clone 2, respectively. The size of each band is shown in kb.


A
B
10.3
9.0
8.8
6.2
5.0
6.7
3.4

intensely with the 3' root sequence (Figure 14A). Thus, the sequence located on the 3' end of the Lb locus appears to be a member of a multigene family. It is not known whether these sequences encode proteins or represent repeated transcribed elements.

B. Structure and Chromosomal Arrangement of the Leghemoglobin genes in Kidney Bean

(1). Construction of a partial cDNA library

Although soybean cloned Lb cDNA cross-hybridizes with the kidney bean genome (Brisson et al., 1982), in order to increase the hybridization efficiency in screening for genomic sequences I constructed a partial cDNA library of kidney bean nodule mRNA. Double-stranded cDNA was prepared from poly(A)$^+$ RNA of 21-day nodules and cloned into the PstI site of pBR322. This yielded about 250 ampicillin-sensitive clones. This partial library was screened with radiolabelled inserts of a soybean Lb cDNA clone, pLb23, constructed by F. Fuller. Of the 250 clones examined, 22 clones hybridized to the probes.

In order to determine the size of the insert of each Lb-hybridizing clone, plasmid DNA was isolated from all Lb cDNA clones, digested with restriction enzyme PstI, and subjected to electrophoresis on 5% polyacrylamide gels. A Lb cDNA clone, pJSLb15, was found to contain an insert of about 720 bp long, representing almost the complete leghemoglobin sequence. Partial nucleotide sequence of the insert of pJSLb15 was determined. The amino acid sequence deduced from the nucleotide sequence (from codon 32 to 130) matched the known amino acid

sequence (Lehtovaata and Ellfolk, 1975b) with only minor diferences (Figure 15; see also below). Nick-translated inserts of pJSLb15 were used as probes for screening a kidney bean genomic library and for genomic hybridizations to determine the number and possible organization of leghemoglobin genes on the kidney bean genome.

(2). Isolation of a kidney bean leghemoglobin gene

A kidney bean MboI partial genomic library, constructed in lambda 1059 as a vector (a gift of J. Slightom) (Sun et al., 1981), was screened with the insert of pJSLb15. Six positive clones were independently isolated from the genomic library. Phage DNA from each clone was prepared and mapped with the restriction enzymes, EcoRI, BamHI and HindIII. It revealed that all the clones are identical. Figure 16 shows Southern hybridization of one of clones (PvLb1) after digestion with various enzymes. A Lb cDNA-hybridizing region was localized on the clone as is shown in Figure 17. Nucleotide sequence of the Lb-containing region on a 2.8 kb BamHI fragment, which had been subcloned on pBR322 was analyzed by using the dideoxy chain termination method as illustrated in Figure 18.

(3). Structure of a kidney bean leghemoglobin gene

The nucleotide sequence of the kidney bean Lb-containing region shown with that of one of the soybean Lb genes is presented in Figure 19. The sequence indicates that the leghemoglobin gene is complete and interrupted by three intervening sequences. They are located at codons 32, 68-69 and 103-104, the same positions as those of soybean leghemoglobin genes (see

Figure 15. Partial nucleotide sequence of the Lb cDNA insert of pJSLb15.

The nucleotide sequence from codon 32 to 130 was determined.

```
............TCG ATA CTG GAG AAA GCA CCA GCA GCA AAG AAC TTG TTT TCA TTT
            Ser Ile Leu Glu Lys Ala Pro Ala Ala Lys Asn Leu Phe Ser Phe

CTT GCT AAT GGA GTA GAT CCC ACT AAT CCT AAG CTC ACG GCT CAT GCT GAA AGC
Leu Ala Asn Gly Val Asp Pro Thr Asn Pro Lys Leu Thr Ala His Ala Glu Ser

CTT TTT GGA TTG GTG CGT GAT TCA GCT GCA CAA CTT CGA GCA AAT GGA GCA GTG
Leu Phe Gly Leu Val Arg Asp Ser Ala Ala Gln Leu Arg Ala Asn Gly Ala Val

GTG GCT GAT GCT GCA CTT GGT TCT ATC CAC TCT CAA AAA GGA GTA AAC GAA TCT
Val Ala Asp Ala Ala Leu Gly Ser Ile His Ser Gln Lys Gly Val Asn Asp Ser

CAG TTT CTG GTG GTA AAA GAG GCT TTG CTT AAA ACA TTA AAG GAA GCA GTT GGA
Gln Phe Leu Val Val Lys Glu Ala Leu Leu Lys Thr Leu Lys Glu Ala Val Gly

GAC AAA TGG ACT GAT GAA CTC AGC ACT GCT CTC GA..........................
Asp Lys Trp Thr Asp Glu Leu Ser Thr Ala Leu Glu
```

Figure 16. Southern hybridization of a genomic clone isolated from an MboI partial library of kidney bean genomic DNA. DNA digested with restriction enzymes EcoRI, BamHI and HindIII as indicated on each lane, was hybridized with the nick-translated insert of pJSLb15.

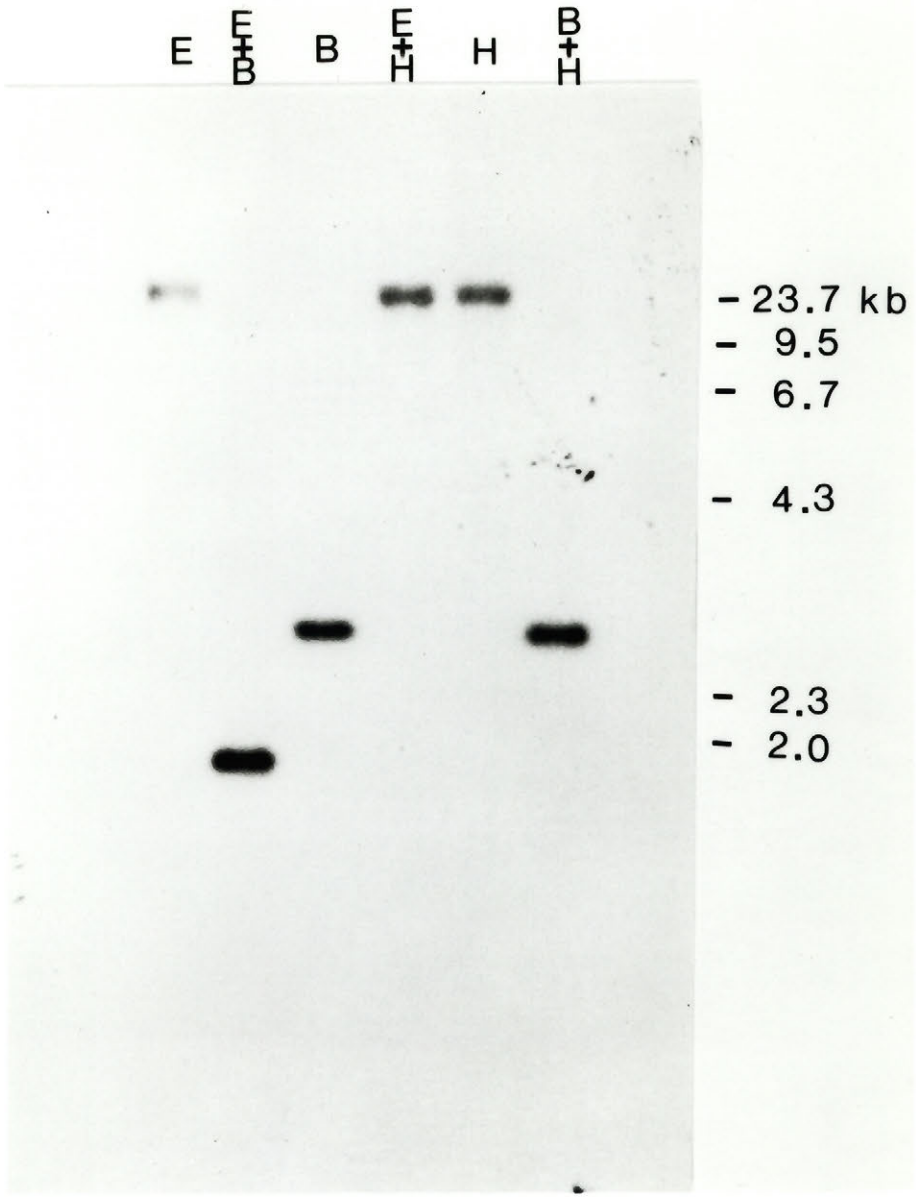
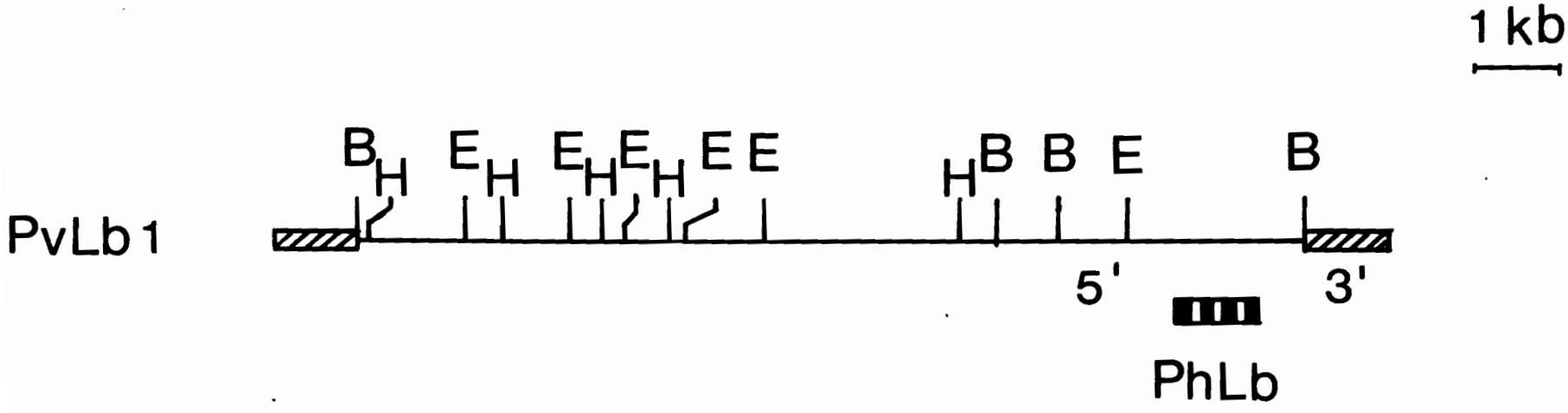

E
E+B
B
E+H
H
B+H
23.7 kb
9.5
6.7
4.3
2.3
2.0

Figure 17. Restriction map of clone PvLb1. The map has been constructed with the restriction enzymes EcoRI, BamHI and HindIII. The hatched boxes indicate lambda arms. A Lb cDNA-hybridizing region was localized on the clone.



B: BamHI

E: EcoRI

H: HindIII

Figure 18. Detailed restriction map of the Lb-containing region on clone PvLb1 and sequencing strategy. A 1.7 kb BamHI-EcoRI fragment of the 2.8 kb BamHI fragment which had been subcloned into pBR322 was subjected to the determination of nucleotide sequence using the dideoxy chain termination method. Horrizontal arrows indicate the extent and direction of sequencing.

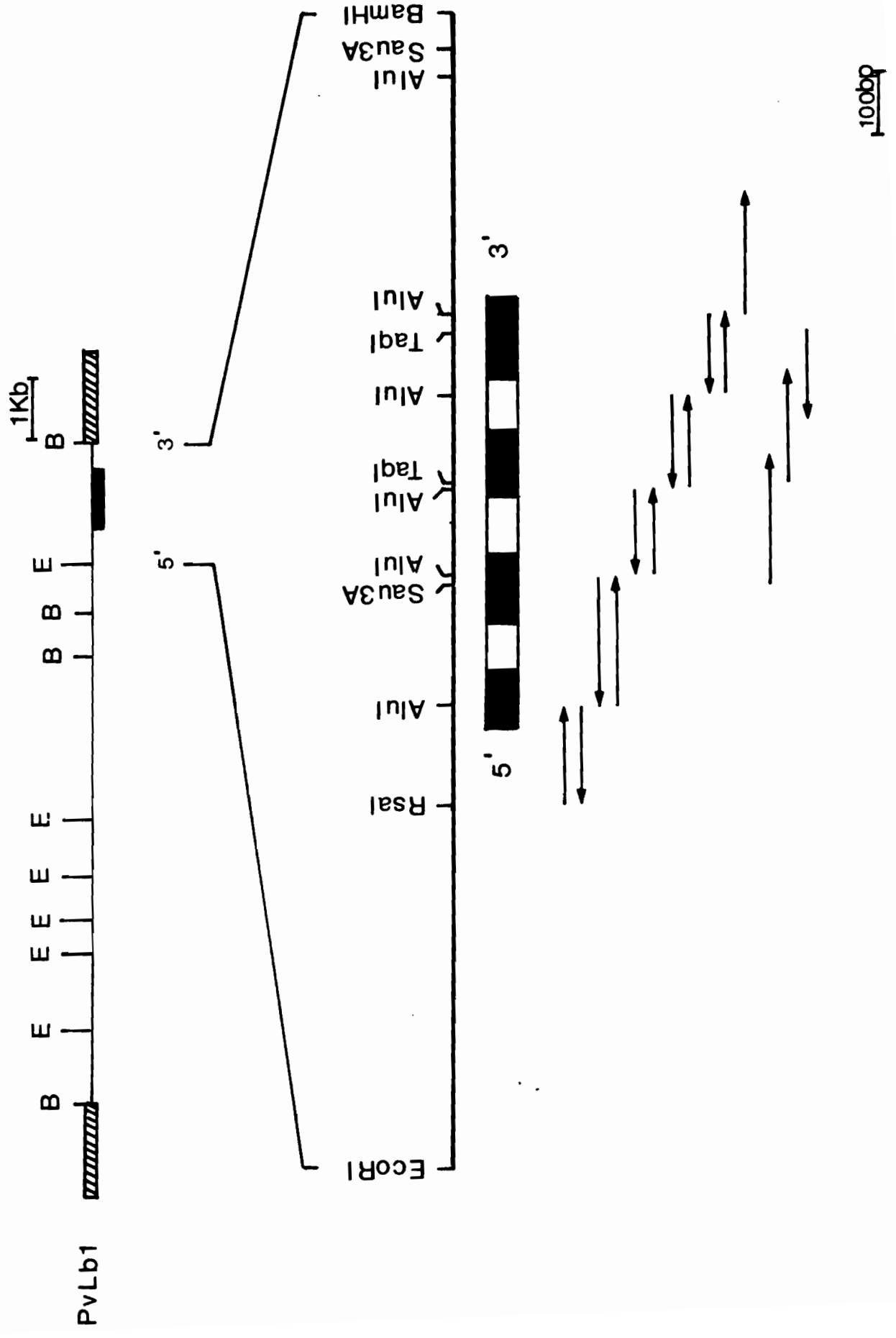
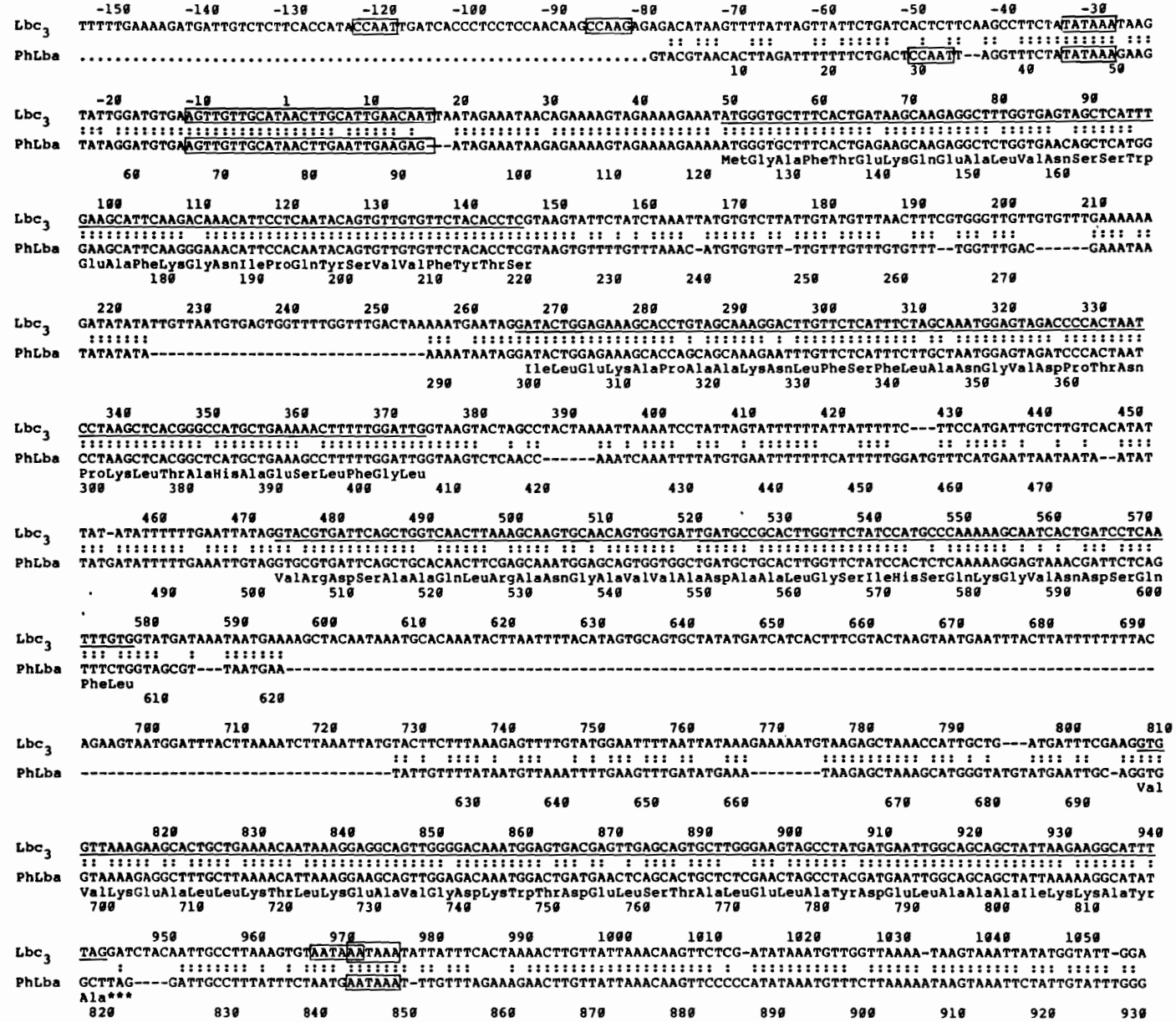

PvLb1
1Kb
B
B E E E E E E B B B E
5'
3'
EcoRI
BamHI
Sau3A
AluI
AluI
TaqI
AluI
TaqI
AluI
Sau3A
AluI
RsaI
AluI
5'
3'
100bp

Figure 19. 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; also, see below) are underlined. 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 kidney bean Lb gene are indicated.


Lbc3
PhLba
MetGlyAlaPheThrGluLysGlnGluAlaLeuValAsnSerSerTrp
GluAlaPheLysGlyAsnIleProGlnTyrSerValValPheTyrThrSer
IleLeuGluLysAlaProAlaAlaLysAsnLeuPheSerPheLeuAlaAsnGlyValAspProThrAsn
ProLysLeuThrAlaHisAlaGluSerLeuPheGlyLeu
ValArgAspSerAlaAlaGlnLeuArgAlaAsnGlyAlaValValAlaAspAlaAlaLeuGlySerIleHisSerGlnLysGlyValAsnAspSerGln
PheLeu
Val
ValLysGluAlaLeuLeuLysThrLeuLysGluAlaValGlyAspLysTrpThrAspGluLeuSerThrAlaLeuGluLeuAlaTyrAspGluLeuAlaAlaAlaIleLysLysAlaTyr
Ala***

Brisson and Verma, 1982). The corresponding intervening sequences are 78, 95 and 88 nucleotides long, respectively. The length of the three intervening sequences of the kidney bean leghemoglobin gene are shorter than those of the corresponding intervening sequences of soybean leghemoglobin genes.

The amino acid sequence deduced from the nucleotide sequence matches the known amino acid sequence of PhLba consisting of 145 amino acids (Lehtovaata and Ellfolk, 1975b), except at positions 98 (Ser-Asn), 99 (Asn-Asp), 100 (Asp-Ser) and 124 (Gln-Glu). These differences were also found in the amino acid sequence deduced from the cDNA sequence (Figure 15), suggesting that the changes at postions 98, 99 and 100 could be due to errors in the protein sequence. The change at position 124 may be due to the transition (CAA to GAA).

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

(4). Hybridization of kidney bean genomic DNA with pJSLbl5

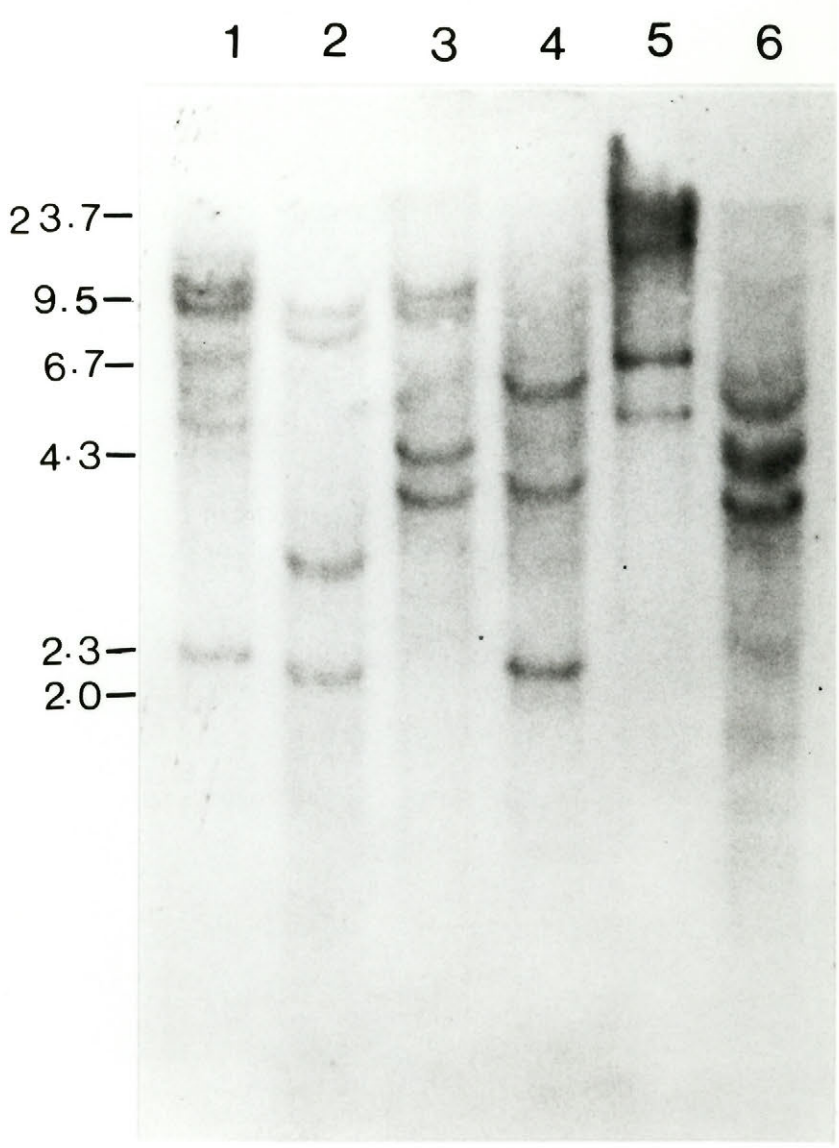
Genomic DNA from kidney bean embryonic axes was digested with various restriction enzymes, separated on a 0.9% agarose gel, trasferred onto GeneScreen paper and hybridized with the nick-translated insert of pJSLbl5. As shown in Figure 20, it revealed several distinct bands, suggesting a family of related sequences. The presence of four hybridizing bands in most lanes, whether from single or double digestions, indicated that there are four leghemoglobin sequences on the chromosome of kidney bean.

In kidney bean, there is only one major leghemoglobin 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 have silenced, or all four genes code for one identical component. It is possible that the increase in the kidney bean genome size resulted in the loss of expression of three genes which are quite isolated from the regulatory sequences that are involved in the derepression of the expression of the leghemoglobin genes in the nodules.

(5). Search for the 3' repeat sequence in kidney bean and in *Glycine soja* genomes

As shown in Figures 4 and 5, four leghemoglobin loci in soybean share a common sequence specific to the 3' end. Therefore, using as a probe the 2.7 kb *Hin*dIII fragment next to the Lbc_3 gene of soybean, I searched genomic DNAs of two other legumes for the presence of this sequence. Kidney bean (*Phaseolus vulgaris*) is closely related phylogenetically to

Figure 20. 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, EcoRI; 2, EcoRI and HindIII; 3, HindIII; 4, EcoRI and BamHI; 5, BamHI; and 6, HindIII and BamHI. The positions and sizes (in kb) of the HindIII-digested lambda DNA are shown.


1
2
3
4
5
6
23.7
9.5
6.7
4.3
2.3
2.0

soybean (Glycine max). They belong to the same tribe Phaseoleae and their leghemoglobins share an extensive homology in amino acid sequence compared to others as shown in Table I. Glycine soja (wild soybean) is the closest ancestral relative of the modern soybean (Hymowitz and Newell, 1980).

Figure 21A shows the hybridization patterns of kidney bean genomic DNA digested with various restriction enzymes and probed with the nick-translated 2.7 kb HindIII fragment. It revealed a single band on each lane, suggesting the presence of only one copy of the sequence. As shown in Figure 21B, however, there are as many as four hybridizing bands in the case of Glycine soja. Since it was found that the 2.7 kb fragment contains two sequences, only one of which is common to all four loci in soybean, the presence of several hybridizing bands from double digestions may indicate that this sequence (found only in the Lba and Lbc_2) is also present on the genome of G. soja. The presence of almost the same number of hybridizing bands with this sequence, as that in soybean, may suggest that leghemoglobin genes in G. soja are arranged on the genome similarly to those of soybean.

(6). 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 at the 5' upstream region which is expressed preferentially in root and leaf tissues. Accordingly, I searched for this sequence on the genomes of two other legume species.

Figure 22 shows the hybridization pattern of genomic DNAs of soybean, kidney bean and G. soja digested with BamHI or EcoRI using as a probe the 1.5 kb HindIII fragment from the Lbc_2 locus. It revealed that G. soja

Figure 21. Southern blots of kidney bean genomic DNA (A) and Glycine soja DNA (B) for the search of the 3' repeat sequence. Genomic DNAs from two other species were digested with combinations of EcoRI, HindIII and BamHI and hybridized with the nick-translated 2.7 kb fragment next to the Lbc_3 gene of soybean (Figure 4). The position and size (in kb) of the markers (HindIII-digested bacteriophage lambda DNA) are shown. Lanes: 1, Soybean DNA cut with EcoRI; 2, EcoRI; 3, EcoRI and HindIII; 4, HindIII; 5, EcoRI and BamHI; 6, BamHI; and 7, HindIII and BamHI.

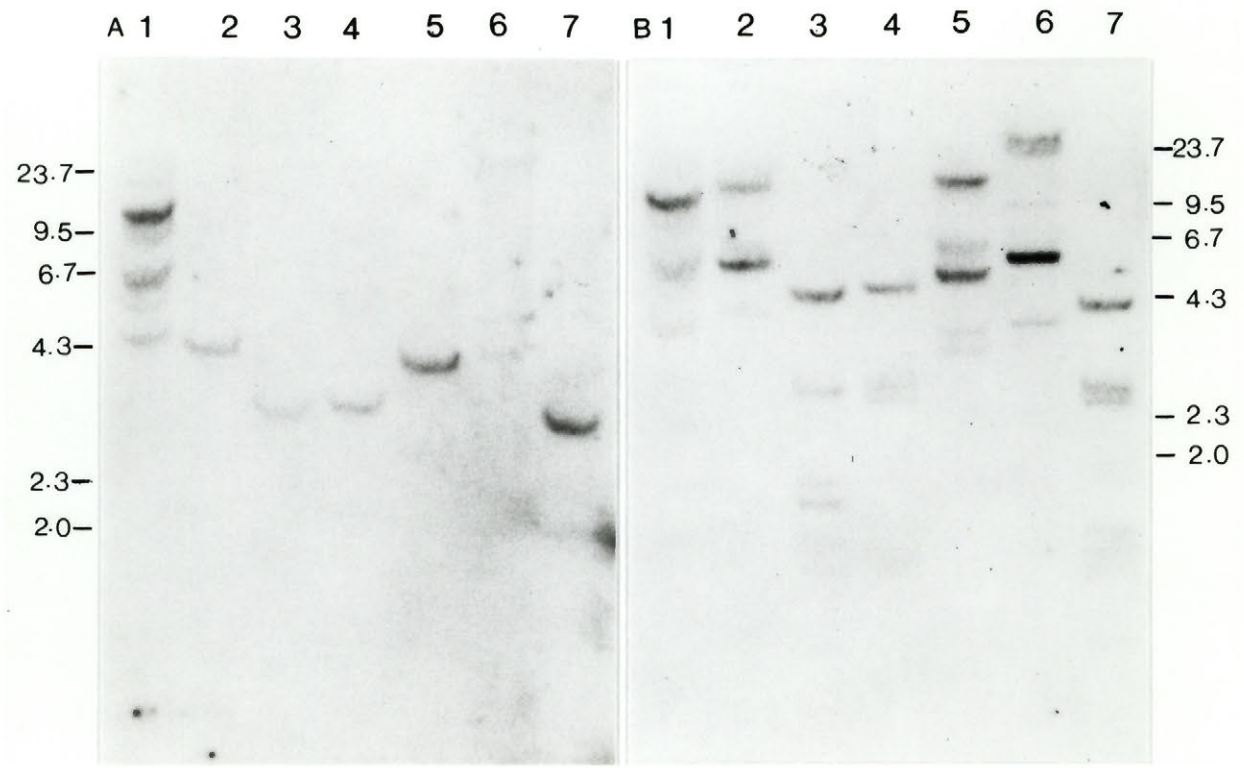
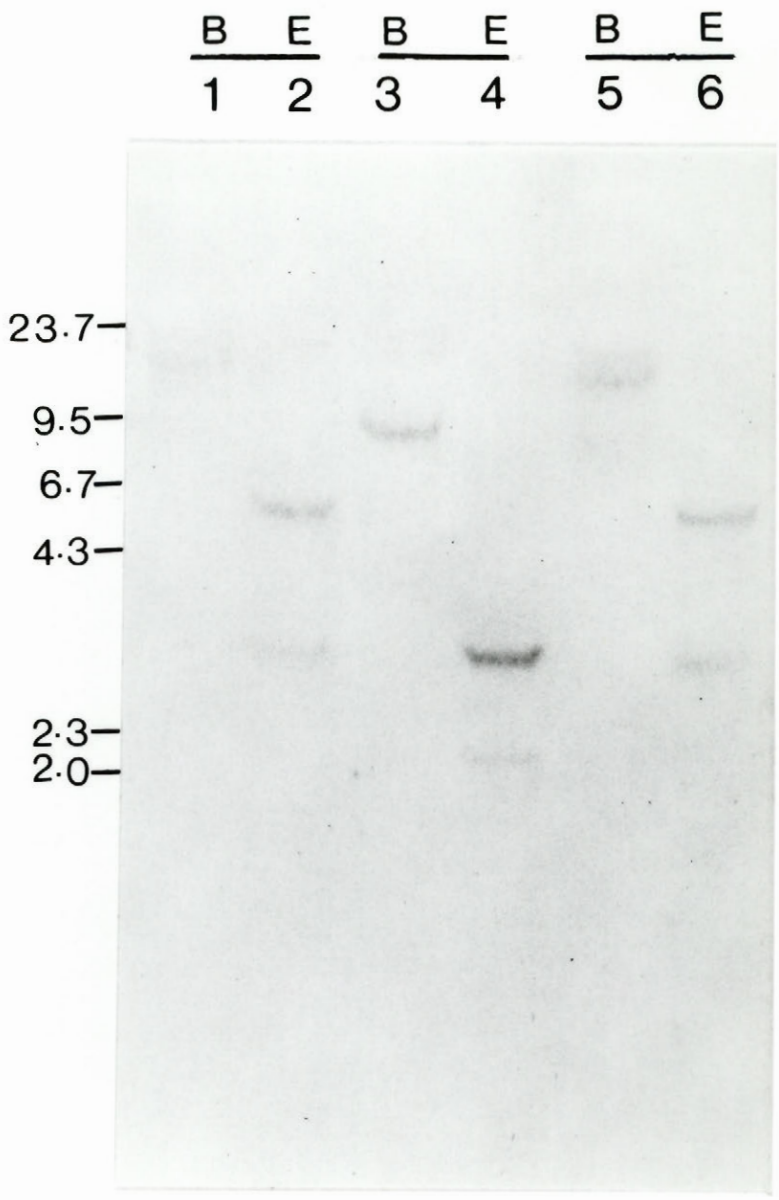

A 1 2 3 4 5 6 7
B 1 2 3 4 5 6 7
23·7
9·5
6·7
4·3
2·3
2·0
23.7
9.5
6.7
4.3
2.3
2.0

Figure 22. Southern blot of genomic DNAs of soybean, kidney bean and *Glycine soja* for the search of the 5' gene expressed in root and leaf tissues of soybean. Genomic DNAs of three species were subjected to Southern hybridization after digestion with *Bam*HI (B) or EcoRI (E). Lanes: 1 and 2, soybean; 3 and 4, kidney bean; 5 and 6, *Glycine soja*. Nick-translated 1.5 kb *Hind*III fragments were used as a probe. The position and size (in kb) of the markers (*Hind*III-digested bacteriophage lambda DNA) are indicated.


B E B E B E
1 2 3 4 5 6
23·7
9·5
6·7
4·3
2·3
2·0

carries two copies of this sequence as does G. max. However, only one copy of this sequence seems to be present on the kidney bean genome. The two small hybridizing fragments on the EcoRI digested lane could be due to polymorphism while one fragment hybridized in the BamHI lane. The sequence was not present at the 5' upstream region of the kidney bean leghemoglobin gene on the genomic clone PvLb1 (Figure 17).

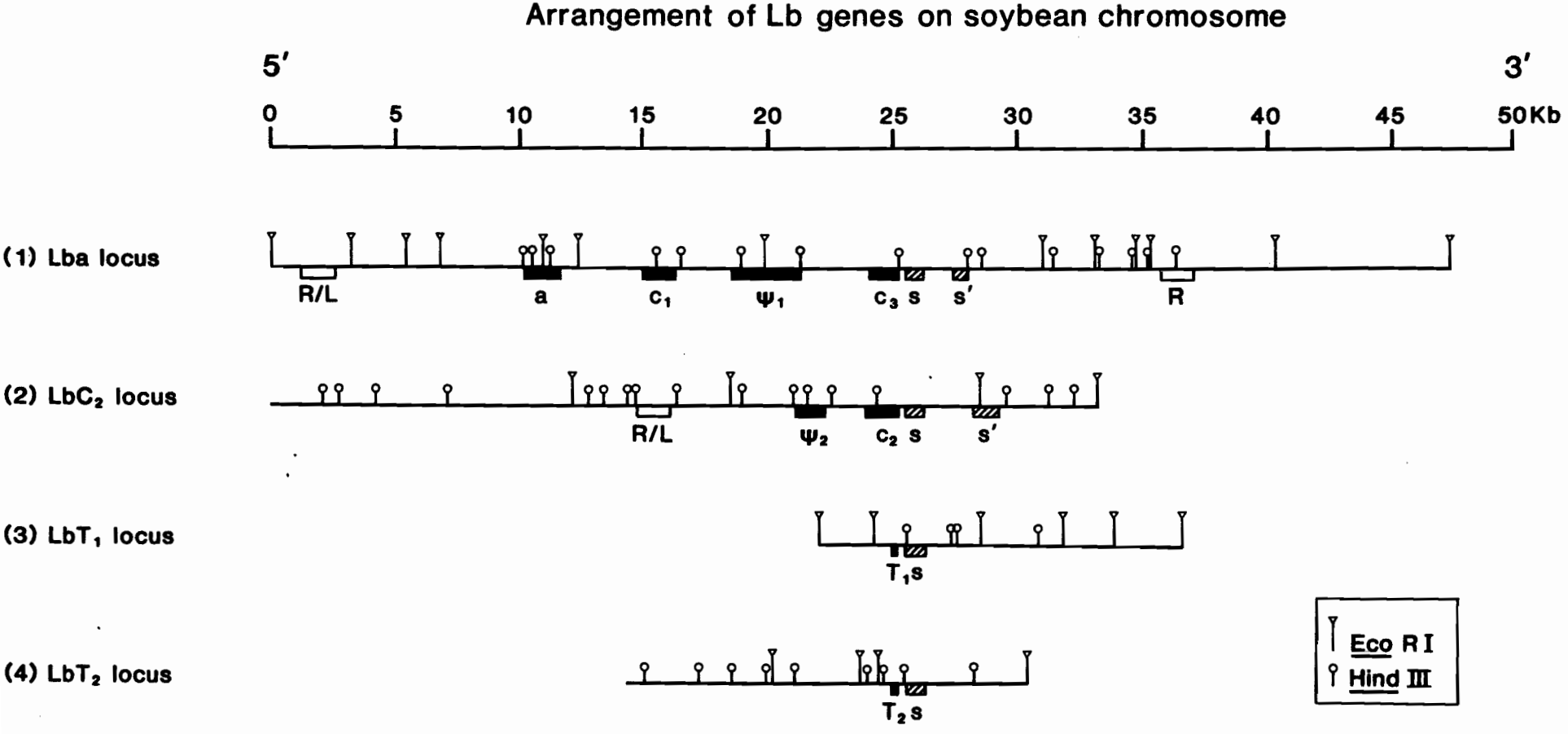
IV. DISCUSSION

A. Chromosomal Arrangement of Leghemoglobin Genes in Soybean

(1). Multiple loci of leghemoglobin genes

Figure 23 summarizes the results from the chromosomal walks around leghemoglobin genes in soybean. There were found four leghemoglobin loci in soybean. In the Lba locus, four leghemoglobin genes are clustered within a region of 15 kb and are interrupted by intergenic regions of an average size of 2.5 kb. In the Lbc_2 locus, two leghemoglobin genes are also clustered with a intergenic region of 2.5 kb. The two truncated leghemoglobin sequences, one (LbT_1) on Gm4 and the other (LbT_2) on clones 36 and 43 (Figures 5B and C), do not appear to be contiguous to the other loci. No detectable rearrangements were observed in the phage containing these truncated sequences, indicating that they represent an <u>in vivo</u> situation. Since no linkage has been found among any of these regions, it appears that leghemoglobin genes exist at multiple loci in soybean. Since the main locus is flanked by two different genes at its 3' and 5' ends which are expressed in tissues other than nodules, these genes seem to demark the boundary of this main locus. The dispersal of the incomplete loci may be due to unequal crossing over in the main locus (Jeffrey and Harris, 1982). The mechanism by which the additional leghemoglobin loci arose is unclear. One intriguing possibility is that these loci appeared initially as a result of tetraploidization events which took place during the recent evolutionary history of soybean (Hadley and Hymowitz, 1973; for more details, see below).

Figure 23. Chromosomal arrangement of leghemoglobin genes in soybean. The EcoRI and HindIII restriction maps of the regions carrying leghemoglobin sequences were derived from the detailed analyses of lambda clones of those 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 Figures 4 and 5 for more details).


Arrangement of Lb genes on soybean chromosome
5′
3′
0
5
10
15
20
25
30
35
40
45
50Kb
(1) Lba locus
R/L
a
c1
ψ1
c3 s
s′
R
(2) LbC2 locus
R/L
ψ2
c2 s
s′
(3) LbT1 locus
T1s
(4) LbT2 locus
T2s
Eco R I
Hind III

In soybean, some members of small multigene families encoding glycinin (Fisher and Goldberg, 1982), a 15 kd protein (Fisher and Goldberg, 1982), lectin (Goldberg et al., 1983) and actin (Meagher et al., 1983) have been isolated from genomic libraries. However, they did not show any close intrafamily linkage within a 10 kb chromosomal domain, suggesting that they are also present in multiple loci on the chromosome and are probably due to the tetraploidization. Also, genes coding for glycinin and 15 kd protein have been found flanked by sequences which are expressed without obvious tissue specificity (Fisher and Goldberg, 1982; also see below).

(2). Flanking regions of the leghemoglobin loci

Two types of sequences have been identified in the flanking region of the leghemoglobin loci. The first is a sequence which is found only adjacent to the 3' ends of some of the leghemoglobin genes. This sequence is not found in any of the intergenic regions, but only at one end of the loci (designated as 's' in Figure 23). The second type of sequences found in the main locus is actually expressed at low levels in various tissues. Two different sequences of this kind were found. The sequence present on the 3' end is a member of a multigene family. However, the other members of this 3' sequence do not appear to be closely associated with leghemoglobin loci. Another expressed sequence is present at the 5' region upstream to the Lba and Lbc_2 loci. Two copies of this 5' sequence are present in the genome, indicating that they are closely associated with two leghemoglobin loci. However, it is not clear why and how these sequences are present in the regions carrying leghemoglobin sequences and what their funtions are.

This observation that there are closely linked genes in soybean chromosomes which are differentially expressed was not unexpected. Calculations based on the soybean genome size, the fraction of single-copy DNA sequences, and the number of active transcription units, predicted an average of one structural gene every 10 kb in soybean chromosomes (Fisher and Goldberg, 1982). A sequence expressed in low levels and without obvious tissue specificity, has also been observed flanking soybean seed protein genes (Fisher and Goldberg, 1982). These findings suggest that each plant gene (family) possesses a cryptic, tightly-linked sequence which is responsible for regulating its expression during development. What these sequences are, and how they interact with the physiological processes inside the cell, remain to be determined.

B. Chromosomal Arrangements of Leghemoglobin Genes in Kidney Bean and Wild Soybean (*G. soja*)

(1). Structure of a leghemoglobin gene in kidney bean

The nucleotide sequence of a kidney bean leghemoglobin gene indicated that the gene is complete and is interrupted by three intervening sequences which are located at the same positions as those in soybean leghemoglobin genes (Brisson and Verma, 1982). However, the length of the three intervening sequences are shorter than those of the corresponding intervening sequences of soybean genes. The difference in the number of introns also found in the kidney bean Lb gene is consistent with the idea that these genes are ancestral to the animal globin genes (Lewin, 1981)

and that introns would have been removed during evolution (Blake, 1983) as have been observed in the cases of the actin genes and the preproinsulin gene of eucaryotes.

The structural similarity between leghemoglobin genes of soybean and kidney bean indicates their close evolutioary relationship. Also, it suggests that they have been placed under certain constraints which prevent them from diverging. Similar comparison between seed storage proteins of soybean and kidney bean (Schuler et al., 1983) showed that although conglycinin and phaseolin genes have diverged from a single ancestral gene, their level of divergence is much greater than that of leghemoglobin genes of two species.

(2). Chromosomal arrangement of leghemoglobin genes in kidney bean

As shown in Figure 2Ø, genomic hybridization indicates the presence of four leghemoglobin sequences in the kidney bean genome. The kidney bean genome also contained two sequences found in the Lba locus of soybean. They are a 3' repeat sequence commonly found in all four loci in soybean and a 5' sequence in root and leaf tissues located at the 5' upstream of the Lba and Lbc_2 loci. These two sequences appear to be closely associated with the leghemoglobin loci. The presence of these two sequences as well as four leghemoglobin genes on the kidney bean genome suggests that four leghemoglobin genes in kidney bean might be arranged in 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 to those of soybean genes. 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 comparison of 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) (Wilbot and Goldberg, 1978) is almost identical to that of kidney bean genome (2n = 22) (1.9×10^9 bp) (Sun et al., 1980; Evans, 1980).

The presence of four Lb genes in the kidney bean genome but only one major Lb component in its nodules suggests that gene duplication may have been followed by silencing other copies by sequence divergence. Kidney bean leghemoglobin may be similar to human myoglobin in this respect. Human myoglobin is encoded by a single gene from a family of related sequences, whose chromosomal arrangement is not yet known (Weller et al., 1984).

(3). Chromosomal arrangement of leghemoglobin genes in wild soybean

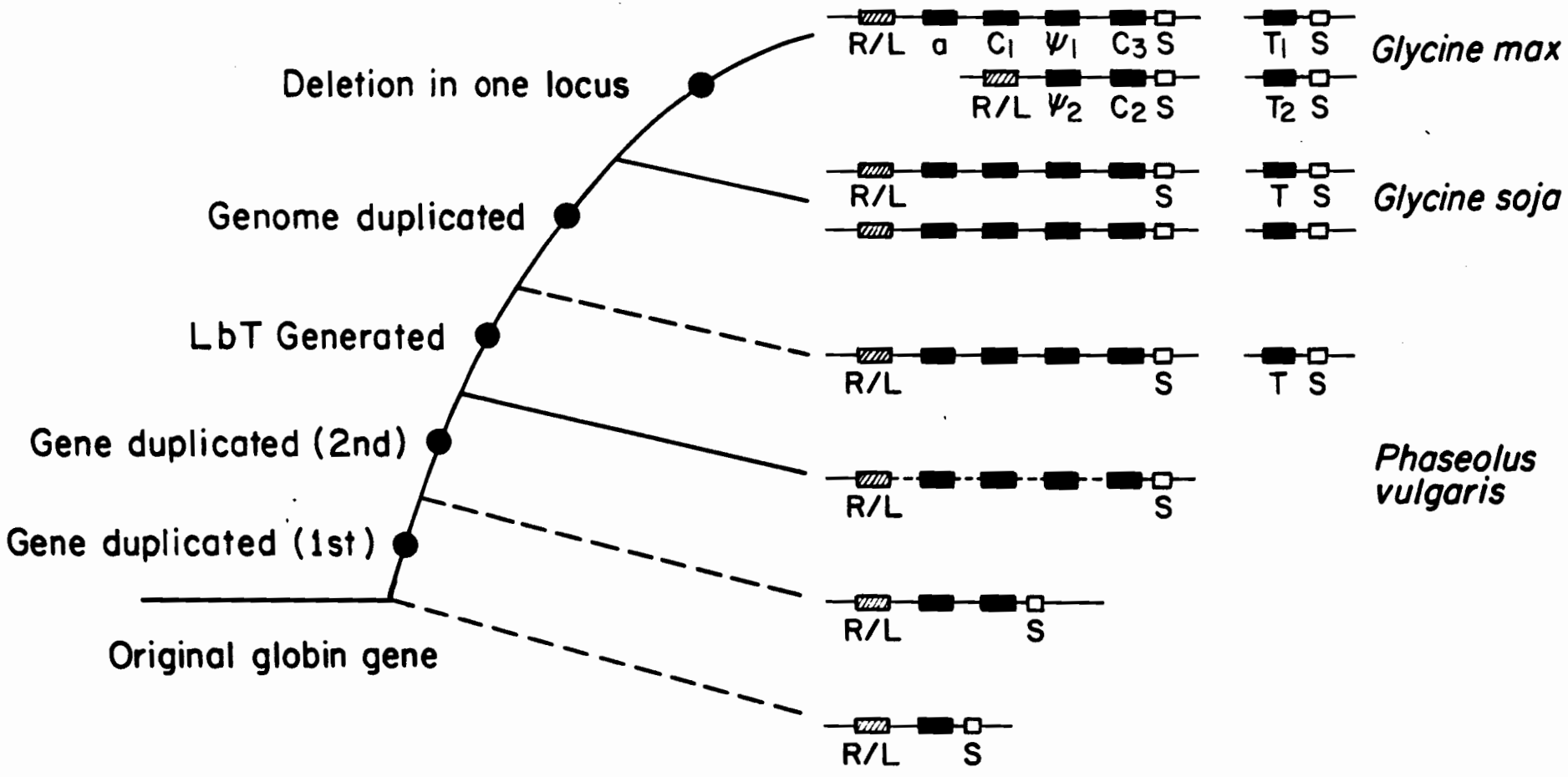
Glycine soja is the closest ancestral relative of the modern soybean (Glycine max). G. soja and G. max both have 40 chromosomes and the two species can be intercrossed (Hymowitz and Newell, 1980). Therefore, I searched the genomic DNA of G. soja for the presence of the 5' and 3' sequences in order to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

C. Evolutionary Relationship Among Leghemoglobin Loci of Three Species

The arrangement of leghemoglobin genes in four loci in soybean raises the question concerning how these four loci have been generated during evolution. Therefore, I compared chromosomal arrangements of leghemoglobin genes of three evolutionarily related species, soybean, kidney bean and wild soybean. The results obtained from this study made it possible to trace the evolution of leghemoglobin gene loci among three species as illustrated in Figure 24. It suggests three major steps in the evolution of leghemoglobin gene loci: (1) gene duplication, (2) genome duplication and finally (3) deletion in one of the Lb loci.

Before Glycine species and kidney bean diverged, a single primordial plant globin gene might have duplicated to generate a locus, carrying four leghemoglobin 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 species 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 was 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. Then the chromosomal duplication by tetraploidization resulted in the generation of two loci, each carrying four leghemoglobin genes, and also two identical truncated genes. This probably occurred before the divergence of two Glycine species. In soybean, a deletion on one four-gene locus removed about 12 kb DNA to generate the Lbc_2 locus

Figure 24. Phylogeny of the leghemoglobin loci. The chromosomal arrangements deduced from this study revealed possible events which could have occurred through the evolutionary time.



PHYLOGENY OF THE LEGHEMOGLOBIN GENE LOCI

carrying a functional and a pseudo (truncated) genes. It is also probable that the deletion event might have occurred before two Glycine species diverged.

This type of evolution is not unique in these 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 have occurred before genome duplication, the extensive homologies among soybean leghemoglobin genes are maintained by concerted evolution (Zimmer et al., 1980; Brown et al., 1984). However, close comparison of the 5' non-coding region of the Lb ψ_2 gene on the Lbc_2 locus with that of the Lba gene revealed 96% homology between the two genes. Also, a similarity was found between the Lbc_2 and Lbc_3 genes with respect to the number of amino acids. They code for one extra amino acid than the Lba and Lbc_1 genes. Thus, a deletion in the main (Lba) locus may have given rise to the Lbc_2 locus. The mechanism by which the deletion occurred is unclear. The presence of a few copies of a short repeat sequence at the breakpoint (see arrow in Figure 19) may suggest the possibility of the involvement of these sequences in homologous recombination. However, it is rather unlikely because coding regions could be involved in homologous recombination as in the human β-globin gene cluster (Flavell et al., 1978). Instead, it is more probable that the deletion is the result of a non-homologous exchange during replication as recently observed in the human β- globin cluster (Vanin et al., 1983).

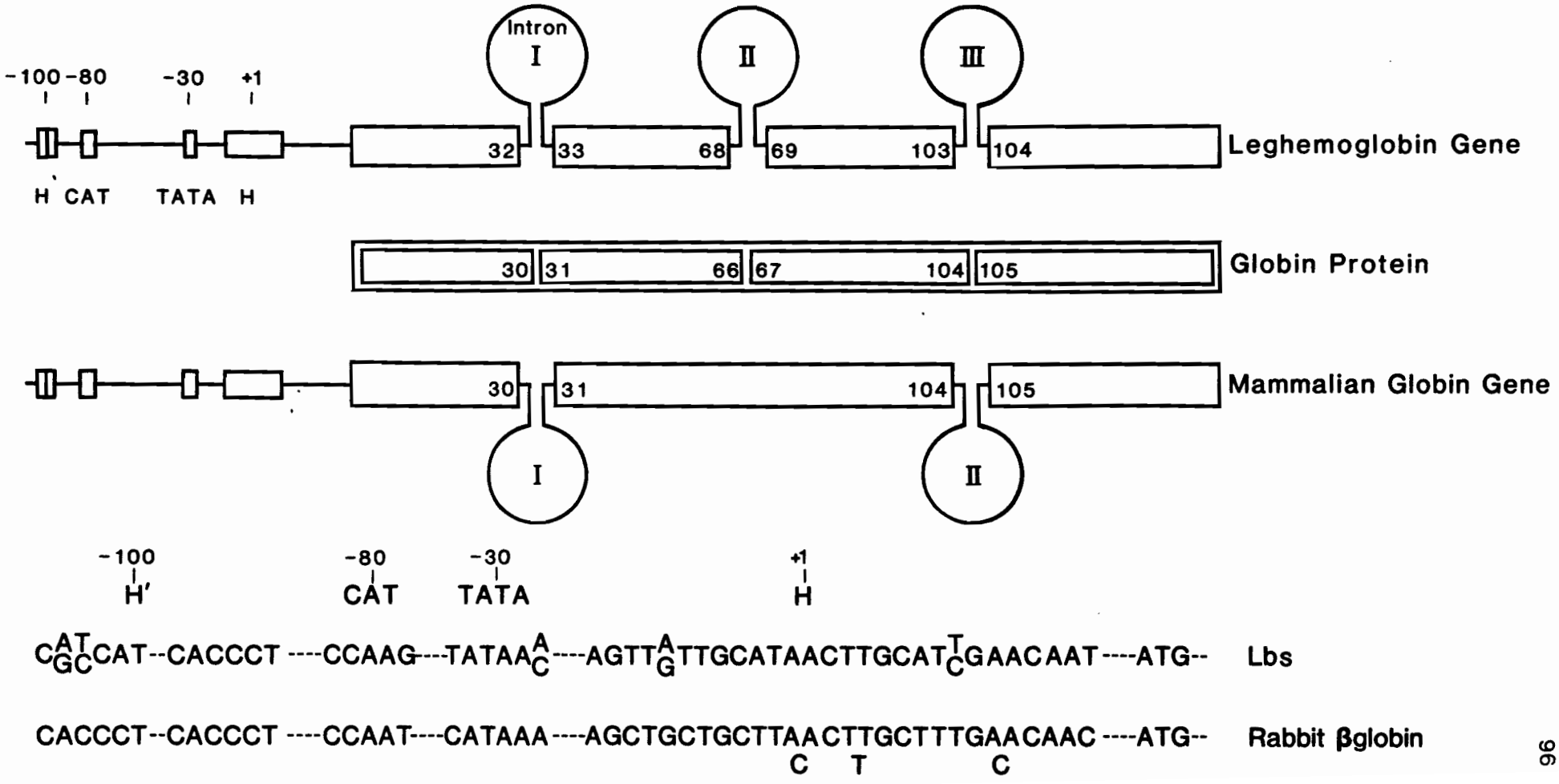
Polyploidy creates an abundance of raw genetic material, which can be exploited by subsequent mutation and selection (Ohno, 1970). It has been suggested that after polyploidization there is a strong tendency to evolve into a diploid state by chromosomal rearrangement and by sequence diversification (Leipoldt and Schmidtke, 1982). This diploidization of the tetraploid state occurs also on the level of the quantitation of gene expression. The deletion in the Lbc_2 locus 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 reduced rate although their multiplicity, variability and temporal expression indicate a possible distinct role for each component during nodule development (Fuchsman et al., 1976; Fuchsman and Appleby, 1979; Verma et al., 1979).

D. Comparison of Leghemoglobin Genes with Animal Globin Genes

(1). Structural comparison

The comparison of the structure of plant and animal genes shows that the intragenic organization of the two globin genes is very similar as shown in Figure 25. In addition to the two intervening sequences, common to all animal genes, the leghemoglobin genes contain a third, central intervening sequence. The fact that an α-globin pseudogene lacking two intervening sequences has been found (Nishioka et al., 1980), may indicate that the excision of a central intron had occurred before the myoglobin and hemoglobin genes were duplicated (Blanchetot et al., 1983). Thus, the leghemoglobin genes probably resemble the primitive globin gene.

Figure 25. Positions of the introns in leghemoglobins and globins in relation to the globin structural units as determined by Gō (1981) and the presence of two homologous sequences (H and H') in the 5' region of these genes. H indicates consensus sequences derived from the region surrounding the putative 'cap' site of β-globin genes (Efstratiadis et al., 1980) and four leghemoglobin genes (see Brown et al (1984) for details). For H', see Dierks et al. (1983).


Intron
I
II
III
-100 -80 -30 +1
32 33 68 69 103 104
H CAT TATA H
Leghemoglobin Gene
30 31 66 67 104 105
Globin Protein
30 31 104 105
I
II
Mammalian Globin Gene
-100 -80 -30 +1
H' CAT TATA H
C(AT/GC)CAT--CACCCT ----CCAAG---TATAA(A/C)----AGTT(A/G)TTGCATAACTTGCAT(T/C)GAACAAT----ATG-- Lbs
CACCCT--CACCCT ----CCAAT----CATAAA----AGCTGCTGCTTAACTTGCTTTGAACAAC ----ATG-- Rabbit βglobin
C T C

Although no significant stretches of homology between the leghemoglobin and animal globin genes could be identified in either the introns or the 3' flanking regions, except that of the polyadenylation signal (Brisson and Verma, 1982; Hyldig-Nielsen et al., 1982), several stretches of homology were observed in the 5' non-coding region. In addition to the general eucaryotic regulatory sequences, TATAAA and CCAAT, two regions of possible regulatory significance specific to globin families were found in the 5' non-coding region. A homology (about 30 bp) (designated as H in Figure 25) is found in the region surrounding the 'cap' site (see Brown et al., 1984). This conserved sequence appears to be specific to globin genes since such a homology is not apparent around the 'cap' site of other animal genes. The other sequence, found at the 5' region, is an imperfect tandemly repeated sequence (designated as H' in Figure 25) at the -100 region which has been suggested to be essential for optimal promoter functions in rabbit β-globin genes (Dierks et al., 1983). The conservation of these two sequences in plant and animal globin genes may suggest that they are essential for the expression of globin genes.

(2). Organization of plant and animal globin gene families

As shown in this study, leghemoglobin genes are located in four regions on the soybean chromosome. The main locus, the Lba locus, carries four leghemoglobin genes which are interrupted by intergenic regions of an average size of 2.5 kb. These intergenic regions are shorter than those in most animal globin loci (Efstratiadis et al., 1980). However, the organization of leghemoglobin genes in the main locus is very similar to a

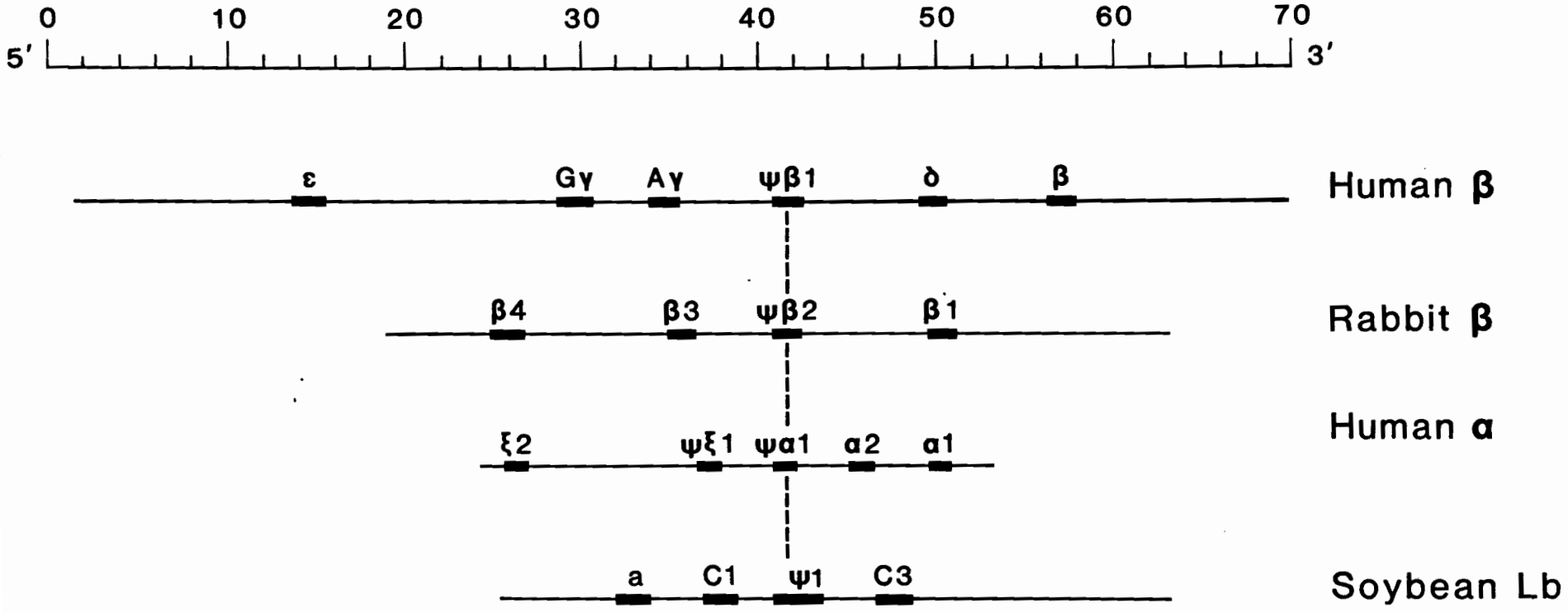
typical animal globin locus (Figure 26). As is the case with many other globin loci (Efstratiadis et al., 1980), this locus consists of a central pseudogene flanked by related true genes. Unlike the pseudogenes in most animal globin loci, Lb ψ_1 does not show high homology to a particular member of the leghemoglobin gene family. This suggests that the duplication event which gave rise to this sequence, occurred independently to or concomitant with, the other leghemoglobin gene duplication events.

The temporal sequence of induction of animal globin loci is in the 5' to 3' direction (Efstratiadis et al., 1980). In the Lba locus in soybean however, the temporal induction occurs in the reverse direction. The Lbc_3 gene, which is located on the 3' end of the locus, appears to be induced before the Lba gene (Verma et al., 1979; Fuchsman and Appleby, 1979), which is located on the 5' end. The reason for this difference between plant and animal globin loci is not apparent. Since active leghemoglobin genes are located in two loci, these two loci may share a unique sequence which may be involved in the interaction with a Rhizobium gene product. Also, it suggests that the expression of leghemoglobin genes is regulated at two levels; the separate induction of two leghemoglobin loci and the temporal induction of legehmoglobin genes in the main locus.

(3). Evolution of plant and 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, and the monomeric leghemoglobins in legumes (Hunt et al., 1978). Although plant and animal globin genes may have diverged between 900 million and

Figure 26. Comparison of the organization of the leghemoglobin main locus with representative mammalian globin gene loci (Efstratiadis et al., 1980). The genes have been aligned for comparison. The direction of transcription is left and scale in kb.


0
10
20
30
40
50
60
70
5′
3′
ε
Gγ
Aγ
ψβ1
δ
β
Human β
β4
β3
ψβ2
β1
Rabbit β
ξ2
ψξ1
ψα1
α2
α1
Human α
a
C1
ψ1
C3
Soybean Lb

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 may suggest that they have been under evolutionary constraints, rather than that they have evolved similarly by accident.

It is generally believed that duplication of hemoglobin and myoglobin genes had occurred about 700 million years ago before the divergence of the alpha and beta globin genes about 500 million years ago. In contrast with hemoglobin, which functions as a tetrameric protein, myglobin is monomeric and serves to facilitate the diffusion of oxygen within muscle fibers. In some diving mammals such as whales and seals, elevated level of myoglobin also acts as a significant muscle oxygen store (Wittenberg, 1970). In this respect, it is similar to leghemoglobin. However, the three exon and two intron organization of the myoglobin genes in seal (Blanchetot et al., 1983) and humans (Weller et al., 1984) rules out the possibility that the leghemoglobin gene might have derived from the myoglobin gene (Lewin, 1983).

Recently, it has been reported that a dimeric hemoglobin occurs in the nitrogen-fixing root nodules of Parasponia (Appleby et al., 1983), which is a non-leguminous plant nodulated by strains of Rhizobium. This raises 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, it 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 about 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 replacement, causing base substitutions, is over 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 (or all) higher and lower plant families. 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 this group of 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.

V. CONCLUDING REMARKS

Leghemoglobins in soybean are encoded by a family of closely related sequences (Sullivan et al., 1981) which exist as functional, pseudo or truncated genes (Brisson et al., 1982). This study demonstrated that these leghemoglobin sequences are located in four loci on the chromosome. The structure and chromosomal arrangement of leghemoglobin genes in kidney bean revealed three possible events of evolution which might have resulted in the generation of the four Lb loci in soybean. They are gene duplication, genome duplication by tetraploidization, and a deletion in one of the 2 four gene loci.

Leghemoglobin gene structure was found to be very similar to that of mammalian globin genes with respect to the position of two introns common to all globin genes, as well as several regulatory sequences on the 5' end of these genes. Furthermore, the arrangement of leghemoglobin genes on the chromosome is similar to mammalian globin loci. These suggest their close evolutionary relationship.

The occurance of hemoglobin genes in a non-leguminous plant (Appleby et al., 1983) suggests that they play an important role in symbiosis and also pertains to the origin of plant globin genes. If hemoglobins in the non-leguminous plant and leghemoglobins have overall homology of their tertiary structures and amino acid sequences, and also have the same gene structure including the central intron, then the ancestral globin genes must have occurred in all higher plants rather than having been transferred from an animal. The accumulation of more information on the structures and chromosomal arrangements of globin genes in other

leguminous as well as non-leguminous plants may provide further understanding of the origin, evolution and maintenance of this group of plant genes which are essential for symbiotic nitrogen fixation.

LITERATURES CITED

Appleby, C. A., Tjepkema, J. D. and Trinick, M. J. (1983). Hemoglobin in a nonleguminous plant, Parasponia. Possible genetic origin and function in nitrogen fixation. Science 220, 951-953.

Appleby, C. A., Turner, G. A. and MacNichol, P. K. (1975). Involvement of oxyleghemoglobin and cytochrome P-450 in an efficient oxidative phosphorylation pathway which support nitrogen fixation in Rhizobium. Biochim. Biophys. Acta 387, 461-475.

Atkins, C. A., Rainbird, R. M. and Pate, J. S. (1980). Evidence for a purine pathway of ureide synthesis in N_2-fixing nodules of cowpea (Vigna unguiculata L. Walp.). Z. Pflanzen physiol. 97, 249-260.

Auger, S. and Verma, D. P. S. (1981). Induction and expression of nodule-specific host genes in effective and ineffective root nodules of soybean. Biochemistry 20, 1300-1306.

Auger, S., Baulcombe, D. and Verma, D. P. S. (1979). Sequence complexities of the poly(A)-containing mRNA in uninfected soybean root nodule tissue developed due to the infection by Rhizobium. Biochim. Biophys. Acta 563, 496-507.

Avissar, V. J. and Nadler, K. D. (1978). Stimulation of tetrapyrrole formation in Rhizobium japonicum by restricted aeration. J. Bacteriol. 135, 782-789.

Awonaike, K. O., Lea, P. J. and Miflin, B. J. (1981). The location of the enzymes of ammonia assimilation in root nodules of Phaseolus vulgaris L. Plant Science Lett. 23, 189-195.

Bauer, W. D. (1981). Infection of legumes by rhizobia. Ann. Rev. Plant Physiol. 32, 407-449.

Baulcombe, D. and Verma, D. P. S. (1978). Preparation of a complementary DNA for leghemoglobin and direct demonstration that leghemoglobin is encoded by the soybean genome. Nucl. Acids Res. 5, 4141-4153.

Bendich, A. J., Anderson, R. S. and Ward, B. L. (1980). Plant DNA: Long, pure and simple. In: Genome Organization and Expression in Plants. ed. C. J. Leaver, Plenum Press, New York, pp. 31-33.

Benton, W. D. and Davies, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196, 180-182.

Bergersen, F. J. (1980). Leghemoglobin, oxygen-supply and nitrogen fixation: Studies with soybean root nodules. In: Nitrogen Fixation. eds. W. D. P. Stewart and J. S. Gallon, Academy Press, London. pp. 139-160.

Bergersen, F. J. and Appleby, C. A. (1981). Leghemoglobin within bacteroid-enclosing membrane envelopes from soybean root nodules. Planta 152, 534-543.

Bergersen, F. J. and Briggs, M. J. (1958). Studies on the bacterial component of soybean root nodules: Cytology and organization of host tissue. J. Gen. Microbiol. 19, 482-490.

Bergersen, F. J. and Goodchild, D. J. (1973). Aeration pathways in soybean root nodules. Aust. J. Biol. Sci. 26, 729-740.

Bergersen, F. J. and Nutman, P. S. (1957). Symbiotic effectiveness in nodulated red clover. IV. The influence of the host factors i, and ie upon nodule structure and cytology. Heredity 11, 175-184.

Bergersen, F. J. and Turner, G. L. (1967). Nitrogen fixation by bacteroid fraction of brees of soybean root nodules. Biochim. Biophys. Acta 141, 507-515.

Bergmann, H., Preddie, E. and Verma, D. P. S. (1983). Nodulin-35: A subunit of specific uricase localized in uninfected cells of nodules. The EMBO J. 2, 2333-2339.

Beringer, J. E., Brewin, N. J. and Johnston, A. W. B. (1980). The genetic analysis of Rhizobium in relation to symbiotic nitrogen fixation. Heredity 45, 161-186.

Bisseling, T., Been, C., Klugkist, J., van Kammen, A. and Nadler, K. (1983). Nodule specific host proteins in effective and ineffective root nodules of Pisum sativum. The EMBO J. 2, 961-966.

Blake, C. (1983). Exons - present from the beginning? Nature 306, 535-537.

Blanchetot, A., Wilson, V., Wood, D. and Jeffreys, A. J. (1983). The seal myoglobin gene: An unusually long globin gene. Nature 301, 732-734.

Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. Furlong, L.-A., Grunwald, D. J., Kieffer, D. O., Moore, D. D., Schumm, J. Sheldon, E. L. and Smithies, O. (1977). Charon phages: Safe derivatives of bacteriophages lambda for DNA cloning. Science 196, 161-169.

Bohlool, B. B. and Schimidt, E. L. (1974). Lectins: A possible basis for specificity in the Rhizobium-legume root nodule symbiosis. Science 185, 269-271.

Boland, M. J. and Schubert, K. R. (1982). Purine biosynthesis and catabolism in soybean root nodules: Incorporation of ^{14}C from $^{14}CO_2$ into xanthine. Arch. Biochem. Biophys. 213, 486-491.

Boland, M. J., Farnden, K. J. F. and Robertson, J. G. (1980). In: Nitrogen Fixation. eds. W. E. Newton and W. H. Orme-Johson. University Park Press, Baltimore. Vol. II, p 33.

Boland, M. J., Hanks, J. F., Reynolds, P. H. S., Blevins, D. G., Tolbert, N. E. and Schubert, K. R. (1982). Subcellular organization of ureide biogenesis from glycolytic intermediates and ammonium in nitrogen-fixing soybean nodules. Planta 155, 45-51.

Bond, G. (1967). Fixation of nitrogen by higher plants other than legumes. Ann. Rev. Plant Physiol. 18, 107-126.

Bond, G. (1974). Root nodules symbioses with actinomycete-like organisms. In: The Biology of Nitrogen Fixation. ed. A. Quispel. North-Holland Pub. Co., Amsterdam-Oxford, pp. 342-380.

Brill, W. (1980). Biochemical genetics of nitrogen fixation. Microbiol. Rev. 44, 449-469.

Brisson, N. (1982). Ph. D. Thesis, McGill University, Montreal, Canada.

Brisson, N. and Verma, D. P. S. (1982). Soybean leghemoglobin gene family: Normal, pseudo and truncated genes. Proc. Natl. Acad. Sci. USA 79, 4055-4059.

Brisson, N., Pombo-Gentile, A. and Verma, D. P. S. (1982). Organization and expression of leghemoglbin genes. Can. J. of Biochem. 60, 272-278.

Broughton, W. J. and Dilworth, M. J. (1971). Control of leghemoglobin synthesis in snake bean. Biochem. J. 125, 1075-1080.

Brown, G. G., Lee, J. S., Brisson, N. and Verma, D. P. S. (1984). Evolution of a plant globin family. J. Mol. Evol. (in press).

Burns, R. C. (1977). Mechanism of dinitrogen reaction. In: A Treatise on Dinitrogen Fixation, Sec. II. R. W. F. Hardy and W. S. Silver, Wiley-Interscience Pub., New York. pp. 491-514.

Burus, R. C. and Hardy, R. W. F. (1975). Description and classification of diazotrophs. In: Nitrogen Fixation in Bacteria and Higher Plants. eds. A. Kleinzeller and H. G. Whittmann. Springer-Verlag, New York. pp. 14-38.

Caldwell, B. E. and Vest, H. G. (1977). Genetic aspects of nodulation and dinitrogen fixation by legumes: The macrosymbiont. In: A Treatise on Dinitrogen Fixation. Sec. III, eds. R. W. F. Hardy and W. S. Silver, Wiley-Interscience Pub. New York. pp. 557-576.

Christeller, J. T., Lainning, W. A. and Sutton, W. D. (1977). Carbon dioxide fixation by lupin root nodules. 1. Characterized association with phosphoenolpyruvate carboxylase, and correlation with nitrogen fixation during nodule development. Plant Physiol. 60, 47-50.

Christensen, T. M. I. E. and Jochimsen, B. U. (1983). Enzymes of ureide synthesis in pea and soybean. Plant Physiol. 72, 56-59.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980). Promoter sequences of eucaryotic protein-coding genes. Science 209, 1406-1413.

Cullimore, J. V., Lara, M., Lea, P. J. and Miflin, B. J. (1983). Purification and properties of the two forms of glutamine synthetase from the plant fraction of Phaseolus root nodules. Planta 157, 245-253.

Cutting, J. A. and Schulman, H. M. (1969). The site of heme synthesis in soybean root nodules. Biochim. Biophys. Acta 192, 486-493.

Dart, P. J. (1974). The infection process. In: The Biology of Nitrogen Fixation. ed. A. Quispel. Elsevier-North-Holland, Amsterdam. pp. 381-429.

Dart, P. J. (1977). Infection and development of leguminous nodules. In: A tratise on dinitrogen fixation. Sec. III, eds. R. W. F. Hardy and W. S. Silver. Wiley-Interscience Pub. New York. pp. 367-472.

Dazzo, F. B. and Hubbell, D. (1975). An antigenic differences between infective and noninfective strains of Rhizobium trifolii. Appl. Microbiol. 30, 172-177.

Dazzo, F. B. and Hubbell, D. (1981). Control of root hair infection. In: Nitrogen Fixation in Rhizobium. ed. W. Broughton, Oxford University Press, London. Vol. 2, pp. 274-310

Dart, P. J. and Mercer, F. V. (1964). The legume rhizosphore. Arch. Mikrobiol. 47, 344-378.

Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J. and Weissman, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell 32, 695-706.

Dilworth, M. J. (1969). The plant as the genetic determination of leghemoglobin production in the legume root nodule. Biochim. Biophys. Acta 184, 432-441.

Dilworth, M. J. and Kidby, D. K. (1968). Localization of iron and leghemoglobin in the legume root nodule by electron microscope autoradiography. Exptl. Cell Res. 49, 148-159.

Dixon, R. O. D. (1967). Ann Bot. N. S. 31, 179

Dullaart, J. (1967). Quantitative estimation of indoleacetic acid and indolecarboxylic acid in root nodules and roote of Lupinus luteus L. Acta Bot. Neerl. 16, 222-230.

Ellfolk, N. (1972). Leghemoglobin, a plant haemoglobin. Endeavour 31, 139-142.

Ellfolk, N. and Sievers, G. (1971). The primary structure of soybean leghemoglobin. Acta Chem. Scand. 25, 3532-3534.

Efstratiadis, A., Posakony, J. W. Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Barelle, R. E., Shoulders, C. C. and Proudfoot, N. J. (1980). The structure and evolution of the human β-globin gene family. Cell 21, 653-668.

Evans, A. M. (1980). Structure, variation, evolution and classification in Phaseolus. In: Advances in Legume Science. eds. R. J. Summerfield and A. H. Bunting, Royal Botanic Gardens, Kew. pp. 337-347.

Fisher, R. L. and Goldberg, R. B. (1982). Structure and flanking regions of soybean seed protein genes. Cell 29, 651-660.

Fuchsman, W. and Appleby, C. A. (1979). Separation and determination of relative concentrations of the homogenous components of soybean leghemoglobins by isoelectric focusing. Biochim. Biophys. Acta 579, 314-324.

Fuchsman, W. H., Barton, C. R., Stein, M. M., Thompson, J. T. and Willett, R. M. (1976). leghemoglobin: Different roles for different components? Biochim. Biophys. Res. Commun. 68, 387-392.

Fujihara, S. and Yamaguchi, M. (1978). Effects of allopurinol on the metabolism of allantoin in soybean plants. Plant Physiol. 62, 134-138.

Fuller, F. and Verma, D. P. S. (1984). Appearance and accumulation of nodulin mRNAs and their relationship to the effectiveness of root nodules. Plant Mol. Biol. 3, 21-28.

Fuller, F., Kunstner, P., Nguyen, T. and Verma, D. P. S. (1983). Nodulin genes of soybean: Analysis of cDNA clones reveals several abundant sequences in nitrogen fixing root nodules. Proc. Natl. Acad. Sci. USA 80, 2594-2599.

Gibson, A. M. (1962). Aust. J. Agric. Res. 13, 388.

Gibson, A. M. (1964). Genetic control of strain-specific ineffective nodulation in Trifolium subterraneum L. Aust. J. Agric. Res. 15, 37-49.

Gō, M. (1981). Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291, 90-92.

Goldberg, M. (1980). Dissertation. Standford University, Standford, CA.

Goldberg, R. B., Hoschek, G. and Vodkin, L. O. (1983). An insertion sequence blocks the expression of a soybean lectin gene. Cell 33, 465-475.

Goodchild, D. J. and Bergersen, F. J. (1966). Electron microscopy of the infection and subsequent development of soybean nodule cells. J. Bacteriol. 92, 204-213.

Greshoff, P. M. and Rolfe, B. G. (1978). Viability of Rhizobium bacteroids isolated from soybean nodule protoplasts. Planta 142, 329-333.

Grunstein, M. and Hogness, D. S. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA 72, 3961-3965.

Gunning, B. E. S., Pate., J. S., Minchin, F. R. and Marks, I. (1974). Quantitative aspects of transfer cell structure in relation to vein loading in leaves and solute transport in legume nodules. Symp. Soc. Exp. Biol. 28, 87-126.

Hadley, H. H. and Hymowitz, T. (1973). Speciation and cytogenetics. In: Soybeans: Improvement, Production and Uses. ed. B. E. Caldwell, (Madison, Wisconsin: American Society of Agronomy), pp. 97-114.

Hanks, J. F., Schubert, K. and Tolbert, N. E. (1983). Isolation and characterization of infected and uninfected cells from soybean nodules. Role of uninfected cells in ureide synthesis. Plant Physiol. 71, 869-873.

Hanks, J. F., Tolbert, N. E. and Schubert, K. R. (1981). Localization of enzymes of ureide biosynthesis in peroxisomes and microsomes of nodules. Plant Physiol. 68, 65-69.

Herridge, D. F., Atkins, C. A., Pate, J. S. and Rainbird, R. M. (1978). Allantoin and allantoic acids in the nitrogen economy of the cowpea (Vigna unguiculata). Plant Physiol. 62, 495-498.

Holl, F. B. (1973). Host-determined genetic control of nitrogen fixation with Pisum-Rhizobium symbiosis. Can. J. Genet. Cytol. 15, 659.

Holl, F. B. (1975). Host plant control of the inheritance of dinitrogen fixation with Pisum-Rhizobium symbiosis. Euphytica 24, 767-770.

Holmes, D. S. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114, 193-197.

Hosbach, H. A., Wyler, T. and Weber, R. (1983). The Xenopus laevis globin gene family: Chromosomal arrangement and gene structure. Cell 32, 45-53.

Hurrell, J. G. R., Nicola, N. A. and Leach, S. J. (1979). Evolutionary and structural relationships of leghemoglobins. In: Soil Microbiology and Plant Nutrition. eds. W. J. Broughton, C. K. John, J. C. Rajarao and Beda Lim. University of Malaya Press, Kuala Lumpur. pp. 253-281.

Hurrell, J. G. R., Thulborn, K. R., Broughton, W. J., Dilworth, M. J. and Leach, S. J. (1977). Leghemoglobins: Immunochemistry and phylogenetic relationships. FEBS Lett. 84, 244-246.

Hunt, L. T., Hurst-Calderone, S. and Dayhoff, M. O. (1978). Globins. In: Atlas of Protein Sequence and Structure. Vol. 5, Suppl. 3, 229-251.

Hyldig-Nielsen, J. J., Jensen, E. O., Paludan, K., Wibog, O., Carrett, R., Jorgensen, P. and Marcker, K. A. (1982). The primary structures of two leghemoglobin genes from soybean. Nucl. Acids Res. 10, 689-701.

Hymowitz, T. and Newell, C. A. (1980). Taxonomy, speciation, domestication, dissemination, germplasm resources and variation in the genus Glycine. In: Advances in legume science. eds. R. J. Summerfield and A. H. Bunting, Royal Botanic Gardens, Kew. pp. 251-264.

Imamura, T., Riggs, A. and Gibson, Q. H. (1972). Equilibria and kinetics of legand binding by leghemoglobin from soybean root nodules. J. Biol. Chem. 247, 521-526.

Jeffreys, A. J. (1982). Evolution of globin genes. In: Genome evolution. eds. G. A. Dover and R. B. Flavell, Academic Press, London. pp. 157-176.

Jeffreys, A. J. and Harris, H. (1982). Processes of gene duplication. Nature 296, 9-10.

Jeffreys, A. J., Wilson, V., Wood, D. and Simons, J. P. (1980). Linkage of adult α- and β-globin genes in X. laevis and gene duplication by tetraploidization. Cell 21, 555-564.

Jegorov, C. A., Feigina, M. I., Kazakov, V. K., Shahparonov, M., Mitaleva, S. V. and Ouchinnikov, I. A. (1976). The amino acid sequences of L. luteus LbI. Biorg. Khim. 2, 125-128.

Jegorov, C. A., Kazakov, V. K., Shahparonov, M. I., Feigina, M. I. and Kostetsky, P. V. (1978). The amino acid sequence of L. luteus LbII. Biorg. Khim. 4, 476-480.

Jensen,, E. O., Paludan, K., Hyldig-Nielsen, J. J., Jorgensen, P. and Marcker, K. A. (1981). The structure of a chromosomal leghemoglobin gene from soybean. Nature 291, 677-679.

Jing, Y., Paau, A. S. and Brill, W. J. (1982). Leghemoglobins from alfalfa (Medicago sativa L. Vernal) root nodules. I. Purification and in vitro synthesis of five leghemoglobin components Plant Sci. Lett. 25, 119-132.

Kefford, N. P., Brockwell, J. and Zwar, J. A. (1960). The symbiotic synthesis of auxin by legumes and nodule bacteria and its role in nodule development. Aust. J. Biol. Sci. 13, 456-467.

Keilin, D. and Wang, Y. L. (1945). Haemoglobin in the root nodules of leguminous plants. Nature 155, 227-229.

Kennedy I. R., Parker, C. A., and Kidby, D. K. (1966). The probable site of nitrogen fixation in root nodules of Ornithopus sativum. Biochim. Biophys. Acta 130. 517-519.

Kidby, D. K. and Goodchild, D. J. (1966). Host influence in the ultrastructure of root nodules of Lupinus luteus and Ornithopus sativum. J. Gen. Microb. 45, 147-152.

Kijne, J. W. (1975). Physiol. Pl. Path. 5, 75.

Kondorosi, A. and Johnston, A. W. B. (1981). The genetics of Rhizobium. Int. Rev. Cytol. supp. 13, pp. 191-224.

Kubo, H. (1939). Uber das Hamoprotein aus den Wurzellknollchen von leguminosen. Acta Phytochim. 11, 195-200.

Lara, N., Cullimore, J. V., lea, P. J., Miflin, B. J., Johnston, A. W. B. and Lamb, j. W. (1983). Appearance of a novel form of plant glutamine synthetase during nodule development in Phaseolus vulgaris L. Planta 157, 254-258.

Lawrie, A. C. and Wheeler, C. T. (1975). Nitrogen fixation in the root nodules of Vicia fava L. in relation to the assimilation of carbon II. the dark fixation of carbondioxide. New Phytol. 74, 429.

Lea, P. J. and Miflin, B. J. (1980). Transport and metabolism of asparagine and other nitrogen compounds within the plant. in the biochemistry of plants. ed. B. J. Miflin, Vol. 5, pp. 569-607.

Lee, J. S., Brown, G. G. and Verma, D. P. S. (1983). Chromosomal arrangement of leghemoglobin genes in soybean. Nucl. Acids. Res. 11, 5541-5553.

Legocki, R. P. and Verma, D. P. S. (1979). A nodule-specific plant protein (Nodulin-35) from soybean. Science 205, 190-193.

Legocki, R. P. and Verma, D. P. S. (1980). Identification of "nodule-specific" host proteins (Nodulins) involved in the development of Rhizobium-legume symbiosis. Cell 20, 153-163.

Lehtovaara, P. and Ellfolk, N. (1975a). Purification and properties of Phaseolus vulgaris leghemoglobin (PhLb). Acta Chem. Scand. B29, 56-64.

Lehtovaara, P. and Ellfolk, N. (1975b). The amino-acid sequence of leghemoglobin component a from P. vulgaris (Kidney bean). Eur. J. Biochem. 54, 577-584.

Lehtovaara, P., Lappalainen, A. and Ellfolk, N. (1980). The amino acid sequence of pea (Pisum sativum) leghemoglobin. Biochim. Biophys. Acta 623, 98-106.

Leipoldt, M. and Schmidtke, J. (1982).Gene expression in phylogenetically polyploid organizms. In: Genome Evolution. eds. G. A. Dover and R. B. Flavell, Academic Press, London, pp. 219-236.

Leong, S. A., Ditta, G. S. and Helinski, D. R. (1982). Heme biosynthesis in Rhizobium. J. Biol. Chem. 257, 8724-8730.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980). Are snRNPs involved in splicing? Nature 283, 220-224.

Lewin, R. (1981). Evolutionary history written in globin genes. Science 214, 426-429.

Lewin, R. (1983). Myoglobin gene is a big surprise. Science 219, 1312.

Libbenga, K. R. and Bogers, R. J. (1974). Root-nodule morphogenesis. In: The Biology of Nitrogen Fixation. ed. A. Quispel. North-Holland Pub. Amsterdam. pp. 430-472.

Libbenga, K. R. and Torrey, J. G. (1973). Hormone-induced endoreduplication prior to mitosis in cultured pea nodule cortex cells. Am. J. Bot. 60, 293-299.

Libbenga, K. R., van Iren, F., Bogers, R. J. and Shraag-Lamers, M. F. (1973). The role of hormones and gradients in the initiation of cortex proliferation and nodule formation in Pisum sativum L. Planta 114, 29-39.

Lie, T. A. (1971). Symbiotic nitrogen fixation under stress conditions. Plant soil (special volume), pp. 117-127.

Ljones, T. (1974). The enzyme system. In: The Biology of Nitrogen Fixation. ed. A. Quispel. North-Holland Pub. Co., Amsterdam. pp. 617-638.

Ljunggren, H. (1969). In: Mechanism and Pattern of Rhizobium Invasion into Leguminous Root Hairs. eds. C. Bloms and A. B. Batryckeri, p.18.

Mandel, M. and Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. 53, 159-162.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Conell, C., Quon, D., Sim, D. K. and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell 15, 687-701.

Matsumoto, T., Yatazaa, M. and Yamamoto, Y. (1977). Allantoin metabolism in soybean plants as influenced by grafts, a delayed inoculation with Rhizobium, and a late supply of nitrogen-compounds. Plant and Cell Physiol. 19, 1161-1168.

McCormack, D. K., Farnden, K. J. F. and Boland, M. J. (1982). Purification and properties of glutamine synthetase from the plant cytosol fraction of lupin nodules. Arch. Biochim. Biophys. 218, 561-571.

McDonnell, M. W., Simon, M. N. and Studier, (1977). Analysis of restriction fragments of T4 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. 11Ø, 119-146.

McParland, N., McCarter, L., Artz, S. and Kustu, S. (1981). Nitrogen regulatory locus glnR of entric bacteria is composed of cistrons ntrB and ntrC: Identification of their protein products. Proc. Natl. Acad. Sci. USA 78, 2135-2139.

Meagher, R. B., Hightower, R. C., Shah, D. M. and Mozer, T. (1983). Plant actin is encoded by diverse multigene families. In: Advances in Gene Technology: Molecular Genetics of Plants and Animals. eds. K. Downey, R. W. Voellmy, R. W. F. Ahmad and J. Schultz, Academic Press, New York. pp. 171-187.

Messing, J., Geraghty, D., Heidecker, G., Hu, H.-T., Kridl, J. and Rubenstein, I. (1983). Plant gene structure. In: Genetic Engineering of Plants. eds. T. Kosuge, C. Meredith and A. Hollaender, Plenum Press, New York. pp. 211-227.

Miflin, B. J. and Lea, P. J. (1976). The pathway of nitrogen assimilation in plants. Phytochemistry 15, 873-885.

Nadler, K. D. and Avissar, Y. J. (1977). Heme synthesis in soybean root nodules. Plant Physiol. 60, 433-436.

Nash, D. and Schulman, H. (1976). Leghemoglobin and nitrogenase activity during soybean root nodule development. Can. J. Bot. 54, 2790-2797.

Newcomb, E. H. and Tandan, S. R. (1981). Uninfected cells of soybean root nodules: Ultrastructure suggests key role in ureide production. Science 219, 1394-1396.

Newcomb, W. (1976). A correlated light and electron microscope study of symbiotic growth and differentiation in Pisum sativum root nodules. Can. J. Bot. 54, 2163-2186.

Newcomb, W., Sippel, D. and Petersen, R. L. (1979). The early morphogenesis of Glycine max and Pisum sativum root nodules. Can. J. Bot. 57, 2603-2616.

Nishioka, Y., Leder, A. and Leder, P. (1980). An unusual alpha globin-like gene that has cleanly lost both globin intervening sequences. Proc. Natl. Acad. Sci. USA 77, 2806-2809.

Nutman, P. S. (1954). Symbiotic effectiveness in nodulated red clover. A major gene for ineffectiveness in the host. Heredity 8, 47-60.

Nutman, P. S. (1956). The influence of the legume in root nodule symbiosis. A comparative study of host determinants and functions. Biol. Rev. Camb. Philos. Soc. 31, 109-161.

Nutman, P. S. (1959). Symp. Soc. Exp. Biol. 13, 42.

Nutman, P. S. (1981). Hereditary host factors affecting nodulation and nitrogen fixation. In: Cuurent Perspectives in Nitrogen Fixation. eds. A. H. Gibson and W. E. Newton. Aust. Academy Science Canberra. pp. 194-204.

Ohyama, T. and Kumazawa, K. (1978). Incorporation of ^{15}N into various nitrogenenous compounds in intact soybean nodules after exposure to $^{15}N_2$ gas. Soil Sci. Plant Nutr. 24, 525-533.

Ohyama, T. and Kumazawa, K. (1980). Nitrogen assimilation in soybean nodules I. the role of GS/GOGAT system in the assimilation of ammonia produced by N_2 fixation. Soil Sci. Plant Nutr. 26, 109-115.

Ohno, (1970). Evolution by gene duplication. Springer, Berlin, Heidelberg, New Nork.

Pate, J. S. (1958). Aust. J. Biol. Sci. 11, 516.

Pate, J. S., Gunning, B. E. S. and Briarty, L. (1969). Ultrastructure and functioning of the transport system of the leguminous root nodule. Plant 85, 11-34.

Postgate, J. R. (1982). The fundamentals of nitrogen fixation. Cambridge University Press, Cambridge, U. K.

Proudfoot, N. J. and Brownlee, G. G. (1976). 3' non-coding region sequences in eucaryotic messenger RNA. Nature 203, 211-214.

Proudfoot, N. J., Gil, A. and Maniatis, T. (1982). The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell 31, 553-563.

Radley, M. (1961). Gibberellin-like substances in plants. Nature 191, 684

Reinbothe, H. and Mothes, K. (1978). Urea, ureides and guanidines in plants. Ann. Rev. Plant Physiol. 13, 129-150.

Reynolds, P. H. S. and Farnden, K. J. F. (1979). The involvement of aspartate aminotransferases in ammonium assimilation in lupin nodules. Phytochemistry 18, 1625-1630.

Reynolds, P. H. S., Blevins, D. G., Boland, M. J., Schubert, K. R. and Randall, D. D. (1982). Enzymes of ammonia assimilation in legume nodules: a comparison between ureide- and amide-transporting plants. Physiol. Plant 55, 255-260.

Richardon, M., dilworth, M. J. and Scarven, M. D. (1975). The amino acid sequence of leghemoglobin I from root nodules of broad bean (Vicia faba L.) FEBS Lett. 51, 33-37.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113, 237-251.

Robertson, J. G., Farden, K. J. F., Warburton, M. P. and Banks, J. M. (1975). Induction of glutamine synthetase during nodule development in lupin. Aust. J. Plant Physiol. 2, 265-272.

Robertson, J. G., Littleton, P., Bullivant, S. and Grayston, G. F. (1978). Membranes in lupin root nodules. I. The role of Golgi bodies in the biogenesis of infection threads and peribacteroid membranes. J. Cell. Sci. 30, 129-149.

Sanger, F., Nicklen, S. and Coulsor, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Schubert, K. R. (1981). Enzymes of purine biosynthesis and catabolism in Glycine max. I. Comparison of activities with N_2 fixation and composition of xylem exudate during nodule development. Plant Physiol. 68, 1115-1122.

Schuler, M. A., Doyle, J. and Beachy, R. N. (1983). Nucleotide homologies between the glycosylated seed storage proteins of Glycine max and Phaseolus vulgaris. Plant Mol. Biol. 2, 119-127.

Scott, D. B., Robertson, J. G. and Farnden, K. J. F. (1976). Ammonia assimilation in lupin nodules. Nature 263, 703-708.

Shelp, B. J., Atkins, C. A., Storer, P. J. and Canvin, D. T. (1983). Cellular and subcellular organization of pathways of ammonia assimilation and ureide synthesis in nodules of cowpea (Vigna unguiculata L. Walp.). Arch. Biochem. Biophys. 224, 429-441.

Sievers, G., Huhtala, M-L. and Ellfolk, N. (1978). The primary structure of soybean (Glycine max) leghemoglobin c. Acta Chem. Scand. B 32, 380-386.

Smith, J. D. (1949). The concentration and distribution of haemoglobin in the root nodules of leguminous plants. Biochem. J. 44, 585-591.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Streeter, J. G. (1972). Nitrogen nutrition of field-grown soybean plants 1. Seasonal variations in soil nitrogen and nitrogen composition of stem exudates. Agron. J. 64, 311-314.

Sullivan, D., Brisson, N., Goodchild, B., Verma, D. P. S. and Thomas, D. Y. (1981). Molecular cloning and organization of two leghemoglobin genomic sequences of soybean. Nature 289, 516-518.

Sun, S. M., Slightom, J. L. and Hall, T. C. (1981). Intervening sequences in a plant gene-comparison of the partial sequence of cDNA and genomic DNA of French bean phaseolin. Nature 289, 37-41.

Sutton, W. D. (1983). Nodule development and senescence. In: Nitrogen Fixation: Legumes. ed. W. J. Broughton. Oxford University Press, New York. Vol. 3, pp. 144-212.

Sutton, W. D., Pankhurst, C. E. and Craig, A. S. (1981). The Rhizobium bacteroid state. ed. K. Jeon. Academic Press. Int. Rev. Cytol. Suppl. 13, 149-177.

Tajima, S. and Yamamoto, Y. (1975). Enzymes of purine catabolism in soybean plants. Plant and Cell Physiol. 16, 271-282.

Tempest, D. W., Meers, J. L. and Brown, C. M. (1970). Synthesis of glutamate in Aerobactor aerogens by a hitherto unknown route. Biochem. J. 117, 405-407.

Thimann, K. V. (1936). On the formation of nodules in legume roots. Proc. Natl. Acad. Sci. USA 22, 511-514.

Treisman, R., Orkin, S. H. and Maniatis, T. (1983). Specific transcription and RNA splicing defects in five cloned β -thalassaemia genes. Nature 302, 591-596.

Triplett, E. W., Blevins, D. G. and Randall, D. D. (1980). Allontoic acid synthesis in soybean root nodule cytosol via xanthine dehydrogenase. Plant Physiol. 65, 1203-1206.

Triplett, E. W., Blevins, D. G. and randall, D. D. (1980). Purification and properties of soybean xanthine dehydrogenase Arch. Biochem. Biophys. 219, 39-46.

Truchet, G. (1978). Sur l'etat diploide des cellules du meristeme des nodules radiculaires des legumineruses. Ann. Sc. Nat. Bot. Biol. reg. 19, 3-38.

Vainshtein, B. K., Harutyunyan, E. H., Kuranova, I. P., Borisov, V. V., Sosfenov, N. I., Pavolvsky, A. g., Grebenko, A. I. and Konareva, N. (1975). Structure of leghemoglobin from lupin root nodules at 5 A resolution. Nature 254, 163-164.

Vanin, E. F., Henthorn, P. S., Kioussis, D., Grosveld, F. and Smithies, O. (1983). Unexpected relatioships between four large deletions in the human β-globin gene cluster. 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). The organization of a nuclear DNA sequence from a higher plant: Molecular cloning and characterization of soybean ribosomal DNA. Gene 7, 317-334.

Verma, D. P. S. (1982). Host-Rhizobium interactions during symbiotic nitrogen fixation. In: The Molecular Biology of Plant Development. eds. H. Smith and D. Grierson. Blackwell Pub. pp. 437-466.

Verma, D. P. S. and Bal, A. K. (1976). Intracellular site of synthesis and localization of leghemoglobin in root nodules. Proc. Natl. Acad. Sci. USA 73, 3843-3847.

Verma, D. P. S. and Long, S. (1983). The molecular biology of Rhizobium-legume symbiosis. Int. Rev. Cytol. Sup. 14. ed. K. Jeon. Academic Press, pp. 211-245.

Verma, D. P. S. and Nadler, K. D. (1984). Legume-Rhizobium symbiosis: Host's point of view. In: Genes Involved in Microbe-Plant Interactions. eds. D. P. S. Verma and T. H. Hohn, Springer-Verlag, Wien, New York. (in press)

Verma, D. P. S., Ball, S., Guerin, C. and Wanamaker, L. (1979). Leghemoglobin biosynthesis in soybean root nodules. Characterization of nascent and released peptides and the relative rate of synthesis of major leghemoglobins. Biochemistry 18, 476-483.

Verma, D. P. S., Haugland, R., Brisson, N., Legocki, R. and Lacroix, L. (1981). Regulation of the expression of leghemoglobin genes in effective and ineffective root nodules of soybean. Biochim. Biophys. Acta 653, 98-107.

Verma, D. P. S., Hunter, N. and Bal, A. K. (1978a). Asymbiotic association of Rhizobium with pea epicotyls treated with plant hormone. Planta 138, 107-110.

Verma, D. P. S., Kazazian, V., Zogbi, V. and Bal, A. K. (1978b). Isolation and characterization of the membrane envelope enclosing the bacteroids in soybean root nodules. J. Cell Biol. 78, 919-936.

Verma, D. P. S., Nash, D. T. and Schulman, H. M. (1974). Isolation and in vitro translation of soybean leghemoglobin mRNA. Nature 251, 74-77.

Vest, G. (1970). Rj_3-A gene conditioning ineffective nodulation in soybean. Crop Sci. 10, 34-35.

Vest, G. and Caldwell, B. F. (1972). Rj_4-A gene conditioning ineffective nodulation in soybean. Crop Sci. 12, 692-693.

Viands, D. R., Vance, C. R., Heidel, G. H. and Barnes, D. K. (1979). Crop Sci. 19, 905.

Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tazard, R., Naker, S. P., Chick, W. L. and Gilbert, W. (1978). A bacterial clone synthesizing proinsulin. Proc. Natl. Acad. Sci. USA 75, 3727-3731.

Vincent, J. M. (1974). In: The Biology of Nitrogen Fixation. ed. A. Quispel. North-Holland, Amsterdam. pp. 265-341..

Vincent, (1980). Factors controlling the legume-Rhizobium symbiosis. In: Nitrogen Fixation. eds. W. E. Newton and H. H. Orme-Johnson. Univ. Press, Baltimore. Vol. II. pp. 103-129.

Vogt, V. M. (1973). Purification and further properties of single-strand-specific nuclease from Aspergillus oryzae. Eur. J. Biochem. 33, 192-200.

Wahl, G. M., Stern, M. and Stark, G. R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA 76, 3683-3687.

Walbot, V. and Goldberg, R. (1978). Plant genome organization and its relationship to classical plant genetics. In: Nucleic Acids in Plants, eds. J. Davis and T. Hall. Boca Raton, Fla. CRC Press, pp. 3-40.

Weber, C. R. (1966). Agro. J. 58, 43.

Weller, p., Jeffreys, A. J., Wilson, V. and Blanchetot, A. (1984). Organization of the human myoglobin gene. The EMBO J. 3, 439-446.

Whittaker, R. G., Lennox, S. and Appleby, C. A. (1981). Relatioship of the minor soybean leghemoglobins d_1, d_2, d_3 to the major leghemoglobins c_1, c_2, and c_3. Biochem. Int. 3, 117-124.

Wiborg, O., Hyldig-Nielsen, J. J., Jensen, E. O., Palndan, K. and Marcker, K. A. (1982). The nucleotide sequences of two leghemoglobin genes from soybean. Nucl. Acids Res. 10, 3487-3494.

Wiborg, O., Hyldig-Nielsen, J. J., Jensen, E. O., Paludan, K. and Marcker, K. A. (1983). The structure of an unusual leghemoglobin gene from soybean. The EMBO J. 2, 449-452.

Wickens, M. P., Buell, G. N. and Schimke, R. T. (1978). Synthesis of double-stranded DNA complementary to lysozyme, ovomucoid, and ovalbumin mRNAs. J. Biol. Chem. 253, 2483-2495.

Wilbur, W. J. and Lipman, D. J. (1983). Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA 80, 726-730.

Wittenberg, J. B. (1970). Myoglobin-facilitated oxygen diffusion: Role of myoglobin in oxygen entry into muscle. Physiol. Rev. 50, 559-636.

Wittenberg, J. B., Appleby, C. A. and Wittenberg, B. A. (1972). The kinetics of the reactions of leghemoglobin with oxygen and carbon monoxide. J. Biol. Chem. 247, 527-531.

Wittenberg, J. B., Bergersen, F. J., Appleby, C. A. and Turner, G. L. (1974). Facilitated oxygen diffusion. The role of leghemoglobin in nitrogen fixation by bacteroids isolated from soybean root nodules. J. Biol. Chem. 249, 4057-4066.

Woo, S. L. C. (1979). A sensitive and rapid method for recombinant phage screening. Methods Enzymol. 68, 389-395.

Zimmer, E. A., Martin, S. L., Berverley, S. M., Kan, Y. W. and Wilson, A. C. (1980). Rapid duplication and loss of genes coding for the chains of hemoglobins. Proc. Natl. Acad. Sci. USA 77, 2158-2162.