

The genome sequence of hairy root *Rhizobium rhizogenes* strain LBA9402. Bioinformatics analysis suggests the presence of a new opine system in the agropine Ri plasmid

Marjolein Hooykaas¹ and Paul Hooykaas¹

¹Leiden University

February 15, 2021

Abstract

We report here the complete genome sequence of the *Rhizobium rhizogenes* (formerly *Agrobacterium rhizogenes*) strain LBA9402 (NCPBP1855rifR), a pathogenic strain causing hairy root disease. In order to assemble a complete genome we obtained short-reads from Illumina sequencing as well as long-reads from Oxford Nanopore Technology sequencing. The genome consists of a 3,958,212 bp chromosome, a 2,005,144 bp chromid (secondary chromosome) and a 252,168 bp Ri plasmid (pRi1855), respectively. The primary chromosome was very similar to that of the avirulent biocontrol strain K84, but the chromid showed a 724 kbp deletion accompanied by a large 1.8 Mbp inversion revealing the dynamic nature of these secondary chromosomes. The sequence of the agropine Ri plasmid was compared to other types of Ri and Ti plasmids. Thus we identified the genes responsible for agropine catabolism, but also a unique segment adjacent to the TL-region that has the signature of a new opine catabolic gene cluster including the three genes that together encode an opine dehydrogenase. Our sequence analysis also revealed a novel gene at the very right end of the TL-DNA, which is unique for the agropine Ri plasmid. The protein encoded by this gene was most related to the succinamopine synthases of chrysopine and agropine Ti plasmids and thus may be involved in synthesis of the unknown opine that can be degraded by the adjacent catabolic cluster. The available sequence will facilitate the use of *R. rhizogenes* and especially LBA9402 in both the laboratory and for biotechnological purposes.

The genome sequence of hairy root *Rhizobium rhizogenes* strain LBA9402. Bioinformatics analysis suggests the presence of a new opine system in the agropine Ri plasmid

Marjolein J.G. Hooykaas, Paul J.J. Hooykaas*

Institute of Biology, Leiden University, Sylviusweg 72, Leiden, The Netherlands

*Corresponding author

Abstract

We report here the complete genome sequence of the *Rhizobium rhizogenes* (formerly *Agrobacterium rhizogenes*) strain LBA9402 (NCPBP1855rifR), a pathogenic strain causing hairy root disease. In order to assemble a complete genome we obtained short-reads from Illumina sequencing as well as long-reads from Oxford Nanopore Technology sequencing. The genome consists of a 3,958,212 bp chromosome, a 2,005,144 bp chromid (secondary chromosome) and a 252,168 bp Ri plasmid (pRi1855), respectively. The primary chromosome was very similar to that of the avirulent biocontrol strain K84, but the chromid showed a 724 kbp deletion accompanied by a large 1.8 Mbp inversion revealing the dynamic nature of these secondary chromosomes. The sequence of the agropine Ri plasmid was compared to other types of Ri and Ti plasmids. Thus we identified the genes responsible for agropine catabolism, but also a unique segment adjacent to the TL-region that has the signature of a new opine catabolic gene cluster including the three genes that together

encode an opine dehydrogenase. Our sequence analysis also revealed a novel gene at the very right end of the TL-DNA, which is unique for the agropine Ri plasmid. The protein encoded by this gene was most related to the succinamopine synthases of chrysopine and agropine Ti plasmids and thus may be involved in synthesis of the unknown opine that can be degraded by the adjacent catabolic cluster. The available sequence will facilitate the use of *R. rhizogenes* and especially LBA9402 in both the laboratory and for biotechnological purposes.

Key words: *Agrobacterium rhizogenes*; chromid; opine; ornithine cyclodeaminase; Ri-plasmid; succinamopine synthase

Introduction

Hairy root, a neoplastic plant disease with a wide host range, is characterized by the formation of adventitious roots from infected wound sites (Riker 1930). It was originally encountered as a problem in tree nurseries, but nowadays the disease is also increasingly causing problems in the greenhouse by inducing extensive root mats, thereby reducing the harvest of cucumbers and tomatoes (Weller *et al.* 2000). The causal agent, a bacterium that was for long called *Agrobacterium rhizogenes*, contains a large, about 200 kbp root-inducing (Ri) plasmid, which contains the essential virulence determinants (White and Nester 1980). The molecular mechanism by which hairy root is induced is similar to that which is used by the related bacterium *Agrobacterium tumefaciens* to induce crown gall tumors in plants. During infection, part of the Ri plasmid (the T-region) is transferred to plant cells and integrated into the plant genome (Bevan and Chilton 1982). Expression of *rol*-genes located in the T-DNA leads to the transformation of normal cells into tumor cells that develop into roots that can grow in *in vitro* culture in the absence of added plant growth regulators (Jouanin *et al.* 1987). In the cells of hairy roots unusual compounds called opines are produced, which are specific condensates of amino acids and keto acids or sugars (Petit *et al.* 1983). These opines, which are formed by enzymes encoded by the T-DNA, support the growth of the pathogen, which contains the catabolic genes usually in a region adjacent to the T-region on the Ri plasmid (Dessaux *et al.* 1993). On the basis of the specific opines formed and degraded, agropine, cucumopine, mannopine, and mikimopine Ri plasmids are nowadays distinguished.

Like *A. tumefaciens*, *A. rhizogenes* has been disarmed by deleting the T-DNA genes in order to convert it into a vector useful for plant genome engineering (Collier *et al.* 2018). Also, the bacterium as such is used for biotechnological research and application; in research for instance for gene function and gene expression analysis in roots (Ron *et al.* 2014), in industry to obtain roots that can be grown in bioreactors for the production of secondary metabolites (Mehrotra *et al.* 2015).

It has become apparent over the years that in nature various bacteria of the *Rhizobiaceae* family may cause hairy root or crown gall. On the basis of their physiological properties three different groups (biovars 1-3) were distinguished already long ago (Kerr and Panagopoulos 1977). The species name *Rhizobium rhizogenes* is now commonly used for the bacteria belonging to biovar 2. *R. rhizogenes* strains have two megabase DNA circles, a chromosome and a plasmid-derived megacircle, sometimes called a chromid (Jumas-Bilak *et al.* 1998; Harrison *et al.* 2010). Draft genome sequences of several strains are available on NCBI, and one draft genome sequence has been published for *R. rhizogenes* strain ATCC15834 consisting of 43 scaffolds (Kajala *et al.* 2014). However, up to now only one complete genome sequence is available for *R. rhizogenes*, that of the avirulent agrocin-producing biocontrol agent Kerr 84 (Slater *et al.* 2009). Here we present a second complete genomic sequence of *R. rhizogenes*, that of the hairy root inducing strain LBA9402. This strain is a rifampicin-resistant derivative of wild-type strain NCPPB1855, which is one of the most widely used laboratory strains (Desmet *et al.* 2020). By comparing the chromosome and chromid of LBA9402 with those of strain K84 we found that the chromosome was very similar, but that the chromid showed large differences due to a large 724 kbb deletion accompanied by a large inversion, underscoring the dynamic nature of the chromid. Analysis of the sequence of the agropine Ri plasmid of LBA9402 revealed that this had a few unique areas including one that we predict encodes a new opine catabolic cluster, including the three genes characteristic for defining an opine dehydrogenase. A candidate gene for a novel opine synthase was identified at the very right end of the TL-DNA.

Materials and methods

2.1 Organism

R. rhizogenes strain LBA9402 is a rifampicin-resistant derivative of wild-type strain NCPPB1855, which was isolated in our laboratory (Hooykaas 1979), and which causes prolific hairy roots on infected plants. The bacterium was grown on TY medium (Difco tryptone 5 g/l, Difco yeast extract 3 g/l, CaCl₂·6H₂O 1.3 g/l). The bacterium was tested for virulence by puncturing the plant stems with a sterile wooden toothpick that had been dipped into a colony of the bacterium.

2.2 Sequencing methods

R. rhizogenes strain LBA9402 was cultured in TY medium (Beringer, 1974), followed by genomic DNA isolation using QIAGEN gravity-flow columns. The genome of LBA9402 was sequenced using a combination of Illumina and Oxford Nanopore Technologies platforms. Nanopore sequencing was done in house, but Illumina sequencing was performed at the Leiden Genome Technology Center (LGTC) of the Leiden University Medical Center (Leiden, The Netherlands), where TruSeq DNA Libraries were sequenced on an Illumina HiSeq 2000 machine. The Oxford Nanopore sequencing library was generated with 200 ng DNA using the SQK-RBK004 Rapid Barcoding Kit. The library was pooled with another library, followed by in-house sequencing on a MinION flow cell (version R9.4.1).

2.3 Data processing methods

After basecalling with Albacore (version 2.3.4) the MinION reads were demultiplexed (with Epi2me). The total yield for LBA9402 was 298,712 reads, totaling 1,027,720,149 bp, with a mean read length of 3441 bp. Nanopore reads were end-trimmed and filtered on average quality (>Q10) and length

(>5000 bp) with NanoFilt (64-fold coverage after filtering). A total of 4,518,191 99-nucleotide paired-end Illumina reads were quality and adapter trimmed using Cutadapt (70-fold coverage). Hybrid assembly was performed using Unicycler version 0.4.7. Besides three contigs representing the two chromosomes and the Ri plasmid a fourth contig of 5386 bp was identified. This represented the bacteriophage ΦX174 genome sequence, which is spiked-in at low concentration during Illumina library preparation. This contig was therefore removed from the assembly. The assembly was annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAP). In addition, PHASTER was used to annotate prophage sequences (Arndt et al. 2016). For the functional characterization of the encoded proteins eggNOG-Mapper was employed (Huerta-Cepas et al. 2017). Insertion elements (IS elements) were identified using ISEScan (Xie and Tang 2017). In figure 2 and supplementary figure S5, only complete insertion sequences, i.e. including inverted repeats, are shown. IslandViewer was used to predict genomic islands (Dhillon et al. 2015) and CGView was used to generate a circular map of pRi1855 (Stothard and Wishart 2005). Mauve (progressiveMauve) was used to align the LBA9402 and K84 genomes (Darling et al. 2010). BRIG was used to compare pRi1855 with other Ri and Ti plasmids (with BLASTN, e-value cut-off 1e-10) and to visualize the hits in concentric rings (Alikhan et al. 2011). For the comparisons between erythritol catabolism regions and between pRi-1855 and *Rhizobium lusitanum* strain 629, BLASTn was run locally with BLAST version 2.9.0+. Protein alignments were performed with MAFFT version 7.471, L-INS-I method (Katoh 2013) and visualized with Jalview version 2.11.1.2 (Waterhouse et al. 2009) and Adobe Illustrator. Percentage identities as shown in Table 1 were calculated with the R package seqinr.

2.4 Data availability

The complete genome sequence of *R. rhizogenes* LBA9402 was deposited in GenBank under accession numbers CP044122, CP044123 and CP044124. The raw reads are deposited in the Sequence Read Archive under accessions numbers SRR10177303 and SRR10177304.

Results

3.1 The genomic sequence of *Rhizobium rhizogenes* strain LBA9402

As Illumina sequencing data alone were not sufficient to obtain a high quality and complete genome sequence of strain LBA9402, we additionally obtained long reads by Nanopore sequencing. Unicycler was used to obtain a hybrid assembly. This resulted in three circular contigs of 3,958,212 bp, 2,005,144 bp and 252,168 bp, respectively. More than 99.9% of both MinION and Illumina reads align to the assembly, indicating that the assembly is complete. The G+C content of the genome was 60%. The assembly was annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAP). In total, 5822 coding sequences, 9 rRNA genes (3 operons), and 53 tRNA genes were annotated. COG categories were assigned to predicted coding sequences with eggNOG-mapper (supplementary figure S1, Supplementary Material online).

3.2 Comparison to the sequence of *R. rhizogenes* strain K84

The only other complete genome of *R. rhizogenes* obtained thus far is that of the avirulent biocontrol strain K84 (Slater *et al.* 2009). This latter strain does not contain an Ri plasmid, but the two LBA9402 megacircles could be aligned to those of K84 using MUMmer (NUCmer) (Kurtz *et al.* 2004). As can be seen in supplementary figure S2 the largest contig of LBA9402 is largely collinear with the primary chromosome of K84 (96% of the sequence can be aligned to that of K84 at >95% sequence identity). It contains the genes for replication such as for a DnaA replication protein, repair and DNA recombination, for cell division, for transcription and translation including the 3 rRNA clusters of the bacterium and the 53 tRNA genes. We annotated an extra tRNA in one of the unique regions of the LBA9402 primary chromosome. The chromosome contains a putative genomic island of about 100 kbp with a set of genes for conjugative DNA transfer encoding not only a Type4 secretion system for mating pair formation (from position 3,324,071-3,314,501), but also the enzymes necessary for DNA transfer and replication (from position 3,302,269-3,307,564). The genomic island contains a gene for a putative integrase and is surrounded by a direct repeat of 15 bp, which may be (the remains of) two *att* -sites. Larger differences between the chromosomes of LBA9402 and K84 are mainly due to presence/absence of other mobile elements. For example, various proteins encoded in the unique segment of DNA from position 745,195-784,185 in LBA9402, have homology to phage proteins (Rhizobium phage vB_RleM_PPF1 and other tailed phages) as revealed by the phage search tool PHASTER (Arndt *et al.* 2016) and thus seems due to the insertion of a prophage (supplementary figure S3).

The second-largest replicons are less similar (supplementary figure S2, fig. 1), but still 85% of the LBA9402 sequence aligns to that of K84 (and 64% of K84 aligns to LBA9402). The LBA9402 sequence is smaller, mainly due to a large approximately 724 kb deletion, which seems accompanied by a large 1.8 Mbp inversion (fig. 1). The large deletion did not affect one functional category of genes in particular as can be seen in supplementary figure 1, demonstrating that there are no major differences in the categories of genes present according to eggNOG-mapper analysis.

This secondary megacircle has a plasmid-like RepABC replication system, but has a similar GC content as the primary chromosome. Such secondary megacircles are considered (developing) secondary chromosomes that over evolutionary time exchange genes with the primary chromosome and have been coined ‘chromids’ (Slater *et al.* 2009; Harrison *et al.* 2010). The chromid of LBA9402 contains many metabolic genes, but also genes for the production of cell wall polysaccharides and fimbriae/pili. We found in the chromid a set of genes homologous to the erythritol region in *Sinorhizobium meliloti* and *Rhizobium leguminosarum* (supplementary figure S4) including a transport operon with genes *eryEFG*, a catabolic operon with genes *eryABCD* and *adeoR* -type regulator (also called *eryR*) followed by genes called *eryH* and *eryI* (Yost *et al.* 2006; Geddes and Oresnik 2012; Barbier *et al.* 2014). The ability to catabolize erythritol is one of the key characteristics distinguishing biotype 1 and biotype 2 agrobacteria (Kerr and Panagopoulos 1977). The presence of erythritol catabolic genes was thus expected, but it was remarkable that they were present on the more dynamic chromid instead of the chromosome. Finally the 252 kbp circle represents the agropine pRi1855 plasmid, which is very different from the large nopaline catabolic plasmid carried by strain K84, and will be described below.

3.3 General properties of the Ri plasmids

The pRi1855 plasmid comprises 252,168 bp. It has an approximately 4% lower GC content than the rest of the

genome. In total, 236 protein coding sequences were identified with an average size of 898 bp (supplementary figure S5). We compared the agropine Ri plasmid pRi1855 sequence to publicly available Ri and Ti plasmid sequences: octopine Ti plasmid pTiAch5 (CP007228; Henkel et al. 2014; Huang et al. 2015), nopaline Ti plasmid pTiC58 (AE007871; Goodner et al. 2001; Wood et al. 2001), succinamopine Ti plasmid pTiEU6 (KX388535; Shao et al. 2019), agropine Ti plasmid pTiBo542 (DQ058764; Oger et al. 2001), chrysopine Ti plasmid pTiChry5 (KX388536; Shao et al. 2018), mannopine Ri plasmid pRi8196 (Weisberg et al. 2020; our own unpublished results), cucumopine Ri plasmid pRi2659 (NZ_CP019703.3; Valdes Franco et al. 2016; Tong et al. 2018), and mikimopine Ri plasmid pRi pRi1724 (NC_002575; Moriguchi et al. 2001). The conservation of the different areas in these plasmids is visualized in figure 2 and we shall discuss these in the following parts.

In previous studies dealing with the closely related agropine Ri plasmid pRiA4, the replication (*repABC*) and conjugative transfer (*tra*, *trb*) genes in the agropine Ri plasmid were already described (Nishiguchi et al. 1987; Wetzel et al. 2015) and are very similar to those in other Ti and Ri plasmids.

The agropine Ri plasmid has two T-regions, one of which, the TL-region contains the *rol*-genes that are necessary and sufficient for the formation of hairy roots. Other Ri plasmids have only one T-region with very similar genes (fig. 3; Otten 2018). In the agropine Ri plasmid, however, a copy of IS630 has inserted between *orf3* and *orf8*. At the very left end in the agropine TL-region and the mannopine T-region a gene for agrocinopine synthase is present, but only remnants of these genes are still present in the cucumopine and mikimopine Ri plasmids. At the very right end of the T-region one (in the cucumopine and mikimopine Ri plasmids) or two (in the mannopine Ri plasmids) non-conserved genes are present. These encode the cucumopine (Valdes Franco et al. 2016) and mikimopine synthase (Moriguchi et al. 2001), respectively, and the two genes necessary for mannopine synthesis in the mannopine Ri plasmid (fig. 3). The TL-region of the agropine Ri plasmid pRiA4 was previously sequenced (Slightom et al. 1986) and this revealed at the right end the presence of *orf15* /*rolD* and three smaller *orfs*. In our pRi1855 sequence we find besides *orf15/rolD* only one larger *orf*, hereinafter called *orf16*. The function of *orf16* is unknown, but may encode an unknown opine synthase as will be discussed in a next paragraph. The agropine Ri plasmid has besides the conserved T-region (TL-region) an additional T-region (TR-region) containing *aux*-genes involved in biosynthesis of the auxin indole acetic acid (Offringa et al. 1986) and the genes *mas1*, *mas2* and *ags* for agropine biosynthesis (Bouchez and Tourneur 1991).

The virulence region of pRi1855 responsible for transfer of the T-DNA into plant cells contains the essential virulence genes *virA-virD5*, but although *virE3* is present the *virE1* and *virE2* genes are missing and replaced by a new *orf* with some similarity to nopaline pTi *virF*. The sequence of the cucumopine and mikimopine Ri plasmids in this area is almost identical suggesting that the deletion of *virE2* occurred once before the divergence of these Ri plasmids. The *virE2* gene is functionally replaced by a gene called *GALLS* as reported before for pRiA4 (Hodges et al. 2004). The *GALLS* gene was previously also identified in the cucumopine and mikimopine Ri plasmids by Southern analysis, but is not present in the mannopine Ri plasmid, which still carries the *virE2* gene (Hodges et al. 2004). Together with a *tzs* gene, which encodes an enzyme that catalyzes the synthesis of the cytokinin zeatin riboside 5'-phosphate (Krall et al. 2002), *GALLS* is located outside the *vir*-region, about 65 kbp clockwise from the *virE3* gene, near opine catabolic genes in the cucumopine and mikimopine Ri plasmids. We could now identify and locate the *GALLS* gene in the pRi1855 sequence at a completely different location next to the *traG* gene almost 90 kbp counterclockwise from the *virE3* gene. Besides the core set of *vir* genes mentioned above, the *vir* region of the agropine Ri plasmid contains next to the *virA* gene the *tzs* gene, *virK*, a second nopaline Ti-like *virF* gene, and finally *virH1*, *virH2*. It resembles in this respect the *vir*-region of the nopaline Ti plasmid and like in the nopaline Ti plasmids *virJ* is absent. Other types of Ri plasmids have similar *vir*-regions, but a *virH2* gene is absent from the cucumopine and mikimopine Ri plasmids, and as mentioned above a *tzs* gene is present, but located in an entirely different area of the plasmid.

3.4 Agropine and agrocinopine catabolism genes

Hairy roots formed by agropine strains contain agropine, agropinic acid, mannopinic acid and mannopine

(Petit et al. 1983). The agropine Ri plasmid enables host strains to degrade agropine. *R. rhizogenes* strains such as A4, but not 1855 or HRI contain a second, catabolic plasmid with genes for catabolism of the other three mannityl opines (Petit et al. 1983). We have now identified the genes for agropine transport and catabolism in pRi1855, which are located in a segment of the plasmid adjacent to the TR-region (fig. 2). This region contains genes with high similarity to the genes described by Kim and Farrand (1996) on the octopine Ti plasmid involved in agropine uptake and degradation. These genes encode an agropine permease and also comprise *agcA* for the delactonase converting agropine into mannopine, *mocC* for oxidizing mannopine into deoxyfructosyl glutamine and *mocDE* determining the deconjugase liberating an amino acid and a phosphorylated sugar. In the octopine Ti plasmid the genes *mocA* and *mocB* encode enzymes with weak homology to glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydratase (Kim and Farrand 1996). These are probably involved in further catabolism of the release phosphorylated sugar. However, while an intact homolog of *mocA* was present in pRi1855, as well as a homolog of *mocC*, in between only a truncated remnant of a gene homologous to *mocB* was present due to a deletion of more than 1 kbp. Regulators closely related to *mocS* and *mocR* are present in an identical position in front of *mocA* and between *mocC* and *mocD*. Our detection of *amocD* gene in pRi1855 was remarkable as such gene was thought to be absent from the agropine Ri plasmid (Baek et al. 2005). The agropine Ri plasmid has adjacent to the left end of the TL-region in pRi1855 a set of *acc* genes for agrocinopine catabolism (fig. 2), which matches the presence of an *acs* gene for the biosynthesis of agrocinopine in the TL-region. These genes are also present in the mannopine Ri plasmid, but absent from the cucumopine and mikimopine Ri plasmids.

3.5 Genes for a novel opine system in pRi1855

The pRi1855 plasmid has several regions with genes of unknown function. It shares a large region of about 65 kbp (area from 88-153 kbp on the map of figure 2) with the other Ri plasmids. This region contains genes putatively involved in sugar transport, glycerol metabolism and encodes several transcription regulators and two chemoreceptors (Moriguchi et al. 2001). Besides, the pRi1855 plasmid has several unique areas with genes that are found in none of the other types of Ri and Ti plasmids described. These include transposable elements (fig.2) and two larger areas of about 20 kbp (area from 20-40 kbp on the map adjacent to the right border of the TL-region in fig.2) and about 24 kbp (area from 64-88 kbp on the map adjacent to the agropine catabolic genes). The latter area contains mainly metabolic genes and may have been introduced into pRi1855 by transposition as it is surrounded by IS5-like insertion sequences. It may have originated from the chromosome of another *Rhizobium* species, as a very similar stretch of DNA was detected in the recently sequenced chromosome of *Rhizobium lusitanum* strain 629 (supplementary figure S6). The 20 kbp segment adjacent to the right border of the TL-region may be involved in the transport and catabolism of a new opine. In this area we identified all the three characteristic genes that together code for a putative flavin-containing opine dehydrogenase (supplementary figures S7-S9). Flavin-containing opine dehydrogenases such as octopine, nopaline and succinamopine dehydrogenase consist of three subunits OdhABC that are encoded by *ynoxABC/ooxABC* -like genes arranged in tandem in the genome (Watanabe et al. 2015). The three genes in pRi1855 (F3X89_28345, F3X89_28350, F3X89_28355) encode closely related proteins in which the characteristic binding sites for the FAD and FMN co-factors and the Fe-S cluster have fully been conserved (supplementary figures S7-S9). These three genes are surrounded on both sides by genes for a transport system and a LysR-type regulator. In Ti and Ri plasmids genes encoding an opine dehydrogenase are often accompanied by genes encoding the permease required for uptake of specific opine into the bacterial cell. Besides in the vicinity often genes are present encoding enzymes involved in the catabolism of the specific products released by the opine dehydrogenase. In this area of pRi1855 genes encoding such metabolic proteins are also present including genes encoding a putative saccharopine dehydrogenase and a putative amino adipate semialdehyde dehydrogenase, which may form part of a catabolic pathway of the amino acid lysine (de Mello Serrano et al. 2012). A gene for an AsnC/Lrp regulator is located at the end of this DNA segment. The Lrp family of transcriptional regulators is known to control amino acid metabolism in bacteria (Brinkman et al. 2003).

If these genes are involved in the catabolism of an opine, a gene for an unknown opine synthase should be

present in the T-region of pRi1855. Genes for agrocinopine synthase are located at the extreme left end of the T-region in Ti and Ri plasmids, while genes for nopaline synthase, octopine synthase, and succinamopine synthase are located at the extreme right end of the T-region in Ti plasmids. In cucumopine, mikimopine, and mannopine Ri plasmids the genes for cucumopine, mikimopine, and mannopine synthesis are likewise located immediately next to the right border repeat. We find at the very right end of the TL-region of pRi1855 two related genes (fig. 3): *orf15/rolD* and *orf16*, which share 55% identity (fig. 4, table 1). These genes are not present in the T-regions of any of the other types of Ri plasmids (fig. 3). The *orf15* has been called *rolD*; the encoded RolD protein has weak sequence homology with ornithine cyclodeaminases and indeed can convert ornithine into proline (Trovato et al. 2001). The role of *rolD* in hairy root formation is marginal, but the gene can influence plant development by its metabolic activity (Trovato et al. 2018). Using BLASTP with the proteins encoded by *orf15* and *orf16* as a query we picked up the succinamopine synthases encoded by the T-region of chrysopine pTiChry5 (Shao et al. 2018) and the agropine pTiBo542 (Oger et al. 2001) as the most related proteins. The proteins encoded by *orf15* and *orf16* share 44-47% identity with the two succinamopine synthases, which themselves share 93% identity (table 1). All these proteins (encoded by *orf15/rolD*, *orf16*, *susL*) are evolutionary related to ornithine cyclodeaminases (encoded by *ocd* genes). They share, for instance, about 19-21% identity with the ornithine cyclodeaminase encoded by the nopaline Ti plasmid. That ornithine cyclodeaminase (*ocd*) genes can evolve novel biochemical functions during evolution is known for some time. For instance its function has been reported to evolve into an alanine dehydrogenase in *Archaeoglobus fulgidus* and into a tauropine dehydrogenase in *Halichondria japonica* (Sharma et al. 2013; Watanabe et al. 2014). Apparently, it can evolve also in an opine (succinamopine) synthase. Therefore, it would seem possible that either or both of the *ocd*-like genes at the right end of pRi1855 (*orf15*, *orf16*) similarly have evolved a novel opine synthase function, producing an unknown opine that can be degraded by the putative opine dehydrogenase encoded in the area with genes of unknown function located adjacent to the right border of the TL-region.

Discussion

The second complete genomic sequence of a *Rhizobium rhizogenes* strain and the first of a virulent strain enabled us to make a comparison with the sequence of the previously sequenced biocontrol strain K84. This revealed a high conservation of the primary chromosome, but showed large differences in the secondary megacircle, the chromid. It has been described that chromids have a plasmid-like RepABC replication system, but have a similar GC content as the primary chromosome and this is also the case in LBA9402. It has been proposed that chromids are plasmids that evolve into secondary chromosomes and over time exchange genes with the primary chromosome (Slater et al. 2009; Harrison et al. 2010). We found that the chromid of strain LBA9402 was much smaller than that of strain K84 due to the absence of a segment of 724 kbp that may have been deleted in LBA9402 or inserted in K84. Also we found that this insertion/deletion was accompanied by a large inversion of a segment of 1.8 Mbp. The presence of such complex rearrangements is in line with their plasmid descent and the genes which they carry being mostly non-essential.

Our genomic sequence includes the complete sequence of the agropine pRi1855 plasmid. Over the years sequences have already been published dealing with specific parts of the closely related agropine Ri plasmid pRiA4 and earlier this year a draft of the completed pRiA4 sequence was published (Thompson et al. 2020). When compared our pRi1855 sequence is indeed closely related, but still differs at numerous areas, both by base substitutions and small insertions/deletions.

Hairy roots formed by agropine strains contain agropine, agropinic acid, mannopinic acid and mannopine (Petit et al. 1983). The agropine Ri plasmid, however, enables host strains to degrade agropine, but not the other mannityl opines. *R. rhizogenes* strains such as A4, but not 1855 or HRI, contain a second, catabolic plasmid with genes for catabolism of the other three mannityl opines (Petit et al. 1983). We have now identified the genes for agropine catabolism in pRi1855 adjacent to the right border of the TR-region. The region embraces *anagcA* gene for the delactonase converting agropine into mannopine, *mocC* for oxidizing mannopine, and *mocD* and *mocE* together determining the enzymes that can release the amino acid and a phosphorylated sugar from the conjugate. This pathway would allow the bacterium to degrade both agropine

and mannopine. However, it is known that the bacteria carrying pRi1855 cannot degrade mannopine. This may be because mannopine cannot induce the catabolic genes or because the bacterium cannot import mannopine in the cell. Indeed the pRi1855 plasmid contains genes for an agropine permease, but not for a mannopine transport system.

Agrobacteria induce neoplasias in which opines are formed that serve as a nutritional source of the bacteria. All the different types of Ti and Ri plasmids described so far have a gene coding for an opine synthase at the very right end of the T-region. Our sequence now shows that also the agropine Ri plasmid has one larger *orf* (*orf16*) at the very right end of the TL-region, which shares 55% identity with the neighboring *rolD* gene. This gene encodes a protein that is evolutionary related to ornithine cyclodeaminase (*ocd*) and which still has ornithine cyclodeaminase activity (Trovato et al. 2001). Here we discovered that both *rolD* and *orf16* have also significant identity of 44-47% with the *susL* genes encoding succinamopine synthase in the agropine and chrysopine Ti plasmids. Ornithine cyclodeaminase encoding (*ocd*) genes have been reported to have evolved into genes encoding new enzymatic activities such as alanine dehydrogenase activity in *Archaeoglobus fulgidus* and tauroopine dehydrogenase activity in *Halichondria japonica* (Sharma et al. 2013; Watanabe et al. 2014). We hypothesize that an *ocd*-like gene may have evolved in the agropine Ri plasmid into a gene for a new opine synthase. Opine catabolic genes are often located close to the synthase gene, but on the other side of the right border, an arrangement seen in many different Ti and Ri plasmids. Three of the genes located here together indeed have the signature of the trios of genes that are known to encode the octopine and nopaline dehydrogenase (Watanabe et al. 2015), and thus may encode the opine dehydrogenase needed for catabolism of the novel unknown opine.

In view of its frequent application, the available sequence will facilitate the use of *R. rhizogenes* and especially LBA9402 in both the laboratory and for biotechnological purposes.

Supplementary Material

Additional Supporting Information may be found online in the supporting information tab for this article.

Acknowledgements

We are grateful to Amke den Dulk-Ras and Shuai Shao for help with DNA isolation, Dr Marcel van Verk and Dr Chris Henkel for contributions to the initial analyses of the Illumina data. P.J.J.H. was supported by the fellowship associated with the appointment as Academy Professor by the Royal Netherlands Academy of Sciences.

Data availability statement

The complete genome sequence of *R. rhizogenes* LBA9402 was deposited in GenBank under accession numbers CP044122, CP044123 and CP044124. The raw reads are deposited in the Sequence Read Archive under accession numbers SRR10177303 and SRR10177304.

Literature cited

Alikhan NF, Petty NK, Zakour NLB, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402

Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44(W1): W16-W21.

Baek C, et al. 2005. Genes for utilization of deoxyfructosyl glutamine (DFG), an amadori compound, are widely dispersed in the family Rhizobiaceae. *FEMS Microbiol Ecol* 53: 221-233.

Barbier T, et al. 2014. Erythritol feeds the pentose phosphate pathway via three new isomerases leading to D-erythrose-4-phosphate in *Brucella*. *Proc Natl Acad Sci USA* 111: 17815-17820.

Beringer JE. 1974. R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* 84:188-198.

- Bevan MW, Chilton MD. 1982. T-DNA of the *Agrobacterium* Ti and Ri Plasmids. *Ann Rev Genet* 16: 357-384.
- Brinkman AB, Ettema TJ, de Vos WM, van der Oost J. 2003. The Lrp family of transcriptional regulators. *Mol Microbiol* 48: 287-94.
- Collier R, Thomson J, Thilmony R. 2018. A versatile and robust *Agrobacterium*-based gene stacking system generates high quality transgenic Arabidopsis plants. *Plant J* 95: 573-583.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLOS ONE* 5: e11147.
- de Mello Serrano GC, e Silva Figueira TR, Kiyota E, Zanata N, Arruda P. 2012. Lysine degradation through the saccharopine pathway in bacteria: LKR and SDH in bacteria and its relationship to the plant and animal enzymes. *FEBS Lett* 586: 905-911.
- Desmet S, et al. 2020. Rhizogenic agrobacteria as an innovative tool for plant breeding: current achievements and limitations. *Appl Microbiol Biotechnol* 104: 2435-2451.
- Dessaux Y, Petit A, Tempe J. 1993. Chemistry and biochemistry of opines, chemical mediators of parasitism. *Phytochem* 34 : 31-38.
- Dhillon BK, et al. 2015. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucl Acids Res* 43: W104-W108.
- Geddes BA, Oresnik IJ. 2012. Genetic characterization of a complex locus necessary for the transport and catabolism of erythritol, adonitol and l-arabitol in *S.meliloti*. *Microbiology* 158: 2180-2191.
- Goodner B, et al. 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* 294: 2323-2328.
- Harrison PW, Lower RPJ, Kim NKD, Young JPW. 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol* 18: 141-148.
- Henkel CV, den Dulk-Ras A, Zhang X, Hooykaas PJ. 2014. Genome sequence of the octopine-type *Agrobacterium tumefaciens* strain Ach5. *Genome Announc.* 2: e00225-14.
- Hodges LD, Cuperus J, Ream W. 2004. *Agrobacterium rhizogenes* GALLS protein substitutes for *Agrobacterium tumefaciens* single-stranded DNA-binding protein VirE2. *J Bacteriol* 186: 3065-3077.
- Hooykaas PJJ. 1979. The role of plasmid determined functions in the interactions of Rhizobiaceae with plant cells. A genetic approach. PhD Thesis. Leiden University, The Netherlands.
- Huang YY, et al. 2015. Complete genome sequence of *Agrobacterium tumefaciens* Ach5. *Genome Announc.* 3: e00570-15.
- Huerta-Cepas J, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 34: 2115-2122.
- Jouanin L, et al. 1987. Transfer of a 4.3 kb fragment of the TL-DNA of *Agrobacterium rhizogenes* strain A4 confers the pRi transformed phenotype to regenerated tobacco. *Plant Science* 53: 53-63.
- Jumas-Bilak E, Michaux-Charachon S, Bourg G, Ramuz M, Allardet-Servent A. 1998. Unconventional genomic organization in the alpha subgroup of the Proteobacteria. *J Bacteriol* 180: 2749-2755.
- Kajala K, Coil DA, Brady SM. 2014. Draft genome sequence of *Rhizobium rhizogenes* strain ATCC15834. *Genome Announcements* 2: e01108-14.
- Katoh S. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772-780.

- Kerr A, Panagopoulos CG. 1977. Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their biological control. *Phytopath. Z.* 90: 172-179.
- Kim KS, Farrand SK. 1996. Ti plasmid-encoded genes responsible for catabolism of the crown gall opine mannopine by *Agrobacterium tumefaciens* are homologs of the T-region genes responsible for synthesis of this opine by the plant tumor. *J Bacteriol* 178:3275-3284.
- Krall L, Raschke M, Zenk MH, Baron C. 2002. The Tzs protein from *Agrobacterium tumefaciens* C58 produces zeatin riboside 5'-phosphate from 4-hydroxy-3-methyl-2-(E)-butenyl diphosphate and AMP. *FEBS Lett* 527: 315-318.
- Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biology* 5:R12.
- Mehrotra S, Srivastav V, Rahman LU, Kukreja, AK. 2015. Hairy root biotechnology-indicative timeline to understand missing links and future outlook. *Protoplasma* 252: 1189-1201.
- Moriguchi, et al. 2001. The complete nucleotide sequence of a plant root-inducing (Ri) plasmid indicates its chimeric structure and evolutionary relationship between tumor-inducing (Ti) and symbiotic (Sym) plasmids in Rhizobiaceae. *J Mol Biol*: 307: 771-784.
- Nishiguchi R, Takanami M, Oka A. 1987. Characterization and sequence determination of the replicator region in the hairy-root-inducing plasmid pRiA4b. *Mol Gen Genet* 206: 1– 8.
- Offringa IA, et al. 1986. Complementation of *Agrobacterium tumefaciens* tumor-inducing aux mutants by genes from the TR-region of the Ri plasmid of *Agrobacterium rhizogenes* *Proc Natl Acad Sci USA* 83: 6935-6939.
- Oger PM, Reich C, Olsen GJ, Farrand SK. 2001. Complete nucleotide sequence and analysis of pTiBo542: what genomics tells us about structure and evolution of plasmids in the family *Rhizobiaceae* . *Plasmid* 45: 169-170.
- Otten L. 2018. The *Agrobacterium* phenotypic plasticity (*Plast*) genes. *Current Top Microbiol Immunol* 418: 375-420.
- Petit A, et al. 1983. Further extension of the opine concept: Plasmids in *Agrobacterium rhizogenes* cooperate for opine degradation. *Mol gen Genet* 190: 204-214.
- Riker AJ. 1930. Studies on infectious hairy root of nursery apple trees. *J Agr Res* 41: 507-540.
- Ron M, et al. 2014. Hairy root transformation using *Agrobacterium rhizogenes* as a tool for exploring cell type-specific gene expression and function using tomato as a model. *Plant Physiol* 166: 455-469.
- Shao S, Zhang X, van Heusden GPH, Hooykaas PJJ. 2018. Complete sequence of the tumor-inducing plasmid pTiChry5 from the hypervirulent *Agrobacterium tumefaciens* strain Chry5. *Plasmid* 96-97:1-6.
- Shao S, van Heusden GPH, Hooykaas PJJ. 2019. Complete sequence of succinamopine Ti-plasmid pTiEU6 reveals its evolutionary relatedness with nopaline-type Ti-plasmids. *Genome Biology and Evolution* 11: 2480–2491.
- Sharma S, Shinde S, Verslues PE. 2013. Functional characterization of an ornithine cyclodeaminase-like protein of *Arabidopsis thaliana* . *BMC Plant Biol* 13: 182.
- Slater SC, et al. 2009. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J Bacteriol* 191: 2501-2511.
- Slightom JL, Durand-Tardif M, Jouanin L, Tepfer D. 1986. Nucleotide sequence analysis of TL-DNA of *Agrobacterium rhizogenes* agropine type plasmid. Identification of open reading frames. *J Biol Chem* 261(1):108-21.

Stothard P, Wishart DS, 2005. Circular genome visualization and exploration using CGView. *Bioinformatics* 21: 537-539.

Thompson MG, et al. 2020. Draft genome sequence of *Agrobacterium fabrum* ARqua1. *Microbiology Resource Announcements* 9: e00506-20.

Tong X, et al. 2018. The complete genome sequence of cucumopine-type *Agrobacterium rhizogenes* strain K599 (NCPFB2659), a nature's genetic engineer inducing hairy roots. *Int J Agric Biol* 20: 1167-1174.

Trovato M, Maras B, Linhares F, Costantino P. 2001. The plant oncogene *rolD* encodes a functional ornithine cyclodeaminase. *Proc Natl Acad Sci USA* 98: 13449-13453.

Trovato M, Mattioli R, Costantino P. 2018. From *A. rhizogenes* *rolD* to plant P5CS: exploiting proline to control plant development. *Plants* 7:108.

Valdes Franco et al. 2016. Draft genome sequence of *Agrobacterium rhizogenes* strain NCPFB2659. *Genome Announcements* 4: e00746-16.

Watanabe S, Tozawa Y, Watanabe Y. 2014. Ornithine cyclodeaminase/ μ -crystallin homolog from the hyperthermophilic archaeon *Thermococcus litoralis* functions as a novel Δ 1-pyrroline-2-carboxylate reductase involved in putative *trans*-3-hydroxy-l-proline metabolism. *FEBS Open Bio* 4: 2211-5463

Watanabe S, Sueda R, Fukumori F, Watanabe Y. 2015. Characterization of flavin-containing opine dehydrogenase from bacteria. *PLOS One* 10: e0138434.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189-1191.

Weisberg et al. 2020. Unexpected conservation and global transmission of agrobacterial virulence plasmids. *Science* 368: eaba5256.

Weller SA, Stead DE, O'Neill TM, Hargreaves D, McPherson GM. 2000. Rhizogenic *Agrobacterium* biovar 1 and cucumber root mat in the UK. *Plant Pathol* 49: 43-50.

Wetzel ME, Olsen GJ, Chakravartty V, Farrand SK. 2015. The *repABC* plasmids with quorum-regulated transfer systems in members of the Rhizobiales divide into two structurally and separately evolving groups. *Genome Biology and Evolution*. 7: 3337-3357.

White FF, Nester EW. 1980. Hairy root: plasmid encodes virulence traits in *Agrobacterium rhizogenes*. *J Bacteriol* 141: 1134-1141.

Wood DW, et al.. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science*. 294: 2317-2323.

Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33: 3340-3347.

Yost CK, Rath AM, Noel TC, Hynes MF. 2006. Characterization of genes involved in erythritol catabolism in *Rhizobium leguminosarum* bv. *Viciae*. *Microbiology* 152: 2061-2074.

Legends

Fig. 1

Large rearrangements in the chromid of *R. rhizogenes*. The chromid/secondary chromosome of LBA9402 was aligned to that of K84 using progressiveMauve. The aligned blocks were visualized with the R package genoplots. Red areas are regions with the same orientation, whereas blue areas align in inverse orientation.

Fig. 2

Circular representation of pRi1855 and comparative analysis to other Ri and Ti plasmids.. The inner rings show BLASTn comparisons between Ri1855 and various Ri and Ti plasmids. Color intensity indicates the degree of sequence similarity, as shown in the legend. A number of regions and genes are indicated on the outermost ring.

Fig. 3

Unique genes located at the very right end of the T-regions of Ri plasmids. Schematic overview of genes in the T-regions of different Ri plasmids (pRi1724: mikimopine type, pRi2659: cucumopine type, pRi8196: mannopine type, pRi1855: agropine type. Homologous genes have the same color.

Fig. 4

The proteins encoded by *orf15* and *orf16* are evolutionary related to succinamopine synthase and more distantly to ornithine cyclodeaminase. Multiple sequence alignment of proteins encoded by *orf15* and *orf16* of pRi-1855, *susL* from *A. tumefaciens* strains Chry5 and Bo542, and *ocd* from *A. tumefaciens* C58 and *Pseudomonas putida* . Colors are according to the Clustal X color scheme.

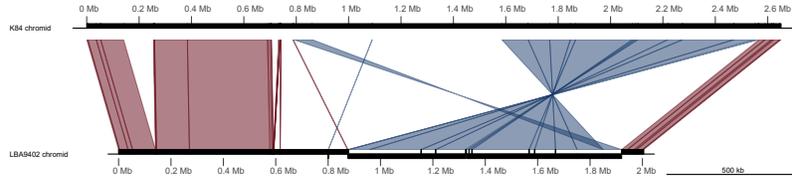


Figure 1

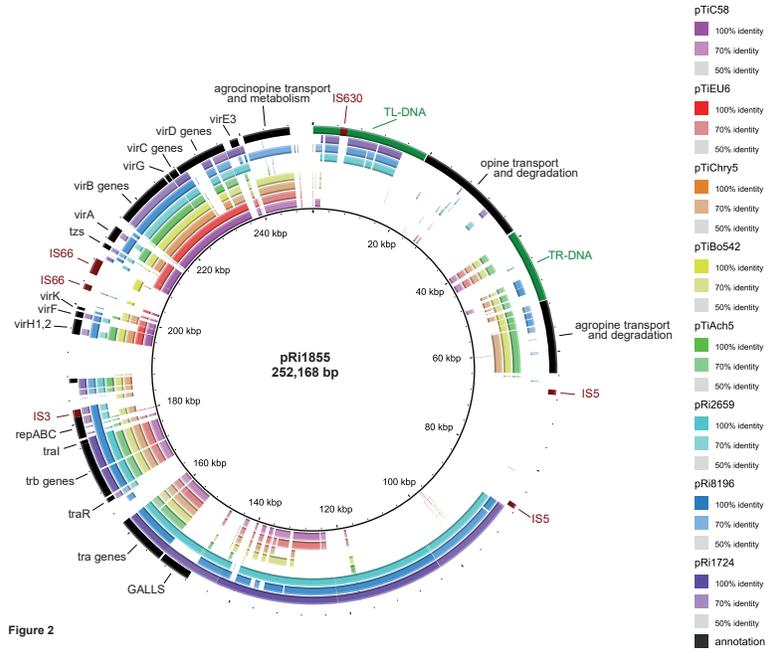


Figure 2

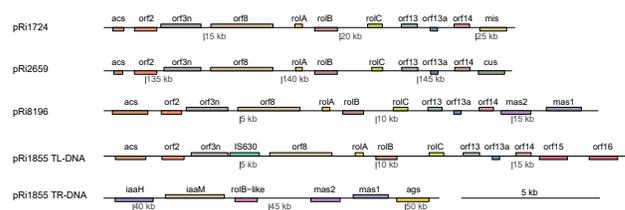


Figure 3

```

LB49402_orf15 1 M-----KGLCEVHWTESASNL-----DITPTAF--VDEA 90
LB49402_orf16 1 M-----ADQHEQCWTEESAF-----DVIYTPEL--VADA 30
Bo542_susL 1 M-----TTEISERHWSEOKLOEL-----GVAITPAL--VQDS 30
Chry5_susL 1 M-----AMEINELHWSSEOKLEL-----GVAITPTL--VQDS 30
C58_ood 1 MIGARAAAYTINKGGKMPALANLNIVRFISVENMMDLAVSTGLENFVQLAGY 54
Pputida_ood 1 M-----TYFIDVPTMSDLVHDTGVAFFIGELAA 29

LB49402_orf15 11 KS---YWFVFAAKKABHYKALTFVNW-----PGVTEGALIGYFGRYSG 74
LB49402_orf16 31 LKDL--YWHLFAEGRRLHNKVFVLDSE-----GQTEGALVGFQDYSGV 72
Bo542_susL 31 LQD---YWD--SQFEKQVKKWLQFDSIF-----NGWREGVLGCFKDYSGV 72
Chry5_susL 31 LQD---YWE--SQFDKQVKKWLQFDSIF-----NGWREGVLGCFKDYSGV 72
C58_ood 55 IEEDFRRWE-----SFDKIPRIASHRSDGVIELMPTSDDTLYGF 93
Pputida_ood 30 RQDFKRW-----AFKSRVASHSEVGVIELMPTADKSRVAF 68

LB49402_orf15 75 QDIFAFPTNAATKPLQHSDFILRDVSGTLLMSVEGVAISNGSTGWFSLACVN 128
LB49402_orf16 73 KSIHNSPNSARYKEPTSHIDVVLRNRTGRRLFSVDGVAISERRTGWLALACLN 126
Bo542_susL 73 KNIHFSPNSERNNLPLRHIDILLRRSTGERLLSLEKKAISFGRTAWFAPACID 126
Chry5_susL 73 KNIHFSPNSEQKNLPLRHIDILLRRSTGERLLSLEKKAISFGRTAWFAPACVD 126
C58_ood 94 KYVNGHPKNTKSRQTVAFQVLSVDVSGYPLLSMTILTALTAATSAIAAK 147
Pputida_ood 69 KYVNGHPANTARNLITVMAFQVLDVDSGYPLVLELTALATLTAATSLMAAQ 122

LB49402_orf15 129 LLLQGRDIDVFLFGAKVAEAVILSLNYGAARIRKVAVLSRGGKSNFELVKE 182
LB49402_orf16 127 LLLQGRSDINVFLFGAKVAEAILLALNSGASRIQRIAILRCNOSNHELVR 180
Bo542_susL 127 LLLNAKDKIEIFLFGAGLGREIVRALNENTSCRKKIWIWLSRGSSNERLVRE 179
Chry5_susL 127 LLLNAKDKIEVFLFGAGLGREIVRALNENTSCRKKIWIWLSRGSSNERLVRE 179
C58_ood 148 YLARKSRTHALIGNEAQSFEQALAFKALIGVORJLVDIDPEATRCERN 198
Pputida_ood 123 ALARPNALKMALIGNEAQSFEQALAFKHLG--IEEIVAMDTPLATAKLIAN 173

LB49402_orf15 183 LRDCVTFSLAEVQDRSLRYSQFVIMATNYGKPVFEAAETAPNAVTLSLGIDDM 236
LB49402_orf16 181 LQPEVKIISKAVNNRAYLSKSKFIIITAINSNKPVFEAAETAPNAVTLSLGIDDM 234
Bo542_susL 180 LADTVKIPLAAVWETSALSSADLVITAISSPPAFNPGDLAKAVTLSSMNDV 232
Chry5_susL 181 LADTVKIPLAAVWETSALSSADLVITAISSPPAFNPGDLAKAVTLSSMNDV 232
C58_ood 199 LQRFDFQIEAGTSAEQWVEADITATADK--HNATILSDNMIGPGVHINGV 249
Pputida_ood 174 LKEYSGLTIARRASSVAEAVKQVDITITVADK--AYATILTFDMLPEGMHLNAV 225

LB49402_orf15 237 PPD-----YIEHVLSDGLIVADLLVAMEARNVDAVALYYSRRGMKLTQHGKR 284
LB49402_orf16 236 PAD-----YFDHVLKLSGIVVADDMAMETRNIDSLALHYSRRHLKTKHGRD 282
Bo542_susL 233 PQR-----YLESLEAKAIIICDDMLAMEERNVALALLFSKGGKLESEIYFS 280
Chry5_susL 233 PQR-----YLESLEAKAIIICDDMLAMEERNVALALLFSKGGKLESEIYFS 280
C58_ood 250 GDCPGKTEHHRDILLRSDIFVEFPQRIEGEQQLARDHPVTELRVMTQQD 303
Pputida_ood 226 GDCPGKTEHLHADVLRNARVFEYEPQRIEGEQQLPADEFVVDLWRVLRQET 279

LB49402_orf15 285 DGIKNYTEILDDEALMNDLKAWKBPANFSPVGLASIDVAVAAHVYETLLNKLAE 338
LB49402_orf16 283 NGIRNYSERLDRLVLMQRLISWNGPANFSAVGLASVYDLAVAVHLYEKLAITVQ 336
Bo542_susL 281 VEVKHYSAIRAKSLFDALSTRSGPANTFTVGLASMDI AVAAHVYEMJIARIN 333
Chry5_susL 281 VEVKHYSAIRAKSLFDALSTRSGPANTFTVGLASMDI AVAAHVYEMJIARIN 333
C58_ood 304 VBRSDKDI-----LFDVSGFAIEDFSAIRYVDRVVESSHS 341
Pputida_ood 280 EBRSDSQV-----LFDVSGFAIEDYTVLRYVLDQAEKRMS 317

LB49402_orf15 339 PMSHG----- 344
LB49402_orf16 337 KYMHAASVLLA-----D 349
Bo542_susL ----- 344
Chry5_susL ----- 344
C58_ood 342 SLDL-----LLA-DDEPRDLFGMLLRQAFRLG-G 371
Pputida_ood 318 TKLD-----LVVWEDDPKDLFSHTRRAKRRIRRVA 350

```

Figure 4