

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**ISOLATION AND DNA SEQUENCE ANALYSIS OF
A *RHIZOBIUM LOTI* GENE REQUIRED FOR
EFFECTIVE NODULATION OF
*LOTUS PEDUNCULATUS***

A thesis presented in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy in Microbiology at
Massey University, Palmerston North, New Zealand

LAWRENCE JAMES HENRY WARD
1989

FRONTISPIECE

Transverse section through a swollen infection thread in a *Lotus pedunculatus* nodule. The plant was inoculated with the *Rhizobium loti* mutant strain PN239.

Magnification approximately 3500 times

Photograph courtesy of Clive Pankhurst



ERRATA

- p4 line 17 'Spanik' (twice) should be 'Spaink'
- p11 line 7 'is indicated' should be 'is indicated by X'
- p17 line 19 'Sodium chloride, 0.5' should be 'Sodium chloride, 5.0'
- p18 line 14 '*lacI*⁹' should be '*lacI*^q'
- p20 line 17 ' $\text{MnCl}_2 \cdot 4\text{Cl}_2\text{O}$ ' should be ' $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ '
- p22 line 21 'Tn5' should be 'Tris'
- p27 line 12 '0.5 $\mu\text{g/ml}$ ' should be '0.5 mg/ml '
- p33 line 14 'BC1G' should be 'BCIG'
- p36 line 1 'Whitfield' should be 'Whitfeld'
- p51 line 4 '*bar*' should be '*bar*'
- p54 line 6 '(2 $\mu\text{l/ml}$)' should be '(2 $\mu\text{g/ml}$)'
- p54 line 21 Insert '(Fig. 14)' after '*Xho*I'
- p66 Fig. 18 Add 'Vertical lines indicate the location of Tn5 insertions which result in a Fix^+ (longer vertical line) and Fix^- (shorter vertical line) phenotype on *Lotus pedunculatus*'
- p76 Fig. 22 '(arrowed)' should be '(two examples of which are arrowed)'
- p77 Fig. 23 'the sequence read from them (B)' should be 'part of the sequence read from them (B)'
Add 'The sequences given in B (left to right) read from bottom to top in the segments of gels from the + strand and from top to bottom in the segments of gels from the - strand. Due to poor reproduction not all bands are readily visible.'
- p81 line 7 'cosmids' should be 'plasmids'
- p81 line 33 'base repeat' should be 'base direct repeat'
- p83 Fig. 26 'sequence of two Tn5 - *Rhizobium* junctions.' should be 'sequences of Tn5 - *Rhizobium* DNA junctions from two different mutants.'

ABSTRACT

A *Rhizobium loti* gene required for effective nodulation of the host *Lotus pedunculatus* has been identified by transposon Tn5 mutagenesis. Cosmids from a *R. loti* gene library which complemented a previously isolated mutant strain, PN239, (Chua *et al* 1985; J. Bacteriol. 162; 335-343) at this locus were identified by *in planta* complementation. A physical map of these cosmids was constructed and the site of insertion of the Tn5 was mapped to a 7.5 kb *Eco*RI fragment common to all cosmids which complemented the mutation. This 7.5 kb *Eco*RI fragment was subcloned into pBR328 and pLAFR1 and a more detailed physical map constructed. The 7.5 kb *Eco*RI fragment in pLAFR1 was able to complement the Tn5 mutation when introduced into strain PN239.

Further Tn5 mutagenesis of the 7.5 kb *Eco*RI fragment was carried out in *E. coli* and the mutations were homogenotised into *R. loti* NZP2037. Three additional mutations were isolated which caused a Fix⁻ phenotype on *Lotus pedunculatus*. The Tn5 inserts which caused a Fix⁻ phenotype were mapped to positions adjacent to the position of the original mutation in strain PN239. All other Tn5 insertions isolated in the 7.5 kb *Eco*RI fragment gave a Fix⁺ phenotype on *Lotus pedunculatus*.

A region was sequenced which was involved in effective nodulation of *Lotus pedunculatus* as indicated by the position of the Tn5 insertions. Analysis of the consensus sequence of 2307 bases identified a potential open reading frame (ORF) of 576 base pairs, coding for a putative protein of 21.2 kD. The positions of the Tn5 insertions causing a Fix⁻ phenotype and the adjacent Tn5 insertions which did not affect fixation were determined in the sequence. The position and orientation of the ORF identified was consistent with the sequenced positions of these Tn5 insertions.

A fragment containing most of the ORF identified from the sequence was used as a hybridization probe to various strains of rhizobia. Homology was only demonstrated with DNA from other *R. loti* strains. *R. loti* strains containing Tn5 insertions which were Fix⁻ on *Lotus pedunculatus* were found to be fully effective on *Lotus corniculatus*. These observations suggest that the gene characterised in this investigation may be involved in the host specificity of *R. loti* for *Lotus pedunculatus*.

ACKNOWLEDGEMENTS

I wish to thank my supervisors, Assoc. Prof. B.D.W. Jarvis and Prof. D.B. Scott for their encouragement and guidance throughout this project.

I would also like to thank Prof. D.F. Bacon for his interest and encouragement.

Thanks are also due to the academic, technical and secretarial staff (past and present) of the Department of Microbiology and Genetics for their help and interest; in particular Assoc. Prof. T.J. Brown for use of space in his laboratory and Dr B. Mansfield for assistance with the Department's computers.

The co-operation, assistance and encouragement of my fellow postgraduate students throughout the course of this project has been appreciated (I'm not going to mention names - I'm sure to miss somebody!).

Thanks are also due to the staff of the DSIR computer unit and to Mr Mark Pritchard (DSIR) for help with the VAX computer and Staden programs and to the staff of the Massey University Computer Centre, in particular Mr Glen Eustace, for their co-operation in transferring data and assistance with using the Prime computer.

I would also like to thank my friends and colleagues at the N.Z. Dairy Research Institute for their interest and encouragement during the final stages of this thesis.

The co-operation and assistance of Mr Paul Le Ceve (DRI) and Mrs Veronica Fieldsend with the photographs and typing respectively is gratefully acknowledged.

TABLE OF CONTENTS

	PAGE
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	vi
LIST OF TABLES	xi
LIST OF FIGURES	xiii
 <u>INTRODUCTION</u>	 1
1.1 Biological nitrogen fixation and its importance	1
1.2 Significance of <i>Lotus</i> spp. in New Zealand agriculture	2
1.3 Characteristics of rhizobia which nodulate <i>Lotus</i> spp.	2
1.4 Nodule development	3
1.4.1 Nodule formation in <i>Lotus</i>	7
1.5 Genetics of nodulation and nitrogen fixation	8
1.5.1 Nodulation genes	10
1.5.2 Nitrogen fixation genes	13
1.6 Background and aims of this investigation	15
 <u>MATERIALS AND METHODS</u>	 17
2.1 Bacterial strains and plasmids	17
2.2 Growth of bacteria	17
2.3 Media	17
2.3.1 Luria broth (LB)	17
2.3.2 TY medium	17
2.3.3 M9 medium	20
2.3.4 Thornton's medium	20
2.3.5 Top agar	20
2.4 Buffers and solutions	21
2.4.1 Electrophoresis (E) buffer	21
2.4.2 Tris-Borate - EDTA (TBE) buffer	21
2.4.3 Standard saline citrate (SSC)	21
2.4.4 Phenol/chloroform solution	21

Table of Contents Cont'd.

	PAGE
2.4.5 STET buffer	21
2.4.6 Solutions for plasmid preparation	22
2.4.6.1 Solution I	22
2.4.6.2 Solution II	22
2.4.6.3 Solution III	22
2.4.7 <i>Hae</i> III buffer	22
2.4.8 SDS dye mixture	22
2.4.9 TE buffer	22
2.4.10 Scintillation fluid	22
2.4.11 TEC buffer	22
2.4.12 Hybridization buffer	23
2.4.13 Formamide dye mixture	23
2.5 Plant nodulation tests	23
2.5.1 Reisolation of bacteria from nodules	23
2.6 Isolation of total DNA from <i>Rhizobium</i>	24
2.7 Plasmid isolation methods	24
2.7.1 Rapid boiling method	24
2.7.2 Cosmid miniprep method	25
2.7.3 Preparative plasmid isolation	26
2.8 Purification of DNA samples by extraction with phenol/chloroform	27
2.9 Precipitation of DNA with ethanol	27
2.10 Determination of DNA concentration and purity	28
2.11 Restriction endonuclease digests	28
2.12 Horizontal gel electrophoresis	29
2.13 Recovery of DNA fragments from agarose gels	29
2.13.1 U tube electroelution	29
2.13.2 Electroelution from a dialysis sac	31
2.13.3 Phenol freeze extraction	31
2.14 DNA ligations	32
2.15 Preparation and transformation of competent cells	32
2.15.1 <i>E. coli</i> strain HB101	32
2.15.2 <i>E. coli</i> strain JM101	33

Table of Contents Cont'd.

	PAGE
2.16 Southern blot technique	33
2.17 Preparation of random primers	36
2.18 Preparation of ^{32}P -labelled probe DNA	35
2.19 Hybridization of probe DNA to Southern blots	36
2.20 DNA sequencing	39
2.20.1 Preparation of sequencing templates	39
2.20.2 Annealing template and primer DNA	40
2.20.3 Labelling the annealed template DNA	40
2.20.4 Preparation of d/ddNTP stock solutions	41
2.20.5 Acrylamide gel electrophoresis of the products of the sequencing reaction	43
<u>RESULTS</u>	45
3.1 Construction of a physical map of the cosmids complementing the Bar mutation	45
3.2 Subcloning of the common 7.5 kb <i>EcoRI</i> fragment	50
3.3 Complementation of the Bar mutation by the subcloned 7.5 kb <i>EcoRI</i> fragment	52
3.4 Restriction enzyme mapping of the 7.5 kb <i>EcoRI</i> fragment	54
3.5 Mapping the position of the Tn5 insert (PN239) in the 7.5 kb <i>EcoRI</i> fragment	61
3.5.1 Insertion of further Tn5's into the 7.5 kb <i>EcoRI</i> fragment and mapping the position of these insertions	65
3.6 Hybridization of restriction fragments from the <i>bar</i> region to total genomic blots of DNA from other rhizobia	67
3.6.1 Hybridization of the 7.5kb <i>EcoRI</i> fragment from pPN318	67
3.6.2 Hybridization of the 0.5kb <i>SphI</i> - <i>AluI</i> fragment containing the <i>bar</i> gene	69

Table of Contents Cont'd.

	PAGE
3.7 Subcloning fragments from the 7.5kb <i>EcoRI</i> fragment into M13 vectors for sequencing	70
3.8 Sequencing of the subcloned fragments and assembly into a contiguous sequence	72
3.9 Resolution of compressions observed in the sequence	75
3.9.1 Reverse transcriptase	75
3.9.2 40% formamide 7M urea gels	75
3.10 Determination of the position of Tn5 insertions within the sequence	81
3.10.1 Subcloning Tn5:: <i>Rhizobium</i> DNA junction fragments into M13mp9	81
3.10.2 Sequencing the Tn5:: <i>Rhizobium</i> DNA junction fragments	81
3.11 Computer analysis of the sequence data	84
3.11.1 Open reading frames	84
3.11.2 TestCode analysis	84
3.11.3 Third position compositional bias	87
3.11.4 Construction of a <i>Rhizobium</i> codon usage table	91
3.11.5 Codon preference	97
3.11.6 Rare codon usage	98
3.12 Features of the sequence	106
3.12.1 A potential ribosome binding site	106
3.12.2 Possible promoter region for the gene	107
3.12.3 Possible termination structures	109
3.12.4 Direct and inverted repeats upstream from the presumed ATG initiation codon for the gene	110
3.12.5 Putative protein translation	111
3.12.6 Comparison of the nucleic acid and protein sequences to the databases	113

Table of Contents Cont'd.

	PAGE
<u>DISCUSSION</u>	115
4.1 The bacterial release (Bar) mutant	116
4.2 Complementation of the Bar mutant	116
4.3 Mapping the cosmids complementing the Bar mutant	117
4.4 The common 7.5kb <i>Eco</i> RI fragment	117
4.5 Location of Tn5 Insertions	118
4.6 DNA sequencing	119
4.7 Sequence compressions	121
4.8 Sequencing the position of Tn5 insertions	122
4.9 Analysis of the sequence	123
4.9.1 Open reading frames	123
4.9.2 TestCode	123
4.9.3 Third position compositional bias	124
4.9.4 Codon preference	124
4.9.5 Rare codon usage	126
4.10 Additional evidence for the <i>bar</i> gene	126
4.11 A potential ribosome binding site	127
4.12 A possible promoter region for the gene	127
4.13 Direct and inverted repeats	128
4.14 Termination structures	128
4.15 Amino acid translation of the sequence	128
4.16 Possible function of the gene	129
<u>CONCLUSIONS</u>	131
<u>REFERENCES</u>	132
<u>APPENDIX</u>	156
Paper published:	
Ward <i>et al.</i> , 1989	
Molecular Plant-Microbe Interactions 2: 224-232	

LIST OF TABLES

TABLE	PAGE
1 Proposed steps in <i>Rhizobium</i> -legume symbiosis.	9
2 Bacterial strains and plasmids used in this investigation.	18
3 Deoxynucleotide triphosphate mixtures.	41
4 Dideoxynucleotide triphosphate mixtures.	42
5 Mixture of dideoxynucleotides used for sequencing with DNA polymerase I.	42
6 Deoxy/dideoxynucleotide triphosphate mixture used for sequencing with reverse transcriptase.	43
7 Fragment sizes obtained from cosmids pPN318, pPN319 and pPN320 digested with <i>Eco</i> RI and <i>Hind</i> III.	49
8 Fragment sizes which result from digestion of the 9.2 kb <i>Hind</i> III fragment.	56
9 Hybridization of four restriction fragments from the 7.5 kb <i>Eco</i> RI fragment with double and triple digests of pPN28.	58
10 Fragments cloned into M13 vectors for determination of the DNA sequence.	71
11 Sequences used in the construction of <i>Rhizobium</i> codon usage tables.	92

List of Tables Cont'd.

TABLE	PAGE
12 Codon usage table constructed from the <i>Rhizobium</i> genes listed in Table 11.	95
13 CORRESPOND analysis of codon frequency tables.	97
14 Comparison of the ribosome binding site (RBS) "rules" proposed by Stormo <i>et al.</i> , (1982) and the putative RBS for the <i>bar</i> gene.	107
15 Amino acid composition of the putative protein.	111

LIST OF FIGURES

FIGURE		PAGE
1	Arrangement and direction of transcription of nodulation genes.	11
2	Glass "U" tube apparatus for recovery of DNA fragments from agarose gels.	30
3	Apparatus for the transfer of DNA from agarose gels to nitrocellulose filters.	35
4	"Minispin" column apparatus.	38
5	Exploded view of assembly of filters, film and intensifying screens in an X-ray cassette.	38
6	Fragments obtained from the digestion of cosmid pPN318 hybridized to probes from pPN318 and pPN320.	46
7	Fragments obtained from the digestion of cosmid pPN319 hybridized to probes from pPN318 and pPN320.	47
8	Fragments obtained from the digestion of cosmid pPN320 hybridized to probes from pPN318 and pPN320.	48
9	Physical map of the <i>bar</i> gene region of <i>R. loti</i> strain NZP2037.	51

List of Figures Cont'd.

FIGURE		PAGE
10	Agarose gel electrophoresis of <i>Eco</i> RI digests of A; the cosmids which complemented the <i>Bar</i> ⁻ mutant and B; plasmids into which the common 7.5kb <i>Eco</i> RI fragment was subcloned from pPN318.	53
11	<i>Lotus pedunculatus</i> plants inoculated with <i>R. loti</i> strains.	55
12	Fragments electroeluted from a <i>Xho</i> I/ <i>Pst</i> I/ <i>Eco</i> RI triple digest of pPN28.	57
13	Mapping of the 7.5kb <i>Eco</i> RI fragment by probing restriction endonuclease digests of pPN28 with fragments obtained from a <i>Xho</i> I/ <i>Pst</i> I/ <i>Eco</i> RI triple digest of pPN28.	59
14	Physical map of the 7.5 kb <i>Eco</i> RI fragment containing the <i>bar</i> region of <i>R. loti</i> strain NZP2037.	60
15	Hybridization of the 7.5kb <i>Eco</i> RI fragment from pPN28 to total genomic DNA from strains NZP2037 and PN239.	62
16	Relationship of restriction fragments in <i>R. loti</i> strain NZP2037 and strain PN239.	63
17	Determination of the position of the Tn5 insert in strain PN239.	64
18	The position of Tn5 insertions isolated in the 7.5kb <i>Eco</i> RI fragment.	66

List of Figures Cont'd.

FIGURE		PAGE
19	Hybridization of the <i>bar</i> gene to total genomic DNA from various rhizobia.	68
20	The region of M13mp8 and M13mp9 used for cloning fragments for sequencing.	73
21	Strategy for sequencing the <i>bar</i> region of <i>R. loti</i> strain NZP2037 showing the location and direction of templates sequenced and used to construct the consensus sequence.	74
22	Resolution of sequence compressions using reverse transcriptase in the sequencing reaction.	76
23	Resolution of sequencing compressions using formamide gels.	77
24	The consensus sequence of the <i>bar</i> gene region.	78
25	Sequence mapping of Tn5 inserts in the <i>bar</i> locus.	82
26	DNA sequence of <i>Rhizobium</i> ::Tn5 junction fragments.	83
27	Output plot from the UWGCG program FRAMES indicating the open reading frames in both orientations of the consensus sequence.	85

List of Figures Cont'd.

FIGURE		PAGE
28	Output plot from the UWGCG program TESTCODE for the consensus sequence in the 5' to 3' direction.	86
29	Third position GC bias plots of the consensus sequence (5' to 3' direction).	88
30	Output from the UWGCG program CODON PREFERENCE on the sequence in the 5' to 3' direction; (A) using the codon table Ecohigh.cod; (B) using the codon table Ecolow.cod; and (C) using the codon table Rhizall.cod.	99
31	Output plot from the UWGCG program CODON PREFERENCE on the sequence in the 3' to 5' direction, (A) using the codon table Ecohigh.cod; and (B) using the codon table Rhizall.cod.	102
32	Output from the UWGCG program CODON PREFERENCE indicating the rare codons throughout the consensus sequence, (A) using the codon table Ecohigh.cod and (B) using the codon table Rhizall.cod.	104
33	Stem-loop structures downstream of the <i>bar</i> gene.	109
34	The putative coding region of the <i>bar</i> gene sequence and the amino acid sequence derived from it.	112