NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Investigation of Nodulin Genes Expressed During Soybean Root Nodule Development

BY

Sylvia Gabriele Gottlob

Thesis submitted to
the School of Graduate Studies and Research
in partial fulfillment of the requirements for the Ph.D
degree in Biology

University of Ottawa
Ottawa, Ontario

ISBN 0-315-60116-7

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ACKNOWLEDGEMENTS

This thesis would not have been completed without the guidance, encouragement and support of Doug Johnson. His advice and patience were greatly appreciated.

I would like to thank Dr. V. Iyer, Dr. P. Moreau, Dr. M. Tenniswood and Dr. M. McBurney who served as members of my research committee. In addition, I would like to thank Dr. Linda Bonen and Dr. Jiro Hattori for their helpful suggestions, advice, and comments.

Thanks to Madeleine Lévesque, Steve Bertrand, Stephe Baird, Tim Gleeson, and Jiro Hattori for providing a cheerful working environment.

Special thanks go to Rhonda Daman and Kathy MacKenzie without whose efforts this thesis would not have been typed. I would also like to thank Rhonda for her friendship and encouragement.

I would like to thank N.S.E.R.C and Doug Johnson for financial support during the course of these studies.

Thanks also go to Paul Brunon and Jacques Hélie for their assistance in making figures.

Lastly, I would like to thank my family for their encouragement, especially my husband Doug, for his love and support during the course of these studies.

ABSTRACT

The infection of soybean roots by the soil bacterium Bradyrhizobium japonicum results in the formation of specialized organs on the roots called nodules. Within the nodules, a differentiated form of the bacterium (the bacteroid) reduces atmospheric nitrogen to ammonia. Nodule development involves a complex interaction requiring the co-ordinated expression of nodule-specific bacterial (bacteroidin) and nodule-specific plant (nodulin) genes. A number of nodulin gene products from soybean have been characterized, but the functional roles of only a few have been determined.

This thesis describes the isolation and characterization of four nodule-specific cDNA clones (and a leghemoglobin clone) from a cDNA library that was derived from soybean nodule poly $(A)^+$ enriched RNA. mRNA corresponding to these cDNA clones, could be detected in RNA isolated from the soybean nodule; but could not be detected in RNA isolated from the root. The pattern of expression of the genes corresponding to the cDNA clones was investigated at various times following infection of the soybean Glycine max (L.) Merrill by Bradyrhizobium japonicum. It appears that these nodulin mRNAs are expressed just prior to the onset of nitrogen fixation and accumulate to their maximum levels shortly thereafter.

Sequence characterization of the four clones indicated that they represented members of a previously described nodulin multi-gene family: the "Nodulin A" family (Verma and Delauney, 1988). The "Nodulin-A" family represents a group of abundantly transcribed nodulins, whose function is unknown. Three of these clones, 9-11-B, 36-1-A, and 6-9-F, support the previously published sequences of nodulin-23 (Mauro et al.,1985; Sengupta-Gopalan et al., 1986), nodulin-22 (Sandal et al.,1987), and nodulin-44 (Sengupta-Gopalan et al.,1986) respectively. A fourth clone, 15-9-A, represents a new member of this multi-gene family. Northern blot hybridization with oligo-nucleotide probes specific to this cDNA clone, indicates that 15-9-A represents an abundantly transcribed 1 kb mRNA. Our analyses suggest that 15-9-A is a member of a subfamily of three closely related genes within the larger "Nodulin-A" family. 15-9-A is related to the nodulin-20 sequence of Sandal et al. (1987), which our analyses show is also a member of this subfamily. The 15-9-A cDNA is 100% identical to nodulin-20 in the region that is 5' to the position of the nodulin-20 intron, and 89% similar in the region that is 3' to the intron. Although the region of greatest divergence between the two sequences is in the 3' coding region, each sequence codes for a proline-rich carboxy terminus. In addition, we have determined that a 0.8 kb

mRNA that cross-hybridizes to 15-9-A represents a third gene within this subfamily.

We have also detected the presence of "Nodulin-A"-like sequences in the DNA of alfalfa, several actinorhizal plants, spinach, and corn. This raises the possibility that "Nodulin-A" like sequences have a functional role in plants other than soybean. We have isolated four alder genomic clones which cross hybridize to the cDNA clone 9-11-B, as an approach to characterizing a "Nodulin-A" sequence from another plant. Further characterization of these clones should determine their relatedness to the soybean sequences and may provide clues as to their possible functional significance.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 3

LIST OF TABLES

LIST OF ABBREVIATIONS

A	-	adenine
bp	-	base pairs
C	-	cytosine
cDNA	-	complementary DNA
DNA	-	deoxyribonucleic acid
dNTP	-	deoxyribonucleotide
DTT	-	dithiothreitol
E. coli	-	Escherichia coli
EDTA	-	ethylene-diaminetetra-acetic acid
G	-	guanine
kb	-	kilobases (kilobase pairs)
Lb	-	leghemoglobin
mRNA	-	messenger RNA
MOPS	-	morpholino propanesulphonic acid
PBM	-	peribacteroid membrane
RNA	-	ribonucleic acid
SDS	-	sodium dodecyl sulphate
SDS PAGE	-	SDS polyacrylamide gel electrophoresis
SSC (20X)	-	3 M NaCl, 0.3 M Na citrate, pH7.0
T	-	thymine
TCA	-	trichloroacetic acid
TE	-	10 mM Tris, 1mM EDTA
tRNA	-	transfer RNA
UTR	-	untranslated region of an mRNA

CHAPTER ONE

INTRODUCTION

Biological nitrogen fixation is by far the most important source of fixed nitrogen; however, the capacity to fix atmospheric nitrogen (reduce dinitrogen to ammonia) is limited to a relatively few genera of bacteria and cyanobacteria (Postgate,1982). No eucaryotic organisms have developed this capability; however some plants are able to participate indirectly by forming symbiotic associations with nitrogen-fixing procaryotes. These associations may range from loose associations, such as, that of the bacterium Beijerinkia with the roots of sugar cane, to much more complex associations that involve the development of specific structures on the plants (Postgate,1982). Perhaps one of the most interesting and complex nitrogen-fixing symbioses involves the development of structures on the root called nodules. This occurs in legumes following infection with either Rhizobium or Bradyrhizobium (Postgate,1982), or actinorhizal plants following infection with bacteria of the genus Frankia (See Newcomb and and Wood,1987).

The legume (e.g. soybean, pea, clover, bean, alfalfa) symbiosis with Rhizobium or Bradyrhizobium has been extensively investigated. A complex interaction between the bacterium and its host, requiring the expression of selected genes from both the Rhizobium (bacteroidin genes) and the plant (nodulin genes), results in the development of a

highly specialized structure: the root nodule. Within the nodules a differentiated form of Rhizobium, the bacteroid, reduces atmospheric nitrogen to ammonia. The ammonia is excreted into the plant cytoplasm for assimilation; thus permitting growth of the plant under nitrogen-deficient conditions. This ability to grow without added nitrogen has made the legume-Rhizobium symbiosis agriculturally very important. However, the development of the root nodules is interesting not only because nodules are the site of symbiotic nitrogen fixation; but because nodule development is the result of very specific interactions between two very different organisms -- one a procaryote and the other a eucaryote. How two such very different organisms are able to interact and co-ordinate gene expression is an intriguing problem. To understand the molecular interactions occurring during nodule development, the identification of the symbiotic genes of both partners is necessary. In recent years, a great deal of progress has been made in identifying numerous Rhizobium genes (See Rolfe and Shine,1984; Long,1989). In contrast, relatively few symbiotic plant genes (nodulins) have been characterized (See Verma and Delauney,1988).

In this chapter, several aspects of nodulation, including, nodule morphogenesis, metabolism, and gene expression will be discussed; with emphasis on the plant and plant products. The aspects discussed will focus on the

interaction between soybean and Bradyrhizobium japonicum with reference to other legume-Rhizobium interactions.

1.1 Overview of Nodule Initiation and Development

The initial step towards nodule development requires the recognition of the legume host by Rhizobium or Bradyrhizobium. A prominent feature of the symbiosis is that it exhibits specificity: particular strains of Rhizobium will infect some legumes and not others. For example: Rhizobium japonicum nodulates soybean, R. trifolii nodulates clover, R. meliloti nodulates alfalfa and R. leguminosarum nodulates peas. The host specificity of Rhizobium has formed the basis for their taxonomic classification. Rhizobium species generally fall into two classes of bacteria based on their growth characteristics: a fast growing, and a slow growing class. Generally, it is the fast growing class of rhizobia that demonstrates a great deal of host specificity. The slower growing class of rhizobia is generally found to have a broader host range and has been reclassified into a separate group known as Bradyrhizobium (Jorden,1982). (For the sake of simplicity, hereafter, both Rhizobium and Bradyrhizobium are referred to as Rhizobium).

The initial stages of the infection process involve the induction of Rhizobium nodulation genes, which is mediated by specific plant flavanoid compounds found in root exudates (For review, see Downie and Johnston,1988; Rolfe and Gresshoff,1988; Long,1989). The induction of Rhizobium nodulation genes by plant compounds may, at least in part, contribute to the specificity of the interaction (Horvath et al.,1987). Rhizobium generally gains entry into the plant by invading legume root hairs (root hairs are slender tubular extensions of root epidermal cells). Attachment of rhizobia to root hairs is thought to be mediated by plant lectins bound to root hair surfaces which interact with specific Rhizobium cell surface components (Diaz et al.,1989; reviewed by Dazzo and Gardiol, 1984 and Halverson and Stacey,1986). The first visible sign of infection by compatible Rhizobium is marked curling or contortion of the root hairs. The rhizobia become entrapped in the curl of the root hair. This is accompanied by an induction of cortical cell division adjacent to the root hair cell (Newcomb,1981; Calvert et al.,1984). Invasion by Rhizobium involves the ingrowth of the plant cell wall to form an infection thread; a tube-like structure consisting of an outer plant cell wall and membrane, and an inner matrix containing the invading bacteria (see Newcomb,1981, Sutton,1983). The rhizobia proliferate within the infection thread. The thread passes from the root hair cell to the

newly dividing cortical cells where branching of the infection thread results in the penetration of many, but not all, of the dividing cortical cells (Turgeon and Bauer, 1982). Rhizobia are released from the thin areas at the tips of the infection threads, into the cytoplasm of dividing cells, by a process resembling endocytosis.

In several legumes, division of the cortical cells ceases after release of the rhizobia; but in soybean several more cell divisions occur. Endocytosis of the rhizobia leads to the enclosure of the bacteria by the infection thread membrane, which at this stage becomes known as the peribacteroid membrane (PBM) (see Newcomb,1981; Sutton,1983). Within the peribacteroid membrane, the rhizobia continue to divide until the cytoplasm becomes virtually filled with bacteria. The membrane increases in rough proportion. In the soybean nodule, as many as eight rhizobia may be surrounded by one peribacteroid membrane. Within the peribacteroid membrane, the rhizobia are known as bacteroids which are morphologically different from the free-living rhizobia (See Sutton et al.,1981).

In maturing soybean nodules, cortical cell division generally ceases and subsequent increases in the size of the nodules are due to cell enlargement. The increase in the volume of numerous infected cells causes the nodule to expand radially resulting in a spherically shaped structure. In soybean, about half of the nodule cells

remain uninfected. These uninfected cells are smaller and structurally different from the adjacent infected cells but play an important role in nodule metabolism.

1.2 Overview of Metabolic Interactions

The legume root nodule is a highly specialized organ in which a differentiated form of Rhizobium, the bacteroid, converts atmospheric nitrogen to ammonia. Extensive metabolic cooperation between the two partners is required in order to obtain an effective symbiosis.

1.2.1 Nitrogenase

The reduction of atmospheric nitrogen to ammonia within the bacteroids is catalyzed by the enzyme nitrogenase. Nitrogenase is an enzyme complex composed of two proteins: a molybdenum-iron protein which contains the active site for nitrogen reduction; and an iron containing protein which transfers the electrons required for the nitrogenase reaction to the MoFe protein (See Yates,1980). The nitrogenase complex has two major requirements so that it can effectively reduce nitrogen:

1) protection from oxygen, since the nitrogenase is irreversibly inactivated by oxygen and;

2) a continual supply of ATP and reducing power for the nitrogenase reaction, since the overall energy requirements are considerable (Yates,1980).

1.2.2 Control of oxygen levels

Within the soybean nodule, vigorous oxygen consumption by the bacteroids (Bergerson,1962), and the restricted oxygen entry into the nodule (Tjepkema and Yokum,1973; Tjepkema and Yokum,1974), maintain a low oxygen tension. Although the low O_2 levels are well suited for preserving nitrogenase (Appleby,1984), oxygen is required for generating, in the bacteroid, the ATP necessary for the nitrogenase reaction. An abundant plant protein, leghemoglobin, found exclusively within the infected nodule tissue, facilitates the diffusion of oxygen to the bacteroids at a sufficient rate to support nitrogenase activity (Bergerson et al.,1973; Wittenberg et al.,1974; Appleby,1984).

Leghemoglobin is a myoglobin-like protein with an apoprotein of a molecular weight of about 16,000 daltons linked to a heme prosthetic group. In soybean nodules, there are four major leghemoglobins, each of which is a different gene product (See Section 1.4.1). It is of interest that the apoprotein of Lb is a plant gene product (See Verma and Delauney,1988), whereas the heme group of Lb is produced by the bacteroids (Cutting and Schulman,1969; O'Brian et al.,1987). This is a true example of the symbiosis at the molecular level.

1.2.3 Nitrogen assimilation

The nitrogen fixed by the bacteria is excreted into the plant cell cytoplasm for assimilation by the plant enzyme glutamine synthetase (for review see Miflin and Cullimore,1984; Schubert,1986). Nitrogen-fixing legumes can be classified into two groups based on the nitrogenous products exported from the nodule. They may be either amide (e.g. asparagine or glutamine) exporters or ureide (e.g. allantoin or allantoic acid) exporters (see Schubert,1986). In the nodulated soybean, the major products of nitrogen assimilation are the ureides, allantoin and allantoic acid. Synthesis of ureides is the result of a pathway involving de novo purine synthesis followed by purine catabolism (see Schubert,1986). Both infected and uninfected cells are thought to be involved in ureide biosynthesis (Schubert and Boland,1984; Schubert,1986).

1.2.4 Carbon metabolism

The establishment of an effective symbiosis requires a continual supply of plant photosynthates. The products of plant photosynthesis are transported to the nodule, to support the growth and maintenance of nodule tissue, to supply the large energy requirements of the nitrogenase reaction and ammonia assimilation, and to serve as carbon skeletons for nitrogen assimilation and export.

In soybean, sucrose is the major carbohydrate transported to the nodules (Reibach and Streeter,1983). The major carbon sources transported from the plant to the bacteroids for the support of nitrogenase activity, appear to be the C_4-dicarboxylates of the Krebs cycle (McRae *et al.*,1989).

1.3 Symbiotic Genes

The understanding of the molecular basis of the *Rhizobium*-legume symbiosis requires an understanding of the symbiotic genes of both *Rhizobium* and the legume. Substantial progress has been made in isolating and characterizing numerous *Rhizobium* genes involved in infection, nodule formation and nitrogen fixation (See Pühler,1983; Downie and Johnston,1988; Long, 1989). In contrast, relatively little progress has been made in identifying the roles of the plant genes involved in the symbiosis.

Classical genetic experiments have indicated that specific plant genes are involved in nodule formation and effectiveness (Holl and LaRue,1976; Caldwell and Vest,1977; Nutman,1981). The host appears to influence the symbiosis at several stages including: recognition of the *Rhizobium* strain, the time of nodule appearance, nodule abundance and size, intracellular organization, and the level of nitrogen fixation (Nutman,1981; Gresshoff and Delves,1986; Vance *et al.*,1988). There are very likely numerous nodule-specific

genes (nodulins) involved in the symbiosis. For instance, it is likely that nodulins expressed in the early stages of infection include proteins required for infection thread synthesis and cellular division. As the nodule develops, there will be an increase in the levels of components involved in the synthesis of various structural components of the nodules such as the PBM, as well as enzymes involved in carbon metabolism and nitrogen assimilation. In addition, proteins such as Lb, which are involved in facilitating nitrogen fixation, will be induced.

1.4 Nodulins

About twenty nodule-specific polypeptides, termed "nodulins", were first detected in soybean nodules by immunoprecipitation of *in vitro* translation products of soybean nodule and root mRNA with nodule-specific antiserum (Legocki and Verma, 1980). Since that time, electrophoretic and immunological techniques have been used to detect nodulins in several other legumes. These include: pea (Bisseling *et al.*,1983; Govers *et al.*,1985), alfalfa (Lang-Unnasch and Ausubel,1985; Vance *et al.*,1985; Lullien *et al.*,1987), french bean *Phaseolus vulgaris* (Campos *et al.*,1987), the stem nodulated legume *Sesbania rostrata* (de Lajudie and Huguet,1988), vetch (Moerman *et al.*,1987) and broadbean, *Vicia faba* (Mohaptra *et al.*,1987). In several of these studies, the presence of nodulins at different times

during nodule development were investigated. In addition, bacterial mutants that arrest nodule development at different stages have been useful in determining when various nodulins are expressed (see Verma et al.,1986; Vance et al.,1988; Verma and Delauney,1988 for review).

The first cloned nodulin cDNA was that of Lb from soybean nodules (Truelson et al.,1979). cDNA cloning of nodule-specific mRNAs other than Lb was also first accomplished with soybean (Fuller et al.,1983). Recently, cDNAs corresponding to nodulin genes have been cloned from pea (Govers et al.,1987), alfalfa (Barker et al.,1988); Dunn et al.,1988), Sesbania rostrata (Strittmatter et al.,1989), Phaseolus vulgaris (Cullimore,1984; Bennett et al.,1989) and lupin (Konieczny et al.,1989).

cDNA cloning of nodule-specific sequences is useful; because cDNA clones can be used to evaluate gene structure, regulation, and expression. In addition, analysis of the nucleotide sequence may provide some insight as to function and subcellular location of the nodulins; however, until recently most of the advances in the study of nodulin gene structure and function have been limited to soybean (see Verma and Delauney,1988).

1.4.1 Leghemoglobin

Leghemoglobin is the most extensively studied and well characterized of the nodulins. It is a very abundant protein found within the nodule tissue and is involved in facilitating the diffusion of O_2 to the bacteroids (Appleby,1984). Leghemoglobin had been thought to occur exclusively in the cytoplasm of the infected cells (see Verma and Delauney,1988). However, a recent study, in which immunogold labeling was used to search for Lb in the nodule cells of soybean, has detected Lb in the adjacent uninfected cells; but at a lower level than is found in the infected cells (Vanden Bosch and Newcomb,1988). It is uncertain whether the Lb in the uninfected cells represents the holoprotein or the apoprotein only. Van den Bosch and Newcomb (1988) have suggested that in addition to facilitating O_2 diffusion to the bacteroids, Lb may play a role in delivering oxygen to plant organelles throughout the infected region of the nodule.

In soybean, leghemoglobin is encoded by a small multi-gene family. Soybean nodules contain four major leghemoglobin species called Lba, Lbc_1, Lbc_2, Lbc_3 which differ only slightly from each other in amino acid sequence (Sievers et al.,1978; Fuchsman and Appleby,1979). Several minor Lb species have also been detected in the soybean nodule that are likely derived from post translational modifications of each of the major components.

Genes representing the four major leghemoglobin components have been isolated and characterized (Brisson and Verma,1982; Hyldig-Nielsen et al.,1982; Wiborg et al.,1982). In addition, the leghemoglobin gene family includes two pseudogenes and two truncated genes (Brisson and Verma,1982; Bojsen et al.,1983; Wiborg et al.,1983). Six of the Lb genes are found in two separate clusters in the soybean genome (Lee et al.,1983; Bojsen et al.,1983). One cluster contains Lba, Lbc_1, Lbc_3, and a pseudogene. The second cluster contains Lbc_2 and the second pseudogene. The two truncated genes are located at two independent loci (see Verma and Nadler,1984 for review).

Leghemoglobin mRNA is detected in the nodule several days before the onset of nitrogen fixation (Verma et al.,1981; Fuller and Verma,1984; Marcker et al.,1984). Marcker et al. (1984) have shown that Lb mRNA expression is activated over a few days, early in nodule development, followed by a dramatic increase in transcription shortly before the onset of nitrogen fixation.

Jensen et al. (1986) have investigated the expression of a chimeric Lbc_3 gene in transgenic Lotus corniculatus (a legume). It was found that the chimeric gene was activated in the nodule of Lotus corniculatus and followed the correct developmental timing. These results suggested that the mechanisms responsible for the activation of leghemoglobin genes are conserved in different Rhizobium-

legume associations. Such a transformation system may be a valuable tool for the identification of regulatory sequences (Section 1.5).

1.4.2 Nodulins involved in nitrogen assimilation

The ammonia produced by the bacteroids is excreted into the plant cytoplasm for assimilation by the host. In the soybean, the major products of nitrogen fixation are the ureides -- allantoin and allantoic acid (for review see Schubert,1986). Therefore, one would expect either elevated levels of enzymes that can be found in other plant tissues, or new nodule-specific forms of the enzymes (nodulins) involved in nitrogen assimilation and ureide biosynthesis. In the soybean nodule, it has been established that the activities of several nitrogen assimilating enzymes do increase in the nodule tissue (Reynolds et al.,1982; Triplett,1985; see Robertson and Farnden,1980; Schubert,1986).

Glutamine synthetase is the major enzyme responsible for the first step in the assimilation of the ammonia excreted by the bacteroids (see Miflin and Cullimore,1984). An apparently nodule-specific form of glutamine-synthetase has been reported in Phaseolus vulgaris (Lara et al.,1983). The corresponding cDNA has been isolated and used to determine that this glutamine synthetase is expressed in the nodule (Cullimore et

al.,1984; Gebhardt et al.,1986) and not the root. A recent report, however, indicates that this form of glutamine synthetase is not entirely nodule-specific as was previously thought, since it was also detected at low levels in stems, petioles and green cotyledons (Bennett et al.,1989). An apparently nodule-specific form of glutamine synthetase has been detected in alfalfa (Dunn et al.,1988).

In soybean, studies by Sengupta-Gopalan and Pitas (1986) had shown that increases in glutamine synthetase were due to the synthesis of nodule-specific forms of the enzyme. However, Hirel et al. (1987) have demonstrated that the increase in glutamine synthetase in soybean nodules is due to an increase in the levels of at least two glutamine synthetase mRNAs that can also be detected at a low level in the root. The studies by Hirel et al. (1987) in soybean suggest that the level of NH_4^+ in the nodule may play a role in stimulating the soybean glutamine synthetase genes.

In pea, glutamine synthetase levels are also elevated in the nodules compared to the roots, but no nodule-specific isoform has been detected (Tingey et al.,1987). Thus, it appears that although glutamine synthetase activity is greatly enhanced in the nodules of many legumes, this enhancement is likely due to an enhanced level of forms that can be detected in other plant tissues. Therefore, glutamine synthetase is strictly speaking, at least in some legumes (and possibly all legumes), not

nodule-specific. However, because of its importance in ammonia assimilation, and the fact that it is developmentally controlled in the nodule, the control of glutamine synthetase genes requires further study (see Cullimore and Bennett,1988 for review).

In the soybean nodule, several enzymes involved in ureide biosynthesis appear to be nodulins. One such enzyme, xanthine dehydrogenase, which is involved in purine catabolism, has been localized to the infected cells of the soybean nodule (Triplett et al.,1985); however, this enzyme may also not strictly speaking be nodule-specific, since it has been detected at low levels in other soybean tissue.

The enzyme uricase II has been identified as a nodule-specific form of uricase in soybean (Legocki and Verma,1979; Bergmann et al.,1983; Nguyen et al.,1985). Uricase, an important enzyme for ureide production, catalyzes the conversion of uric acid to allantoin (Schubert,1986). Uricase II is composed of four 35 kDa subunits and was originally detected as nodulin-35, the second most abundant cytoplasmic protein in soybean nodules (Legocki and Verma,1979).

Uricase II (nodulin-35) has been localized to the peroxisomes of the uninfected cells within the soybean nodule (Nguyen et al.,1985; van den Bosch and Newcomb,1986). The fact that this nodulin is located in the uninfected

cells of the soybean nodule reiterates the complexity of the metabolic interaction occuring in the nodule.

Analysis of the nodulin-35 gene indicates that it is about 5 kb in length and contains 7 introns (Nguyen et al, 1985).

1.4.3 Nodulins involved in carbon metabolism

Relatively little information is available about nodule-specific proteins involved in carbon metabolism. The activities of a few enzymes (invertase, phosphoenolpyruvate carboxylase and sucrose synthethase) involved in carbon metabolism are detected at elevated levels in nodules (see Robertson and Farnden,1980; Morell and Copeland,1985; Verma and Delauney,1988).

Thummler and Verma (1987) have characterized a nodule-specific cDNA corresponding to nodulin-100 (Fuller et al.,1983), and found that it codes for sucrose synthetase. Thummler and Verma (1987) have also isolated an abundant 90 kDa protein which has the properties of sucrose synthetase. Antibodies raised against this protein detected low levels of cross-reacting protein in the uninfected roots but not in leaves or stems. Therefore, it is possible that levels of sucrose synthetase are elevated in the nodule compared to the root, and sucrose synthetase is not entirely nodule-specific.

Further study is required to determine the biochemistry and expression of other enzymes involved in the metabolism of photosynthates in the nodule.

1.4.4 Peribacteroid membrane nodulins

Within the legume nodule, the bacteroids are surrounded by the peribacteroid membrane (PBM). The peribacteroid membrane plays an important role in the symbiosis since it segregates the bacteroids from the host cytoplasm. In addition, it must have nodule-specific functions that involve the exchange of metabolites between the host and the bacteroid. The area of the peribacteroid membrane in an infected soybean cell is about 20 - 40 fold higher than the area of the plasma membrane in the same cell (Verma et al.,1978). Therefore, it is reasonable to assume that soybean nodules would contain elevated levels of enzymes involved in PBM synthesis as well as structural components of the membrane itself.

Mellor et al. (1986) have determined that the soybean nodule tissue contains much higher choline kinase activity than uninfected tissue. Choline kinase is thought to be involved in PBM biosynthesis. Two dimensional gel electrophoretic analysis of infected and uninfected tissue, suggested that the soybean nodule tissue contains a nodule-specific form of choline kinase (choline kinase II). This raises the possibility that a novel form of an enzyme

necessary for membrane synthesis is expressed specifically in the nodule (Mellor et al.,1986).

Fortin et al. (1985) have raised nodule-specific antisera against the PBM from soybean and demonstrated the presence of several nodulins. Verma's group has characterized three PMB nodulins: nodulin-24 (Katinakis and Verma,1985), nodulin-26 (Fortin et al.,1987), and nodulin-23 (Mauro et al.,1985).

The nodulin-24 gene codes for a polypeptide with an apparent molecular weight of 24 kDa (as determined by SDS PAGE). Sequence analysis of the nodulin-24 gene indicates that it has 5 exons separated by 4 introns, and has a coding capacity for a polypeptide of 15.1 kDa (Katinakis and Verma,1985). The discrepancy between the predicted and the apparent molecular weight is apparently due to the hydrophobic nature of the protein, causing it to run anomalously in SDS PAGE gels (Katinakis and Verma,1985). Analysis of the possible secondary structure of nodulin-24 has suggested that the protein may reside on the PBM with the hydrophobic half of the protein embedded in the membrane and the hydrophilic half in the peribacteroid fluid (Fortin et al.,1987). Although nodulin-24 appears to be a PMB protein, its function is unknown.

The characterization of a cDNA clone corresponding to a different PBM nodulin, nodulin-26, has indicated that it is a transmembrane protein (Fortin et al.,1987).

Estimation by SDS-PAGE, had indicated that it had a weight of 26.5 kDa but the nodulin-26 cDNA clone only encoded a polypeptide of 22.5 kDa molecular weight (Fortin et al.,1987). It was postulated that the discrepancy in the real and the apparent molecular weights might be due to the extreme hydrophobic nature of the protein. Recently, Sandal and Marcker (1988) have also reported the characterization of a cDNA clone corresponding to nodulin-26. The sequence differs from the sequence published by Fortin et al. (1987) by an extension of 66 amino acids at the amino terminus of the protein (Sandal and Marcker,1988) which would increase the predicted molecular weight to about 29.7 kDa. It has recently been reported that nodulin-26 shares significant amino acid sequence homology and structure similarity with the major intrinsic protein (MIP) of the bovine lens fibre membrane (Sandal and Marcker,1988) and the MIP of the rat lens fibre membrane (Shiels et al.,1988). MIP is believed to be a component of lens fibre membrane gap junctions; therefore, Sandal and Marcker (1988) have proposed that nodulin-26 in the PBM might be involved in transporting small molecules by forming channels through the PBM.

The third PBM nodulin, nodulin-23, was found to represent a 24.3 kDa protein (Mauro et al.,1985; see Verma and Delauney, 1988). Characterization of the gene for nodulin-23 has indicated that it contains one intron near the 3' end of the coding region. The function of nodulin-

23 is unknown. Nodulin-23 was found to be a member of a small nodulin multi-gene family (Section 1.4.5).

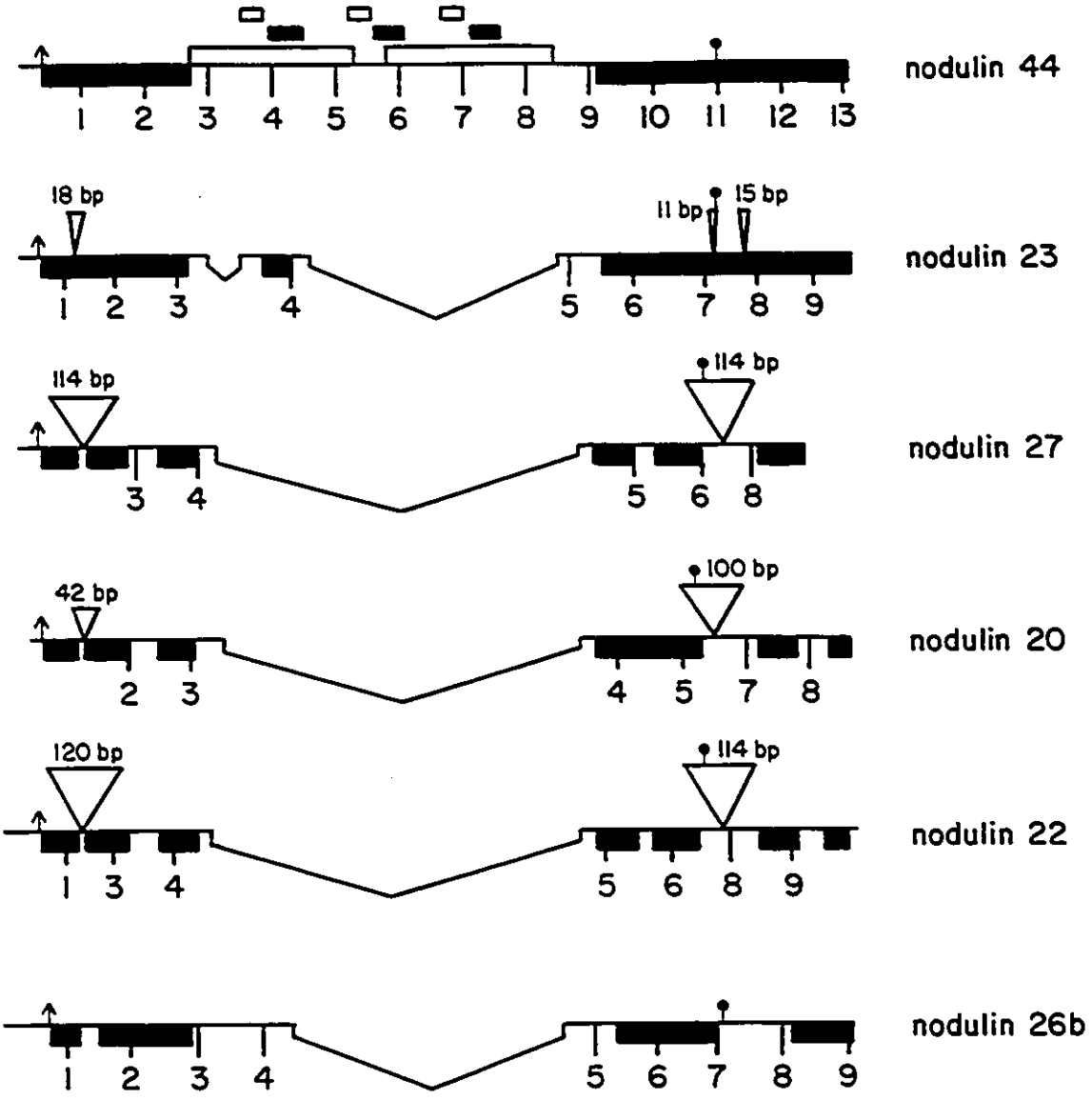
1.4.5 The "Nodulin-A" family

A number of soybean nodulins have been found to be members of a small abundantly transcribed multi-gene family, the "Nodulin-A" family ("Nodulin A" is the nomenclature used by Verma and Delauney (1988), and will hereafter be the term used to refer to this family). The work of several different groups (Sengupta-Gopalan et al.,1986; Jacobs et al.,1987; Sandal et al.,1987) has identified the presence of at least six members of the "Nodulin-A" family: nodulin-20, nodulin-22, nodulin-23, nodulin 26b, nodulin 27, and nodulin-44. The genes corresponding to nodulin-20, nodulin-22 (Sandal et al.,1987) and nodulin-23 (Mauro et al.,1985) have been characterized and each has been found to contain one intron close to the 3' end of the coding region. Nodulin-26b, nodulin-27 (Jacobs et al.,1987), and nodulin-44 (Sengupta-Gopalan et al.,1986) have been characterized at the cDNA level. Nodulin-23 has also been characterized at the cDNA level (Sengupta-Gopalan et al.,1986). DNA sequence analysis has indicated that these nodulins each have two domains which are 65 - 95% conserved among the various nodulins, separated by a unique region of varying length (see Figure 1.1, Sengupta-Gopalan et al.,1986; Jacobs et al.,1987; Sandal et al.,1987; Verma and Delauney,1988).

These two domains are however, significantly different from each other. The amino terminal sequences of these nodulins have hydrophobic peptides resembling the signal peptides required for the translocation of proteins across membranes (Sengupta-Gopalan,1986; Jacobs et al.,1987; Sandal et al.,1987); however, these putatative signal peptides may not be functional in all of these nodulins. Immunoprecipitation of the translation products of hybrid-selected mRNAs with antisera against PBM and against soluble proteins have suggested that nodulin-23 is located in the peribacteroid membrane and nodulin-27 is located in the cytosol (Jacobs et al.,1987). The functions of these nodulins are unknown. Amino acid sequence alignments by Sandal et al. (1987) of nodulins 20, 22, 23 and 44 have demonstrated that a number of amino acids are conserved among the various nodulins. Each of the two conserved domains contain four cysteine residues that can be arranged into structures resembling metal binding domains. Sandal et al. (1987) have proposed that the binding of metal ions is important for the function of these nodulins.

FIGURE 1.1 **Schematic alignment of "Nodulin-A" Family cDNA Sequences**

The cDNA sequences of nodulin-23 (nod C51) (Mauro et al.,1985; Sengupta-Gopalan et al.,1986), nodulin-27 (Jacobs et al.,1987), nodulin-20 (Sandal et al., 1987), nodulin-22 (Sandal et al.,1987), and nodulin-26b (Jacobs et al.,1987) were aligned for maximum homology to the nod E27 (nodulin-44) sequence of Sengupta-Gopalan et al. (1986). The nodulin-23 sequence used for the alignment, was the nod C51 cDNA sequence of Sengupta-Gopalan et al. (1986). The nodulin-20 and nodulin-22 sequences were derived from the gene sequences (Sandal et al.,1987) with the introns removed. Regions of greater than 70% homology to the nodulin-44 sequence are indicated by solid bars beneath the lines. Open and closed bars above the nodulin-44 sequence, represent repetitive sequences. Insertions are indicated by the open triangles. The initiation codons are indicated by the arrows, and the termination codons are indicated by the closed circles. The numbers indicate the positions of the nucleotides (as multiples of 100 nucleotides) from either the start of the cDNA clone, or the transcription initiation site of the gene. Sequence similarity indicate that nodulin-22 and nodulin-27 likely represent alleles of the same gene. This figure has been adapted from Verma and Delauney (1988).


nodulin 44
1 2 3 4 5 6 7 8 9 10 11 12 13
18 bp
11 bp
15 bp
nodulin 23
1 2 3 4 5 6 7 8 9
114 bp
114 bp
nodulin 27
3 4 5 6 8
42 bp
100 bp
nodulin 20
2 3 4 5 7 8
120 bp
114 bp
nodulin 22
1 3 4 5 6 8 9
nodulin 26b
1 2 3 4 5 6 7 8 9

1.4.6 Early nodulins

In soybean nodules, several nodulins have been detected that are transiently expressed, early in nodule development, when the nodule structure is being formed (Gloudemans et al.,1987). Induction of these nodulins does not require the presence of bacteria inside the root tissue. This is in contrast to "late" nodulins such as Lb or the "Nodulin-A" family members which are expressed shortly before the onset of nitrogen fixation and are probably involved in nodule function (Fuller and Verma,1984; Marcker et al.,1984; Sengupta-Gopalan et al.,1986). Franssen et al. (1987) have reported the characterization of a cDNA (ENOD2) corresponding to a 1.2 kb mRNA species detected early in nodule development. The cDNA clone was used to hybrid select RNA that was in vitro translated to produce two nodulins with an apparent molecular weight of 75 kDa (nodulin-75). The amino acid sequence predicted by the cDNA sequence indicates that proline accounts for 45% of the amino acid residues of nodulin-75. Franssen et al.,(1987) have suggested that nodulin-75 may be a structural protein involved in nodule morphogenesis. Nodulin-75 appears to be conserved among legumes. mRNA homologous to the nodulin-75 cDNA sequence has been detected in infected tissue from pea (Govers et al.,1986), vetch (Moerman et al.,1987), and Sesbania rostrata (Strittmatter et al.,1989).

Table I CHARACTERIZED NODULINS FROM SOYBEAN

Nodulin	Apparent Size(kD)	Actual Size(kD)	Subcellular Location	Function
CK[1]	60	-	?	pbm synthesis
GS[2]	38,40	-	?	glutamine synthetase
Lb[3]	13	15.7	cytoplasm	oxygen diffusion
N-20[4]	-	20	?	?
N-22,36-1-A[5]	-	22.7	cytoplasm	?
N-23,C51, 9-11-B[6]	24.5,29	24.3	pbm	?
N-26b[7]	25.5	23.5	cytoplasm?	?
N-27[8]	27	22.7	cytoplasm	?
N-44,E27 6-9-F[9]	42	39	pbm?	?
N-24[10]	24	15.1	pbm	pbm surface membrane protein
N-26[11]	26.5	22.5	pbm	transport across pmb?
N-35[12]	33	35.1	peroxisome of uninfected cell	uricase II
N-75[13]	75	<45	?	nodule morphogenesis?
N-100[14]	90	-	cytoplasm	sucrose synthase
XDH[15]	141	-	?	xanthine dehydrogenase

1. Mellor et al.,1986.
2. Sengupta-Gopalan & Pitas,1986; Hirel et al.,1987.
3. Four different Lbs genes from soybean have been characterized. Brisson & Verma,1982; Hyldig-Nielson et al.,1982; Wiborg et al.,1982
4. Sandal et al.,1987. 5. Sandal et al.,1987; this thesis.
6. Mauro et al.,1985; Sengupta-Gopalan et al.,1986; this thesis.
7. Jacobs et al.,1987. 8. Jacobs et al.,1987.
9. Sengupta-Gopalan et al.,1986; this thesis
10. Katinakis & Verma,1985; Fortin et al.,1985.
11. Fortin et al.,1987; Sandal & Marcker,1988.
12. Bergmann et al.,1983; Nguyen et al.,1985. 13. Franssen et al.,1987.
14. Thummler & Verma,1987. 15. Triplett et al.,1985.

1.5 Regulation of Nodulin Gene Expression

The availability of bacterial mutants which arrest the development of the nodules at various stages has provided a tool for studying nodulin gene expression at specific stages in development (for review see Govers et al.,1987; Vance et al.,1988; Verma and Delauney,1988; Long, 1989). Specific Rhizobium signals may be involved in activating nodulin gene expression during the different stages of nodule development (see Verma and Delauney,1988; Long,1989). Also, the physiological conditions of the nodule environment may play a role in the induction of specific genes (see Verma and Delauney,1988).

With the isolation and characterization of nodulin genes, it has become possible to investigate the cis-regulatory sequences necessary for activation of nodulin genes. The expression of chimeric nodulin genes has been studied in transgenic Lotus corniculatus (Stougaard et al.,1986; Jorgensen et al.,1987). A chimeric Lbc_3 gene was found to be expressed in a nodule-specific manner and followed the correct developmental timing of expression. These results indicated that the mechanism for Lb induction is similar among different legumes (Stougaard et al.,1987). Deletion analysis of the 5' region of the Lbc_3 gene has defined cis-regulatory regions required for promoter activity and nodule specificity (Stougaard et al.,1987). Similarly, a chimeric nodulin-23 gene was found to be

expressed in a nodule-specific manner in two transgenic legumes: Lotus corniculatis and Trifolium repens (Jorgensen et al.,1988). These results also support the notion that some of the molecular mechanisms involved in activating nodulin genes may be conserved in various Rhizobium-legume symbioses. Promoter deletion analysis of the nodulin-23 gene also delimited putative regulatory sequences involved in promoter activity and organ specificity (Jorgensen et al.,1988).

Recently, gel retardation assays have been used to study the interaction of trans-acting factors from nuclear extracts of soybean nodules, leaves and roots with the 5' upstream region of Lbc_3 (Jensen et al.,1988). These studies indicated that two distinct regions in the 5' upstream region (at positions -246 to -223 and positions -176 to -161) of the soybean Lbc_3 gene interacted with the same, nodule-specific trans-acting factor (Jensen et al.,1988). The promoter deletion analysis by Stougaard et al.,(1987) suggests these elements may be positive control elements involved in promoter activity. The element for nodule-specificity was located at -139 to -102 relative to the start point of transcription (as defined by Stougaard et al.,1987) but this region did not appear to bind to any nuclear factors (Jensen et al.,1988). Therefore, it is clear that further investigation of factors leading to nodule-specific expression of leghemoglobin genes is

necessary. Also, the factors which regulate the expression of Lb need not be similar to those regulating other nodulin genes.

Although the mechanisms involved in nodulin gene expression are unknown, it is clear from these recent studies that progress is being made in deciphering the underlying molecular mechanisms involved in nodulin gene expression. The ability to generate chimeric genes, and the fact that many legumes are now available for transgenic studies, should result in rapid progress in elucidating some of the molecular mechanisms of nodulin gene induction.

1.6 Thesis Objectives

The development of the legume root nodule is a complex process involving the coordinated expression of bacterial and plant genes. At the time this project was initiated, significant progress had been made in identifying bacterial genes involved in the symbiosis but relatively little progress had been made in identifying plant genes. Immunological analysis of soybean root and nodule tissue using nodule-specific antiserum had detected the presence of about 20 nodule-specific polypeptides other than leghemoglobin (Legocki and Verma, 1980). Also, RNA - cDNA hybridization kinetic studies by Auger and Verma (1981) suggested the existence of from 20 - 40 nodule-specific sequences within the moderately abundant mRNA population of

soybean nodules. But, except for leghemoglobin and nodulin-35 from soybean, most host gene products were uncharacterized.

As an approach to investigating some of the changes that occur during soybean nodule development, nodule-specific cDNA clones were isolated from a cDNA library that had been constructed from poly $(A)^+$ RNA derived from 2 - 3 week old soybean nodules following infection with Bradyrhizobium japonicum strain 61A76. These cDNA clones were then used as molecular probes to study the expression of their corresponding genes during nodule development in soybean.

Sequence analysis of these cDNA clones indicated that they were all members of the abundantly transcribed "Nodulin-A" gene family (Fuller et al.,1983; Verma and Delauney, 1988). Sequencing of three of the cDNA clones has supported the sequences of previously published "Nodulin-A" family members. However, a fourth cDNA clone represents a new "Nodulin-A" family member. We have determined that this clone represents a member of a subfamily of genes within the larger "Nodulin-A" gene family.

The analysis of the sequences of these "Nodulin-A" family members has revealed limited information as to their function. As an approach to investigating the possible functional significance, we have attempted to isolate a "Nodulin-A" homologue from a non-legume -- alder.

If a functional "Nodulin-A" gene can be isolated from alder, comparison of the alder sequence to the soybean sequences might reveal conserved amino acids that may be necessary for protein function. In addition, alder forms a symbiotic association with Frankia resulting in root nodule development, but very little is known about the induction of plant gene expression during alder nodule development. It would be particularly exciting if a "Nodulin-A" family homologue was expressed during alder nodule development. In this thesis, we report the detection of "Nodulin-A" like sequences in the DNA of several non-legumes. In addition, we have isolated four alder genomic clones that cross-hybridize to a "Nodulin-A" cDNA. We are currently investigating these clones to determine their relatedness to the soybean "Nodulin-A" family.

CHAPTER 2

MATERIALS AND METHODS

2.1 Growth and Collection of Plant Tissues

Soybean seeds (Glycine max (L.) Merrill cv. Maple Arrow) were obtained from H. Volding, Agriculture Canada, Ottawa. Seeds were surface sterilized in a 20% solution of commercial bleach for 10 minutes. The seeds were then rinsed with distilled water and dispersed on sterile trays covered with moistened Whatman 3MM paper. The seeds were germinated in the dark at 25°C for 3 days. Seeds that did not germinate, or seedlings that appeared abnormal were discarded. Healthy seedlings were planted in 6 inch pots (5 plants/pot) containing vermiculite (day 0). Each seedling was inoculated with 3 ml of a late log phase culture of Bradyrhizobium japonicum strain 61A76. The plants were grown at 25°C with a 16 hour light/8 hour dark photoperiod. The plants were watered with a nitrogen-free nutrient solution (Cutting and Schulman,1969). All plant tissues were hand-picked, immediately frozen in liquid nitrogen and stored at -80°C until further use.

Mature nodules used for northern blotting experiments were hand-picked from 3-4 week old plants. Control root tissue was obtained from 3 day old soybean seedlings and from 17 and 25 day old uninoculated plants.

Tissues collected for developmental studies were harvested at day 4, 7, 10, 13, 16, 24 and 34 following inoculation. At day 4, no nodules were visible and the top 4 cm of the 6 cm total root length were excised. At day 7, very small brownish nodules were visible on the tap root. Approximately 5 cm of the 9 cm root system was collected. After 10 days, the nodules were visible in the top 4-6 cm of 10-12 cm root systems. The top 4-6 cm of the nodulated root system was harvested at this time point. The nodules from the 13, 16, 24 and 34 day time points were individually removed from the root tissue and frozen in liquid nitrogen.

2.2 Nitrogenase Assay

Nitrogenase activity was measured by the acetylene reduction assay (Hardy et al.,1968). Root systems were removed from the plants at the base of the stem and placed in 30 ml serum stoppered bottles. 3 ml of acetylene was injected. 3 ml samples were removed from the bottle at 0, 30 and 60 minutes and injected into 3 ml vacutainers. Ethylene production was measured by injecting 0.5 ml of the vacutainer sample into a Carle Intruments GC9700 Basic Gas Chromatograph (Agriculture Canada, Ottawa). Known amounts of ethylene were run as calibration standards. Nodules were then removed, dried, and weighed and nitrogenase activity was calculated as nmoles ethylene produced/minute/gram of

dry weight of nodule. The data presented are the mean of three different plants.

2.3 Bacterial Strains, Plasmids and Phage

Bradyrhizobium japonicum strain 61A76 from Nitragin Corporation, Wisconsin, was used to nodulate soybeans.

Escherichia coli strain HB101, RR1 (described in Maniatis et al.,1982) and DH5α (BRL) were used as hosts for the growth and purification of plasmids. E. coli strain JM101 was used as the host for M13 vectors (Messing et al.,1981). E. coli strain WA802 (Raleigh et al.,1988) was used as the host for lambda GEM-11 (Promega).

Plasmid cloning vectors included pBR322 (Maniatis et al.,1982) and the pGEM 3Z/pGEM 4Z vectors from Promega.

M13 mp18 and M13 mp19 (Yanisch-Perron et al.,1985) phages were used as sequencing vectors.

2.4 Isolation of DNA

2.4.1 Isolation of plant DNA

High molecular weight soybean DNA was isolated from soybean leaves or sprouts. Soybean tissue was ground to a fine powder under liquid nitrogen using a mortar and pestle. 10 ml of extraction buffer (100 mM Tris-HCl, 50 mM EDTA, 1% SDS, pH 8.0) were added per gram of powdered tissue and the slurry was poured into an 50 ml Oakridge centrifuge

tube or 250 ml centrifuge bottle. The slurry was extracted with an equal volume of phenol (buffered in 50 mM Tris, 10 mM EDTA) and centrifuged for 10 minutes at 8,000 rpm at 20°C. This was followed by two phenol-chloroform (prepared as in Maniatis et al.,1982) extractions. Total nucleic acids were precipitated overnight at -20°C following the addition of 0.1 volumes of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol. The precipitate was then collected by centrifugation at 8,000 rpm for 20 minutes at 4°C. The pellet was dissolved in TE and treated with RNAse (50 µg/ml) for 30 minutes at room temperature. Following extraction with an equal volume of phenol-chloroform, the nucleic acid was precipitated with sodium acetate and ethanol overnight at -20°C. The precipitate was collected by centrifugation at 8,000 rpm for 20 minutes and washed 2X with 70% ethanol. The pellet was dissolved in 25 ml of TE and 25 grams of CsCl and 1 ml of ethidium bromide (10mg/ml) were added. DNA was purified by CsCl equilibrium gradient centrifugation for 18 hours at 44,000 rpm using a 50 VTi rotor (Beckman). Following centrifugation, the ethidium bromide was removed by butanol extraction (Maniatis et al.,1982). Three volumes of TE were added to the aqueous phase, and following the addition of 2.5 volumes of ethanol, the DNA was precipitated overnight at -20°C.

DNA from plants other than soybean was isolated using the procedure described in Appendix II.

2.4.2 Preparation of plasmid DNA

Plasmid DNA was prepared both on a large scale and small scale by the alkaline lysis procedure of Birnboim and Doly (1979) as described in Maniatis et al.(1982). Following large scale isolation, plasmids were purified by CsCl/ethidium bromide centrifugation.

2.4.3 Isolation of DNA fragments

DNA restriction fragments were isolated 1) from agarose gels by using a GENECLEAN kit as recommended by the manufacturer (Bio 101), or 2) from low melting temperature agarose gels by phenol extraction as described below.

Following electrophoresis of restriction digests in a low melting temperature gel, the gel was stained with ethidium bromide and the appropriate band was excised from the gel with a scalpel and put into an eppendorf tube. The volume of the gel slice was estimated following incubation of the gel slice at 65°C for 2 minutes. 1-2 volumes of TE were added to the liquified gel. This was incubated at 65°C for 15-30 minutes with occasional vortexing. Agarose was extracted from the sample by adding an equal volume of hot phenol (65°C) buffered with 50 mM Tris-HCl pH 8.0, 10mM EDTA, followed by immediate vortexing. This was left for 5 minutes at room temperature with occasional vortexing, followed by a 5 minute centrifugation. The aqueous phase was removed into a clean eppendorf tube and extracted once

with phenol chloroform and once with chloroform (chloroform - isoamyl alcohol, 24:1). The aqueous phase from the chloroform extraction was then removed into a clean eppendorf tube and the fragment was precipitated by adding 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol.

2.4.4 Isolation of lambda DNA

Lambda phage was grown, titred and purified as described in Maniatis et al. (1982). Extraction of lambda phage DNA was also as described in Maniatis et al. (1982).

2.5 RNA Isolation Procedures

2.5.1 Isolation of total RNA

Total RNA from soybean roots or nodules was isolated using the method of Chirgwin et al. (1979) or an updated version of this isolation procedure (MacDonald et al.,1987). Briefly, the isolation procedure involves the homogenization of 1 gram of plant tissue per 5 ml of a 4 M guanidinium thiocyanate solution at a high speed in an omni-mixer, followed by selective precipitation of the RNA with ethanol. The precipitated RNA was pelleted and then dissolved in a 7.5 M solution of guanidine-HCl, and again selectively precipitated. This was then followed by a second precipitation from guanidine-HCl and the final RNA

pellet was processed as described in Chirgwin et al. (1979) or MacDonald et al. (1987).

2.5.2 Isolation of poly (A)$^+$ RNA

Polyadenylated RNA was isolated from total RNA by oligo (dT)-cellulose chromatography (Aviv and Leder,1972). Poly (A)$^+$ RNA selection was performed essentially as described in Maniatis et al. (1982). The oligo (dT) column was equilibrated with sterile loading buffer which consisted of 10 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 1 mM EDTA, 0.1% SDS and loaded onto BIO-RAD polypropylene Econo-Columns. RNA was dissolved in sterile H_2O, heated to 65°C and cooled on ice. An equal volume of 2X loading buffer was added to the RNA. The final concentration of RNA in 1X loading buffer was about 1 mg/ml. 80-100 mg of oligo (dT)-cellulose was used per mg of total RNA. The RNA was applied to the column three times. The column was then washed with 10 column volumes of loading buffer, followed by 5 column volumes of loading buffer containing 0.1 M NaCl. Poly (A)$^+$ RNA was eluted with 2-3 column volumes of sterile elution buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA). The poly (A)$^+$ RNA was precipitated at -20°C following the addition of 0.1 volume of 3 M sodium acetate (pH 5.5) and 2.5 volumes of ethanol.

2.6 Labeling of Probes

2.6.1 Preparation of radioactively labeled plasmids

Plasmid DNAs were radiolabeled by nick translation (Rigby et al., 1977) using kits purchased from Amersham. [α-^{32}P]-dCTP (3000 Ci/mmol) was used as the radioactive label. Following nick translation of the plasmid DNA, unincorporated nucleotides were removed by passage over a Sephadex G-50 column. Specific activities of the probes were generally 2-6 x 10^8 cpm/μg. Prior to use, the required amount of probe was heated at 100°C for 5 minutes and fast cooled on ice for 5 minutes and then added to the hybridization mix.

2.6.2 Preparation of radioactively labeled fragments

Isolated restriction fragments were labeled by the random primer method (Feinberg and Vogelstein,1983) using kits purchased from Amersham. The radioactive label was [α-^{32}P]-dCTP (3000 Ci/mmol). Unincorporated nucleotides were removed by passing the labeled fragments through Sephadex G-50 by the spun-column procedure as indicated in Maniatis et al. (1982). Prior to use, the required amount of labeled fragment was heated to 100°C for 2 minutes, fast cooled on ice and added to the hybridization solution.

2.6.3 End-labeling of RNA

Soybean poly $(A)^+$ RNA that had been isolated by oligo (dT)-cellulose chromatography, or soybean ribosomal RNA that had been isolated by sucrose density gradient centrifugation were prepared for end-labeling. 5 μg of RNA were incubated in 30 μl of 0.1M NaOH for 40 minutes at 4°C to yield fragments of about 200 nucleotides in length. The RNA was precipitated by adding 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol.

The RNA was end-labeled as described by Williams and Lloyd (1979). 1 μg of cleaved RNA was incubated for 30 minutes at 37°C in 20 μl of buffer containing 50 mM Tris-HCl (pH 7.6), 5 mM $MgCl_2$, 10 mM mercaptoethanol and 10 μCi of [γ-^{32}P]-ATP (New England Nuclear, specific activity: 7000 Ci/mmol) and 0.5 units of T_4 polynucleotide kinase. The reaction was stopped by increasing the volume of the reaction mix to 100 μl with TE, heating briefly at 65°C and then cooling on ice. Incorporation was assayed by TCA precipitation (Maniatis et al.,1982). The RNA was extracted once with phenol and once with phenol-chloroform. Unincorporated nucleotides were removed by passing the labeled RNA over a DEAE-Sephacel column or Sephadex G-50 column (Maniatis et al.,1982). The specific activity was generally about 1 - 2 x 10^7 cpm/μg.

2.6.4 End-labeling of oligonucleotides

200 ng of oligonucleotide (24 pmol) was incubated for 30 minutes at 37°C in 30 µl of a reaction mix containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT, 10 units of T_4 polynucleotide kinase and 130 µCi (approximately 24 pmol) of [γ-^{32}P]-dATP (Amersham, specific activity: 5000 Ci/mmol). The reaction was stopped by adding 0.9 µl of 500 mM EDTA (pH 7.8). Unincorporated nucleotides were removed using a NACS ion exchange column as recommended by the manufacturer (BRL).

Before use, the probes were centrifuged for 5 minutes and then incubated for 2 minutes at 100°C. The probes were then added immediately to the hybridization solution.

2.7 Screening of cDNA Library

A cDNA library had been constructed from 2-3 week soybean nodule poly (A)$^+$ RNA (D.A. Johnson, unpublished). The cDNA was synthesized using techniques described in Maniatis et al. (1982). The cDNA was tailed with dCTP, annealed to the oligo (dG)-tailed Pst I site of pBR322 and transformed into E. coli strain RR1 (D.A. Johnson, unpublished). The library, representing about 5400 clones, was stored in an ordered array in microtitre wells. The library was displayed on Whatman 541 filters as described in Gergen et al., (1979). Each filter represented 192

independent clones. The cDNA library was screened by the colony hybridization procedure (Grunstein and Hogness, 1975).

2.7.1 Screening with leghemoglobin

The library was screened with 5 x 10^5 cpm/ml of a nick-translated 550 bp Lb insert containing the 3' untranslated region of Lb and most of the coding region. Filters were hybridized as described in Gergen et al. (1979) in 50% formamide, 5X SSC pH 7.0, 250 µg/ml herring sperm DNA for 16 hours at 37°C. The blots were washed 3 times for 30 minutes in 2X SSC at room temperature and exposed overnight at -70°C with a Picker Spectra Blue intensifying screen.

2.7.2 Screening with rRNA

End-labeled 18S and 25S rRNA from soybean was used to screen the library for the positions of ribosomal clones. 1 x 10^5 cpm/ml of end-labeled rRNA was hybridized to filters in 50% formamide, 5X SSC, 250 µg/ml herring sperm DNA and 100 µg/ml poly (A). Hybridization was for 16 hours at 37°C. Filters were washed 3 times for 30 minutes at room temperature in 2X SSC, and exposed overnight at -70°C with a Picker Spectra Blue intensifying screen.

2.7.3 Screening by differential hybridization

The cDNA library was screened by differential hybridization (Williams and Lloyd,1979). About 0.5 X 10^6 cpm/ml of end-labeled nodule poly $(A)^+$ RNA was hybridized to the filters according to Gergen et al. (1979) for 45 hours at 37°C in 50% formamide, 5X SSC, 250 µg/ml herring sperm DNA, 100 µg/ml poly (A). The filters were washed for 30 minutes in 2X SSC and RNAse (10 µg/ml) at room temperature. The filters were then washed 3 times for 30 minutes in 2X SSC at room temperature. The filters were reused after removal of the probe with NaOH as described in Gergen et al. (1979). The filters were treated with 0.05% diethyl pyrocarbonate to remove residual RNAse activity. The library was rescreened in the same manner using end-labeled root poly $(A)^+$ RNA as a probe. Positive clones were selected based on the differential intensities of the hybridizing nodule and root probes as determined by autoradiography.

2.7.4 Screening by hybridization-competition

Screening of the cDNA library by hybridization-competition (Mangiarotti et al.,1981) was accomplished by adding approximately 0.1 µg of end-labeled nodule poly $(A)^+$ RNA per filter to a hybridization solution containing an excess of unlabeled root RNA. 10 ml of hybridization solution contained 50% formamide, 5X SSC, 250 µg/ml herring

sperm DNA, 100 µg/ml poly (A), 0.1 µg of end-labeled probe and 1 mg of total root RNA (competitor). The filters were hybridized for 45 hours at 37°C. Filters were washed in 2X SSC, and RNAse (10 µg/ml) at room temperature, followed by 3 washes in 2X SSC at room temperature. Filters were exposed for 2-3 days at -70°C with Picker Spectra Blue intensifying screen.

2.8 Agarose Gel Electrophoresis

2.8.1 Gel electrophoresis of DNA

Restriction fragments of DNA were separated by electrophoresis through horizontal agarose slab gels. The concentration of the gels varied from 0.8-1.2% depending on the size range of the fragments to be resolved. The electrophoresis buffer contained 40 mM Tris-acetate, 2 mM EDTA pH 8.0 (Maniatis et al.,1982). Lambda Hind III digests or pBR322 Taq I digests were generally used as size markers.

2.8.2 Gel electrophoresis of RNA

RNA was separated by electrophoresis through 1.0-1.5% agarose gels containing formaldehyde (Lehrach et al., 1977). RNA samples for electrophoresis were prepared as described in Maniatis et al. (1982). The gels contained 20 mM MOPS (pH 7.0), 5 mM sodium acetate, 1 mM EDTA, and 2.2 M formaldehyde. The running buffer contained 20 mM MOPS (pH 7.0), 5 mM sodium acetate, 1 mM EDTA and 1.1 M formaldehyde. Following electrophoresis, the gels were washed with 4-5

changes of H_2O for 2 hours and two changes of 0.1 M ammonium acetate for one hour. The gel was stained with 0.5 µg/ml of ethidium bromide in 0.1 M ammonium acetate for 45 minutes. The gels were destained for 2 hours with 2-3 changes of 0.1 M ammonium acetate and then photographed. E. coli 23S, 16S, and 5S rRNAs were used as size standards.

2.9 Transfer of DNA to BIOTRANS (Biodyne)

Restricted DNA was separated electrophoretically in agarose gels. Following electrophoresis, the gels were stained with ethidium bromide and photographed. Gels containing plant genomic DNA restriction digests were then depurinated with 0.25 M HCl for 5 minutes to facilitate transfer of large molecular weight fragments. Gels containing restriction digests of phage and plasmid DNA were not depurinated. The DNA in the gels was denatured by gentle agitation in several volumes of 1.5 M NaCl, 0.5 M NaOH for 30 minutes. This step was repeated for an additional 30 minutes. The denaturation buffer was then replaced by 2 changes of neutralization buffer containing 3 M sodium acetate (pH 5.2). The DNA was transferred to BIOTRANS membranes by the method according to Southern (1975). Following transfer, the DNA was immobilized onto the membrane by U.V. crosslinking for 5 minutes (Church and Gilbert,1984).

2.10 Hybridization of Radiolabeled Probes to DNA Immobilized on BIOTRANS Filters

2.10.1 Standard hybridization conditions

Filters containing immobilized DNA were prehybridized in 4 ml of hybridization solution per 100 cm^2 of membrane in sealable bags for a minimum of one hour. Hybridization solution contained 5X SSPE (0.9 M NaCl, 0.05 M sodium phosphate pH 8.3, 5 mM EDTA), 0.2% SDS, 5X Denhardt's buffer (0.1% ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin) and 500 µg/ml herring sperm DNA. If the probe to be used contained either a (GC) homopolymeric region and/or (AT) homopolymeric region, poly (C) and/or poly (A) were added to a concentration of 50 µg/ml. Blots were then hybridized in 2 ml of hybridization solution per 100 cm^2 of membrane with 1-5 X 10^6 cpm/ml of radiolabeled probe for 16-20 hours at 65-68°C . Blots were washed under moderate stringency conditions in a buffer containing 5 mM sodium phosphate (pH 7.0), 1 mM EDTA and 0.2% SDS at room temperature. High stringency wash conditions were sometimes necessary. These blots were washed with 0.1 X SSC and 0.1% SDS at 65-68°C.

2.10.2 Low stringency hybridization conditions

The DNA blots were prehybridized in 4 ml per 100 cm^2 of hybridization solution containing 33% formamide, 5X Denhardt's buffer, 5X SSC, 50 mM sodium phosphate (pH 7.0), 0.5% SDS, 100 µg/ml herring sperm DNA. If the probe

contained either (GC) or (AT) homopolymeric regions, poly (C) or poly (A) (50 μg/ml) was added to the hybridization solution. The blots were hybridized to 1-5 X 10^6 cpm/ml of random primer labeled probe in 2 ml of hybridization solution per 100 cm^2 of membrane, for 16-20 hours, at 37°C. Following hybridization, the blots were washed in 6 changes of 2X SSC, 0.1% SDS for 30 minutes each at 37°C.

2.11 Hybridization to DNA in Dried Agarose Gels

2.11.1 Preparation of dried agarose gels

Dried agarose gels were prepared by the procedure of Tsao et al. (1983). Following electrophoresis of DNA samples in 0.8-1% agarose gels, gels were stained with ethidium bromide and photographed. DNA samples were denatured by gently shaking the gel in 2 changes of 0.5 M NaOH, 0.15 NaCl for 30 minutes each. The gel was then neutralized by soaking for 1 hour in 2 changes of 0.5 M Tris-HCl (pH 7.7), 0.15 M NaCl. The gel was placed on two sheets of Whatman 3MM paper and covered with Saran wrap. It was placed on a gel dryer and dried with only the vacuum until the gel was almost flat (about 30 minutes). The heater was then turned to 60°C and the gel was dried for an additional hour. The dried gel was stored at room temperature covered with plastic wrap. Before using the gel, it was removed from the filter paper by soaking it in H_2O.

2.11.2 Hybridization of oligonucleotide probes to DNA by direct gel hybridization

End-labeled oligonucleotides (25mers) were used for direct gel hybridization. 5 X 10^6 cpm/ml of oligonucleotide probe was hybridized to dried agarose gels for 20 hours at 65°C in a hybridization buffer containing 5X SSPE (0.9M NaCl, 50 mM sodium phosphate, 5 mM EDTA), 0.5% SDS, 100 µg/ml tRNA. Prior to use, the hybridization solution was heated and filtered through a disposable 0.45 µm filter. 6 ml of hybridization buffer were added per 140 cm^2 of gel. Following hybridization, the gel was washed 3 times for 30 minutes in 6X SSC, 0.1% SDS. This was followed by a stringent wash for 30-60 minutes in 6X SSC, 0.1% SDS at T_d-5°C. This is the temperature which should differentiate single base differences between the oligonucleotide and the target DNA sequence. T_d was estimated (see Wallace and Miyada,1987) using the formula:

$$T_d\ (°C) = 4\ (G + C) + 2(A + T).$$

Blots were then wrapped in plastic and exposed to Kodak XAR-5 film at -70°C using a Picker Spectra Blue intensifying screen.

2.11.3 Hybridization of long DNA probes to DNA by direct gel hybridization

2 X 10^6 cpm/ml of a random-primer labeled fragment was hybridized for 16-20 hours to the agarose gel in a buffer containing 5X SSPE, 5X Denhardt's buffer, 0.5% SDS, 100 µg/ml herring sperm DNA, 50 µg/ml poly (C), 50 µg/ml poly (A) at 65°C. Following hybridization, the gel was washed in 2 changes of 1X SSPE, 0.1% SDS at 60°C for 1 hour each. This was followed by a one hour wash in 0.1X SSPE, 0.1% SDS. The blot was then wrapped in plastic and exposed at -70°C using a Picker Spectra Blue intensifying screen.

2.12 Transfer of RNA to BIOTRANS Membrane

RNA was electrophoretically separated in formaldehyde gels. Following electrophoresis, the gel was washed with several changes of H_2O for 2 hours. Marker lanes were cut from the gel and stained as indicated in Section 2.8.2. The remainder of the gel was transferred overnight to BIOTRANS membrane as described in Thomas (1983) using 20X SSC.

2.13 Hybridization of Radiolabeled Probes to RNA Immobilized on BIOTRANS Filters

2.13.1 Hybridization of nick-translated or random-primer labeled probes to RNA blots

BIOTRANS filters containing immobilized RNA, were prehybridized in sealable bags containing 4 ml per 100 cm^2 of membrane of a hybridization solution that contained 5X Denhardt's buffer, 5X SSC, 50 mM sodium phosphate pH 6.8, 0.1% SDS, 250 µg/ml herring sperm DNA, and 50% formamide for a minimum of one hour. The blots were hybridized using 1-5 X 10^6 cpm/ml of nick-translated or random primer labeled probe in 2 ml of hybridization solution per 100 cm^2 of membrane. The blots were hybridized 16-20 hours at 42°C. Following hybridization, the blots were washed in 3 changes of 2X SSC, 0.1% SDS for 15 minutes each and two changes of 0.1X SSC, 0.1% SDS for 30 minutes each at 50°C. For some experiments, it was necessary to use more or less stringent hybridization and wash conditions than those that have been indicated above. For conditions of higher stringency, hybridization was done at 45°C or 50°C and the final wash was in 0.1X SSC, 0.1% SDS at 55°C. Low stringency hybridizations were done at 30°C or 37°C and the final wash was in 2X SSC, 0.1% SDS at room temperature.

2.13.2 Hybridization of oligonucleotide probes to Northern blots

Northern blots were prehybridized in 4 ml per 100 cm^2 of membrane of filtered hybridization buffer containing, 5X SSPE, 5X Denhardt's, 0.5% SDS, 200 μg/ml tRNA for a minimum of one hour. The blots were then hybridized to 5 X 10^6 cpm/ml of an end-labeled oligonucleotide probe in 2 ml of hybridization buffer per 100 cm^2 of membrane, 16-20 hours at 60°C. The blots were then washed in three changes of 6 X SSC, 0.1% SDS for one hour at room temperature and followed by a stringent wash for 30 minutes in 6 X SSC, 0.1% SDS at T_d - 5°C.

2.14 DNA Dot Blots

Purified plasmid DNA was linearized with Eco RI, extracted once with phenol and once with phenol-chloroform and precipitated overnight at -20°C with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol. The precipitated plasmid was resuspended in 100 μl of TE. An equal volume of 20X SSC was added to the sample and 1 μg aliquots were applied to a BIOTRANS (Biodyne) membrane using a Schleicher and Schuell dot blot apparatus. Following application of the DNA to the membrane, the dot blot was laid on Whatman 3MM paper saturated with 0.5 M NaOH, 1.5 M NaCl for 10 minutes. The membrane was neutralized by laying it on a Whatman 3MM paper saturated with 3 M sodium acetate

(pH 5.5). The DNA was immobilized on the filter by U.V. crosslinking for 5 minutes (Church and Gilbert,1984).

Filters were prehybridized in 4 ml per 100 cm^2 of hybridization solution containing, 5X SSC, 5X Denhardt's buffer, 50 mM sodium phosphate pH 6.5, 0.1% SDS, 250 μg/ml of herring sperm DNA, 200 μg/ml poly (A), 50% formamide for 1 hour. Filters were hybridized to 0.5 μg of end-labeled nodule RNA (specific activity of about 1 X 10^7 cpm/μg) for 45 hours in 2 ml per 100 cm^2 of hybridization solution. Following hybridization, the blots were washed 4 times in 2X SSC, 0.1% SDS for 15 minutes each at room temperature and 2 times in 0.1X SSC, 0.1% SDS for 30 minutes at 50°C. Filters were exposed to Kodak XAR-5 film for several days at -70°C using an intensifying screen. The probe was stripped from the filters and the filters were hybridized to 0.5 μg of end-labeled root RNA (specific activity of about 1 X 10^7 cpm/μg) in the manner indicated above. The filters were then exposed for several days at -70°C to XAR-5 film using an intensifying screen.

2.15 Screening of an Alder Genomic Library

An alder genomic library had been constructed by cloning a partial Mbo I digest of alder genomic DNA into lambda GEM-11 (Promega) using E. coli strain WA802 as the host (M. Lévesque, unpublished). A filter (23 cm x 23 cm), representing about 5 X 10^5 genomic clones, was screened

using a random primer labeled Pst I - Bam HI fragment of cDNA clone 9-11-B (Figure 3.4) under low stringency hybridization conditions as described in section 2.10.2. The filter represented about half of the total alder genomic library. Following exposure to a Kodak XAR-5 film, the probe was stripped from the filter by incubation for 2 hours at 65°C in 100 ml of 10 mM sodium phosphate (pH 7.0), 50% formamide per 100 cm^2 of membrane. The formamide was removed from the filter by washing twice in 250 ml of 2X SSC, 0.1% SDS per 100 cm^2 of membrane for 15 minutes with vigourous agitation. The filter was then rehybridized with the same probe. Only phage that appeared to give a positive hybridization signal on both screenings were chosen for further analysis. Positive plaques were picked. Eighteen plaques that gave positive hybridization signals were pooled into six groups of three clones each, and replated. The plaques were subjected to two subsequent rounds of plaque purification. Following the final screening, 4 hybridizing plaques derived from different pools were chosen for further analysis. These clones were amplified and the DNA was purified.

2.16 Preparation of Templates for DNA Sequencing

Restriction fragments for sequencing were subcloned into either M13mp18/mp19 vectors (Yannish-Perron et al.,1980), or pGEM-3Z/pGEM-4Z (Promega) plasmid vectors.

Fragments for subcloning were determined by restriction analysis of the cDNA clones.

In some cases, restriction of the clones did not yield useful restriction sites for subcloning. To determine initial sequence information, cDNA inserts were excised from the pBR322 vectors using the Pvu I restriction site at position 3735 of pBR322 and the Bgl I site at position 3482 of pBR322. Fragments were purified using a GENECLEAN kit. 0.4 μg of isolated insert was then digested with 0.4 units of Bal 31 enzyme (Amersham) at 30°C in 40 μl of reaction buffer from IBI containing 600 mM NaCl, 12.5 mM $CaCl_2$, 12.5 mM $MgCl_2$, 20 mM Tris-HCl (pH 8.0), 1mM EDTA. 10 μl reaction samples were withdrawn at 2 minutes, 5 minutes, 8 minutes and 10 minutes and reactions were stopped by adding 1 μl of 120 mM EGTA. The 2 and 5 minute samples and the 8 and 10 minute samples were combined and the Bal 31 fragments were purified using GENECLEAN. Fragments were blunt-ended by filling in the ends using Klenow (Maniatis et al., 1982). The fragments were then ligated into an M13mp19 vector, that had been restricted with Sma I or Hinc II, using T4 DNA ligase (Maniatis et al., 1982).

All ligations into M13 phage vectors were transformed into E. coli strain JM101 by the method of Lederberg and Cohen (1974). Recombinant phage were grown and isolated as described in the M13 cloning and sequencing handbook provided by Amersham, except that Luria broth was

used instead of 2X TY broth to grow the bacteria and phage. Single-stranded template for sequencing was prepared as described in the Amersham handbook.

Ligation into plasmid vectors pGEM3Z/pGEM4Z were transformed into E. coli strain DH5α (BRL) by the method of Lederberg and Cohen (1974). Plasmids were isolated by the alkaline lysis method (Birnboim and Doly,1977) and purified on CsCl/ethidium bromide gradients.

2.17 DNA Sequencing

DNA was sequenced using the dideoxy chain termination method (Sanger et al.,1977). Sequencing of M13 recombinant phage DNA was as described in the Amersham M13 cloning and sequencing handbook using the Klenow fragment of E. coli DNA polymerase I. In circumstances where there appeared to be some difficulty in determining sequence due to possible secondary structure, the reaction temperature was raised to 50°C.

Double-stranded plasmid DNA was sequenced according to the protocols supplied with a kit purchased from Promega.

Most sequencing reactions were performed using [α-^{32}P]-dATP (800 Ci/mmol) as the radioactive label. Sequence information was manipulated using MicroGenie software purchased from Beckman.

2.18 Quantitation of RNA by Slot Blot Analysis

2.18.1 Generation of RNA standards

In vitro RNA transcripts were generated to quantitate the amount of RNA corresponding to specific cDNA clones within total RNA isolated from nodulated tissue. Specific fragments corresponding to isolated cDNA clones were subcloned in pGEM3Z or pGEM4Z vectors (see below). RNA was generated exactly as described by Promega (Synthesis of large amounts of RNA), but ^{3}H-CTP was included to assay for incorporation.

2-5 μg of plasmid DNA was linearized with an appropriate enzyme, extracted with phenol-chloroform, and precipitated at -20°C with 0.1 volume of 3M sodium acetate (pH 5.2) and 2.5 volumes of ethanol. Following precipitation, the plasmid was centrifuged for 10 minutes. The pellet was then washed twice with 70% ethanol and once with 95% ethanol and resuspended in TE buffer at a concentration of 1 μg/μl. The plasmid served as a template for RNA synthesis in 100 μl of a transcription reaction containing 40 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, 100 units of RNasin ribonuclease inhibitor, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 0.5 mM UTP, 20 units of Sp6 or T7 RNA polymerase and 15 or 20 μCi of [^{3}H]-CTP (specific activity : 28 Ci/mmol). The reaction was incubated for 90 minutes at 40°C. Following incubation, the DNA template was removed by adding 5 units

of RNase-free DNase I. This was incubated for 15 minutes at 37°C. At this point two 2 µl and two 10 µl samples were removed and set aside to assay for incorporation. The transcription reaction was extracted with phenol-chloroform. The aqueous phase was removed and the phenol-chloroform interface was back extracted with 80 µl of distilled water to ensure that transcribed RNA was not retained at the interface. The two aqueous phases were combined and the RNA was precipitated overnight at -20°C with the addition of 0.1 volume of 3M sodium acetate (pH 5.2) and 2.5 volumes of ethanol.

Incorporation of [^{3}H]-CTP, as measured by TCA precipitation, was used to determine the amount of RNA synthesized. The two 2 µl samples were each spotted on GF/C glass fibre filters. Each of the two 10 µl samples was precipitated with cold TCA as described in Maniatis et al. (1982), and the precipitate was collected by filtering through a GF/C glass filter disc using a filtration manifold. The amount of radioactivity on the glass fibre discs was determined by liquid scintillation counting using the procedure recommended by New England Nuclear. The GF/C filters were placed in glass liquid scintillation vials. After the addition of 0.5 ml of Protosol (NEN) tissue solubilizer to each sample, the vials were incubated at 60°C for 30 minutes. The vials were then cooled to room temperature and 20 µl of glacial acetic acid and 10 ml of

Econofluor (NEN) liquid scintillation fluid were added. The amount of tritiated material was then determined by scintillation counting. The two values obtained from TCA precipitation onto the glass fibre discs never varied by more than 2%. The average of these two values was used to determine the percent incorporation of tritiated CTP. The amount of RNA synthesized was calculated based on the actual (C) content of the RNA molecule synthesized (this could be determined because the sequence of the insert and the pGEM polylinker was known). The amount of RNA synthesized was calculated based on the fact that 2.89 nmol of ribonucleotides is equivalent to 1 μg. 2-8 μg of RNA was synthesized using this procedure depending on the plasmid template used.

The fragments that had been subcloned for generation of RNA were as follows: Pst I - Sau 3A fragment containing a portion of the coding region from the leghemoglobin cDNA clone 47-11-E (Figure 3.8) cloned into pGEM4Z; a Pst I - Bam HI fragment from cDNA clone 9-11-B cloned into pGEM3Z (Figure 3.4); a Xho I - Bgl II fragment containing a portion of the coding region of cDNA clone 15-9-A cloned into pGEM4Z (Figure 3.7); a Sau 3a - Pst I fragment from the cDNA clone 6-9-F was cloned into pGEM3Z (Figure 3.5); a Pst I fragment from cDNA clone 36-1-A (Figure 3.6) was cloned into pGEM3Z. In each case the sense strand of the RNA was synthesized.

Following transcription, it was determined by electrophoresis that RNA corresponding to the cDNA clones 47-11-E, 15-9-A, 36-1-A and 6-9-F migrated as one band; however, the RNA generated from the 9-11-B clone migrated as a smear. This same results was observed after 4 different attempts to generate RNA from 9-11-B; therefore, it was not possible to accurately determine the amount of full length 9-11-B RNA transcript synthesized. Thus, 9-11-B RNA transcripts were not used as standards on slot blots.

2.18.2 RNA Slot Blots

Total RNA isolated from nodulated tissues at various times following inoculation was denatured in a solution containing 50% deionized formamide, 2.2 M formaldehyde, 20 mM MOPS (pH 7.0), 5 mM sodium acetate, 1 mM EDTA for 15 minutes at 55°C and chilled on ice. 5 µg of total RNA was denatured per 40 µl of denaturation solution. Two-fold serial dilutions of the RNA were made using distilled water.

The RNA standards were denatured in the manner described above except 1 µg of RNA was denatured per 40 µl of denaturation solution, and 5-fold serial dilutions were made in distilled water.

An equal volume of 20X SSC was added to each dilution so that the final concentration in each sample was

10X SSC. tRNA was also added as a carrier to each sample with less than 5 μg of RNA to increase the amount to 5 μg.

The RNA dilutions were then applied to BIOTRANS membranes using a Schleicher and Schuell Slot Blotter apparatus. No more than 5 μg of RNA were applied per 7 mm^2 slot because this amount of RNA was approaching the maximum binding capacity of the BIOTRANS membrane (80 - 100 $\mu g/cm^2$, ICN). The RNA was immobilized to the membrane by U.V. cross-linking for 5 minutes (Church and Gilbert,1984).

The blots were hybridized to cDNA specific probes as follows: 1) to determine the amount of Lb sequences, 5 X 10^6 cpm/ml of a random-primer labeled Pst I - Sau 3A fragment containing coding sequence from cDNA clone 47-11-E (Figure 3.8) were hybridized to RNA slot blots using standard hybridization conditions used for northern blots; 2) to determine the amount of 36-1-A specific RNA, 5 x 10^6 cpm/ml of a Pst I fragment (Figure 3.6) were hybridized to RNA slot blots under the conditions of high stringency used for northern blots (50% formamide, 45°C) and washed in 0.1 X SSC, 0.1% SDS at 55°C; 3) to determine the amount of 6-9-F specific RNA, 5 x 10^6 cpm/ml of a Sau 3A - Hinc II fragment were hybridized to RNA slot blots under conditions of high stringency (50% formamide, 50°C) and washed in 0.1X SSC, 0.1% SDS at 55°C; 4) to determine the amount of 15-9-A specific RNA, 5 x 10^6 cpm/ml of a sequence specific oligonucleotide (probe D in Figure 3.14) were hybridized to

RNA slot blots using the conditions described for the hybridization of oligonucleotides to RNA transfers.

To establish standard curves, probe hybridization to the dilutions of ^{3}H-labeled RNA standards was quantified by liquid scintillation of RNA slots cut from the filter and/or densitometric scanning of various exposures using an LKB Ultrascan XL enhanced laser densitometer, and plotted versus the amount (ng) of standard RNA applied to the filter. The amount of mRNA in the nodule RNA samples was determined by converting hybridizable counts and/or the area determined by densitometry to ng of RNA using the standard curve. In order to compare the levels of the different nodulin RNAs at a given time, the number of nmoles of RNA was calculated. All values were calculated only from dilutions of RNA that fell within the linear range of the standard curves.

CHAPTER 3

RESULTS

3.1 Screening of a Soybean cDNA Library for Nodule-Specific Clones.

The development of soybean root nodules is a complex process which involves the induction and expression of specific soybean genes (Auger and Verma, 1981). As an approach to investigating some of the changes that occur during nodule development, cDNA clones corresponding to nodule specific mRNAs were isolated from a nodule cDNA library. The cDNA library had been constructed from poly $(A)^+$ RNA isolated from 2 - 3 week old soybean nodules (D.A. Johnson, unpublished).

3.1.1 Preliminary screening of the library for leghemoglobin and rRNA clones.

About 5400 cDNA clones stored in microtitre dishes were first screened using the colony hybridization procedure (Grunstein and Hogness, 1975) for the presence of leghemoglobin (Lb) cDNA clones. Leghemoglobin mRNA constitutes 12 - 15% of the poly $(A)^+$ RNA in soybean nodules (Fuller et al., 1983); therefore it was expected that Lb cDNA clones would constitute a relatively large portion of the cDNA library. Preliminary identification of the Lb cDNA clones facilitated further screening of the cDNA library, since we were interested in isolating nodule-specific clones

other than leghemoglobin. The initial screening of the library was attempted using a 200 bp Taq I - Sau 3A fragment derived from the Lb clone 7-56 (a gift from K. Marcker; Jensen et al.,1981). However, this fragment comprised only a portion of the coding region of the leghemoglobin cDNA. In order to determine the position of all of the Lb clones in the library, it was necessary to have a longer probe. The Lb 7-56 cDNA had been inserted into pBR322 by AT tailing; thus, the full insert could not be cleanly excised from the pBR322 vector (Truelsen et al.,1979). Therefore, the Taq I - Sau 3A fragment from 7-56 was hybridized to one filter from our cDNA library that represented 192 individual clones. One of the clones, that hybridized strongly to the 7-56 fragment, was found to contain a 550bp insert that could be excised using the restriction enzyme Pst I. Since the size of the Lb mRNA in the nodule is approximately 750 bases, it was assumed that this clone (named 47-11-E) would be suitable as a probe for detecting the presence of Lb cDNA clones in the library. In fact, only about 2.5% of the library hybridized to 47-11-E. Since it has been estimated that the soybean nodule poly $(A)^+$ message is 12-15% Lb message (Fuller, et al., 1983), 2.5% representation in our library was lower than expected. The under-representation of leghemoglobin sequences in our library may have been due to a number of reasons:

1) The cDNA library contained pBR322 sequences that contained no inserts;

2) Some leghemoglobin inserts may have been so short that they could not form stable hybrids, or the hybridization signal could not be detected above background;

3) The Lb clone 47-11-E was not completely full length, and therefore was not able to detect certain small Lb cDNA clones that contained sequences that were not contained within the 47-11-E probe;

4) The RNA isolation procedure (Chirgwin et al.,1979) resulted in some DNA contamination of the RNA; thus some of the cDNA clones may have contained genomic DNA inserts (For an example of this, see Appendix I);

5) Some clones represented rRNA contaminants. Ribosomal RNA clones are common contaminants of oligo dT-primed cDNA libraries (Williams and Lloyd,1979; Sullivan et al.,1980), since ribosomal RNA may be present in the poly $(A)^+$ fraction even after several rounds of oligo-dT chromatography. These clones may, therefore, be identified as false positives when screening a cDNA library. To identify the positions of such ribosomal clones, the library was screened with $[\gamma\text{-}^{32}P]$-end-labeled soybean 25S and 18S rRNA that had been isolated by sucrose gradient centrifugation. Screening in this manner indicated that about 0.4% of the library represented ribosomal clones.

The preliminary identification of Lb and rRNA clones in the library was useful for further screening of the library since these particular clones could be used as hybridization controls. Also, because of their prior identification, they would not be mistakenly identified as new nodule-specific clones.

3.1.2 Screening for nodule-specific cDNA clones.

The library was screened for the presence of nodule-specific clones using two procedures. The first approach involved the differential hybridization (Williams and Lloyd,1979) of [γ-^{32}P]-end-labeled root and nodule poly (A)$^+$ RNA to filters. The second approach involved a competition-hybridization procedure (Mangiarotti et al, 1981) in which the filters were hybridized to end-labeled nodule poly (A)$^+$ RNA in the presence of a several hundred fold excess of root RNA.

Twenty-four clones were selected that had not hybridized to the Lb probe and appeared to hybridize specifically to nodule poly (A)$^+$ using both the differential and competition hybridization approach. There was some problem using the Whatman 541 paper because there was often significant background on the filter sometimes making it difficult to discern the nodule-specific clones. But the twenty-four clones that were chosen seemed to hybridize

significantly above background using both hybridization approaches.

Figure 3.1 shows an example of the screening of one filter representing 192 clones. Figure 3.1A demonstrates the result of hybridization to the Lb clone 47-11-E. Five putative Lb clones were detected. Figure 3.1 B and C show the result of a differential hybridization experiment. In Figure 3.1B, the nodule probe hybridized significantly above background to the five Lb clones and to a clone designated "C". In Figure 3.1C, the root poly $(A)^+$ probe did not significantly hybridize to the leghemoglobin clones or clone "C". Both probes did, however, hybridize to the ribosomal clone designated "R" in Figure 3.1 indicating that both probes did contain some rRNA. This clone had been identified as ribosomal by prior screening with end-labeled rRNA.

To confirm that the twenty-four chosen clones were in fact nodule-specific, plasmid DNA isolated from the clones was dot blotted onto BIODYNE filters and hybridized to end-labeled root and nodule probes (Figure 3.2). Figure 3.2 demonstrates that twenty-three of the twenty-four clones selected from the library, hybridized specifically to nodule RNA and not to root RNA. The Lb clone 47-11-E and two ribosomal clones selected from the library were included on the dot blot as positive hybridization controls and the plasmid pBR322 was included as a negative hybridization

FIGURE 3.1 SCREENING OF A SOYBEAN NODULE cDNA LIBRARY

A Whatman 541 filter representing 192 different clones from the cDNA library, was hybridized sequentially using the following probes:

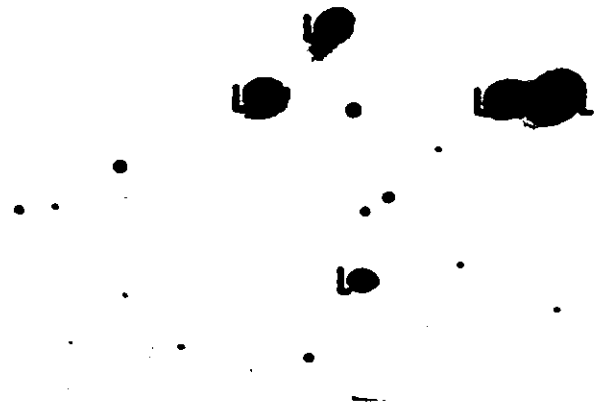
A) Lb 47-11-E Pst I insert;

B) [γ-^{32}P] - end-labeled nodule poly (A)$^+$ RNA;

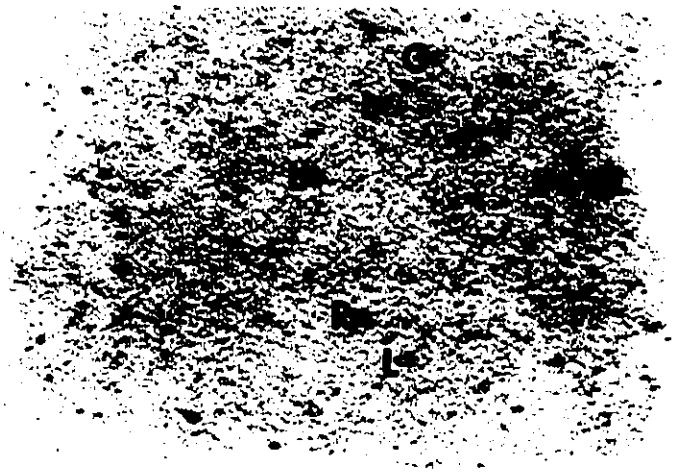
C) [γ-^{32}P] - end-labeled root poly (A)$^+$ RNA;

The letter "L" designates the position of Lb clones, "R" designates the position of a ribosomal clone and "C" designates the position of a putative nodule-specific cDNA clone.

A



B



C

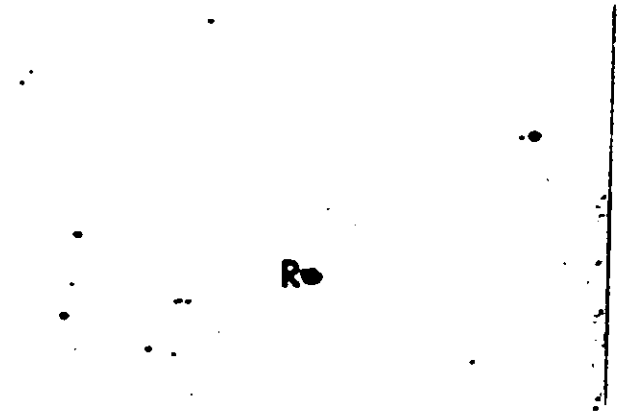
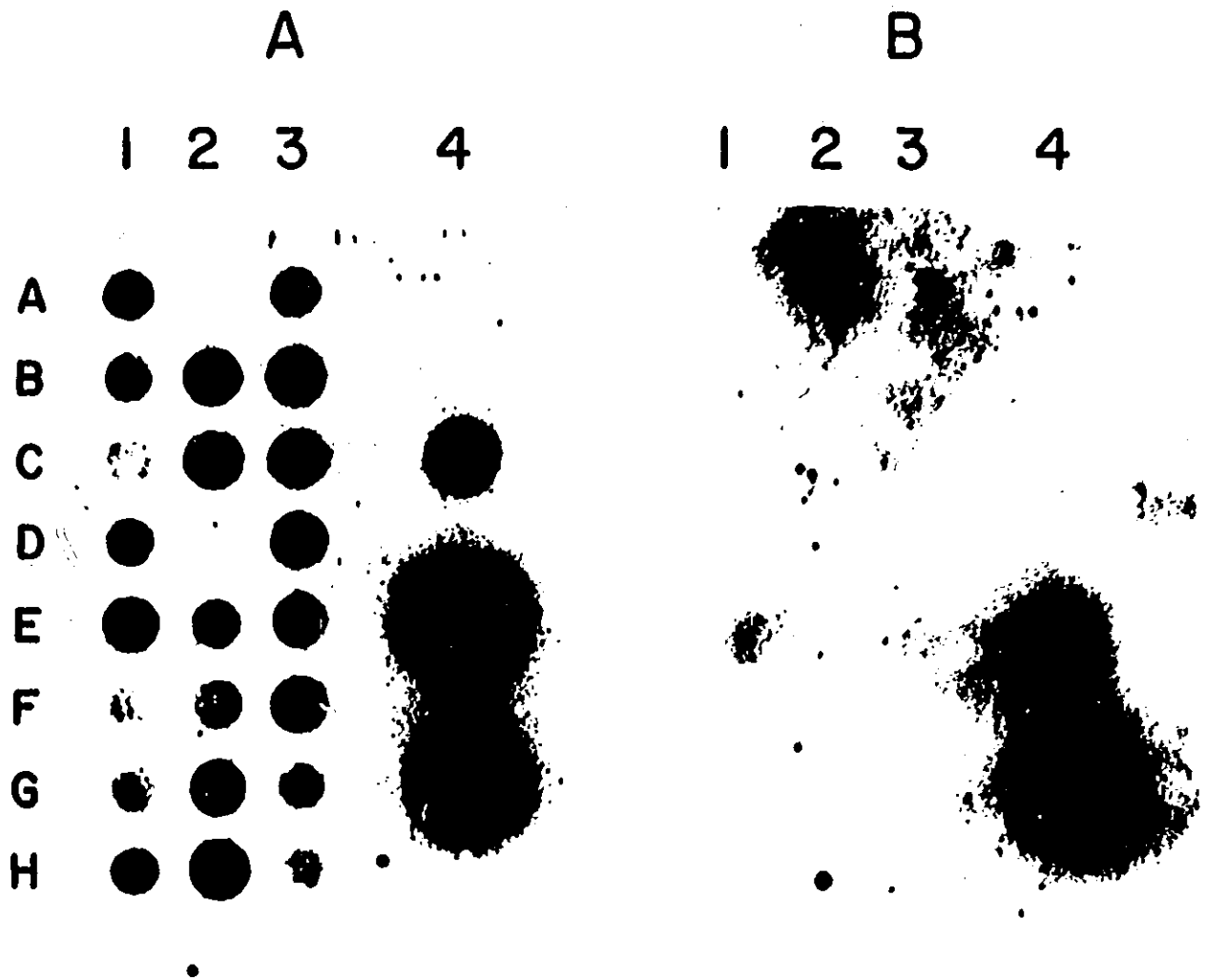


FIGURE 3.2 VERIFICATION OF NODULE-SPECIFIC CLONES BY DOT BLOT ANALYSIS

1 μg samples of plasmid DNA containing cDNA inserts were digested with Eco RI, and dot blotted onto BIODYNE membranes using a Schleicher and Schuell Minifold filtration manifold, and hybridized to end-labeled nodule (A) or end-labeled root (B) poly $(A)^+$ enriched RNA. Dots A-H in Lanes 1, 2 and 3 represented 24 cDNA clones that had been selected by screening of the cDNA library. Dot A-4 contains pBR322 plasmid DNA (vector). Dot C-4 contains clone 47-11-E (Lb) plasmid DNA. Dots E-4 and G-4 contain two different ribosomal clones selected from the library.


A
B
1 2 3 4
1 2 3 4
A
B
C
D
E
F
G
H

control. Thus, the results of dot blot hybridization indicated that almost all of the clones isolated by screening of the library were nodule-specific.

3.2 Characterization of the Nodule-Specific cDNA Clones.

To determine how many of the selected clones represented unique nodulin sequences, the clones were digested with the enzyme Pst I to liberate their inserts. Labeling of the inserts and hybridization to the other clones should have helped discern which clones represented related sequences. Unfortunately, virtually all of the clones had only one or no Pst I sites. Therefore, most of the Pst I sites had not been regenerated following (GC) homopolymer tailing and insertion of the the cDNA into pBR322 vector. The inability to isolate the specific inserts made it impossible to use cross-hybridization of the inserts as a method of differentiating between the clones.

Restriction analysis of the clones with various enzymes did not yield much more useful information as to the relatedness of the various clones.

Therefore, to distinguish between the twenty-four cDNA clones, each of the cDNA containing plasmids was labeled by nick-translation and used as a hybridization probe against total soybean DNA digested with Eco RI and against total RNA isolated from soybean nodules and roots.

Based on these northern and Southern blot analyses, the clones were classified into five groups summarized in Table II. Each group has one representative clone listed. The representative clone chosen was the one which appeared to have the largest insert of each group based on restriction analysis of the plasmid. Northern blot analysis clearly demonstrated the nodule specificity of the cDNA clones (Figure 3.3). The clones hybridized specifically to nodule RNA and no hybridization to 3 day root RNA was seen even after prolonged exposure. Also, no hybridization could be detected to RNA isolated from 17 day or 25 day roots (data not shown). However, it is possible that these nodule sequences may be expressed in the root at a level below our level of detection.

Under the standard hybridization conditions used, the group I clones cross-hybridized to two messages of 1.1 and 1.5 kb, but hybridization to the 1.1 kb message was always stronger. Increasing the stringency of hybridization and washing eliminated the hybridization to the 1.5 kb message, indicating that this group of clones coded for a 1.1 kb message. Similarly, under standard hybridization conditions, the clones in group II hybridized to a 1.5 kb message but some of the clones also hybridized to a 1.1 kb message. Hybridization to the 1.1 kb message was never seen under higher stringency hybridization conditions. The fact

Table II CHARACTERIZATION OF SOYBEAN NODULIN CLONES

Group	Representative Clone	No. of Clones[1]	Size of Hybridizing RNA (kb)[2]	Size of Genomic Eco R1 fragments (kb)[3]
I	9-11-B	6	1.1* (1.5)	7.1*, 1.6, 0.6
II	6-9-F	11	1.5* (1.1)	7.1*, 1.6, 0.6
III	15-9-A	2	1.0 0.8	11.5, 4.7, 4.4
IV	36-1-A	2	1.2	5.5
V	Lb	2	0.75	1.4, 4.3, 5.3, 6.0, 7.5, 9.7, 12

1) The clones were classified into each group based on at least two independent determinations of RNA size and the sizes of hybridizing soybean genomic Eco R1 fragments.

2) The molecular weight of the hybridizing RNA species was calculated from the mobilities of E. coli 23S, 16S and 5S RNA and Lb RNA in the same gel. The sizes indicated are the average of at least two independent determinations.

3) The sizes of the genomic Eco R1 fragments were calculated from the mobilities of lambda Hind III markers in the same gel. The sizes indicated represent the averages of at least two independent determinations.

* Indicates that the band hybridizes much more strongly than the other hybridizing bands indicated.

FIGURE 3.3 NORTHERN BLOT ANALYSIS OF RNA CORRESPONDING TO NODULIN cDNA CLONES

10 μg of total RNA from 3 day soybean roots (R) or 3 - 4 week nodules (N) were electrophoresed on a 1.2% agarose gel containing formaldehyde. The transferred RNAs were probed with 1 x 10^6 cpm/ml of the indicated nick-translated cDNA clone. 25S and 18S indicate the positions of the plant large ribosomal RNAs.

6-9-F 9-11-B 15-9-A 36-1-A 47-11-E

R N R N R N R N R N

that the clones in groups I and II cross-hybridized to similar sized bands suggested that these two groups of clones shared significant sequence homology. Sequencing of the group I representative, 9-11-B, and the group II representative, 6-9-F, verified that this was the case (See Section 3.3).

The clones from group III hybridized to two bands of about 0.8 and 1.0 kb in size even using stringent hybridization criteria. The clones also hybridized strongly to three distinct Eco RI fragments on a Southern blot suggesting that the group III clones were members of a multi-gene family. Further work presented in Section 3.6 demonstrates that this is in fact the case.

The fourth group of clones hybridized to a 1.2 kb message. The clones hybridized strongly to a 5.5 kb Eco RI genomic DNA fragment. Under standard conditions of washing, the clones also hybridized to 11.5, 7.2, 4.7 and 4.4 kb Eco RI fragments; but increasing the stringency of washing removed these additional hybridizing bands. These results suggested that the group IV clones also represented members of a multi-gene family.

The fifth group of clones includes two leghemoglobin sequences. These sequences did not cross-hybridize to the Lb 47-11-E probe originally used to screen the library, but northern and Southern hybridization analyses indicated that these clones represented Lb clones.

Partial sequencing of one of these clones, 7-3-H, indicated that it represented a short insert containing the 5' region of a Lb message. Since the clone 47-11-E did not contain this region, it would explain why the 7-3-H clone had not been detected by hybridization to the 47-11-E probe.

Thus, analysis of the twenty-four cDNA clones revealed that, other than Lb, they represent at least 4 different classes of nodulin clones.

3.3 Sequence Analysis of Nodulin cDNA Clones

Each of the representative clones from the four groups of nodulins was sequenced by the dideoxy chain-termination method (Sanger et al.,1977). Fragments of the cDNA clones were subcloned into M13mp18/mp19 or the pGem3/4 (Promega) vectors by means of various internal restriction sites (determined by restriction analysis) and sites within the pBR322 vector. Because the Pst I sites were frequently not regenerated, the use of the Pst I site for subcloning was limited. Initial restriction analysis also did not always yield useful internal restriction sites (or in some cases, specific restriction fragments would not clone into the M13 vectors). Therefore, to determine some initial sequence information, fragments containing the cDNA inserts were excised from the pBR322 vector using restriction sites within the vector, and subjected to Bal 31 digestion. Following Bal 31 digestion for different lengths of time,

the fragments were cloned into M13mp19 and sequenced. This approach was useful for obtaining some initial sequence information for the clones 6-9-F and 15-9-A.

Sequence analysis of the four nodulin clones indicated that these clones were related, and that they represented four members of a multi-gene family that had been previously described (Jacobs et al., 1987; Sandal et al.,1987).

3.3.1 Sequence of 9-11-B

The sequence and sequencing strategy of 9-11-B are presented in Figure 3.4. The 9-11-B cDNA clone has an 825 bp insert (excluding the (GC) homopolymer tails). Comparison of this sequence to published sequences indicated that 9-11-B corresponded to the sequence of nodulin-23 (Mauro et al.,1985). Two versions of the nodulin-23 sequence have been published (Mauro et al.,1985; Sengupta-Gopalan et al.,1986). The sequence of 9-11-B corresponds to positions 12 - 836 of the nod C51 (nodulin-23) sequence of Sengupta-Gopalan et al. (1986). 9-11-B differs from the nod C51 sequence at only one position (position 234 of 9-11-B, See Figure 3.4) but the change is a third position substitution that does not alter the amino acid sequence. This difference between the two sequences may only reflect the fact that 9-11-B was isolated from a different soybean cultivar than was used by Sengupta-Gopalan et al. (1986).

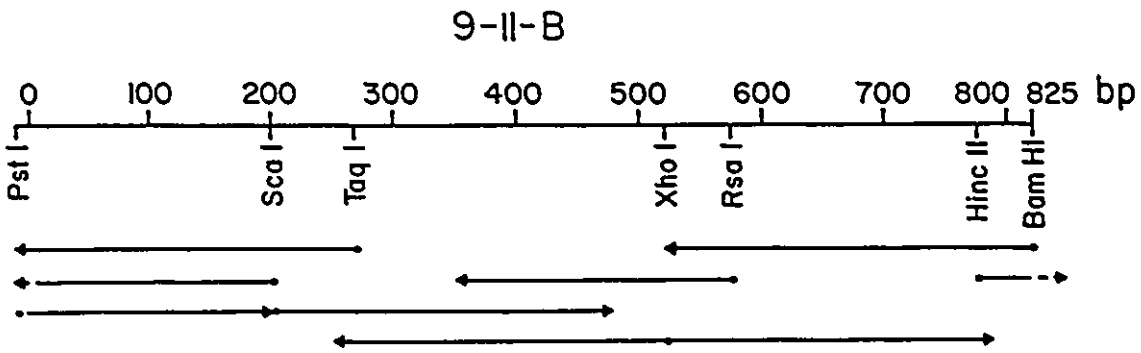
FIGURE 3.4 SEQUENCE OF THE cDNA INSERT FROM CLONE 9-11-B

A) Restriction map and strategy used to sequence 9-11-B. The arrows represent the direction and extent of sequencing from the positions indicated by solid dots. The position 0 indicates the first nucleotide of the cDNA clone adjacent to the (GC) homopolymer tail. The dashed line indicates the sequence of the (GC) homopolymer and the pBR322 vector.

B) Nucleotide sequence and predicted amino acid sequence of 9-11-B.

The nucleotides are numbered on the left hand side. Nucleotide 1 is the first base of the cDNA clone. The amino acids encoded by the open reading frame are shown below the nucleotide sequence. The asterisks (*) mark the position of the termination codon. The one base difference between this sequence and that of the nod C51 sequence of Sengupta-Gopalan et al. (1986) is shown above the nucleotide sequence at position 234.

A



B

```
1                                                    AGGTGCTAATTATTAGCGGCTAGCTAAGTAACCAAGTAATTA

43   ATG GAG AAA ATG AGG GTG ATA GTA ATT ACT GTA TTC CTA TTT ATA GGT GCA GCA ATT GCA
      M  E  K  M  R  V  I  V  I  T  V  F  L  F  I  G  A  A  I  A

103  GAA GAT GTT GGT ATT GGT CTC CTT AGC GAA GCT GAG GCG TAT GTG TCT CCT AAG TTA AAA
      E  D  V  G  I  G  L  L  S  E  A  E  A  Y  V  S  P  K  L  K

163  AAG TTC ATC ACA CCT TGC ACT TCG CAT GTT GGT GAA ACA TGC AGT ACT ACT AGT AGT AGT
      K  F  I  T  P  C  T  S  H  V  G  Q  T  C  S  T  T  S  S  S
                A
223  GGA AGT GAA GCG TTA ATG CAG AAC CAG GGT GGG TTG GCT CTT TGC CTT TTC GAT TCT ATG
      G  S  E  A  L  M  Q  N  Q  G  G  L  A  L  C  L  F  D  S  M

283  GAG AGA TGC TTG GTA GAC CAT GGT GCC CAA CTT TAT CAA ACA AGT GTT ACT AAC CTT CAA
      E  R  C  L  V  D  H  G  A  Q  L  Y  Q  T  S  V  T  N  L  Q

343  GTT GAA CCT TCT GAA GTA TTT CCA AGG AAG AAT AAT CCA CAA GGT GGA CGT AAG TCC AAA
      V  E  P  S  E  V  F  P  R  K  N  N  P  Q  G  G  R  K  S  K

403  TTA GAT GAC CAT CAA GTT CAA CCC TTA TCA TTT CGA TTA CCA CCA TTT CGA TTA CCA CCA
      L  D  D  H  Q  V  Q  P  L  S  F  R  L  P  P  F  R  L  P  P

463  ATG CCA AAA CTA GGA CCA ACA AGT CCG ATT ATA AGA ACG ATT CCA TCA CCA CCC ATA GCT
.     M  P  K  L  G  P  T  S  P  I  I  R  T  I  P  S  P  P  I  A

523  CCT CGA GAT TTG TCA CTC ATT GAG ACT ATA CAA TTA CGA ACT GCC TTG AGA ACC TGT ACT
      P  R  D  L  S  L  I  E  T  I  Q  L  R  T  A  L  R  T  C  T

583  CAT GTC ACT GCA CGA ACT TGT CTC ACT GCT CCA AAT GTT GCC ACA TCT GAT TTA GAG GCT
      H  V  T  A  R  T  C  L  T  A  P  N  V  A  T  S  D  L  E  A

643  TGT CTC ACT CCA TCC ATG AAT CAA TGC ATC TAT CCT CGT GGA GCT GAA TAT GGT AGC CCG
      C  L  T  P  S  M  N  Q  C  I  Y  P  R  G  A  E  Y  G  S  P

703  CCT ATT AGG GCG TAA ATTATTTCAGGCATAAAGGAAAATATGTTGGAAGTATTGCTAGTAGAAGAATAGCCAGA
      P  I  R  A  ***

777  CATGCGTCAACTAGATTTAAGCTAGCTAGCTTGTTTTATGATGAAGGAT
```

9-11-B includes the full open-reading frame of the nodulin-23 sequence including the initiation and termination codons. 9-11-B is, however, missing a portion of the 5' and 3' untranslated regions compared to nodulin-23 or nod C51, likely due to incomplete first and second strand synthesis during construction of the cDNA library. The 9-11-B sequence has a Pst I site at the 5' end of the insert (generated during construction of the cDNA library). Fortunately, sequencing from the Pst I site through the (GC) homopolymer region was possible using the Klenow enzyme and increased reaction temperatures. The Bam HI site at the 3' end of the sequence is not contained within the nod C51 sequence of Sengupta-Gopalan et al. (1986). The Bam HI site is followed immediately by the (GC) homopolymer; thus, it appears that the Bam H1 site was fortuitously generated by the cDNA cloning.

3.3.2 Sequence of 6-9-F

The sequence of the clone 6-9-F is presented in Figure 3.5. The 6-9-F cDNA insert is 589 bp in length (excluding the (GC) homopolymer tail). When the sequence of 6-9-F was compared to published sequences, it was found that 6-9-F was a truncated version of the nod E27 (nodulin-44) sequence of Sengupta-Gopalan et al. (1986). It corresponded to positions 694-1283 of the nod E27 sequence. The sequence of 6-9-F has one nucleotide difference from nod E27 at

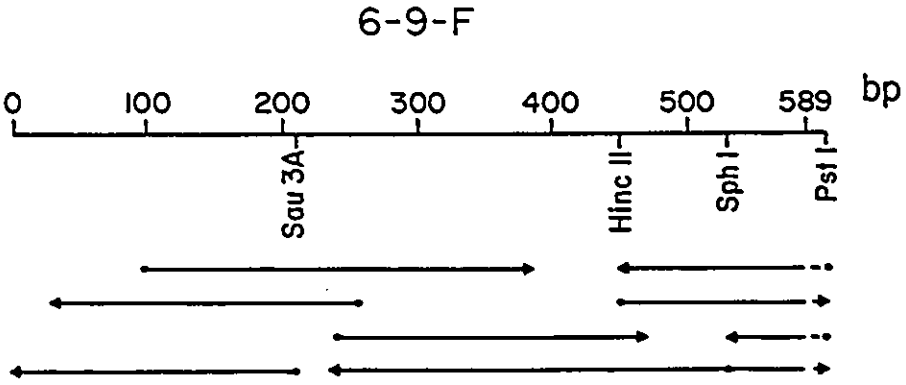
FIGURE 3.5 SEQUENCE OF THE cDNA INSERT FROM CLONE 6-9-F

A) Restriction map and strategy used to sequence 6-9-F. The arrows represent the direction and extent of sequencing from the positions indicated by the solid dots. The positions of the solid dots do not necessarily correspond to restriction sites because Bal 31 digested cDNA inserts were also subcloned and sequenced. The position 0 indicates the first nucleotide of the cDNA clone beside the (GC) homopolymer and 589 indicates the last nucleotide of the cDNA clone. The dashed line represents the sequence of the (GC) homopolymer tail.

B) Nucleotide sequence and predicted amino acid sequence of 6-9-F.

The nucleotides are numbered on the left hand side. Nucleotide 1 represents the first base of the cDNA insert. The dots at the beginning of the sequence indicate that the sequence is truncated. The amino acid sequences encoded by the open reading frame are below the nucleotide sequence. The asterisks indicate the position of the termination codon. "Underline" represents the putative polyadenylation signal. The one base difference in the nod E27 sequence of Sengupta-Gopalan et al.(1986) is indicated above the nucleotide sequence at position 262.

A



B

```
1      ..T CCA TTT TCA CCT CGT GGT CGT CGC TCC AAA TTA GAT AAC CAT CAA ACA GAT GCT GGA
           P   F   S   P   R   G   R   R   S   K   L   D   N   H   Q   T   D   A   G

59     ACT CTT GGA AGA GTA ATT CCA CTT CCA CCG ATT CGA CCA GGA CCG CCA TTA AAA ATA ATT
        T   L   G   R   V   I   P   L   P   P   I   R   P   G   P   P   L   K   I   I

119    CCA TTT CCA GGA ACT AAC ATA GTT CCA TTT CCA AAA CCT TAC ATT CCT CAT TCA TAC AAT
        P   F   P   G   T   N   I   V   P   F   P   K   P   Y   I   P   H   S   Y   N

179    ACA ATT ACG AAC CTT CTT GGC GCA CGA CGT GAT CAA GTT CAA GAT TTG CCA CTC CTC ATT
        T   I   T   N   L   L   G   A   R   R   D   Q   V   Q   D   L   P   L   L   I
                                    A
239    CAG ACT ACA CAA TTA CGA ACC GTC TTG GGA ATC TGT AGT CAT GTC ACT GCA CGA ACT TGT
        Q   T   T   Q   L   R   T   V   L   G   I   C   S   H   V   T   A   R   T   C

299    CTC ACT GCT CCA AAT GTT GCT ACA TCA GAT TTA GAG GCT TGT CTC ACT CCT TCC ATG AAT
        L   T   A   P   N   V   A   T   S   D   L   E   A   C   L   T   P   S   M   N

359    CAA TGC GTC TAT CCT CCT GGA GCT GAA TCT GGT AGC CCG CCA ATA TAA TTAGGGCATACAGGA
        Q   C   V   Y   P   P   G   A   E   S   G   S   P   P   I  ***

422    AAATATGTTGGAAGAATAAATGGACATGTCAACTAGATTTAAGCAAGCTAGCTTGTTTTATGATGAAGGATTTTGTATC

501    CCTTCAAAATTCCAATAGCTGCATCACTGGCATGCATAAATAAATAGCTGAATGTTACGAACGATATGGTGTTCCTACT

580    AAATAAATAG
```

position 262 (Figure 3.5), but this change does not alter the amino acid sequence. This difference again may only reflect the different cultivars used in our work and that of Sengupta-Gopalan et al. (1986). The sequence of 6-9-F did not contain a poly (A) region. The partial sequence of an additional clone from our cDNA library called 23-12-A (another member of the group II class of nodulins - Table II), was identical in sequence to 6-9-F; but contained an additional 21 bases at the 3' end before a poly (A) stretch of sequence. These additional 21 bases were also identical to those of Sengupta-Gopalan et al. (1986). The putative polyadenylation signal, is located at position 582 of the 6-9-F sequence. Unfortunately, the (GC) homopolymer region at the 5' end of the 6-9-F sequence caused difficulties in sequencing even after increasing the reaction temperature. Therefore, the 5' end of the sequence could only be determined for one strand. Sequencing of the strand was repeated several times and was found to be quite unambiguous. In addition, this region completely supported the nod E27 sequence of Sengupta-Gopalan et al. (1986).

Comparison of the 9-11-B (nodulin-23) sequence to 6-9-F (nodulin-44) indicates that these two clones share extensive regions of homology. Analyses by others (Sengupta-Gopalan et al.,1986; Jacobs et al.,1987; Sandal et al.,1987) and this work demonstrate that nodulin-23 (nod C51) and nodulin-44 (nod E27) share regions of greater than

85% sequence homology at their 5'and 3' ends. This significant homology then explains why the clones cross-hybridize to similar sized messages on a northern blot (using standard hybridization conditions). Nodulin-23 and nodulin-44 are members of a small multi-gene family in soybean (Jacobs et al.,1987; Sandal et al.,1987) called the "Nodulin-A" family (Verma and Delauney,1988).

3.3.3 Sequence of 36-1-A

The sequence of the 36-1-A insert presented in Figure 3.6 is 490 bp in length. 36-1-A was found to be a truncated version of nodulin-22, another "Nodulin-A" family member (Sandal et al.,1987). The sequence of 36-1-A is contained within the coding region of nodulin-22 and is identical to the sequence published by Sandal et al. (1987). The sequence corresponds to positions 198 - 606 of the nodulin 22 gene sequence (Sandal et al.,1987). 36-1-A is also very similar to the sequence of nodulin-27 (Jacobs et al.,1987 - See Figure 3.6). Nodulin-22 and nodulin-27 have been treated as two genes (Verma & Delauney,1988); but given that nodulin-22 and nodulin-27 are more than 98% similar, it is possible that they could represent alleles of the same gene (See Discussion).

The ends of the 36-1-A insert (Figure 3.6A) could not be sequenced on both strands because the (GC) homopolymer tails proved refractory to conventional

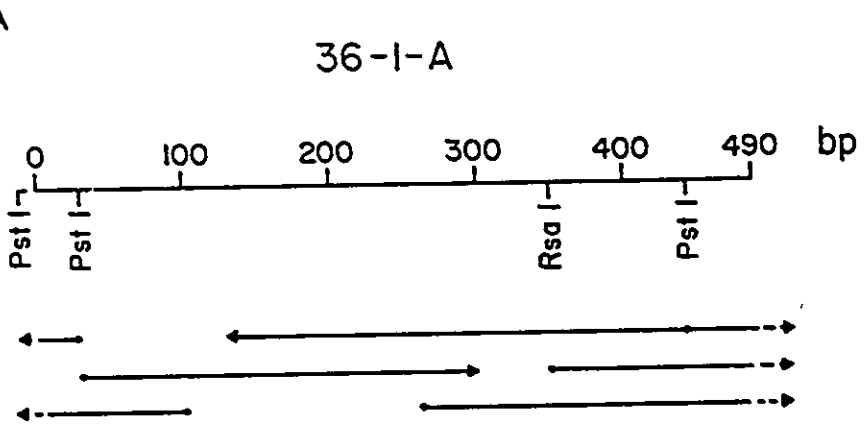
FIGURE 3.6 SEQUENCE OF THE cDNA INSERT FROM CLONE 36-1-A

A) Restriction map and strategy used to sequence 36-1-A. The arrows represent the direction and extent of sequencing from the positions indicated by the solid dots. The position 0 indicates the first base of the cDNA insert and the position 490 represents the last base of the insert. The dashed line represents the (GC) homopolymer tail and/or pBR322 vector.

B) Nucleotide sequence and predicted amino acid sequence of 36-1-A.

The nucleotides are numbered on the left hand side. Nucleotide 1 represents the first nucleotide of the cDNA insert. The dots at the beginning of the sequence indicate that the cDNA is truncated. The differences of the nodulin-27 sequence compared to 36-1-A are indicated above the nucleotide sequence. A,C,G or T above the 36-1-A nucleotide sequence represents a base change. $(N)^{*}$(where N = A,C,G or T) represents a base insertion in the nodulin-27 (Jacobs et al.,1987) sequence compared to 36-1-A. $(N)^{-}$ represents a base deletion in nodulin-27 compared to 36-1-A.

A



B

```
                                                                  C-          T-
1     ..A GAA GCT GGT GGT ATT GGT GAT GCC ATT ACT CCT GCA GAA GGC AAA GCC ACT AAT CTT
        E   A   G   G   I   G   D   A   I   T   P   A   E   G   K   A   T   N   L

      A-                                                  C
59    CAA GCG TAT GAG TCA GCT AGA TTC AAA AAG TTT GTG ACA CAT TGC AGC TCA CAT GTT GCT
        Q   A   Y   E   S   A   R   F   K   K   F   V   T   H   C   S   S   H   V   A

                                                    G-              G-
119   CAA ACA TGC AGT GGA AAT GAT CCA TTG CAT CAT CAG GAA GGT GGC CAT GGA ATA AAC GTT
        Q   T   C   S   G   N   D   P   L   H   H   Q   E   G   G   H   G   I   N   V

179   CCA CTT GGG TTG TCA TTT TGC CTT TTT GAT TCT ATG GAG AAA TGC TTG GGA GAC CAT GAA
        P   L   G   L   S   F   C   L   F   D   S   M   E   K   C   L   G   D   H   E

                      G
239   GCC AAA CTT ATA GAT CCC AAC CCA GGT CCC ATG TCG GCT ATT CCT AAT TCA ATC CAA TCT
        A   K   L   I   D   P   N   P   G   P   M   S   A   I   P   N   S   I   Q   S

299   CAG CAA CTC CTC ATT GAG ACT GTA AAA TTC AGA ACT GTC TTG AAA ACC TGT ACT CGT GTT
        Q   Q   L   L   I   E   T   V   K   F   R   T   V   L   K   T   C   T   R   V

              A                   G
359   AGT GCA CAA TTT TGT TTA ACT GCT CCT AAC GTT GAT ACA TCG GTT TTA CCG GCA TGT CTC
        S   A   Q   F   C   L   T   A   P   N   V   D   T   S   V   L   P   A   C   L

419   GGG CCA TCT CTC AAT CAA TGT GTT TAT CCT GCA GCT GAT GCA TTT ACA CCT GGC CCG CCG
        G   P   S   L   N   Q   C   V   Y   P   A   A   D   A   F   T   P   G   P   P

479   CTC GAA CTC CCA
        L   E   L   P
```

sequencing methods. Sequencing of the one strand was, however, repeated several times and the sequence was quite unambiguous. Also, the sequence of the ends completely supported the sequence of Sandal et al. (1987).

3.3.4 Sequence of 15-9-A

The cDNA clone 15-9-A contains a 653 bp insert (excluding the flanking (GC) homopolymer regions and a 3' poly (A) region). The sequence and sequencing strategy of 15-9-A is presented in Figure 3.7. Unfortunately, there were no Pst I sites regenerated in this clone. Therefore, subclonings of 15-9-A for sequencing made use of internal sites within the clone as well as restriction sites within pBR322, in addition to Bal 31 digestion of the cloned insert. The homopolymer regions (homopolymeric (A) region generated from the poly(A) tail, and a (C) region generated from the cDNA cloning) at the 3' end of 15-9-A made it impossible to sequence the second strand for the last 58 bp of the sequence; however, sequencing of several different subclones from this region repeatedly yielded the sequence presented in Figure 3.7. 15-9-A is a truncated cDNA which has an open reading frame for the first 353 bases followed by a 300 bp 3' non-coding region. The termination codon TAG is at position 354 and a putative polyadenylation signal occurs at position 626, 29 bases 5' to the poly (A) tail on the sequence. The sequence of 15-9-A is similar to the

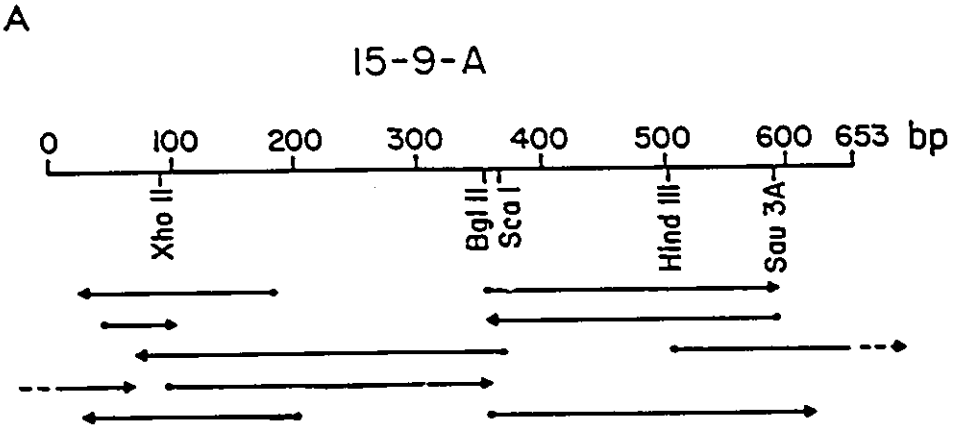
FIGURE 3.7 SEQUENCE OF THE cDNA INSERT FROM CLONE 15-9-A

A) Restriction map and strategy used to sequence 15-9-A. The arrows represent the direction and extent of sequencing from the positions indicated by solid dots. The position 0 indicates the start of the cDNA insert and 653 is the last base of the cDNA insert before the (A) homopolymeric region. The dashed line represents the sequence of the (GC) homopolymer tails, the (A) homopolymer tail, and/or the pBR322 vector.

B) Nucleotide sequence and predicted amino acid sequence of 15-9-A.

The nucleotides are numbered on the left hand side. Nucleotide 1 is the first base of the cDNA clone. The dot at the beginning of the sequences indicates that the sequence is truncated. The amino acids encoded by the open reading frame are indicated below the nucleotide sequence. The asterisks indicate the position of the termination codon. The "underline" indicates the position of the putative polyadenylation signal.

A



B

```
1     .CC AGT AGA ATG AAT AGT CCA TTT GGG TTG TCT TTT TGC CTT TTT GAT TCT ATG GAG AAG
          S   R   M   N   S   P   F   G   L   S   F   C   L   F   D   S   M   E   K

60    TGC TTG GCA GAC CAT AAA GCC TCA CTT AAA GAT CCC CAA GAT AAC AAC AAC CTA GCT TCA
       C   L   A   D   H   K   A   S   L   K   D   P   Q   D   N   N   N   L   A   S

120   ATG TCG TCT CTT CCT GGC TCA ATC CAA AAT CAG CCA CTC CTC ATT GAG ACT GTA AAA TTC
       M   S   S   L   P   G   S   I   Q   N   Q   P   L   L   I   E   T   V   K   F

180   AGA GCC GTC TTG AAA ACC TGT TCC CAT GTC AGT GCA CGA TAT TGT TTC ACT AAT CCT AAC
       R   A   V   L   K   T   C   S   H   V   S   A   R   Y   C   F   T   N   P   N

240   GTT GCT ACA TCG GCT TTA GCG GAT TGT CTC ATG CCA TCT CTC ACT CAT TGT GTT TAT CCT
       V   A   T   S   A   L   A   D   C   L   M   P   S   L   T   H   C   V   Y   P

300   TCT AGT AGC CCT ACG CCC CCG CCA CCA CCC CCG CCA CCC CCG CCA CCT GGT ATT TAG ATC
       S   S   S   P   T   P   P   P   P   P   P   P   P   P   P   P   G   I  ***

360   TTTGAAGTACTTGCTAATAGAAGAATAACTGGGTTTGTTGAATCAGCCTATGAACTAATCCTTTGCCTAATCAAGGGAC

439   AAGTCTTTATCTCTCTTGAAACATCCCTAGCGATATGTCAATAATACATATGATTGACTTGCAAGCTTGTTTTATGGTG

518   AAGGATTTATTATACCTTCAATATTCCAATTAGCTACATCTGGTTTGTGTAAGAAATCTCTCGTCATGAATCATGATCG

597   ATAATGACTTTCCTACTAAATAGTAATTAATAAAACTTGCTTTACATCTTCTTCTTT
```

previously described sequence of nodulin-20 (Sandal et al.,1987), another "Nodulin-A" family member, but significant differences reported in this thesis suggest that it represents a new "Nodulin-A" family member (see Section 3.5).

3.3.5 Sequence of 47-11-E

The Pst I fragment of the clone 47-11-E was partially sequenced (Figure 3.8) and found to correspond exactly to the sequence of leghemoglobin c_1 (Hyldig-Nielsen et al.,1982). The Pst I site at the 3' end of the clone was generated by the cDNA cloning. The 3' end of the clone also contained a poly (A) stretch suggesting that the 47-11-E clone contained the full 3' non-coding sequence of the Lbc_1 gene; however this region of 47-11-E was not sequenced. The Pst I site at the 5' region of the cDNA clone was a site that is located within the leghemoglobin coding region. Thus, the fragment that had been used to screen the library was missing about 170 bases of the 5' end of the leghemoglobin mRNA sequence including the 5' non-coding region and some of the 5' coding region.

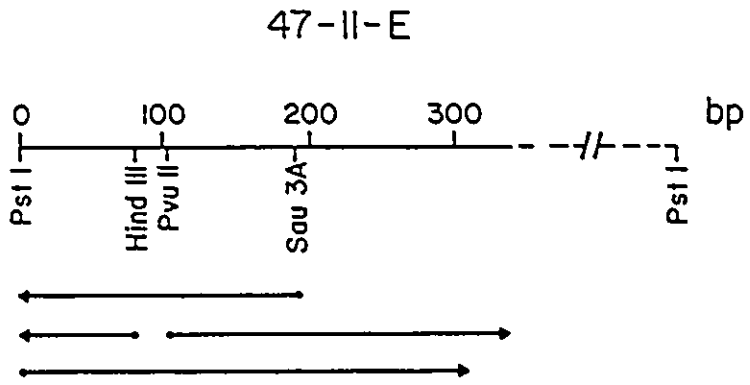
FIGURE 3.8 SEQUENCE OF THE cDNA INSERT FROM CLONE 47-11-E

A) Restriction map and strategy used to sequence 47-11-E. The arrows represent the direction and extent of sequencing from the positions indicated by the solid dots. The position 0 indicates a Pst I site internal to the cDNA sequence. The dashed line indicates the region of 47-11-E that was not sequenced including the 3' untranslated region, a homopolymeric (A) region generated from the poly (A) tail, and the (GC) homopolymeric tail.

B) Nucleotide sequence and predicted amino acid sequence of 47-11-E.

The nucleotides are numbered on the left hand side. Nucleotide 1 is the first base of a Pst I site located within the 47-11-E cDNA insert. The amino acids encoded by the open reading frame are located below the nucleotide sequence. Asterisks mark the position of the termination codon.

A



B

```
1     .CT GCA GCA AAG GAC TTG TTC TCA TTT CTA GCA AAT GGA GTA GAC CCC ACT AAT CCT AAG
          A   A   K   D   L   F   S   F   L   A   N   G   V   D   P   T   N   P   K

60    CTC ACG GGC CAT GCT GAA AAG CTT TTT GCA TTG GTG CGT GAC TCA GCT GGT CAA CTT AAA
       L   T   G   H   A   E   K   L   F   A   L   V   R   D   S   A   G   Q   L   K

120   ACA AAT GGA ACA GTG GTG GCT GAT GCT GCA CTT GTT TCT ATC CAT GCC CAA AAA GCA GTC
       T   N   G   T   V   V   A   D   A   A   L   V   S   I   H   A   Q   K   A   V

180   ACT GAT CCT CAG TTC GTG GTG GTT AAA GAA GCA CTG CTG AAA ACA ATA AAG GAA GCT GTT
       T   D   P   N   F   V   V   V   K   E   A   L   L   K   T   I   K   E   A   V

240   GGC GGC AAT TGG AGT GAC GAA TTG AGC AGT GCT TGG GAA GTA GCC TAT GAT GAA TTG GCA
       G   G   N   W   S   D   E   L   S   S   A   W   E   V   A   Y   D   E   L   A

300   GCA GCA ATT AAA AAG GCA TAATTAGGATCTA
       A   A   I   K   K   A  ***
```

3.4 Developmental Expression of Nodule-Specific Clones.

In order to investigate the expression of nodulin genes, 3 day old soybean seedlings were infected with Bradyrhizobium japonicum strain 61A76 and harvested at various times following infection. Under the growth conditions we have used, the first swellings were barely visible at the 5th day following infection. They were detectable as small brown bumps by 7 days and were 0.5 - 1.0 mm in size and pink in colour (due to the presence of leghemoglobin) by the 10th day following infection (Figure 3.9). Nitrogenase activity was first detected (using the acetylene reduction assay, Hardy et al.,1968) at 10 days following infection (Figure 3.10).

Northern blot analysis had clearly demonstrated that the isolated nodulin cDNA clones detected RNA in the mature nodule that was not detectable in the root tissue (Figure 3.3). To determine the time of appearance and the relative levels of nodulin mRNAs during nodule development, total RNA, isolated from uninfected roots and nodule tissue of various ages, was blotted onto BIOTRANS membranes using a slot blotter apparatus. These blots were hybridized with radiolabeled probes specific to the cDNA clones 47-11-E (Figure 3.11), 15-9-A, 36-1-A, and 6-9-F (see Materials and Methods). Sense strand RNA corresponding to the various cDNA clones was synthesized using an in vitro RNA

FIGURE 3.9 THE DEVELOPMENT OF SOYBEAN ROOT NODULES

Photographs of soybean roots were taken 4 days (A), 7 days (B), 10 days (C), 13 days (D), 16 days (E), 24 days (F), and 34 days (G) following inoculation of 2 day old soybean seedlings with Bradyrhizobium japonicum strain 61A76. The scale is indicated to the left of the photograph.

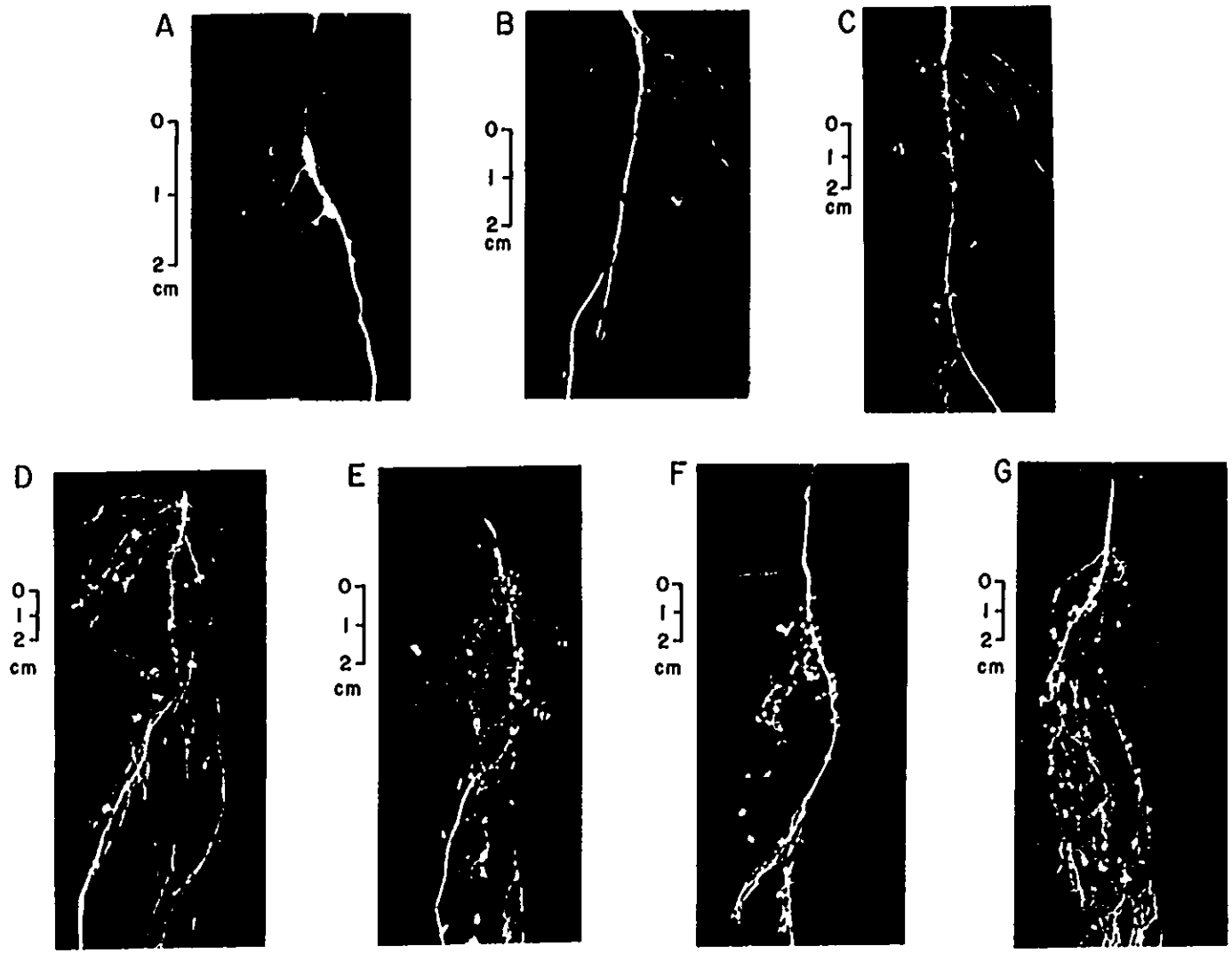

A
0
1
2
cm
B
0
1
2
cm
C
0
1
2
cm
D
0
1
2
cm
E
0
1
2
cm
F
0
1
2
cm
G
0
1
2
cm

FIGURE 3.10 NITROGEN FIXATION ACTIVITY AT VARIOUS DAYS AFTER INOCULATION

The root systems of soybean plants (Glycine max (L.) Merr. cv. Maple Arrow) inoculated with Bradyrhizobium japonicum strain 61A76, were assayed for acetylene reduction activity at the time points indicated. Three sets of soybean root systems at each time point were assayed for acetylene reduction activity expressed as nmoles of acetylene reduced/min/gram of dry weight of nodule. The error bars represent the standard deviation.

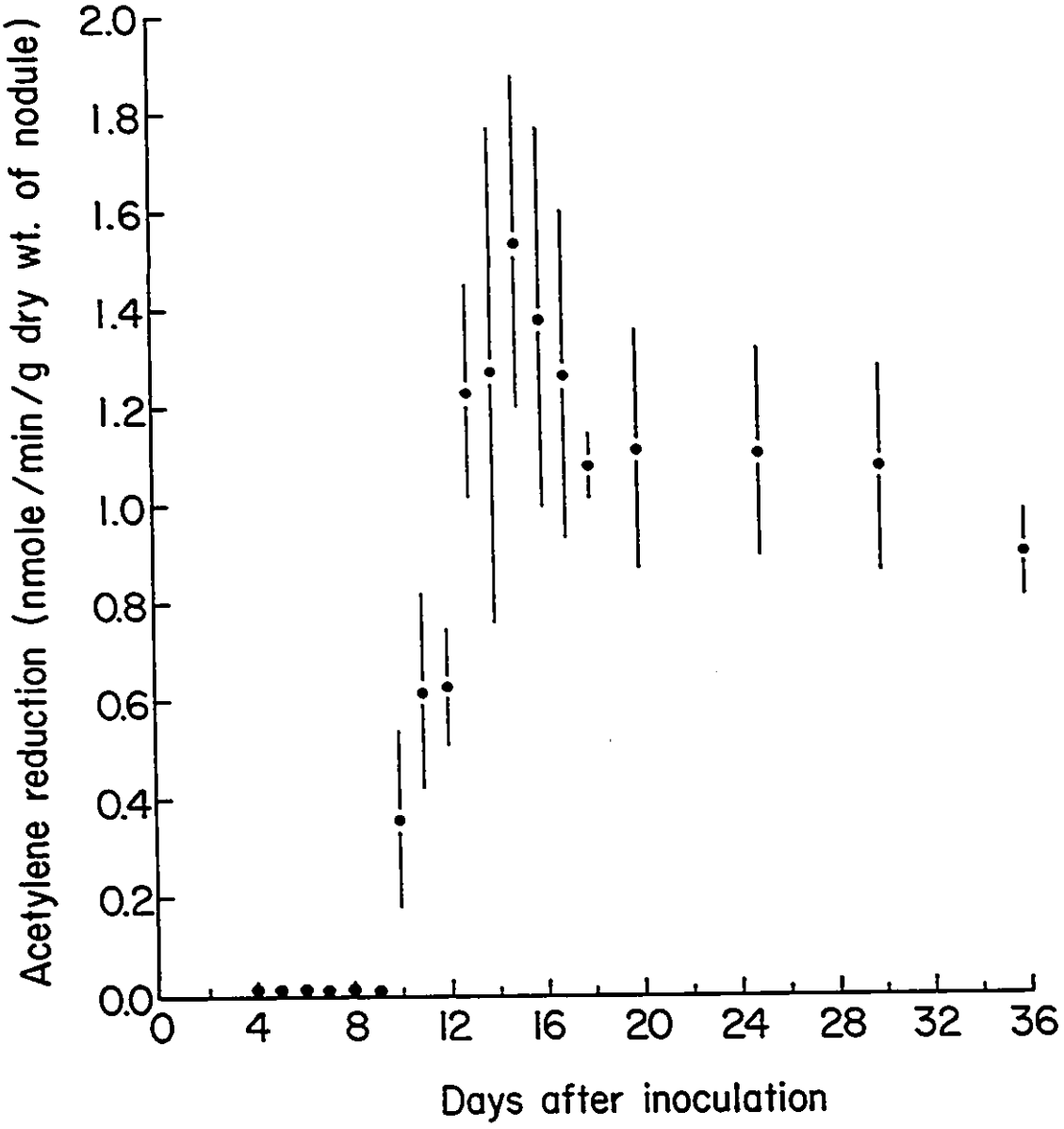

Acetylene reduction (nmole/min/g dry wt. of nodule)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
4
8
12
16
20
24
28
32
36
Days after inoculation

FIGURE 3.11 DEVELOPMENTAL ACCUMULATION OF LB SEQUENCES AS DETERMINED BY RNA SLOT BLOT ANALYSIS

Two-fold dilutions of total RNA (5 μg, 2.5 μg, 1.25 μg, 0.63 μg) and 5-fold dilutions of RNA standards were applied to BIOTRANS membranes using a Schleicher and Schuell Slot Blotter apparatus; each set of dilutions was applied in duplicate. The blot was hybridized to 5 x 10^5 cpm/ml of a labeled Pst I - Sau 3A fragment from the coding region of Lb cDNA, 47-11-E (Figure 3.8). The filter was exposed for 2 hours at -70°C with an intensifying screen.
Nodulated tissues were collected (as indicated in Materials and Methods) at the various times indicated.

Lane A,B, Rows 1-4: 2-fold serial dilutions of 7 day RNA
Lane A,B, Rows 5-8: 2-fold serial dilutions of 10 day RNA
Lane A,B, Rows 9-12: 2-fold serial dilutions of 13 day RNA
Lane A,B, Rows 13-16: 2-fold serial dilutions of 16 day RNA
Lane A,B, Rows 17-20: 2-fold serial solutions of 24 day RNA
Lane A,B, rows 21-24: 2-fold serial dilution of 34 day RNA
Lane C, Rows 1-7 and 8-14: 5-fold serial dilutions of a Lb RNA standard generated by in vitro RNA transcription. Rows 1, 8 contain 50 ng of RNA standard. Rows 7, 14 contain 3 pg of RNA standard.
Lane C, Rows 15, 16: 5 μg of tRNA
Lane C, Rows 17, 18: blank
Lane C, Rows 17, 19: 5 μg of root RNA
Lane C, Rows 18, 20: 2.5 μg of root RNA
Lane C, Rows 21, 23: 5 μg of 4 day RNA
Lane C, Rows 22, 24: 2.5 μg of 4 day RNA

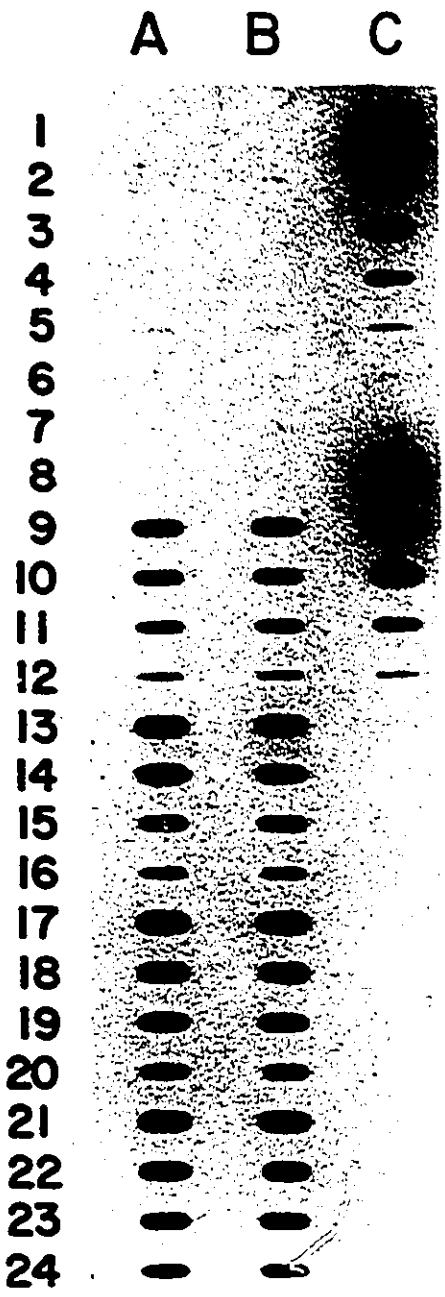

A
B
C
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

FIGURE 3.12 RELATIVE ACCUMULATION OF NODULIN RNAs DURING SOYBEAN ROOT NODULE DEVELOPMENT

The graph summarizes the results of RNA slot blot experiments. Total RNA was extracted from nodules at 13 days, 16 days, 24 days and 34 days following infection. The RNA for the 10 day time period was isolated from 5 cm sections of nodulated root. Isolated RNAs were slot-blotted and hybridized at high-stringency to random-primed cDNA inserts (as indicated in Materials and Methods) or for the case of 15-9-A, a specific oligonucleotide (probe D of Figure 3.14). The amount of RNA for each nodulin was quantified using standard curves (see Materials and Methods) for each time point by liquid scintillation counting and/or densitometric scanning of the autoradiogram (various exposures). The relative amounts were plotted as a fraction of the maximum amount of leghemoglobin RNA detected at day 24.

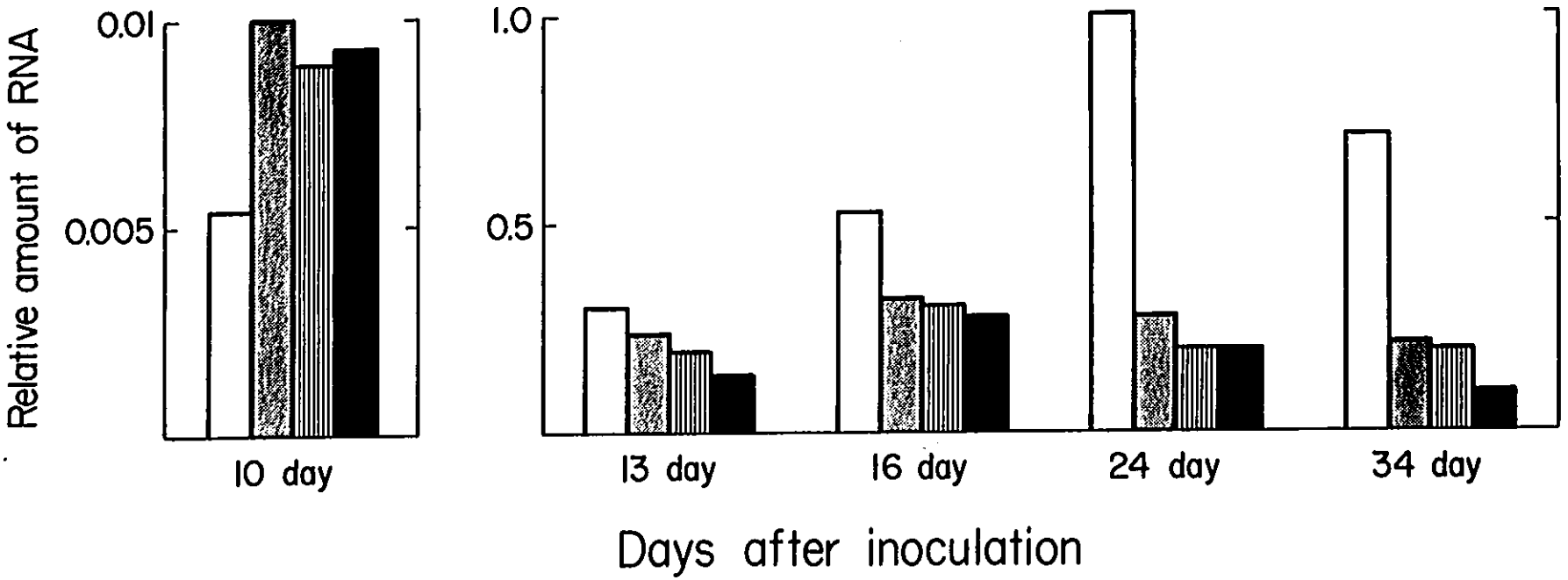
The various nodulin clones are indicated as follows:

open bar - Lb

stippled bar - 15-9-A

vertical lines - 36-1-A

solid bar - 6-9-F


Relative amount of RNA
0.01
0.005
1.0
0.5
10 day
13 day
16 day
24 day
34 day
Days after inoculation

transcription system (see Materials and Methods). Dilutions of these RNAs were included on the blots as standards to determine the amounts of a given nodulin RNA at a particular time in nodule development. The amounts of the nodulin RNAs at various times have been plotted relative to the maximum amount of Lb detected at 24 days following infection (Figure 3.12). The level of 9-11-B mRNA could not be determined. Repeated attempts to generate 9-11-B RNA as a standard were unsuccessful. The RNA generated by in vitro transcription each time migrated as a smear on a formaldehyde gel instead of as a distinct band (see Materials and Methods). Thus, it was impossible to determine with any accuracy, the amount of full length 9-11-B transcript generated in the in vitro system.

No nodulin RNA could be detected above background in either the 4 day or 7 day tissue. Background was determined to be the level of hybridization to 5 μg of total root RNA as determined by scintillation counting and/or densitometry. All of the nodulins gave a signal with root RNA following several days of exposure at -70°C; but this was presumed to be non-specific, since no hybridization to root RNA had ever been detected by northern blotting.

Nodule-specific mRNA for all four clones was first detected at 10 days (Figures 3.11, 3.12) following infection, which is concomitant with the onset of nitrogen-fixation activity. Several other groups have investigated

the expression of leghemoglobin and members of the "Nodulin-A" family during nodule development (Fuller and Verma, 1984; Marcker et al.,1984; Sengupta-Gopalan et al.,1986; Jorgensen et al.,1988). In these studies, nodulin RNA could be detected 2 - 3 days prior to the onset of nitrogen fixation. However, we did not detect nodulin mRNA above background levels prior to 10 days following infection. This may have been due to the limit of detection of our experiment, since total RNA was analyzed from root tissue containing the developing nodules (see Materials and Methods). Any nodulin mRNA would have been diluted by the presence of a large amount of root RNA. This is evident from the result obtained with RNA isolated from nodulated root tissue 10 days following infection. Although we could detect a signal at 10 days significantly above background (Figure 3.11), the signal was relatively low, largely due to the excess of root RNA in the isolated RNA. Total RNA was not isolated from 8 or 9 day nodulated tissue, but based on the results of others (Fuller and Verma, 1984; Marcker et al.,1984; Sengupta-Gopalan et al.,1986), it is highly likely that these RNAs were expressed at least 7 - 9 days following infection.

Changes in the relative amounts of the 15-9-A, 36-1-A and 6-9-F mRNAs compared to Lb mRNA in the maturing nodule, indicated that the "Nodulin-A" mRNAs and Lb mRNA accumulate in a different manner. At day 10, the relative

level of leghemoglobin mRNA was significantly less than the steady-state levels of the other three nodulins (Figure 3.12). It appears that the "Nodulin-A" RNAs accumulate to their maximum levels (16 days following infection) earlier than leghemoglobin (24 days following infection) (Figure 3.12). The fact that 15-9-A, 36-1-A and 6-9-F are detected at a greater level relative to leghemoglobin early in nodule development and their accumulation as a group appears to be similar, suggests that they may be regulated in a different manner than leghemoglobin.

The decline in overall nodulin RNA levels at 34 days may reflect the onset of nodule senescence and/or changes in the relative amounts of poly $(A)^+$ RNA within the total nodule RNA. The onset of nodule senescence is often associated with the pod-filling stage of reproductive growth (see Sutton,1983). Under our conditions of growth, the first flowers were detected at about 27 days following infection, and the first pods appeared at about 42 days following infection.

3.5 15-9-A and Nodulin-20 are Distinctly Different Sequences.

Sequence comparison of 15-9-A to other nodulin sequences demonstrated that 15-9-A was similar to the sequence of nodulin-20 (Sandal et al.,1987), a member of the "Nodulin-A" family (Figure 3.13). The two sequences are identical for most of the coding region and then suddenly diverge just prior to the 3' end of the coding region and the 3' untranslated region. The sequence divergence correlates with the position of the intron in the nodulin-20 gene (designated by the arrow in Figure 3.13). The sequence of nodulin-20 5' to the location of the intron in nodulin-20, is 100% identical to 15-9-A and the sequence 3' to the nodulin-20 intron is about 89% similar. The divergence of the two sequences therefore seems to be limited to the second exon. (Since 15-9-A and nodulin-20 are so similar, it is assumed that if they represent different genes, the introns would be located at a similar position in each gene). These results raise the possibility that 15-9-A and nodulin-20 represent recently duplicated genes in which the first exons are being homogenized by a mechanism such as gene conversion (See Discussion). The possibility that

15-9-A and nodulin-20 are two distinct genes is also supported by the fact that multiple hybridizing bands are observed in both Southern and northern blot analyses (see Table II) which suggests that 15-9-A is a member of a small

FIGURE 3.13 SEQUENCE COMPARISON OF THE cDNA 15-9-A VERSUS NODULIN-20

A) Comparison of the cDNA 15-9-A sequence with the nodulin-20 sequence of Sandal et al. (1987).

The nucleotides are numbered on the left hand side. The nodulin-20 cDNA sequence used for comparison was generated by removing the intron from the nodulin-20 gene sequence of Sandal et al. (1987) using the MicroGenie software programs. The vertical arrow represents the position of the intron within the nodulin-20 gene. The asterisks designate the positions of mismatches as determined by the MicroGenie Make Search Program. The underline beneath TAG represents the position of the termination codon. The underline beneath AATAAA represents the position of the putative polyadenylation signal. The nucleotides are numbered on the left hand side.

B) Amino acid sequence of the carboxyterminus of 15-9-A and nodulin-20.

Amino acids encoded by the divergent regions of 15-9-A and nodulin 20, following the position of the intron are presented below the nucleotide sequence. The sequences presented correspond to nucleotides 306 - 356 of the 15-9-A sequence presented in (A) and nucleotides 538 - 582 of the nodulin-20 sequence in (A). TAG indicates the termination codon in each sequence.

A

15-9-A
Nodulin-20 cDNA

```
1      CCAGTAGAATGAATAGTCCATTTGGGTTGTCTTTTTGCCTTTTTGATTCTATGGAGAAGT
233    CCAGTAGAATGAATAGTCCATTTGGGTTGTCTTTTTGCCTTTTTGATTCTATGGAGAAGT

61     GCTTGGCAGACCATAAAGCCTCACTTAAAGATCCCCAAGATAACAACAACCTAGCTTCAA
293    GCTTGGCAGACCATAAAGCCTCACTTAAAGATCCCCAAGATAACAACAACCTAGCTTCAA

121    TGTCGTCTCTTCCTGGCTCAATCCAAAATCAGCCACTCCTCATTGAGACTGTAAAATTCA
353    TGTCGTCTCTTCCTGGCTCAATCCAAAATCAGCCACTCCTCATTGAGACTGTAAAATTCA

181    GAGCCGTCTTGAAAACCTGTTCCCATGTCAGTGCACGATATTGTTTCACTAATCCTAACG
413    GAGCCGTCTTGAAAACCTGTTCCCATGTCAGTGCACGATATTGTTTCACTAATCCTAACG

241    TTGCTACATCGGCTTTAGCGGATTGTCTCATGCCATCTCTCACTCATTGTGTTTATCCTT
473    TTGCTACATCGGCTTTAGCGGATTGTCTCATGCCATCTCTCACTCATTGTGTTTATCCTT

301    CTAGTAGCCCTACGCCCCCGCCACCACCCCCGCCACCCCCGCCACCTGGTATTTAGATCT
533    CTAGTT CTAT CTTATTA CCACCCCCGCCACCACCCCCGCCAC T  TATTTAGATCT
          ** ** * *******     *  *  *              * **

361    TTGAAGTACTTGCTA ATAGAAGAATAACTGGGTTTGTTGAATCAGCCTATGAACTAATC
587    TTGAAGTACTTGCTATAGAAAAATATAACTGGGTTTGCTCAATCAGCCTATGTAATAATC
                  * * *  **            * *             * *

420    CTTTGCCTAATCAAGGGACAAGTCTTTATCTCTCTTGAAACATCCCTAGCGATATGTCAA
647    CTTTGCCTAGTCAAGGGACAAGTCTTAATCTATCTTGAAACATCCATAGTTATATGTCAA
          *                 *    *              *   **

480    TAATACATATGATTGACTTGCAAGCTTGTTTTATGGTGAAGGATTTATTATACCTTCAAT
707    TAATACATATGATTGACTTGCAAGCTTGTTTTATGATGAAGGATTTATTATACCTTCAAT
                                        *

540    ATTCCAATTAGCTACATCTGGTTTGTGTAAGAAATCTCTCGTCATGAATCATGATCGATA
767    ATTCCAATTAGCTACATCTGGTTTGTGTAAGAAATCACTTGTCATGAATCATGATCGTTA
                                       *  *                 *

600    ATGACTTTCCTACT AA A  TAGTAATTAATAAAACTTGCTTTACATCTTCTTCTTT
827    ATGGCTTTCCTACTTAATAGCTAGTAATTAATAAAACATGTTTTACATCTATTTTTTT
          *          *  * **                 *  *          **  *
```

B

15-9-A

```
AGC CCT ACG CCC CCG CCA CCA CCC CCG CCA CCC CCG CCA CCT GGT ATT TAG
Ser Pro Thr Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Gly Ile
```

Nodulin-20

```
TCT ATC TTA TTA CCA CCC CCG CCA CCA CCC CCG CCA CTT ATT TAG
Ser Ile Leu Leu Pro Pro Pro Pro Pro Pro Pro Pro Leu Ile
```

FIGURE 3.14 SCHEMATIC REPRESENTATION OF 15-9-A cDNA VERSUS NODULIN-20

A, B, C, D indicate the positions of oligonucleotide probes used for Southern and northern blot experiments. The vertical arrow represents the position of the intron in the nodulin-20 gene.

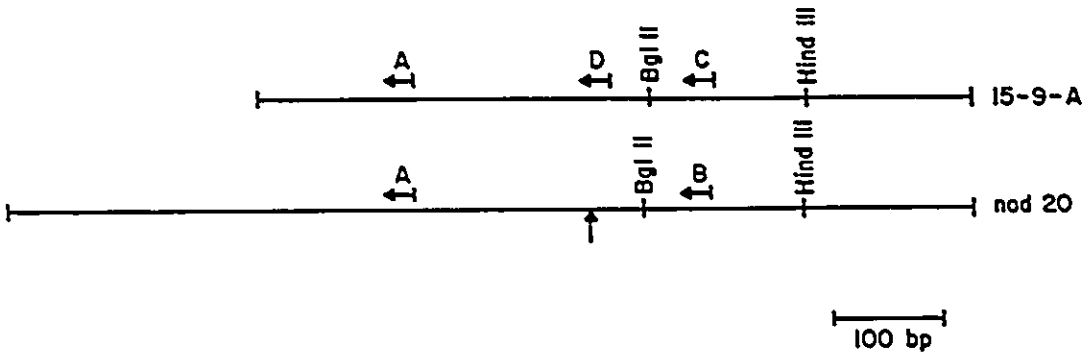

A
D
Bgl II
C
Hind III
15-9-A
A
Bgl II
B
Hind III
nod 20
100 bp

FIGURE 3.15 DETECTION OF 15-9-A mRNA USING AN OLIGONUCLEOTIDE PROBE

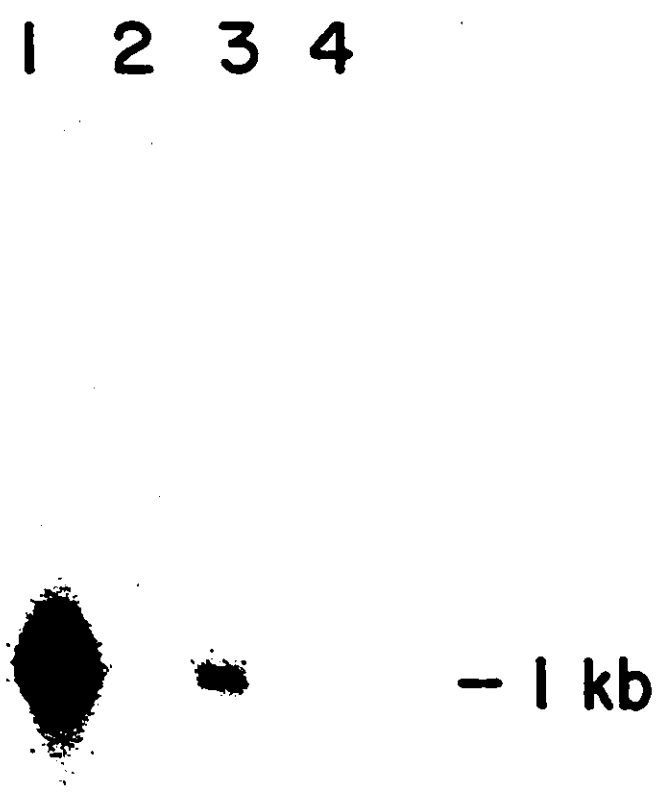
RNA from 3 day soybean roots or 3 - 4 week soybean nodules was electrophoresed on a 1.2% agarose gel containing formaldehyde and transferred to a BIOTRANS nylon membrane. The blot was hybridized to 5 x 10^6 cpm/ml of the 15-9-A sequence specific oligonucleotide D (Figure 3.14).

Lane 1: 1.5 μg of nodule poly $(A)^+$ RNA

Lane 2: 1.5 μg of root poly $(A)^+$ RNA

Lane 3: 10 μg of total nodule RNA

Lane 4: 10 μg of total root RNA


1 2 3 4
– 1 kb

multi-gene family. Although the region of greatest divergence occurs directly after the position of the intron, each sequence codes for a proline-rich carboxy terminus and uses the termination codon TAG (Figure 3.13B).

To establish that 15-9-A was in fact a unique sequence, and not a cloning artifact generated from nodulin-20 during the construction of the cDNA library, a sequence specific oligonucleotide (oligo D in Figure 3.14) was used as a probe against nodule RNA. The antisense oligonucleotide D, corresponding to the position 294-318 of the 15-9-A sequence (the region that is most divergent between nodulin-20 and 15-9-A), hybridized specifically to a 1.0 kb mRNA (Figure 3.15). Therefore, 15-9-A did appear to represent a new and unique member of the "Nodulin-A" family.

3.6 15-9-A and Nodulin-20 are Members of a Subfamily Within the Larger " Nodulin-A" Gene Family.

Hybridization of the 15-9-A cDNA clone to RNA blots, using stringent hybridization conditions, detected bands of about 0.8 and 1.0 kb in size. Sandal et al. (1987) have determined that the nodulin-20 sequence represents a 1.0 kb message. This work has determined that 15-9-A also represents a 1.0 kb message. Therefore, the 0.8 kb message represents a third message closely related to nodulin-20 and 15-9-A. There are a number of possibilities that can explain these results:

1) All three of the mRNAs may be transcribed from only one (or two) gene(s), but the different mRNAs arise from the use of different transcription initiation and termination sites, or alternative splice sites;

2) Each mRNA is transcribed from a unique gene.

To investigate which of these possibilities might explain the observed results, sequence-specific antisense oligonucleotide probes (25 mers) were used as probes against soybean genomic DNA and soybean nodule RNA.

3.6.1 Hybridization of specific oligonucleotide probes to soybean genomic DNA.

Specific oligonucleotides (A,B & C in Figure 3.14) were synthesized that would hybridize to the common 5' region of 15-9-A and nodulin-20 (probe A) or to the distinct 3' regions of the nodulin-20 (probe B) or 15-9-A (probe C) sequences.

Probe A is a specific oligonucleotide which corresponds to nucleotides 114-138 of the 15-9-A sequence, a region where 15-9-A is identical to nodulin-20 (Figure 3.13, 3.14). Hybridization of this oligonucleotide to soybean genomic DNA detected three distinct fragments with each of the restriction enzymes used, except Bst I where it detected two fragments (see below) (Figure 3.16B).

FIGURE 3.16 HYBRIDIZATION OF OLIGONUCLEOTIDE AND cDNA PROBES TO SOYBEAN DNA BY DIRECT GEL HYBRIDIZATION

Each panel represents the results of direct gel hybridization to specific probes.

(A) Hybridization to 2 x 10^6 cpm/ml of random-primer labeled 15-9-A cDNA insert.

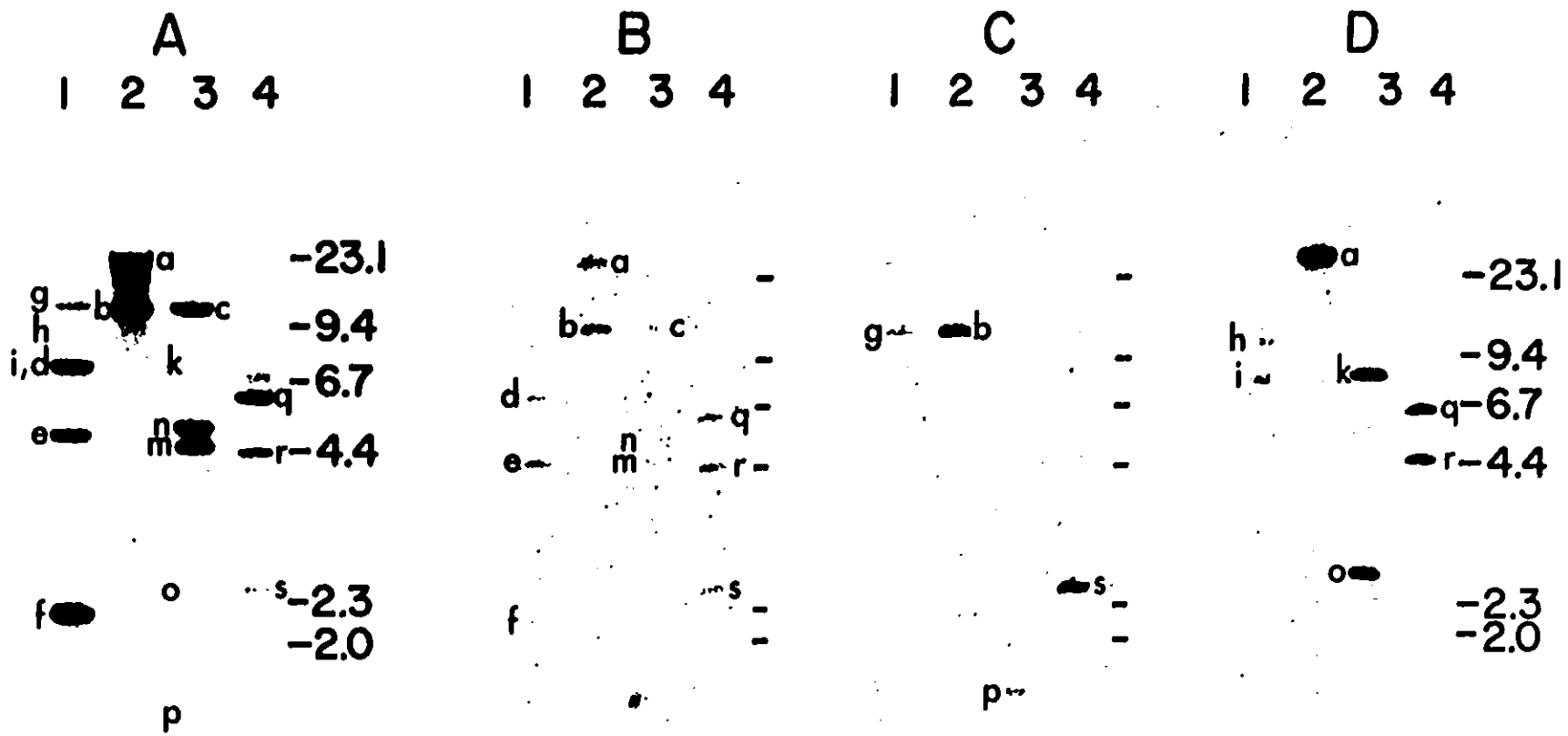
(B) Hybridization to 5 x 10^6 cpm/ml of probe A (Figure 3.14).

(C) Hybridization to 5 x 10^6 cpm/ml of the 15-9-A specific probe C (Figure 3.14).

(D) Hybridization to 5 x 10^6 cpm/ml of the nodulin-20 specific probe B (Figure 3.14).

10 μg of soybean DNA was digested with various enzymes. Restriction enzymes: Lane 1, Bgl II; Lane 2, Bst I; Lane 3, Eco RI; Lane 4, Hind III. All restriction digests in Panels B, C, and D were electrophoresed on the same gel. The restriction digests in Panel A were electrophoresed on a separate gel. The sizes of the lambda Hind III markers are indicated to the right of Panel A and Panel D.

The letters indicate corresponding bands on the different panels.


A
1 2 3 4
a
g
b
c
h
i,d
k
q
e
n
m
r
f
o
s
p
-23.1
-9.4
-6.7
-4.4
-2.3
-2.0
B
1 2 3 4
a
b
c
d
q
e
n
m
r
f
s
C
1 2 3 4
g
b
s
p
D
1 2 3 4
a
h
i
k
q
r
o
-23.1
-9.4
-6.7
-4.4
-2.3
-2.0

The size of each of the hybridizing fragments detected by probe A (Figure 3.16B) corresponded to the sizes of fragments detected when the full 15-9-A cDNA insert was used as a probe (Figure 3.16A). These results therefore strongly suggest the presence of three distinct genes.

The 15-9-A 3' specific probe C corresponds to positions 390-414 of the 15-9-A sequence. This oligonucleotide sequence differs from the nodulin-20 sequence by four nucleotides. When probe C was hybridized to soybean genomic DNA, it hybridized to only one fragment with each of the four restriction enzymes used (Figure 3.16C). Probe C hybridized to a 14.5 kb Bst I fragment and a 2.5 kb Hind III fragment. (Figure 3.16C - Lanes 2, 4). The 5' probe A and the full length cDNA were found to hybridize to Hind III and Bst I fragments of similar size. Since there are no Bst I sites contained within the 15-9-A cDNA sequence, and the Hind III site within the 15-9-A sequence is 3' to both regions defined by probe A and C (Figure 3.14), the results suggest that a soybean gene containing a 5' common region and 15-9-A - specific 3' region are contained on a 2.5 kb Hind III fragment and a 14.5 kb Bst I fragment.

Evidence for the presence of an intron in 15-9-A comes from the results of hybridization to Eco RI digested soybean DNA. The size of the 1.6 kb Eco RI fragment hybridizing to probe C (Figure 3.16C - Lane 3), is different

from any of the three Eco RI fragments (11.5 kb, 4.7kb, 4.4 kb) hybridizing to probe A (Figure 3.16B - Lane 3). Since there are no Eco RI restriction sites within the 15-9-A cDNA sequence, the difference in the sizes of the hybridizing fragments can be explained by the presence of an intervening sequence with an Eco RI site somewhere between the positions of the two oligonucleotides. Because of the striking similarity of the 15-9-A and nodulin-20 sequences, the intron of 15-9-A most likely occurs at the same relative position as in the nodulin-20 gene. The difference in the size of the Bgl II fragment detected by probe C (Figure 3.16C, Lane 1) relative to any of the sizes of the Bgl II fragments detected by probe A (Figure 3.16B, Lane 1), reflects the fact the Bgl II cleaves the 15-9-A sequence between the positions of the two oligonucleotides (see Figure 3.14).

The oligonucleotide probe B is a 25 mer that corresponds to position 618 - 641 of nodulin-20. This oligonucleotide differs by 4 nucleotides from the 3' specific probe of 15-9-A. Probe B hybridized to two distinct fragments with each of the enzymes used, except Bst I, where it hybridized to one very large band.

The fact that two genomic fragments were detected with each enzyme used indicated that there were two regions in the soybean genome that were identical to the nodulin-20 3' specific oligonucleotide.

Two hybridizing Hind III fragments of 6.5 kb and 4.3 kb detected with the probe B (Figure 3.16D, Lane 4) could also be detected using the 5' common probe A (Figure 3.16B, Lane 4). Hind III cleaves nodulin-20 at a position 3' to the positions of the two oligonucleotides and there is no Hind III site within the nodulin-20 intron (Sandal et al.,1987). Therefore, the fact that both 3' and 5' oligonucleotide probes detect Hind III fragments of similar size, suggests that the nodulin-20 gene is contained on either a 6.5 kb or 4.3 kb Hind III fragment. The fact that two hybridizing fragments are detected with the enzymes, Bgl II, Eco RI, and Hind III suggests that, in addition to nodulin-20, there is a second nodulin-20 "like" gene in the soybean genome.

Both probe B (Figure 3.16D, Lane 2) and probe A (Figure 3.16B, Lane 2) hybridize to a large Bst I fragment. This raises the possibility that the two nodulin-20 like sequences are linked on one large Bst I fragment. Alternatively, they may also be present on two large comigrating Bst I fragments, as resolution in this region of the gel is poor.

It should be noted that hybridization of a full length cloned insert of 15-9-A to restricted soybean DNA accounts for every fragment hybridizing to the oligonucleotide probes (Figure 3.16).

All of these results taken together suggest that the soybean genome contains 3 closely related nodulin-20 "like" sequences -- a subfamily within the "Nodulin-A" multi-gene family. It is possible that two of these sequences may represent different alleles of the same gene. However, commercial Glycine max (L.) Merr. cultivars are highly inbred. The parents (cv. Harosoy 63 and cv. 840-7-3) of the Maple Arrow cultivar used in this study, had virtually the same genomic hybridization patterns as Maple Arrow when the 15-9-A cDNA was used as a probe (data not shown). Another Glycine max cultivar, Evans, and the wild soybean, Glycine soja, also had very similar hybridization patterns. Therefore, although the possibility that two of these sequences represent different alleles cannot be entirely ruled out, it seems unlikely.

One property of the 15-9-A full length cDNA probe was that the region 3' to the termination codon gave a much weaker hybridization signal on Southern blots than hybridization with the 5' region. In fact, the fragments hybridizing to the 3' region had not been detected on the original Eco RI digested genomic DNA blots that identified 15-9-A as a unique nodule-specific clone (Section 3.2). It was found that a stringent wash in 0.1xSSC, 0.1% SDS at 65°C could remove virtually all of the hybridization signal arising from the 3' region. Even when the wash was not as stringent (0.1 xSSC, 0.1%SDS, at room temperature), the

hybridization signal was always significantly less than the signal obtained by the 5' region of the probe (Figure 3.16A). A similar phenomenon was observed in northern blot hybridization experiments when the 3' region was used as a probe (See Section 3.6.2). The reason for these observations is probably due to the difference in G + C content between the sequences that are 5' to the intron (42%) and the sequences that are 3' to the termination codon (32%). The lower G + C content would result in not only decreased stability of the hybrid (Marmur and Doty,1962); but also a decreased specific activity of the probe in this region, since [α - ^{32}P] dCTP was used to radioactively label the DNA.

3.6.2 Northern blot analysis of 15-9-A & Nodulin-20

Hybridization of the 15-9-A cDNA clone to nodule RNA revealed the presence of a 1.0 kb and a 0.8 mRNA. (Figure 3.3). Sandal et al. (1987) observed a similar hybridization pattern when nodulin-20 was used as a probe. Further hybridization studies have indicated that the 0.8 kb mRNA hybridized strongly to the 5' region of the 15-9-A cDNA even under stringent hybridization conditions (not shown), but did not hybridize to the 3' region of the 15-9-A cDNA. Only a 1.0 kb mRNA could be detected when the 3' region of the cDNA was used as a probe (Figure 3.17A). Sandal et al.

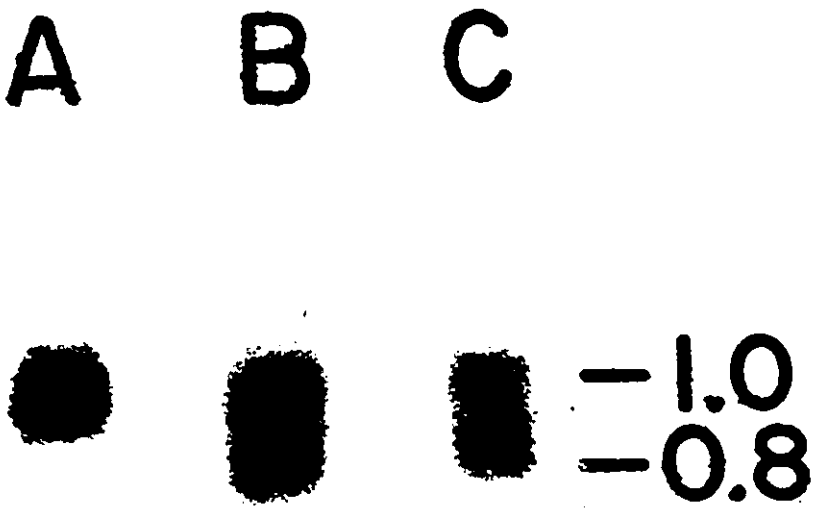
FIGURE 3.17 NORTHERN BLOT HYBRIDIZATION ANALYSIS OF THE NODULIN-20 SUBFAMILY

10 μg of total nodule RNA per lane was electrophoresed on a 1.5% agarose gel containing formaldehyde and transferred to a BIOTRANS filter.

(A) Hybridization to a 2 x 10^6 cpm/ml random-primer labeled fragment containing the 3' UTR region of 15-9-A. Hybridization was using standard hybridization conditions.

(B) Hybridization to 5 x 10^6 cpm/ml of end-labeled oligonucleotide probe A.

(C) Hybridization of 5 x 10^6 cpm/ml of end-labeled oligonucleotide probe B.


A
B
C
1.0
0.8

(1987) had made similar observations when they used the nodulin-20 5' and 3' sequences as probes against nodule RNA. These results together suggest that part of the sequence of the 0.8 kb mRNA is very similar to 15-9-A and nodulin-20, but that the 3' regions of 15-9-A and nodulin-20 are significantly different from the 0.8 kb mRNA sequence. These results raise the question as to the origin of the 0.8 kb mRNA.

A number of explanations could account for the presence of the 0.8 kb mRNA:

1) It may arise as the result of differential use of transcription and termination sites of either the 15-9-A and/or nodulin-20 gene;

2) Alternative splicing results in the addition of an entirely different 3' region to the 5' sequences of nodulin-20 or 15-9-A. Alternative splicing must be considered especially in light of the fact that sequence analysis suggests that the 5' exons of 15-9-A and nodulin-20 are identical and the 3' exons are divergent. It is possible for instance, that one 5' exon is spliced onto three different 3' exons; thus all three of the detected mRNAs could have arisen from one gene;

3) It is the product of a unique gene.

If the 0.8 kb message is the product of a unique gene, it is likely that it arises from one of the two hybridizing genomic sequences detected by probe B. (Figure

3.16). Since the 0.8 kb message appeared to contain sequences that were very homologous to the 5' region of 15-9-A and nodulin-20, and since the 5' probe A hybridized to three genomic fragments, two of which code for 15-9-A and nodulin 20, the 5' oligonucleotide probe A was used as a probe against nodule RNA. As expected, probe A was found to cross-hybridize to both a 1 kb and a 0.8 kb mRNA (Figure 3.17, Lane B). If the 0.8 kb mRNA is encoded by the third gene, it should cross-hybridize to the oligonucleotide probe B which detected two hybridizing fragments in the soybean genomic DNA. Therefore, probe B was hybridized to soybean nodule RNA. Probe B was also able to detect the presence of both a 1.0 kb and 0.8 kb mRNA (Figure 3.17, Lane C). These results strongly suggest that the 0.8 kb message arose from the third hybridizing gene detected by soybean genomic DNA hybridization analysis. Because the nodulin-20 specific oligonucleotide detected the 0.8 kb mRNA, this mRNA must share some homology to the nodulin-20 3' region. It is possible that the 0.8 kb message has sequences similar to the 3' region of 15-9-A and nodulin-20 that are not detected under standard hybridization conditions because of the reduced hybrid stability of the 3' region (discussed earlier). An attempt to reduce the stringency of hybridization in order to determine if this could be the case did not clarify the situation since strong hybridization to the 1.0 kb message obscured the resolution

of the 1.0 kb and 0.8 kb message on northern blots (data not shown). However, when the 3' untranslated region of the 15-9-A cDNA was used as a probe against genomic DNA, it was able to detect bands of the same size as those detected by the 3' specific oligonucleotide probes B and C (data not shown). These results suggested that the gene corresponding to the 0.8 kb mRNA, contained sequences similar to the 3' regions of nodulin-20 and 15-9-A.

3.7 Nodulin-A Family Homologues in Plants Other than Soybean.

The isolation of cDNA clones from soybean resulted in the detection of a multi-gene family that is specifically expressed during nodulation (Sengupta-Gopalan et al.,1986; Sandal et al.,1987; Jacobs et al.,1987; this work). Even though several genes and cDNAs corresponding to this family have been characterized, virtually nothing is known about their function. Sandal et al. (1987) have suggested that these nodulins may have metal-binding domains that are required for their function. Although these nodulin genes are induced during soybean nodulation, it is likely that these genes arose from a common ancestral gene that may have had no role in the nodulation process. This raises the possibility that similar gene sequences may be present and in fact functional in other plants. This possibility is not unprecedented, since hemoglobin which was generally thought to be present only in the nodules of legumes, has been

detected in several nonleguminous plants (Appleby et al.,1984; Bogusz et al.,1988; Appleby et al.,1988).

Therefore, a "Nodulin-A" family member was used as a probe to investigate the presence of homologous sequences in non-leguminous plants. The clone 9-11-B was chosen since it was the only clone that contained both the 5' and the 3' conserved domains and in addition, the full insert could be easily excised from the pBR322 vector. Hybridization of 9-11-B to Eco RI digested DNA from several different plants under conditions of low stringency, revealed the presence of several cross-hybridizing fragments. Cross-hybridizing sequences were detected in alfalfa (a legume) and several non-legumes that form symbiotic associations with the soil bacterium Frankia (sea buckthorn, russian olive and alder). As well, cross-hybridizing sequences were detected in a dicot which is not nodulated (spinach) and monocot (corn) (Figure 3.18). These results raise the possibility that "Nodulin-A" like genes exist in plants other than soybean.

The 9-11-B probe was hybridized to the various plant DNAs in at least two different Southern blots to verify the results presented in Figure 3.18. 10 μg of plant DNA was used for each track and ethidium bromide staining of the gels confirmed that approximately the same amount of DNA had been loaded into each lane (data not shown). Although the intensity of the hybridization signal in Figure 3.18B is

FIGURE 3.18 DETECTION OF "NODULIN-A"-LIKE SEQUENCES IN PLANTS OTHER THAN SOYBEAN

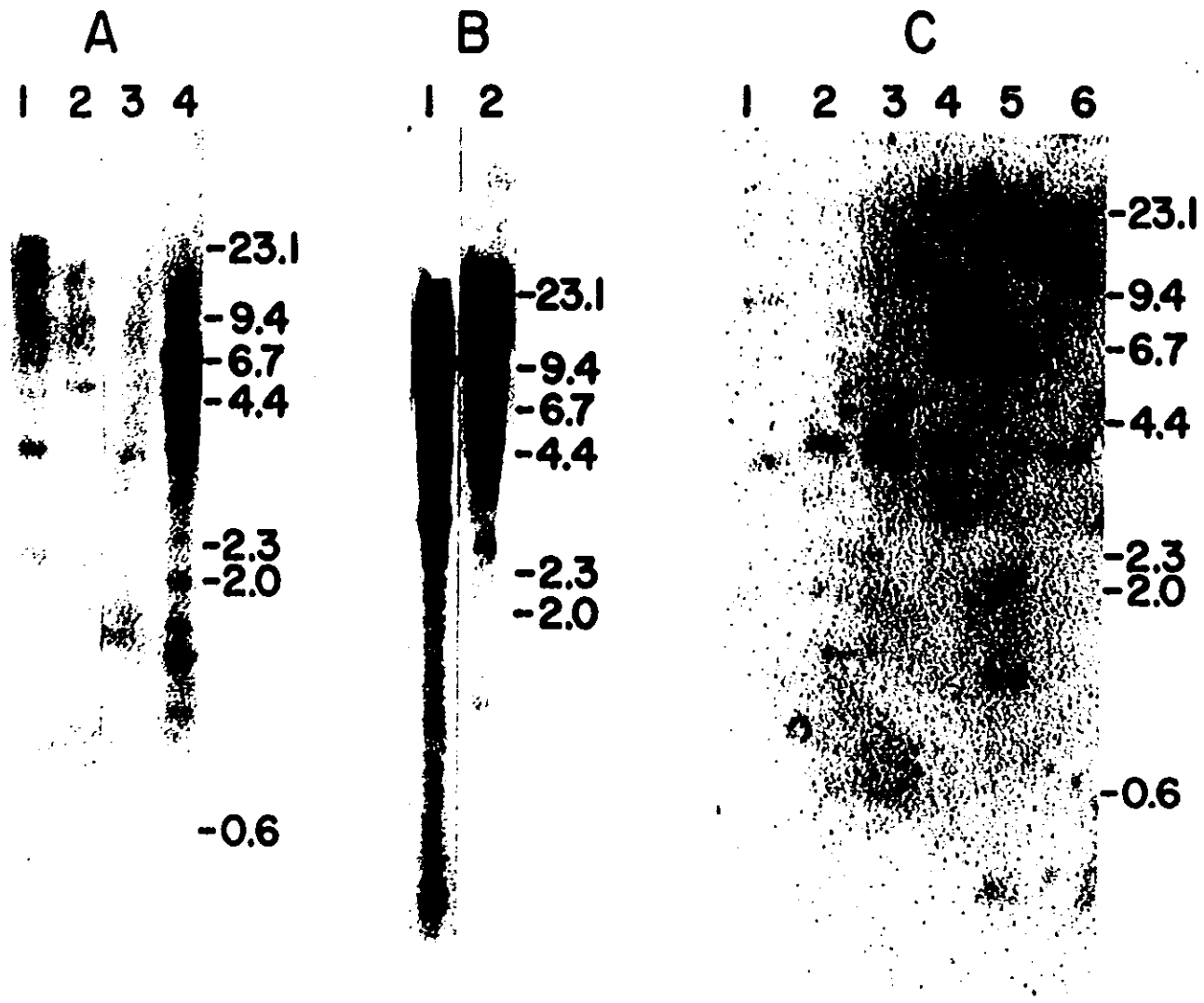
Plant DNAs were digested with restriction enzymes, electrophoresed on a 0.9% agarose gel and transferred to a BIOTRANS filter. The blots were hybridized to 2 x 10^6 cpm/ml of random-primer labeled 9-11-B insert under conditions of low stringency (33% formamide, 37°C). Washing was done in 2xSSC, 0.1% SDS at 37°C.

10 μg of each plant DNA in (A) & (B) were restricted with Eco RI.

A) Lane 1 - Hippophae rhamnoides (sea buckthorn);
Lane 2 - Eleagnus angustifolia (russian olive);
Lane 3 - Medicago sativa L. (alfalfa);
Lane 4 - Glycine max L. Merrill (soybean).

B) Lane 1 - Spinacia vulgaris (spinach)
Lane 2 - Zea mays (corn)

C) 3 μg of Alnus incana (alder) DNA was restricted with various enzymes: Lane 1 - Bam HI; Lane 2 - Bgl II; Lane 3 - Eco RI; Lane 4 - Hind III; Lane 5 - Msp I; Lane 6 - Pst I.


A
1 2 3 4
-23.1
-9.4
-6.7
-4.4
-2.3
-2.0
-0.6
B
1 2
-23.1
-9.4
-6.7
-4.4
-2.3
-2.0
C
1 2 3 4 5 6
-23.1
-9.4
-6.7
-4.4
-2.3
-2.0
-0.6

considerable, the hybridization signal obtained with spinach and corn DNA was considerably less than the hybridization signal obtained with the soybean DNA control that had been included on the same gel. In no case was hybridization to a heterologous plant stronger than the soybean control.

3.7.1 Screening of an alder genomic library.

The detection of homologous sequences in plants other than soybean is by no means conclusive evidence for the presence of "Nodulin A" gene family homologues. As an approach to investigating the presence of homologues in non-legumes, an alder genomic library (prepared by Madeleine Lévesque - unpublished) was screened.

A filter representing about 5×10^5 plaques was screened using the 9-11-B insert. The probe was then removed from the filter and the filter was rescreened. Only plaques that demonstrated a positive hybridization signal on both screenings were chosen for further investigation. Eighteen such alder clones were pooled into six groups of three plaques each, and subjected to further rounds of plaque purification (Figure 3.19). To make the analysis somewhat manageable, four different alder clones were chosen for further analysis. DNA was isolated from these clones and digested with various restriction enzymes. Southern hybridization analysis (Figure 3.20) indicated that these clones contained regions that specifically hybridize to the

9-11-B probe. Further analysis of these clones should provide some insight as to how these clones are related to the soybean 9-11-B probe; that is, whether the hybridization is due to a short region of high homology that is unrelated to the "Nodulin-A" family, or to a more extensive region of homology related to the "Nodulin-A" family. The clone alder - 6 is of particular interest since this clone hybridizes to two restriction enzyme fragments when different restriction enzymes are used (Figure 3.20, Lanes 6,11,15). The fact that two cross-hybridizing fragments are seen raises the possibility that 9-11-B is homologous to the alder - 6 clone over a more extended region, rather than just a short sequence.

The isolation of these alder clones supports the idea that alder may contain "Nodulin-A" like sequences; however at this point the results are not conclusive. Further analysis and sequencing of these clones will indicate whether any of these clones do in fact represent a "Nodulin-A" like sequence.

FIGURE 3.19 PURIFICATION OF ALDER LAMBDA CLONES CARRYING SEQUENCES HOMOLOGOUS TO 9-11-B, A "NODULIN-A" FAMILY MEMBER

Screening of a lambda genomic library derived from alder DNA was accomplished using 2 x 10^6 cpm/ml of a random-primer labeled cDNA insert of 9-11-B. Hybridization conditions were low stringency (33% formamide, 37°C) and wash conditions were 2xSSC, 0.1% SDS at 37°C.

A) First screening of an alder genomic library;

B) Purification of a plaque isolated from the first screening (A) (indicated by arrow);

C) Purification of a plaque isolated from the second screening (B) (indicated by arrow)

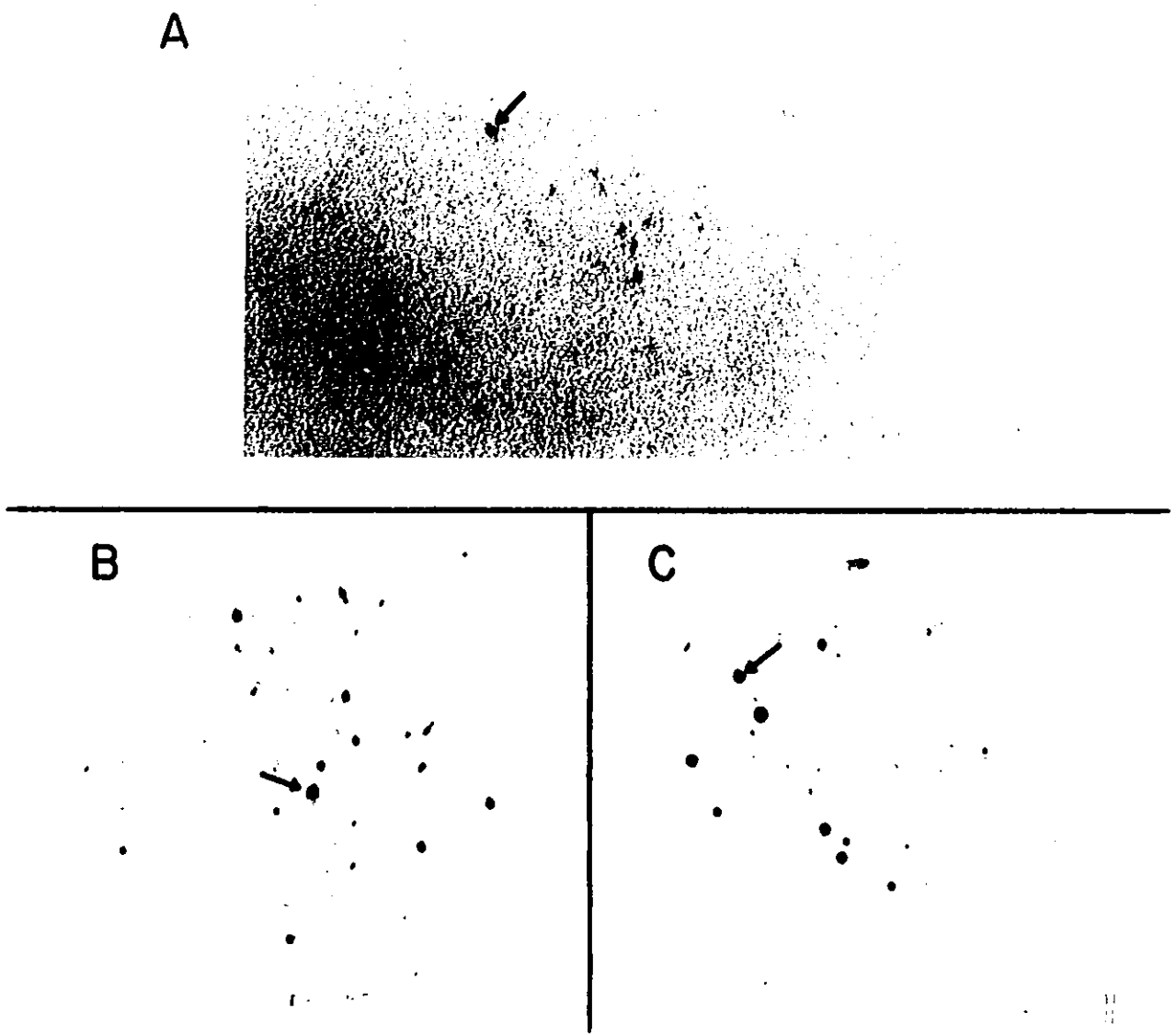

A
B
C

FIGURE 3.20 SOUTHERN BLOT HYBRIDIZATION OF DNA ISOLATED FROM ALDER CLONES HOMOLOGOUS TO THE 9-11-B cDNA INSERT

Lambda DNA was isolated from four different plaques that had hybridized to the 9-11-B insert during screening of the alder genomic library. The DNAs were restricted with various enzymes, separated in a 1% agarose gel and transferred to BIOTRANS. Blots were hybridized to 2 x 10^6 cpm/ml of random-primer labeled 9-11-B cDNA insert under conditions of low stringency (33% formamide, 37°C) and washed in 2xSSC, 0.1% SDS at 37°C.

Restriction enzymes: lanes 1 - 4, Bam HI; lanes 5 - 8, Bgl II; lanes 10 - 13, Pst I; lanes 14 - 17, Pvu II.

The clones are designated alder-3, alder-6, alder-8 and alder-10.

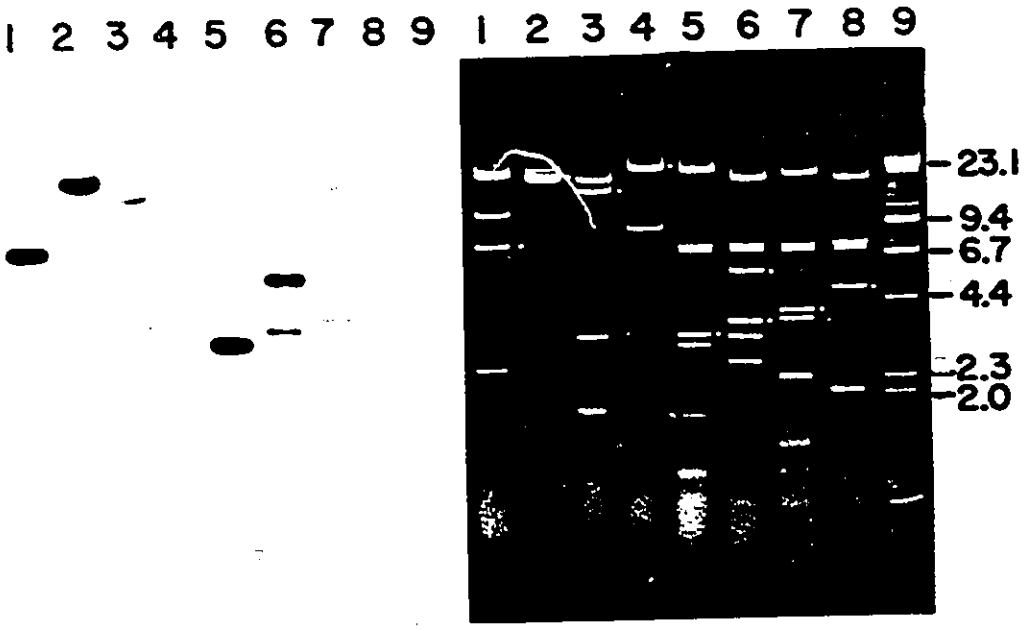
Alder-3: Lanes 1, 5, 10, 14;

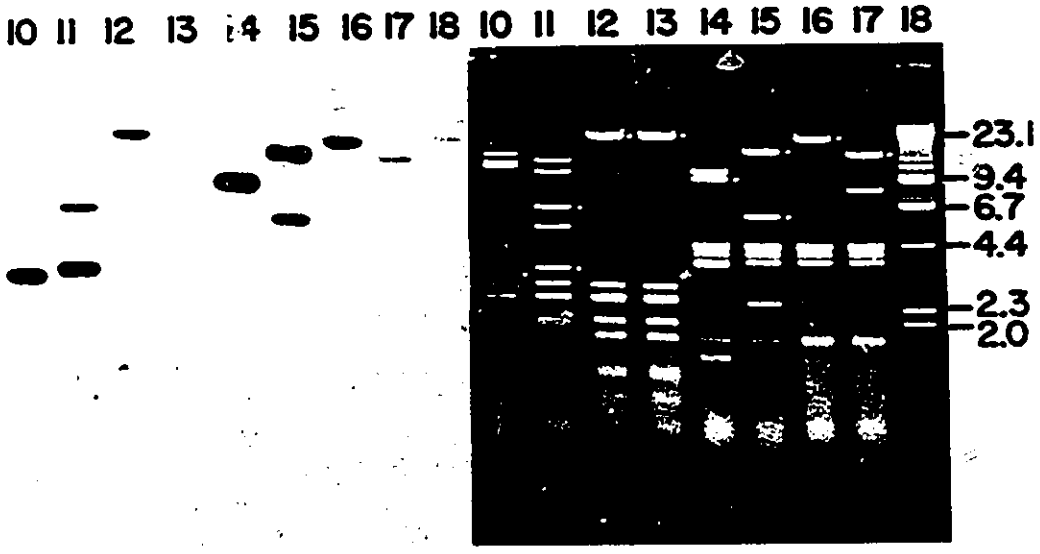
Alder-6: Lanes 2, 6, 11, 15;

Alder-8: Lanes 3, 7, 12, 16;

Alder-10: Lanes 4, 8, 13, 17;

Lanes 9 and 18 contain lambda Hind III size markers. The white dots marked on the gel photographs correspond to the bands on the Southern blots.


1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9
23.1
9.4
6.7
4.4
2.3
2.0


10 11 12 13 14 15 16 17 18
10 11 12 13 14 15 16 17 18
23.1
9.4
6.7
4.4
2.3
2.0

CHAPTER 4

DISCUSSION

4.1 Sequence Relationships of Nodulin cDNA clones.

The work in this thesis describes the isolation and characterization of five different soybean cDNA clones, each representing a gene that is specifically expressed in soybean root nodules. One of these clones (47-11-E) corresponds to the leghemoglobin c_1 gene (Hyldig-Neilson et al.,1982). The other four clones represent members of a small multi-gene family in soybean that has been designated as the "Nodulin-A" family (Verma and Delauney,1988). Studies by several different groups indicate that this gene family contains at least six members: nodulin-44 (nod E27) (Sengupta-Gopalan et al.,1986), nodulin-23 (Mauro et al.,1985; Sengupta-Gopalan et al.,1986), nodulin-26b (Jacobs et al.,1987), nodulin-27 (Jacobs et al.,1987), nodulin-20 (Sandal et al.,1987) and nodulin-22 (Sandal et al.,1987).

Sequence analysis of our four cDNA clones indicates that three of these cDNA clones, 9-11-B, 6-9-F, and 36-1-A, confirm previously published sequences of "Nodulin-A" family members. The fourth cDNA clone, 15-9-A, represents a new "Nodulin-A" family member.

The cDNA clone 9-11-B contains the complete coding region of nodulin-23, but is truncated in the 5' and 3'

untranslated regions. The sequence of 9-11-B is identical to the nodulin-23 (nod C51) sequence of Sengupta-Gopalan et al. (1986) except for one nucleotide difference; however, this difference is a third position substitution and does not cause any alteration of the amino acid sequence proposed by Sengupta-Gopalan et al. (1986). The one base difference likely only reflects the fact that two different soybean cultivars were used in these studies: Glycine max cv. Maple Arrow in our study and Glycine max cv. Prize in the studies by Sengupta-Gopalan et al. (1986). The sequence of the nodulin-23 gene (Mauro et al.,1985) is very similar to the 9-11-B sequence and the nod C51 sequence (Sengupta-Gopalan,1986); however, there are several significant differences. Deletions of a few nucleotides in the 5' coding region of nodulin-23 compared to the nod C51 sequence of Sengupta-Gopalan et al. (1986) shift the reading frame, causing the two sequences to differ for about fifteen amino acids. In addition, there are a few other single base differences distributed throughout the sequences. Apparently, the nodulin-23 sequence of Mauro et al.,(1985) has been revised, since Jacobs et al. (1987) state that the nodulin-23 sequence differs from the nod C51 sequence by only three nucleotides, although no sequence was presented. These remaining three differences again may only reflect cultivar differences.

The cDNA clone 6-9-F is a truncated form of the nodulin-44 sequence. It contains about 35% of the nodulin-44 coding region and most of the 3' untranslated region. The sequence of 6-9-F is identical to the nodulin-44 (nod E27) sequence of Sengupta-Gopalan et al. (1986) except for one nucleotide difference which does not alter the amino acid sequence. The difference again may just reflect the fact that the cDNA libraries in our study and that of Sengupta-Gopalan et al. (1986) are derived from two different soybean cultivars.

36-1-A is a truncated version of the nodulin-22 gene. The 490 bp sequence of 36-1-A is identical to 490 bp in the coding region of the nodulin-22 gene (Sandal et al.,1987). The sequence is also very similar to the sequence of the nodulin-27 cDNA (Jacobs et al.,1987, see Figure 3.6). Although nodulin-27 and nodulin-22 are treated as different gene products (Verma and Delauney,1988), our comparison of the published cDNA sequence of nodulin-27 to the nodulin-22 gene sequence (with the intron removed), indicates that these sequences are more than 98% similar (determined using the Micro Genie Make Search Program). Differences between the two sequences can be explained by soybean cultivar differences and/or possible sequencing errors. But if nodulin-22 and 27 in fact represent different genes, these genes have likely duplicated very

recently and/or sequence homogeneity is being maintained by a mechanism such as gene conversion.

Thus, our analyses of three of the nodule cDNA clones agree with the previously published analyses of other groups.

4.2 Developmental Expression of Nodule-specific genes.

It was very clear from northern blot analyses (Figure 3.3), that mRNAs corresponding to Lb and the four isolated cDNA clones were expressed in a nodule-specific manner.

To determine both the timing of expression of these cDNA clones and their relative amounts, total RNA was isolated from nodulated tissue at various days following infection and hybridized to radiolabeled probes specific to Lb, 15-9-A, 36-1-A and 6-9-F. We were first able to detect the presence of nodulin RNA corresponding to all four clones at 10 days following infection (Figure 3.11, 3.12), about the time we also first detected nitrogenase activity (Figure 3.10); but were unable to detect nodulin RNA at the earlier time point; 7 days after infection.

In soybean nodules, there are at least two distinct groups of nodulins: one group includes the early nodulins (e.g. nodulin-75) that do not require the presence of Rhizobium in the root tissue, and the second group includes late nodulins that require the presence of

Rhizobium within the infection thread and are probably involved with nodule function (see Verma and Delauney,1988). Studies by several groups have indicated that Lb and the "Nodulin-A" family members are late nodulins. They are expressed about 2 - 3 days prior to the onset of nitrogen fixation activity (Fuller and Verma,1984; Marcker et al.,1984; Sengupta-Gopalan et al.,1986; Jorgensen et al.,1988). These studies suggest that we should have been able to detect the RNA corresponding to our cDNA clones at about 7 - 9 days following infection.

The fact that we were unable to detect mRNA corresponding to our cDNAs at 7 days following infection is likely due to the fact that the nodule RNA from 7 day tissue was diluted with a large amount of root RNA; therefore, the presence of any nodulin mRNA may have been below our level of detection. The level of detection could be improved by isolating poly $(A)^+$ RNA from this time point. Also, RNA from 8 and 9 day nodulated roots was not isolated.

Our results indicate that the mRNAs corresponding to the "Nodulin-A" family members accumulate differently from leghemoglobin. Although Lb is the most abundant mRNA in the mature nodule (Auger et al.,1979; Fuller et al., 1983), at 10 days following infection, Lb mRNA was detectable at a significantly lower level than the levels of 36-1-A, 15-9-A or 6-9-F. This may reflect differences in the stability of Lb RNA compared to the "Nodulin-A" mRNAs or

a difference in the kinetics of induction of Lb genes and the "Nodulin-A" family members. Our results are consistent with those of Marcker et al. (1984), which indicate that Lb gene expression is regulated somewhat differently than other late nodulins. Marcker et al. (1984) have determined that leghemoglobin transcripts can be detected at a low level relatively early in nodule development. But, there is a dramatic increase in leghemoglobin transcription just prior to the onset of nitrogen fixation. This dramatic increase occurs at about the same time as the accumulation of nodule-specific mRNAs corresponding to "Nodulin-A" family members (Marcker et al.,1984; Fuller and Verma,1984; Sengupta-Gopalan et al.,1986). Thus, it appears that although Lb and the "Nodulin-A" family members are late nodulins, there are some differences in their induction kinetics (Marcker et al.,1984; Jorgensen et al.,1988).

The mRNA levels of 15-9-A, 36-1-A and 6-9-F accumulate in a similar manner and appear to reach their maximum levels earlier than leghemoglobin (Figure 3.12). It is difficult to quantitatively compare the steady-state amounts of RNA between the different time points because it is possible that the relative percentage of poly $(A)^+$ RNA in the total RNA is different in the nodules harvested at different times. However, the levels of RNA can be quantitatively compared within a given time point and it is very clear from our data, that leghemoglobin levels increase

relative to the other nodulin RNAs until 24 days following infection. The differences in accumulation of leghemoglobin mRNA compared to that of the other nodulins, may reflect either that the leghemoglobin mRNA is a much more stable mRNA than that of the other nodulins, and/or differences in the levels of transcription of the different genes. It is also clear from our studies that the "Nodulin-A" RNAs are abundantly transcribed. These results are consistent with the results of Fuller et al. (1983) who have estimated that 12 - 15% of the polysomal poly $(A)^+$ RNA from 21 day soybean nodules represents Lb mRNA and 6% of the mRNA represents two "Nodulin-A" sequences, nodulin-23 (6-9-F) and nodulin-44 (9-11-B).

4.3 15-9-A is a Member of a Subfamily Within the "Nodulin-A" Family.

Comparison of the 15-9-A sequence to the nodulin-20 sequence of Sandal et al. (1987) revealed that these two sequences were very similar (Figure 3.13). Conservation is 100% in the region 5' to the position of the intron in nodulin-20, and 89% in the region 3' to the position of the intron. Although the region of greatest divergence between the two sequences occurs directly after the position of the intron, each sequence codes for a proline-rich carboxy terminus and has a termination codon at approximately the same position. These results suggest that 15-9-A and

nodulin-20 may represent a subfamily within the larger "Nodulin-A" family. The fact that both 15-9-A and nodulin-20 have a proline rich carboxy terminus (15-9-A has 11 consecutive proline residues, nodulin 20 has 8 consecutive proline residues, see Figure 3.13) suggests that the proline rich carboxy terminus may have som functional significance; however, we have not been able to determine whether this is the case. Alternatively, the proline-rich termini may not have significant functional roles, but rather are only a consequence of the evolution of the sequences.

Hybridization of an oligonucleotide probe specific to 15-9-A (in the region that is most diverged between the two sequences) to soybean nodule RNA revealed that the 15-9-A represented an abundantly transcribed 1.0 kb mRNA. Northern blot hybridization analysis with the full 15-9-A cDNA insert had indicated the presence of two cross-hybridizing bands of 0.8 kb and 1.0 kb in size, even under stringent hybridization criteria. Since it has been demonstrated that both 15-9-A (this thesis) and nodulin-20 (Sandal et al.,1987) represent 1.0 kb mRNAs, the presence of the 0.8 kb mRNA could not be accounted for by either of these two sequences. The detection of the 0.8 kb message suggested the presence of a third closely related gene within the soybean genome.

4.3.1 Hybridization to soybean genomic DNA suggests the presence of three closely related nodulin-20 like genes.

A specific oligonucleotide that was derived from the common 5' region of nodulin-20 and 15-9-A (probe A) was found to hybridize to three different genomic fragments when different restriction enzymes were used. The sequence of the oligonucleotide used could not be detected in any of the other known nodulin sequences (using the MicroGenie Make Search Program). Since the sizes of each of the fragments hybridizing to the oligonucleotide corresponded to the size of strongly hybridizing fragments observed when the full 15-9-A cDNA insert was used as a probe, the results strongly suggest the presence of three different closely related genes.

An oligonucleotide specific for the 3' untranslated region of 15-9-A (probe C) hybridized to only one soybean genomic fragment with each of the enzymes used. The size of each of the hybridizing fragments were similar to sizes of fragments that could be detected when the full 15-9-A cDNA insert was used as a probe. These results, therefore, indicated that the 15-9-A 3' sequence is present at only one genomic location in the soybean. The fact that both the 5' and 3' oligonucleotides hybridized to the same size of fragment when either the enzyme Hind III or Bst I is used, suggests that the 5' and 3' regions detected by the oligonucleotides, are linked together and are separated by a

distance of no greater than 2.5 kb (size of the hybridizing Hind III fragment).

The fact that all of the Eco RI fragments that hybridize to the 5' oligonucleotide, are different from the 1.6 kb fragment that hybridizes to the 15-9-A 3' oligonucleotide (Figure 3.16), suggests that an intron is located somewhere between the positions of the two oligonucleotides, since 15-9-A does not contain an Eco RI site within the cDNA sequence. Since 15-9-A is very similar to nodulin-20, it is highly probable that the intron is located at approximately the same position as in the nodulin-20 gene -- near the 3' untranslated region. The genes of two other "Nodulin-A" family members have been isolated, nodulin-23 (Mauro et al.,1985) and nodulin-22 (Sandal et al.,1987). Like nodulin-20, these genes also contain an intron within the coding region near to the 3' untranslated region.

An oligonucleotide specific to the 3' region of the nodulin-20 gene (probe B) hybridized to two distinct soybean genomic fragments from each restriction digest (except Bst I). Fragments of similar size could be detected using the full length cDNA clone as a probe. These results suggest that there are at least two regions in the soybean genome that contain the sequence of the nodulin-20 3' specific oligonucleotide. The fact that fragments of the same size could be detected using either the full length

cDNA clone, or the 3' UTR region of 15-9-A as probes, suggests that the similarity between these two regions extends beyond the 25 bp sequence of the oligonucleotide. This, therefore, suggests the existence of another 3' exon closely related to the 3' exon of nodulin-20 and 15-9-A. It was expected that the 3' exons would be at least somewhat similar, since all of the 3' untranslated regions of the "Nodulin-A" gene family members share regions of similarity (see Figure 1.1; Sengupta-Gopalan et al.,1986; Jacobs et al.,1987; Sandal et al.,1987).

The sizes of the two Hind III fragments hybridizing to the nodulin-20 specific oligonucleotide were similar in size to two of the three Hind III fragments detected using the 5' oligonucleotide as a probe (Figure 3.16). As well, both oligonucleotide probes (probe A and probe B) hybridized to a large molecular weight Bst I fragment. In addition, we know that the nodulin-20 specific 3' oligonucleotide and the 5' oligonucleotide are linked together within at least one gene, the nodulin-20 gene, because the gene has been isolated by Sandal et al. (1987). These results, taken together, suggest that in addition to the nodulin-20 gene, the soybean genome contains a second closely related gene which contains the 5' oligonucleotide sequence and the 3' nodulin-20 specific oligonucleotide sequence.

The fact that both the 5' oligonucleotide and the nodulin-20 oligonucleotide hybridized to only one large Bst I fragment, instead of two distinct fragments, raises the possibility that the nodulin-20 gene and the nodulin-20 "like" sequence are linked on one Bst I fragment. However, it is also possible that these fragments are located on two different large co-migrating Bst I fragments. To distinguish between these possibilities, it will be necessary to isolate the genes and their flanking sequences to try to find regions of overlap. It would be interesting to determine if 15-9-A could also be linked to these other two sequences.

The results presented here suggest the existence of a subfamily of three genes within the larger "Nodulin-A" family. It is possible that two of the sequences represent alleles of the same gene. However, Glycine max cultivars are highly inbred. Also, similar Southern hybridization patterns were observed in several other soybean cultivars, including the parents of Maple Arrow. In addition, a similar pattern was observed in Glycine soja, a wild soybean relative of Glycine max. Therefore, it seems unlikely that two of these sequences could represent alleles of the same gene.

4.3.2 Three nodulin-20 like genes code for three different mRNAs.

Hybridization of the full length 15-9-A cDNA insert and specific oligonucleotide probes to soybean nodule RNA demonstrated the presence of at least three closely related mRNAs -- two that are about 1.0 kb in size corresponding to nodulin-20 (Sandal et al.,1987) and 15-9-A (this work), and a third mRNA that is about 0.8 kb in size. The 0.8 kb mRNA contains sequences that are very similar to the 5' cDNA sequence of 15-9-A (or nodulin-20). The fact that the 5' specific oligonucleotide hybridizes to the two RNA species reinforces this point. The nodulin-20 3' specific oligonucleotide was also able to detect a 1.0 kb and 0.8 kb mRNA. The fact that both the 5' common and 3' nodulin-20 specific oligonucleotides were able to detect the two messages suggest that the 0.8 kb mRNA arises from either nodulin-20 or a nodulin-20 "like" gene. Northern hybridization of the 3' region of the nodulin-20 gene (Sandal et al.,1987) did not detect the 0.8 kb message suggesting that this message did not arise from the nodulin-20 gene. Therefore, the additional nodulin-20 "like" sequence that was detected by genomic DNA hybridization may be an active gene coding for the 0.8 kb mRNA. The degree of divergence of the 3' region of the 0.8 kb message from either nodulin-20 or 15-9-A is unclear. It is possible that in fact this region is quite similar; but because of the relative instability of the hybrids formed with the (AT)

rich 3' UTR region during standard hybridization experiments, the similarity was not detected. In fact, 15-9-A and nodulin-22 are 84% similar over a stretch of 300 bp in their 3' end but in our experiments, no message corresponding to the size of 36-1-A (nodulin-22) has been detected using standard hybridization conditions and the 15-9-A 3' UTR region as a probe.

Both the nodulin-20 and 15-9-A genes code for a 1.0 kb mRNA. However the nodulin-20 "like" gene appears to contain 5' and 3' sequences that are similar to 15-9-A and nodulin-20; yet it encodes an mRNA of only 0.8 kb in size. The fact that the 0.8 kb mRNA is smaller than the 1.0 kb mRNA may be due to a number of reasons:

1) The mRNA arises as a result of alternative splicing of the nodulin-20 or 15-9-A 5' exon to a different 3' exon. This possibility cannot be entirely ruled out but is unlikely, since we have demonstrated that the 0.8 kb mRNA likely arises from a unique gene;

2) The 0.8 kb mRNA arises from the third gene which contains a deletion(s) relative to the nodulin-20 or 15-9-A sequences;

3) The mRNA arises from the third gene but has different transcription initiation and/or termination signals or polyadenylation signals relative to the positions of those in the nodulin-20 or 15-9-A sequence. The possibility that alternative polyadenylation signals may be used is suggested

by the sequence of the 15-9-A and nodulin-20 3' regions. The 3' UTR sequences of 15-9-A and nodulin-20 contain sequences that are similar to 3' end processing signals found in animal cells. The positions of these sequences within the 3' UTR region raise the intriguing possibility that 3' end processing of the 0.8 kb mRNA may contribute to its decreased size in relation to the nodulin-20 and 15-9-A mRNAs.

Investigation of 3' end formation in animal cells has defined two cis-acting elements that are important for processing of the primary transcript: a highly conserved AAUAAA sequence located 10 - 30 bases upstream of the poly (A) addition site (Proudfoot and Brownlee,1976), and a less conserved (GU) rich region about 30 bases downstream of the polyadenylation signal (Gil and Proudfoot,1984; McLauchlan et al.,1985; McDevitt et al.,1986; Gil and Proudfoot,1987). However, virtually no information is available about 3' end processing in plants. Sequence analysis of a compilation of plant genes indicates that they have a conserved AATAAA-like sequence, although most plant genes do not contain a perfect AATAAA signal (Joshi,1987). Plant genes also contain a consensus sequence YGTGTTTT (similar to the mammalian YGTGTTYY consensus sequence (McLauchlan et al.,1985) downstream of the AATAAA motif. However, the position of this (GT) rich sequence did not appear at a

fixed position relative to the polyadenylation signal (Joshi et al.,1987).

Recent studies involving the deletion analysis of 3' regions of the potato proteinase inhibitor II gene (An et al.,1989) and the octopine synthase gene (Inglebrecht et al.,1989) have demonstrated that 3' regions downstream of the polyadenylation signal strongly influenced the level of mRNA expression in plants. In both of these studies, the 3' control region contains a YGTGTTTT like sequence, although it is not clear that this particular motif per se has a functional role. However, the analogy to the animal system suggests that this sequence motif may have an important role (Ingelbrecht et al.,1989; An et al.,1989).

The 3' sequence of 15-9-A contains a putative polyadenylation signal at position 625 (Figure 3.13). In addition, the sequence AATAAT can be found at position 478 of the 15-9-A sequence. This sequence differs by only one base from the AATAAA consensus sequence. A (T) rich motif, CTTGTTTT, is located at position 504, 26 bp downstream of the AATAAT sequence. This motif differs by only one base from the YGTGTTTT consensus sequence. Both of these motifs are also located in the 3' UTR of the nodulin-20 gene. It is intriguing to speculate that the 3' region of the gene encoding the 0.8 kb message may have diverged from the 15-9-A or nodulin-20 sequences sufficiently that the AATAAT sequence at position 478 of the 15-9-A sequence is

recognized as the major polyadenylation signal. Recognition of this site would reduce the mRNA size by about 150 nucleotides which could account in part or totally for the difference in the sizes of the 0.8 kb and 1.0 kb mRNAs.

It will be necessary to isolate the cDNA and the gene corresponding to the 0.8 kb mRNA to determine the exact nature of its divergence from 15-9-A and nodulin-20.

4.4 Evolution of the "Nodulin-A" Family.

Analysis of members of the "Nodulin-A" family has demonstrated that these family members share a conserved 5' domain and a conserved 3' domain separated by a diverged region of variable length (Figure 1.1, Sengupta-Gopalan et al.,1986; Jacobs et al.,1987; Sandal et al.,1987; Verma and Delauney,1988). DNA sequence comparisons between the various "Nodulin-A" family members indicate that each of the two domains share 65 - 95% homology among the various nodulins; however, numerous insertions and deletions cause transient shifts in the reading frames of the conserved domains (Jacobs et al.,1987; Sandal et al.,1987). Analysis by Sandal et al. (1987) has demonstrated that specific amino acid residues are conserved. Sandal et al.,(1987) have suggested that 4 cysteine residues in each of the conserved domains of nodulin-20, nodulin-22, nodulin-23 and nodulin-44 can be arranged into structures that are similar to metal binding domains, suggesting that the binding of metal ions

may be important in the functioning of these nodulins. However, our analyses indicated that the sequence of nodulin-26b (Jacobs et al., 1987), another "Nodulin-A" family member, is missing one cysteine residue in each of the two domains. Therefore, either all four cysteine residues are not necessary for the functioning of these proteins, or perhaps the nodulin-26b protein is evolving a functional role that is different from the other "Nodulin-A" family members. Even the subcellular location of these nodulins does not appear to be consistent, since nodulin-23 is located in the peribacteroid membrane and nodulin-27 is located in the cytosolic compartment (Jacobs et al.,1987).

Comparison of the DNA and amino acid sequences indicate that nodulin-44 and nodulin-23 are more closely related to each other than to nodulin-20 or 22. Similarly, nodulin-20 and nodulin-22 are more closely related to each other (Sandal et al.,1987). It is possible that these nodulins arose as a result of gene duplication and divergence where nodulin-23 and nodulin-44 have recently duplicated and nodulin-20 and nodulin-22 have also recently duplicated. Another possibility is based on the suggestion that soybean may be an allotetraploid (Hymowitz and Singh,1987), and therefore, some of the "Nodulin-A" members represent alloallelic pairs of homologous genes.

To distinguish between some of these possibilities, it will be necessary to determine the chromosomal organization of the various "Nodulin-A" family members.

The work presented in this thesis demonstrates the presence of two new "Nodulin-A" family members -- nodulin 15-9-A coding for a 1.0 kb mRNA, and a nodulin-20 "like" gene probably coding for a 0.8 kb mRNA. Sequence analysis and hybridization analysis have shown that these two "Nodulin-A" family members are very closely related to nodulin-20 and likely comprise a small subfamily within the larger "Nodulin-A" family. It is interesting to note that the 5' regions of 15-9-A and nodulin-20 are 100% identical for at least 303 bp (Figure 3.13) whereas the 3' regions are significantly diverged. The divergence between the two sequences correlates with the position of the intron in nodulin-20. These results suggest that the two exons may be evolving in a different manner. The 3' exons are gradually accumulating changes and diverging whereas the 5' exons are being homogenized. It is also possible that the nodulin-20 "like" gene coding for the 0.8kb mRNA exhibits a similar phenomenon, since by hybridization analysis, it is clear that the 5' regions are very similar whereas the 3' regions diverge.

Gene homogenization has been documented in many gene families. For example, possible unequal exchange and/or gene conversion events have been demonstrated in the vertebrate globin gene families (Slightom et al.,1980; Liebhaber et al.,1980; Jeffreys,1982). Of note is the report of Slightom et al. (1980), where extensive sequencing of human γ-globin genes has shown that the 5' two-thirds of the $^{A}\gamma$-globin gene is more like the $^{G}\gamma$ gene on its own chromosome than the allelic $^{A}\gamma$ gene on its homologue. Slightom et al. (1980) have proposed that a T-G rich region in the second intron might have "initiated" gene conversion events. Homogenization events also have been documented in plants. Analysis of the five genes of the small subunit of ribulose-1,5-bisphosphate carboxylase in tomato indicate that three of the genes are linked within a 10 kb region (Sugita et al.,1987). These three closely linked genes are virtually identical but this homology does not extend into 5' or 3' flanking sequences, suggesting that the sequence homogeneity of the genes is due to gene conversion events (Sugita et al.,1987).

It is very likely that events such as gene conversion or unequal crossing over are involved in maintaining the homogeneity of the nodulin-20 subfamily. Since the divergence of 15-9-A and nodulin-20 correlates with the position of the intron, it is intriguing to think that the event which caused the homogenization of the 5'

exons was initiated within the intron sequence. Since the 5' exon of the nodulin-20 "like" gene appears to be very similar to the 5' exons of 15-9-A and nodulin-20 (by hybridization), it is intriguing to speculate that the "intron" located within these genes contains a "hot spot" for recombination as was suggested for the γ-globin genes (Slightom et al.,1980). Clearly, the isolation of the 15-9-A and nodulin-20 "like" genes and their flanking sequences should provide some insight into what type of evolutionary events have occurred during the generation of this nodulin-20 subfamily. An attempt was made to isolate nodulin-20 subfamily genes from a soybean genomic library purchased from Clontech. The 5' region of the 15-9-A cDNA was used as a probe to isolate genomic clones from a soybean library. Of 15 clones that were found to hybridize to the 5' sequence of 15-9-A, none were found to hybridize to the 3' sequence. It is possible that none of the clones contained a full nodulin-20 subfamily gene, or that the 5' region was selecting other "Nodulin-A" family members. Isolation of the genes of the nodulin-20 subfamily will be necessary to determine their genomic organization and evolutionary relationships.

4.5 Detection of "Nodulin-A" Family Homologues in Other Plants.

The "Nodulin-A" multi-gene family represents several specific sequences that are expressed during soybean nodule development (Verma and Delauney,1988). Although the analysis of the nucleotide sequences of genes of unknown function often provide clues as to their functional roles, the role of the "Nodulin-A" family has remained elusive. It is possible that the "Nodulin-A" family may have arisen from an ancient membrane protein, since all of the family members appear to contain signal-like sequences (Jacobs et al.,1987; Sandal et al.,1987; Verma and Delauney,1988) of which only some appear to be functional (Jacobs et al.,1987). Also, their functional role may be dependent on a metal-binding activity (Sandal et al.,1987).

One an approach to investigating the possible functional role and evolution of the "Nodulin-A" family members, is to isolate functional homologues from non-leguminous plants.

Southern hybridization experiments, using the cDNA insert of 9-11-B as a probe, detected cross-hybridizing sequences in several different plants. These plants included alfalfa (a legume), several plants that form symbiotic relationships with the soil bacterium Frankia, (sea buckthorn, russian olive, alder), as well as a dicot that is never nodulated (spinach) and a monocot (corn). The fact that a cross-hybridizing sequence could be detected in

another legume, alfalfa, raises the possibility that a "Nodulin-A" like member might be involved in alfalfa nodulation. An attempt to determine if this was the case was unsuccessful. Northern blot hybridization of the four cDNA clones 9-11-B, 6-9-F, 36-1-A or 15-9-A did not detect any cross-hybridizing sequences in alfalfa nodule RNA using low stringency hybridization conditions (data not shown). However, it is possible that if the sequences are expressed in the alfalfa nodule, they have diverged from the soybean sequences to such an extent that they cannot be detected under these particular conditions of northern hybridization. Alternatively, if the hybridizing alfalfa sequence represents a functional gene, it may be expressed at a very low level in the nodule, or it may have a role that is entirely unrelated to nodulation.

It is entirely possible that a "Nodulin-A" like protein could have a functional role in plants other than soybean that is not related to symbiotic nitrogen fixation. For example, leghemoglobin, which was formerly thought to be located only in the nodules of legumes (Appleby,1974), has been detected in the nodules of several non-leguminous plants (Appleby et al.,1984; Tjepkema,1983; Kortt et al.,1988). In addition, hemoglobin protein has been detected in the roots of Trema, a member of the Ulmaceae which has never been shown to be nodulated (Bogusz et al.,1988; Appleby et al.,1988). This raises the possibility

that hemoglobin proteins may be involved in oxygen transport in some plants that are not involved in symbiotic nitrogen fixation.

The detection of sequences in other plants, that are homologous to "Nodulin-A" family members, does not provide evidence for a functional "Nodulin-A" like gene in these plants. It will be necessary to isolate and evaluate these sequences to determine their relatedness to the "Nodulin-A" family. As an approach to isolating a "Nodulin-A" family homologue, an alder genomic library was screened using the cDNA insert of 9-11-B. Four different genomic clones containing sequences that hybridize to 9-11-B have been isolated, and are currently being analyzed for their relatedness to the soybean "Nodulin-A" family. Further restriction, hybridization, and sequence analysis should determine how these clones are related to the 9-11-B probe. Of particular interest is the clone alder - 6, which hybridized to two restriction fragments when several different restriction enzymes were used (Figure 3.20). This suggests that the alder - 6 clone is homologous to 9-11-B over an extended region, rather than just a short region. Since the various alder clones appear quite different by restriction analysis and hybridization, it is possible that they could represent multiple genes or alleles within the alder genome. This possibility should be elucidated by further investigation of the alder clones. If any of these

clones do in fact appear to be a "Nodulin-A" like gene, it will be necessary to determine if it is expressed. Since alder forms a symbiotic association with Frankia, it would be particularly exciting if a "Nodulin-A" message could be detected in the RNA of alder root nodules. Also, detection and analysis of a non-soybean homologue may help provide some insight as to the functional role of "Nodulin-A" proteins.

REFERENCES

An, G., Mitra A., Choi, H.K., Costa, M., An, K., Thornburg, R.W., Ryan, C.A., (1989) Functional analysis of the 3' control region of the potato wound-inducible proteinase inhibitor II gene. The Plant Cell 1, 115 - 122.

Appleby, C.A. (1974) Leghemoglobin. in Quispel, A. (ed.) The Biology of Nitrogen Fixation. North Holland: Amsterdam, 521 - 554.

Appleby, C.A.,(1984) Leghemoglobin and Rhizobium respiration. Ann. Rev. Plant Physiol. 35, 443 - 478.

Appleby, C.A., Tjepkema, J.D., Trinick, M.J.,(1984) Hemoglobin in a non-leguminous plant, Parasponia. Possible genetic origin and function in nitrogen fixation. Science 220, 951 - 953.

Appleby, C.A., Bogusz, D. Dennis, E.S., Peacock, W.J., (1988) A role for haemoglobin in all plant roots? Plant Cell Environ. 11, 359 - 367.

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

Auger, S., 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.

Aviv, H., Leder, P., (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid cellulose. Proc. Natl. Acad. Sci. 69, 1408 - 1412.

Barker, D.G., Gallusci, P., Lullien, V., Kahn, H., Gherardi, M., Huguet, T.,(1988) Identification of two groups of leghemoglobin genes in alfalfa (Medicago sativa) and a study of their expression during root nodule development. Plant Mol. Biol. 11, 761 - 772.

Bennett, M.J., Lightfoot, D.A., Cullimore, J.V.,(1989) cDNA sequence and differential expression of the gene encoding the glutamine synthetase γ polypeptide of Phaseolus vulgaris L. Plant Mol. Biol. 12, 553 - 565.

Bergerson, F.J.,(1962) The effects of partial pressure of oxygen upon respiration and nitrogen fixation by soybean root nodules. J. Gen. Microbiol. 29, 113 - 125.

Bergerson, F.J., Turner, G.L., Appleby, C.A.,(1973) Studies on the physiological role of leghemoglobin in soybean root nodules. Biochim. Biophys. Acta 292, 271 - 282.

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

Birnboim, H.C., Doly, J., (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acid Res. 7, 1513 - 1520.

Bisseling, T., Been, C., Klugkist, J., Van Kammen, A., Nadler, K., (1983) Nodule-specific host proteins in effective and ineffective root nodules of Pisum sativum. EMBO J. 2, 961 - 996.

Bogusz, D., Appleby, C.A., Landsmann, J., Dennis, E.S., Trinick, M.J., Peacock, W.J., (1988). Functioning haemoglobin genes in non-nodulating plants. Nature 331, 178 - 180.

Bojsen, K., Abiladsten, D., Jensen, E.O., Paluden, K., Marcker, K.A., (1983) The chromosomal arrangement of six soybean leghemoglobin genes. The EMBO J. 2, 1165 - 1168.

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

Caldwell, B.E., Vest, H.G.,(1977). Genetic aspects of nodulation and dinitrogen fixation by legumes: the macrosymbiont In: Hardy, R.W.F., Silver, W.S., (eds.). A Treatise on Dinitrogen Fixation, Section 111. John Wiley & Sons: New York, 557 - 576.

Calvert, H.E., Pence, M.K., Pierce, M., Malik, N.S.A., Bover, W.D. (1984). Anatomical analysis of the development and distribution of Rhizobium infections in soybean roots. Can. J. Bot. 62, 2375 - 2384.

Campos, F., Padilla, J., Vazquez, M., Ortega, J.L., Enriquez, C. and Sanchez, F. (1987). Expression of nodule-specific genes in Phaseolus vulgaris L. Plant Mol. Biol. 9, 521 - 532.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J.,(1979) Isolation of biologically active ribonucleic acid from sources enriched in endonuclease. Biochem 18, 5294 - 5299.

Church, G., Gilbert, W. (1984) Genomic sequencing. Proc. Natl. Acad. Sci. 81, 1991 - 1995.

Cullimore, J.V., Gebhardt, C., Saarelainen, R., Miflin, B.J., Idler, K.B., Barker, R.F.,(1984) Glutamine synthetase of Phaseolus vulgaris L.: Organ specific expression of a multigene family. J. Mol. Appl. Genet. 2, 589 - 599.

Cullimore, J.V., Bennett, M.J.,(1988) The molecular biology and biochemistry of plant glutamine synthetase from root nodules of Phaseolus vulgaris L. and other legumes. J. Plant Physiol 132, 387 - 393.

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

Dazzo, F.B., Gardiol, A.E.,(1984) Host-specificity in Rhizobium-legume interactions in Verma, D.P.S., Hohn, T.H. (eds.) Genes Involved in Microbe-Plant Interactions. Springer-Verlag: New York, 3 - 31.

de Lajudie, P., Huguet, T.,(1988) Plant gene expression during effective and ineffective nodule development of the tropical stem-nodulated legume Sesbania rostrata. Plant Mol. Biol. 10, 537 - 548.

Diaz, C.L., Melchers, L.S., Hooykaas, P.J.J., Lugtenberg, B.J.J., Kijne, J.W. (1989) Root lectin as a determinant of host-plant specificity in the Rhizobium-legume symbiosis. Nature 338, 579 - 581.

Downie, J.A., Johnston, A.W.B. (1988) Nodulation of legumes by Rhizobium. Plant Cell Environ. 11, 403 - 412.

Dunn, K., Dickstein, R., Feinbaum, R., Burnett, B.K., Peterman, T.K., Thoidis, G., Goodman, H.M., Ausubel, F.M.,(1988) Developmental regulation of nodule-specific genes in alfalfa root nodules. Mol. Plant Microbe Interact., 2, 66 - 74.

Feinberg, A.P., Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6 - 13.

Fortin, M.G., Zelechowska, M., Verma, D.P.S.,(1985) Specific targeting of membrane nodulins to the bacteroid-enclosing compartment in soybean nodules. EMBO J. 4, 3041 - 3046.

Fortin, M.G., Morrison, N.A., Verma, D.P.S.,(1987) Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. Nucl. Acids Res. 15, 813 - 824.

Franssen, H.J., Nap, J.P., Gloudemans, T., Stiekema, W., Van Dam, H., Govers, F., Louwerse, J., Van Kammen, A., Bisseling, T., (1987) Characterization of cDNA for nodulin-75 of soybean: a gene product involved in early stages of root nodule development. Proc. Natl. Acad. Sci. U.S.A. 84, 4495 - 4499.

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

Fuller, F., Kunstner, P., Nguyen, T., Verma, D.P.S., (1983) Soybean nodulin genes: Analysis of cDNA clones reveals several major tissue-specific sequences in nitrogen-fixing root nodules. Proc. Natl. Acad. Sci. U.S.A. 80, 2594 - 2598.

Fuller, F., 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.

Gebhardt, C., Oliver, J.E., Forde, B.G.,Saarelainen, R., Miflin, B.J.,(1986) Primary structure and differential expression of glutamine synthetase genes in nodules, roots and leaves of Phaseolus vulgaris. EMBO J. 5, 1429 - 1435.

Gergen, J.P., Stern, R.H., Wensink, P.C.,(1979) Filter replicas and permanent collections of recombinant DNA plasmids. Nucl. Acids Res. 7, 2115 - 2136.

Gil, A., Proudfoot, N.J.,(1984) A sequence downstream of the AAUAAA is required for rabbit β-globin mRNA 3'-end formation. Nature 312, 473 - 474.

Gil, A., Proudfoot, N.J. (1987) Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β-globin mRNA 3' end formation. Cell 49, 399 - 406.

Gloudemans, T., de Vries, S., Bussink, H.J., Malik, N.S.A., Franssen, H.J., Louwerse, J., Bisseling, T., (1987) Nodulin gene expression during soybean (Glycine max) nodule development. Plant Mol. Biol. 8, 395 - 403.

Govers, F., Gloudemans, T., Moerman, M., van Kammen, A., Bisseling, T.,(1985) Expression of plant genes during the development of pea root nodules EMBO J. 4, 861 - 867.

Govers, F., Moerman, M., Downie, J.A., Hooykaas, P., Franssen, H.J., Louiwerse, J., van Kammen, A., Bisseling, T., (1986) Rhizobium nod genes are involved in inducing an early nodulin gene. Nature 323, 564 - 566.

Govers, F., Nap, J.P., Moerman, M. Franssen, H.J., Van Kammen, A., Bisseling, T.,(1987) cDNA cloning and developmental expression of pea nodulin genes. Plant Mol. Biol. 8, 425 - 435.

Govers, F., Nap, J.P., van Kammen, A., Bisseling, T., (1987) Nodulins in the developing root nodule. Plant Physiol. Biochem. 25, 309 - 322.

Gresshoff, P.M., Delves, A.C.,(1986) Plant genetic approaches to symbiotic nodulation and nitrogen fixation in legumes In: Blonstein, A.D., King, P.J. (eds.). A Genetic Approach to Plant Biochemistry. Springer-Verlag: New York, 159 - 206.

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

Halverson, L.J., Stacey, G. (1986). Signal exchange in plant-microbe interactions. Microbiol. Rev. 50, 193 - 225.

Hardy, R.W.F., Holsten, R.D., Jackson, E.K., Burns, R.C., (1968) The acetylene-ethylene assay for N_2 fixation: laboratory and field evaluation. Plant Physiol. 43, 1185 - 1207.

Hirel, B., Bouet, C., King, G., Layzell, D., Jacobs, F., Verma, D.P.S.,(1987) Glutamine synthetase genes are regulated by ammonia provided externally or by symbiotic nitrogen fixation. EMBO J. 6, 1167 - 1171.

Holl, F.B., LaRue, T.A.,(1976). Genetics of legume host plants In: Newton, W.E., Nyman, C.J., (eds.), Proceedings of the 1st International Symposium on Nitrogen Fixation. Washington State University Press: Washington, 391 - 399.

Horvath, B., Christian, W.B.B., Schell, J., Kondorosi, A., (1987) Host specific regulation of nodulation genes in Rhizobium is mediated by a plant-signal, interacting with the nodD gene product. EMBO J. 6, 841 - 848.

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

Hymowitz, T., Singh, J., (1987) Taxonomy and Speciation in Wilcox, J.R. (ed.) Soybeans: Improvement, Production and Uses (2nd ed.) American Society of Agronomy: Madison, 23 - 48.

Ingelbrecht, I.L.W., Herman, L.M.F., Dekeyser, R.A., Van Montagu, M.C., Depicker, A.G., (1989) Different 3' end regions strongly influence the level of gene expression in plant cells. The Plant Cell 1, 671 - 680.

Jacobs, F.A., Zhang, M., Fortin, M.G., Verma, D.P.S., (1987) Several nodulins of soybean share structural domains but differ in their subcellular localization. Nucl. Acids Res. 15, 1271 - 1280.

Jeffreys, A.J. (1982) Evolution of Globin Genes in Dover, G.A., Flavell, R.B. (eds.) Genome Evolution Academic Press, London, 157 - 176.

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

Jensen, E.O., Marcker, K.A., Schell, J., de Bruijn, F.J., (1988) Interaction of a nodule specific, trans-acting factor with distinct DNA elements in the soybean leghaemoglobin lbc_3 5' upstream region. EMBO J. 7, 1265 - 1271.

Jensen, J.S., Marcker, K.A., Otten, L., Schell, J.,(1986) Nodule-specific expression of a chimaeric soybean legheamoglobin gene in transgenic Lotus corniculatus. Nature 321, 669 - 674.

Jordan, D.C. (1982) Transfer of Rhizobium japonicum Buchanan 1980 to Bradyrhizobium japonicum gen. nov., a genus of slow-growing root nodule producing bacteria from leguminous plants. Int. J. Syst. Bacteriol. 32, 136 - 139.

Jorgensen, J.E., Stougaard, J., Marcker, A., Marcker, K.A., (1988) Root nodule specific gene regulation: analysis of the soybean nodulin N23 gene promoter in heterologous symbiotic systems. Nucl. Acids Res. 16, 39 - 50.

Joshi, C.P. (1987) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. Nucl. Acids Res. 15, 9627 - 9640.

Katinakis, P., Verma, D.P.S., (1985) Nodulin-24 gene of soybean codes for a peptide of the peribacteroid membrane and was generated by tandem duplication of a sequence resembling an insertion element. Proc. Natl. Acad. Sci. U.S.A. 82, 4157 - 4161.

Konieczny, A., Szczyglowski, K., Boron, L., Przybylska, M., and Legocki, A.B.,(1988) Expression of lupin nodule genes during root nodule development. Plant Sci. 55, 145 - 149.

Kortt, A.A., Inglis, A.S., Fleming, A.I., Appleby, C.A., (1988) Amino Acid sequence of hemoglobin I from root nodules of the non-leguminous Casaurina glauca - Frankia symbiosis. FEB 231, 341 - 346.

Lang-Unnasch, N., Ausubel, F.M., (1985) Nodule-specific polypeptides from effective alfalfa root nodules and from ineffective nodules lacking nitrogenase. Plant Physiol. 77, 833 - 839.

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

Lederberg, E.M., Cohen, S.N. (1974) Transformation of Salmonella typhimurium by plasmid deoxyribonucleic acid. J. Bacteriol. 119, 1072 - 1074.

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

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

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

Lehrach, H.D., Diamond, D., Wozney, J.M., Boedtker, H. (1977) RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochem. 16, 4743 - 4751.

Liebhaber, S.A., Goossens, M., Kan, Y.W. (1981) Homology and concerted evolution at the α1 and α2 loci of human α-globin. Nature 290, 26 - 29.

Long, S.R.(1989) Rhizobium-legume nodulation: life together in the underground. Cell 56, 203 - 214.

Lullien, V., Barker, D.G., de Lajudie, P., Huguet, T., (1987) Plant gene expression in effective and ineffective root nodules of alfalfa (Medicago sativa L.). Plant Mol. Biol. 9, 469 - 478.

MacDonald, R.J., Swift, G.H., Przybyla, A.E., Chirgwin, J.M. (1987) Isolation of RNA using guanidinium salts. Meth. Enzymol. 152, 219 - 227.

Mangiarotti, G., Chung, S., Zuker, C., Lodish, H.F., (1981) Selection and analysis of cloned developmentally-regulated Dictyostelium discoideum genes by hybridization-competition. Nucl. Acids. Res. 9, 947 - 963.

Maniatis, T., Fritsch, E.F., Sambrook, J., (1982) Molecular Cloning. Cold Spring Harbor Laboratory: Cold Spring Harbour, N.Y.

Marcker, A., Lund, M., Jensen, E.O., Marcker, K.A.,(1984) Transcription of the soybean leghemoglobin genes during nodule development. EMBO J. 3, 1691 - 1695.

Marmur, J., Doty, P.,(1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J. Mol. Biol. 5, 109 - 118.

Mauro, V.P., Nguyen, T., Katinakis, P., Verma, D.P.S., (1985) Primary structure of the soybean nodulin-23 gene and potential regulatory elements in the 5' flanking regions of nodulin and leghemoglobin genes. Nucl. Acids Res. 13, 239 - 249.

McDevitt, M.A., Hart, R.P., Wong, W.W., Nevins, J.R., (1986) Sequences capable of restoring poly (A) site function define two distinct downstream elements. EMBO J. 5, 2907 - 2913.

McLauchlan, J., Gaffney, D., Whitton, J.L., Clements, J.B., (1985) The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucl. Acids Res. 13, 1347 - 1368.

McRae, D.G., Miller, R.W., Berndt, W.B., Joy, K., (1989) Transport of C_4-dicarboxylates and amino acids by Rhizobium meliloti bacteroids. Molec. Plant-Microbe Interact. 2, 273 - 278.

Mellor, R.B., Christensen, T.M.I.E., Werner, D., (1986) Choline kinase II is present only in nodules that synthesize stable peribacteroid membranes. Proc. Natl. Acad. Sci. U.S.A. 83, 659 - 663.

Messing, J., Crea, R., Seeburg, P.H., (1981) A system for shotgun DNA sequencing. Nucl. Acids Res. 9, 309 - 322.

Miflin, B.J., Cullimore, J.V.,(1984) Nitrogen assimilation in the legume-Rhizobium symbiosis: a joint endeavor in Verma, D.P.S., Hohn, J.H. (eds.). Genes Involved in the Plant Microbe Interaction. Springer-Verlag: New York, 129 - 178.

Moerman, M., Nap, J.P., Govers, F., Schilperoort, R., van Kammen, A., Bisseling, T.,(1987) Rhizobium nod genes are involved in the induction of two early nodulin genes in Vicia sativa root nodules. Plant Mol. Biol. 9, 171 - 179.

Mohapatra, S.S., Perlick, A., Puhler, A.,(1987) Nodule specific polypeptides of broadbean (Vicia faba L.) Symbiosis 3, 219 - 232.

Morell, M., Copeland, L.,(1985) Sucrose synthase of soybean nodules. Plant Physiol. 78, 149 - 154.

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

Newcomb, W. (1981) Nodule morphogenesis. Int. Rev. Cytol. Supp. 13, 246 - 298.

Newcomb, W., Wood, S.M. (1987) Morphogenesis and Fine Structure of Frankia (Actinomycetales): The microsymbiont of nitrogen-fixing actinorhizal root nodules. Int. Rev. Cytol. 109, 1 - 88.

Nguyen, T., Zelechowska, M., Foster, V., Bergmann, H., Verma, D.P.S.,(1985) Primary structure of the soybean nodulin-35 gene encoding uricase II localized in the peroxisomes of uninfected cells of nodules. Proc. Natl. Acad. Sci. U.S.A. 82, 5040 - 5044.

Nutman, P.S.,(1981) Hereditary host factors affecting nodulation and nitrogen fixation In: Gibson, A.H., Newton, W.E., (eds.)., Current Perspectives in Nitrogen Fixation. Aust. Acad. Sci.: Canberra, 105 - 116.

O'Brian, M.R., Kirshbom, P.M., Maier, R.J.,(1987) Bacterial heme synthesis is required for expression of the leghemoglobin holoprotein but not the apoprotein in soybean root nodules. Proc. Natl. Acad. Sci. USA 84, 8390 - 8393.

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

Proudfoot, N.J., Brownlee, G.G., (1976) 3' non-coding region sequences in eukaryotic messenger RNA. Nature **263**, 211 - 214.

Pühler, A., (ed.)., (1983) Molecular Genetics of the Bacteria-Plant Interaction. Springer-Verlag: Berlin.

Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J., Hanahan, D. (1988) McrA and McrB restriction phenotypes of some E. coli strains and implications for cloning. Nucl. Acid Res. **16**, 1563 - 1575.

Reibach, P.M., Streeter, J.G.,(1983) Metabolism of ^{14}C-labelled photosynthate and distribution of enzymes of glucose metabolism in soybean nodules. Plant Physiol. **72**, 634 - 640.

Reynolds, P.H.S., Boland, M.J., Blevins, D.G., Schubert, K.R., Randall, D.D.,(1982) Enzymes of amide and ureide biogenesis in developing soybean nodules. Plant Physiol. **69**, 1334 - 1338.

Rigby, P.W.J., Dieckmann, M., Rhodes, C., 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., Farnden, K.J.F.,(1980) Ultrastructure and metabolism of the developing legume root nodule in: Miflin, B.J. (ed.). The Biochemistry of Plants. A Comprehensive Treatise. Vol. 5. Academic Press: New York: 65 - 113.

Rolfe, B.G., Shine, J.,(1984) Rhizobium-Leguminsae symbiosis: the bacterial point of view in Verma, D.P.S., Hohn, T.H., (eds.) Genes Involved in Microbe Plant Interactions. Springer-Verlag, New York, 95 - 128.

Rolfe, B.G., Gresshoff, P.M.,(1988) Genetic analysis of legume nodule initiation. Ann. Rev. Plant Physiol. Plant Mol. Biol. **39**, 297 - 319.

Sandal, N.N., Bojsen, K. Marcker, K.A., (1987) A small family of nodule-specific genes from soybean. Nucl. Acids Res. **15**, 1507 - 1519.

Sandal, N. N., Marcker, K.A.,(1988) Soybean nodulin 26 is homologous to the major intrinsic protein of the bovine lens fiber membrane. Nucl. Acids Res. **16**, 9347.

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

Schubert, K.R., Boland, M.J.,(1984) The cellular and intracellular organization of the reactions of ureide biogenesis in nodules of tropical legumes In: Veeger, C., Newton, W.E., (eds.)., Advances in Nitrogen Fixation Research, Nijoff/Junk: The Hague, 445 - 451.

Schubert, K.R., (1986) Products of biological nitrogen fixation in higher plants. Ann. Rev. Plant Physiol. 37, 539 - 574.

Sengupta-Gopalan, C., Pitas, J.W.,(1986) Expression of nodule-specific glutamine synthetase genes during nodule development in soybeans. Plant Mol. Biol. 7, 189 - 199.

Sengupta-Gopalan, C., Pitas, J.W., Thompson, D.V., Hoffman, L.M., (1986) Expression of host genes during root nodule development in soybeans. Mol. Gen. Genet. 203, 410 - 420.

Shiels, A., Kent, N.A., McHale, M., Bangham, J.A.,(1988) Homology of MIP26 to Nod26. Nucl. Acids Res. 16, 9348.

Sievers, G., Huhtala, M.L., Ellfolk, N.,(1978) The primary structure of soybean (Glycine max) leghemoglobin c*. Acta Chem. Scand. B32, 380 - 386.

Slightom, J.L., Blechl, A.E., Smithies, O., (1980) Human fetal $^{G}\gamma$ and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can exchanged between these duplicated genes. Cell 21, 627 - 638.

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

Stougaard, J., Sandal, N.N., Gron, A., Kuhle, A., Marcker, K.A.,(1987) 5' analysis of the soybean leghaemoglobin lbc_3 gene: regulatory elements required for promoter activity and organ specificity. EMBO J., 6, 3565 - 3569.

Strittmatter, G., Chia, T.F., Trinh, T.H., Fumiaki, K., Kuhlemeier, C., Chua, N.H.,(1989). Characterization of nodule-specific cDNA clones from Sesbania rostrata and expression of the corresponding genes during the initial stages of stem nodules and root nodules formation. Mol. Plant-Microbe Inter. 2, 122 - 127.

Sugita, M., Manzara, T., Pichersky, E., Cashmore, A., Gruissem, W. (1987) Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5 - bisphosphate carboxylase/oxygenase from tomato. Mol. Gen. Genet. 209, 247 - 256.

Sullivan, D.E., Brisson, N., Verma, D.P.S.,(1980) Reverse transcription of 25S soybean ribosomal RNA in the absence of exogenous primer. Biochem. Biophys. Res. Comm. 94, 144 - 150.

Sutton, W.D., Pankhurst, C.E., Craig, A.S. (1981) The Rhizobium bacteroid state. Int. Rev. Cyt. Suppl. 13, 149 - 177.

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

Thomas, P.S. (1983) Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Meth. in Enzymol. 100, 255 - 266.

Thummler, F., Verma, D.P.S., (1987) Nodulin-100 of soybean is the subunit of sucrose synthetase regulated by the availability of free heme in nodules. J. Biol. Chem. 262, 14730 - 14736.

Tingey, S.V., Walker, E.L., Coruzzi, G.M.,(1987) Glutamine synthetase genes of pea encode distinct polypeptides which are differentially expressed in leaves, roots, and nodules. EMBO J. 6, 1 - 9.

Tjepkema, J.D., Yokum, C.S.,(1973) Respiration and oxygen transport in soybean nodules. Planta 115, 59 - 72.

Tjepkema, J.D., Yokum, C.S.,(1974) Measurement of oxygen partial pressure within soybean nodules by oxygen microelectrodes. Planta 119, 351 - 360.

Tjepkema, J.D.,(1983) Hemoglobin in the nitrogen-fixing root nodules of actinorhizal plants. Can. J. Bot. 61, 2924 - 2929.

Triplett, E.W.,(1985) Intracellular nodule localization and nodule specificity of xanthine dehydrogenase in soybean. Plant Physiol. 77, 1004 - 1009.

Truelsen, E., Gausing, K., Jochimsen, B. Jorgensen, P., Marcker, K.A., (1979) Cloning of soybean leghemoglobin structural gene sequences synthesized in vitro. Nucl. Acids Res. 6, 3061 - 3072.

Tsao, S.G.S., Brunk, C.F., Pearlman, R.E. (1983) Hybridization of nucleic acids directly in agarose gels. Anal. Biochem. 131, 365 - 372.

Turgeon, B.G., Bauer, W.D.,(1982) Early events in the infection of soybean by Rhizobium japonicum: Time course and cytology of the initial infection process. Can. J. Bot. 60, 152 - 161.

Vance, C.P., Boylan, K.L.M., Stade, S., Somers, D.A., (1985) Nodule specific protein in alfalfa (Medicago sativa L.). Symbiosis 1, 69 - 84.

Vance, C.P., Egli, M.A., Griffith, S.M., Miller, S.S., (1988) Plant regulated aspects of nodulation and N_2 fixation. Plant Cell Environ. 11, 413 - 427.

van den Bosch, K.A., Newcomb, E.H., (1986) Immunogold localization of nodule-specific uricase in developing soybean nodules. Planta 167, 425 - 436.

van den Bosch, K.A., Newcomb, E.H.,(1988) The occurrence of of leghemoglobin protein in the uninfected interstitual cells of soybean root nodules. Planta 175, 442 - 451.

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

Verma, D.P.S., Haugland, R., Brisson, N., Legocki, R. 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., Nadler, K.,(1984) Legume-Rhizobium symbiosis: Host's point of view in: Verma, D.P.S., Hohn, T., Genes Involved in Microbe-Plant Interaction. Springer Verlag: New York, 57 - 93.

Verma, D.P.S., Fortin, M.G., Stanley, J., Mauro, V.P., Purohit, S., Morrison, N.,(1986) Nodulins and nodulin genes of Glycine max. Plant Mol. Biol. 7, 51 - 61.

Verma, D.P.S., Delauney, A.,(1988) Root Nodule Symbiosis: nodulins and nodulin genes. In: Verma, D.P.S., Goldberg, R.B., (eds.), Temporal and Spatial Regulation of Plant Genes., Springer Verlag: New York, 169 - 199.

Wallace, R.B., Miyada, C.G. (1987) Oligonucleotide probes for the screening of recombinant DNA libraries. Meth. in Enzymol. 152, 432 - 442.

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

Williams, J.G., Lloyd, M.M.,(1979) Changes in the abundance of polyadenylated RNA during slime mould development measured using cloned molecular hybridization probes. J. Mol. Biol. 129, 19 - 35.

Wittenberg, J.B., Bergerson, F.J., Appleby, C.A., 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.

Yannish-Perron, C., Vieira, J., Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC 19 vectors. Gene 33, 103 - 119.

Yates, M.G. (1980) Biochemistry of nitrogen fixation in Miflin, B.J. (ed)., The Biochemistry of Plants, Vol. 5, Academic Press: New York, 2 - 65.

APPENDIX I

Appendix I is a draft of a paper that has been submitted to Genome. It describes the characterization of a clone, isolated from the soybean nodule cDNA library; that contains one intact and one truncated copy of the soybean 5S gene sequence, separated by non-coding (or intercistronic) DNA. This paper describes the characterization of this clone including sequence, and its use to investigate the 5S gene structure in soybean. As described in the manuscript, the soybean 5S genes are clustered, tandemly repeated, and heavily methylated (as determined by the use of restriction enzyme isoschizomers). In addition, we show that the 5S gene structure is conserved in higher plants. Also, this clone is a direct demonstration that genomic DNA can be a contaminant of a cDNA library.

I initially isolated this clone from the cDNA library as a clone which appeared to hybridize equally well to end-labeled root and nodule poly $(A)^+$ enriched RNA; and determined by northern blot analysis, that the clone represented a low molecular weight poly $(A)^-$ RNA species.

Sequencing of the clone indicated that it contained 5S gene sequences. In addition, I supervised summer students and fourth year students who isolated plant DNAs and did some of the methylation studies.

Organization of the 5S rRNA genes in the soybean Glycine max (L.) Merrill and conservation of the 5S rDNA repeat structure in higher plants

Gottlob-McHugh, S.G., Lévesque, M., MacKenzie, K., Olson, M., Yarosh, O. and D.A. Johnson. Department of Biology, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5

Corresponding author: Dr. D.A. Johnson, 1-613-564-4236.

Abstract

The 5S rRNA gene of the soybean Glycine max (L.) Merr. has been cloned on a 556bp fragment of DNA and sequenced. This fragment contains two copies of the soybean 5S rDNA sequence, one intact and one truncated, separated by non-coding DNA. We have used this clone to investigate the organization of the 5S genes within the soybean genome and the extent of their methylation. Our results demonstrate that soybean 5S genes are clustered, organized into tandem repeats of 330bp, and extensively methylated. Hybridization of the 5S sequence to Southern transfers of soybean DNA digested with BamH1 reveals a striking "ladder-like" pattern. Hybridization of the soybean 5S sequence to a wide variety of plant DNAs results in similar patterns, suggesting that the 5S rDNA sequence, gene organization, and methylation pattern are conserved in many higher plants.

Keywords: 5S rDNA, sequence, methylation, soybean, repeat conservation.

Introduction

Studies of the 5S rDNA in a number of higher eucaryotes have shown that, in most eucaryotes, the genes are organized into clusters of tandem repeats of several hundred base pairs, consisting of a 120 base pair (bp) coding region and a non-coding spacer region (Long and Dawid 1980). In many higher plants; for example corn: (Mascia et al. 1981), wheat (Gerlach and Dyer 1980), rice (Hariharan et al. 1987), flax (Goldsbrough et al. 1982), and lupin (Rafalski et al. 1982), this organization is conserved.

Within a plant species, several size classes of repeat may be found. Such variability is due to differences in the sizes of the spacer regions (Gerlach and Dyer 1980; Khvyrleva et al. 1988; Reddy and Appels 1989). Also, methylation of cytosine residues within the repeat has been reported (Mascia et al. 1981; Goldsbrough et al. 1982; Ellis et al. 1988). For these two reasons, Southern hybridization of plant DNA probes derived from 5S RNA or 5S rDNA sequences, reveal complex patterns unexpected for a simple repeating structure. In many plants, a ladder-like pattern is observed when the DNA has been digested with BamH1 (Mascia et al. 1981; Goldsbrough et al. 1982; Rafalski et al. 1982; Khvyrleva et al. 1988; Scoles et al. 1988).

Sequence comparisons of the published 5S repeat units of various plant species have demonstrated that the 5S coding regions are highly conserved among different plant species (Ellis et al. 1988; Scoles et al. 1988; Wolters and Erdmann 1988); but in general, the intercistronic regions lack sequence homology. This is to be expected for structures with functional genes separated by non-coding sequences.

In this study we have cloned and sequenced the 5S rDNA repeat from the soybean Glycine max (L.) Merrill (cv. Maple Arrow) and investigated the organization and extent of methylation of the 5S repeat in the soybean genome. In addition we have used the soybean 5S gene as a probe to investigate 5S gene organization in over 30 species of plants.

Materials and Methods

Plant Materials

The classification of plants throughout this study was suggested to us by J. McNeill and is based upon the work of Cronquist et al. (1966) and Wittaker and Margulis (1978). Most plants were collected locally and identified by the Plant Systematics Division of Agriculture Canada. Seeds were purchased from F.W. Schumacher Co. In., Sandwich Mass., USA.

DNA preparation

The method of Hattori et al. (1987) was used to isolate high molecular weight DNA from the following plants:

(a) Magnoliophyta, Magnoliopsida(dicots): Acer rubrum, Albizia lebbek, Alnus incana, Betula papyrifera, Cassia fistula, Casuarina equisetifolia, Ceratonia siliqua, Curcurbita sativus, Cynara scolymus, Datisca glomerata, Elaeagnus angustifolia, Gleditsia triacanthos, Glycine max (L.) Merr., Gymnocladus dioicus, Myrica gale, Myrica pensylvanica, Malus sp., Nicotiana tabacum, Phaseolus vulgaris, Populus deltoides, Prosopis alba, Spinacia vulgaris, Syringa vulgaris, Trifolium repens;

(b) Magnoliophyta, Liliopsida(monocots): Allium cepa, Dracaena margina, Hordeum vulgare, Polygonatum multiflora, Zea mays;

(c) Pinophyta, Pinopsida: Juniperus viginiana, Picea glauca;

(d) Pinophyta, Cycadopsida: Cycas revoluta;

(e) Polypodiophyta: Athyrium felix-femina;

(f) Equisetophyta: Equisetum arvense;

(g) Bryophyta, Bryopsida: Bryum sp.;

(h) Volvophyta: Chlamydomonas segnis

DNA techniques

Restriction endonuclease digestions of DNA were performed according to the manufacturer's recommendations and digested DNA was electrophoresed through 1.25% agarose gels (Maniatis et al. 1982). When methylation sensitive enzymes were used (BamH1, HpaII, Msp1, Sau3A), lambda DNA was added to the reaction mixture as an internal control for digestion. Simple cloning protocols were taken from Maniatis et al. (1982). Plasmid DNA was prepared by the procedure of Birnboim and Doly

(1979) followed by centrifugation in CsCl gradients containing ethidium bromide.

For Southern hybridization experiments, fragments were transferred to BIOTRANS membrane (ICN) according to Southern (1979). Hybridization was in 5 X SSC (1 X SSC is 0.15 M NaCl, 0.015 M sodium citrate), 50 mM sodium phosphate pH 7.0, 5 mM EDTA, 5 X Denhardt's (1 X Denhardt's is 0.02% ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 0.2% SDS and 100μg/mL denatured herring sperm DNA, at 65°C for 16 h. Filters were washed at room temperature for 3 X 30 min in 5 mM sodium phosphate pH 7.0, 1 mM EDTA, 0.2% SDS with a final wash in the same buffer for 30 min at 65°C. Hybridization probes were derived from the cloned 5S repeat (see Figure 1). Cloned fragments (in plasmids or M13 RF DNAs) were labelled by nick-translation (Rigby et al. 1977) and purified fragments were labelled by the random primer method (Feinberg and Vogelstein 1983) with the aid of kits (Amersham).

Fragment sizes were determined using as standards, lambda DNA restricted with HindIII, pBR322 DNA digested with TaqI, and ϕX174 DNA cut with HaeIII. Repeat lengths for different plant 5S genes were calculated from Southern blots of BamH1 digested DNA. Starting with the band of lowest molecular weight, a number was assigned to each band and the size of the band in the ladder was divided by this repeat number to arrive at a value for the size of the repeat. These values were averaged for all of the bands detected by Southern hybridization. Soybean DNA, digested with BamH1, was included on most gels as an internal standard, giving rise to a ladder with a repeat length of 330bp. Autoradiograms were traced using an LKB Ultrascan XL densitometer.

Cloning and sequencing of soybean 5S repeats

A clone 19-10-D containing 5S sequences, was fortuitously isolated from a soybean cDNA library. Restriction fragments of 19-10-D were subcloned into M13mp18 or M13mp19 (Yannish-Perron et al. 1985) RF DNA and introduced into Escherchia coli strain JM101 by transformation. Single-stranded DNA was isolated and sequenced by the dideoxy chain termination procedure (Sanger et al. 1977). The 19-10-D insert was

sequenced in both orientations. Sequence analysis was performed with MicroGenie computer software (Beckman).

To confirm the sequence of the soybean 5S rDNA repeat, additional 5S genes were isolated following the ligation of BamH1 digested soybean DNA into M13mp19. The cloned DNA was transformed into E. coli strain DH5α mcr (BRL) and plated on E. coli JM109. DH5α mcr is a recA1 strain that will not restrict methylated plant DNA, because it lacks the McrA and McrB methylcytosine-specific restriction systems present in E. coli K12 strains. Recombinant phage were screened by plaque hybridization and M13 clones, which hybridized to the 330 bp BamH1 fragment from 19-10-D (Figure 1), were further purified and sequenced.

Results

Sequence organization and comparisons

The clone 19-10-D was fortuitously cloned during the construction of a soybean cDNA library. Initial northern hybridization experiments indicated that the 19-10-D insert hybridized to a low molecular weight RNA species present in total and poly A$^-$ RNA preparations. Subsequent sub-cloning of overlapping fragments followed by sequencing, indicated that the clone contained 5S gene sequences. The sequence of the insert is presented in Figure 1.

Within the 556 (bp) insert, there is one complete copy of a 5S repeat (330 bp) consisting of the 5S gene and a spacer, a truncated copy of the spacer 5' to the gene, and an incomplete copy of the 5S gene at the 3' end. The coding sequence is highly homologous to other plant and animal 5S gene sequences. By comparison to plant sequences (Vandenberghe et al. 1984; Ellis et al. 1988; Wolters and Erdmann 1988), we infer that the coding region starts 5' with the AGG (labelled position +1) that has also been observed in other legumes (Vandenberghe et al. 1984; Barciszewska et al. 1987; Ellis et al.1988; Hemleben and Werts 1988), and ends with CTC (position +119) to give a 5S RNA size of 119 nucleotides. Previously it has been shown that the 3' end of the 5S gene from yellow lupin can be variable (Rafalski et al. 1982) and therefore our assignment of an exact 3' end, and RNA size, must be considered tentative. The sequence also contains a T at position +118

which appears to be a feature that is common to legumes (Ellis et al. 1988).

It has been postulated that conserved sequences are involved in the transcription of plant 5S genes. Similar sequences are found in this soybean sequence at their expected locations. The sequence AATATA found at position -30 to -25, resembles the "TATA" region or AT rich sequence shown to be required for transcription in Neurospora (Selker et al. 1986), and found in other plants 5' to the intiation site for transcription (Hemleben and Werts 1988). TATA like sequences can be found in a similar location in at least three other legumes: yellow lupin (Rafalski et al. 1982), pea (Ellis et al. 1988) and mung bean (Hemleben and Werts; 1988). Scoles et al. (1988) have observed a similar AT rich conserved sequence, ATAAG found at position -25 to -29, in members of the Triticeae. Also, we find clusters of T residues 3' to the coding sequence (+120 to +144). Such T rich sequences have been implicated in transcription termination of Xenopus 5S genes (Korn 1982), and are found in many plant 5S repeats (Hemleben and Werts 1988; Scoles et al. 1988).

To confirm the sequence of the 5S gene repeat from 19-10-D, as well as to investigate heterogeneity of 5S sequences in soybean, several additional 5S isolates were sequenced. The sequences of eight isolates compared to the 19-10-D repeat sequence are presented in Figure 2. The sequences in Figure 2 demonstrate that the soybean 5S repeats are relatively homogeneous. One isolate (isolate "a") is identical to the 19-10-D sequence. The other seven isolates showed some sequence heterogeneity. The most significant region of heterogeneity was found to occur in the 11 bp T cluster (+128 to +138) 3' to the coding region. In seven of the eight isolates, there were either insertions or deletions of T residues in this region. In three of the isolates (isolates "c","d", and "h"), the 5S repeat sequence was identical to 19-10-D, except for additional T residues in the T cluster. The remaining four isolates contained deletions within the T cluster. These four isolates also contained a total of fourteen base changes scattered throughout the repeat; however, there did not appear to be any trend as to where these changes were located. Seven of these 14 changes were C

to T transitions. Compared to 19-10-D, there were no insertions or deletions within any of the isolates except within this T rich region; however, the heterogeneity in this region caused variability in the size of the repeat from 321 bp (isolate "c") to 333 bp (isolates "d" and "g").

Genomic organization and methylation pattern

In all of the higher plant species that have been investigated, the 5S genes are tandemly arranged, clustered and extensively methylated. Southern hybridization experiments, using the cloned gene as a probe, suggest that in soybean this is also the case. When soybean DNA was digested with BamHl, a ladder of bands, with sizes that are multiples of the monomer size, was observed (Figure 3A). This observation is consistent with a tandem arrangement of the repeat (Mascia et al. 1981; Goldsbrough et al. 1982; Rafalski et al. 1982; Scoles et al 1988). This pattern was not due to incomplete digestion, since repeated extraction of digested DNA followed by redigestion (performed 5 times) did not change the observed pattern. The monomer size of 340bp, as determined by gel electrophoresis, is in agreement with the size of 330bp calculated from the sequence. Up to 25 bands can be observed suggesting that the 5S genes are clustered. Clustering is also indicated by the results of digestion with EcoRI, HindIII, PstI or XhoI. Cleavage with any one of these enzymes gives a high molecular weight smear (not shown) indicating that no cleavage sites for these enzymes are found within the repeat. Double digestions of soybean DNA with BamHI and any of these enzymes did not change the sizes of the bands observed (Figure 3A) when compared to digestion with BamHI alone. These results indicated that the 5S genes must be tandemly repeated and clustered.

The pattern observed with BamHI restriction could be due to sequence heterogeneity and/or methylation. Densitometer tracing of the bands seen in Figure 3A suggest that only 12% of the DNA is found in the monomer fraction. Thus either methylation is extensive, or few 5S genes contain the BamHI restriction site seen in the sequence. Soybean DNA (McClelland, 1983), like all plant DNA (Gruenbaum et al. 1981), is highly methylated at C residues in the sequences CpG and CpNpG.

Methylation of the soybean 5S gene was investigated by the use of isoschizomers which are differentially sensitive to methylation of a base(s) within the recognition sequence. The sequence CCGG occurs two times within the repeat (positions +28 and +285) and is recognized by HpaII and MspI. HpaII does not cut either mCCGG or CmCGG while MspI does not cut when the external C is methylated (Nelson and McClelland 1989). Southern blots of digested soybean DNA indicate that, at this level of detecton, HpaII does not cleave within the repeat but MspI cleaves to produce a ladder identical to the BamHI ladder (Figure 3B). These results suggest that the HpaII/MspI restriction sites are extensively methylated. Digestion with MspI suggests that the external C is variably methylated. It is also possible that the MspI pattern is partly due to sequence heterogeneity. This possibility is supported by the sequence of isolate "f" which has a T instead of a C residue at position +29, thereby altering the MspI site located at position +28 from CCGG to CTGG.

The sequence GATC (found 4 times, at positions +7, +31 (within the BamHI sequence), +214, and +267) is recognized by MboI but not by Sau3A when the C is methylated (Nelson and McClelland 1989). Methylation at this site also blocks cleavage by BamHI of GGATCC. Digestion with Sau3A produces a ladder identical to the BamHI ladder, while digestion with MboI produces fragments of 187bp and 74bp (Figure 3C). From the sequence data, sizes of 183bp and 70bp were predicted assuming that fragments of 53bp and 24bp were lost from the analysis. These results suggest that the DNA containing the 5S genes is extensively methylated at the GATC sites and that the hybridization patterns observed are not due to extensive heterogeneity of the recognition sequences.

Digestion with BamHI, or MspI, or Sau3A gives apparently identical ladder patterns due to variable methylation. We interpret our results as follows: the single BamHI site is variably methylated; one MspI site is always methylated and the second site variably methylated; three Sau3A sites are always methylated and one is variably methylated. To determine directly which site(s) is differentally methylated, double digestions were performed using combinations of enzymes that individually gave ladder-like patterns. Digestion with BamHI plus

Sau3A, or BamHI plus MspI, or MspI plus Sau3A, did not alter the pattern observed following restriction with one enzyme alone (Figure 3B). Thus the observed Southern patterns must be due to differential methylation of the overlapping MspI and BamHI recognition sites beginning at nucleotide +28(CCGGATCC). Extensive methylation at this sequence in maize (Mascia et al. 1981) and barley (Khvyrleva et al. 1988) has been previously reported .

Conservation of the 5S repeat structure

The tandemly repeated structure of the 5S genes is found in many higher eucaryotes. In higher plants a ladder of bands is often observed following Southern transfers of BamHI digested DNA, suggesting extensive modification of the BamHI site within a tandemly repeated 5S sequence (Mascia et al. 1981; Goldsbrough et al. 1982; Rafalski et al. 1982; Khvyrleva et al. 1988; Scoles at al. 1988; this work). To determine the extent of conservation of the 5S sequence and organization among plants species, we hybridized the soybean 5S sequence to BamHI restricted DNA isolated from a wide variety of plants representing many divisions. Using stringent hybridization criteria, the 5S probe hybridized to all of the plant species tested, including four species of lower plants. In addition, a ladder pattern was observed in all of the higher plant species tested (see Figure 4 for examples). These higher plants include: 24 members of the Magnoliophyta class Magnoliopsida (dicots); 5 members of the Magnoliophyta, class Liliopsida (monocots); 2 members of the Pinophyta, class Pinopsida (conifers); and 1 member of the Pinophyta, class Cycadopsida(cycads). The calculated repeat sizes for 29 of these higher plant species are presented in Table 1. Typically a simple pattern is observed as seen for soybean, but more complicated double (eg. poplar) or triple (eg. alder or birch, Figure 4A) banding patterns have been seen. The sizes of the repeats range from 215 for the kentucky coffeetree to 1572 for the cycad. The size observed for the cycad is much larger than the repeat sizes reported for other plant species (Table 1; references in Introduction). Since the size of the 5S coding sequence is conserved between widely divergent organisms, our results suggest that the spacer region of cycads are much larger than for any other reported 5S repeat.

The sizes of repeats determined in our study, are similar to sizes reported in other studies. The size of 323 bp for the maize repeat, corresponds to the size of 320 reported by Mascia et al. (1981); and the sizes of 304 and 442 determined for the barley repeats, correspond to the sizes of 301 and 450 determined by Khvyrleva et al. (1988). Three of the plant species, golden shower, the flowering crab apple and white spruce, had ladder patterns that demonstrated periodicity but were too complex to determine repeat sizes. These more complex patterns are likely due to the presence of more than one class of repeat within these plant species.

We have detected hybridization to lower plants but the characteristic ladder pattern is no longer seen; however, we have observed multiple bands that suggest the 5S sequence is repeated in these plants (see Figure 4D, lane 2 for example). This group includes one member each of the Polypodiophyta (Athyrium felix-femina - lady fern; Figure 4, lane 2), Equisetophyta (Equisetum arvense - horsetail, not shown), Bryophyta (Bryum sp.- moss, not shown). A single band was observed for Chlamydomonas segnis, a member of the Volvophyta. Although our survey is biased towards the higher plants, we conclude that the 5S gene sequence is conserved in all plants; but that the characteristic ladder pattern, suggesting clustering and tandem repetition of the 5S repeats, is restricted to higher plants.

The intercistronic region is thought not to contain coding sequences although in rice the spacer can be folded into a tRNA-like structure (Hariharan et al. 1987). The possibility that the spacers may contain conserved and possibly functional sequences was investigated by hybridizing an intercistronic probe derived from soybean (the HpaI to HpaII fragment (see Figure 1) cloned into M13mp19) to the blots used in the above experiments. In soybean, as expected, the identical ladder pattern was observed but we did not observe hybridization to the same set of bands, visualized by the 5S gene probe, in any other plant DNA. Thus under stringent hybridization conditions no homologous sequences were detected, indicating that the intercistronic regions have diverged.

Discussion

These studies indicate that the arrangement of the 5S genes in soybean is identical to that in other higher plants: tandemly repeated, clustered and highly methylated. At the sequence level, the coding region is highly homologous to other plant 5S gene sequences (Wolters and Erdmann 1988); and therefore should be able to fold into the typical structures postulated for 5S RNAs (Vandenberghe et al 1984; Wolters and Erdmann 1988). The soybean coding region is identical to the sequence of the mung bean Vigna radiata (Hemleben and Werts 1988) and differs from other legume sequences by only one or two nucleotides (Ellis et al. 1988; Wolters and Erdmann 1988). Soybean 5S repeats appear to have relatively little sequence heterogeneity. Most of the variation occurs in the T rich 3' transcription termination region, causing this region to range from 16-28 bp in length. It is possible that deletions and insertions within this region are generated by slippage during DNA replication. Variation in the length of the T stretch likely does not alter transcription termination, since this region exhibits considerable variation in length among different plant species.

Southern hybridization following digestion of soybean DNA with a variety of restriction enzymes (BamHI, MspI, Sau 3A), gives rise to a striking pattern. This pattern is largely due to extensive methylation of the DNA sequences within the recognition sites for these enzymes. The extent of methylation and the exact positions of the methylated sites were investigated by the use of isoschizomers. Results following digestion with Sau3A or MboI suggest that there is little sequence heterogeneity at the GATC sites within the repeat, and that modification to GATmC must occur frequently. HpaII did not cleave within the 5S region suggesting that, at this level of detection, the two CCGG sites in the repeat are modified at at least one C residue. Digestion with MspI gives rise to a ladder pattern based upon a monomer size of 330bp. These results suggest that the internal C of the sequence CCGG is heavily methylated (cleaved by MspI, not HpaII), whereas the external C is variably methylated with a high enough frequency of methylation to explain the observed MspI pattern. However, some sequence heterogeneity may also contribute to the observed pattern.

As shown in Figure 3B, double digestion of soybean DNA with various restriction enzymes did not change the pattern or the sizes of the bands observed with single digestion alone. This result allows us to fix the sites of variable modification for each pair of isoschizomers within the 5S sequence as, the CCGG at position +28, and the GATC at position +31. In a typical repeat, all other sites are modified so as to prevent cleavage: mCCGG (or mCmCGG) at position +285; and GATmC at positions +7, +214, and +267. The fact that three of the GATC sites appear to be extensively modified, thereby preventing cleavage with Sau3A, can in each case be explained by extensive methylation at a CpG doublet. The Sau3A sites are preceeded by a C residue (+7, +267) to give the sequence CGATC or followed by a G residue (+214) to give the sequence GATCG. Methylation of the C residues in the CpG doublet would generate hemi-methylated sites that inhibit the cleavage of these sites by Sau3A (Nelson and McClelland 1989). Also the results obtained with HpaII/MspI suggested that the C residue, within the CpG doublet of the HpaII/MspI site, is extensively methylated. Our results indicate that the Sau3A site located at +31 is variably methylated. This may be due to variable methylation of the C residue complimentary to the G residue at +31. This C residue falls within the triplet CpCpG. The C residue at this position also corresponds to the external C of the MspI site. Variable methylation of this C residue could explain the results observed following digestion with Sau3A, MspI, or BamHI. The fact that methylation at this site appears to be variable suggests that methylation at the CpNpG triplet is not as extensive as at the CpG doublet within the soybean 5S repeat. We were unable to explain why the MspI site located at position +285 did not appear to be variably methylated. It is possible that MspI cleavage at this site is affected by surrounding nucleotides (Nelson and McClelland 1989). Since MboI does not restrict the sequence GmATC (Nelson and McClelland 1989), modification of the A within this sequence does not occur.

Our results do not eliminate the possibility that a minor fraction of the soybean 5S rDNA exists in a class of repeat with a different spacer length, since overexposure of our blots reveals additional faint bands which may be due to such repeat heterogeneity (results not shown).

Several size classes of repeat have been reported in various plant species including wheat (Gerlach and Dyer 1980), flax (Goldsbrough et al. 1982), barley (Khvyrleva et al. 1988) and rye (Lawrence and Appels 1986, Reddy and Appels 1989). We have studied the nature of the 5S repeat in over 30 species of higher plants. Most of the plants within our study appear to contain only one class of repeat; but Southern hybridization analysis of BamHI digested plant DNA, has suggested the presence of more than one class of repeat in several plant species. In our study, double and triple band hybridization patterns were observed in a number of plants (Table I, Figure 4). However, repeat heterogeneity is not always the correct explanation for multiple banding patterns, since we have shown that the triple band patterns detected in birch and alder (Figure 4A) are due to the presence of a second BamHI site within the repeat unit (Johnson, unpublished). Differential methylation of the two BamHI sites can account for the observed repeat pattern.

The fact that 5S sequences could be detected in a wide variety of plants under conditions of stringent hybridization, indicate that the 5S coding sequences are very conserved in both higher and lower plants. This is confirmed by comparisons of the various available 5S rRNA and rDNA sequences which indicate that the 5S coding sequences among plants are very similar (Wolters and Erdman 1988). Our analyses indicate however, that the conservation does not extend into the non-transcribed spacer region, since the the sequences from the soybean spacer region do not hybridize to the DNA of any of the plant species investigated. Our results are consistent with those of Scoles et al. (1988) who have demonstrated that even within the closely related members of the Triticeae, the 5S coding sequence is conserved whereas the intergenic spacer has diverged considerably. These results indicate that the 5S sequence is under strong evolutionary constraints, but that this constraint does not extend into the spacer region except perhaps in regions required for gene regulation and transcription.

We have also demonstrated that the 5S gene organization is conserved in a wide variety of higher plants, but not the lower plants. We have observed ladder patterns in all of the higher plants

investigated, suggesting that the 5S genes are clustered and tandemly repeated. Hybridization to BamHI digested DNA from lower plants, demonstrates the presence of multiple bands; but the periodicity of the pattern observed in the higher plants was not observed in the lower plants. The lack of periodicity in the hybridization pattern may be due to the fact that some lower plants lack the BamHI site (Wolters and Erdmann 1988). But, Equisetum arvense (horsetail) contains the BamHI site (Wolters and Erdmann, 1988); and yet no ladder-like hybridization pattern was detected. These results suggest that the organization and/or the methylation of 5S genes is different in lower plants, when compared to higher plants. All of the species, which in our study revealed a ladder pattern, may have arisen from a single ancestor, the Progymnosperm (Margulis and Schwartz, 1982); therefore, tandem duplication of a repeat containing a BamHI site may have occurred before divergence from this progenitor.

Acknowledgements

The authors would like to thank their colleagues for either sending plant materials or helping with the collection and identification of plants gathered locally. Special thanks to John Carlson for helpful discussions. This work was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada. S. Gottlob-McHugh was a holder of an NSERC scholarship and M. Lévesque and M. Olson were holders of NSERC summer studentship awards. We thank P. Brunon and J. Hélie for processing of material for figures.

References

Birnboim, H. C. and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7: 1513-1523.

Barciszewska, M. Z., Naleskowska, M., Mashova, T.D. 1987. Heterogeneity in the 3'-portion of the Papilonaceae 5S rRNAs. The primary structure of alfalfa 5S rRNA. Nucl. Acids Res. 15: 362.

Cronquist, A., Takhtajan, A.L., and Zimmerman W. 1966. On the higher taxa of embryobionta. Taxon 15: 129-134.

Ellis, T.H.N., Lee,D., Thomas, C.M., Simpson, P.R., Cleary, W.G., Newman, M.A., Burcham, K.W.G. 1988. 5S rRNA genes in Pisum: Sequence, long range and chromosomal organization. Mol. Gen. Genet. 214: 333-342.

Feinberg, A.P., Vogelstein, B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.

Gerlach, W.L. and Dyer, T.A. 1980. Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. Nucleic Acids Res. 8: 4851-4865.

Goldsbrough, P.B., Ellis, T.H.N., and Lomonosoff GP. 1982. Sequence variation and methylation of the flax 5S-rRNA genes. Nucleic Acids Res. 10: 4501-4514.

Gruenbaum Y., Naveh-Many T., Cedar H., Razin A. 1981. Sequence specificity of methylation in higher plant DNA. Nature 292: 860-862.

Hariharan, N., Reddy, P.S., and Padayatty, J.D. 1987. 5S-rRNA genes in rice embryos. Plant Mol. Biol. 9: 443-451.

Hattori, J., Gottlob-McHugh, S.G., and Johnson, D.A. 1987. The isolation of high molecular weight DNA from plants. Anal. Biochem. 165: 70-74.

Hemleben, V. and Werts, D. 1988. Sequence organization and putative regulatory elements in the 5S rRNA genes of two higher plants (Vigna radiata and Matthiola incana). Gene 62: 165-169.

Khvyrleva, Ts.D., Gazumyan, A.K., Anan'ev, E.V. 1988. Organization and primary nucleotide sequence of the 5S rRNA genes in barley (Hordeum vulgare). Genetika 24: 1830-1840.

Korn, L.J. 1982. Transcription of Xenopus 5S ribosomal RNA genes. Nature 295: 101-105.

Lawrence, C.J., Appels, R. 1986. Mapping the nucleolus organizer region, seed protein loci and isozyme loci on chromosome 1R in rye. Theor. Appl. Genet. 71: 742-749.

Long, E.O., Dawid, I.B. 1980. Repeated genes in eukaryotes. Ann. Rev. Biochem. 49: 727-764.

Maniatis, T., Fritsch, E.F., and Sambrook, J. 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, New York.

Margulis, L. and Schwartz, K.V. 1982. Five kingdoms: An illustrated guide to the phyla of life on earth. W.H. Freemand and Company, New York.

Mascia, P.N., Rubenstein, I., Philips, R.L., Wang, A.S., and Xiang, L.Z. 1981. Localization of the 5S rRNA genes and evidence for diversity in the 5S rDNA region of maize. Gene 15: 7-20.

McClelland, M. 1983. The frequency and distribution of methylatable DNA sequences in leguminous plant protein coding genes. J. Mol. Evol. 19: 346-354.

Nelson M., McClelland M. 1989. Effect of site-specific methylation on DNA modification methyltransferases and restriction endonucleases. Nucl. Acids Res. Suppl. 17: r389-r415.

Rafalski, J.A., Wiewiorowski, M., and Soll, D. 1982. Organization and nucleotide sequence of nuclear 5S rRNA genes in yellow lupin (Lupinus luteus). Nucleic Acids Res. 10: 7635-7642.

Reddy, P., Appels, R. 1989. A second locus for the 5S multigene family in Secale L.: sequence divergence in two lineages of the family. Genome 32: 456-467.

Rigby, P.W.J., Dieckman, M., Rhodes, C., 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.

Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74: 5463-5467.

Scoles, C.J., Gill, B.S., Xin, Z.-Y., Clarke, B.C., McIntyre, C.L., Chapman, C., Appels, R. 1988. Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of the Triticeae. Pl. Syst. Evol. 160: 105-122.

Selker, E.U., Morzycha-Wroblewska, E., Steven, J.N., and Metzenberg, R.L. 1986. An upstream signal is required for in vitro transcription of Neurospora 5S RNA genes. Molec. Gen. Genet. 205: 189-192.

Southern, E. 1979. Gel electrophoresis of restriction fragments. In: Wu R (ed.) Methods in Enzymology, Vol. 68. Academic Press, New York pp. 152-176.

Vandenberghe, A., Chen, M.-W., Dams E., de Baere, R., de Roeck, E., Huysmans, E., de Wachter, R. 1984. The corrected nucleotide sequences of 5 S RNAs from six angiosperms. FEBS 171: 17-23.

Whittaker, R.H. and Margulis, L. 1978. Protist classification and the kingdoms of organisms. Biosystems 10: 3-18.

Wolters, J., and Erdmann, V.A. 1988. Compilation of 5S rRNA and 5S rDNA gene sequences. Nucleic Acids Res. 16, supplement r1-r70.

Yannish-Perron, C., Viera, J., Messing,J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC vectors. Gene 33: 103-119.

Table 1. 5S rDNA repeat sizes in selected plant species

Name of Plant species	Size of 5S repeat (bp)
A. Magnoliophyta, Magnoliopsida (dicots):	
Acer rubrum (red maple)	338
Albizia lebbek (siris tree)	477
Alnus incana (speckled alder)	479
Betula papyrifera (paper birch)	479
Cassia fistula (golden shower)	#
Casuarina equisetifolia (Australian pine)	346
Ceratonia siliqua (carob)	330
Curcurbita sativus (cucumber)	341
Cynara scolymus (artichoke)	320,330,480
Datisca glomerata (durango root)	219
Elaeagnus angustifolia (russian olive)	330
Gleditsia triacanthos (honey locust)	278
Glycine max (L.) Merrill (soybean)	330
Gymnocladus dioicus (Kentucky coffeetree)	215
Myrica gale (bog myrtle)	362, 413
Myrica pensylvanica (bayberry)	404
Malus sp. (flowering crab apple)	#
Nicotiana tabacum (tobacco)	416, 617
Phaseolus vulgaris (broad bean)	389
Populus deltoides (poplar)	543, 634
Prosopis alba (mesquite)	293
Spinacia vulgaris (spinach)	336
Syringa vulgaris (lilac)	353
Trifolium repens (white clover)	330
B. Magnoliophyta, Liliopsida (monocots):	
Allium cepa (onion)	336
Dracaena margina (dragon-tree)	486
Hordeum vulgare (barley)	304, 442
Polygonatum multiflora (Solomon's seal)	486
Zea mays (corn)	323
C. Pinophyta, Pinopsida:	
Juniperus viginiana (eastern red cedar)	378
Picea glauca (white spruce)	#
D. Pinophyta, Cycadopsida:	
Cycas revoluta (cycad or sago palm)	1,572

# indicates periodicity, but the banding pattern was too complex to determine the repeat size.

FIG. 1 DNA sequence of the soybean 5S gene repeat contained within the clone 19-10-D. The postulated first nucleotide in the 5S RNA has been given the number +1 with the second repeat starting at +331. Underlined sequences are possible regulatory regions as described in the text. Restriction sites for enzymes used in this study are shown for one repeat as B=BamHl, H=HpaI, M=MspI/HpaII, S=Sau3A/MboI.
The probes of 5S gene structure used in this study were the following:
A= fragment defined by the two BamHI sites (repeat probe);
B= fragment defined by the HpaI site at +195 and the HpaII site at position +285(intercistronic probe).

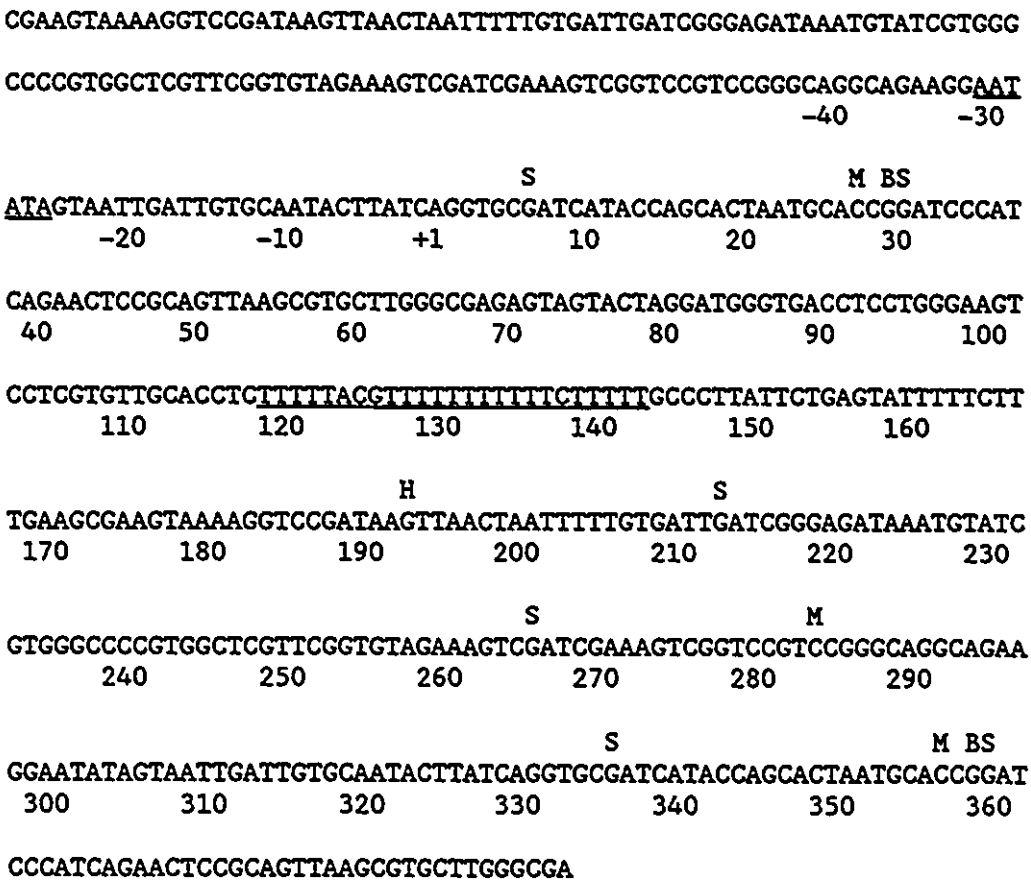

CGAAGTAAAAGGTCCGATAAGTTAACTAATTTTTGTGATTGATCGGGAGATAAATGTATCGTGGG
CCCCGTGGCTCGTTCGGTGTAGAAAGTCGATCGAAAGTCGGTCCGTCCGGGCAGGCAGAAGGAAT
-40
-30
S
M BS
ATAGTAATTGATTGTGCAATACTTATCAGGTGCGATCATACCAGCACTAATGCACCGGATCCCAT
-20
-10
+1
10
20
30
CAGAACTCCGCAGTTAAGCGTGCTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCTGGGAAGT
40
50
60
70
80
90
100
CCTCGTGTTGCACCTCTTTTTACGTTTTTTTTTTTCTTTTTGCCCTTATTCTGAGTATTTTTCTT
110
120
130
140
150
160
H
S
TGAAGCGAAGTAAAAGGTCCGATAAGTTAACTAATTTTTGTGATTGATCGGGAGATAAATGTATC
170
180
190
200
210
220
230
S
M
GTGGGCCCCGTGGCTCGTTCGGTGTAGAAAGTCGATCGAAAGTCGGTCCGTCCGGGCAGGCAGAA
240
250
260
270
280
290
S
M BS
GGAATATAGTAATTGATTGTGCAATACTTATCAGGTGCGATCATACCAGCACTAATGCACCGGAT
300
310
320
330
340
350
360
CCCATCAGAACTCCGCAGTTAAGCGTGCTTGGGCGA

FIG. 2 Sequence comparison of soybean 5S rDNA isolates. The sequence of eight different 5S sequences (a-h) from soybean are compared to the isolate 19-10-D. The differences between these isolates and 19-10-D are indicated below the 19-10-D sequence. Insertions of nucleotides are indicated by "t" and deletions are indicated by the asterisks (*).

```
19-10-D AGGTGCGATCATACCAGCACTAATGCACCGGATCCCATCAGAACTCCGCAGTTAAGCGTG  60
a       ------------------------------------------------------------  60
b       ----------------------------------T-------T-----------------  60
c       ------------------------------------------------------------  60
d       ------------------------------------------------------------  60
e       ---------------------------------------------T--------------  60
f       ----------------------------T-------------------------------  60
g       ------------------------------------------------------------  60
h       ------------------------------------------------------------  60
19-10-D CTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCTGGGAAGTCCTCGTGTTGCACCTCT 120
a       ------------------------------------------------------------ 120
b       ------------------------------------------------------------ 120
c       ------------------------------------------------------------ 120
d       ------------------------------------------------------------ 120
e       ------------------------------------------------------------ 120
f       ------------------------------------------------------------ 120
g       ----------------------------------T-----------T------------ 120
h       ------------------------------------------------------------ 120
19-10-D TTTTACGTTTTTTTTTTTCTTTTTGCCCTTATTCTGAGTATTTTTCTTTGAAGCGAAGTA 180
a       ------------------------------------------------------------ 180
b       ----------*********------------------------------------------ 171
c       -------tt-------------------------------------------------- 182
d       -------ttt------------------------------------------------- 183
e       -------*--------------------------------------------------- 179
f       -------**-------------------------------------------------- 178
g       -------*--------------------------------------------------- 179
h       -------ttt------------------------------------------------- 183
19-10-D AAAGGTCCGATAAGTTAACTAATTTTTGTGATTGATCGGGAGATAAATGTATCGTGGGCC 240
a       ------------------------------------------------------------ 240
b       ----T-----------A------------------------------------------- 231
c       ------------------------------------------------------------ 242
d       ------------------------------------------------------------ 243
e       ------------------------------------------------------------ 239
f       ------------------------------------------------------------ 238
g       ------------------------------------------------------------ 239
h       ------------------------------------------------------------ 243
19-10-D CCGTGGCTCGTTCGGTGTAGAAAGTCGATCGAAAGTCGGTCCGTCCGGGCAGGCAGAAGG 300
a       ------------------------------------------------------------ 300
b       ----------------------A------------------------------------- 291
c       ------------------------------------------------------------ 302
d       ------------------------------------------------------------ 303
e       ------------------------------------------------------------ 299
f       ------------------------T-------------------TA-------------- 298
g       ------------------------------------------------------------ 299
h       ------------------------------------------------------------ 303
19-10-D AATATAGTAATTGATTGTGCAATACTTATC                                330
a       ----------------------------                                 330
b       ----------------------------                                 321
c       ----------------------------                                 332
d       ----------------------------                                 333
e       -------------------------G--                                 329
f       ----------------------------                                 328
g       -------C--------------------                                 329
h       ----------------------------                                 330
```

FIG. 3 Analysis of the soybean 5S rDNA repeat structure by Southern hybridization. Soybean DNA (1μg) was digested with the appropriate enzyme(s), electrophoresed, transferred to BIOTRANS membrane and hybridized to the 330bp repeat probe defined in Figure 1. DNA was restricted with the following enzymes: (A). lane 1, BamHI; lane 2, BamHI plus HindIII; lane 3, BamHI plus PstI; lane 4, BamHI plus XhoI. (B). lane 1, Sau3A; lane 2, Msp1; lane 3, HpaII; lane 4, Sau3A plus BamH1; lane 5, Msp1 plus BamH1; lane 6, Sau3A plus Msp1. (C). lane 1, BamHI; lane 2, Sau3A; lane 3, MboI.

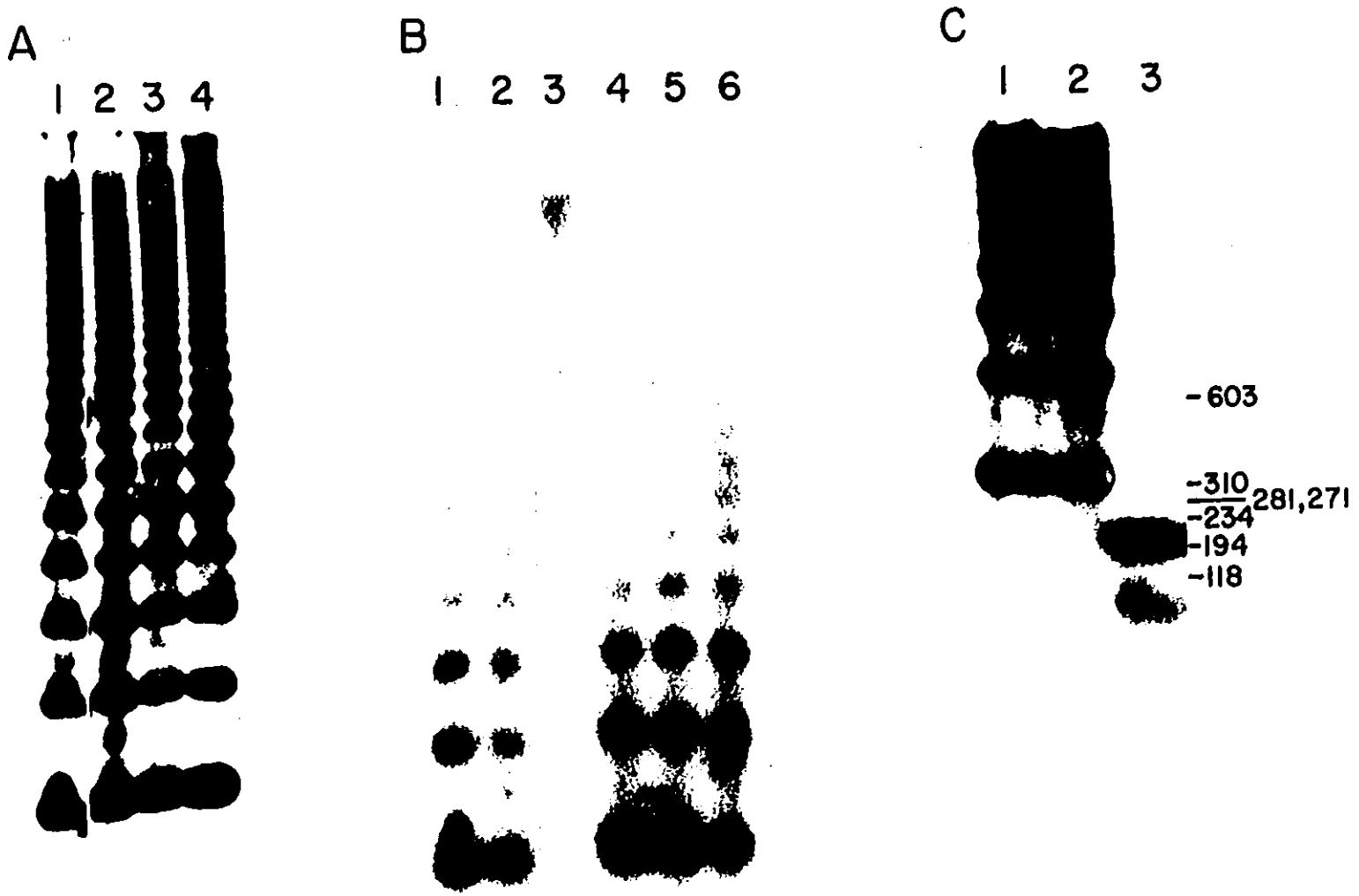
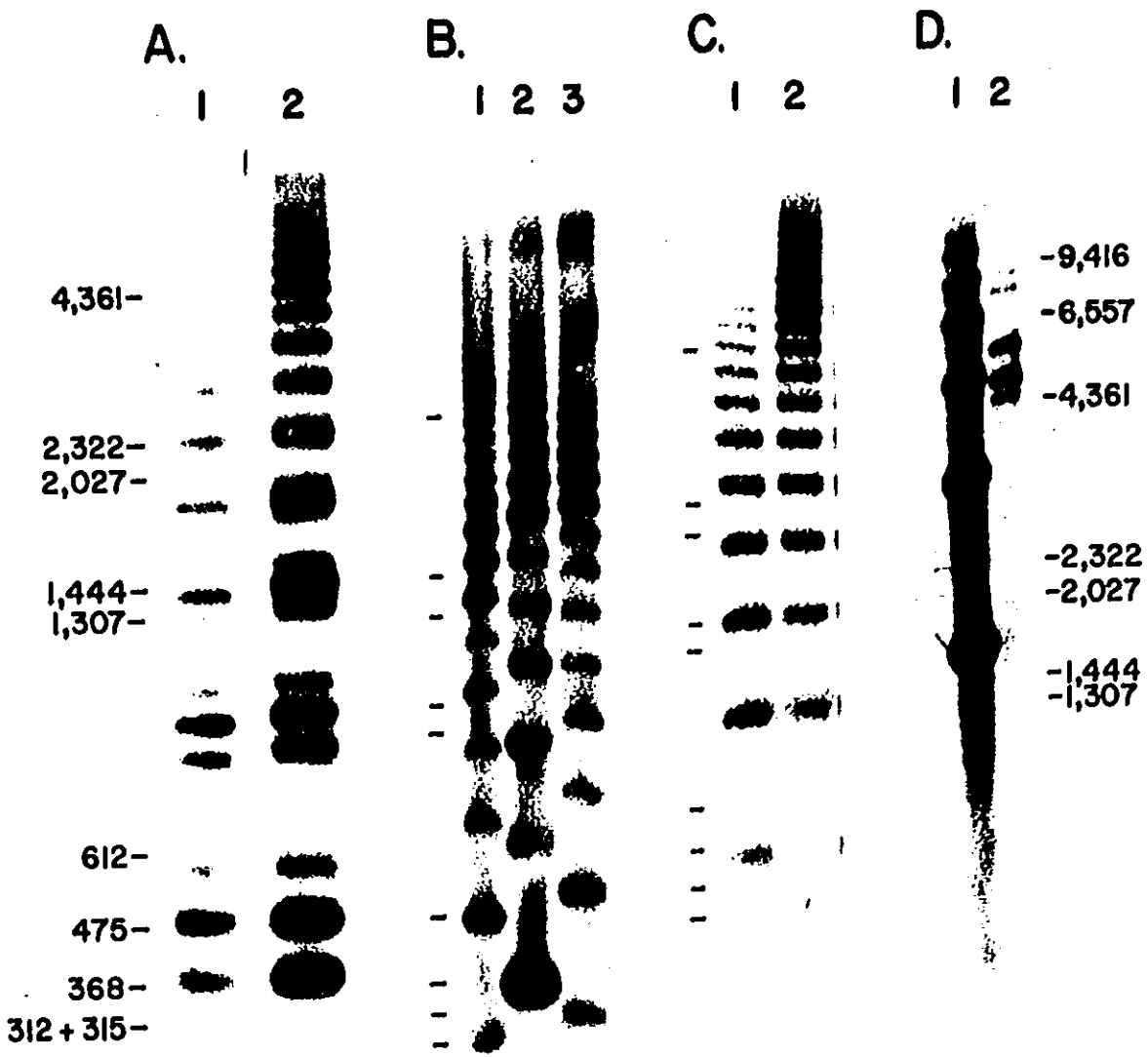

A
1 2 3 4
B
1 2 3 4 5 6
C
1 2 3
-603
-310
-234
281,271
-194
-118

FIG. 4 Detection of 5S rDNA sequences in different plant species by Southern hybridization. 1μg of plant DNA was digested with BamH1, electrophoresed, transferred to BIOTRANS membrane and hybridized to the 330bp repeat probe defined in Figure 1. (A). lane 1, Alnus rugosa; lane 2, Betula papyrifera. (b). lane 1, Prosopsis alba; lane 2, Myrica pensylvanica; lane 3, Elaeagnus angustifolia. (C). lane 1, Dracaena margina; lane 2, Polygonatum multiflora. (D). lane 1, Cycas revoluta; lane 2, Athyrium felix-femina.


A.
1
2
4,361-
2,322-
2,027-
1,444-
1,307-
612-
475-
368-
312+315-
B.
1 2 3
C.
1 2
D.
1 2
-9,416
-6,557
-4,361
-2,322
-2,027
-1,444
-1,307

APPENDIX II

During the course of the experiments described in this thesis and Appendix I, as well as other experiments conducted in the lab, we found it necessary to improve on plant DNA isolation procedures. This paper represents our efforts, and is a general procedure that can be applied to many different plant species. It was published in Analytical Biochemistry **165**, 70 - 74 (1987).

I worked with Dr. Hattori in modifying a number of steps involved in the isolation procedure. In addition, I supervised summer students who isolated DNA from several different plant species.

ANALYTICAL BIOCHEMISTRY 165, 70–74 (1987)

The Isolation of High-Molecular-Weight DNA from Plants

J. HATTORI, S. G. GOTTLOB-MCHUGH, AND D. A. JOHNSON

Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

Received October 17, 1986

A procedure to isolate high-molecular-weight DNA from plant materials has been devised. With this procedure, high-molecular-weight DNA suitable for Southern transfer experiments has been isolated from over 30 plant species including angiosperms (both dicots and monocots), a gymnosperm, members of other divisions, and two microorganisms.

KEY WORDS: plant DNA; purification.

The isolation of high-molecular-weight DNA (HMW[1] DNA) is required for many experiments in molecular biology such as Southern transfers and the construction of genomic libraries. The isolation of HMW DNA from animal tissues is, in general, relatively easy. In contrast, the isolation of HMW DNA from plant materials is considerably more difficult, due in part to the rigid cell wall and high carbohydrate content. In the past, several procedures for the isolation of HMW DNA have been reported (1–3). We have tried several of these procedures to isolate HMW DNA from plants but have obtained satisfactory results with only some plants.

We have developed a new procedure to isolate DNA from plants which is relatively easy and quick. Using this procedure, we have successfully isolated HMW DNA at reasonable yields from leaves, roots, and seeds of over 30 plant species including angiosperms (both dicots and monocots), a gymnosperm, members of other divisions, and microorganisms. DNA thus isolated has been cleaved by several restriction enzymes and used for Southern transfer experiments. In one case DNA has also been used to construct a genomic library. In addition, from the same preparation, RNA of sufficient quality for RNA blots can be isolated.

In this communication, we present a detailed description of the DNA isolation procedure and describe some of the results obtained using the DNA prepared according to this procedure.

[1] Abbreviations used: HMW, high-molecular-weight; PEG, polyethylene glycol; PVP, polyvinylpolypyrrolidone; SDS, sodium dodecyl sulfate.

MATERIALS AND METHODS

Chemicals. Chemicals used were polyethylene glycol (PEG) 8000 (Carbowax PEG 8000; Fisher); sodium dodecyl sulfate (SDS; BDH "specially pure" grade); polyvinylpolypyrrolidone (PVP; Sigma), insoluble in water, washed with 0.1 M HCl, neutralized with 50 mM Tris–HCl, pH 8.0, and filtered through Whatman No. 1 paper to obtain a slurry; solid phenol (BDH ACS grade) dissolved in 50 mM Tris–HCl, 10 mM EDTA, pH 8.0; Sevag, a 24:1 mixture of chloroform (BDH AnalaR) and iso-amyl alcohol (Fisher Certified); and phenol/Sevag, a one-to-one mixture of phenol and Sevag. Other chemicals used were ACS grade or equivalent.

Buffers. Buffers were sterilized by autoclaving. The extraction buffer contained 0.1 M Tris–HCl, 0.05 M EDTA, 1% SDS, pH 8.0; the TE buffer contained 10 mM Tris–HCl, 1

0003-2697/87 $3.00

mM EDTA, pH 8.0; and the KAc buffer contained 20 g/100 ml potassium acetate in 10 mM Tris–HCl, 5 mM EDTA, pH 8.0.

Procedure for DNA isolation. The procedure described below has been adjusted for 20 g of starting material. The procedure can be scaled up or down as required. Not all steps described below are required for all samples. Our experience is that for most of the samples only steps 1 through 12 are required. Comments on the procedure are included under Notes.

1. Grind the leaves, radicles, whole seeds, etc., in liquid nitrogen with the aid of a mortar and pestle. Add a small amount of extraction buffer and freeze the sample with liquid nitrogen. Repeat the step until a homogeneous paste is obtained (usually three times).

2. Add extraction buffer to a final volume of about 200 ml and transfer to two sterile 250-ml polypropylene centrifuge bottles (see Note 1).

3. Extract with 20–40 ml per bottle of phenol (see Note 2). Centrifuge at 7000 rpm, 4°C, for 10 min to separate the aqueous and phenol phases.

4. Back-extract the phenol phase with 20 ml extraction buffer.

5. Extract the combined aqueous phases sequentially with (a) 40 ml of phenol, (b) 40 ml of phenol/Sevag, and (c) a small volume (10–20 ml) of ether.

6. Chase the ether from the aqueous phase by placing the centrifuge bottles in a water bath at about 60°C for 10–15 min.

7. Mix with PVP (slurry from 2 g dry wt PVP) for a few minutes at room temperature and centrifuge at 7000 rpm for 10 min (see Note 3. Filter the supernatant through Whatman No. 1 paper and transfer the filtrate to 50-ml polyallomer centrifuge tubes.

8. Add NaCl (solid or 5 M stock solution) to a final concentration of 0.5 M and PEG 8000 to 10% (w/v). Leave the solution overnight at 4°C (see Note 4).

9. Collect the pellet by centrifugation at 8000 rpm, 10 min, 4°C. Drain the tubes thoroughly to remove PEG. Resuspend the pellet in 5–10 ml of TE buffer. Gentle shaking aids resuspension. Run 50 μl of the sample on a mini-gel to check quickly for the presence and size of DNA. If the DNA looks acceptable, proceed to the next step (see Note 5).

10. Add DNase-free RNase to 50 μg/ml and incubate for 30 min at 37°C.

11. Extract with 5 ml of phenol/Sevag. Centrifuge at 7000 rpm for 5 min. To the aqueous phase add ammonium acetate (7.5 M stock solution) to a final concentration of 2.5 M. Precipitate with 2 vol of ethanol overnight at −20°C.

12. Centrifuge at 10000 rpm for 30 min at 4°C. Wash the pellets twice with 70% (v/v) ethanol and dry the pellets at 37°C for not more than 10 min. (If the pellets are dried excessively, they will become difficult to resuspend.) Resuspend in a small volume (1–5 ml) of TE buffer and read A_{260}/A_{280}. Digest DNA with a restriction enzyme. Run on a mini-gel to check for the size and proper digestion by the enzyme (see Note 6).

13. Perform DEAE–Sephacel column chromatography as follows: Equilibrate the column with TE buffer. Apply the DNA to the column in TE buffer. Wash the column with 2 vol of TE buffer. Wash the column with 2 vol of TE buffer + 0.3 M NaCl. Elute the DNA with 3–5 vol of TE buffer + 0.6 M NaCl. Precipitate the DNA with 2 vol of ethanol overnight at −20°C (see Note 7). Collect the pellets by centrifugation at 10,000 rpm for 30 min.

14. Add 10 ml of KAc buffer to a Beckman SW41 ultracentrifuge tube. Layer 0.5–1 ml of DNA in TE buffer on top and centrifuge at 30,000 rpm, 15°C, for 16–18 h. Drain the tube carefully by inversion and then re-

move as much liquid as possible with a sterile cotton swab. Resuspend the pellet in 1 ml of TE with gentle shaking. Precipitate DNA with 2.5 M ammonium acetate, pH 7.5, and 2 vol of ethanol overnight at −20°C (see Note 8).

NOTES:

1. If the ground tissue is incubated with proteinase K prior to phenol extraction, the yield of DNA increases. However, the size of DNA will decrease. Therefore, the incubation with proteinase K is not recommended if HMW DNA is desired.

2. We use solid phenol as supplied by the manufacturer. We have not found it necessary to use redistilled phenol for this work.

3. PVP removes phenolic compounds often found in significant amounts in plant samples.

4. Precipitation with PEG separates nucleic acids from compounds which interfere with restriction enzyme digestion. Precipitation with ethanol at this stage, though possibly more efficient in precipitating the DNA, often does not remove these compounds. This will necessitate a much larger column at step 13.

5. RNA may be prepared from this step by precipitating nucleic acids with LiCl/urea solution following the procedure of Auffray and Rougeon (4).

6. If the DNA solution is viscous due to polysaccharides, centrifuge the solution in a microcentrifuge for 10–15 min. This will remove high-molecular-weight polysaccharides and colored material.

7. DEAE–Sephacel column chromatography is performed according to Maniatis *et al.* (5). This step removes compounds which interfere with restriction enzyme digestion and which have not been removed by the PEG precipitation step (step 8). We have found this step necessary with only a few species. Batch-wise elution, though somewhat more time-consuming than the standard column elution, is more efficient in eluting HMW DNA from DEAE–Sephacel beads.

8. Modified from the procedure described in Maniatis *et al.* (5). This step is used if it is necessary to remove smaller-sized DNA caused by degradation. Centrifugation time may be changed depending on the size of the DNA desired.

Other procedures. Southern transfers (6) were performed as described in (5). Hybridizations were carried out using the probe 2-2-G, a soybean large rRNA clone isolated from a soybean nodule cDNA library constructed by insertion of double-stranded cDNA into the *Pst*1 site of pBR322 via dG–dC tailing (unpublished results).

RESULTS AND DISCUSSION

Using the procedure described under Materials and Methods, we have isolated HMW DNA suitable for Southern transfer experiments from the following plants (the number after the name indicates the plant part from which DNA has been isolated: 1, leaves; 2, seeds; 3, roots; 4, whole plants; 5, catkins).

(a) Angiosperms (dicots)—*Acacia sp.* (acacia), 1, 2; *Acer sp.* (maple), 1; *Albizia lebbek* (siris tree), 1; *Alnus rugosa* (speckled alder), 1, 2; *Betula papyrifera* (paper birch), 1, 2; *Cassia fistula* (golden shower), 2; *Casuarina equisetifolia* (Australian pine), 2; *Ceratonia siliqua* (carob), 1, 4; *Cynara scolymus* (artichoke), 1; *Datisca glomerata* (durango root), 2; *Elaeagnus angustifolia* (Russian olive), 2; *Gleditsia triacanthos* (honey locust), 1; *Glycine max* (soybean), 1, 2, 3; *Gymnocladus dioicus* (Kentucky coffee tree), 1; *Hamamelis sp.* (witch hazel), 1; *Myrica gale* (bog myrtle), 5; *Myrica pensylvanica* (bayberry), 2; *Malus sp.* (flowering crab), 1; *Nicotinia tabacum* (tobacco), 1; *Phaseolus vulgaris* 1, 2; *Populus sp.* (poplar), 1, 2; *Prosopis alba* (mesquite), 2; *Spinacia sp.*

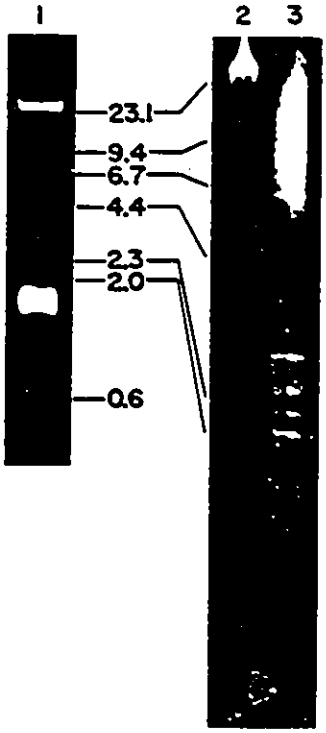


FIG. 1. Carob (*C. siliqua*) DNA electrophoresed on a 1% agarose gel. Lane 1, DNA after step 9 (see Materials and Methods). Lane 2, DNA after step 12. Lane 3, step 12 DNA digested with *Eco*RI. Size markers (kb) were λ DNA digested with *Hin*dIII.

(spinach), 1; *Syringa sp.* (lilac), 1; *Trifolium repens* (white clover), 2.

(b) Angiosperms (monocots)—*Allium cepa* (onion), 2; *Dracaena margina* (dragon tree), 1; *Polygonatum sp.* (Solomon's seal), 1; *Zea mays* (maize), 2, 3.

(c) Gymnosperm—*Juniperus virginiana* (eastern red cedar), 1.

(d) Other divisions—*Athyrium felix-femina* (lady fern), 1; *Bryum sp.* (moss), 4; *Cycas sp.* (cycad or sago palm), 1; *Equisetum sp.* (horsetail), 4; *Psilotum nudum* (whisk fern), 4.

(e) Microorganisms—*Frankia, Chlamydomonas segnis.*

Of the 36 plant species from which we have attempted to isolate DNA, we have been unsuccessful with only one species (*Ginkgo biloba*). Different DNA contents of the plant samples will change the yields of DNA from sample to sample. Actual yields obtained were .10–130 μg per gram starting material. The average yield was about 50 μg per gram starting material.

As can be seen in Fig. 1, the DNA isolated from young carob (entire plant) is mostly high molecular weight. After RNase treatment and reextraction with phenol, the DNA can be used for restriction analysis. In most cases the DNA at this stage can be digested satisfactorily with the restriction enzymes *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Sma*I, and *Xho*I. Only with a few species was a further purification by DEAE–Sephacel chromatography necessary. Lower molecular-weight bands noticeable in Fig. 1 (lane 1) are RNA and are removed by RNase treatment (compare to lane 2). This material has been used directly without further purification for RNA transfer experiments.

The quality of the DNA preparations was assessed by Southern transfer experiments. Figure 2 shows the results of an experiment in which various plant DNAs were digested with *Bam*HI, electrophoresed, transferred to

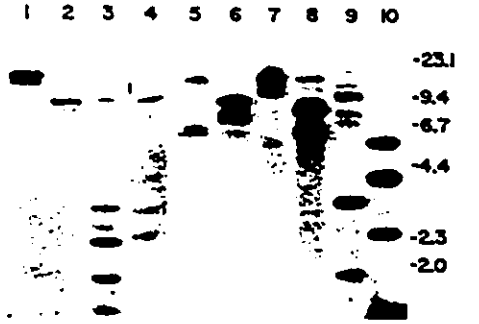


FIG. 2. Southern blots of plant DNA hybridized with ^{32}P-labeled soybean 25 S rRNA probe 2-2-G. One microgram each of the following DNAs were probed: Lane 1, dragon tree; lane 2, Solomon's seal; lane 3, artichoke; lane 4, fern; lane 5, lilac; lane 6, tobacco; lane 7, maple; lane 8, spinach; lane 9, apple; lane 10, soybean. Size markers (kb) were λ DNA digested with *Hin*dIII.

Biodyne membrane, and hybridized with the ^{32}P-labeled cDNA of the soybean 25 S rRNA. The sizes of the bands observed for soybean were in agreement with those published by Friedrich *et al.* (7). Specific hybridization was also observed for plant DNAs from various species of angiosperms (dicots and monocots) and a fern, indicating that the DNA preparations are of sufficient quality for Southern transfer experiments. These results also demonstrated that the 25 S rRNA genes are highly conserved.

In addition to the experiments reported here we have used soybean leaf DNA to construct a genomic library in the vector EMBL4 (Johnson, unpublished results). Thus we have demonstrated that the procedure described in this paper is useful for the isolation of reasonable quantities of HMW DNA from a wide range of plant species, that the DNA is suitable for restriction enzyme analysis, Southern transfers, and cloning, and that in some preparations it is possible to recover RNA which can be used for RNA transfers.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada and a contract from the Canadian Forestry Service. The authors thank M. Lévesque and M. Olson (NSERC summer students) for the isolation of many of the DNAs. The assistance of all those who provided the plant material to us is gratefully acknowledged.

REFERENCES

1. Locy, R. D., and Hall, R. H. (1978) *Canad. J. Bot.* 56, 1453–1457.
2. Blin, N., and Stafford, D. W. (1976) *Nucleic Acids Res.* 3, 2303–2308.
3. Goldberg, R. B. (1978) *Biochem. Genet.* 16, 45–68.
4. Auffray, C., and Rougeon, F. (1980) *Eur. J. Biochem.* 197, 303–314.
5. Maniatis, T., Fritsh, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
6. Southern, E. M. (1975) *J. Mol. Biol.* 98, 503–517.
7. Friedrich, H. H., Hemleben, V., and Key, J. L. (1979) *Plant Syst. Evol. Suppl.* 2, 73–88.