Available online at www.sciencedirect.com

ScienceDirect

Genomics 89 (2007) 237–247

GENOMICS

www.elsevier.com/locate/ygeno

Syntenic arrangements of the surface polysaccharide biosynthesis genes in *Rhizobium leguminosarum*

Jarosław E. Król, Andrzej Mazur *, Małgorzata Marczak, Anna Skorupska

Department of General Microbiology, Institute of Microbiology and Biotechnology, University of Maria Curie Skłodowska, 19 Akademicka Street, 20-033 Lublin, Poland

Received 7 July 2006; accepted 29 August 2006
Available online 2 October 2006

Abstract

We applied a genomic approach in the identification of genes required for the biosynthesis of different polysaccharides in *Rhizobium leguminosarum* bv. *trifolii* TA1 (RtTA1). Pulsed-field gel electrophoresis analyses of undigested genomic DNA revealed that the RtTA1 genome is partitioned into a chromosome and four large plasmids. The combination of sequencing of RtTA1 library BAC clones and PCR amplification of polysaccharide genes from the RtTA1 genome led to the identification of five large regions and clusters, as well as many separate potential polysaccharide biosynthesis genes dispersed in the genome. We observed an apparent abundance of genes possibly linked to lipopolysaccharide biosynthesis. All RtTA1 polysaccharide biosynthesis regions showed a high degree of conserved synteny between *R. leguminosarum* bv. *viciae* and/or *Rhizobium etli*. A majority of the genes displaying a conserved order also showed high sequence identity levels.

Keywords: *Rhizobium leguminosarum* core-accessory genome; Polysaccharide biosynthesis genes; BAC library

Introduction

Members of the genus *Rhizobium* and of related genera, referred to as rhizobia, are gram-negative α-proteobacteria that establish nitrogen-fixing symbioses with the roots of leguminous plants. Rhizobial genomes are complex and are composed of a chromosome plus none to many large plasmids, the set of which may represent up to 50% of the total genome [1]. The completion of nucleotide sequences of several rhizobial genomes confirmed that the genes participating in symbiosis, i.e., nodulation (*nod*) and nitrogen fixation (*nif-fix*), either are carried by the large plasmids, which are therefore called symbiotic plasmids (pSym) as in *Sinorhizobium meliloti* pSymA [2], *Rhizobium* sp. NGR234-pNGR234a [3], *Rhizobium etli* CFN42 p42d [4,5], and *Rhizobium leguminosarum* bv. *viciae* 3841 pRL10 [6], or are incorporated into particular stretches of the chromosome called symbiotic islands, as in *Mesorhizobium loti* MAFF303099 [7] and *Bradyrhizobium japonicum* USDA110 [8].

Among the known rhizobial genes required for symbiosis with legume plants, the genes responsible for the production of different types of cell-surface polysaccharides play a major role. The rhizobial surface is characterized by a variety of polysaccharides, such as lipopolysaccharides (LPS), capsular polysaccharides (CPS), and exopolysaccharides (EPS) [9,10]. In addition, cyclic β-(1,2)-glucans can be found in the periplasmic space, playing an important role in osmotic adaptation of bacteria [11]. Recently, Laus et al. [12] reported the presence of additional surface polysaccharides in *R. leguminosarum*: a high-molecular-weight neutral polysaccharide (HMW NP or glucomannan) and gel-forming polysaccharide (GPS). Among these, EPS is of special interest. It most likely plays numerous biologically relevant functions, including a mechanistic role in protecting the bacterium against environmental stresses or plant antimicrobial compounds and involvement in plant infection [9,12,13]. The genomic era revealed that genes directing the biosynthesis of surface polysaccharides can form large clusters located either on the

* Corresponding author.
E-mail address: mazur@hektor.umcs.lublin.pl (A. Mazur).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.08.015

chromosome, for instance, *M. loti, B. japonicum* [7,8] or on megaplasmids, for instance, pSymb in *S. meliloti* [14].

R. leguminosarum comprises two biovars, *viciae* and *trifolii*, that differ in their host specificity, and it is a close relative of *R. etli* (formerly the third biovar, *phaseoli*). Despite the completion of two genome projects for *R. etli* CFN42 (*Rhe*) [5] and *R. leguminosarum* bv. *viciae* 3841 (*Rlv*) [6], the data concerning the synthesis of surface polysaccharides in *R. leguminosarum* are still fragmentary.

The 7,751,309-bp *Rlv* genome consists of a circular chromosome and six plasmids. Most of the polysaccharide biosynthesis genes were mapped to the chromosome [6]. The *R. etli* genome (6,530,228 bp) is also partitioned and similarly comprises seven replicons: a circular chromosome (containing most of the genes implicated in the synthesis of the extracellular envelope) and six large plasmids [5].

In *R. leguminosarum*, the largest and probably best recognized region is the EPS biosynthesis (Pss-I) region, comprising several chromosomal genes, well characterized in bvs. *viciae* and *trifolii* and recently reviewed by Skorupska et al. [15]. It contains genes encoding glycosyltransferases responsible for EPS subunit synthesis and EPS modifying enzymes [16–18]. Moreover, genes whose products form the EPS assembly and export system [19,20] and other loci implicated in EPS modification were found within this region [21]. The *Rhe* Pss-I region contains additional open reading frames (ORFs) encoding putative epimerases, deacetylases, and glycosyltransferases [5]. The Pss-I locus is highly conserved between *R. etli* and *R. leguminosarum* but is absent from the other sequenced rhizobial genomes.

Additional genes with a function in surface polysaccharide biosynthesis are found dispersed in the *R. leguminosarum* genomes. This is the case for *exoB*, involved in the biosynthesis of various heteropolysaccharides (EPS, CPS, LPS, GPS, and glucomannan) [12,22], the conserved *pssA* gene [18,23], and regulatory genes *psi* and *psr* [24], *rosR* and *exoR* [25,26], and *pssB* [27]. The data regarding other surface polysaccharides in *R. leguminosarum* are incomplete. At least four potential *lps* regions, designated α, β, γ, and δ, have been identified, and some of them have been studied in detail [28–30]. Moreover, the *acpXL*, *lpxXL*, *lpxE* [31], and *lpxQ* genes [32] were recognized and characterized as being involved in LPS biosynthesis. Most of those genes are located on the chromosome; nevertheless a plasmid-borne β-*lps* region has been described in *R. etli* [33]. Several *cel* genes engaged in the cellulose fibril biosynthesis process in *R. leguminosarum* bv. *trifolii* have also been described [34].

Here we describe the application of a genomic approach in the identification of potential genes required for the biosynthesis of different types of polysaccharides in *R. leguminosarum* bv. *trifolii* TA1 (RtTA1). Sequencing of selected RtTA1 BAC library clones in conjunction with *R. leguminosarum* bv. *viciae* 3841 and *R. etli* CNF24 genome annotation-driven PCR amplification of hypothetical polysaccharide genes from the RtTA1 genome led to the identification of several large regions and gene clusters, as well as many separate potential polysaccharide biosynthesis genes dispersed in the genome. Pulsed-field gel electrophoresis (PFGE) of intact RtTA1 DNA followed by multiple Southern hybridizations allowed us to assess the plasmid sizes and localization of the identified polysaccharide biosynthesis genes. We also provide a computational analysis and a genomic perspective interpretation for the identified polysaccharide genes.

Results

Genome size determination and replicon identification

The PFGE analysis of intact genomic DNA revealed that the RtTA1 genome consists of five replicons: a chromosome and four plasmids, pRtTA1d, pRtTA1c, pRtTA1b. and pRtTA1a. Plasmid sizes, estimated on the basis of comparison of band migration with replicons of known size (*R. leguminosarum* bv. *viciae* 3841), are approximately 798, 646, 598, and 497 kb, respectively (Fig. 1).

Series of Southern hybridization experiments on undigested RtTA1 genomic DNA allowed the determination of the function of replicons. The *nod-nif* region was identified on pRtTA1a, which was thus designated symbiotic plasmid (pSym). Thiamine biosynthesis gene *thiC* was mapped to this replicon as well, whereas the *nadA* and *bioA* genes also required for the biosynthesis of group B vitamins were found on pRtTA1b and the chromosome, respectively. Surprisingly, *fixGHI* genes are not carried by pSym, but instead are present on the chromosome and pRtTA1c. Ribosomal RNA genes, *dnaK*, and the previously described *pss* genes were mapped to the chromosome (Table 1). Using a heterologous probe derived from the *Rlv repC* gene, we demonstrated that all four RtTA1 plasmids possess at least a putative *repC* gene-based replication apparatus; however, the presence of the entire *repABC* replication system that is most common in α-proteobacteria cannot be excluded.

Characterization of RtTA1 genomic library and STS marker identification

A BAC library of RtTA1 total genomic DNA consisting of 2591 clones was constructed with *Hin*dIII. Size determination,

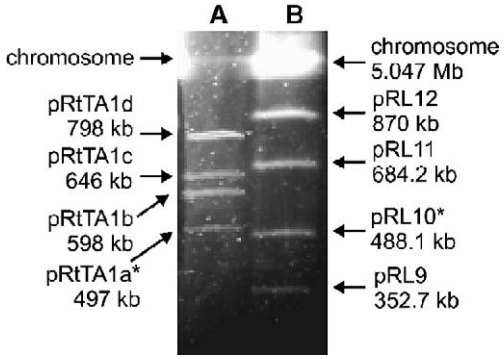

Fig. 1. PFGE analysis of undigested *R. leguminosarum* genomic DNA: (A) bv. *trifolii* TA1; (B) bv. *viciae* 3841. The plasmid sizes calculated for RtTA1 are the average of at least five separate experiments. pRL8 of *Rlv* is not visible under the applied experimental conditions. pSym plasmids are marked with an asterisk.

by *Hin*dIII digestion, of 298 randomly sampled clones indicated an average insert size of 40 kb, with a range of 10-108 kb. Consequently, a fivefold coverage of the entire genome would require the analysis of approximately 915 clones.

For further analysis, we have randomly selected 89 BAC clones containing the largest inserts (greater than 80 kb), the sizes of which were additionally confirmed by *Xba*I cleavage and PFGE (data not shown). Most of them were subjected to end sequencing using universal primers. A collection of unique 118 STS markers was thus obtained. BAC end sequences (STS) were searched against a nonredundant database from the National Center for Biotechnology Information (NCBI) using the BLASTX algorithm. Sequence similarities with known (and hypothetical) proteins were found for 82 of the query sequences, whereas in 36 cases, matched proteins of unknown function or no matches were found. Most sequences shared similarity with proteins involved in ammonium metabolism, surface polysaccharide biosynthesis and/or cell wall biogenesis, transport across the membranes, and regulatory functions (including transcription regulators). Moreover, sequences related to genes with roles in signal transduction, nucleotide and carbohydrate metabolism, mobility and chemotaxis, DNA replication, recombination, and repair, and RNA synthesis were found. Additionally, 2 STS markers matching transposon-related function were identified. Such a diversity of matches may be evidence for a good representation of genomic sequences in the constructed library. Most of the sequences share similarities with Rhizobiaceae family proteins (including hypothetical proteins), with the following hit distribution: 26 sequences from *Agrobacterium*, 22 from *S. meliloti*, 11 from *Mesorhizobium*, 5 from *Bradyrhizobium*, 3 from *R. leguminosarum*, and 2 from *Rhizobium* NGR234. All sequences were deposited in the GSS (Genome Survey Sequences) database at NCBI (GenBank Accession Nos. DX355039-DX355156).

Identification of BAC clones carrying polysaccharide biosynthesis genes

Subsequent analysis allowed us to identify the BAC clone carrying the previously described *pss* genes involved in EPS production. The pool of previously selected 298 BAC clones was probed with two sets of primers for the *pssT* and *pssL* genes, respectively. Only one clone (designated BAC-G1), with an insert size of approximately 81 kb, gave a strong signal in the PCRs. The presence of other *pss* genes within the BAC-G1 insert was further confirmed. Specific PCR amplicons of predicted sizes were obtained for *pssN*, *pssP*, *prsD*, and *pssD* but not for *pssB* and *pssA* genes (data not shown). This is in agreement with the previous observations [18,23] that *pssA* and *pssB* genes are not closely linked to other *pss* genes, although they are present in the RtTA1 genome. The BAC-G1 insert was further sequenced, either by subcloning of restriction fragments into plasmid vectors or by primer walking. This led to the identification of a 33.7-kb region (designated Pss-I) (Fig. 2A, Table 1), comprising 24 polysaccharide biosynthesis genes, 11 of which were previously described in RtTA1. The other identified genes were *regA, mgl2, pssV*, *pssW*, *pssS*, *pssR, pssM, pssK, pssJ*, *pssI*, *pssF1*, *orf1*, and *exo5* (Fig. 2A, Table 1).

The working pool of 298 BAC clones was further analyzed using PCR primers whose design was based on sequences of different rhizobial polysaccharide genes available in the GenBank database. This approach allowed us to select several BAC clones whose inserts were then partially sequenced by subcloning and primer walking. Approximately 100 kb of different plasmid subclones were read in total. The BAC-G11 insert carries a large polysaccharide biosynthesis region, designated Pss-IV (Table 1). It consists of two separate but closely linked clusters, designated Pss-IVA and Pss-IVB, respectively. Cluster Pss-IVA contains four genes, *ddhA, ddhB, orf7*, and *rfbC1*, whereas six genes, namely, *purU, exoZ, orf9, egl1, wzz10*, and *orf8,* make up Pss-IVB (Fig. 2D, Table 1). The Pss-V, or *rfb*, region consisting of four genes (*rfbA, rfbD, rfbB,* and *rfbC*) was identified in the BAC-G223 insert (Fig. 2E). RtTA1 BAC library screening also led to the identification of an *lpc* cluster containing putative *lpcA* and *lpcB* loci (BAC-G5) and several other genes possibly involved in LPS biosynthesis: *lpxQ* and *kdtA*, located in BAC-G99 and BAC-G40, respectively. A putative *exoR* regulatory gene was identified in BAC-G59, whereas BAC-G68 contains a paralogue of *pssA*, designated *pssA2*. Additionally, the cellulose synthesis gene *celA* was found to be carried by BAC clones designated G57 and G72 (Table 1).

With Southern hybridization analyses, Pss regions I, IV, and V and the *lpc* cluster were mapped to the RtTA1 chromosome (Table 1). Similarly, most of the polysaccharide biosynthesis genes identified during BAC library screening were found dispersed throughout the chromosome, except for the *lpxQ* locus, which was located on the pRtTA1b plasmid.

Identification of other potential polysaccharide genes in RtTA1

The search for polysaccharide biosynthesis genes in the RtTA1 genome was also conducted with an alternative approach based on the outputs from automatic annotation of the sequenced *R. leguminosarum* bv. *viciae* genome, generated by the PEDANT (Protein Extraction, Description, and Analysis Tool) algorithm. Significant sequence similarity between the *viciae* and *trifolii* biovars allowed us to design a series of PCR primer pairs that were further used for the amplification of potential polysaccharide genes. We have isolated two large clusters of polysaccharide biosynthesis genes, designated Pss-II and Pss-III, located on the chromosome and pRtTA1b, respectively (Figs. 2B and C, Table 1). Pss-II (14 kb) contains 10 genes: *gltA8, orf5, pssF2, gltA2, pssP2, pssY, crp1, gltA1, gltB1,* and *pssL2*. The 8.5-kb sequence of Pss-III comprises 7 putative genes, *pssF, gltB2, gltC1, orf4, orf5, orf6*, and *pssM*. Moreover, many other genes, possibly with a function in the biosynthesis of different surface polysaccharides, were found, being part of putative larger clusters as well as existing as separate ORFs on the RtTA1 chromosome (Table 1). It should be noted that the *lpsB2* gene was mapped to the pRtTA1b plasmid.

Table 1
Database accessible genes and genetic regions positioned by Southern hybridization in the *R. leguminosarum* bv. *trifolii* TA1 genome

Gene, cluster, or region name/ GenBank Accession No.	Sequence source [a]/ genome location [b]	Protein product description, protein family (if known)	Closest protein similarity, GenBank Accession No.
Pss-I/DQ384110	BAC-G1/ch		
regA		Putative regulatory protein, similar to helix-turn-helix XRE- like family proteins	ORF 1.1 *Rlv* AAK77322
mgl2		Putative methyltransferase	RHE_CH03209 *Rhe* YP_470701
pssV		Putative polysaccharide biosynthesis	PssV *Rlv* AAK77323
pssW		Putative glycosyl hydrolase family 10	PssT *Rlv* AAK77325
pssS		Putative glycosyl transferase group 1	PssS *Rlv* AAK77325
pssR		Putative acetyltransferase	PssR *Rhe* YP_470705
pssM		Putative ketal pyruvate transferase	PssM *Rlv* AAK77315
pssL		Polysaccharide transporter/flippase, Wzx-like exopolysaccharide translocase	PssL *Rlv* AAK77316
pssK		Putative polysaccharide pyruvyl transferase	PssK *Rhe* YP470708
pssJ		Putative galactosyl transferase	PssJ *Rlv* AAB88891
pssI		Putative glycosyl transferase	PssI *Rlt* AAB95424
pssF1		Putative glycosyl transferase group 2	PssF *Rlv* CAA73298
pssC		Putative glycosyl transferase	PssC *Rlv* CAA73297
pssD		Putative glycosyl transferase	PssD *Rlv* AAB88897
pssE		Putative glucuronosyl transferase, related to UDP-glucuronosyltransferase	PssE *Rlv* AAB88898
orf3		Putative acyl-CoA-thioesterase, SGNH hydrolase subfamily	ORF 3.1 *Rlv* AAK77318
prsE		HlyD family secretion protein	PrsE *Rlv* CAA73293
prsD		ABC transporter protein	PrsD *Rlv* CAA73292
pssP		Exopolysaccharide polymerization protein, involved in EPS chain-length determination, MPA1 (Wzc) family	PssP *Rhe* YP470724
pssO		Putative exopolysaccharide membrane protein	PssO *Rhe* YP470725
pssN		Putative exopolysaccharide export outer membrane protein, OMA (Wza) family	PssN *Rhe* YP470726
pssT		Wzy-like exopolysaccharide polymerase	PssT2 *Rhe* YP470727
orf1		Hypothetical protein	
exo5		Putative UDP-glucose dehydrogenase	RkpK *Rhe* YP470729
Pss-II/DQ384109	PCR/ch		
gltA8		Putative glycosyl transferase family 8, RfaJ-like protein, possibly involved in LPS biosynthesis	RHE_*CH03391 Rhe* YP470881
orf5		Hypothetical inner membrane protein	RHE_CH03389 *Rhe* YP470879
pssF2		Putative exopolysaccharide biosynthesis/export outer membrane protein, OMA (Wza) family	ExoF2 *Rhe* YP470878
gltA2		Putative glycosyl transferase family 2	RHE_CH03387 *Rhe* YP470877
pssP2		Putative MPA1 (Wzz) family protein, possibly involved in polysaccharide chain-length determination	RHE_CH03386 *Rhe* YP470876
pssY		Putative glycosyl transferase, possibly involved in LPS biosynthesis	ExoY *Rhe* YP470875
crp1		Putative transcriptional regulator protein	RHE_CH03384 *Rhe* YP470874
gltA1		Putative glycosyl transferase group 1	RHE_CH03383 *Rhe* YP470873
gltB1		Putative glycosyl transferase group 1	RHE_CH03382 *Rhe* YP470872
pssL2		Putative polysaccharide biosynthesis	RHE_CH03381 *Rhe* YP470871
Pss-III/DQ417329	PCR/pRtTA1b		
pssF		Putative exopolysaccharide biosynthesis/export outer membrane protein, OMA (Wza) family	RL110389 *Rlv* Q1M5Z7
gltB2		Putative glycosyl transferase family 2	RL110390 *Rlv* Q1M5Z6
gltC1		Putative glycosyl transferase family 1, possibly involved in LPS biosynthesis	RL110391 *Rlv* Q1M5Z5
orf4		Hypothetical integral inner membrane protein	RL110392 *Rlv* Q1M5Z4
orf5		Hypothetical cytoplasmic protein	RL110393 *Rlv* Q1M5Z3
orf6		Hypothetical integral inner membrane protein	RL110394 *Rlv* Q1M5Z2
pssM		Putative glycosyltransferase family 2	RL110395 *Rlv* Q1M5Z1
Pss-IVA/DQ683725	BAC-G11/ch		
ddhA		Putative glucose-1-phosphate cytidylyltransferase	RL0246 *Rlv* Q1MMR6
ddhB		Putative CDP-glucose 4,6-dehydratase	RL0245 *Rlv* Q1MMR7
orf7		Putative NDP-hexose 3-*C*-methyltransferase	RL0244 *Rlv* Q1MMR8
rfbC1		Putative dTDP-4-dehydrorhamnose 3,5-epimerase, contains RfbC conserved domain	RL0243 *Rlv* Q1MMR9

Table 1 (*continued*)

Gene, cluster, or region name/ GenBank Accession No.	Sequence source[a]/ genome location[b]	Protein product description, protein family (if known)	Closest protein similarity, GenBank Accession No.
Pss-IVB/DQ778610	BAC-G11/ch		
purU		Putative formyltetrahydrofolate deformylase	RL0233 *Rlv* Q1MMT0
exoZ		Putative acetyltransferase	RL0234 *Rlv* Q1MMS9
orf9		Putative transmembrane exopolysaccharide biosynthesis protein	RL0235 *Rlv* Q1MMS8fs
egl1		Putative endoglucanase (cellulase)	RL0236 *Rlv* CAK05725
wzz10		Putative polysaccharide transport protein, Wzz family	RL0237 *Rlv* Q1MMS7
orf8		Hypothetical protein	RL0238 Rlv Q1MMS5
Pss-V (*rfb*)/DQ679959	BAC-G223/ch		
rfbA partial cds		Putative glucose-1-phosphate thymidylyltransferase protein	RfbA *Rlv* Q1MIU2
rfbD		Putative dTDP-4-dehydrorhamnose reductase protein	RfbD *Rlv* Q1MIU1
rfbB		Putative dTDP-glucose 4,6-dehydratase protein	RfbB *Rlv* Q1MIU0
rfbC partial cds		Putative dTDP-rhamnose-3,5-epimerase protein	RfbC *Rlv* Q1MIT9
lpc cluster/DQ471905	BAC-G5/ch		
lpcA partial cds		Putative galactosyl transferase	LpcA *Rlph* CAA64421
lpcB partial cds		Putative CMP KDO transferase	LpcB *Rlph* CAA64420
acpXL/ DQ140392	PCR/ch	Putative acyl carrier protein XL	AcpXL *Rhe* YP_469982
lpsB2 partial cds/ DQ677348	PCR/pRtTA1b	Putative DTDP-glucose 4,6-dehydratase, O-antigen biosynthesis protein	
lpxQ partial cds/ DQ836933	BAC-G99/ pRtTA1b	Putative lipid A oxidase	LpxQ *Rlv* AY228164
kdtA partial cds/ DX355097	*BAC-G40/ch*	3-deoxy-D-manno-octulosonic acid transferase protein	KdtA *Rhe* ABC89655
exoR/ DQ347956	BAC-G59/ch	Putative exopolysaccharide production negative regulator	ExoR *Rlv* Q52822
pssA2/ DQ515971	BAC-G68/ch	Putative exopolysaccharide production protein, paralogue of *pssA*	PssA3 *Rhe* YP_471341
ndvA partial cds/ DQ471907	PCR/ch	Putative cyclic β 1-2 glucan synthetase protein	NdvB *Rhe* YP471468
celA partial cds/ DX355109	BAC-G57/ chromosome	Putative cellulose synthase	CelA *Rlt* AAD28574
Symbiotic genes			
fixGHI region DQ314612	PCR/ch and pRtTA1c		
fixI partial cds		Putative nitrogen fixation protein	FixI *Rlv* VF39 O33533
fixH		Putative cation pump membrane protein	FixH *Rlv* VF39 CAA04806
fixG partial cds		Putative ferrodoxin oxidase protein	FixG *Rlv* VF39 CAA04805
nif region DQ471906	PCR/pRtTA1a		
nifN partial cds		Putative nitrogenase iron-molybdenum cofactor biosynthesis protein	NifN *Rhe* NP_659840
nifE partial cds		Putative nitrogenase molybdenum-cofactor synthesis protein	NifE *Rhe* NP_659839
nodA partial cds/*AY904443*	PCR/pRtTA1a	Putative nodulation protein A	NodA *Rlv* P04338
Other genes or sequences			
dnaK partial cds/ DQ535895	PCR/ch	Putative heat shock protein 70 family	DnaK *Rhe* AAW82901
nadA partial cds/ DQ521662	PCR/pRtTA1b	Putative quinolinate synthetase A	NadA *Rhe* ABC93876
bioA partial cds/ DQ535896	PCR/ch	Putative adenosylmethionine-8-amino-7-oxononanoate aminotransferase	RHE_CH03090 *Rhe* ABC91856
thiC partial cds/ DQ535897	PCR/pRtTA1a	Putative thiamine biosynthesis protein	ThiC *Rhe* AAC45972
16-23 S rRNA intergenic region/ DQ639765	PCR/ch		*Rlv* AY491949

Note. *Rlv, R. leguminosarum* bv. *viciae*; *Rhe*, *R. etli*; *Rlt*, *R. leguminosarum* bv. *trifolii*; *Rlph*, *R. leguminsorum* bv. *phaseoli*.

[a] Given as BAC clone number if originated from RtTA1 BAC library or as PCR if amplified from RtTA1 genomic DNA.

[b] ch, chromosome.

Computational analysis of the identified polysaccharide biosynthesis genes

During the primary stage of analysis, we searched for sequence identities of the putative polysaccharide biosynthesis genes and aimed to identify orthologues in the sequenced genomes of the closely related rhizobial species *Rlv* and *Rhe*. Next, we examined the clustered genes for synteny, i.e., a conserved gene arrangement on bacterial replicons. We also performed GC3S (G+C content of synonymous third position)

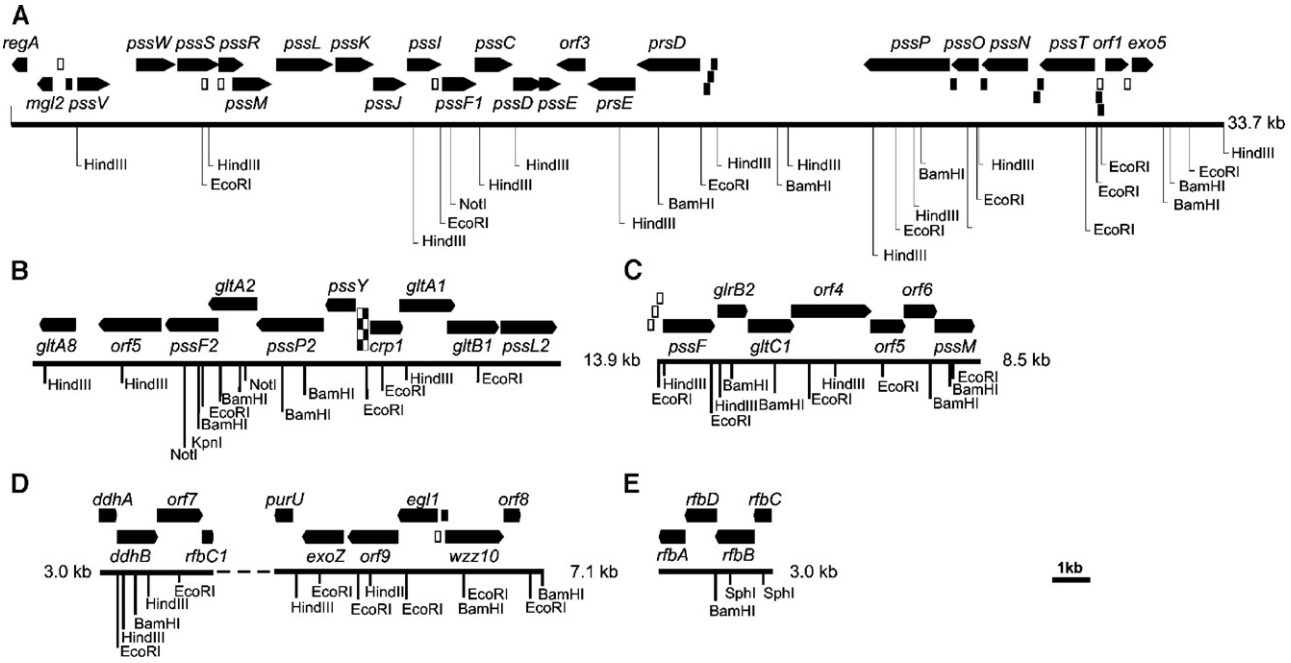

Fig. 2. Physical and genetic map of Pss regions of *R. leguminosarum* bv. *trifolii* TA1. Predicted promoter sequences whose probability of functioning as a promoter is greater than 0.9 are indicated as white boxes for forward strand and as black boxes for reverse strand. Selected restriction sites are marked. (A) Pss-I, (B) Pss-II, (C) Pss-III, (D) Pss-IV, (E) Pss-V.

analysis to ascribe a core or accessory [6] genomic role to the polysaccharide biosynthesis regions.

The largest identified RtTA1 polysaccharide biosynthesis region is Pss-I. It is connected with EPS biosynthesis, as determined in a number of recent studies describing mutations located inside this region that resulted in an EPS-deficient phenotype [16,19]. Pss-I showed high sequence identity with the homologous chromosomal regions of *Rlv* and *Rhe*, and orthologues of all RtTA1 *pss* genes were identified. Most of the predicted Pss-I chromosomal orthologues were syntenic when RtTA1, *Rlv*, and *Rhe* respective regions were compared using a WebACT program (Fig. 3A). Despite the high degree of conservation between the *Rlv* and *Rhe* Pss-I regions, important differences in coding region organization were found within RtTA1 Pss-I (Fig. 3A). First, the *regA* and *mgl2* genes are located upstream of the RtTA1 *pssV* gene. Orthologues of *regA* were found in both *Rhe* and *Rlv*, but *mgl2* was not found in the *Rlv* Pss-I cluster. Next, the pseudogene *pssU* present downstream of *pssV* in both *Rlv* and *Rhe* is missing in RtTA1. The most significant difference in the RtTA1 Pss-I organization involved the absence of two potential glycosyltranferase-encoding genes (*pssH* and *pssG*) that are located in both the *Rlv* and *Rhe* genomes downstream of *pssI* (Fig. 3A). Resequencing of the RtTA1 segment between *pssI* and *pssC* confirmed the differences mentioned above, and *pssH* and *pssG* were not found in the entire genome. In *Rhe*, downstream of *pssC* a putative ORF (named *yhch00973*) with similarity to O-antigen polymerase WzyC was detected; however, this locus was found in neither RtTA1 nor *Rlv*. Significant sequence variations and changes in the RtTA1 gene order were observed in the DNA stretch between the *prsD* and *pssP* genes. In the corresponding regions of *Rlv* and *Rhe*, at least two ORFs with similarity to known bacterial proteins PlyA and RapA are located (Fig. 3A). None of those ORFs were identified in RtTA1. Detailed sequence analysis revealed that several basepair deletions, insertions, or substitutions in this RtTA1 segment (confirmed not to be sequencing errors) are responsible for the local differences, which resulted in the prediction of some short, residual ORFs sharing similarity with PlyA and RapA proteins.

A synteny was observed in the case of the RtTA1 chromosomal Pss-II region and the corresponding regions of *Rlv* and *Rhe*, although upstream of *gltA8* in RtTA1, the orthologue of *Rlv orf61* and *Rhe yhch01007* is missing (Fig. 3B). Both of these genes coded for very short hypothetical products sharing no similarity to known proteins. In Pss-II, three genes for putative glycosyltransferases (*gltA8*, *gltA1*, and *gltB1*) distantly related to lipopolysaccharide biosynthesis proteins RfaJ, WcaO, and WbaZ, respectively, as well as a hypothetical PssL2 protein with similarity to WzxC, were found. The presence of these genes may suggest that Pss-II may be engaged in LPS biosynthesis. On the other hand, a pair of *wzx* and *wza* homologues may also indicate Pss-II involvement in EPS or CPS biosynthesis. The sequence of RtTA1 Pss-II is almost identical with the corresponding part of the *Rlv* chromosome and only minor changes were detected in comparison to *Rhe*. Most were confined to the intergenic sequences, especially in the segment between *pssY* and *crp1* of RtTA1 (Fig. 3B), where hypothetical promoters on both DNA strands were detected (Fig. 2B). This allowed the prediction that Pss-II might be organized as two separate transcriptional units: one comprising *crp1*, *gltA1*, *gltB1*, and *pssL2* and another with the *pssY*, *pssP2*, *gltA2*, *pssF2*, *orf5* and *gltA8* genes.

Pss-III is the RtTA1 plasmid-borne polysaccharide biosynthesis region. Functional predictions for putative Pss-III genes

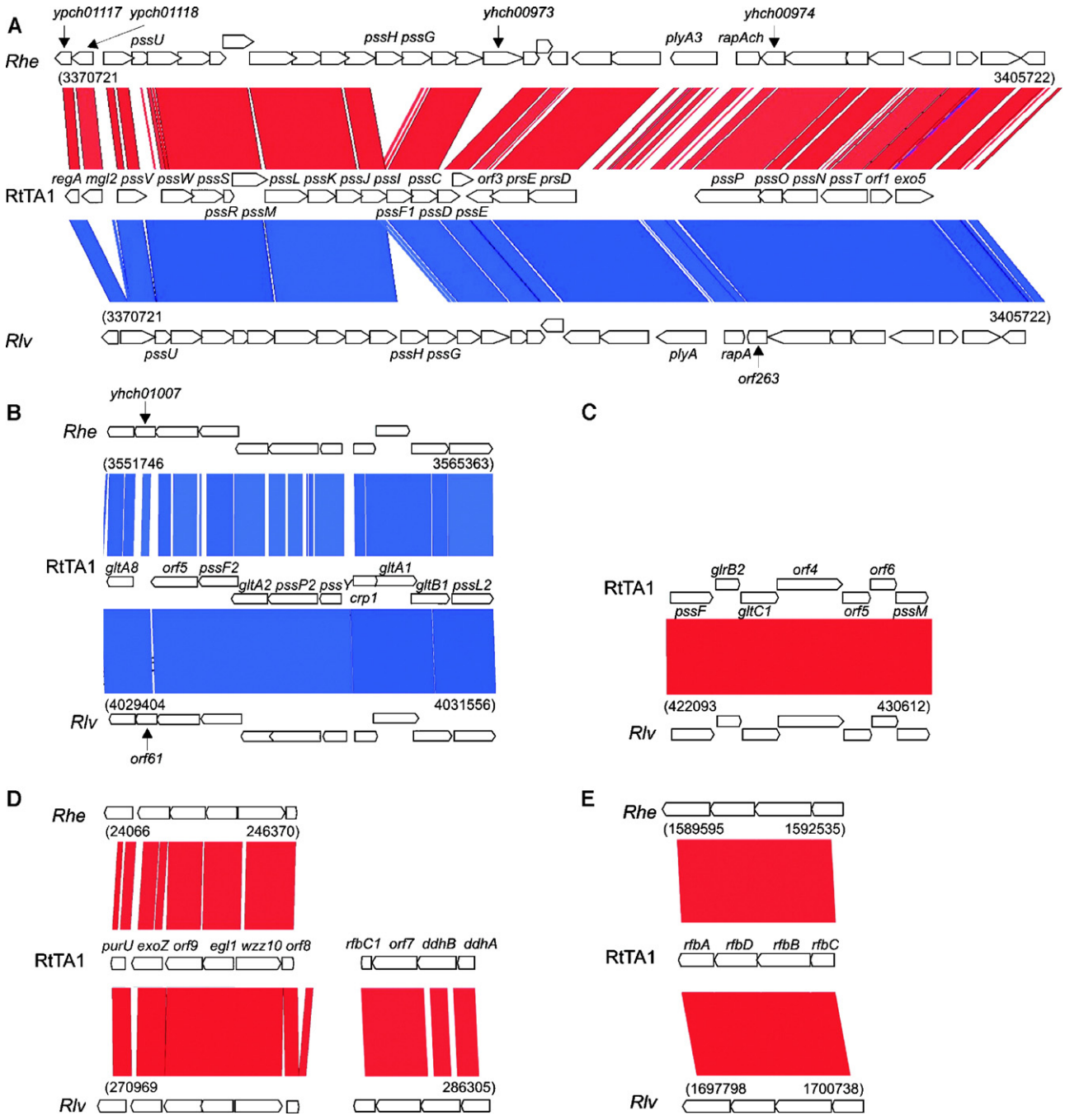

Fig. 3. Synteny between Pss regions of RtTA1 and related regions of *Rhe* and *Rlv*. Red and blue blocks match syntenic regions. Blue blocks refer to sequences that were inverted to obtain maximal colinearity. The numbers in parentheses indicate the location of syntenic regions on the respective replicons of *Rhe* and *Rlv*. For *Rhe* and *Rlv* Pss regions only, additional (not found in RtTA1) genes are indicated. (A) Pss-I, (B) Pss-II, (C) Pss-III, (D) Pss-IV, (E) Pss-V.

are summarized in Table 1. For Pss-III, DNA sequence identity greater than 95% and absolute synteny with the corresponding region of *Rlv* pRL11 were detected (Fig. 3C). A homologous segment was not found in the *Rhe* genome, neither on the chromosome nor on the plasmids. Only a slight similarity was shown for the RtTA1 *pssF* gene product with ExoF1 and ExoF2 reiterated on *Rhe* chromosome. Upstream of *pssF*, three potential promoters were predicted, raising the possibility that the whole region is transcribed as a single unit (Fig. 2C).

Pss-IV is yet another chromosomal polysaccharide biosynthesis region of RtTA1, divided into two closely linked clusters, Pss-IVA and IVB (Fig. 2D). Functional predictions of the putative ORFs encoded by Pss-IVA strongly suggested its role in LPS biosynthesis (Table 1). Similarly to Pss-III, highly syntenic orthologues of Pss-IVA were found only in the *Rlv* genome (Fig. 3D); therefore, these two regions seem to be specific to RtTA1 and *Rlv*. In both instances, extraordinary sequence identity between the biovars was observed. On the other hand, orthologues of RtTA1 Pss-IVB were found in both

the *Rlv* and *Rhe* genomes. Moreover, the *purU* gene coding for a putative formyltetrahydrofolate deformylase was found on the chromosome and pRtTA1c. The same repeats were identified in the *Rlv* and *Rhe* genomes on pRL11 and p42e, respectively. Significant sequence identity and conserved gene arrangement of Pss-IVB putative ORFs have been detected (Fig. 3D). As it is linked with Pss-IVA, its possible function in LPS biosynthesis cannot be excluded. Upstream of the *egl1* and *wzz10* genes, potential promoter sequences have been predicted. Thus, it is possible that Pss-IVB could be organized as two separate, divergently transcribed units (Fig. 2D).

Based on significant sequence similarity between the hypothetical proteins encoded by the chromosomal region Pss-V (*rfbADBC*) with proteins such as RmlADB from *Escherichia coli* and RmlC from *Shigella boydii* whose role in LPS O-antigen biosynthesis is well established and experimentally verified, their function in LPS biosynthesis could be proposed. The Pss-V region is very similar to the respective regions in *Rlv* and *Rhe* (95 and 92% DNA sequence identity), and the same was observed when gene arrangement was compared (Fig. 3E).

It should be noticed that no synteny of RtTA1 Pss regions with other (other than *Rlv* and *Rhe*) rhizobial genomes was observed.

For the *Rlv* genome, the GC content and especially the GC3S variations were recently proposed as being sensitive determinants of a particular chromosome region functioning as a core or accessory segment [6]. A core genome is higher in G+C content and is generally chromosomal, whereas an accessory component displays lower G+C content and is located mainly on plasmids and chromosomal islands. Table 2 summarizes the numerical data for the RtTA1 polysaccharide biosynthesis regions and their syntenic equivalents in *Rlv* and *Rhe*. The highest G+C and GC3S values were calculated for the chromosomal Pss-V (*rfb*), Pss-IVA, Pss-IVB, and Pss-II regions and *lpc* clusters in all the tested genomes, whereas both these parameters were decreased for plasmid-borne Pss-III in RtTA1 and *Rlv* (Table 2). Surprisingly, the lowest G+C value and GC3S percentage were computed for the chromosomal Pss-I region: in all three genomes, they were even below the level of G+C and GC3S values calculated for plasmid-located Pss-III regions in RtTA1 and *Rlv* (Table 2). On the other hand, the plasmid-located *lpsB2* gene showed a high GC3S content in RtTA1, similarly to *Rlv* and *Rhe* genomes, and was comparable to the values for a chromosomal paralogue of *lpsB2* in *Rlv* (Table 2).

Table 2
Nucleotide composition of chromosomally and extrachromosomally located polysaccharide biosynthesis genes of RtTA1, *Rlv*, and *Rhe*, shown as G+C and GC3S [a]

Region name/ genome location	% G+C/GC3S		
	RtTA1	*Rlv*	*Rhe*
Pss-I/ch	56.6/54.1	56.8/55.3	57.5/56.7
Pss-II/ch	59.8/59.0	59.9/60.5	64.6/60.1
Pss-III/pl	60.0/57.3	60.1/56.2	n.d.
Pss-IVA/ch	63.5/66.2	63.7/64.4	n.d.
Pss-IVB/ch	61.4/61.9	62.5/60.7	61.7/62.3
Pss-V (or *rfb*)/ch	61.6/66.1	60.9/64.0	60.6/61.7
lpc cluster/ch	61.8/59.7	62.8/61.0	60.7/65.6
lpsB2/pl	59.2/60.2	57.7/67.9	57.3/68.3
lpsB2 [b]/ch	n.d	57.3/68.1	n.d.

Note. ch, chromosome; pl, plasmid; n.d., not detected.
[a] GC3S is the G+C content of the silent third position of codons.
[b] The *lpsB2* paralogue in the *Rlv* chromosome.

Discussion

We applied a genomic approach to identify potential genes required for biosynthesis of different types of polysaccharides in the *R. leguminosarum* bv. *trifolii* TA1 genome. As a first stage, we performed PFGE analyses of intact RtTA1 genomic DNA. The genome is partitioned and consists of five replicons: a circular chromosome and four large plasmids. It thus represents a genomic architecture typical for *R. leguminosarum*. The RepABC replicator is possibly present in the RtTA1 extrachromosomal DNA, similar to plasmids of *Rhe* and *Rlv* [5,6] and other rhizobia. Most of the symbiotic genes were mapped to the smallest plasmid, pRtTA1a, thus designated pSym. We showed that not all the symbiotic genes reside on pSym. Two copies of *fixGHI* genes were located on the chromosome and on pRtTA1c, respectively. Additional copies of symbiotic genes in the remaining replicons of the genome, aside from pSym, were previously reported for the *fixNOQP* and *fixGHIS* operon in *R. leguminosarum* VF39 [35], Mitsch et al. unpublished results, GenBank Accession No. AJ001522]. Southern hybridizations showed that in RtTA1, similarly to *Rlv* and *Rhe*, the majority of polysaccharide biosynthesis genes are chromosomal and there is no distinguishable pExo plasmid such as that described previously for *S. meliloti* carrying *exo/exs* clusters [14]. Despite this, the plasmids do not appear to be entirely dispensable, at least with respect to surface structure biosynthesis. Some polysaccharide synthesis loci were found to be located extrachromosomally: the Pss-III region and *lpxQ* and *lpsB2* on pRtTA1b. The possible advantages of multipartite rhizobial genomes could be explained in the manner recently proposed for *R. etli*: such arrangements might enhance the adaptive potential of the bacterium, allowing the reassortment of essential, nonessential, and redundant functions to contend with challenging environments [5].

The combined genomic approach reported here, comprising sequencing of selected RtTA1 BAC library clones and PCR amplification of polysaccharide biosynthesis genes from the RtTA1 genome, proved to be very useful in both detecting new genes and assigning their putative functions. Orthologues with conserved replicon localization and neighborhood predominated among the identified genes. Computational analysis showed that of the five large polysaccharide biosynthesis regions identified in this work, three could be potentially engaged in LPS biosynthesis (Table 1). Moreover, many other putative LPS biosynthesis genes and small clusters (*acpXL*, *lpc*, *lpxQ*, *lpsB*, and *kdtA*) were found dispersed throughout the genome. Such an apparent abundance of genes linked to LPS biosynthesis was previously reported for *R. leguminosarum*. In *R. etli*, four potential *lps* genetic regions designated α, β, γ, and δ (both chromosomal and plasmid-borne) have been identified

[29,30,33]. However, this raises a further and more general question as to the significance of the presence of numerous polysaccharide biosynthesis genes. It might be just a demonstration of the metabolic redundancy typical for rhizobial genomes, reported recently for *Rhe* and *Rlv* [5,6], and also observed in other rhizobial genomes sequenced to date, such as *Bradyrhizobium* and *Mezorhizobium* [7,8]. It is also possible that in the heterogeneous soil environment where the rhizobia thrive, and are exposed to different substrates and potential stresses, a genome that encodes numerous potential metabolic pathways is advantageous and versatile. Similarly this could be the case for polysaccharide biosynthesis genes. The surface of a rhizobial cell is characterized by a variety of polysaccharides [12] for which many potential biological functions have been postulated (for a review, see [15]). The reason for an abundance of polysaccharide biosynthesis genes could simply be a reflection of the necessity for rhizobial versatility. On the other hand, the diversity and amount of the identified RtTA1 polysaccharide biosynthesis genes seems to be inadequate, considering the complexity of the rhizobial cell surface. For example, RtTA1 (like *Rhe*) lacks *rkp* cluster homologues for CPS found in other rhizobia [10]. Therefore, it is possible that one of the identified Pss loci might encode an alternate pathway for the synthesis of another surface polysaccharide. In *R. leguminosarum,* acidic CPS has a structure that is similar or even identical to the acidic EPS [36,37], differing only in noncarbohydrate residues [38]. Thus, both Pss-I (or its components) and Pss-II or Pss-III containing *wza* homologues might be a good potential candidate for CPS biosynthesis in *R. leguminosarum*. Consequently, the putative functional overlapping of polysaccharides could be the reason for pleiotropic effects exerted by mutations in *pss* genes such as *exo5* [37], *exoB* [12,22], *pssD*, and *pssP* [39]. Further functional studies are necessary to verify this hypothesis.

We showed that all the large polysaccharide biosynthesis regions described here were syntenic and in all cases, except for Pss-III, chromosomal. A majority of the genes displaying a conserved order also showed high sequence identity levels. It had been previously established that in Rhizobiales, syntenic genes encode a high proportion of the essential cell functions and are chromosomal [40]. Thus, the tendency of syntenic gene products to retain higher identity levels could reflect functional constraints, as well as their possible essential character. This could explain the sequence and gene arrangement conservation of most of the RtTA1 Pss chromosomal regions since surface structure biosynthesis (at least some instances) could undoubtedly be considered a housekeeping process, indispensable for cell viability. Essentiality could also explain the chromosomal location of other polysaccharide biosynthesis genes. This concerns, e.g., *exoB* involved in EPS/CPS, LPS, and GPS biosynthesis [12,22] and *ndv* for the synthesis of cyclic β-(1,2)-glucans that play an important role in osmotic adaptation of bacteria [11].

It was also proposed [40] that the preservation of a conserved gene order is not a random event because an appropriate gene neighborhood confers an adaptive advantage to the cell, as the synthesized proteins are likely to perform related functions. This probably could be the reason for the synteny of plasmid-borne RtTA1 and *Rlv* Pss-III regions. However, the *Rhe* plasmids, in contrast to the chromosome, lack synteny [5].

Genome-wide sequence analysis of Rhizobiales [40] showed that syntenic chromosomal genes had a strong tendency to form operons. On the basis of promoter prediction, we showed that at least for some of the identified Pss regions, coupled transcription of their constituents was plausible, irrespective of replicon localization (e.g., the RtTA1 Pss-III region) and the postulated function, e.g., in EPS or LPS biosynthesis.

Young et al. [6] hypothesized that the *Rlv* genome can be considered to have two main components: a core component, which has a higher G+C content, is mostly chromosomal, and is shared with related organisms, and an accessory component located on the plasmids and chromosomal islands, which has a lower G+C content and is more sporadic in distribution. Our calculations of G+C and GC3S values for the polysaccharide biosynthesis regions of RtTA1 and the related *Rlv* and *Rhe* generally agreed with this assumption (Table 2). Plasmid-borne RtTA1 Pss-III has a lower GC3S content and a homologous region was found only in *Rlv*; thus, it could be regarded as an accessory genome part. However, there are some exceptions. Pss-IVA, despite chromosomal localization and the highest G+C and GC3S content indicating a core genome, was detected only in RtTA1 and *Rlv*. Nevertheless, some interesting patterns could be discerned: the regions Pss-II, Pss-IV, and Pss-V/*rfb* and the *lpc* cluster predicted to be potentially connected with LPS biosynthesis generally fulfill the requirements for a core genome. Surprisingly, the chromosomal Pss-I region, which is strongly conserved and found in all three genomes analyzed, showed G+C and GC3S values even below those calculated for the plasmid-borne Pss-III. Chromosomal variation in GC3S, such as that recently reported for *Rhe* and *Rlv*, often reveals chromosomal islands (e.g., genes of plasmid origin) incorporated into the chromosome, e.g., via mobile elements or phages, as has been described for a symbiosis island of *M. loti* [7]. Although no signs of mobile elements or other indications of horizontal gene transfer events were detected in the proximity of Pss-I, it cannot be excluded that this region was in fact transferred to the chromosome via potential rearrangements between replicons. As previously shown for pNGR234a and the *R. etli* genome, repeated DNA sequences can promote homologous recombination and thus lead to genomic rearrangements [5,41]. For *Rhe*, more than 200 reiterated DNA families could be found, indicating that genomic rearrangements might indeed be frequently occurring events [4,42], resulting in a structural complexity of rhizobial replicons [4,5,43]. DNA reiterations were detected, e.g., for RtTA1 *fixGHI* genes and also for polysaccharide biosynthesis genes. For example, *purU* was shown to be present both on the chromosome and on pRtTA1c. The chromosome of *Rlv* also possesses many distinct gene islands with typical accessory characteristics (low GC3S). It is likely that they form part of the accessory genome residing, probably temporarily, in the chromosome [6]. Nevertheless, our analysis of polysaccharide

Table 3
Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant characteristics	Source or reference
R. leguminosarum		
bv. *trifolii* RtTA1	Wildtype, Str^r, Rif^r	[49]
bv. *viciae* 3841	Wildtype, Str^r	[6]
E. coli		
DH5α	*supE*44 Δ*lac*U169 (ϕ80*lacZ*ΔM15) *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1	[44]
JW480	Gene Hods (Invitrogen) *araC* pBAD-*trf254* integrated at λatt	[50]
Plasmids		
pIndigoBAC5/oriV	*lacZ oriV oriS repE parABC* cos loxP Cm^r	[50]
pGEM 3Zf+/− pGEM T-Easy	*ori colE1*,*lacZ*, Ap^r	Promega

biosynthesis genes seems to be a useful indictor of the core and accessory genome, at least to some extent.

Materials and methods

Bacterial strains and growth conditions

Bacterial strains and plasmids used in this work are listed in Table 3. *E. coli* strains were grown at 37°C in Luria-Bertani or SOC medium [44]. For library storage, the freezing medium (LB medium supplemented with 13 mM KH_2PO_4, 36 mM K_2HPO_4, 1.7 mM sodium citrate, 6.8 mM $(NH_4)_2SO_4$, 0.4 mM $MgSO_4$, and 4.4% v/v glycerol) was used. Antibiotics were added at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 12.5 μg/ml; rifampicin, 40 μg/ml. When needed, 50 μg/ml 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-Gal), 25 μg/ml isopropyl β-D-thiogalactoside (IPTG), and 0.01% L-arabinose were added.

DNA methods

Standard techniques were used for plasmid isolation, restriction enzyme digestion, agarose gel electrophoresis, DNA cloning, labeling of DNA, and Southern hybridization [44]. The DNA of the BAC vector and BAC clones was extracted by alkaline lysis, followed by chloroform treatment and isopropanol precipitation. We routinely prepared DNA from 4.5 ml of overnight culture. Purified DNA was resuspended in 30 μl TE containing DNase-free RNase (20 μg/ml). PFGE was performed with the contour-clamped homogenous electric field mode with the Bio-Rad system (Model CHEF-DRIII). DNA samples were separated in 1% agarose gels in 0.5× TBE buffer, refrigerated at 12–14°C, with switch times of 30-80 s, angle 120°, voltage gradient 6 V/cm for 22 h. DNA probes for Southern hybridization were labeled with a nonradioactive DIG DNA Labeling and Detection Kit (Roche). Automatic sequencing was performed using the BigDye Terminator Cycle Sequencing Kit and ABI PRISM 310 sequencer (Applied Biosystems).

Preparation of high-molecular-weight DNA

The plugs were formed with 25 ml of a 48-h culture of RtTA1, which after centrifugation was resuspended in TE buffer and mixed with 2% LMP agarose (Sigma). Agarose-embedded cells were incubated at 37°C with TE and lysozyme (1.5 mg/ml) for 16 h and then for an additional 48 h in cell lysis buffer (1% sodium lauryl sarcosine, 50 mM EDTA, 50 mM Tris-HCl, pH 8.0) supplemented with proteinase K (0.5 mg/ml) at 50°C. Proteinase K was inactivated by PMSF (0.4 mg/ml) at 37°C for 1 h. Plugs were washed three times (30 min) with TE buffer and finally stored in TE at 4°C. The quality and quantity of DNA were analyzed by PFGE.

BAC library construction and insert sizing

The BAC library was constructed with *Hin*dIII. For partial digestion, five agarose plugs were incubated for 30 min in 100 μl *Hin*dIII reaction buffer on ice. Then, the buffer was replaced and 10 U of *Hin*dIII enzyme was added and incubated on ice for 15 min, followed by incubation for 7 min at 37°C; then to stop the partial digestion, the agarose plugs were placed on ice and 1/10 vol of 0.5 M EDTA was added. The agarose plugs were melted at 65°C for 10 min and digested with Agarase (Fermentas) according to the manufacturer's instructions. The 200 ng of the digested DNA was ligated with 100 ng of *Hin*dIII-digested, dephosphorylated IndigoBac5/oriV using 10 U T4 DNA Ligase (Fermentas) in a total volume of 100 μl (12°C for 16 h). The ligation mixture was heated at 65°C for 10 min and than drop-dialyzed for 90 min against MilliQ water, using VS 0.025-μm membranes (Millipore). Desalted DNA was electrotransformed into *E. coli* JW480 with Electro Cell Manipulator 630 (BTX) with settings of 2500 V, 150 Ω, and 50 μF in 2-mm-wide electroporation cuvettes (Bio-Rad). Transformed cells were diluted in 600 μl SOC medium, incubated for 1 h at 37°C, and plated on LB with Cm, X-gal, and IPTG. White recombinant colonies were manually picked, cultured, and stored at −80°C.

Insert size was determined by agarose electrophoresis after *Hin*dIII cleavage with the λ *Eco*RI/*Hin*dIII or λ *Hin*dIII as MW marker. Sizes of inserts were estimated using the Bio1D Gel Analysis Program (Vilber Lourmat). Selected clones were cleaved with *Xba*I and separated by PFGE.

Bioinformatic tools

Sequence data were analyzed with DNASTAR-Lasergene analysis software and Artemis, a DNA sequence viewer and annotation tool [45]. WebACT was used for sequence comparison and visualization [46]. Database searches were performed with the BLAST and FASTA programs at NCBI (Bethesda, MD) and the European Bioinformatic Institute (Hinxton, UK). The PEDANT [47] outputs of automatic analysis of genomic sequences of *R. leguminosarum* bv. *viciae* 3841 were used for polysaccharide gene detection. The GC3S content of gene clusters was calculated using CODONW [48]. Neural Network Promoter Prediction version 2.2 (Berkeley *Drosophila* Genome Project) was used for putative promoter sequence detection, with the minimum promoter score set to 0.9.

Acknowledgments

We thank Waclaw Szybalski (University of Wisconsin) for the BAC vector. Thanks are extended to Peter Young (University of York) and Andrew Johnston (University of East Anglia) for sending the *Rlv* 3841 strain. This work was supported by Grant 2 P04A 034 26 from the Committee for Scientific Research.

References

[1] R. Palacios, W.E. Newton, Genomes and genomics of nitrogen-fixing organisms, Springer, Dordrecht, 2005.
[2] M.J. Barnett, et al., Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 9883–9888.
[3] C. Freiberg, et al., Molecular basis of symbiosis between *Rhizobium* and legumes, Nature 387 (1997) 394–401.
[4] V. Gonzalez, et al., The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments, Genome Biol. 4 (2003) R36.
[5] V. Gonzalez, et al., The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 3834–3839.
[6] J.P. Young, et al., The genome of *Rhizobium leguminosarum* has recognizable core and accessory components, Genome Biol. 7 (2006) R34.
[7] T. Kaneko, et al., Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti, DNA Res. 7 (2000) 331–338.

[8] T. Kaneko, et al., Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110, DNA Res. 9 (2002) 189–197.
[9] A. Becker, A. Pühler, Production of exopolysaccharides, in: H.P. Spaink, A. Kondorosi, P.J.J. Hooykaas (Eds.), Rhizobiaceae, Kluwer Academic, Dordrecht/London, 1998, pp. 97–118.
[10] A. Becker, N. Fraysse, L. Sharypowa, Recent advances in studies on structure and symbiosis-related function of rhizobial K-antigens and lipopolysaccharides, Mol. Plant Microbe Interact. 18 (2005) 899–905.
[11] K.J. Miller, E.P. Kennedy, V.N. Reinhold, Osmotic adaptation by gram-negative bacteria: Possible role for periplasmic oligosaccharides, Science 231 (1986) 48–51.
[12] M.C. Laus, et al., Novel polar surface polysaccharide from *Rhizobium leguminosarum* binds host plant lectin, Mol. Microbiol. 59 (2006) 1704–1713.
[13] N. Fraysse, F. Couderc, V. Poinsot, Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis, Eur. J. Biochem. 270 (2003) 1365–1380.
[14] T.M. Finan, et al., The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont *Sinorhizobium meliloti*, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 9889–9894.
[15] A. Skorupska, M. Janczarek, M. Marczak, A. Mazur, J. Krol, Rhizobial exopolysaccharides: Genetic control and symbiotic functions, Microb. Cell Factors 5 (2006) 7.
[16] J. Król, et al., Molecular characterization of *pssCDE* genes of *Rhizobium leguminosarum* bv. *trifolii* strain TA1: *pssD* mutant is affected in exopolysaccharide synthesis and endocytosis of bacteria, Mol. Plant Microbe Interact. 11 (1998) 1142–1148.
[17] W.A.T. van Workum, et al., Cloning and characterization of four genes of *Rhizobium leguminosarum* bv. *trifolii* involved in exopolysaccharide production and nodulation, Mol. Plant Microbe Interact. 10 (1997) 290–301.
[18] M.R. Sadykov, T.V. Ivashina, A.A. Kanapin, M.G. Shliapnikov, V.N. Ksenzenko, Structure-functional organization of exopolysaccharide biosynthetic genes in *Rhizobium leguminosarum* bv. *viciae* VF39, Mol. Biol. 32 (1998) 797–804.
[19] A. Mazur, J.E. Król, J. Wielbo, T. Urbanik-Sypniewska, A. Skorupska, *Rhizobium leguminosarum* bv. *trifolii* PssP protein is required for exopolysaccharide biosynthesis and polymerization, Mol. Plant Microbe Interact. 15 (2002) 388–397.
[20] A. Mazur, J.E. Król, M. Marczak, A. Skorupska, Membrane topology of PssT, the transmembrane protein component of the type I exopolysaccharide transport system in *Rhizobium leguminosarum* bv. *trifolii* strain TA1, J. Bacteriol. 185 (2003) 2503–2511.
[21] C. Finnie, N.M. Hartley, K.C. Findlay, J.A. Downie, The *Rhizobium leguminosarum prsDE* genes are required for secretion of several proteins, some of which influence nodulation, symbiotic nitrogen fixation and exopolysaccharide modification, Mol. Microbiol. 25 (1997) 135–146.
[22] M.W. Breedveld, et al., Polysaccharide synthesis in relation to nodulation behavior of *Rhizobium leguminosarum*, J. Bacteriol. 175 (1993) 750–757.
[23] M. Janczarek, A. Skorupska, Exopolysaccharide synthesis in *Rhizobium leguminosarum* bv. *trifolii* is related to various metabolic pathways, Res. Microbiol. 154 (2003) 433–442.
[24] J.W. Latchford, D. Borthakur, A.W.B. Johnston, The products of *Rhizobium* genes, *psi* and *pss*, which affect exopolysaccharide production, are associated with the bacterial cell surface, Mol. Microbiol. 5 (1991) 2107–2114.
[25] M.A. Bittinger, J.L. Milner, B.J. Saville, J. Handelsman, *rosR*, a determinant of nodulation competitiveness in *Rhizobium etli*, J. Bacteriol. 10 (1997) 180–186.
[26] W.G. Reeve, M.J. Dilworth, R.P. Tiwari, A.R. Glenn, Regulation of exopolysaccharide production in *Rhizobium leguminosarum* biovar *viciae* WSM710 involves *exoR*, Microbiology 143 (1997) 1951–1958.
[27] M. Janczarek, A. Skorupska, *The Rhizobium leguminosarum* bv. *trifolii pssB* gene product is an inositol monophosphatase that influences exopolysaccharide synthesis, Arch. Microbiol. 175 (2001) 143–151.
[28] J.R. Cava, P.M. Elias, D.A. Turowski, K.D. Noel, *Rhizobium leguminosarum* CFN42 genetic regions encoding lipopolysaccharide structures essential for complete nodule development on bean plants, J. Bacteriol. 171 (1989) 8–15.
[29] D. Allaway, B. Jeyaretnam, R.W. Carlson, P.S. Poole, Genetic and chemical characterization of a mutant that disrupts synthesis of the lipopolysaccharide core tetrasaccharide in *Rhizobium leguminosarum*, J. Bacteriol. 178 (1996) 6403–6406.
[30] R.W. Carlson, B. Reuhs, T.B. Chen, U.R. Bhat, K.D. Noel, Lipopolysaccharide core structures in *Rhizobium etli* and mutants deficient in O-antigen, J. Biol. Chem. 270 (1995) 11783–11788.
[31] M.J. Karbarz, S.R. Kalb, R.J. Cotter, C.R. Raetz, Expression cloning and biochemical characterization of a *Rhizobium leguminosarum* lipid A 1-phosphatase, J. Biol. Chem. 278 (2003) 39269–39279.
[32] N.L. Que-Gewirth, M.J. Karbarz, S.R. Kalb, R.J. Cotter, C.R. Raetz, Origin of the 2-amino-2-deoxy-gluconate unit in *Rhizobium leguminosarum* lipid A. Expression cloning of the outer membrane oxidase LpxQ, J. Biol. Chem. 278 (2003) 12120–12129.
[33] A. Garcia-de los Santos, S. Brom, Characterization of two plasmid-borne *lps* beta loci of *Rhizobium etli* required for lipopolysaccharide synthesis and for optimal interaction with plants, Mol. Plant Microbe Interact. 10 (1997) 891–902.
[34] N. Ausmees, H. Jonsson, S. Hoglund, H. Ljunggren, M. Lindberg, Structural and putative regulatory genes involved in cellulose synthesis in *Rhizobium leguminosarum* bv. *trifolii*, Microbiology 145 (1999) 1253–1262.
[35] A. Schluter, et al., Functional and regulatory analysis of the two copies of the *fixNOQP* operon of *Rhizobium leguminosarum* strain VF39, Mol. Plant Microbe Interact. 10 (1997) 605–616.
[36] B.K. Robertsen, P. Aman, A.G. Darvill, M. McNeil, P. Albersheim, Host-symbiont interactions. V. The structure of acidic extracellular polysaccharides secreted by *Rhizobium leguminosarum* and *Rhizobium trifolii*, Plant Physiol. 67 (1981) 389–400.
[37] M.C. Laus, et al., Involvement of *exo5* in production of surface polysaccharides in *Rhizobium leguminosarum* and its role in nodulation of *Vicia sativa* subsp. *nigra*, J. Bacteriol. 186 (2004) 6617–6625.
[38] M.A. O'Neill, A.G. Darvill, P. Albersheim, The degree of esterification and points of substitution by *O*-acetyl and *O*-(3-hydroxybutanoyl) groups in the acidic extracellular polysaccharides secreted by *Rhizobium leguminosarum* biovars viciae, *trifolii*, and *phaseoli* are not related to host range, J. Biol. Chem. 266 (1991) 9549–9555.
[39] J. Wielbo, et al., Complexity of phenotypes and symbiotic behaviour of *Rhizobium leguminosarum* biovar *trifolii* exopolysaccharide mutants, Arch. Microbiol. 182 (2004) 331–336.
[40] G. Guerrero, et al., Evolutionary, structural and functional relationships revealed by comparative analysis of syntenic genes in Rhizobiales, BMC Evol. Biol. 5 (2005) 55–74.
[41] M. Flores, et al., Genomic instability in *Rhizobium phaseoli*, J. Bacteriol. 170 (1988) 1191–1196.
[42] M. Flores, et al., Reiterated DNA sequences in *Rhizobium* and *Agrobacterium spp*., J. Bacteriol. 169 (1987) 5782–5788.
[43] M.L. Girard, et al., Structural complexity of the symbiotic plasmid of *Rhizobium leguminosarum* bv. *phaseoli*, J. Bacteriol. 173 (1991) 2411–2419.
[44] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[45] K. Rutherford, Artemis: Sequence visualisation and annotation, Bioinformatics 16 (2000) 944–945.
[46] J.C. Abbott, et al., WebACT—An online companion for the Artemis Comparison Tool, Bioinformatics 21 (2005) 3665–3666.
[47] D. Frishman, et al., The Pedant genome database, Nucleic Acids Res. 31 (2003) 207–211.
[48] J.F. Peden, Analysis of Codon Usage, PhD thesis, University of Nottingham, Nottingham, 1999.
[49] A.K. Chakravorty, W. Żurkowski, J. Shine, B.G. Rolfe, Symbiotic nitrogen fixation: Molecular cloning of *Rhizobium* genes involved in exopolysaccharide synthesis and effective nodulation, J. Mol. Appl. Genet. 1 (1982) 585–596.
[50] J. Wild, Z. Hradecna, W. Szybalski, Conditionally amplifiable BACs: Switching from single-copy to high-copy vectors and genomic clones, Genome Res. 12 (2002) 1434–1444.