ANALYSIS OF CARBOHYDRATE-RELATED GENE EXPRESSION

IN *RHIZOBIUM LEGUMINOSARUM* BIOVAR *VICIAE*

DURING SYMBIOTIC PEA NODULE DEVELOPMENT

by

HYONG SOOK KIM

(Under the Direction of Russell W. Carlson)

ABSTRACT

The carbohydrate-related gene expression of *Rhizobium leguminosarum* biovar *viciae* strain 3841, a gram-negative, alpha-proteobacteria forming symbiotic indeterminate nodules on peas, was analyzed with publicly available whole-genome microarray data set at subsequent time points during rhizobial pea root invasion and bacteriod development. A statistical analysis was performed on the expression ratios of free-living culture and bacteroid state to study a broad sample of rhizobial carbohydrate-related ortholog genes compiled from various databases (NCBI COG, KEGG, CAZy, and GeneDB) and reported literatures. (i) We examined known carbohydrate classes in *R. leguminosarum* and validated previous findings, e.g. the up-regulation of Nod factor genes and down-regulation of EPS genes. (ii) Unclassified carbohydrate-related genes that were either up- or down-regulated were studied. (iii) A number of either expressed or suppressed genes were selected as candidates for future potential genetic and biochemical analysis of their roles in carbohydrate biosynthesis and symbiotic development.

INDEX WORDS: *Rhizobium leguminosarum*, microarrya, carbohydrates, symbiosis, genome, transcriptome, bioinformatics

ANALYSIS OF CARBOHYDRATE-RELATED GENE EXPRESSION IN *RHIZOBIUM LEGUMINOSARUM* BIOVAR *VICIAE* DURING SYMBIOTIC PEA NODULE DEVELOPMENT

by

HYONG SOOK KIM

B.S., The University of Georgia, 2008

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2010

ANALYSIS OF CARBOHYDRATE-RELATED GENE EXPRESSION IN *RHIZOBIUM LEGUMINOSARUM* BIOVAR *VICIAE* DURING SYMBIOTIC PEA NODULE DEVELOPMENT

by

HYONG SOOK KIM

Major Professor:	Russell W. Carlson
Committee:	Elmar L. Kannenberg
	William S. York

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2010

DEDICATION

I would like to dedicate everyone who has influenced me to be who I am now.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my major professor, Dr. Carlson, and the professors in my committee, Dr. Kannenberg and Dr. York, for their time and effort to make this study possible. I deeply appreciate advices and suggestions from Dr. Carlson's lab members, and I cannot forget the wonderful trip we had during the North American Nitrogen Fixation Conference at Columbia, MO. I am also thankful for all students and faculty members I met in the Institute of Bioinformatics for their friendship and rapport. I must acknowledge Dr. Jaxk Reeves, Mr. Huanyao Gao, and Ms. Guannan Wang from the Statistics Consulting Center for their help with statistical analysis of the microarray data and even extra help and attention after the class. Thanks to all my family members for their unending prayers and support until the end of my study. Above all, thanks to our God.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

BACKGROUND AND INTRODUCTION

1.1 *Rhizobium*-legume Symbiosis

Rhizobium bacteria are gram-negative, alpha-proteobacteria that can live in symbiosis with legume plants (60). Rhizobia have fascinated researchers now for more than 100 years ever since symbiotic nitrogen fixation was discovered by Hellriegel and Wilfarth in 1888 (24). The Rhizobia live and dwell in the soil and as nitrogen-fixing endoysmbionts in legume host plants. During symbiosis with legumes, the host plants form a new organ, the root nodule, while the rhizobia develop into specialized symbiotic forms, the bacteroids. Bacteroids are normally found within nodule cells in organell-like compartments, called symbiosomes. They fix atmospheric nitrogen to ammonia and supply it to the host plant as a nitrogen source. Successful survival of rhizobia depends, therefore, on being capable of invading and adapting to very different legume plant and soil environments.

In case of the symbiosis of *Rhizobium leguminosarum* biovar *viciae* with pea plants, symbiosis starts in the pea root rhizosphere with the attachment of rhizobia to a newly emerging root hair and the formation of a biofilm (20, 22). The root hair deforms and entraps the rhizobia. The root hair tip growth is redirected placing cell wall materials in apposition to the rhizobial attachment point. The root hair and the bacterial microcolony interact, and a cell wall invagination is formed that develops into an inwardly growing tunnel, the infection thread. The infection thread is filled with a column of rhizobia embedded in polysaccharide matrix that is bordered by a cell wall layer. The growing (eventually also branching) infection thread leads the

bacteria to a newly formed meristem in the pea root cortex. When an infection thread is making contact with a meristematic cell, the apposition of cell wall material is in these so called infection droplets is thinned, and rhizobia can make contact with the cytoplasmic membrane of the cell. They are then endocytosed and immature symbiosomes are formed that multiply and mature into functional sybiosomes occupied by the endosymbiotic, polymorphic, nitrogen-fixing *Rhizobium* bacteroids. Pea symbiosomes contain only one bacteroid that is surrounded by a specialized plant-derived membrane, the peribacteroid membrane (22).

The *Rhizobium* pea symbiosis is an important plant bacterium interaction model that can teach us a great deal about the molecules that are involved in the interactions between the eukaryotic pea host and its prokaryotic *Rhizobium* symbiont. While in the past two decades, the focus of research has been largely on the initial steps of the symbiotic interactions that underlay the initiation of nodule development, the interest has recently shifted towards later stages of nodule development and what happens to bacteria once they occupy the plant environment. One approach to address questions relating to these later stages is to interrupt and investigate nodule development at different time points after plant inoculation. Bacterial root invasion and nodule development are, to a degree, synchronized through the simultaneous inoculation of the plant host roots with a massive number of rhizobia (i.e. 10^8 to 10^9 bacteria per plant); hence, nodule development can be studied through investigating nodules at different stages of symbiosis. For example, in young nodules at 7 days post inoculation (d.p.i.) rhizobia are largely confined to infection threads, while in fully developed nodules at 28 d.p.i. the vast majority of rhizobia are mature, nitrogen-fixing bacteroids. In between these two time points, the ratio of invading rhizobia to mature bateroids varies.

1.2 Carbohydrates in *Rhizobium leguminosarum*

Glycans and glycoconjugates are primarily located in the cell wall, in membranes and on the cell surface of bacteria and have important adaptive roles for cell survival under various environmental stresses, such as alterations in the surrounding osmolarity, temperature, depletion of oxygen, lack of nutrients, or desiccation conditions (37). In addition to this protective role of bacterial glycans under the various environmental stresses, carbohydrates found in many pathogenic microbes that invade host tissues or even cells serve also as virulence factors in their respective hosts, e.g. mimic host carbohydrate structures to protect themselves from host immune response during invasion (50). As stated above, *Rhizobium* bacteria also invade plant tissues and cells of their host legume plants; but in contrast to pathogenic bacteria which cause detrimental consequences to their hosts, the symbiotic *Rhizobium* bacteria form an intimate relationship that is mutually beneficial and stable for weeks. The rhizobial carbohydrates, too, are 'virulence factors', but they are expected to be involved during the symbiotic invasion in modulating plant development and in the suppression of plant defense reactions (19, 21). The rhizobial cell surface is expected to be important for a direct cell-to-cell attachment or recognition processes and diffusible carbohydrate molecules could be responsible for indirect or distant signaling between rhizobia and their legume hosts (31).

The known cell carbohydrate of *Rhizobium leguminosarum* is comprised of number of different carbohydrate classes: lipochitinoligosaccharide (LCO), exopolysaccharide (EPS), lipopolysaccharide (LPS), capsular polysaccharide (CPS), cyclic glucans, and peptidoglycan (PG). The LCOs, or Nod factors, are plant hormone-like lipooligosaccharides that are comprised of an acylated chitin oligomeric backbone with various functional groups and serve as signaling molecules at an early stage of the symbiosis process with plants; they are formed by rhizobia in

response to flavanoid signal molecules released by the legume host and initiate root hair deformation and meristem induction in the host root cortex (22, 61).

A well studied rhizobial surface carbohydrate is EPS. It is an acidic heteropolymer composed of glucose, glucuronic acid, and galactose with pyruvyl and acetyl substituents (37, 51). It thought to have a protective role under various environmental conditions such as osmotic, oxidative and ionic stress, and desiccation. In *Rhizobium*-legume symbiosis, it has been suggested to play a role in biofilm formation and in suppression of the host plant immune response (37).

Similar to LPSs from other gram-negative bacteria, the rhizobial LPS is the major component of the outer leaflet of the cell outer membrane and consists of three general structural regions. It is anchored in the outer membrane via its lipid A (LA) which consists of an acylated disaccharide backbone. In *Rhizobium leguminosarum*, attached to the LA via a unique glycosyl residue, Kdo (3-deoxy-D-manno-2-octulosonic acid), is a core oligosaccharide (COS), and attached to the COS via another Kdo residue, is the O-chain polysaccharide (OPS) (7). This rhizobial LPS structure is unique for several reasons: it completely lacks phosphate in its lipid A backbone; instead, the lipid A carries a galacturonosyl residue on the distal backbone glucosamine and can have its proximal glucosaminosyl residue oxidized to 2-aminogluconic acid; in addition, the LA contains a unique acyl chain attached to its carbohydrate backbone, 27-hydroxyoctacosanoic acid, 27-$OHC_{28:0}$. This unique very long chain hydroxyl fatty acid (VLCFA) is likely to span both halves of the outer membrane bilayer; it is considered important for the soil and intracellular, symbiotic life style of *Rhizobium* bacteria, but may also play a similar role in a range of pathogenic, intracellular living bacteria such as *Brucella* and *Bartonella* (7-8). Taken together, in *Rhizobium* the LPS is likely to be important in the stress adaptation of

bacterial cells to soil and plant environments and an intact LPS proved to be essential for the normal symbiotic development to proceed (7). However, much remains to be learned about LPS function. Recently, it has been implicated also in the suppression of host defense responses (6, 35, 37, 62-63). In addition, structural modifications in the LPS during symbiosis were observed (18, 35, 43, 56) that are not yet explained but could contribute to an observed increase in overall cell and LPS hydrophobicity in symbiotic nodule bacteria (12, 33). The LPS functions during symbiosis are still largely unknown and are currently an area of intense investigation.

Rhizobial cyclic glucans, comprised of β-(1,2) linked glucose, are primarily located in the periplasm; they are involved in the cellular osmoregulation and storage (49). It has been reported that they also act as a suppressors of the host immune system (4, 37). Cyclic glucan synthesis possibly occurs throughout the nodule formation process and its importance for *Rhizobium*-legume symbiosis was noted in several studies (4, 10, 37). Another known class of polysaccharides in *Rhizobium* is capsular polysaccharide (CPS or KPS). It is anchored to cell membrane via a lipid anchor. Its structure is not common among the strains of the same species (35, 42). The CPS is thought to protect the bacterium against desiccation and bacteriophages; different types of CPSs are produced depending on growth conditions *in vitro* (19, 42). It is necessary for effective nodulation and infection thread development and has a possible role in the adaptation to the symbiotic host (2, 37).

Other known carbohydrates of *R. leguminosarum* include a gel-forming polysaccharide (GPS) which is a neutral polysaccharide secreted under stationary growth condition (58, 68), glucomannan carbohydrates thought to be involved in *Rhizobium* attachment to plant cells (16, 40, 58), cellulose fibrils responsible for irreversible attachment and aggregation of rhizobia around root hairs to form a biofilm (37, 41), and trehalose which is a type of carbohydrate

secreted by rhizobia as an osmoprotective molecule under osmotic and desiccation stress (9). Finally, the peptidoglycan (PG) sacculus, also a carbohydrate, forms a layer between the inner and outer membrane of the bacteria cell wall. It is composed of repeating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) disaccharides to form long strands which are connected through short amino acid bridges. The PG, placed between the outer and inner membrane, envelopes the rhizobial cell and provides structural integrity to the bacteria, e.g. prevents possible rupture of the inner membrane under extreme environment conditions such as osmotic changes and hypotonic conditions (45). The function of the PG during symbiosis has not yet been reported.

1.3 Purpose of the Study

With regard to carbohydrate formation during symbiotic development, all of the above carbohydrate classes have been studied only in isolated settings, i.e. mostly only an individual class of carbohydrate in a singular aspect of nodule development. However, the recent elucidation of whole rhizobial genome sequences combined with modern transcription profiling analysis allows for the analysis of carbohydrate-related gene expression at a number of different developmental stages, e.g. it allows for the study of carbohydrate-related gene expression at various time points of symbiotic nodule development. In this study, we used a recently sequenced *Rhizobium leguminosarum* biovar *viciae* strain 3841, a spontaneous streptomycin-resistant mutant of strain isolate 300 (3, 23, 27, 60), to analyze carbohydrate-related gene expression during rhizobial pea root invasion and bacteroid development at 7, 15, 21, and 28 d.p.i. The starting point of our analysis was the publicly available whole genome microarray data set of *R. leguminosarum* strain 3841 gene expression during symbiotic pea nodule development

(36, 67). Carbohydrate-related gene orthologs were compiled for *R. leguminosarum* 3841 from various databases (NCBI COG, KEGG, CAZy, and GeneDB) and literature reports. We analyzed carbohydrate-related gene expression using a statistical approach to identify significant changes in the expression levels of these carbohydrate-related genes during *Rhizobium* infection and bacteroid development. Then, the gene expression levels of different classes of known carbohydrates as well as genes involved in fatty acid metabolism as a function of the rhizobial infection process were determined. A number of either expressed or suppressed genes of *R. leguminosarum* glycome were identified for potential future genetic and biochemical analysis of their roles during pea nodule and bacteroid development.

CHAPTER 2

MATERIAL AND METHODS

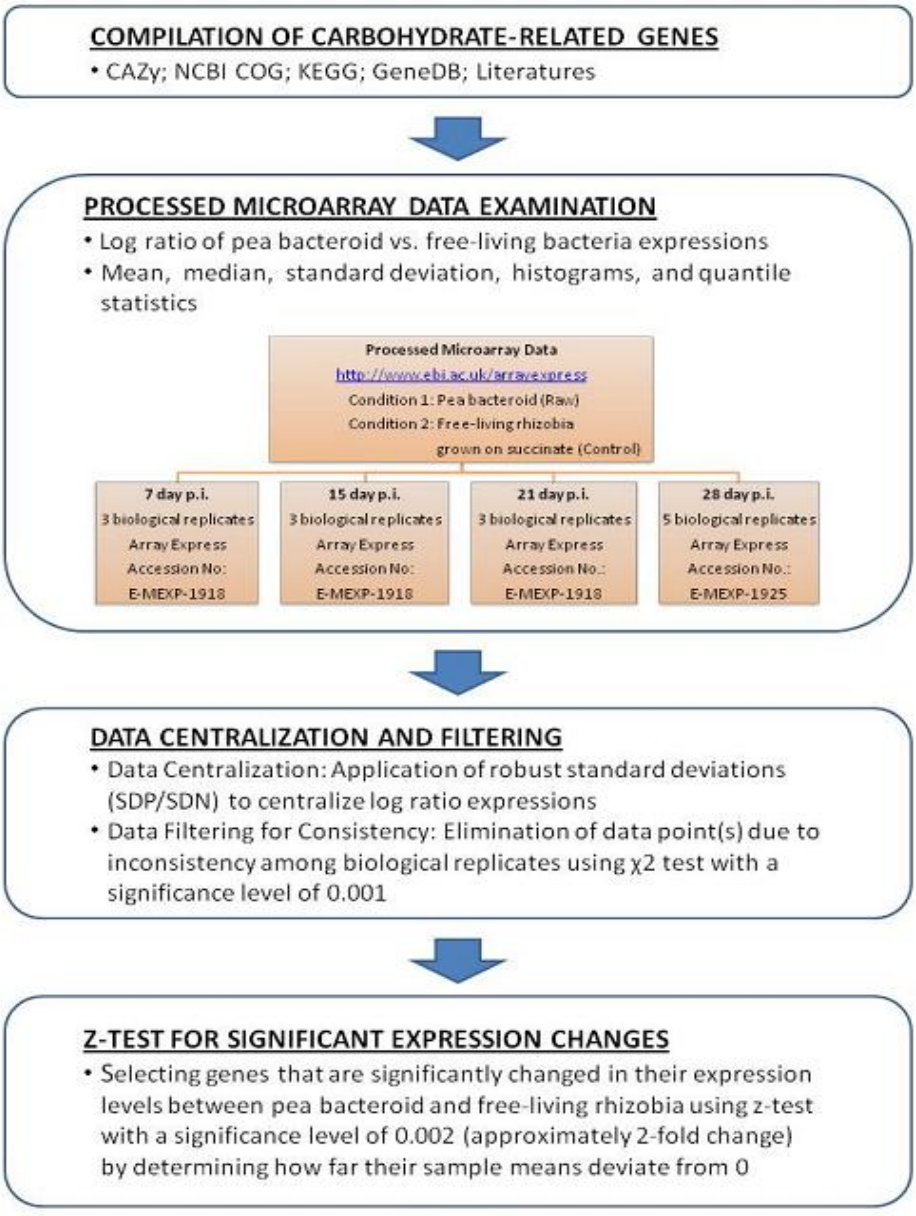


Figure 1. Schematic Representation of Experimental Procedure.

2.1 Microarray Data Source

The whole-genome microarray data of *R. leguminosarum* biovar *viciae* strain 3841 (36) deposited at Array Express (http://www.ebi.ac.uk/arrayexpress) with accession numbers E-MEXP-1918 (7, 15, 21 days post inoculation (d.p.i.); 3 replicates on each day) and E-MEXP-1925 (28 d.p.i; 5 replicates) were used in this study and are shown in Table 1. See Figure 1 for a flow chart of the analysis and (36) for details on the microarray experiment.

Table 1. General Summary of the Microarray Data Experiment of *R. leguminosarum* Deposited on http://www.ebi.ac.uk/arrayexpress (36)

Array Express Accession no.	Condition 1[a,b] (Raw)		Condition 2[a] (Control)	
	Sample[e]	Cell type	Sample[c]	Cell type
E-MEXP-1918	7 day (SA)	Pea bacteroid	Succinate	Free living
E-MEXP-1918	15 day (SA)	Pea bacteroid	Succinate	Free living
E-MEXP-1918	21 day (SA)	Pea bacteroid	Succinate	Free living
E-MEXP-1925	28 day (SA)[d]	Pea bacteroid	Succinate	Free living

[a]SA, sense amplification of RNA was performed. Three biological replicates were performed.
[b]Seeds inoculated and at 7-, 15-, and 21-days post inoculation (d.p.i.) bacteroid samples were analyzed.
[c]For growth of free-living cells, the carbon source added to the medium is shown.
[d]Five biological replicates were performed.

2.2 Compilation of *Rhizobium* Carbohydrate- and Fatty Acid-related Gene Sample

Carbohydrate-related genes from the *R. leguminosarum* biovar *viciae* 3841 genome sequence were compiled using various bioinformatics database sources including Carbohydrate-Active Enzymes database CAZy (http://www.cazy.org/) (5), Kyoto Encyclopedia of Genes and Genomes KEGG (http://www.genome.jp/kegg/) (28-29), National Center for Biotechnology Information: Clusters of Orthologous Groups of proteins NCBI COGs (http://www.ncbi.nih.gov/COG) (59), GeneDB (http://www.genedb.org/Homepage) (25) and a number of literature reports (1, 17, 26, 30, 38, 46-48, 55, 57, 65-66, 69). Gene annotations and

Riley classifications from the original *R. leguminosarum* genome sequencing paper (67) were used for gene information.

2.3 Statistical Analysis of Microarray Data

A statistical computing and graphics software R (http://www.r-project.org/) was used to perform the following statistical analyses on the processed microarray data from the Array Express website (http://www.ebi.ac.uk/arrayexpress). The microarray data were optimized by the following two steps: distribution normalization and consistency improvement. After microarray data optimization, classification of the normalized gene expression ratios of pea bacteroids versus free-living rhizobial culture (logarithm ratios) was done via a z-test.

For the distribution normalization step, the distribution of logarithm bacteroid/bacteria gene expression ratios for each d.p.i. was examined for centrality and normality. Since most genes likely are unchanged in their bacteroidal and bacterial expression levels, one would expect typical ratios to be near 1, so that typical log-ratios would be near zero. In fact, Q-Q plots and quantile histograms for replicates for each d.p.i. indicated that the population means for replicates were reasonably close to $\mu_0 = 0$, but the distributions of the data were clearly not normal and skewed slightly to one side. To improve the distributions of the microarray data for each d.p.i., the logarithm ratios of the replicates were centralized using its median, and the skewness was adjusted by different robust standard deviation estimates, SDP or SDN, where SDP and SDN were the robust estimators of standard deviation of the replicates for each d.p.i. for positive and negative residuals, respectively. SDP and SDN were calculated by taking a difference between the 95^{th} quantile for SDP, or 5^{th} quantile for SDN, and the median of the logarithm ratios and dividing by the standard z-value (0.95) = 1.645 for SDP, or the z-value

(0.05) = -1.645 for SDN, as shown in equation (1) and (2). The logarithm ratios were then centralized as shown in equation (3), where X is a logarithmic ratio of the replicates for each d.p.i., and $X_{[0.50]}$ is a 50^{th} quantile or a median of the logarithm ratios for each d.p.i.

$$SDP = \frac{X_{[0.95]} - X_{[0.50]}}{1.645} \tag{1}$$

$$SDN = \frac{X_{[0.05]} - X_{[0.50]}}{-1.645} \tag{2}$$

$$X_{centralized} = \begin{cases} (X - X_{[0.50]}) / SDP & if \quad X > X_{[0.50]} \\ (X - X_{[0.50]}) / SDN & if \quad X < X_{[0.50]} \end{cases} \tag{3}$$

For the consistency improvement step, the consistency of the data was examined by performing a χ^2 test for each gene for each d.p.i. under an assumption that the logarithm ratio of each biological replicate for a specific gene should be independently identically distributed and the replicates' logarithm ratios should be consistent for each gene for each d.p.i. The sample standard deviation of a gene (denoted as s in equation (4)) for a d.p.i. was calculated using all replicates (this was n=3 replicates for genes observed 7, 14, and 21 d.p.i., and n=5 for those observed 28 d.p.i.) The population standard deviation (denoted as σ in equation (4)) was defined as the median of all sample standard deviations calculated over all genes for that d.p.i. The χ^2 statistics for each gene for each day were calculated as shown in equation (4), where n is the number of biological replicates, and the χ^2 statistics were assumed to follow a χ^2 (n) distribution if there were no outliers among the n standardized log-ratios for that gene. The χ^2 test at a significance level of 0.001 was performed for each gene for each d.p.i. under an assumption that the logarithm ratios were independently distributed as N (0, σ^2) The hypothesis for the χ^2 test are:

H_0: the sample standard deviation, s, of a gene for a particular d.p.i. is not significantly different from the population standard deviation, σ, for the d.p.i.;

H_a: the sample standard deviation, *s*, of a gene for a particular d.p.i. is significantly different from the population standard deviation, σ, for the d.p.i.

$$\chi^2 = (n-1)\left(\frac{s}{\sigma}\right)^2 \qquad (4)$$

When a gene's chi-square statistic led to rejection of the null hypothesis, the logarithm ratio of a replicate that was most deviant from the median of all replicates' logarithm ratio was eliminated. Then, the sample and population standard deviations were re-calculated, and this step was repeated until no gene rejects the null hypothesis or no more than one replicate for 7, 15, 21 d.p.i. or no more than two replicates for 28 d.p.i. was/were eliminated. The main point of this consistency check was to ensure that any gene which will be judged to be significantly expressed (or repressed) is not unduly influenced by one very high or very low ratio reading.

For the classification of gene expression levels, the optimized logarithm ratios for genes on a particular d.p.i. were examined by a standard z-test. The gene expression of optimized logarithm ratios were interpreted in three ways depending on their values: (i) the logarithm ratio was declared to be insignificantly different from zero if the gene expression levels were approximately the same for pea bacteroids and free-living cell culture, (ii) the logarithm ratio was declared 'positive' when the gene expression of pea bacteroids was significantly higher than the gene expression of free-living cell culture, and (iii) the logarithm ratio was declared 'negative' when the gene expression of fee-living cell culture was significantly higher than the gene expression of pea bacteroids. The thresholds for the z-test with $\mu_0 = 0$ at a significance level of 0.002 (double tail) on each gene for a particular d.p.i. were implemented using equation (5). Hypothesis of the z-test using equation (5), where *n* is the number of biological replicates for a d.p.i., are:

H_0: a sample mean, $\bar{x}$, is not significantly different from $\mu_0=0$;

H_a: a sample mean, $\bar{x}$, is significantly different from $\mu_0=0$.

$$z = \frac{\bar{x} - \mu_o}{\sigma / \sqrt{n}} \qquad (5)$$

According to the z-statistic calculated for each gene for each d.p.i., the gene's expression level was classified. If the z-score was greater than 3.09 (about 2-fold increase in the gene expression of pea bacteroids over the gene expression of free-living rhizobial culture), then that gene was judged to be significantly expressed, or up-regulated. If the z-score was less than -3.09 (about 1/2-fold decrease in the gene expression of pea bacteriods over the gene expression of free-living rhizobial culture), then that that gene was judged to be significantly repressed, or down-regulated. If the z-score was between 3.09 and -3.09, then that gene was judged to be not significantly expressed nor suppressed, but constitutively expressed between the gene expressions of pea bacteriods and free-living rhizobia culture.

CHAPTER 3

RESULTS

3.1 Overview of the Broad Sample of Compiled *Rhizobium* Carbohydrate-Related Genes (CRGs)

To observe gene expression changes of the carbohydrate- and fatty acid-related genes in *Rhizobium leguminosarum* biovar *viciae* (Rlv) strain 3841 during rhizobial infection process and pea nodule development the gene expression data at progressive four time points, 7, 15, 21, and 28 days post inoculation (d.p.i.), available at the Array Express website (http://www.ebi.ac.uk/arrayexpress), were evaluated. A broad spectrum of *Rhizobium* genes was compiled from numerous primary literature reports and several bioinformatic databases (CAZy, KEGG, NCBI COGs, and GeneDB). Fatty acid-related genes were included as part of the CRGs since a number of rhizobial carbohydrates are glycolipids, e.g. lipopolysaccharide (LPS) and lipochitin oligosaccharide nod factors (NOD). A total of 1471 CRGs were identified which accounts for approximately 20% of the Rlv 3841 genome (total 7276 genes). The compiled CRGs were analyzed based on their Riley classification system described in (54, 67) for *R. leguminosarum* (Figure 2). Among the different Riley classifications of the compiled CRGs, approximately 20% of the CRGs were categorized as "other Riley classes". These "other Riley classes" genes were those that were identified as CRGs using the above databases, but their Riley class description did not indicate a close relationship with *R. leguminosarum* carbohydrate or fatty acid metabolism. The large gene categories other than the "other Riley class" genes include the CRGs that encode transport/binding proteins (~35%), those identified as involved in 'pool, multipurpose conversion of intermediate metabolism' (~11%), and those identified as surface

polysaccharide and antigen CRGs (~7.1%). These three major categories accounted for about 53% of the CRGs.

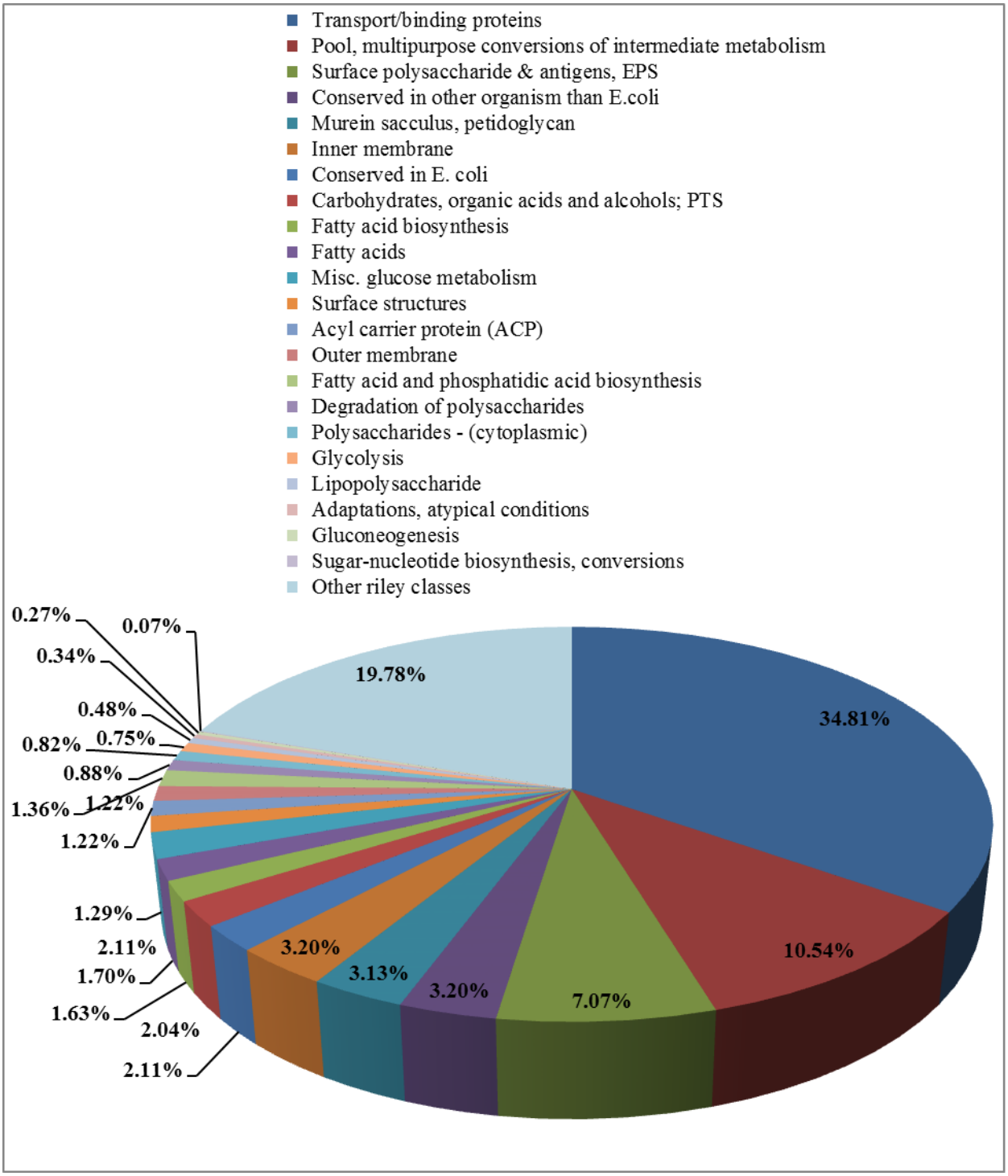


Figure 2. Classification of Compiled 1472 Carbohydrate-related *Rhizobium* Genes (according to the Riley System). A broad sample of carbohydrate- and fatty acid-related genes

compiled from CAZy, KEGG, NCBI COGs, GeneDB, and primary literature were classified according to their Riley classification as described in (67). Not all Riley classes are shown. Riley classes related to carbohydrate and fatty acids are selected here.

Table 2. Summary of Gene Expression in the Whole Genome and Compiled Carbohydrate-related Genes of *R. leguminosarum*. After the statistical analysis on the publicly available microarray data for four time points during pea nodule development, the gene expression of the total genome and the CRGs were classified as "Up-regulated", approximately 2 x or larger actual expression over free-living culture; "Constitutive", no significant expression change; "Down-regulated", approximately 2 x or less actual expression as compared to free-living culture.

Total Genes	**7 d.p.i.**	**15 d.p.i.**	**21 d.p.i.**	**28 d.p.i.**
Up-regulated	1055	1241	760	769
Constitutive	4918	4695	5644	5356
Down-regulated	1303	1340	872	1151
Total	7276	7276	7276	7276
CRGs	**7 d.p.i.**	**15 d.p.i.**	**21 d.p.i.**	**28 d.p.i.**
Up-regulated	163	174	104	125
Constitutive	1033	1007	1188	1126
Down-regulated	275	290	179	220
Total	1471	1471	1471	1471

To observe gene expression changes of the carbohydrate- and fatty acid-related genes in *Rhizobium leguminosarum* biovar *viciae* (Rlv) strain 3841 during the rhizobial infection process and pea nodule development the gene expression data at progressive four time points, 7, 15, 21, and 28 days post inoculation (d.p.i.), available at the Array Express website (http://www.ebi.ac.uk/arrayexpress), were evaluated. A broad spectrum of *Rhizobium* genes was compiled from numerous primary literature reports and several bioinformatic databases (CAZy, KEGG, NCBI COGs, and GeneDB). Fatty acid-related genes were included as part of the CRGs since a number of rhizobial carbohydrates are glycolipids, e.g. lipopolysaccharide (LPS) and lipochitin oligosaccharide nod factors (NOD). A total of 1471 CRGs were identified which accounts for approximately 20% of the Rlv 3841 genome (total 7276 genes). The compiled

CRGs were analyzed based on their Riley classification system described in (54, 67) for *R. leguminosarum* (Figure 2). Among the different Riley classifications of the compiled CRGs, approximately 20% of the CRGs were categorized as "other Riley classes". These "other Riley classes" genes were those that were identified as CRGs using the above databases, but their Riley class description did not indicate a close relationship with *R. leguminosarum* carbohydrate or fatty acid metabolism. The large gene categories other than the "other Riley class" genes include the CRGs that encode transport/binding proteins (~35%), those identified as involved in 'pool, multipurpose conversion of intermediate metabolism' (~11%), and those identified as surface polysaccharide and antigen CRGs (~7.1%). These three major categories accounted for about 53% of the CRGs.

The gene expression comparison as a function of d.p.i. during nodule development between the total genes in the Rlv 3841 genome and the CRGs is shown in Table 2. The results show that 10 - 17% of the total genome is up-regulated compared to 8 - 12% of the CRGs, while 12 - 18% of the total genome is down-regulated compared to 12 - 20% of the CRGs. Thus, the up/down regulated ratio is greater for the total genome than for the CRGs. The data also indicate that the percentage of up- and down-regulated genes for the total genome as well as for the CRGs generally decreases as a function of nodule development; i.e. 32% and 35% at 7 and 15 d.p.i. compared to 22% and 27% at 21 and 28 d.p.i. for the total genome; and 30% and 32% CRGs at 7 and 15 d.p.i. compared with 19% and 23% CRGs at 21 and 28 d.p.i. Perhaps this reflects a general overall decrease in metabolic activity as more and more bacteroids are formed within the nodule.

3.2 Expression of Known *Rhizobium* Carbohydrate-Related and Nitrogen Fixing Genes

Table 3. Summary of *Rhizobium* Gene Expression in Known Carbohydrate Classes. For the definition of expression levels, see Table 2. The publicly deposited *Rhizobium* microarray of four time points during pea nodule development. LPS, lipopolysaccharide; EPS, exopolysaccharide; PG, peptidoglycan; CEL, cellulose; CGL, cyclic glucan; TRE, trehalose; CPS, capsular polysaccharide; OTHERS, genes not assigned to any known classes of carbohydrates; NOD, nod factor; FIX, fix genes.

Gene Expression at 7 d.p.i.													
Classification	**EPS**	**LPS**	**PG**	**FA**	**CGL**	**CPS**	**GLM**	**CEL**	**GPS**	**TRE**	**OTHERS**	**NOD**	**FIX**
Up-regulated	2	5	1	25	0	0	0	0	0	0	131	11	36
Constitutive	31	51	13	105	4	0	2	6	1	4	825	26	11
Down-regulated	11	35	12	32	2	2	0	3	0	1	179	3	0
Total	44	91	26	162	6	2	2	9	1	5	1135	40	47

Gene Expression at 15 d.p.i.													
Classification	**EPS**	**LPS**	**PG**	**FA**	**CGL**	**CPS**	**GLM**	**CEL**	**GPS**	**TRE**	**OTHERS**	**NOD**	**FIX**
Up-regulated	5	11	1	15	0	0	0	1	0	1	142	3	25
Constitutive	27	56	8	107	3	0	2	6	1	3	801	29	14
Down-regulated	12	24	17	40	3	2	0	2	0	1	192	8	8
Total	44	91	26	162	6	2	2	9	1	5	1135	40	47

Gene Expression at 21 d.p.i.													
Classification	**EPS**	**LPS**	**PG**	**FA**	**CGL**	**CPS**	**GLM**	**CEL**	**GPS**	**TRE**	**OTHERS**	**NOD**	**FIX**
Up-regulated	3	5	1	13	0	0	0	0	0	0	83	5	29
Constitutive	32	71	9	121	4	0	2	8	1	4	944	29	18
Down-regulated	9	15	16	28	2	2	0	1	0	1	108	6	0
Total	44	91	26	162	6	2	2	9	1	5	1135	40	47

Gene Expression at 28 d.p.i.													
Classification	**EPS**	**LPS**	**PG**	**FA**	**CGL**	**CPS**	**GLM**	**CEL**	**GPS**	**TRE**	**OTHERS**	**NOD**	**FIX**
Up-regulated	3	2	1	18	0	0	0	0	0	0	101	4	31
Constitutive	26	56	11	118	4	0	2	9	1	4	901	32	16
Down-regulated	15	35	14	21	2	2	0	0	0	1	133	4	0
Total	44	93	26	157	6	2	2	9	1	5	1135	40	47

The CRGs were next evaluated in a more focused manner according to known carbohydrate classes in Rlv 3841. These carbohydrates include the lipopolysaccharide (LPS), extracellular and capsular polysaccharide (EPS and CPS), a gel-forming polysaccharide (GPS), cellulose (CEL), cyclic glucan (CGL), peptide glycan (PG), and trehalose (TRE). In addition, the

lipochitin oligosaccharide nod factor (NOD) and fixation genes (FIX) were included as positive controls to compare with the rest of known carbohydrate types. Genes encoding for FA synthesis were included due to the fact that some of the carbohydrates, e.g. the LPS are acylated. The CRGs for these known rhizobial components totaled 336 genes out of these 1471 genes. These CRGs for these known genes include LPS (91 genes), EPS (44 genes), CPS (2 genes), GPS (1 gene), CEL (9 genes), CGL (6 genes), PG (26 genes), TRE (5 genes), and FA (162 genes). There are some duplicates among these CRGs.

The genes recognized to encode proteins required for the synthesis of Nod factors, *nod* genes, are known to be plasmid-encoded and induced by flavonoid molecules from the host plant. The Nod factors cause root hairs to curl and are in the root cortex involved in nodule meristem formation; they confer the earliest stages of nodule formation and development. In the case of Rlv 3841, the Nod factor is a fatty acylated chitin pentasaccharide in which the fatty acid is unsaturated and the pentasaccharide is modified by acetylation. Included as positive control were also a group of chromosomal proteins that are suspected to be involved in Nod factor biosynthesis or export or are nodule-expressed, but are not involved regulation and formation of the nitrogen-reducing enzyme complex. A total of 40 genes involved in Nod factor synthesis or nodule-expressed proteins were identified as a positive control group. At the early time of 7 d.p.i., the expression of 11 *nod* genes increased while 3 decreased; Figure 3 and Table 3. The *nod* gene whose expression increased to the greatest extent at 7 d.p.i. was *nodC* (pRLl00187) which encodes the chitin synthase that makes the chitin pentasaccharide backbone of the nod factor. The expression of another nodule-related gene, RL2270, also showed a large increase at 7 and 15 d.p.i. Gene RL2270 encodes a protein that has similarity to nodulin-21, a plant nodule-specific protein that may be involved in symbiotic nitrogen fixation (13). This protein also belongs to a

family that is related to CCC1, a yeast vacuole transmembrane protein that functions as an iron and manganese transporter (44). At later times, the ratio of up- to down-regulated *nod* gene ratio decreased from approximately 4:1 at 7 d.p.i. to 1:1 by 21 d.p.i. This result is consistent with the fact that nod factor synthesis is an early event that decreases as symbiosis proceeds.

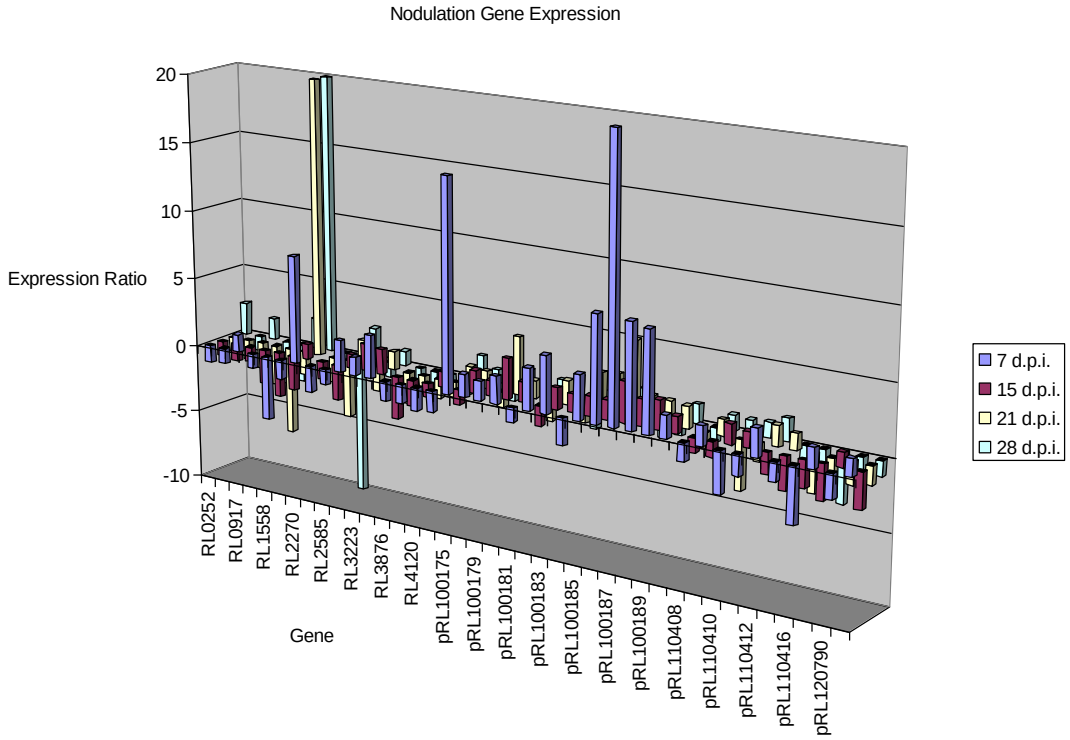


Figure 3. Gene Expression Ratio of Nodulation Genes in *R. leguminosarum* during Four Stages of Nodule Development. After the statistical analysis on the publicly available *Rhizobium* microarray data during four time points of pea nodule development, each gene expression ratio of pea bacteroids versus free-living rhizobial culture of noducation genes are graphed separately. X-axis, gene locator; Y-axis, gene expression ratio of pea bacteroids versus free-living rhizobial culture; Z-axis, four stages: Blue, 7 days; Red, 15 days; Yellow, 21 days; Light Green, 28 days. All gene IDs may not shown in graphs.

Since the *nif/fix* genes, coding for the regulatory and structural proteins of the nitrogen-fixation enzyme complex, would be expected to be up-regulated in mature bacteroids at the later stages of nodule development, a number of plasmid- and chromosomally-located genes coding

for these proteins were also observed as a positive control. It was found that of these 47 genes, >50% were classified as up-regulated genes and >20% of *nif*/*fix* genes, primarily chromosomally borne, fell into the constitutive category in all four time points during nodule development, Table 3. Three clusters of these genes, all located on plasmids, primarily on plasmid pRL10, were highly up-regulated; particularly at 21 and 28 d.p.i., Figure 4. These genes include *nifNEKDH* (pRL100158 - pRL100162), *fixBANOQP* (pRL100199 - pRL100200 and pRL100205 - pRL100207), and a homologous *fixPQON* gene region (pRL90016 - pRL90018). With an exception of 15 d.p.i., no down-regulation of *nif*/*fix* genes was found. These results are consistent with the nitrogen-fixing purpose of rhizobia–legume symbioses and a particularly strong expression of nitrogen fixation-related genes in the fully developed nodules.

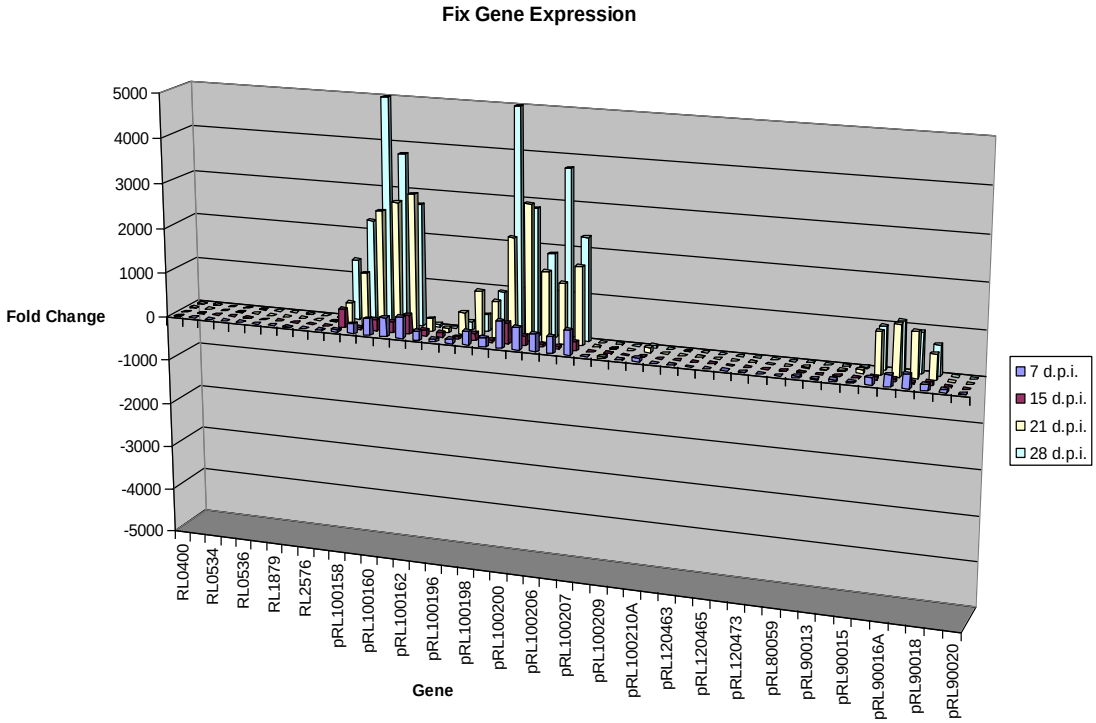


Figure 4. Gene Expression Ratio of Nif and Fix Genes in *R. leguminosarum* during Four Stages of Nodule Development. After the statistical analysis on the publicly available

Rhizobium microarray data during four time points of pea nodule development, each gene expression ratio of pea bacteroids versus free-living rhizobial culture of *fix* genes are graphed separately. X-axis, gene locator; Y-axis, gene expression ratio of pea bacteroids versus free-living rhizobial culture; Z-axis, four stages: Blue, 7 days; Red, 15 days; Yellow, 21 days; Light Green, 28 days. All gene IDs may not shown in graphs.

The EPS genes are mostly located on the chromosome with a few EPS genes found in plasmids pRL9, pRL11, and pRL12. Except for the EPS genes on pRL9, the expression of plasmid-localized EPS genes does not change during *Rhizobium* nodulation, and the EPS genes on the chromosome are largely down-regulated. The expression of EPS genes as a function of d.p.i. are shown in Figure 5. Table 3 shows that there are 44 EPS-related genes. The majority of these genes are relatively unchanged during nodulation. Changes in expression occur to 13, 17, 12, and 18 genes, respectively, at 7, 15, 21 and 28 d.p.i. The up/down gene ratios are 0.2, 0.4, 0.3, and 0.2 at these respective time points and illustrate that these genes have decreased in their expression. The RL3642 – RL3665 gene region contains various homologs encoding glycosyl transferases, acetyl transferases, polysaccharide synthases, polysaccharide polymerization, and polysaccharide export that are all down-regulated to various degrees at all d.p.i. with the largest decrease for most of these genes at 21 and 28 d.p.i. These genes include homologues of *exoV* (RL3645), *exoM* (RL3653), *exoP* (RL3662), *pssO* (RL3663), *pssN* (RL3664), and *pssT* (RL3665). The latter four genes are all putatively involved in EPS export. Down-regulated EPS genes also include RL3752 (*pssA*, a putative glycosyl transferase). The adjacent gene RL3751 (*pssB*, putative inositol monophosphatase) is up-regulated at all d.p.i. The most down-regulated EPS gene is RL4081 (*exoN*, a putative putative UTP-glucose-1-phosphate uridylyltransferase responsible for the synthesis of UDP-Glc). These results are consistent with the observation that the production of rhizobial EPS decreases with progressing nodule development and is likely to

be completely shut down by bacteroids in the symbiosomes (39, 53). Only two plasmid genes, pRL90144 and pRL90146, were altered during nodulation and their expression increased at 15 and 21 d.p.i., but slightly decreased at 7 and 28 d.p.i. The increase in these polysaccharide synthesis plasmid genes at early times may indicate that interaction with the plant induces the production of new polysaccharide or modification of an existing polysaccharide with a potential role in infection thread formation and tissue invasion.

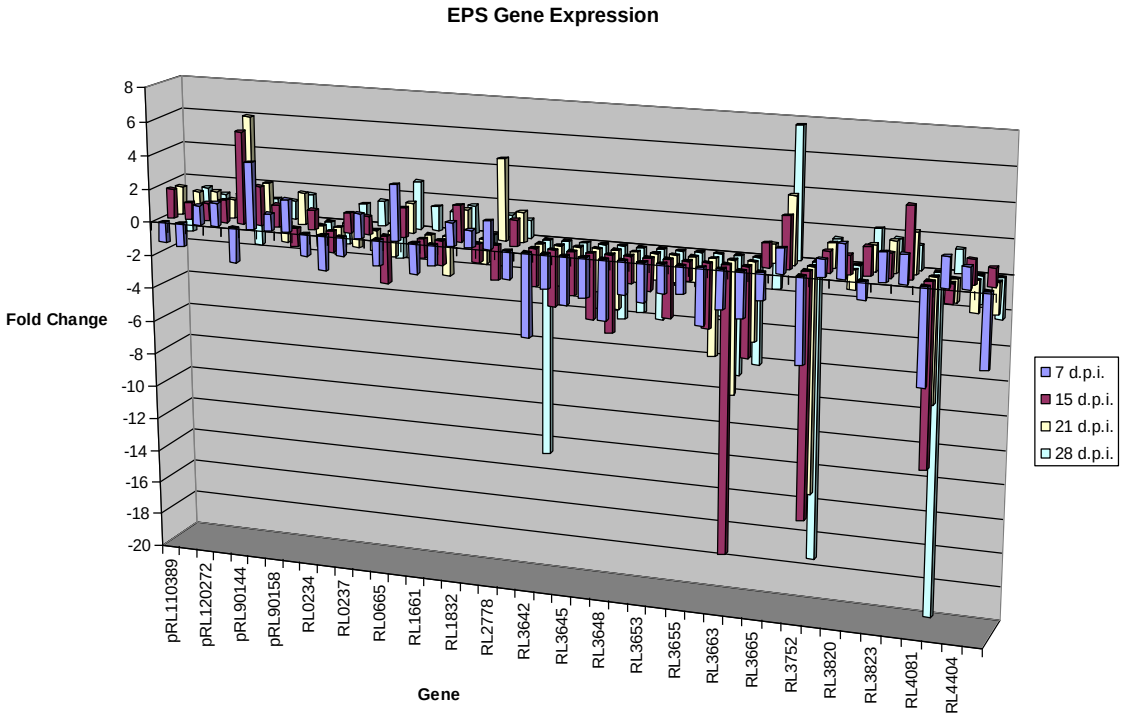
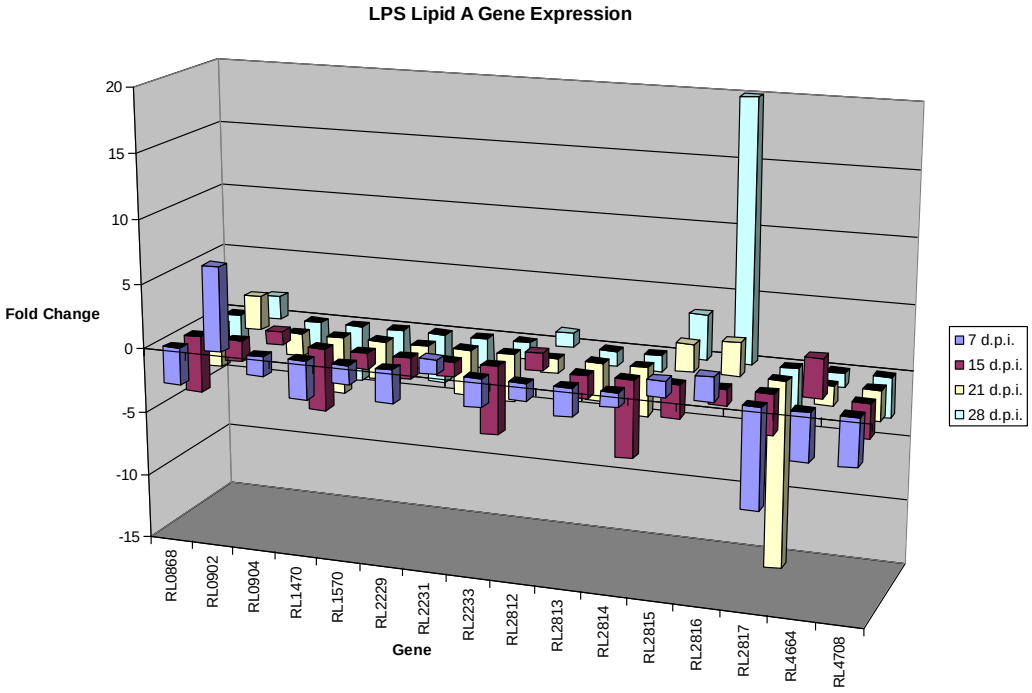


Figure 5. Gene Expression Ratio of EPS Genes in *R. leguminosarum* during Four Stages of Nodule Development. After the statistical analysis on the publicly available *Rhizobium* microarray data during four time points of pea nodule development, each gene expression ratio of pea bacteroids versus free-living rhizobial culture of EPS genes are graphed separately. X-axis, gene locator; Y-axis, gene expression ratio of pea bacteroids versus free-living rhizobial culture; Z-axis, four stages: Blue, 7 days; Red, 15 days; Yellow, 21 days; Light Green, 28 days. All gene IDs may not shown in graphs.

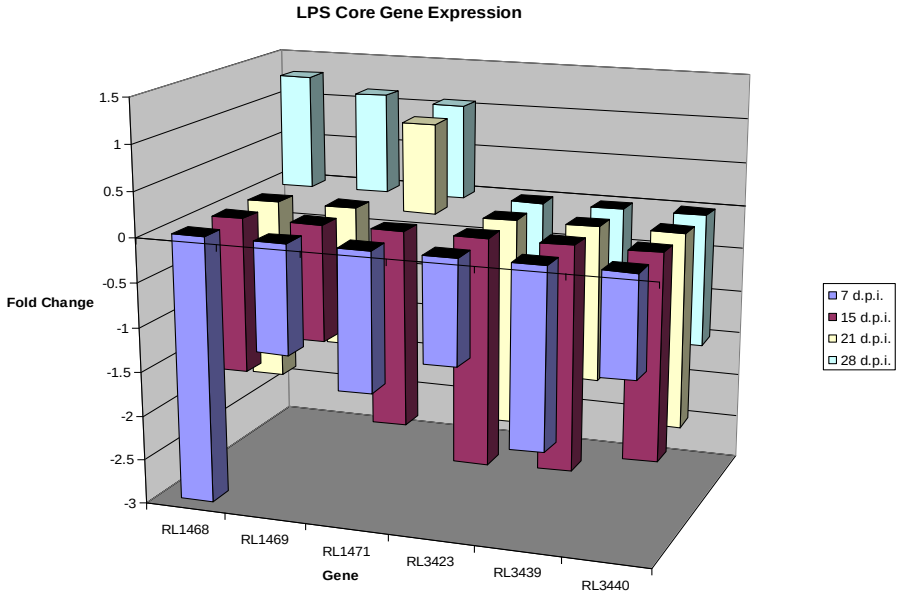
The LPS consists of three structural regions; lipid A, core oligosaccharide, and O-chain polysaccharide. Changes in the expression of the lipid A genes in response to nodulation are shown in Figure 6A. As with EPS genes, most of the lipid A genes decrease in expression as a function of nodulation. At 7 d.p.i., only one gene, RL0902 (kdtA which encodes the Kdo transferase) increases. In addition, at 21 and 28 d.p.i. the genes RL2816 (coding for FabZXL, the hydroxymyristoyl-[acp] dehydratase) and RL2815 (coding for FabF2XL, a 3-oxoacyl-[acp] synthase), both involved in the biosynthesis of the LA VLCFA, show increased expression. Those genes with the greatest decrease in expression are RL3297 (*lpxC*, UDP-3-O-[3-hydroxymyristoyl] *N*-acetylglucosamine deacetylase) and RL2817 (*acpXL*, acyl carrier protein for the very long chain fatty acyl moiety). The decrease in most of these lipid A genes, particularly at the later d.p.i., likely reflects a decrease in bacterial cell division and an increase in terminal bacteroid formation. However, unexplained is why at 21 and 28 d.p.i. *fabZXL* (RL2816) and *fabF2XL* (RL2815) are more highly expressed when most of the other lipid A genes are expressed at a decreased level. The genes responsible for synthesis of the core oligosaccharide slightly decrease in their expression at most d.p.i. None of the core genes are affected to the extent of the lipid A genes (Figure 6B). The O-chain polysaccharide genes are located in the RL0794-RL0826 gene region. The majority of these genes are decreased in their expression level as a function of nodulation with the greatest reduction at 28 d.p.i. (Figure 6C). The greatest decrease occurs in gene RL0800 and the RL0822 –RL0825 genes. RL0800 encodes a hypothetical protein, while RL0822-RL0824 putatively encode aminotransferase, acetyl transferase, and oxidoreductase proteins, respectively, which possibly are required for the synthesis of a unique glycosyl residue found in Rlv 3841 OPS, rhizoaminuronic acid (3-

acetimidoylamino-3-deoxy-D-glucuronic acid) (6), and RL0825 encodes for GDP-mannose-4,6-dehydratase which is a part of the fucosyl synthetic pathway. Thus, it seems likely that the O-chain polysaccharide is reduced and/or modified during bacteroid formation which is consistent with the observation of the production of bacteroid-specific O-polysaccharide mAb epitope changes (6, 35). Consistent with the observation that changes to the OPS or that a novel polysaccharide may be produced during nodulation is the observation of expression changes that occur to a putative second set of O-chain polysaccharide genes on plasmid pRL9 (Figure 6D). These genes include those in the pRL90051 - pRL90053 and pRL90132 - pRL90158 regions. A decrease in the expression of two genes, pRL90051 (which encodes a putative galactosyl tranferase) and pRL90053 (a putative O-antigen ligase) occurs; the former at 7 and 15 d.p.i. and the latter at all four d.p.i. Interestingly, of pRL90132-pRL90158 genes a number have increased levels of expression at various d.p.i. Most of these increases occur at the 15 and 21 d.p.i. time points. Gene pRL90144 encodes a putative EPS biosynthesis protein, pRL90150 a putative acetyl transferase, and pRL90146 a putative UDP-galactosyl lipid carrier transferase. Overall, the number of LPS synthesis genes are 91 and the majority are down-regulated during nodulation (Table 3); 35, 24, 15, and 34 at 7, 15, 21, and 28 d.p.i., respectively. In summary, these results imply that the total amount of LPS decreases during bacteroid formation and that there are likely modifications to the O-chain polysaccharide or the synthesis of an additional polysaccharide.

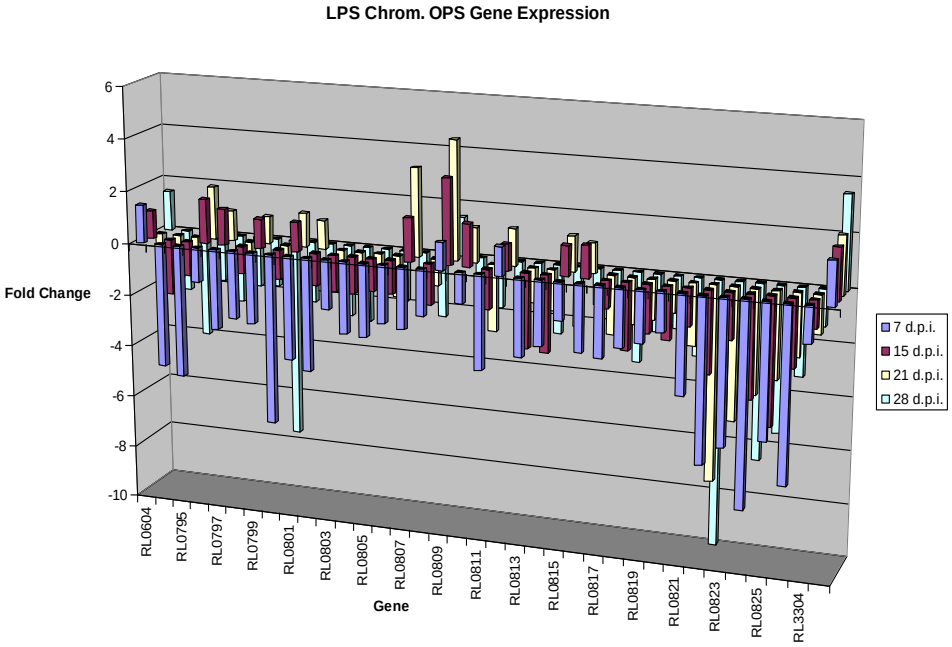
A.


LPS Lipid A Gene Expression
Fold Change
20
15
10
5
0
-5
-10
-15
RL0868
RL0902
RL0904
RL1470
RL1570
RL2229
RL2231
RL2233
RL2812
RL2813
RL2814
RL2815
RL2816
RL2817
RL4664
RL4708
Gene
7 d.p.i.
15 d.p.i.
21 d.p.i.
28 d.p.i.

B.


LPS Core Gene Expression
Fold Change
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
RL1468
RL1469
RL1471
RL3423
RL3439
RL3440
Gene
7 d.p.i.
15 d.p.i.
21 d.p.i.
28 d.p.i.

C.



D.

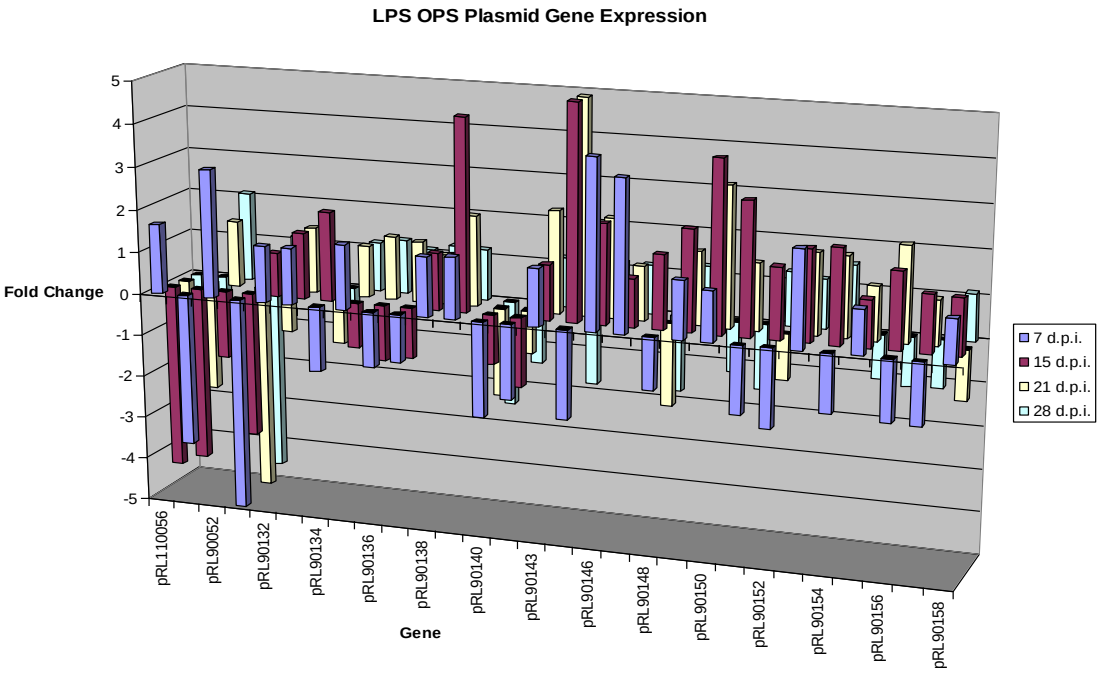


Figure 6A-D. Gene Expression Ratio of LPS Genes in *R. leguminosarum* during Four Stages of Nodule Development. A, LPS Lipid A genes. B, LPS Core genes. C, LPS OPS genes in chromosome. D, LPS OPS genes in plasmids. After the statistical analysis on the

publicly available *Rhizobium* microarray data during four time points of pea nodule development, each gene expression ratio of pea bacteroids versus free-living rhizobial culture of LPS genes are graphed separately. X-axis, gene locator; Y-axis, gene expression ratio of pea bacteroids versus free-living rhizobial culture; Z-axis, four stages: Blue, 7 days; Red, 15 days; Yellow, 21 days; Light Green, 28 days. All gene IDs may not shown in graphs.

A majority of known genes involved with PG is found to be on the chromosome. Except for the only up-regulated PG gene (RL1878; 'putative peptidoglycan binding protein') in *R. leguminosarum* during symbiotic pea nodule development, the PG genes are either down-regulated or constitutively expressed. The down-regulation of the PG genes occurs to the greatest extent at 15 and 21 d.p.i. In most stages of infection process, the proportion of down-regulated PG genes is higher than the proportion of constitutively regulated PG genes (Table 3 and Figure 7). Again, the down-regulation of PG synthesis genes is consistent with the fact that Rlv 3841 bacteroids cease cell division. Interestingly, a putative PG binding protein that is encoded by the one up-regulated gene, RL1878, has a BON superfamily domain which is present in a family of osmotic shock protection proteins; e.g. OsmY (64). The RL1878 gene is adjacent to three genes that putatively encode nitrogen fixation regulatory proteins. Thus, it is possible that the product of RL1878 is involved in protecting the bacterium from changes in osmolarity that occur during infection of the legume host.

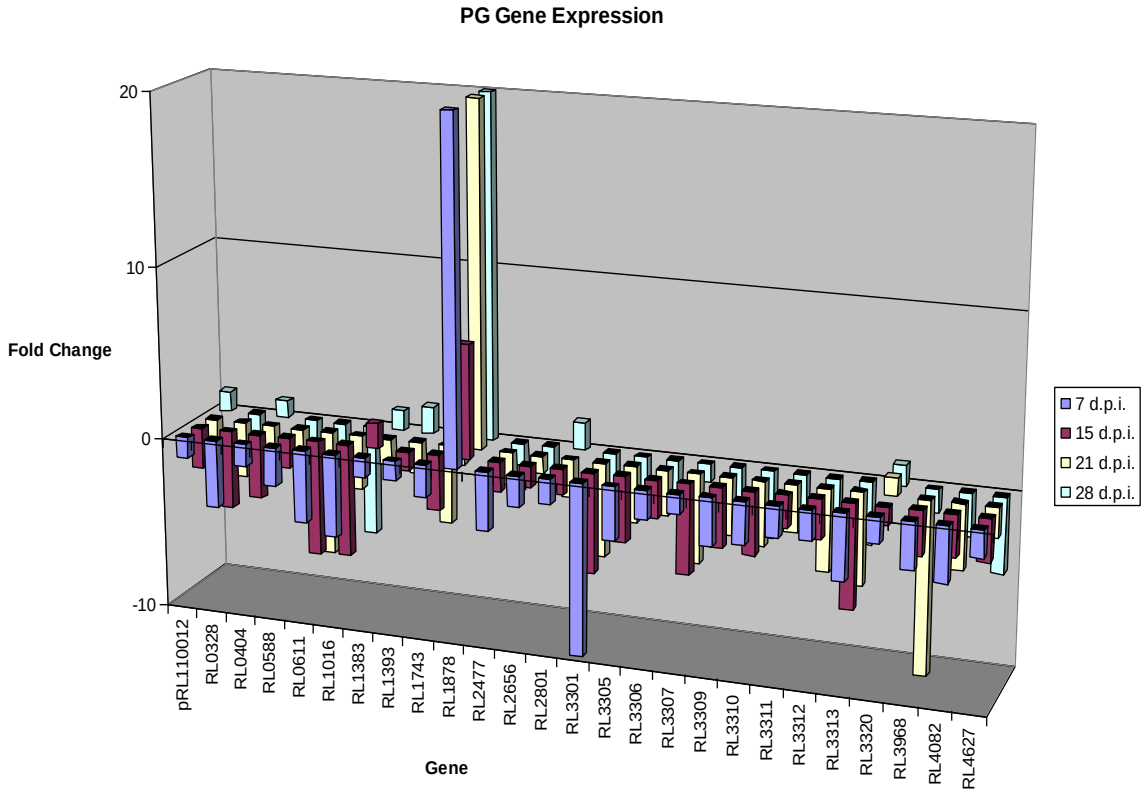


Figure 7. Gene Expression Ratio of PG Genes in *R. leguminosarum* during Four Stages of Nodule Development. After the statistical analysis on the publicly available *Rhizobium* microarray data during four time points of pea nodule development, each gene expression ratio of pea bacteroids versus free-living rhizobial culture of PG genes are graphed separately. X-axis, gene locator; Y-axis, gene expression ratio of pea bacteroids versus free-living rhizobial culture; Z-axis, four stages: Blue, 7 days; Red, 15 days; Yellow, 21 days; Light Green, 28 days. All gene IDs may not shown in graphs.

The fatty acid (FA) biosynthesis and metabolism genes were included in this analysis since one of the primary carbohydrates in the bacterium is the lipopolysaccharide (LPS), which undergoes some changes in its lipid A acylation pattern of the LPS, and it is also known that both the LPS and the cell surface become hydrophobic during symbiosis (11, 33). Approximately 162 FA-related genes were identified. Several FA genes are significantly down-regulated. One of these is RL1559, *acpP*, encodes the synthesis of AcpP, the constitutive acyl carrier protein. This

gene is down-regulated primarily at 28 d.p.i. which probably reflects the fact that the mature bacteroids are no longer undergoing cell division. Other down-regulated genes include RL2089, which putatively encodes for the synthesis of AccB (acetyl-CoA carboxylase) required for the synthesis of malonyl-CoA, an early precursor of FA synthesis, and RL4621 which encodes PhaA (acetyl-CoA acetyl transferase) that converts acetoacetyl-CoA to acetyl-CoA and is possibly involved in fatty acid degradataton. Thus, in bacteroids, certain genes that encode for FA synthesis as well as degradation are down-regulated which probably reflects either a slowing of the general metabolism or a specialized metabolism in mature bacteroids. Several plasmid FA genes are also down-regulated; pRL110386, pRL120524, and pRL120235, encode for acyl-CoA, alcohol, and fatty aldehyde dehydrogenases, respectively, and are also likely involved in fatty acid degradation. Two chromosomal genes that are up-regulated are RL1876 and RL4267. Gene RL1876 is significantly up-regulated at 7, 21, and 28 d.p.i. and encodes a putative alcohol dehydrogenase. Interestingly, this gene is part of a four gene cluster that includes the PG gene RL1878 which, as described above, may encode a protein that is involved in a response to osmotic shock. As mentioned above, this four gene cluster is adjacent to three genes that putatively encode nitrogen-fixation regulatory proteins. Thus, it is possible that both RL1876 and RL1878 are increased in response to a change in osmolarity that occurs during symbiosis. Gene RL4267 is up-regulated on all d.p.i. and putatively encodes AcoD, a putative acetaldehyde dehyrogenase whose possible function in symbiosis is difficult to predict. Several genes encoded on plasmids are also up-regulated. Gene pRL100145 is up-regulated on 7 d.p.i. and encodes a putative acyl-CoA dehydrogenase which is likely involved in fatty acid degradation. Gene pRL110313 is up-regulated at 7, 21, and 28 d.p.i. and putatively encodes Cds, a transmembrane cardiolipin synthetase. It is known that the cell surface of bacteroids is more hydrophobic; thus,

it is likely that changes occur to various membrane lipids during symbiosis, and this gene may be part of that response. Two other plasmid genes that are up-regulated are pRL110622 and pRL90074. Gene pRL110622 is up-regulated at all d.p.i. and encodes a putative transmembrane fatty acid desaturase, while gene pRL90074 is up on 7 d.p.i. and encodes a glycerol-3-P-dehydrogenase involved in lipid metabolism. Both of these genes could be involved in the synthesis or modification of phospholipids that may occur in the membrane changes to bacteroids.

The gene expression for additional known carbohydrates in *R. leguminosarum* other than the genes related to EPS, LPS, and PG are those for cellulose (CEL), capsular polysaccharides (CPS), glucomannan (GLM), gel-forming polysaccharides (GPS), cyclic glucans (CGL), and trehalose (TRE). Three CEL genes are down regulated at 7 d.p.i. These genes are RL1646, RL1647, and RL1729. Genes RL1646 and RL1647 putative encode for the cellulose synthase catalytic unit and cellulose synthase protein, respectively, while RL1729 encodes a putative two-component response regulator transcriptional regulatory protein. The implication is that cellulose synthesis is down-regulated early in symbiosis. A gene involved in the transport of cyclic glucan (CGL), RL4640, is down-regulated at all time points but greatest at 7 d.p.i. There is little change in two genes identified as being involved in glucomannan (GLM) synthesis and transport, RL1661 and RL1664, respectively, two genes related to CPS synthesis, pRL120461 and RL3664, or the gel-forming polysaccharide (GPS) related gene, RL4404. One gene related to trehalose synthesis, pRL120722, is down-regulated at 7 d.p.i.

3.3 Expression of Other Putative *Rhizobium* Carbohydrate Genes

Table 4. Summary of 1135 *Rhizobium* "Other" Carbohydrate-Related Genes that Have Not Been Assigned to the Known Classes of *Rhizobium* Carbohydrates. A putative carbohydrate-related gene portion of the broadly compiled carbohydrate- and fatty acid-related gene sample was analyzed as either up- or down-regulation. Top table, up-regulated genes; bottom table, down-regulated genes; up- and down-regulations as defined in Table 1. The total numbers of genes in different putative gene function are recorded inside parenthesis.

d.p.i. / Putative Gene Function	7	15	21	28	Number of Genes Up-regulated at All Stages
Transport/binding Proteins (511)	62	63	39	52	8
Glycoconjugate Metabolism (296)	24	41	21	21	2
Fatty Acids Metabolism (20)	4	3	1	2	1
Others (308)	41	35	22	26	4
Total	131	142	83	101	15

d.p.i. / Putative Gene Function	7	15	21	28	Number of Genes Down-regulated at All Stages
Transport/binding Proteins (511)	69	64	35	32	12
Glycoconjugate Metabolism (296)	54	51	36	46	19
Fatty Acids Metabolism (20)	6	4	5	5	3
Others (308)	50	73	32	50	16
Total	179	192	108	133	50

Table 5. "Other" Carbohydrate-Related Genes that Change Expression at Least Approximately Two-Fold (Up or Down) for All Time Points. The fold increase (+) or decrease (-) are given for each time d.p.i. Some major "at least 2x up-regulated genes for all time points" are in green, and some major "at least 2x down-regulated genes for all time points" are in red.

GeneID	Gene Name	Putative Protein	7 d.p.i.	15 d.p.i.	21 d.p.i.	28 d.p.i.
Transport Genes						
RL0680	-	bifunctional preprotein translocase subunit SecD/SecF	56.9	40.0	83.5	1715.9
RL1683	-	putative transmembrane component of ABC transporter	3.2	8.8	2.8	2.2
RL2046	tatA	twin arginine translocase protein A	3.1	6.7	12.5	9.4
RL3533	-	putative solute-binding component of ABC transporter	3.9	2.7	11.6	38.4
RL3910	-	putative ATP-binding component of ABC transporter	3.2	2.5	4.8	2.1
RL4244	-	putative transmembrane component of ABC transporter	10.1	4.7	3.0	13.9
RL4274	-	HlyD family transmembrane efflux protein	24.1	3.5	3.0	2.2
RL4530	-	putative ATP-binding component of ABC transporter	6.9	1.8	17.9	56.0
RL4685	-	putative solute-binding component of ABC transporter	3.4	2.7	5.0	3.3
RL0186	-	putative transmembrane component of ABC transporter	-3.7	-3.2	-2.6	-2.6
RL0745	aglE	ABC transporter substrate-binding protein	-6.6	-6.1	-6.3	-8.2
RL2054	-	putative Sec translocase associated protein	-4.4	-9.4	-7.3	-8.4
RL3051	-	putative transmembrane component of ABC transporter	-5.0	-5.9	-4.5	-2.8
RL3540	-	putative solute-binding component of ABC transporter	-3.1	-3.2	-6.2	-4.3
RL3617	-	putative solute-binding component of ABC transporter	-4.8	-6.8	-7.8	-10.1
pRL110372	-	putative permease component of ABC transporter	-3.4	-2.5	-4.4	-4.1
pRL110373	-	putative permease component of ABC transporter	-3.7	-4.9	-6.3	-2.2
pRL110374	-	putative substrate-binding component of ABC transporter	-4.6	-14.6	-9.1	-23.0
pRL120039	thuA	putative sugar uptake related protein	-2.5	-2.5	-2.7	-4.6
pRL90301	-	putative permease component of ABC transporter	-5.3	-2.3	-3.8	-2.3
pRL90302	-	putative ATP-binding component of ABC transporter	-3.5	-2.7	-3.6	-5.9
Glycoconjugate metabolism Genes						
pRL100121	acsA	acetyl-coenzyme A synthetase	7.5	29.7	133.7	74.3
RL3227	-	hypothetical protein	3.1	2.1	4.4	13.3
RL0179	gpmA	phosphoglyceromutase	-2.5	-8.2	-3.8	-2.3

RL0324	-	hypothetical protein	-14.6	-4.1	-4.7	-9.0
RL1086	ppdK	pyruvate phosphate dikinase	-3.4	-3.8	-3.4	-3.7
RL1663	-	putative polysaccharide export protein	-3.1	-3.0	-4.6	-2.2
RL1817	-	putative lipase/esterase	-6.8	-4.6	-3.5	-6.0
RL2239	eno	phosphopyruvate hydratase	-2.9	-6.4	-4.2	-3.7
RL2541	-	putative penicillin-binding protein/D-alanyl-D-alanine carboxypeptidase	-27.7	-9.7	-20.0	-26.0
RL2791	-	hypothetical protein	-5.9	-6.0	-5.5	-2.6
RL3322	pfp	pyrophosphate--fructose-6-phosphate 1-phosphotransferase	-3.8	-4.6	-5.2	-2.2
RL3647	-	putative polysaccharide polymerisation protein	-3.0	-8.6	-7.7	-3.5
RL3677	lspL	putative UDP-glucuronate 5'-epimerase	-12.7	-5.7	-15.9	-12.9
RL4012	-	putative fructose-bisphosphate aldolase	-4.0	-4.6	-5.1	-6.1
RL4116	glgC	glucose-1-phosphate adenylyltransferase	-4.0	-3.7	-2.9	-2.4
RL4118	pgm	phosphoglucomutase	-2.2	-3.6	-4.6	-2.1
RL4562	-	hypothetical protein	-2.2	-2.8	-3.4	-2.8
RL4620	phaB	putative acetoacetyl-CoA reductase	-8.6	-4.3	-5.0	-8.0
RL4727	acs	acetyl-CoA synthetase	-4.1	-4.4	-4.5	-2.5
pRL120540	-	hexapeptide repeat-containing protein	-4.2	-2.7	-4.2	-3.0

"Other Genes"

RL0423	-	putative transmembrane protein	3.4	4.7	3.1	3.8
RL1145	-	putative conjugated bile salt hydrolase	21.2	9.2	656.0	588.3
RL3016	pcaH	protocatechuate 3,4-dioxygenase beta chain (3,4-pcd)	215.5	78.8	448.1	142400.7
RL3130	-	putative phosphatase protein	63.8	13.1	347.2	59.3
RL1589	-	putative ropB outer membrane protein	-4.4	-12.6	-7.2	-4.1
RL2052	-	putative peptidase	-13.4	-4.2	-11.4	-7.0
RL2228	-	putative outer membrane protein	-3.6	-6.5	-6.1	-5.8
RL2392	glnA	putative glutamine synthetase I	-4.3	-7.3	-3.4	-7.3
RL3024	-	putative polysaccharidase	-7.8	-8.6	-5.6	-9.1
RL3267	-	OmpA family outer membrane protein	-9.6	-20.2	-21.9	-10.4
RL3378	cinI	putative autoinducer synthesis protein	-24.1	-8.0	-8.1	-10.0
RL3382	-	hypothetical protein	-3.3	-3.9	-3.8	-4.4
RL3463	lgt	prolipoprotein diacylglyceryl transferase	-2.3	-3.0	-2.7	-2.6

RL3549	glnII	putative glutamine synthetase II	-59.6	-3.5	-15.5	-6.2
RL3962	-	putative heat resistant agglutinin 1	-10.5	-2.6	-4.0	-4.3
RL4011	pgk	phosphoglycerate kinase	-5.1	-5.1	-5.7	-4.1
RL4073	hss	putative homospermidine synthase	-4.8	-2.8	-5.6	-3.9
RL4176	xylA	xylose isomerase	-6.9	-7.4	-8.4	-4.3
RL4423	dapA	putative dihydrodipicolinate synthase	-3.1	-10.7	-3.7	-7.5
pRL90303	-	putative outer membrane lipoprotein	-8.1	-6.4	-11.9	-30.5
Fatty Acid Genes						
RL1240	-	putative lipase	3.0	6.6	16.9	2.8
RL2621	-	putative inositol-1-monophosphatase	-3.2	-2.4	-4.3	-2.4
RL3622	-	putative inositol 2-dehydrogenase	-6.5	-2.3	-4.6	-3.6
RL4025	suhB	putative inositol-1-monophosphatase	-4.0	-7.6	-5.4	-4.8

The above 336 CRGs were identified as being involved in the synthesis of the known rhizobial polysaccharides and fatty acids. Some of these 336 CRGs have been counted more than once as they are assigned to being involved in more than one carbohydrate; thus the total 347 genes in Table 3 are actually 336 CRGs. The remaining 1135 genes were also classified according to the Riley classification as being CRGs. Of these 1135 CRGs approximately 45% (511 CRGs) are involved in various transport processes, 26% (296 CRGs) were putatively involved in some aspect of glycoconjugate metabolism, 1.7% (20 CRGs) in fatty acid metabolism, and 27% (308 CRGs) were designated as “Other” since the class of CRGs could not be assigned (Table 4).

In the comparison between the total numbers of up- or down-regulated genes throughout symbiosis (Table 4), the numbers of total down-regulated genes are greater than the number of total up-regulated genes in all four time points. In general, the numbers of either up- or down-regulation in four categories for earlier infection days (7 and 15 d.p.i.) are greater than that of

later infection days (21 and 28 d.p.i.). The number of genes encoding transport/binding proteins that are up-regulated at all stages of pea nodule development is the highest, and the number of genes (19 genes) encoding proteins involved in glycoconjugate metabolism that are down-regulated is the largest. The up/down ratio of genes affected at all stages of symbiosis is 0.30 indicated that the overall trend of gene regulation is down.

Of the 511 transport CRGs, 8 are up-regulated and 12 and down-regulated at all four time points, Table 5. Six of the 8 up-regulated genes are ABC transporter components. Of these 6 genes, the major up-regulated transport gene is RL0680 which putative encodes the preprotein translocase SecD/SecF subunits that are part of the protein secretory system that transports newly synthesized proteins across the cytoplasmic membrane into the periplasm or out of the bacterial cell. The major down-regulated gene is pRL110374 which putatively encodes the substrate-binding components of an ABC transporter involved in sulfate transport.

Putative glycoconjugate metabolism accounts for 296 genes of these CRGs. Of these, only 2 genes are up-regulated at all time points, while 19 genes are down-regulated at all stages. Of these, one major down-regulated gene includes RL0179 which putatively encodes GpmA, phosphoglyceromutase, which converts glycerate-2-phosphate to glycerate-3-phosphate in a glycolysis step. Two other major down-regulated genes appear to be involved in peptidoglycan synthesis; RL2541 putatively encodes D-alanyl-D-alanine carboxypeptidase which is a key enzyme in PG synthesis that cross-links the individual PG chains, and RL2052 which putatively encodes a PG-binding protein. Thus, this observation is consistent with the previous discussion of PG genes above, most of which were also down-regulated. The fourth gene that is down-regulated to a major degree is RL3677 which putative encodes IspL, a UDP-glucuronate-5'-epimerase, that converts UDP-GlcA to UDP-L-IdoA (iduronic acid). This glycosyl component,

or one derived from UDP-L-IdoA, is not known be part of any of the known *R. leguminosarum* bv. *viciae* carbohydrates and, therefore, may indicated that an as yet unknown carbohydrate is present or modified during symbiosis.

The expression of four genes putatively involved in fatty acid metabolism is up- or down-regulated at all time points; one gene is up- and three are down-regulated. None of these genes are affected to a major degree. However, RL2621 is significantly down-regulated at 21 d.p.i. It putatively encodes for an enzyme involved in the synthesis of phosphatidic acid. The reduction of acidic phospholipids on the bacterial membrane would be consistent with the increased hydrophobicity observed for the nitrogen-fixing the bacteroids (11, 33).

Of the various 308 "other" CRGs, 4 are up and 16 are down-regulated at all time points. Two of the major up-regulated genes are RL1145 and RL3016. Both of these genes are increasingly up-regulated during symbiosis with the major increase occurring in mature bacteroids at 28 d.p.i. Gene RL1145 encodes a putative bile salt hydrolase, choloylglycine hydrolase (CGH). This enzyme is in a number of probiotic microbes and is thought to be secreted when the microbe is within the host in order to protect the microbe from the detergent action of bile salts. This enzyme is also produced by *Brucella abortus*, a pathogen that forms a chronic intracellular infection by surviving within modified phageosomes within the host, as well as by *Brucella melitensis* and *Brucella suis* (14-15). Again, the production and excretion of CGH by these pathogens promotes the infection of mice via the oral route. These pathogens are phylogenetically related to *Rhizobium leguminosarum*. It is not known if "bile salts" are present in pea, however, the increased expression of this gene indicates that such a possibility should be investigated. A second major up-regulated gene is RL3016 which putatively encodes PcaH, protocatechuate 3,4-dioxygenase beta chain. This protein cleaves the aromatic ring of

protocatechuate. While the function of the RL3016 gene product in symbiosis is not known, it is very highly induced at all time points, particularly at 28 d.p.i. One speculation may be that it oxidizes the antimicrobial compound, pisatin, produced by pea in response to potential pathogens. Of the down-regulated genes, RL3549 is greatly down-regulated at 7 d.p.i. and then to much lesser levels at the later time points. This gene putatively encodes GlnII, glutamine synthetase II which converts glutamate to glutamine. It may be that this bacterial enzyme is reduced in order that the newly formed ammonia resulting from nitrogen-fixation would be available for utilization by the plant host.

CHAPTER 4

DISCUSSION AND CONCLUSION

It is a long-standing hypothesis that rhizobial cell and surface carbohydrates are of great importance in *Rhizobium*-legume symbiosis. On the basis of this hypothesis, numerous studies have investigated classes of rhizobial carbohydrates at individual stages of legume nodule development, using mainly biochemical, chemical, molecular, and genetical approaches. However, a comprehensive analysis of the gene expression of all carbohydrate-related genes at different stages of nodule development has not yet been performed. The recent elucidation of whole *Rhizobium* genome sequences combined with modern transcription profiling of gene expression allows for this kind of analysis. We used in this work the model strain *Rhizobium leguminosarum* biovar *viciae* 3841, forming nodules on peas, to perform such a study. A broad sample of carbohydrate-related genes (CRGs) was compiled and the gene expression levels studied at 7, 15, 21, and 28 days post inoculation. The gene expression data for the whole genome were retrieved from the publicly available whole genome microarray data set of *R. leguminosarum* strain 3841 (http://www.ebi.ac.uk/arrayexpress) (36). A statistical approach, that involved data normalization and consistency improvements, was used verify gene expression levels and to identify any CRGs that showed significant changes in their expression levels during *Rhizobium* pea infection and bacteroid development relative to the gene expression level in free-living rhizobia.

From the 7276 identified open reading frames (ORFs) of the *R. leguminosarum* 3841 genome, we identified 1471 (20.2%) ORFs as potentially carbohydrate-related. Based on the

Riley classification for bacterial genes, the sample contained genes that encode for surface polysaccharide and antigen CRGs (~7.1%), genes identified as involved in 'pool, multipurpose conversion of intermediate metabolism' (~11%), genes coding for transport/binding proteins (~35%), and genes classified as "other Riley classes" (~20%) of CRGs. During nodule development, at 7 d.p.i. (at this point a large proportion of the bacteria is still contained extra-cellularly in the infection threads or the plant apoplast) about 67% of the genes of the whole genome were constitutively expressed as compared to 70% of the CRGs (Table 2). At 28 d.p.i. (then, the majority of bacteria are fully developed, symbiosome-occupying bacteroids), the proportion of constitutively expressed genes was significantly higher (73.6% for the whole genome and 76.5% among the CRGs). Overall, the percentage of up- and down-regulated genes for the total genome as well as for the CRGs generally decreases as a function of nodule development; this decrease may reflect the fact that the rhizobia at 7 d.p.i. occupy extra- and intra-cellularly a number of different plant microenvironments to which they have to adapt and which may cause the lower percentage of constitutively expressed genes; at later stages more and more bacteria occupy and adapt as bacteroids exclusively to the symbiosome environment which may be reflected in an overall decrease in metabolic activity and changes in gene expression.

Conceptually, during nodule development rhizobia encounter at least two radically different environments: an extra-cellular environment, which includes infection thread environment and the plant's apoplast, and an intra-cellular environment, the plant organelle-like symbiosome environment. During nodule development, at 7 d.p.i. rhizobia are still largely confined extra-cellularly to the infection threads and the apoplast environment, while at 28 d.p.i. the majority of bacteria are intracellular and occupy symbiosome microenvironment; in the stages in between, the proportion of extra-cellular and intracellular bacterial cells varies. Given

these differences in rhizobial life style, it was of particular interest to investigate gene expression at the different stages of nodule development. Moreover, using biochemistry and genetics, rhizobial adaptation to these environments has been the topic of numerous reports and our knowledge about carbohydrate formation in these environments is significant (19, 22); for pea nodule bacteria, our current picture is as follows: key players in the infection thread environment that originate from rhizobia are the Nod-factors, cellulose fibrils, EPS, CPS, cyclic glucans, and glucomannan polysaccharides; in the symbiosome environment the rhizobia are exposed to a low oxygen environment in which the bacteria presumably do not form extra-cellular and capsule polysaccharides, but the LPS undergoes extensive modifications, and the rhizobia express the nitrogenase enzyme complex allowing them to fix atmospheric nitrogen to ammonia. The expression levels of the genes involved in the biosynthesis of these known classes of carbohydrates or proteins were, therefore, good control systems to test the validity of our gene expression experiment and was used to test our basic assumption, i.e. that the formation of these molecular classes can be monitored through investigating the gene expression levels of the genes involved in their biosynthesis.

In our sample of CRGs from *R. leguminosarum* 3841, 336 genes were classified as being involved in the biosynthesis of known carbohydrate types or in the nitrogenase enzyme complex (i.e. 91 LPS genes, 47 extracellular or capsular polysaccharide genes; 40 Nod factor-related genes, 9 cellulose fibril genes; 6 cyclic glucan genes; 26 peptidoglycan genes; 5 trahalose-related genes, 47 genes related to the nitrogenase complex), but with some duplicate genes among the groups. Two excellent positive controls for the validity of our gene expression experiment were the genes involved in the formation of Nod factor and of the nitrogenase enzyme complex. While the genes involved in Nod factor synthesis were expected to be expressed only at the early stages

of nodule development, the genes of the nitrogenase enzyme complex were expected to be primarily expressed at the later stages of nodule development. As expected, our findings on the overall gene expression of Nod factor-related genes was that they were significantly elevated at the earliest infection stages at 7 d.p.i. and then decreased as the nodule development proceeded. Also in agreement with our expectations was that the genes involved in the formation of the nitrogenase complex were only marginally expressed at the early stages of nodule development, but strongly at the two late stage of 21 and 28 d.p.i. Thus, our findings on the gene expression in this bioinformatics study corroborate nicely what is known on the formation of Nod factor and the nitrogenase enzyme complex using biochemical and genetical methods.

More surprising was the gene expression pattern of genes involved in EPS synthesis. Based on literature findings (39, 53), we expected some level of gene expression for the early stage in nodule development, i.e. at 7 d.p.i., and then a gradual down-regulation of these genes, culminating in no gene expression in mature bacteroids at 28 d.p.i. However, we found that EPS-related genes on the chromosome were largely down-regulated at all staged of nodule development. Somewhat in contrast, the expression of the EPS genes located on plasmids did not change much during *Rhizobium* nodulation except from EPS genes located on pRL9 which were somewhat up-regulated. Based on this gene expression data, one would expect a modified EPS being synthesized by the rhizobia during the passage through the infection threads. This aspect has to date not been studied by other means of investigation (e.g. biochemistry or chemistry). Unexplained is also the strong up-regulation of a putative inositol monophosphatase gene, i.e. RL3751 or *pssB*, which seems a candidate gene for mutational inactivation and functional studies of the effects on nodule development and EPS structure. For the additional known carbohydrates classes including cellulose, capsular polysaccharides, glucomannan, gel-forming

polysaccharides, cyclic glucans, and trehalose, we observed overall down-regulated gene expression levels with progressing nodule development which fits well with a model of a gradual reduction of surface carbohydrates formation during nodule development culminating in their suppression in symbiosome-bound bacteroids.

For *R. leguminosarum* 3841 contained within symbiosomes, it has been shown that the LPS displays structural modifications in the lipid A, and in the O-chain, while the core region was found to be invariant (32-33). These modifications were not seen in bacteria contained in the infection threads or the apoplast. Furthermore, it has been suggested that the total amount of LPS displayed on bacteriods may be reduced by about 50% compared to the amount of LPS displayed in free-living bacteria (33). In line with these biochemical and chemical findings, we found that most genes involved in the lipid A biosynthesis showed decreased levels of gene expression. However, surprising was the finding that this decrease was visible at all stages of nodule expression. This was not an obvious finding; at least for the 7 d.p.i. stage; for the bacterial invasion phase, one could expect an expression level similar to free-living bacteria. This is in particular true for the expression level of gene RL2817, the specialized *acpXL* acyl carrier protein, necessary for the synthesis of the VLCFA of the lipid A. A literature report (33) claims that the VLCFA content in the lipid A of nodule bacteria is increasing which seemingly does not match the finding of this study which finds a decreased level of *acpXL* expression at all stages of nodule development. This discrepancy needs further investigation. Most of the genes responsible for synthesis of the core oligosaccharide slightly decrease in their expression levels at all stages of nodule development. This finding seems to support the observation that nodule rhizobia form reduced amounts of LPS. Surprising was, however, the increased expression at 7 d.p.i. of gene RL0902, a kdtA gene which encodes a Kdo transferase. This finding corroborates another

observation, i.e. that a Kdo-containing epitope is formed in bacterial preparation isolated of nodules harvested at 7 d.p.i. (Kannenberg, unpublished). As this Kdo-containing epitope could be of importance to infection thread development, the increased expression of this gene warrants further genetical investigation. Biochemical evidence suggests that the LPS O-chain is modified in bacteria that occupy symbiosomes (32-33). Our finding here that a large chromosomal gene region, RL0794 to RL0826, involved in O-chain polysaccharide synthesis (34) shows overall a decrease in gene expression levels, seems to confirm the mentioned biochemical findings. On the other hand, the genes on plasmid pRL9 that are also likely to be involved in O-chain biosynthesis show increased levels of gene expression. Taken together, these gene expression data suggest a model for LPS O-chain synthesis in which the LPS O-chain displayed in free-living bacteria is modified or replaced by a new LPS O-chain in bacteroids. This exciting finding warrants future studies of LPS O-chain structure and function in bacteroids.

As FAs are components of the lipid A and the acyl pattern of the lipid A is modified during bacteroid development, the expression level of genes involved in fatty acid metabolism was also considered in this study. Furthermore, past studies have shown that in *R. leguminosarum* 3841 bacteroid development causes both the LPS and the overall bacteroid cell surface to become hydrophobic (33). While it is not clear whether the changes in bacteroid LPS are the cause if the increased bacteroid cell surface hydrophobicity, it is conceivable that the FAs of the lipid A are playing a role in this. In our study, we found a number of chromosomal and plasmid-borne genes involved in fatty acid biosynthesis or break-down are significantly up- or down-regulated. These genes will provide starting points for gene inactivation and genetical analysis and will help to unravel their importance in bacteroid development and functioning.

A long-standing observation of pea nodule development is that rhizobia once taken up into the symbiosomes underdo a number of developmental steps: they initially multiply, then duplicate their DNA without cell division and enlarge in size to become polyploid and polymorph in shape; eventually, they express the nitrogenase enzyme complex and stop cell growth altogether (21). These highly complex series of developmental steps have, to date, not been analyzed with regard to the role the PG is playing in all this. It is highly likely that the polymorphic shapes of rhizobial bacteroids are accompanied or caused by rearrangements in the peptidoglycan. Expression levels of genes involved in PG biosynthesis are probably playing an important role in this. Our data on the gene expression levels of the putative PG-related genes opens, therefore, a new field of investigation into these questions. The majority of the PG genes were either down-regulated or constitutively expressed during all stages of symbiotic pea nodule development. This down-regulation of PG synthesis genes is consistent with the fact that *R. leguminosarum* 3841 bacteroids cease cell division in the course of symbiotic development. More surprising was the down-regulation (which was severe for some genes) happened already at the 7 d.p.i. stage. Also noteworthy is that one gene related to PG biosynthesis, RL1878, a putative peptidoglycan binding protein; was highly up-regulated, suggesting some role in PG rearrangements. However, these observations are not yet understood and much remains to be learned about PG rearrangements in rhizobial bacteroids. The down- and up-regulated genes are good candidates for gene inactivation and genetical analysis for their role in PG rearrangements and function in *Rhizobium* bacteroid devolopment.

From the 1471 CRGs, the 1135 gene sample that could not be directly attributed to known compound classes was comprised of 296 CRGs that belong into the glycoconjugate metabolism, 511 genes were putative genes involved in unspecified transport processes, and 308

genes could not be assigned to any group of the above. Among the group of CRGs involved in glycoconjugate metabolism, we found only two genes up-regulated and 19 down-regulated genes at all four stages. Compared to the few up-regulated genes, the extensive down-regulation of these genes in nodule bacteria may reflect the restricted carbon supply from the host plant which primarily is believed to consist of C4-dicarboxylic acids such as succinate and malate (52), which may make an extensive glycoconjugates metabolism superfluous. Of particular note in this group of genes was the strongly down-regulated the IspL gene, i.e. RL3677 coding for a putative, a UDP-glucuronate-5'-epimerase, that converts UDP-GlcA to UDP-L-IdoA (iduronic acid). To date, a compound containing or derived from L-IdoA has not been described for *R. leguminosarum* bv. *viciae*, indicating that our knowledge of the *R. leguminosarum* glycome is not complete. As extensively described above, rhizobia experience in the soil and in the plant host very different microenvironments. To grow and dwell in all these environments they have to able to adapt to these environmental conditions. This formidable task to be able to grow and survive under all these conditions may help to explain why a good third of all CRGs belong to transport systems. In particular the number of up-regulated genes during nodule development (8 were up-regulated with 6 belonging to ABC transport systems) will provide starting points for gene inactivation and functional analysis of the involved transport systems to work out their role in carbohydrate transport and nodule development. Among the group of CRGs that could not be assigned to any function, a number of genes were up- or down-regulated. However, only gene inactivation and future research into the function of these genes will shed light into the importance of these genes for carbohydrate formation and symbiotic development.

In summary, this comprehensive study on the overall gene expression patterns of all identifiable CRGs of *R. leguminosarum* strain 3841during pea nodule development confirmed

for well known classes of molecules, e.g. the genes involved in Nod factor formation or the nitrogenase complex, our preconceived expectations for their expression levels. Moreover, CRGs and genes clusters were identified that were up- or down-regulated during nodule development. These carbohydrate-related genes with up- or down-regulated gene expression will provide valuable leads for future gene inactivation experiments and functional analysis of their role in carbohydrate formation and pea nodule development. Despite short comings of microarray data experiments and its data processing and interpretation procedures, and despite problems stemming from the complexity of symbiotic nodule formation, it can be concluded that microarray data corroborated through the statistical approach used in this study provide valid insights into gene expression of all cell carbohydrate-related genes as well as for the whole genome. We believe that this study is a step forward in the understanding of the expression and overall trends of the *Rhizobium* glycome during nodule development and symbiosis.

REFERENCES

1. **Ausmees, N., H. Jonsson, S. H"glund, H. Ljunggren, and M. Lindberg.** 1999. Structural and putative regulatory genes involved in cellulose synthesis in Rhizobium leguminosarum bv. trifolii Microbiology **145:**1253-1262.
2. **Becker, A., N. Fraysse, and L. Sharypova.** 2005. Recent Advances in Studies on Structure and Symbiosis-Related Function of Rhizobial K-Antigens and Lipopolysaccharides. Molecular Plant-Microbe Interactions **18:**899-905.
3. **Bhat, U. R., R. W. Carlson, M. Busch, and H. Mayer.** 1991. Distribution and phylogenetic significance of 27-hydroxy- octacosanoic acid in lipopolysaccharides from bacteria belonging to the alpha-2 subgroup of Proteobacteria. International Journal of Systematic Bacteriology **41:**213-217.
4. **Breedveld, M. W., and K. J. Miller.** 1994. Cyclic á-glucans of members of the family Rhizobiaceae Microbiological Reviews **58:**145-161.
5. **Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat.** 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Research **37:**D233-D238.
6. **Carlson, R. W., L. S. Forsberg, and E. L. Kannenberg.** 2010. Lipopolysaccharides in *Rhizobium*-legume symbioses, p. 339-386. *In* P. J. Quinn and X. Wang (ed.), Subcellular Biochemistry: Endotoxins: Structure, function and recognition., vol. 53. Springer.

7. **Carlson, R. W., L. S. Forsberg, and E. L. Kannenberg.** 2010. Lipopolysaccharides in *Rhizobium*-legume Symbiosis, p. 339-386. *In* P. J. Quinn and X. Wang (ed.), Endotoxin: Structure, function and recognition, vol. 53. Springer, London.

8. **Castro, C. D., A. Molinaro, R. Lanzetta, A. Silipo, and M. Parrilli.** 2008. Lipopolysaccharide structures from *Agrobacterium* and *Rhizobiaceae* species. Carbohydrate Research **343:**1924-1933.

9. **Csonka, L. N.** 1898. Physiological and genetic responses of bacteria to osmotic stress. Microbiological Reviews **53:**121-147.

10. **D'Antuono, A. L., T. Ott, L. Krusell, V. Voroshilova, R. A. Ugalde, M. Udvardi, and V. C. Lepek.** 2008. Defects in Rhizobial Cyclic Glucan and Lipopolysaccharide Synthesis Alter Legume Gene Expression During Nodule Development. Molecular Plant-Microbe Interactions **21:**50-60.

11. **D'Haeze, W., C. Leoff, G. Freshour, K. D. Noel, and R. W. Carlson.** 2007. Rhizobium etli CE3 Bacteroid Lipopolysaccharides Are Structurally Similar but Not Identical to Those Produced by Cultured CE3 Bacteria. J. Biol. Chem. **282:**17101-17113.

12. **de Maagd, R. A., A. S. Rao, I. H. M. Mulders, L. Goosen-de Roo, M. C. M. van Loosdrecht, C. A. Wijffelman, and B. J. J. Lugtenberg.** 1989. Isolation and characterization of mutants of Rhizobium leguminosarum bv. viciae 248 with altered lipopolysaccharides: Possible role of surface charge or hydrophobicity in bacterial release from the infection thread. The Journal of Bacteriology **171:**1143-1150.

13. **Delauney, A. J., C. I. Cheon, P. J. Snyder, and D. P. Verma.** 1990. A noduel-specific sequence encoding a methionine-rich polypeptide, nodulin-21. Plant Molecular Biology **14:**449-451.

14. **Delpino, M. V., D. J. Comerci, M. A. Wagner, M. Eschenbrenner, C. V. Mujer, R. A. Ugalde, C. A. Fossati, P. C. Baldi, and V. G. Delvecchio.** 2009. Differential composition of culture supernatants from wild-type Brucella abortus and its isogenic virB mutants. Arch Microbiol **191:**571-581.

15. **Delpino, M. V., M. I. Marchesini, S. M. Estein, D. J. Comerci, J. Cassataro, C. A. Fossati, and P. C. Baldi.** 2007. A Bile Salt Hydrolase of *Brucella abortus* Contributes to the Establishment of a Successful Infection through the Oral Route in Mice. Infection and Immunity **75:**299-305.

16. **Downie, J. A.** 2010. The role of extracellular proteins, polysaccharides and signasl in the interactions of rhizobia with legume roots. FEMS Microbiology Reviews **34:**150-170.

17. **Finnie, C., A. Zorreguieta, N. M. Hartley, and J. A. Downie.** 1998. Characterization of Rhizobium leguminosarum exopolysaccharide glycanases that are secreted via a type I exporter and have a novel heptapeptide repeat motif. Journal of Bacteriology **180:**1691-1699.

18. **Forsberg, L. S., and R. W. Carlson.** 2008. Structural characterization of the primary O-antigenic polysaccharide of the *Rhizobium leguminosarum* 3841 lipopolysaccharide and Identification of a new 3-acetimidoylamino-3-deoxyhexuronic acid glycosyl component: A unique O-methylated glycan of uniform size containing 6-deoxy-3-O-methyl-D-talose, N-acetylquinovosamine, and rhizoaminuronic acid (3-acetimidoylamino-3-deoxy-D-*gluco*-hexuronic acid). J. Biol. Chem. **283:**16037-16050.

19. **Fraysse, N., F. Couderc, and V. Poinsot.** 2003. Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. FEBS Journal **270:**1365-1380.

20. **Fujishige, N. A., M. R. Lum, P. L. De Hoff, J. P. Whitelegge, K. F. Faull, and A. M. Hirsch.** 2008. *Rhizobium* common nod genes are required for biofilm formation. Molecular Microbiology **67:**504-515.

21. **Gage, D. J.** 2004. Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. Microbiol. Mol. Biol. Rev. **68:**280-300.

22. **Gibson, K. E., H. Kobayashi, and G. C. Walker.** 2008. Molecular Determinants of a Symbiotic Chronic Infection. Annual Review of Genetics **42:**413-441.

23. **Glenn, A. R., P. S. Poole, and J. F. Hudman.** 1980. Succinate uptake by free-living and bacteroid forms of *Rhizobium leguminosarum*. Journal of General Microbiology **119:**267-271.

24. **Hellriegel, H., and H. Wilfarth.** 1888. Untersuchungen über die Stickstoffnahrung der Gramineen und Leguminosen. [Studies on the nutrient nitrogen in Gramineae and legumes]. Vereins Rübenzucker-Industrie des Deutschen Reiches**:**1-234.

25. **Hertz-Fowler, C., C. S. Peacock, V. Wood, M. Aslett, A. Kerhornou, P. Mooney, A. Tivey, M. Berriman, N. Hall, K. Rutherford, J. Parkhill, A. C. Ivens, M. Rajandream, and B. Barrell.** 2003. GeneDB: a resource for prokaryotic and eukaryotic organisms. Nucleic Acids Research **32:**D339-D343.

26. **Janczarek, M., J. Jaroszuk-Ściseł, and A. Skorupska.** 2009. Multiple copies of rosR and pssA genes enhance exopolysaccharide production, symbiotic competitiveness and clover nodulation in Rhizobium leguminosarum bv. trifolii. Antonie van Leeuwenhoek **96:**471-486.

27. **Johnston, A. W., and J. E. Beringer.** 1975. Identification of rhizobium strains in pea root nodules using genetic markers. Journal of General Microbiology **87:**343-350.

28. **Kanehisa, M., and S. Goto.** 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research **28:**27-30.

29. **Kanehisa, M., S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa.** 2006. From genomics to chemical genomics: New developments in KEGG. Nucleic Acids Research **34:**D354-D357.

30. **Kanjilal-Kolar, S., S. S. Basu, M. I. Kanipes, Z. Guan, T. A. Garrett, and C. R. H. Raetz.** 2006. Expression Cloning of Three Rhizobium leguminosarum Lipopolysaccharide Core Galacturonosyltransferases. Journal of Biological Chemistry **281:**12865-12878.

31. **Kannenberg, E. L., and N. J. Brewin.** 1994. Host-plant invasion by Rhizobium: the role of cell-surface components. Trends in Microbiology **2:**277-283.

32. **Kannenberg, E. L., N. J. Brewin, R. W. Carlson, G. B. Kiss, and G. Endre.** 1994. Biochemical Separation of Lipopolysaccharides Expressed in Rhizobium Nodule Bacteria, p. 320-320, Proceedings of the 1st European Nitrogen Fixation Conference. Officina Press, Szeged, Hungary.

33. **Kannenberg, E. L., and R. W. Carlson.** 2001. Lipid A and O-chain modifications cause *Rhizobium* lipopolysaccharides to become hydrophobic during bacteroid development. Molecular Microbiology **39:**379-392.

34. **Kannenberg, E. L., S. Perotto, V. Bianciotto, E. A. Rathburn, and N. J. Brewin.** 1994. Lipopolysaccharide epitope expression of Rhizobium bacteroids as revealed by in

situ immunolabelling of pea root nodule sections. The Journal of Bacteriology **176:**2021-2032.

35. **Kannenberg, E. L., B. L. Reuhs, L. S. Forsberg, and R. W. Carlson.** 1998. Lipopolysaccharide and K-antigens: their structures, biosynthesis, and functions, p. 120-154. *In* H. P. Spaink, A. Kondorosi, and J. J. Hooykass (ed.), The Rhizobiaceae. Kluwer Academic Publishers, Boston.

36. **Karunakaran, R., V. K. Ramachandran, J. C. Seaman, A. K. East, B. Mouhsine, T. H. Mauchline, J. Prell, A. Skeffington, and P. S. Poole.** 2009. Transcriptomic analysis of Rhizobium leguminosarum Biovar viciae in symbiosis with host plants Pisum sativum and Vicia cracca. Journal of Bacteriology **191:**4002-4014.

37. **Krol, E., and A. Becker.** 2009. Surface polysaccharides as fitness factors of rhizospheric nitrogen-fixing bacteria, p. 189-211. *In* M. Ullrich (ed.), Bacterial polysaccharides: current innovations and future trends. Caister Academic Press, Norfolk.

38. **Król, J. E., A. Mazur, M. Marczak, and A. Skorupska.** 2007. Syntenic arrangements of the surface polysaccharide biosynthesis genes in Rhizobium leguminosarum. Genomics **89:**237-247.

39. **Latchford, J. W., D. Borthakur, and A. W. B. Johnston.** 1991. The products of Rhizobium genes, psi and pss , which affect exopolysaccharide production, are associated with the bacterial cell surface. Mol.Microbiol. **5:**2107-2114.

40. **Laus, M. C., T. J. Logman, G. E. Lamers, A. A. N. Van Brussel, R. W. Carlson, and J. W. Kijne.** 2006. A novel polar surface polysaccharide from Rhizobium leguminosarum binds host plant lectin. Molecular Microbiology **59:**1704-1713.

41. **Laus, M. C., A. A. N. van Brussel, and J. W. Kijne.** 2005. Role of Cellulose Fibrils and Exopolysaccharides of Rhizobium leguminosarum in Attachment to and Infection of Vicia sativa Root Hairs. Molecular Plant-Microbe Interactions **18:**533-538.

42. **Lepek, V. C., and A. L. D. D'Antouno.** 2005. Bacterial surface polysaccharides and their role in the rhizobia-legume association. Lotus Newsletter **35:**93-105.

43. **Lerouge, I., and J. Vanderleyden.** 2001. O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. FEMS Microbiology Reviews **26:**17-47.

44. **Li, L., O. S. Chen, D. M. Ward, and J. Kaplan.** 2001. CCC1 is a transporter that mediates vacuolar iron storage in yeast. Journal of Biological Chemistry **276:**29515-29519.

45. **Lugtenberg, B. J. J.** 1998. Outer membrane proteins. *In* H. P. Spaink, A. Kondorosi, and P. J. J. Hooykass (ed.), The Rhizobiaceae. Kluwer Academic Publishers, Boston.

46. **Marczak, M., A. Mazur, W. I. Gruszecki, and A. Skorupska.** 2008. PssO, a unique extracellular protein important for exopolysaccharide synthesis in Rhizobium leguminosarum bv. trifolii. Biochimie **90:**1781-1790.

47. **Marczak, M., A. Mazur, J. E. Krol, W. I. Gruszecki, and A. Skorupska.** 2006. Lipoprotein PssN of Rhizobium leguminosarum bv. trifolii: Subcellular Localization and Possible Involvement in Exopolysaccharide Export. J. Bacteriol. **188:**6943-6952.

48. **McIntyre, H. J., H. Davies, T. A. Hore, S. H. Miller, J.-P. Dufour, and C. W. Ronson.** 2007. Trehalose Biosynthesis in Rhizobium leguminosarum bv. trifolii and Its Role in Desiccation Tolerance. Appl. Environ. Microbiol. **73:**3984-3992.

49. **Miller, K. J., E. P. Kennedy, and V. N. Reinhold.** 1986. Osmotic adaptation by gram-negative bacteria: possible role for periplasmic oligosaccharides. Science **231:**48-51.

50. **Nizet, V., and J. D. Esko.** 2009. Bacterial and Viral Infections, p. 537-552. *In* A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler (ed.), Essentials of Glycobiology. Cold Spring Laboratory Press, New York.

51. **Philip-Hollingsworth, S., R. I. Hollingsworth, and F. B. Dazzo.** 1989. Host-range related structural features of the acidic extracellular polysaccharides of Rhizobium trifolii and Rhizobium leguminosarum Journal of Biological Chemistry **264:**1461-1466.

52. **Prell, J., and P. S. Poole.** 2006. Metabolic changes of rhizobia in legume nodules. Trends Microbiol **14:**161-168.

53. **Reuber, T. L., S. Long, and G. C. Walker.** 1991. Regulation of Rhizobium meliloti exo genes in free-living cells and in planta examined by using Tn phoA fusions. The Journal of Bacteriology **173:**426-434.

54. **Riley, M.** 1993. Functions of the Gene Products of *Escherichia coli*. Microbiological Reviews **57:**862-952.

55. **Robledo, M., J. I. JimÃ©nez-Zurdo, E. VelÃ¡zquez, M. E. Trujillo, J. L. Zurdo-PiÃ±eiro, M. H. RamÃ-rez-Bahena, B. Ramos, J. M. DÃ-az-MÃ-nguez, F. Dazzo, E. MartÃ-nez-Molina, and P. F. Mateos.** 2008. Rhizobium cellulase CelC2 is essential for primary symbiotic infection of legume host roots. Proceedings of the National Academy of Sciences **105:**7064-7069.

56. **Roo, L. G., R. A. Maagd, and B. J. Lugtenberg.** 1991. Antigenic changes in lipopolysacharide I of Rhizobium leguminosarum bv. viciae in root nodules of Vicia

sativa subsp. nigra occur during release from infection threads. Journal of Bacteriology **173:**3177-3183.

57. **Russo, D. M., A. Williams, A. Edwards, D. M. Posadas, C. Finnie, M. Dankert, J. A. Downie, and A. Zorreguieta.** 2006. Proteins Exported via the PrsD-PrsE Type I Secretion System and the Acidic Exopolysaccharide Are Involved in Biofilm Formation by Rhizobium leguminosarum. J. Bacteriol. **188:**4474-4486.

58. **Smith, G., S. Swart, B. J. Lutenberg, and J. W. Kijne.** 1992. Molecular mechanisms of attachment of Rhizobium bacteria to plant roots. Molecular Microbiology **6:**2897-2903.

59. **Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, Y. I. Wolf, J. J. Yin, and D. A. Natale.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**1-14.

60. **van Berkum, P., and B. D. Eardly.** 1998. Molecular evolutionary systematic of the Rhizobiaceae, p. 1-24. *In* H. P. Spaink, A. Kondorosi, and P. J. J. Hooykass (ed.), *The Rizobiaceae*. Kluwer Academic Publishers, Boston.

61. **Van Rhijn, P., and J. Vanderleyden.** 1995. The Rhizobium -plant symbiosis. Microbiological Reviews **59:**124-142.

62. **Vanderlinde, E. M., A. Muszynski, J. J. Harrison, S. F. Koval, D. L. Foreman, H. Ceri, E. L. Kannenberg, R. W. Carlson, and C. K. Yost.** 2009. Rhizobium leguminosarum biovar viciae 3841, deficient in 27-hydroxyoctacosanoate-modified lipopolysaccharide, is impaired in desiccation tolerance, biofilm formation and motility. Microbiology **155:**3055-3069.

63. **Vedam, V., J. G. Haynes, E. L. Kannenberg, R. W. Carlson, and D. J. Sherrier.** 2004. A Rhizobium leguminosarum lipopolysaccharide lipid A mutant induces nitrogen-fixing nodules with delayed and defective bacteroid formation. Mol.Plant-Microbe Interact. **17:**283-291.

64. **Weber, A., S. A. Kogl, and K. Jung.** 2006. Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in *Escherichia coli*. Journal of Bacteriology **188:**7165-7175.

65. **Wielbo, J., A. Mazur, J. E. Krol, M. Marczak, and A. Skorupska.** 2004. Environmental modulation of the pssTNOP gene expression in Rhizobium leguminosarum bv. trifolii. Canadian Journal of Microbiology **50:**201-211.

66. **Williams, A., A. Wilkinson, M. Krehenbrink, D. M. Russo, A. Zorreguieta, and J. A. Downie.** 2008. Glucomannan-Mediated Attachment of Rhizobium leguminosarum to Pea Root Hairs Is Required for Competitive Nodule Infection. J. Bacteriol. **190:**4706-4715.

67. **Young, J. P., L. Crossman, A. Johnston, N. Thomson, Z. Ghazoui, K. Hull, M. Wexler, A. Curson, J. Todd, P. Poole, T. Mauchline, A. East, M. Quail, C. Churcher, C. Arrowsmith, I. Cherevach, T. Chillingworth, K. Clarke, A. Cronin, P. Davis, A. Fraser, Z. Hance, H. Hauser, K. Jagels, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M. Simmonds, S. Whitehead, and J. Parkhill.** 2006. The genome of Rhizobium leguminosarum has recognizable core and accessory components. Genome Biology **7:**R34.

68. **Zevenhuizen, L. P. T. M., and A. R. W. van Neerven.** 1983. Gel-forming capsular polysacchrides of Rhizobium leguminosarum and Rhizobium trifolii Carbohydrate Research **124:**166-171.

69. **Zorreguieta, A., C. Finnie, and J. A. Downie.** 2000. Extracellular glycanases of Rhizobium leguminosarum are activated on the cell surface by an exopolysaccharide-related component. Journal of Bacteriology **182:**1304-1312.