(S·V), the optimum is (again) where $\partial \ln S/\partial \ln V = -1$ on the $\ln S$, $\ln V$ trade-off surface. This result is derived, but not really understood, in ref. 6.

It seems that several classic problems in life-history evolution yield a universal trade-off slope of -1 at the optimum because aggregated fitness (Box 1) is naturally expressed as a product of two allocation alternatives (E versus B, $\bar{b}$ versus S, S versus V). Life histories are often treated as complex objects, with numerous possible age-dependent trade-offs[7,8,15]. The approach here reduces them to just a few aggregate variables (S, $\bar{b}$, E)[6]. Although we sacrifice all information about age-dependent allocation decisions[7,8], we gain in finding general rules (hypotheses) about the shape of aggregate (average) trade-off surfaces at the equilibrium. I do not know of any data on trade-offs precise enough to test these differential-invariant predictions, particularly for the size/number or reproductive effort problems[7]. Theory for the optimal age at first reproduction using R_0 as a fitness measure often implicitly invokes the minus-one rule as an intermediate step in the prediction of attributes such as optimal adult body mass[6,7,12]. Thus the minus-one rule is tested, at least qualitatively, whenever these predictions work out. See, for example, the successful prediction of the heights and slopes of the between-species lifespan allometries for various mammal groups (ref. 6, p. 96; ref. 16). Of course, there are many qualitative (and a few quantitative) tests of product maximization for evolution of sex allocation[2,6].

The procedure described here is to reduce darwinian fitness to a function of a few aggregate variables, hoping to find a general form for fitness (here a product) which then yields general rules for the equilibrium. The trick is worth trying for other problems in phenotypic evolution. Economists often use this procedure and, indeed, production or utility function in the form of products are common; $X_1 \cdot X_2$, or more generally $X_1^A \cdot X_2^D$ (see any advanced text on price theory; A and D are > 0 and scale the relative productive value of inputs to X_1 and X_2, respectively). Then products like that of equation (1) may often characterize fitness in non-growing populations with reproductive structure even more complex than simply age. □

Box 1 Fitness is a product

The 'net reproductive rate' R_0, is defined as $R_0 = \int_0^\infty l(x)\cdot b(x)dx$, and calculates the average number of daughters produced over a female's lifespan. ($l(x)$ is the probability of being alive at age x; $b(x)$ is the daughters produced at age x who are alive at independence from mother; $b > 0$ only if $x > \alpha$, the age at first birth, measured from independence). Now, write $b(x) = b(y)$ for $y = x - \alpha$ and denote $l(x)$ for $x > \alpha$ as $l(x) = S(\alpha)\cdot e^{-\phi(x-\alpha)} = S(\alpha)\cdot e^{-\phi(y)}$. (Notice that $\phi(y)$ is zero at $y = 0$ and is increasing with y; $\partial(-\log l(x))/\partial y = \partial\phi(y)/\partial y$, so $\partial\phi/\partial y$ is the instantaneous mortality rate at age y.) $S(\alpha)$ is the chance of living from independence to α. R_0 can be written for this general life history as

$$R_0 = S(\alpha)\cdot\int_0^\infty b(y)\cdot e^{-\phi(y)}dy \qquad (2)$$

Recall from the stable age distribution theory that the proportion of the breeding lifespan spent between ages y and $y + dy$ (the probability density function for the adult ages) is given by

$$\frac{e^{-\phi(y)}}{\int_0^\infty e^{-\phi(y)}dy}dy$$

Now, multiply equation (2) by (1), written as

$$\frac{\int_0^\infty e^{-\phi(y)}dy}{\int_0^\infty e^{-\phi(y)}dy}$$

to yield

$$R_0 = S(\alpha)\left[\int_0^\infty b(y)\cdot\frac{e^{-\phi(y)}}{\int_0^\infty e^{-\phi(y)}dy}dy\right]\left(\int_0^\infty e^{-\phi(y)}dy\right) \qquad (3)$$

$S(\alpha)$ is the chance of living to reproduce at age α; while the term in square brackets is simply $\bar{b}$, the average rate of production of offspring over the reproductive adult life, and the term in curved brackets is simply $E(\alpha)$, the expectation of further life at age α, the average length of the adult lifespan. So equation (3) is really

$$R_0 = S(\alpha)\cdot\bar{b}\cdot E(\alpha) \qquad (4)$$

Equation (4) applies to any age structured life history; R_0 is the simple product of three aggregated terms, each an average. For equation (4) to be used as a fitness measure, the population must not be growing. This makes $R_0 \approx 1$ for typical individuals owing to density dependence[6]. But mutant individuals may have their own $R_0 \neq 1$, and it is fitness and trade-offs for mutants which are discussed here. Thus we use R_0, equation (4), as a fitness measure, with the condition that it must equal unity at the optimum, when mutants are the same as typical[1].

Received 30 January; accepted 3 April 1997.

1. Parker, G. A. & Maynard Smith, J. Optimality theory in evolutionary biology. *Nature* **348,** 27–33 (1990).
2. Charnov, E. L. *The Theory of Sex Allocation* (Princeton Univ. Press, NJ, 1982).
3. Seger, J. & Stubblefield, J. W. in *Adaptation* 93–123 (Academic, New York, 1996).
4. Bulmer, M. G. *Theoretical Evolutionary Ecology* (Sinauer, Sunderland, MA, 1994.).
5. Godfray, H. C. *J. Parasitoids* (Princeton Univ. Press, NJ, 1994).
6. Charnov, E. L. *Life History Invariants* (Oxford Univ. Press, 1993).
7. Stearns, S. C. *The Evolution of Life Histories* (Oxford Univ. Press, 1992).
8. Charlesworth, B. *Evolution in Age-Structured Populations.* (Cambridge Univ. Press, 1994).
9. Charnov, E. L. & Downhower, J. F. A trade-off-invariant life-history rule for optimal offspring size. *Nature* **376,** 418–419 (1995).
10. MacArthur, R. H. in *Theoretical and Mathematical Biology* (eds Waterman, T. H. & Morowitz, H.) 388–397 (Blaisdell, New York, 1965).
11. Williams, G. C. Natural selection, the cost of reproduction, and a refinement of Lack's principle. *Am. Nat.* **100,** 687–690 (1966).
12. Charnov, E. L. Evolution of life history variation among female mammals. *Proc. Natl Acad. Sci. USA* **88,** 1134–1137 (1991).
13. Lloyd, D. G. Selection of offspring size at independence and other size-versus-number strategies. *Am. Nat.* **129,** 800–817 (1987).
14. Mock, D. W. & Parker, G. A. *The Evolution of Sibling Rivalry* (Oxford Univ. Press, 1997).
15. Stearns, S. C. Tradeoffs in life history evolution. *Funct. Ecol.* **3,** 259–268 (1989).
16. Charnov, E. L. & Berrigan, D. Why do female primates have such long lifespans and so few babies? *Evol. Anthropol.* **1,** 191–194 (1993).

Acknowledgements. K. Hawkes, G. Parker, G. Getty and S. Proulx read and improved the arguments.

Correspondence and requests for materials should be addressed to the author.

Molecular basis of symbiosis between *Rhizobium* and legumes

Christoph Freiberg, Rémy Fellay*, Amos Bairoch†, William J. Broughton*, André Rosenthal & Xavier Perret*

Insitut für Molekulare Biotechnologie, Abteilung Genomanalyse, Beutenbergstrasse 11, 07745 Jena, Germany
**Laboratoire de Biologie Moléculaire des Plantes Supérieures, Université de Genève, 1 ch. de l'Impératrice, 1292 Chambésy/Genève, Switzerland*
† Biochimie Médicale, Centre Médical Universitaire, Université de Genève, 1 rue Michel-Servet, 1211 Genève 4, Switzerland

Access to mineral nitrogen often limits plant growth, and so symbiotic relationships have evolved between plants and a variety of nitrogen-fixing organisms. These associations are responsible for reducing 120 million tonnes of atmospheric nitrogen to ammonia each year. In agriculture, independence from nitrogenous fertilizers expands crop production and minimizes pollution of water tables, lakes and rivers. Here we present the complete nucleotide sequence and gene complement of the plasmid from *Rhizobium* sp. NGR234 that endows the bacterium with the ability to associate symbiotically with leguminous plants. In conjunction

with transcriptional analyses, these data demonstrate the presence of new symbiotic loci and signalling mechanisms. The sequence and organization of genes involved in replication and conjugal transfer are similar to those of *Agrobacterium*, suggesting a recent lateral transfer of genetic information.

Many dissimilar organisms live in close association with one another. Nitrogen-fixing symbioses, in which auxotrophs exchange carbohydrates for organic nitrogen with diazotrophs, dominate numerous ecological niches. Members of the plant family Leguminosae, in association with the soil bacteria *Azorhizobium*, *Bradyrhizobium* and *Rhizobium*, are responsible for most of the nitrogen fixed biologically. Plant roots excrete a variety of substances, some of which (especially flavonoids) coordinate the expression of rhizobial nodulation (*nod*) genes[1]. In turn, the Nod proteins direct synthesis of lipochito-oligosaccharidic Nod factors, which initiate nodule formation and allow rhizobia to enter the plant[2,3]. Continued exchange of symbiotic signals is necessary for nitrogen fixation[4].

As a general rule, symbiotic genes are plasmid borne in *Rhizobium* species and are located on the chromosome in *Azo(Brady)rhizobium* strains. To study the molecular control of broad host range in associations between legumes and *Rhizobium*, we use the fast-growing *Rhizobium* sp. NGR234 (ref. 5). This bacterium nodulates more than 110 genera of legumes, as well as the non-legume *Parasponia andersonii* (S. G. Pueppke and W.J.B., unpublished data), and possesses a large plasmid (pNGR234*a*) that carries most symbiotic determinants[6]. Using dye terminators and a thermostable sequenase[7], we sequenced 20 cosmids from the canonical ordered library[8], selected to cover pNGR234*a*.

The symbiotic replicon is 536,165 base pairs long (92% of the genome of *Mycoplasma genitalium*[9]). A total of 416 open reading frames (ORFs) were predicted to encode proteins (Fig. 1, Table 1), 139 of which show no similarity to any known protein. An additional 67 gene fragments were detected that seem to be remnants of functional genes. Many of these have clearly been disrupted by mobile elements. Collectively, potential genes and gene fragments make up 418 kilobases (78% of the replicon). This gene density is slightly lower than that of other bacterial genomes such as *Bacillus subtilis*, *Escherichia coli* and *Haemophilus influenzae*[10] (85–92%; I. Moszer, personal communication), mostly because of the numerous insertion sequences and of their recombinatorial nature. No genes essential to transcription, translation or primary metabolism were found, which is consistent with the observation that NGR234 can be cured of its plasmid, giving strain ANU265 (ref. 11).

In total, 85 proteins belong to families that are represented more than once, even after discounting the many insertion-sequence encoded proteins or those involved in transposition, integration and recombination. There are some large families, including the five members of the short-chain dehydrogenase/reductases[12], one of which (y4vI) contains two homologous domains; four complete and one partial ABC-type transporter operons that each encode at least one ABC-type permease and ABC-type ATP-binding proteins; four cytochrome P450s; and three members of the peptidase family S9A.

Highest similarities were found with proteins of other rhizobia and agrobacteria. Only two eukaryotic proteins are highly homologous to pNGR234*a* counterparts. A glutamate dehydrogenase (y4uF) is significantly more closely related to the mitochondrial form than to that of prokaryotes or to archaea. As mitochondrial proteins are thought to originate from ancestral bacteria, y4uF could thus represent a new subclass of prokaryotic glutamate dehydrogenases. A small protein (y4sK), which belongs to a family (Prosite:PDOC00838) of uncharacterized proteins found in a variety of prokaryotes and eukaryotes, is more closely related to the known mammalian members of this family than to that of other prokaryotes.

Replication of pNGR234*a* probably begins at *oriV*, which is located within the intergenic sequence between the *repC* (y4cI)- and *repB* (y4cJ)-like genes. This locus is highly similar (69–71%) to the origins of replication of *Agrobacterium* and *Rhizobium* plasmids (Fig. 2). In *Agrobacterium*, these intergenic sequences are the determinants of incompatibility. RepABC of pNGR234*a* are 40–60% identical to those of pTiB6S3 (a Ti-plasmid), pRiA4b (an Ri-plasmid) and pRL8JI (a cryptic *Rhizobium* plasmid). A 12-bp portion of the origin of transfer (*oriT*) is identical to that of pTiC58 of *A. tumefaciens*, and highly similar to those of RSF1010 (*E. coli*) and pTF1 (*Thiobacillus ferrooxidans*). This sequence corresponds to the *oriT* of plasmids containing the 'Q-type nick-region' (Fig. 3). Adjacent to this cluster are another 21 predicted genes that are homologous to the conjugal transfer genes of *Agrobacterium* Ti-plasmids. By marking the *nodD2* gene (y4xH), we were able to demonstrate high-frequency conjugal transfer of pNGR234*a* into ANU265. Conjugal transfer of Ti plasmids in *Agrobacterium* is controlled by a family of *N*-acyl-L-homoserine lactone auto-inducers[13,14]. By using established methods[14], we found similar molecules, which interact with the *Agrobacterium traR* gene product, in supernatants of NGR234 cultures.

Carbohydrates are not only constituents of the rhizobial cell wall: they are also morphogens. Short, *N*-acylated tri- to pentamers of *N*-acetyl-D-glucosamine (Nod factors) trigger nodulation responses in homologous legumes at very low concentrations[15]. Thus elements of the biosynthetic pathways leading to cell walls or to Nod factors are common. Most differences are found in the later stages of the pathways that yield specific cell-wall components or Nod factors. Yet even here, the distinction between structural and symbiotic carbohydrates is blurred. Specific extracellular polysaccarides (EPS) are required for nitrogen fixation in certain legume–*Rhizobium* associations[16].

As befits a symbiotic replicon, only 10 ORFs with similarities to polysaccharide synthesis genes (*sensu stricto*) are plasmid borne (Table 1). Sequences homologous to *exo* genes are clearly located on the chromosome (X.P. and V. Viprey, unpublished data). Although loci with weak homologies to *nod*-box::*psiB* of *R. leguminosarum*, and *exoX* of *R. meliloti* exist on pNGR234*a* (y4iR, and y4xQ respectively), these are regulatory rather than structural.

Except for *nodE*, *nodG* and *nodPQ*, which are on the chromosome, most Nod-factor biosynthetic genes are plasmid borne, and are regulated by four transcriptional regulators of the *lysR* family: *nodD*1 (y4aL), *syrM*1 (y4pN), *nodD*2 (y4xH) and *syrM*2 (y4zF). Most *nod* genes share a conserved promoter sequence called the *nod* box. NodC (an *N*-acetylglucosaminyltransferase), which is the first committed enzyme in the Nod-factor biosynthetic pathway, is part of the *hsn*III locus: *nodABCIJnolO* and *noeE* are responsible for synthesis of the core Nod factor as well as the adjunction of 3-(or 4)-*O*-carbomoyl-, 2-*O*-methyl-, and 4-*O*-sulphate-[17] groups, respectively (Fig. 4). The *hsn*I locus encodes enzymes involved in fucosylation of NodNGR factor[18]. NodS and NodU (*hsn*II) *N*-methylate and 6-*O*-carbamoylate NodNGR factors respectively[19], whereas *noIL* is probably an acetyltransferase.

We found 12 additional *nod* box-like sequences. Among these, two are upstream of the transcriptional regulators, *syrM*2 and y4xI, and three (y4hM, *fixF* and y4iR) control expression of genes involved in polysaccharide metabolism. Transcription analysis suggests that at least 11 of the 17 putative *nod* boxes are symbiotically active (Fig. 1). Many *nod* box-dependent genes are under the direct control of NodD1, but y4wM is modulated by NodD2 (data not shown). The presence of both a *nod* box and a NifA-σ^{54}-type promoter upstream of y4vC suggests a new system of transcriptional regulation involving both flavonoid induction and expression in nodules (Fig. 5).

In contrast to the mosaic structure of the *nod* loci, pNGR234*a* contains a single large cluster of 43 *nif* and *fix* genes, including *nifA* (y4uN), which encodes a σ^{54}-dependent regulator. NifA contains two highly conserved regions, a carboxy-terminal part involved in

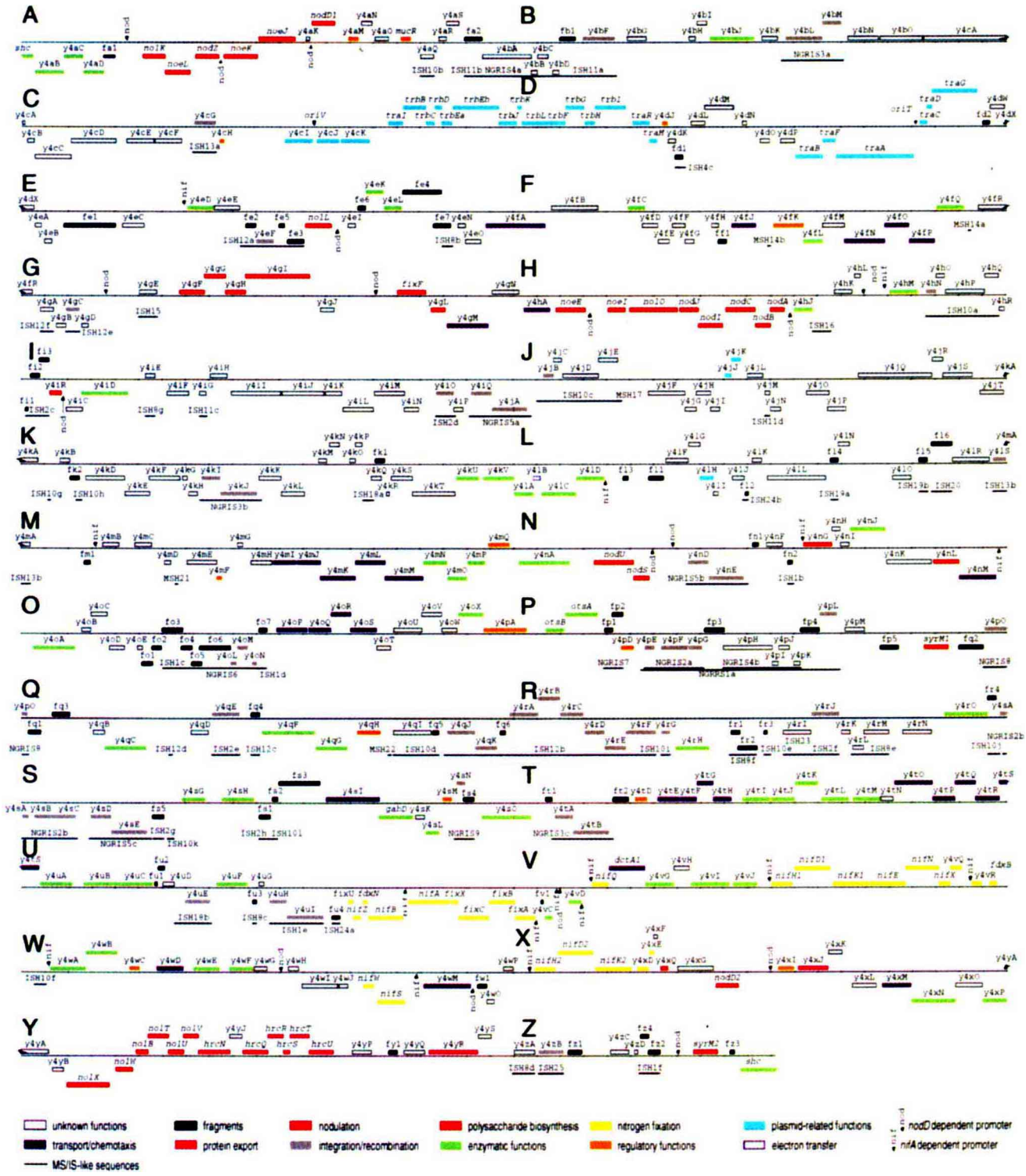

Figure 1 Genetic map of the symbiotic replicon pNGR234*a*. Each line represents 42 kb. ORF names (for example, y4aB) or the names of genes (for example, *shc*) correspond to those in Table 1. The third letter in each ORF name corresponds to one of 26 segments (A to Z), each of which is 21 kb in length except segment Z, which comprises 11,125 bp. The fourth letter of the ORF name indicates the position of the ORF within each segment (for example, y4gI is located on segment G between y4gH and y4gJ). Genes positioned on top of each line are transcribed from left to right, whereas those placed below the lines are encoded by the complementary strand. Colour-coded boxes represent putative coding regions (and gene fragments) grouped according to their presumed functions, except for those involved in nodulation, polysaccharide synthesis, nitrogen fixation, and plasmid-related functions, which are grouped according to their phenotypic class (see Table 1). Positions of the origins of replication and transfer are marked *oriV* and *oriT*, respectively. Regions homologous to consensus sequences of *nod*-box- and NifA-σ^{54} dependent promoters are marked with a black arrowhead followed by nod and nif, respectively. Positions of insertion- and mosaic-sequences are annotated with lines (for example, ISH10b, NGRIS4a or MSH22).

DNA binding, and a central domain that interacts with an alternative sigma factor (σ^{54}) of the RNA polymerase. Mutation of *rpoN* (which encodes σ^{54}) causes a Fix$^-$ phenotype on NGR234 hosts[20]. Genes involved in the synthesis of the MoFe–nitrogenase complex are also present, including two identical copies of the *nifKDH* genes (*nifHDK*1 and *nifHDK*2)[21]. Other *nif* and *fix* genes are involved in elaboration of the electron-transport complex (*fixA*, *fixB*), of various cofactors required for nitrogen fixation (*fixC*, *nifB*, *nifE*, *nifN*) and in the syntheses of ferredoxins (*fdxB*, *fdxN* and *fixX*). Although not directly involved in the fixation process, mutation of the plasmid-borne copy of *dctA* (*dctA*1, y4vF) also impairs nitrogen fixation[22].

In addition, 17 new genes expressed in nodules are part of this *nif* and *fix* cluster (y4vC to y4vJ, with the exception of *dctA1*, y4wA to y4wG, y4wI, y4wJ and y4xQ; see Fig. 5). The predicted functions of six of these (y4vD, *nifQ*, y4vG, y4vI, y4vJ and y4wF) suggest a role in electron transport and oxidation–reduction, but another five are not homologous to any database entry. It thus seems likely that most of the proteins necessary for bacteroid development and synthesis of the nitrogen-fixing complex are coded by pNGR234*a*, although

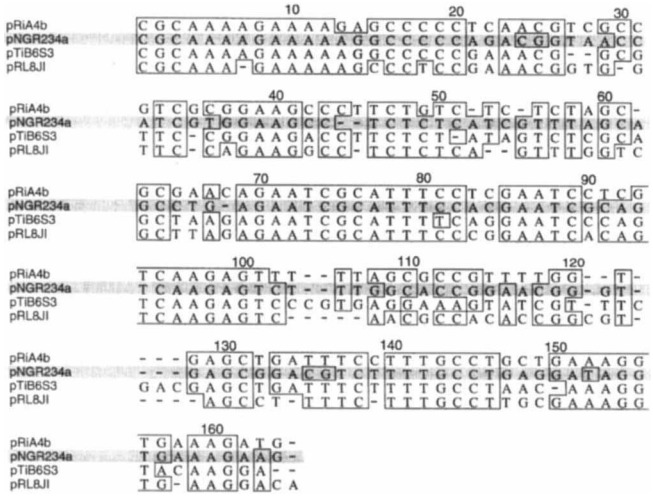

Figure 2 Multiple alignments of the nucleotide sequence of the replication origins of: the Ri plasmid of *Agrobacterium rhizogenes* (pRiA4b); the symbiotic replicon of NGR234 (pNGR234*a*); the Ti plasmid of *A. tumefaciens* BS63 (pTiBS63); and pRL8JI of *R. leguminosarum* bv. *leguminosarum* (pRL8JI). Gaps introduced to give the best sequence alignments are marked with hyphens.

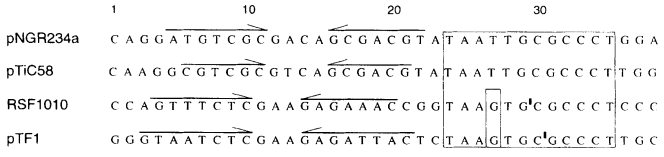

Figure 3 Multiple DNA sequence alignments of loci containing the origin of transfer of: the symbiotic plasmid of NGR234 (pNGR234*a*); the Ti plasmid of *A. tumefaciens* C58 (pTiC58); a mobilizable plasmid of *E. coli* (RSF1010) and a mobilizable plasmid of *Thiobacillus ferrooxidans* (pTFI). Major conserved nucleotide residues are boxed. Known 'nick' sites corresponding to the nucleotide positions where the specific plasmid strand is cleaved are marked with black boxes. Sequence features in the trailing portion of the *oriT* sites include inverted repeats, which are marked with horizontal arrows.

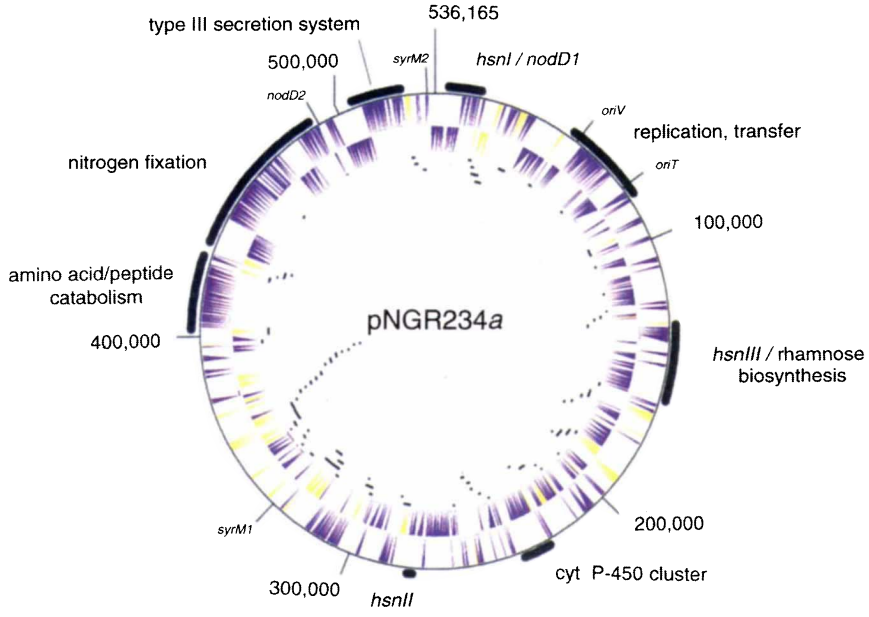

Figure 4 Circular representation of pNGR234*a*. Outer and inner concentric circles indicate coding regions identified on the plus and minus strands, respectively. Within these predicted genes, those coloured in yellow correspond to ORFs that belong to insertion-like elements. Thin concentric black lines represent mosaic sequences as well as complete or partial insertion sequence-like repeats. Major gene clusters are highlighted as thick black concentric lines (for example, *hsn*, host specificity of nodulation loci).

Table1 Functional and phenotypic classification of the predicted proteins encoded by pNGR234*a*

general statistics

total number of predicted proteins	416
mean length of the putative proteins (smallest: 43 aa, largest: 1197 aa)	308 aa
putative subcellular locations of the predicted proteins:	
- cytoplasmic	329
- transmembrane (inner membrane)	62
- periplasmic	20
- outer membrane	2
- lipoprotein attached to a membrane	3
number of proteins grouped into functional classes	234
number of unknown proteins:	
- with database homolog	43
- without any database homolog	139

proteins sorted according to phenotypic classes

functional class	number of proteins
nodulation	
enzymes	14
transporters	2
transcriptional regulators	4
export proteins	1
unknown functions	5
poly-/oligosaccharide synthesis	
enzymes	9
transcriptional regulators	1
nitrogen fixation	
enzymes	13
electron transfer proteins	5
transcriptional regulators	1
unknown functions	7
plasmid-related functions/cell development	
enzymes	1
export proteins	12
transcriptional regulators	2
plasmid control/cell development proteins	12

other proteins

functional class	number of proteins
various enzymatic functions	59
transport	29
protein export/secretion	8
chemotaxis	2
electron transfer	1
transcriptional regulation	14
integration/recombination	44

ORF name	gene name	description of the predicted protein
nodulation		
Nod-factor biosynthesis/modification/transport		
y4aF	*nolK*	pu. NAD-dependent nucleotide sugar epimerase/dehydrogenase
y4aG	*noeL*	GDP-mannose 4,6-dehydratase
y4aH	*nodZ*	fucosyltransferase
y4aI	*noeK*	phosphomannomutase
y4aJ	*noeJ*	mannose-1-phosphate guanylyltransferase
y4eH	*nolL*	sim. to *R. loti* acetyltransferase NolL
y4hB	*noeE*	transfer of activated sulfate to fucose
y4hC	*noeI*	not yet characterized step in Nod-factor synthesis
y4hD	*nolO*	O-carbamoylation of Nod-factors
y4hE	*nodJ*	pr. ABC transporter permease
y4hF	*nodI*	pr. ABC transporter ATP-binding protein
y4hG	*nodC*	N-acetylglucosaminyltransferase
y4hH	*nodB*	chitooligosaccharide deacetylase
y4hI	*nodA*	N-acyltransferase
y4nB	*nodU*	6-O-carbamoylation of Nod-factors
y4nC	*nodS*	methyltransferase involved in Nod-factor biosynthesis
transcriptional regulation		
y4aL	*nodD1*	LysR family H-T-H regulator
y4pN	*syrM1*	LysR family H-T-H regulator
y4xH	*nodD2*	LysR family H-T-H regulator
y4zF	*syrM2*	LysR family H-T-H regulator
protein secretion/unknown functions		
y4yC	*nolX*	unknown function
y4yD	*nolW*	secretion system protein (PulD family)
y4yE	*nolB*	unknown function
y4yF	*nolT*	unknown function
y4yG	*nolU*	unknown function
y4yH	*nolV*	unknown function
poly-/oligosaccharide synthesis		
y4gF		pr. dTDP-glucose 4,6-dehydratase
y4gG		pr. dTDP-4-dehydrorhamnose reductase
y4gH		pr. glucose-1-phosphate thymidylyl-transferase
y4gI		pu. enzyme involved in polysaccharide biosynthesis; sim. to *Myx. xanthus* O-antigen RfbC
y4gK	*fixF*	pu. enzyme involved in polysaccharide biosynthesis; Low similarity to *E. coli* KpsS
y4gL		pr. dTDP-4-dehydrorhamnose 3,5-epimerase
y4iR		possibly involved in polysaccharide biosynthesis; sim. to *R. leguminosarum* PsiB
y4nG		pu. NAD-dependent nucleotide sugar epimerase/dehydrogenase
y4nL		pu. NAD-dependent nucleotide sugar epimerase/dehydrogenase
y4xQ		pu. exopolysaccharide production repressor; sim. to Rhizobium ExoX
nitrogen fixation		
y4uJ	*fixU*	unknown function
y4uK	*nifZ*	unknown function
y4uL	*fdxN*	4Fe-4S ferredoxin
y4uM	*nifB*	nitrogenase cofactor biosynthesis
y4uN	*nifA*	sigma54-dependent H-T-H regulator
y4uO	*fixX*	pr. 3Fe-3S ferredoxin
y4uP	*fixC*	nitrogenase cofactor biosynthesis
y4vA	*fixB*	electron transfer fixAB complex
y4vB	*fixA*	electron transfer fixAB complex
y4vE	*nifQ*	pu. nitrogenase Mo cofactor processing
y4vK	*nifH1*	nitrogenase Fe protein; id. to y4xA
y4vL	*nifD1*	nitrogenase MoFe protein alpha subunit; id. to y4xB
y4vM	*nifK1*	nitrogenase MoFe protein beta subunit; id. to y4xC
y4vN	*nifE*	nitrogenase MoFe cofactor synthesis
y4vO	*nifN*	nitrogenase MoFe cofactor synthesis
y4vP	*nifX*	unknown function
y4vQ		sim. to *nifX-nifW* intergenic ORF in nitrogen-fixing bacteria
y4vR		sim. to N-terminus of NifH
y4vS	*fdxB*	4Fe-4S ferredoxin
y4wK	*nifW*	unknown function
y4wL	*nifS*	class-5 aminotransferase
y4xA	*nifH2*	nitrogenase Fe protein; id. to y4vK
y4xB	*nifD2*	nitrogenase MoFe protein alpha subunit; id. to y4vL
y4xC	*nifK2*	nitrogenase MoFe protein beta subunit; id. to y4vM
y4xD		sim. to *nifX-nifW* intergenic ORF in nitrogen-fixing bacteria
y4xE		sim. to *nifX-nifW* intergenic ORF in nitrogen-fixing bacteria
plasmid-related functions/cell development		
y4cI		pu. replication protein; sim. to *A. rhizogenes* repC
y4cJ		pu. replication protein; sim. to *A. rhizogenes* repB
y4cK		pu. replication protein; sim. to *A. rhizogenes* repA
y4cL	*traI*	pr. autoinducer synthetase
y4cM	*trbB*	pr. conjugal transfer protein (PulE family)
y4cN	*trbC*	pr. conjugal transfer protein (export prot.)
y4cO	*trbD*	pr. conjugal transfer protein (export prot.)
y4cP	*trbEa*	pr. conjugal transfer protein (export prot.)
y4cQ	*trbEb*	pr. conjugal transfer protein (export prot.)
y4dA	*trbJ*	pr. conjugal transfer protein (export prot.)
y4dB	*trbK*	pr. conjugal transfer protein (export prot.)
y4dC	*trbL*	pr. conjugal transfer protein (export prot.)
y4dD	*trbF*	pr. conjugal transfer protein (export prot.)
y4dE	*trbG*	pr. conjugal transfer protein (export prot.)
y4dF	*trbH*	pr. conjugal transfer protein (export prot.)
y4dG	*trbI*	pr. conjugal transfer protein (export prot.)
y4dH	*traR*	LuxR family H-T-H regulator
y4dI	*traM*	pu. repressor
y4dQ	*traB*	pr. conjugal transfer protein
y4dR	*traF*	pr. conjugal transfer protein
y4dS	*traA*	pr. conjugal transfer protein
y4dT	*traC*	pr. conjugal transfer protein
y4dU	*traD*	pr. conjugal transfer protein
y4dV	*traG*	pr. conjugal transfer protein
y4jJ		pu. plasmid stability protein; sim. to *Ps. syringae* StbC
y4jK		pu. plasmid stability protein; sim. to *Ps. syringae* StbB
y4lH		pu. cell filamentation protein; sim. to enterobacterial protein Fic
various enzymatic functions		
y4aA	*shc*	pr. squalene-hopene cyclase
y4aC		pu. phytoene synthase
y4aD		sim. to phytoene synthases
y4bJ		pu. peptidase (S2C family)
y4eD		pu. phosphodiesterase
y4eK		pr. short chain oxidoreductase
y4eL		pr. short chain oxidoreductase
y4fC*		pu. monooxygenase
y4fL		member of inositol monophosphatase family
y4fQ		'ROK' family member
y4gN		sim. to *V. anguillarum* VirA
y4hJ*		pu. oxidoreductase; sim. to N-terminus of anaerobic coproporphyrinogenIII oxidases
y4hM		pu. oxidoreductase; Low similarity to *Z. mobilis* glucose-fructose oxidoreductase
y4iD		pr. monooxygenase
y4kU		pr. geranyltranstransferase
y4kV		pr. cytochrome P450
y4lA		pr. short chain oxidoreductase
y4lC		pr. cytochrome P450
y4lD		pr. cytochrome P450
y4mN*		pr. TPP enzyme (transketolase family; C-terminus)
y4mO*		pr. TPP enzyme (transketolase family; N-terminus)
y4mP		pr. short chain oxidoreductase
y4nA		pr. peptidase (S9A family)
y4nJ		pr. GMC-type oxidoreductase
y4oA		sim. to *E. coli* pMCCC7 microcin biosynthesis protein MccB
y4oX		pr. NAD-dependent oxidoreductase
y4pB	*otsB*	pr. trehalose-phosphatase
y4pC	*otsA*	pr. α,α-trehalose-phosphate synthase
y4qC		sim. to *E. coli* pMCCC7 microcin biosynthesis protein MccB
y4qF		pr. peptidase (S9A family)
y4qG		class-3 aminotransferase
y4rH		pu. ligase; sim. to biotin carboxylase
y4rO		sim. to C-terminus of histidinol-phosphate aminotransferase
y4sG		pu. ligase; sim. to D-ala, D-ala ligases
y4sH		pu. cell wall compound biosynthesis protein; sim. to *B. anthracis* CapA
y4sJ	*gabD*	pr. succinate-semialdehyde dehydrogenase
y4sL		pu. oxidoreductase; sim. to C-terminus of *E. coli* D-amino acid dehydrogenase small subunit
y4sO		pr. peptidase (S9A family)
y4tI		pu. peptidase (M40 family)
y4tJ		pu. threonine dehydratase
y4tK		pu. cyclodeaminase
y4tL		pu. peptidase/hydrolase (M24 family)
y4tM		pu. peptidase/hydrolase (M24 family)
y4uA		pu. cell wall comp. biosynthesis protein; sim. to *B. anthracis* CapA
y4uB		class-3 aminotransferase
y4uC		pr. aldehyde dehydrogenase
y4uF		pr. glutamate dehydrogenase
y4vC		member of HesB/YadR/YfhF family
y4vD		pu. redox enzyme; sim. to *C. boidinii* peroxisomal proteins A/B & *H. influenzae* HI0572
y4vG		pr. cytochrome P450
y4vI		pr. short chain oxidoreductase
y4vJ		pu. monooxygenase; sim. to LuxA/B (bacterial luciferase)
y4wA		pr. zinc peptidase (M16 family)
y4wB		pu. zinc peptidase (M16 family)
y4wE		class-2 aminotransferase
y4wF		pu. monooxygenase; sim. to LuxA/B (bacterial luciferase)

►

Table 1 continued

y4xN		Low similarity to *E. coli* pCOLV-K30 aerobactin synthase subunit lucC
y4xP		pu. cysteine synthase
transport		
probable sugar transport systems		
y4mI		pr. ABC transporter binding prot.
y4mJ		pr. ABC transporter permease
y4mK		pr. ABC transporter ATP-binding protein
y4oP		pr. ABC transporter binding protein
y4oQ		pr. ABC transporter permease
y4oR		pr. ABC transporter permease
y4oS		pr. ABC transporter ATP-binding protein
probable aminoacid/peptide transport systems		
y4tE		pr. ABC transporter binding protein
y4tF		pr. ABC transporter permease
y4tG		pr. ABC transporter permease
y4tH		pr. ABC transporter ATP-binding protein
y4tO		pr. ABC transporter binding protein
y4tP		pr. ABC transporter permease
y4tQ		pr. ABC transporter permease
y4tR		pr. ABC transporter ATP-binding protein
y4tS		pr. ABC transporter ATP-binding protein
other transport systems		
y4fJ		pu. porin; sim. to *R. leguminosarum* OMPIIIA
y4fN		pr. ABC transporter permease
y4fO		pr. ABC transporter ATP-binding protein
y4fP		pr. ABC transporter binding protein
y4gM		pr. ABC transporter ATP-binding protein
y4hA		pu. ionic transporter; sim. to *E. coli* ChaA
y4mL		pu. permease (*E. coli* YiaN/YgiK family)
y4mM		pu. permease (SBR family 7)
y4nM		pu. permease; sim. to *Az. caulidans* NoeC
y4vF	*dctA1*	C4-dicarboxylate transporter
y4wD		permease-type protein; sim. to *R. melliloti* MosC
y4wM		pu. ABC transporter binding prot.
y4xM		*permease-type protein; sim. to E. coli* YceE and to tetracycline transporters
protein export/secretion		
y4yI	*hrcN*	pr. ATP synthase of a secretion system
y4yK	*hrcQ*	pr. secretion system protein (FliN/MopA/SpaO family)
y4yL	*hrcR*	pr. secretion system protein (FliP/MopC/SpaP family)
y4yM	*hrcS*	pr. secretion system protein (FliQ/MopD/SpaQ family)
y4yN	*hrcT*	pr. secretion system protein (FliR/MopE/SpaR family)
y4yO	*hrcU*	pr. secretion system protein (FlhB/HrpN/SpaS family)
y4yR		pr. secretion system protein (FHIPEP family)
y4xJ		secretion system protein (PulD family)
chemotaxis		
y4fA		pr. chemoreceptor
y4sI		pr. chemoreceptor
electron transfer		
y4lB		pu. 3Fe-3S ferredoxin
transcriptional regulation		
y4aM		pu. DNA-binding protein according to *Rsp. rubrum* protein homologue
y4aP		mucR Ros/MucR homologue
y4cH		pr. cold shock regulator
y4dJ		PbsX family H-T-H regulator
y4fK		AraC family H-T-H regulator
y4mF		pr. H-T-H regulator; sim. to phage p22 C2
y4mQ		LysR family H-T-H regulator
y4pA		pr. sigma54-dependent H-T-H regulator
y4pD		Ros/MucR homologue
y4qH		LuxR family H-T-H regulator
y4sM*		AsnC/Lrp family H-T-H regulator
y4tD		AsnC/Lrp family H-T-H regulator
y4wC		pu. DNA-binding protein according to *Rsp. rubrum* protein homologue
y4xI		Signal transduction-type regulator
integration/recombination		
y4bF		pu. transposase (sim. to IS1202)
y4bL		pu. transposase; id. to y4kJ & y4tB (sim. to IS21/IS1162 family)
y4bM		pu. transposase-associated ATP-binding protein; id. to y4kI & y4tA (sim. to IS21/IS1162 family)
y4cG		pr. DNA invertase 'resolvase-type'
y4eF		pu. integrase/recombinase 'phage- type'
y4gC		pu. integrase/recombinase 'phage-type'
y4hN		sim. to IS2/IS1312/IS866 proteins
y4iO*		pu. transposase (sim. to IS1111A/IS1328/IS1533 family)
y4iQ		pu. transposase-associated ATP-binding protein; id. to y4nD & y4sD (sim. to IS21/IS1162 family)
y4jA		pu. transposase; id. to y4nE & y4sE (sim. to IS21/IS1162 family)
y4jB		sim. to IS elements IS2/IS1312/IS866 proteins
y4kI		pu. transposase-associated ATP-binding protein; id. to y4bM & y4tA (sim. to IS21/IS1162 family)
y4kJ		pu. transposase; id. to y4bL & y4tB (sim. to IS21/IS1162 family)
y4lS		pu. integrase/recombinase 'resolvase-type'
y4nD		pu. transposase-associated ATP-binding protein; id. to y4iQ & y4sD (sim. to IS21/IS1162 family)
y4nE		pu. transposase; id. to y4jA & y4sE (sim. IS21/IS1162 family)
y4pE		sim. to *R. fredii* RFRS9 and *Azo. xylinum* IS1268 ORF; id. to y4sA
y4pF		pr. transposase; id. to y4sB (sim. to IS1111A/IS1328/IS1533 family)
y4pG		sim. to *A. xylinum* IS1268 ORFA; id. to y4sC
y4pL		pu. transposase-associated ATP-binding protein (sim. to IS21/IS1162 family)
y4pO		pr. transposase (Mutator family)
y4qE		pr. transposase (sim. to IS1111A/IS1328/IS1533 family)
y4qJ		pu. transposase (sim. to IS801)
y4qK		pu. integrase/recombinase 'phage-type'
y4rA		pu. integrase/recombinase 'phage-type'
y4rB		pu. integrase/recombinase 'phage-type'
y4rC		pu. integrase/recombinase 'phage-type'
y4rD		pu. integrase/recombinase 'phage-type'
y4rE		pu. integrase/recombinase 'phage-type'
y4rF		pu. integrase/recombinase 'phage-type'
y4rG		sim. to IS2/IS1312/IS866 proteins
y4rJ		pu. transposase (sim. to IS1111A/IS1328/IS1533 family)
y4sA		sim. to *R. fredii* RFRS9 and *Azo. xylinum* IS1268 ORF; id. to y4pE
y4sB		pr. transposase; id. to y4pF (sim. to IS1111A/IS1328/IS1533 family)
y4sC		sim. to A. xylinum IS1268 ORFA; id. to y4pG
y4sD		pu. transposase-associated ATP-binding protein; id. to y4iQ & y4nD (sim. to IS21/IS1162 family)
y4sE		pu. transposase; id. to y4jA & y4nE (sim. to IS21/IS1162 family)
y4sN		pu. transposase (IS6501 family)
y4tA		pu. transposase-associated ATP-binding protein; id. to y4bM & y4kI (sim. to IS21/IS1162 family)
y4tB		pu. transposase; id. to y4bL & y4kJ (sim. to IS21/IS1162 family)
y4uH		pu. transposase-associated ATP-binding protein (sim. to IS21/IS1162 family)
y4uI		pu. transposase (sim. to IS21/IS1162 family)
y4uE*		pu. transposase (sim. to IS110 family)
y4zB*		pu. transposase (sim. to IS4 family)
unknown function (with database homologue)		
derived from IS element-like sequences		
y4aQ		sim. to ORFs in *R. meliloti* & *A. tumefaciens* Ti plasmid
y4hO		sim. to ORFC from *R. leguminosarum* symbiotic plasmid
y4hP		sim. to ORFs in *R. meliloti* & *A. tumefaciens* Ti plasmid
y4hQ		sim. to ORF-3 in *A. rhizogenes* plasmid pRIA4B
y4jC		sim. to ORFC from *R. leguminosarum* symbiotic plasmid
y4jD		sim. to ORFs in *R. meliloti* & *A. tumefaciens* Ti plasmid
y4qI		sim. to ORFs in *R. meliloti* & *A. tumefsciens* Ti plasmid
not linked to IS elements		
y4aN		sim. to *A. rhizogenes* pRiA4B replication region ORF3
y4bI		sim. to *H. influenzae* HI1631
y4bK		sim. to *Ps. aeruginosa* PAH cluster ORF1
y4dM		sim. to *E. coli* HipA & *H. influenzae* HI0665
y4dO*		sim. to mitochondrial intron encoded ORFs & *E. coli* ORF319
y4dP		sim. to *A. tumefaciens* Ti plasmid ORF2&3 in conjugal transfer region 1
y4eC		sim. to N-terminus of *E. coli* pRP4 & pR751 TraC
y4fR		sim. to *Sh. flexneri* IpaH 7.8 & *Y. pestis* YopM
y4iC		sim. to *M. tuberculosis* Mtcy373.06
y4kL		pr. AAA-family ATPase
y4kR*		sim. to N-terminus of *Erw. herbicola* ORF6 in *crtE-crtX* region
y4kS		highly sim. to *Br. japonicum* hyp. protein
y4kT		highly sim. to Br.japonicum hyp. protein
y4lL		Member of *E. coli* YegE/YhdA/YhjK/YjcC family
y4lO		sim. to *Ps. syringae* avirulence protein AvrRxv
y4mB		sim. to *E. coli* yciD; *V. cholerae* OmpW & *Ps. oleovorans* AlkL
y4mE		sim. to *E. coli* HipA & *H. influenzae* HI0665
y4nH		sim. to *E. coli* ethidium bromide resistance protein MvrC
y4rN		sim. to *M. tuberculosis* Mtcy50.24
y4sK		pr. important cellular function (Yer057C/Yil051C/YjgF family)
y4wH		sim. to *A. tumefaciens* plasmid pTiA6 ORF in pinF2 region
y4yJ		sim. to R.fredii USDA257 ORF7
y4zC		sim. to *Ps. syringae* avirulence protein AvrPph3
unknown function (no database homologue)		
derived from IS element-like sequences		
y4bA		id. to y4pH
y4bB		id. to y4pI
y4bC		id. to y4pJ
y4bD		id. to y4pK
y4gA		
y4gE*		
y4iE*		
y4iG*		
y4iP*		
y4jE*		
y4jM*		
y4kQ*		
y4mA		
y4oL		
y4oM		
y4oN		
y4pH		id. to y4bA
y4pI		id. to y4bB
y4pJ		id. to y4bC
y4pK		id. to y4bD
y4rI		
y4rL*		
y4rM*		
y4zA*		

not linked to IS elements

y4aK	y4iH	y4mD
y4aO	y4iI	y4mG
y4aR	y4iJ	y4mH
y4aS	y4iK*	y4nF
y4bG	y4iL*	y4nI
y4bH	y4iM*	y4nK
y4bN	y4iN*	y4oB
y4bO	y4jF	y4oC
y4cA	y4jG	y4oD
y4cB	y4jH	y4oE
y4cC	y4jI	y4oT
y4cD	y4jL	y4oU
y4cE	y4jN	y4oV
y4cF	y4jO*	y4oW
y4dK	y4jP*	y4pM
y4dL	y4jQ	y4qB
y4dN	y4jR	y4qD
y4dW	y4jS	y4rK
y4dX	y4jT	y4tN
y4eA	y4kA	y4uD
y4eB	y4kB	y4uG
y4eE	y4kD	y4vH
y4eI	y4kE*	y4wG
y4eN	y4kF*	y4wI
y4eO	y4kG	y4wJ
y4fB	y4kH	y4wO
y4fD	y4kK	y4wP
y4fE	y4kM	y4xF
y4fF	y4kN	y4xG
y4fG	y4kO	y4xK
y4fH	y4kP	y4xL
y4fM	y4lF	y4xO
y4gB	y4lG	y4yA
y4gD	y4lI	y4yB
y4gJ	y4lJ	y4yQ
y4hK	y4lK	y4yS
y4hL	y4lN	y4zD
y4hR	y4lR	
y4iF	y4mC	

The nomenclature system for the ORFs is described in the Fig. 1 legend. General abbreviations: amino acids (aa), identical (id.), probable (pr.), putative (pu.), similar (sim.), hypothetical (hyp.), thiamine pyrophosphate (TPP), alanine (ala), insertion sequence (IS). Abbreviations of organisms: *Agrobacterium* (A.), *Azorhizobium* (Az.), *Azotobacter* (Azo.), *Bacillus* (B.), *Bradyrhizobium* (Br.), *Candida* (C.), *Desulfovibrio* (D.), *Escherichia* (E.), *Erwinia* (Erw.), *Haemophilus* (H.), *Mycobacterium* (M.), *Myxococcus* (Myx.), *Pseudomonas* (Ps.), *Rhizobium* (R.), *Rhodospirillum* (Rsp.), *Shigella* (Sh.), *Streptomyces* (St.), *Vibrio* (V.), *Zymomonas* (Z.). Asterisk indicates possibly fragmentous gene.

some essential *fix* loci are probably carried on the chromosome[23] (V. Viprey, personal communication).

One way of gaining insight into the function of genes is to follow their expression under different conditions. To do this, RNA was hybridized against filters containing amplified portions of 113 genes and gene fragments stretching from y4uA to y4bA. Under the conditions tested, transcripts were produced from most of the ORFs, but only a few genes are expressed in liquid medium (Fig. 5). Induction with daidzein rapidly increased transcript levels of y4vC, y4xL, y4xO, y4yB, y4yP, fy1, y4yQ (all of unknown function), y4xP (cysteine synthase) and y4zF (*syrM2*). Others, such as y4wF (a putative monooxygenase), y4wE (an aminotransferase), y4wM (an ABC-transporter binding protein), y4xI (a signal transduction-type regulator), y4xK (unknown), as well as *nolBTUV* and *hrcNQRST* are induced later. As expected, those genes involved in nitrogen fixation (both copies of *nifKDH*, and *nifE*), as well as y4wA, y4wb (which encodes zinc peptidases), y4wC (a DNA-binding protein), y4wD (a permease), y4aP (*mucR*) and y4aQ (unknown) are strongly expressed in nodules.

In addition to bacterial Nod factors, other signal molecules are probably necessary for the establishment of an effective symbiosis. This is exemplified by a rhamnose-rich extracellular polysaccharide that is involved in bacteroid development and nitrogen fixation in *Vigna*. A single locus encodes the complete biosynthetic pathway of dTDP-rhamnose from glucose-1-phosphate (y4gH, y4gF, y4gL, y4gH and *fixF*), while the y4gI gene product is probably needed for synthesis of rhamonose-rich lipopolysaccharides from dTDP-rhamnose[24].

Polypeptides and proteins are also probably involved in symbiosis. Six ORFs downstream of *nolXWBTUV* show strong homology to components of the type III secretion machinery of animal and plant pathogens. Homologues of these genes (*hrcN*, y4yJ, *hrcQ*, *hrcR*, *hrcS*, *hrcT* and *hrcU*) are responsible for secretion of various proteins, and have been sequenced in the closely related bacterium *R. fredii* strain USDA257 (EMBL accession no. L12251). Secretion of five genistein-induced proteins of USDA257 is dependent on a functional *nolXWBTUV* locus[25]. Both in mammalian and plant pathogens, most of the proteins exported by the type III secretion machinery are pathogenicity determinants. Homologues of YopM of *Yersinia pestis* (y4fR), as well as avirulence genes of *Pseudomonas syringae* (y4lO) and *Xanthomonas campestris* (y4zC), were also found, reinforcing the idea that symbiotic and pathogenic interactions share common molecular mechanisms.

Surprisingly, mosaic sequences and insertion sequences comprise 18% of pNGR234*a*, and resemble those of diverse eubacteria (*Agrobacterium*, *Bacillus*, *Escherichia*, *Pseudomonas* and *Rhizobium*). Many are clustered between nucleotides 300,000 and 390,000 (Fig. 1). Other insertion and mosaic sequences divide the replicon into large blocks of functionally related genes (*oriV-oriT*, *nif*, *fix*, *hsn*; Fig. 4), suggesting that NGR234 has functioned as a 'transposon trap'. The G+C contents of these insertion and mosaic sequences are 3% higher than that of pNGR234*a* (58.5%) which, in turn, is 3.7 mol% less than the 62.2 mol% calculated for the entire genome[26]. Several genes, especially those involved in Nod-factor and polysaccharide synthesis, have low G+C contents (45–55 mol%), raising the possibility that *nod* genes evolved separately from *nif* and *fix* genes (G+C content, 59 mol%). Although transposition of these insertion elements has not been demonstrated, transfer of plasmids amongst rhizobia in the legume rhizosphere[27] and to other non-symbiotic bacteria in fields[28] indicates that lateral transfer of genetic information has helped model symbiotic potential.

The high proportion of insertion and mosaic sequences and the well-conserved *Agrobacterium* conjugal transfer loci have broad implications for the evolution of symbioses between legumes and *Rhizobium*. Both genera are closely related[29], and the similarity of basic plasmid functions suggests a common origin. It is thus possible that an *Agrobacterium*-like progenitor gathered symbiotic genes through transposition with other soil bacteria, or that the progenitor was *Rhizobium* which assimilated opine and hormone genes from the plant[30]. Transposable elements shape evolution in

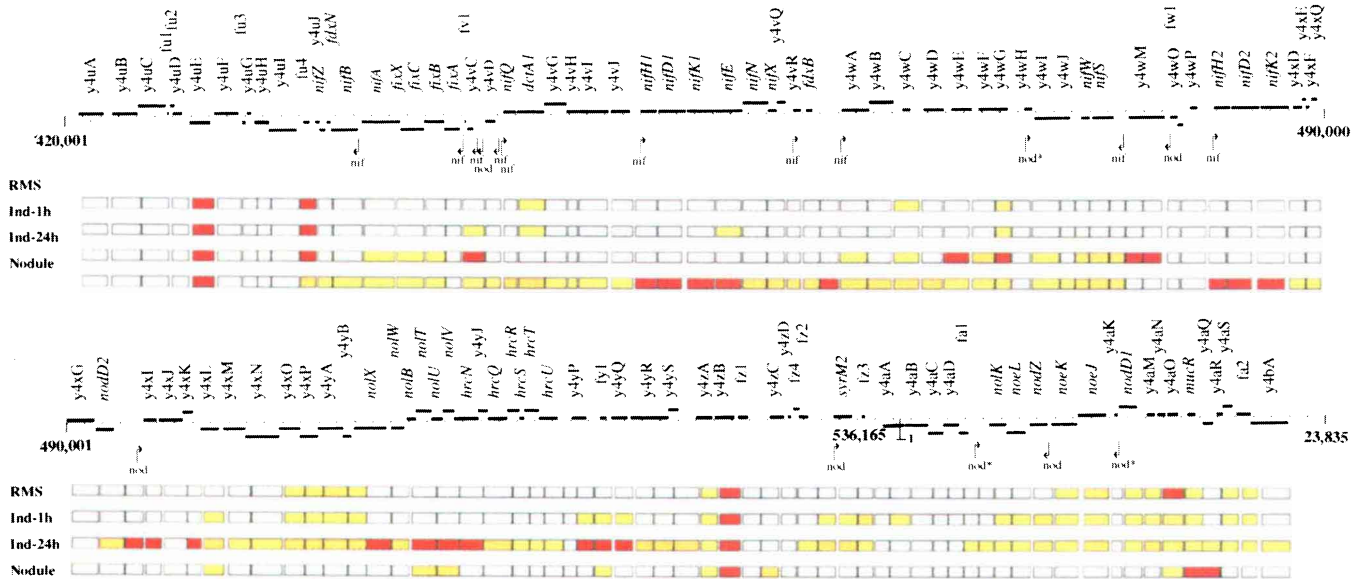

Figure 5 High-resolution transcription map of the 137-kb region encompassing the *nif* and *fix* clusters as well as the *hsnI* locus. Predicted coding regions and gene fragments are shown on each strand (black rectangles). Positions of the corresponding amplified fragments are shown directly under the ORF map. Positions and orientations of the putative *nod*-box (*nod*) and NifA-σ^{54} (*nif*) promoter sequences are marked with arrows. Probable non-functional *nod* boxes are marked with asterisks. RNA probes were prepared from: RMS, NGR234 cells grown at 28 °C in liquid *Rhizobium* minimal supplemented with succinate (RMS)[23] medium; Ind-1h, cells grown in RMS followed by a 1-h induction with 2×10^{-7} M daidzein; Ind-24h, RMS-grown rhizobia collected 24 h after induction with daidzein; Nodule, bacteroids purified from Fix$^+$ nodules of *V. unguiculata* inoculated with NGR234. Intensity of the hybridization signals detected on autoradiograms are colour coded as follows: yellow, weak; orange, medium; red, strong; and white, no signal.

different ways: they could 'pick up and carry' symbiotic or pathogenicity genes from one organism to another by forming composite transposons, or induce deletions, duplications, inversions and replicon fusions[31]. Unfortunately, there is no formal proof that the insertion and mosaic sequences of NGR234 have played these roles, although it seems likely, given the mosaic structure of the replicon. A striking example is the cytochrome P450 cluster (y4kS to y4kV, and y4lA to y4lD) which is 10% richer in G+C content than the rest of the plasmid, and has proteins with 83–93% similarity to those of *Bradyrhizobium*. Thus conjugal transfer and transposition have probably directed the evolution of bacteria which in one case learned genetically to colonize plants, and in the other to enter them. □

Methods

Sequencing and sequence analysis. Dye-terminator methods were used to 'shotgun' sequence a canonical array of cosmids[7]. Coding regions were detected using a mixture of intrinsic and extrinsic methods[32]. GeneMark predictions were based on matrices developed for *R. leguminosarum* and *R. meliloti*. BLAST and FASTA were used for similarity searches against Swiss–Prot, TrEMBL, GenBank and EMBL. Signal sequences, transmembrane segments and other characteristic domains were identified in putative proteins using the PC/Gene package. Definitions of protein families can be retrieved from the Prosite database (http://www.expasy.ch/sprot/prosite.html)[12]. Scans for homologies were last performed in November 1996 using resources at NCBI (http://ncbi.nlm.nih.gov), ExPASy (http://expasy.ch) and EBI (www.ebi.ac.UK). Searches for consensus promoter sequences were performed using FINDPATTERNS (WASP version 8, Genetics Computer Group, Madison, Wisconsin).

Transcription analysis. We made 113 PCR products ranging from ~600 bp to ~2,000 bp between y4uA (5′-end, bp 420,774) and y4bA (3′-end, bp 21,758) using cosmid DNA as the template and primer pairs designed to amplify the discrete ORFs or intergenic regions. In some cases, M13 phage DNA (from the sequencing libraries) was used to produce PCR products. Portions of the amplified fragments were then separated on 0.8% agarose gels, and vacuum-blotted onto GeneScreen Plus Nylon membranes (NEN).

RNA extraction, labelling and hybridizations. Cell cultures at an absorbance at 600 nm of 0.4–0.5 were collected by centrifugation. Bacteroids were isolated from nodules crushed in liquid nitrogen and resuspended in sterile water. Debris was removed by filtration and bacteroids were recovered by centrifugation (4,000*g*, 5 min). Purification, radioactive labelling of RNA from bacteroids and collected cells, and hybridization conditions were as described[18]. No unlabelled competitor RNA was added to the prehybridization solution. Filters were exposed for 6–72 h, and the hybridization signals were grouped according to intensity: no signal, weak, medium and strong.

Received 14 February; accepted 14 April 1997.

1. Fellay, R., Rochepeau, P., Relić, B. & Broughton, W. J. in *Pathogenesis and Host Specificity in Plant Diseases. Histopathological, Biochemical, Genetic and Molecular Bases* (eds Singh, U. S., Singh, R. P. & Kohmoto, K.) 199–220 (Pergamon, Oxford, 1995).
2. Spaink, H. P. Regulation of plant morphogenesis by lipo-chitin oligosaccharides. *Crit. Rev. Plant Sci.* **15,** 559–582 (1996).
3. Dénarié, J., Debellé, F. & Promé, J.-C. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signalling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem.* **65,** 503–535 (1996).
4. Fischer, H.-M. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol. Rev.* **58,** 352–386 (1994).
5. Lewin, A. *et al.* Mutliple host-specificity loci of the broad host-range *Rhizobium* sp. NGR234 selected using the widely compatible legume *Vigna unguiculata*. *Plant Mol. Biol.* **8,** 447–459 (1987).
6. Broughton, W. J., Heycke, N., Meyer z.A., H. & Pankhurst, C. E. Plasmid-linked *nif* and "*nod*" genes in fast-growing rhizobia that nodulate *Glycine max*, *Psophocarpus tetragonolobus*, and *Vigna unguiculata*. *Proc. Natl Acad. Sci. USA* **81,** 3093–3097 (1984).
7. Freiberg, C., Perret, X., Broughton, W. J. & Rosenthal, A. Sequencing the 500-kb GC-rich symbiotic replicon of *Rhizobium* sp. NGR234 using dye terminators and a thermostable "Sequenase": a beginning. *Genome Res.* **6,** 590–600 (1996).
8. Perret, X., Broughton, W. J. & Brenner, S. Canonical ordered cosmid library of the symbiotic plasmid of *Rhizobium* species NGR234. *Proc. Natl Acad. Sci. USA* **88,** 1923–1927 (1991).
9. Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270,** 397–403 (1995).
10. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269,** 496–512 (1995).
11. Morrisoon, N. A. *et al.* Mobilization of a *sym* plasmid from a fast-growing cowpea *Rhizobium* strain. *J. Bacteriol.* **160,** 483–487 (1984).
12. Bairoch, A., Bucher, P. & Hofmann, K. The PROSITE database, its status in 1995. *Nucleic Acids Res.* **24,** 189–196 (1996).
13. Zhang, L., Murphy, P. J., Kerr, A. & Tate, M. E. *Agrobacterium* conjugation and gene regulation by *N*-acyl-L-homoserine lactones. *Nature* **362,** 446–448 (1993).
14. Piper, K. R., Beck von Bodman, S. & Farrand, S. K. Conjugation factor of *Agrobacterium tumefaciens* regulates Ti plasmid transfer by autoinduction. *Nature* **362,** 448–450 (1993).
15. Price, N. P. J. *et al.* Broad-host-range *Rhizobium* species strain NGR234 secretes a family of carbamoylated and fucosylated, nodulation signals that are *O*-acetylated or sulphated. *Mol. Microbiol.* **6,** 3575–3584 (1992).
16. Leigh, J. A. & Walker, G. C. Exopolysaccharides of *Rhizobium*: synthesis, regulation and symbiotic function. *Trends Genet.* **10,** 63–67 (1994).
17. Hanin, M. *et al.* Sulphation of *Rhizobium* sp. NGR234 Nod factors is dependent on *noeE*, a new host specificity gene. *Mol. Microbiol.* (in the press).
18. Fellay, R., Perret, X., Viprey, V., Broughton, W. J. & Brenner, S. Organization of host-inducible transcripts on the symbiotic plasmid of *Rhizobium* sp. NGR234. *Mol. Microbiol.* **16,** 657–667 (1995).
19. Jabbouri, S. *et al.* Involvement of *nodS* in *N*-methylation and *nodU* in 6-*O*-carbamoylation of *Rhizobium* sp. NGR234 Nod factors. *J. Biol. Chem.* **270,** 22968–22973 (1995).
20. van Slooten, J. C., Cervantes, E., Broughton, W. J., Wong, C. H. & Stanley, J. Sequence and analysis of the *rpoN* sigma factor gene of *Rhizobium* sp. strain NGR234, a primary coregulator of symbiosis. *J. Bacteriol.* **172,** 5563–5574 (1990).
21. Badenoch-Jones, J., Holton, T. A., Morrison, C. M., Scott, K. F. & Shine, J. Structural and functional analysis of nitrogenase genes from the broad-host range *Rhizobium* strain ANU240. *Gene* **77,** 141–153 (1989).
22. van Slooten, J. C., Bhuvanasvari, T. V., Bardin, S. & Stanley, J. Two C4-dicarboxylate transport systems in *Rhizobium* sp. NGR234: rhizobial dicarboxylate transport is essential for nitrogen fixation in tropical legume symbiosis. *Mol. Plant Microbe Interact.* **5,** 179–186 (1992).
23. Broughton, W. J. *et al.* Identificaiton of *Rhizobium* plasmid sequences involved in recognition of *Psophocarpus, Vigna*, and other legumes. *J. Cell Biol.* **102,** 1173–1183 (1986).
24. Jabbouri, S. *et al.* in *The Biology of Plant–Microbe Interactions* (eds Stacey, G., Mullin, B. & Gresshoff, P.) 319–324 (International Society of Molecular Plant–Microbe Interactions, St Paul, MN, 1996).
25. Krishnan, H. B., Kuo, C.-I. & Pueppke, S. G. Elaboration of flavonoid-induced proteins by the nitrogen-fixing soybean symbiont *Rhizobium fredii* is regulated by both *nodD1* and *nodD2*, and is dependent on the cultivar-specificity locus, nolXWBTUV. *Microbiology* **141,** 2245–2251 (1995).
26. Broughton, W. J., Dilworth, M. J. & Passmore, I. K. Base ratio determination using unpurified DNA. *Anal. Biochem.* **46,** 164–172 (1972).
27. Broughton, W. J., Samrey, U. & Stanley, J. Ecological genetics of *Rhizobium meliloti*: symbiotic plasmid transfer in the *Medicago sativa* rhizosphere. *FEMS Microbiol.* **40,** 251–255 (1987).
28. Sullivan, J. T., Patrick, H. N., Lowther, W. L., Scott, D. B. & Ronson, C. W. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc. Natl Acad. Sci. USA* **92,** 8985–8989 (1995).
29. Martinez-Romero, E. & Cabellero-Mellado, J. *Rhizobium* phylogenies and bacterial genetic diversity. *Crit. Rev. Plant Sci.* **15,** 113–140 (1996).
30. Tepfer, D. in *Plant Microbe Interactions* (eds Kosuge, T. & Nester, E.) 294–342 (McGraw Hill, New York, 1989).
31. Galas, D. J. & Chandler, M. in *Mobile DNA* (eds Berg, D. E. & Howe, M. M.) 109–162 (American Society of Microbiology, Washington DC, 1989).
32. Borodovsky, M., Rudd, K. E. & Koonin, E. V. Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Res.* **22,** 4756–4767 (1994).

Acknowledgements. We thank E. Michaelis, D. Schnabelrauch, S. Landmann, S. Förste, K. Blechschmidt and G. Nordsiek for technical assistance; B. Drescher and D. Bauer for computational assistance; M. Y. Borodovsky and W. S. Hayes for GeneMark matrices; O. White for helping to create Fig. 1; S. Farrand and P. Murphy for advice on the conjugal transfer experiments; D. Gerber and S. Relić for help with many aspects of this work; and P. Rochepeau and V. Viprey for making their unpublished data available.

Correspondence and requests for materials should be addressed to A.R. (e-mail: arosenth@imb-jena.de). The sequence of pNGR234*a* has been deposited in EMBL/GenBank (accession no. U00090). Putative proteins are annotated in Swiss-Prot. Further information can be retrieved from the World-Wide Web (http://genome.imb-jena.de/archives.html).

Task difficulty and the specificity of perceptual learning

Merav Ahissar*† & Shaul Hochstein‡

* *Center for Higher Brain Functions, Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel*
‡ *Center for Nerual Computation, Department of Neurobiology, Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem 91904, Israel*

Practising simple visual tasks leads to a dramatic improvement in performing them. This learning is specific to the stimuli used for training. We show here that the degree of specificity depends on the difficulty of the training conditions. We find that the pattern of specificities maps onto the pattern of receptive field selectivities along the visual pathway. With easy conditions, learning generalizes across orientation and retinal position, matching the spatial generalization of higher visual areas. As task difficulty increases, learning becomes more specific with respect to both orientation and position, matching the fine spatial retinotopy exhibited by lower areas. Consequently, we enjoy the benefits of

† Present address: Keck Center for Integrative Neuroscience, UCSF, San Francisco, California 94143-0732, USA.