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 & 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 & Chilton, 1982). Expression of rol genes located in the transfer DNA (T-DNA) leads to the transformation of normal cells into tumor cells that develop into roots that can grow 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). Based on the specific opines formed and degraded, agropine, cucumopine, man nopine, and mikimopine Ri plasmids are nowadays distinguished.

Like A. tumefaciens, A. rhizogenes has been disarmed by deleting the T-DNA genes to convert this region 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); and 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, depending on whether they carry a Ri plasmid or a Ti plasmid. Based on their physiological properties, three different groups (biovars 1–3) were distinguished already long ago (Kerr & 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 kb 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 the three subunits of an opine dehydrogenase. A candidate gene for a novel opine synthase was identified at the very right end of the TL-DNA.

2 | MATERIALS AND METHODS

2.1 | Organism

Rhizobium rhizogenes strain LBA9402 is a rifampicin-resistant derivative of wild-type strain NCPPB1855, which was originally isolated from Rosa spp. (Hooymaas, 1979). The bacterium was grown on TY medium (Difco tryptone 5 g/l, Difco yeast extract 3 g/l, CaCl\(_2\)\_6H\(_2\)O 1.3 g/l). The bacterium was tested for virulence by puncturing the stems of Kalanchee daigremontiana and Kalanchee tubiflora with a sterile wooden toothpick that had been dipped into a colony of the bacterium.

2.2 | Sequencing methods

Rhizobium rhizogenes strain LBA9402 was cultured in TY medium (Beringer, 1974), followed by genomic DNA isolation using QIAGEN gravity-flow columns (QIAGEN Genomic-tip 100/G kit Cat No./ID: 10243). 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 base calling 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 3.441 bp. Nanopore reads were end-trimmed and filtered on average quality (>Q10) and length (>5,000 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, the fourth contig of 5,386 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 & Tang, 2017). In Figure 2 and Appendix Figure A5, only complete insertion sequences, that is, 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 & Wishart, 2005). Mauve (progressive-Mauve) (Darling et al., 2010) and MUMmer (NUCmer) (Kurtz et al., 2004) were used to align the LBA9402 and K84 genomes. Average nucleotide identity (ANI) values were calculated with fastANI (Jain et al., 2018), and Digital DNA–DNA hybridization (DDH) values were estimated with GGDC 2.1 (identities/HSP length) (Meier-Kolthoff et al., 2013). 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). Percent identity values between specific regions of pRi1855 and pRiA4 were obtained with EMBOSs needle (Needleman-Wunsch pairwise global alignment). Single-nucleotide polymorphisms (SNPs) and small indels were detected with Snippy (https://github.com/tseemann/snippy, version 4.6.0, ran versus pRiA4 with the contigs option --ctgs). For the comparisons between erythritol catabolism regions and between pRi1855 and Rhizobium lusitanum strain 629, BLASTn was run locally with BLAST version 2.9.0+, and the similarities were visualized with the R package genoplotR (Guy et al., 2010). 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 were calculated with the R package seqinr.

3 | 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, 5,822 coding sequences, 9 rRNA genes (3 operons), and 53 tRNA genes were annotated. COG categories were assigned to predicted coding sequences with eggNOG-mapper (Appendix Figure A1).

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

The LBA9402 genome sequence was uploaded to the Microbial Genome Atlas webserver for taxonomic classification by TypeMat, to verify that it is indeed a R. rhizogenes strain. Average nucleotide identity (ANI) to the type strain NBRC 13257 was very high (98.9%). The digital DNA–DNA hybridization (DDH) value as calculated with Genome-to-Genome Distance Calculator 2.1 (GGDC 2.1) was 91.6% versus the type strain. These high values (>95% ANI, >70% DDH) confirm that LBA9402 is a R. rhizogenes strain. The only other complete, high-quality 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 have high sequence similarity to those of K84 (both genomes have 99% ANI, 92.1% DDH). As can be seen in Appendix Figure A2, 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, 99% ANI). It contains the genes for replication such as for a DNA 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 Type IV secretion system for mating pair formation (from position 3,324,071 to 3,314,501), but also the enzymes necessary for DNA transfer and replication (from position 3,302,269 to 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 the presence/absence of other mobile elements. For example, various proteins encoded in the unique segment of DNA from position 745,195 to 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 (Appendix Figure A3).

The second-largest replicons are less similar (Appendix Figures A2, 1), but still, 85% of the LBA9402 sequence aligns to that of K84 (and 64% of K84 aligns to LBA9402, 99% ANI). The LBA9402 sequence is smaller, mainly due to a large approximately 724 kb deletion, which seems accompanied by a large 1.8 Mbp inversion (Figure 1). The large deletion did not affect any class of genes in particular as can be seen in Appendix Figure A1b, which shows a similar distribution of the predicted proteins in COG categories in both replicons.

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 melliloti and Rhizobium leguminosarum (Appendix Figure A4) including a transport operon with genes eryEFG, a catabolic operon with genes eryABCD and a deoR-type regulator (also called eryR) followed by genes called eryH and eryl (Barbier et al., 2014; Geddes & Oresnik, 2012; Yost et al., 2006). The ability to catabolize erythritol is one of the key characteristics distinguishing biotype 1 and biotype 2 agrobacteria (Kerr & 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-encoding sequences were identified with an average size of 898 bp (Appendix Figure A5). Recently, a draft of agropine Ri plasmid pRiA4 was published (Thompson et al., 2020). This plasmid is slightly smaller than pRi1855 with a size of 249,350 bp, but has a similar restriction profile (Jouanin, 1984) and is indeed very similar to pRi1855 (99% ANI). 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 pTfBo542 (DQ058764; Oger et al., 2001), chrysopine Ti plasmid pTChry5 (KX388556; Shao et al., 2018), mannotline Ri plasmid pRi8196 (Weisberg et al., 2020), cucumopine Ri plasmid pRi2659 (NZ_CP019703.3; Valdes Franco et al., 2016; Tong et al., 2018), and mikiopine Ri plasmid 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.

As can be seen in Figure 2, the replication (repABC) and conjugative transfer (tra, trb) genes are very similar to those of other Ti and Ri plasmids. This was previously already shown for the replication and conjugative transfer genes of the closely related agropine Ri plasmid pRiA4 (Nishiguchi et al., 1987; Wetzell et al., 2015). Indeed, most of these genes are very similar in both agropine Ri plasmids. For example, pRi1855 repABC and tra are 100% identical to those of pRiA4. The traABFH and traCDG genes are 95.3% identical, and the trb operon 88.8%. In contrast, traR and trbR are less conserved (62.7%, 47.5%), also compared to the other 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 (Offringa et al., 1986; White et al., 1985). Other Ri plasmids have only one T region with very similar genes (Figure 3; Otten, 2018). In the agropine Ri plasmid, however, a copy of IS630 is inserted between orf3 and orf8. At the very left end in the agropine Ri TL region and the mannotline Ri T region, a gene for agrocinopine synthase is present, but only remnants of these genes are still present in the cucumopine and mikiopine Ri plasmids. At the very right end of the T region one (in the cucumopine and mikiopine Ri plasmids) or two (in the mannotline Ri plasmids), non-conserved genes are present. These encode the cucumopine (Valdes Franco et al., 2016) and mikiopine synthases (Moriguchi et al., 2001), respectively, and the two genes necessary for mannotline synthase in the mannotline Ri plasmid (Figure 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 the 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 the biosynthesis of the auxin indole acetic acid (Offringa et al., 1986) and the genes mas1, mas2, and ags for agropine biosynthesis (Bouchez and Tournier, 1991).

The virulence region of pRi1855 responsible for the transfer of the T-DNA into plant cells contains the essential virulence genes virA, virB1-virB11, virC, virC1, virC2, and virD1-virD5 in the same order as in other Ri and Ti plasmids, but although virE3 is present close to virD5, 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 mikiopine Ri plasmids in this area is almost identical.

![Diagram](image-url)
suggested 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 (Krallov 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. The virD3 gene, which is very variable, also shows the highest similarity to that of nopaline Ti plasmids. 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, agrocinopic 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 NCPPB1855 contain a second, catabolic plasmid with genes for catabolism of the mannitol three mannopine (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 (Figure 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 deoxyfructoseyl 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 glucos-6-phosphate dehydrogenase and 6-phosphogluconate dehydratase (Kim & Farrand, 1996). These are probably involved in further catabolism of the released 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 a mocD 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 (Figure 2), which matches the presence of an acc 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 (the area 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 (Figure 2) and two larger areas of about 20 kbp (the area 20–40 kbp on the map adjacent to the right border of the TL region in Figure 2) and about 24 kbp (the area 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 ISS-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 (Appendix Figure A6). 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 the three subunits of a putative flavin-containing opine dehydrogenase (Appendix Figures A7–A9). Flavin-containing opine dehydrogenases such as octopine, nopaline, and succinamopine dehydrogenase consist of three subunits OdhABC that are encoded by noxABC-like/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 (Appendix Figures A7–A9). 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 a specific opine into the bacterial cell. Also, genes encoding enzymes necessary for the further catabolism of the products liberated by the activity of the opine dehydrogenase on the opine substrates are often present in the vicinity. In this area of pRi1855 genes encoding such metabolic proteins are also present, including genes encoding a putative saccharopine dehydrogenase and a putative aminoacidopine 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 agropoline 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 synthase 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 (Figure 3): orf15/rolD and orf16, which share 55% identity (Figure 4, Table 1). These genes are not present in the T regions of any of the other types of Ri plasmids (Figure 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 synthetases encoded by the T region of chrysoptin pTiChry5 (Shao et al., 2018) and the agropoline 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 synthetases, which themselves share 93% identity (Table 1). All these proteins (encoded by orf15/rolD, orf16, susL) are evolutionarily 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). 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.

### Table 1

| LBA9402 orf15 | LBA9402 orf16 | Bo542 susL | Chry5 susL | C58 ocd | Pseudomonas putida ocd |
|---------------|---------------|------------|------------|---------|------------------------|
| LBA9402 orf15 | 100%          | 55%        | 46%        | 47%     | 19%                    | 19%                    |
| LBA9402 orf16 | 55%           | 100%       | 44%        | 44%     | 21%                    | 21%                    |
| Bo542 susL    | 46%           | 44%        | 100%       | 93%     | 19%                    | 18%                    |
| Chry5 susL    | 47%           | 44%        | 93%        | 100%    | 19%                    | 16%                    |
| C58 ocd       | 19%           | 21%        | 19%        | 19%     | 100%                   | 56%                    |
| Pseudomonas putida ocd | 19% | 21% | 18% | 16% | 56% | 100% |

### Discussion

The second complete, high-quality 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 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 overtime 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 agropoline pRi1855 plasmid. Over the years, sequences have already been published dealing with specific parts of the closely related agropoline Ri plasmid pRI4 and recently a draft of the completed pRIA4 sequence was published (Thompson et al., 2020). However, it still differs in numerous areas, both by base substitutions, small insertions/deletions (803 differences were detected with variant caller Snippy), and a few larger insertions/deletions encoding complete genes. Also, likely due to technical sequencing differences (the pRIA4-carrying bacterium was solely sequenced with Illumina “short read” technology whereas for LBA9402 we additionally obtained “long reads” with Nanopore sequencing), the repeat-containing GALLS gene sequence is shorter (presumably collapsed) in pRIA4.

Hairy roots formed by agropoline strains contain agropine, argopinic acid, mannopinic acid, and mannopine (Petit et al., 1983). The agropoline Ri plasmid, however, enables host strains to degrade agropine, but not the other mannityl opines. *R. rhizogenes* strains such as A4, but not NCPPB1855, 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
The proteins encoded by orf15 and orf16 are evolutionarily related to succinamopine synthase and more distantly to ornithine cyclodeaminase. Multiple sequence alignment of proteins encoded by orf15 and orf16 of P. putida oc 318 from P. putida oc C58 oc and P. putida oc Chry5 susL from A. tumefaciens MI G. Colors are according to the Clustal X color scheme.

**Figure 4** The proteins encoded by orf15 and orf16 are evolutionarily related to succinamopine synthase and more distantly to ornithine cyclodeaminase. Multiple sequence alignment of proteins encoded by orf15 and orf16 of P. putida oc 318, susL from A. tumefaciens MI G, and orf16 of P. putida oc Chry5 susL and P. putida oc C58 susL from A. tumefaciens Chry5 and P. putida oc C58. Colors are according to the Clustal X color scheme.
embraces an agcA gene for the delactonase converting agopine into mannanopine, mocC for oxidizing mannanopine, 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 agopine and mannanopine. However, it is known that the bacteria carrying pRi1855 cannot degrade mannanopine. This may be because mannanopine cannot induce the catabolic genes or because the bacterium cannot import mannanopine into the cell. Indeed, the pRi1855 plasmid contains genes for an agropine permease, but not for a mannanopine 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 cycloaminase (ocd) and which still has ornithine cycloaminase activity (Trovato et al., 2001). Here, we discovered that both rolD and orf16 have a significant identity of 44%–47% with the susL genes encoding succinamopine synthase in the agropine and chrysopine Ti plasmids. Ornithine cycloaminase encoding (ocd) genes have also been reported to have evolved into genes encoding new enzymatic activities such as alanine dehydrogenase activity in Archaeoglobus fulgidus and tauropine dehydrogenase activity in Halichondria japonica (Sharma et al., 2013; Watanabe et al., 2014). Previously, it was found that enzymes involved in the biosynthesis of mannanopine are related to and thus may have been evolved from the enzymes required for degradation (Kim & Farrand, 1996). In this case, the enzymatic activity of the enzyme remained the same, but acted in the other direction, synthesis instead of degradation. We now find that an ocd-like gene may have evolved into a gene encoding a succinamopine synthase, which catalyzes a very different enzymatic step. Similarly, 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. Indeed, three of the genes residing here together have the signature of the trios of genes such as ooxABC and noxABC (Appendix Figures A7–A9) that are known to encode the three subunits of octopine and nopaline dehydrogenase (Watanabe et al., 2015) and thus may encode the opine dehydrogenase needed for catabolism of the novel unknown opine.

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

ACKNOWLEDGEMENTS

We are grateful to Amke den Dulk-Ras and Shuai Shao for help with DNA isolation, and 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.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Marjolein J. G. Hooykaas: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Funding acquisition (supporting); Investigation (lead); Methodology (lead); Project administration (equal); Resources (supporting); Software (lead); Supervision (supporting); Validation (lead); Visualization (lead); Writing-original draft (equal); Writing-review & editing (equal). Paul J. J Hooykaas: Conceptualization (equal); Data curation (equal); Formal analysis (supporting); Funding acquisition (lead); Investigation (supporting); Methodology (supporting); Project administration (equal); Resources (lead); Software (supporting); Supervision (lead); Validation (supporting); Visualization (supporting); Writing-original draft (equal); Writing-review & editing (equal).

ETHICS STATEMENT

None required.

DATA AVAILABILITY STATEMENT

The complete genome sequence of R. rhizogenes LBA9402 is available 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. BioProject PRJNA566100: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA566100

ORCID

Paul J. J. Hooykaas https://orcid.org/0000-0002-9736-6927

REFERENCES

Alikhan, N. F., Petty, N. K., Zakour, N. L. B., & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics, 12, 402.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research, 44(W1), W16–W21.

Baek, C., Farrand, S. K., Park, D.-K., Lee, K.-E., Hwang, W., & Kim, K.-S. (2005). Genes for utilization of deoxyfructosyl glutamine (DFG), an amadori compound, are widely dispersed in the family Rhizobiaceae. FEMS Microbiology Ecology, 53, 221–233.

Barbier, T., Collard, F., Zúñiga-Ripa, A., Moriyón, I., Godard, T., Becker, J., Wittmann, C., Schaftingen, E. V., & Letesson, J.-J. (2014). Erythritol feeds the pentose phosphate pathway via three new isomerases leading to D-erythrose-4-phosphate in Brucella. Proceedings of the National Academy of Sciences of the United States of America, 111, 17815–17820.

Berenger, J. J. (1974). R factor transfer in Rhizobium leguminosarum. Journal of General Microbiology, 84, 188–198.

Bevan, M. W., & Chilton, M. D. (1982). T-DNA of the Agrobacterium Ti and Ri plasmids. Annual Review of Genetics, 16, 357–384.
Dessaux, Y., Petit, A., & Tempe, J. (1993). Chemistry and biochemistry.

Brinkman, A. B., Ettema, T. J., de Vos, W. M., & van der Oost, J. (2003). The Lrp family of transcriptional regulators. Molecular Microbiology, 48, 287–294.

Collier, R., Thomson, J., & Thilmony, R. (2018). A versatile and robust Agrobacterium-based gene stacking system generates high quality transgenic Arabidopsis plants. The Plant Journal, 95, 573–583.

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One, 5, e11147.

de Mello Serrano, G. C., e Silva Figueira, T. R., Kiyota, E., Zanata, N., & Arruda, P. (2012). Lysis degradation through the saccharopine pathway in bacteria: LKR and SDH in bacteria and its relationship to the plant and animal enzymes. FEBS Letters, 586, 905–911.

Desmet, S., Dhooghe, E., De Keyser, E., Van Huylenbroeck, J., Müller, R., Geelen, D., & Lütken, H. (2020). Rhizogenic agrobacteria as an innovative tool for plant breeding: current achievements and limitations. Applied Microbiology and Biotechnology, 104, 2435–2451.

Dessaux, Y., Petit, A., & Tempe, J. (1993). Chemistry and biochemistry of opines, chemical mediators of parasitism. Phytochemistry, 34, 31–38.

Dhillon, B. K., Laird, M. R., Shay, J. A., Winsor, G. L., Lo, R., Nizam, F., Pereira, S. K., Waglechner, N., McArthur, A. G., Langille, M. G. I., & Brinkman, F. S. L. (2015). IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Research, 43, W104–W108.

Franco, J. A. V., Collier, R., Wang, Y., Huo, N., Gu, Y., Thilmony, R., & Thomson, J. G. (2016). Draft genome sequence of Agrobacterium rhizogenes strain NCPPB2659. Genome Announcements, 4, e00746-16.

Geddes, B. A., & Oresnik, I. J. (2012). Genetic characterization of a complex locus necessary for the transport and catabolism of erythritol, adonitol and l-arabitol in S. melliloti. Microbiology, 158, 2180–2191.

Goodner, H., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B. S., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., lartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., ... Flanagan, C. (2001). Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58. Science, 294, 2323–2328.

Guy, L., Kultima, J. R., & Andersson, S. G. (2010). genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26: 2334–2335.

Harrison, P. W., Lower, R. P. J., Kim, N. K. D., & Young, J. P. W. (2010). Introducing the bacterial ‘chromid’: not a chromosome, not a plasmid. Trends Microbiology, 18, 141–148.

Henkel, C. V., den Dulk-Ras, A., Zhang, X., & Hooykaas, P. J. (2014). Genome sequence of the octopine-type Agrobacterium tumefaciens strain Ach5. Genome Announcements, 2, e00225-14.

Hodges, L. D., Cuperus, J., & Ream, W. (2004). Agrobacterium rhizogenes GALLS protein substitutes for Agrobacterium tumefaciens single-stranded DNA-binding protein VirE2. Journal of Bacteriology, 186, 3065–3077.

Hooykaas, P. J. J. (1979). The role of plasmid determined functions in the interactions of Rhizobacteria with plant cells. A genetic approach. PhD Thesis. Leiden University, The Netherlands.

Huang, Y. Y., Cho, S.-T., Lo, W.-S., Wang, Y.-C., Lai, E.-M., & Kuo, C.-H. (2015). Complete genome sequence of Agrobacterium tumefaciens Ach5. Genome Announcements, 3, e00570-15.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Molecular Biology and Evolution, 34, 2115–2122.

Jain, C., Rodriguez-R, L. M., Phlippppy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications, 9, 5114.

Jouanin, L. (1984). Restriction map of an agropine-type Ri plasmid and its homologies with Ti plasmids. Plasmid, 12, 91–102.

Jouanin, L., Vilaine, F., Tourneur, J., Tourneur, C., Pautot, V., Muller, J. F., & Caboche, M. (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. Journal of Bacteriology, 180, 2749–2755.

Kajala, K., Coil, D. A., & Brady, S. M. (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. Molecular Biology and Evolution, 30, 772–780.

Kerr, A., & Panagopoulos, C. G. (1977). Biotypes of Agrobacterium rhizobiar var. tumefaciens and their biological control. Journal of Phytopathology, 90, 172–179.

Kim, K. S., & Farrand, S. K. (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. Journal of Bacteriology, 178, 3275–3284.

Kral, L., Raschke, M., Zenk, M. H., & 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 Letters, 527, 315–318.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonecchia, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. Genome Biology, 5, R12.

Mehrotra, S., Srivastav, V., Rahman, L. U., & Kukreja, A. K. (2015). Hairy root biotechnology-indicative timeline to understand missing links and future outlook. Protoplasma, 252, 1189–1201.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics, 14, 60.

Moriguchi, K., Maeda, Y., Satou, M., Hardayani, N. S., Kataoka, M., Tanaka, N., & Yoshida, K. (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. Journal of Molecular Biology, 307, 771–784.

Nishiguchi, R., Takenami, M., & Oka, A. (1987). Characterization and sequence determination of the replicator region in the hairy-root-inducing plasmid pRIa4b. Molecular and General Genetics MGG, 206, 1–8.

Offringa, I. A., Melchers, L. S., Regensburg-Tuinck, A. J. G., Costantino, P., Schilpberoort, R. A., & Hooykaas, P. J. J. (1986). Complementation of Agrobacterium tumefaciens tumor-inducing aux mutants by genes from the TR-region of the Ri plasmid of Agrobacterium rhizogenes. Proceedings of the National Academy of Sciences of the United States of America, 83, 6935–6939.

Oger, P. M., Reich, C., Olsen, G. J., & Farrand, S. K. (2001). Complete nucleotide sequence and analysis of pTIB542: 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 Microbiology Immunology, 418, 375–420.

Petit, A., David, C., Dahl, G. A., Ellis, J. G., Guyon, P., Casse-Delbart, F., & Tempé, J. (1983). Further extension of the opine concept: Plasmids in Agrobacterium rhizogenes cooperate for opine degradation. Molecular and General Genetics MGG, 190, 204–214.
Riker, A. J. (1930). Studies on infectious hairy root of nursery apple trees. Journal of Agricultural Research, 41, 507–540.

Ron, M., Kajala, K., Pauluzzi, G., Wang, D., Reynoso, M. A., Zunstein, K., Garcha, J., Winte, S., Masson, H., Inagaki, S., Federici, F., Sinha, N., Deal, R. B., Bailey-Serres, J., & Brady, S. M. (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 Physiology, 166, 455–469.

Shao, S., van Heusden, G. P. H., & Hooykaas, P. J. J. (2019). Complete sequence of succinamopine Ti-plasmid pTiE462 reveals its evolutionary relatedness with nopaline-type Ti-plasmids. Genome Biology and Evolution, 11, 2480–2491.

Shao, S., Zhang, X., van Heusden, G. P. H., & Hooykaas, P. J. J. (2018). Complete sequence of the tumor-inducing plasmid pTiChry5 from the hypervirulent Agrobacterium tumefaciens strain Chry5. Plasmid, 96–97, 1–6.

Sharma, S., Shinde, S., & Verslues, P. E. (2013). Functional characterization of an ornithine cyclodeaminase-like gene of Arabidopsis thaliana. BMC Plant Biology, 13, 182.

Slater, S. C., Goldman, B. S., Goodner, B., Setubal, J. C., Farrand, S. K., Nester, E. W., Burr, T. J., Banta, L., Dickerman, A. W., Paulsen, I., Otten, L., Suen, G., Welch, R., Almeida, N. F., Arnold, F., Burton, O. T., & Du, Z. (2009). Genome sequences of three Agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. Journal of Bacteriology, 191, 2501–2511.

Slightom, J. L., 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. Journal of Biological Chemistry, 261(1), 108–121.

Stothard, P., & Wishart, D. S. (2005). Circular genome visualization and exploration using CGView. Bioinformatics, 21, 537–539.

Thompson, M. G., Cruz-Morales, P., Moore, W. M., Pearson, A. N., Keasling, J. D., Scheller, H. V., & Shih, P. M. (2020). Draft genome sequence of Agrobacterium fabrum ARqua1. Microbiology Resource Announcements, 9, e00506-20.

Tong, X., Li, Y., Xiang, T., Chen, Z., Wang, L., Zhang, C., Feng, M., & Wu, L. (2018). The complete genome sequence of cucumopine-type Agrobacterium rhizogenes strain K599 (NCPPB2659), a nature’s genetic engineer inducing hairy roots. International Journal of Agriculture and Biology, 20, 1167–1174.

Trovato, M., Maras, B., Linhares, F., & Costantino, P. (2001). The plant oncogene rolD encodes a functional ornithine cyclodeaminase. Proceedings of the National Academy of Sciences of the United States of America, 98, 13449–13453.

Trovato, M., Mattioli, R., & Costantino, P. (2018). From A. rhizogenes RolD to plant PCS: exploiting proline to control plant development. Plants, 7, 108.

Watanabe, S., Sueda, R., Fukumori, F., & Watanabe, Y. (2015). Characterization of flavin-containing opine dehydrogenase from bacteria. PLoS One, 10, e0138434.

Watanabe, S., Tozawa, Y., & Watanabe, Y. (2014). Ornithine cyclodeaminase/μ-crystallin homolog from the hyperthermophilic archaeon Thermococcus litoralis functions as a novel Δ1-pyrroline-2-carboxylate reductase involved in putative trans-3-hydroxy-L-proline metabolism. FEBS Open Bio, 4, 2211–2463.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics, 25, 1189–1191.

Weisberg, A. J., Davis, E. W., Tabima, J., Belcher, M. S., Miller, M., Kuo, C.-H., Loper, J. E., Grünwald, N. J., Putnam, M. L., & Chang, J. H. (2020). Unexpected conservation and global transmission of agrobacterial virulence plasmids. Science, 368, eaba5256.

Weller, S. A., Stead, E. D., O’Neill, T. M., Hargreaves, D., & McPherson, G. M. (2000). Rhizogenic Agrobacterium biovar 1 and cucumber root mat in the UK. Plant Pathology, 49, 43–50.

Wetzel, M. E., Olsen, G. J., Chakravarty, V., & Farrand, S. K. (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, F. F., & Nester, E. W. (1980). Hairy root: plasmid encodes virulence traits in Agrobacterium rhizogenes. Journal of Bacteriology, 141, 1134–1141.

White, F. F., Taylor, B. H., Huffman, G. A., Gordon, M. P., & Nester, E. W. (1985). Molecular and genetic analysis of the transferred DNA regions of the root-inducing plasmid of Agrobacterium rhizogenes. Journal of Bacteriology, 164, 33–44.

Wood, D. W., Setubal, J. C., Kaul, R., Monks, D. E., Kitajima, J. P., Okura, V. K., Zhou, Y., Chen, L., Wood, G. E., Almeida, N. F., Jr, Woo, L., Chen, Y., Paulsen, I. T., Eisen, J. A., Karp, P. D., Bovee, D. Sr, Chapman, P., Clendenning, J., Deatherage, G., ... Nester, E. W. (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, C. K., Rath, A. M., Noel, T. C., & Hynes, M. F. (2006). Characterization of genes involved in erythritol catabolism in Rhizobium leguminosarum bv. Viciae. Microbiology, 152, 2061–2074.

How to cite this article: Hooykaas MJ, Hooykaas PJ. 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. MicrobiologyOpen. 2021;10:e1180. https://doi.org/10.1002/mbo3.1180
**APPENDIX**

**FIGURE A1** LBA9402 and K84 proteins were annotated with eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups (OGs)) mapper. Shown in the figure are the frequencies at which the COG functional categories were assigned to proteins (normalized to the total number of proteins). In cases where multiple categories were assigned to a single protein, each category was counted once. On the right, a description of each COG category is shown. (a) The relative frequencies are shown per LBA9402 replicon. (b) Functional categories are shown for the chromids (secondary chromosomes) of K84 and LBA9402.

**Rhizobium rhizogenes K84**

**Rhizobium rhizogenes LBA9402**

**FIGURE A2** The LBA9402 and K84 genomes were aligned using progressiveMauve. Colored boxes are local collinear blocks. Blocks below the center line are aligned in reverse complementary orientation compared to the reference (top) sequence, and blocks above the center line are in the forward orientation. Similarity profiles are shown inside the boxes.
FIGURE A3  Multiple bacteriophage sequences were detected in LBA9402 chromosome 1. Phage search tool PHASTER was used to find prophage sequences. The schematic of the particular genomic region was generated by visualizing the raw output from PHASTER with genoplotR. The default PHASTER legend was slightly adapted by changing the color of protein categories not present in the displayed region to black. Genes annotated as “Hypothetical phage proteins” had a hit in the PHASTER phages database, but no gene function was known, whereas “Hypothetical proteins” did not have a hit in the PHASTER database.

FIGURE A4  Comparison between the erythritol transport and catabolism regions of Sinorhizobium melloti 1021, Rhizobium leguminosarum plasmid pRL12, and LBA9402. The red (same orientation) and blue (reverse orientation) ribbons show BLASTn hits, with darker shading indicating higher similarity. The data were visualized with genoplotR.
FIGURE A5  Map of pRi1855. From center to border, the rings represent GC skew, GC content, locations of genes predicted by PGAP (the colors of the arrows represent the COG functional categories predicted for the proteins; a description of the COG letters is provided in Appendix Figure A1), and finally the location of the number of features with their annotation.

FIGURE A6  The region between two IS5 family insertion sequences shows similarity to a region in the Rhizobium lusitanum strain 629 chromosome. A megablast search, with the sequence between both insertions sequences as a query, against the NCBI RefSeq genomes database (restricted to Rhizobiales sequences), and excluding Ri plasmid sequences, yielded the chromosome of Rhizobium lusitanum strain 629 as best hit (67% query coverage, 94.50% sequence identity). Depicted are the areas with a high similarity between both sequences (plotted with genoplotR). Black arrows represent coding sequences, green arrows insertion sequences, blue arrows rRNAs, and pink arrows tRNAs.
Figure A7: Multiple sequence alignment of α subunits of various opine dehydrogenases. Indicated with orange triangles is the ADP binding motif Gly-X-Gly-X-Gly and with pink triangles the [2Fe-2S] iron–sulfur cluster binding motif (Cys-X-Cys-X_{21-32}-Cys-X_{4-6}Cys). The residues putatively interacting with FMN are indicated with blue triangles. This analysis was based on the crystal structure of *Pyrococcus horikoshii* (Ph) L-Proline dehydrogenase (Ph_L-ProDH) (Tsuge et al. 2005. Crystal structure of a novel FAD-, FMN-, and ATP-containing L-proline dehydrogenase complex from *Pyrococcus horikoshii*. J Biol Chem 280: 31045–31049). Shown are the amino acid sequences of opine dehydrogenase subunits encoded by the *Agrobacterium* Ti plasmids carried by strains Chry5, Bo542, EU6, C58, and Ach5, which function as succinamopine, nopaline, and octopine dehydrogenases, respectively. In addition, sequences from *Pseudomonas putida* (Pp) and *Bradyrhizobium japonicum* (diazoefficiens) USDA110 opine dehydrogenases (which function as nopaline and octopine dehydrogenases, respectively). Finally, the sequence of the L-Proline dehydrogenase of *Pyrococcus horikoshii* (Ph) is shown. The background color of the amino acids is according to the Clustal X color scheme.
FIGURE A8 Multiple sequence alignment of β subunits of various opine dehydrogenases. Indicated with orange triangles is the ADP binding motif Gly-X-Gly-X-Gly. The arginine and tryptophan residues (putatively) interacting with FMN (as based on Pyrococcushorikoshi (Ph-L-ProDH) crystal structure data) are indicated with blue triangles. Abbreviations of species are described in the legend of Appendix Figure A7.
FIGURE A9 Multiple sequence alignment of ϒ subunits of various opine dehydrogenases. Indicated with pink triangles is the [2Fe-2S] iron–sulfur cluster binding motif (Cys-X4-Cys-X2-Cys-X11-12-Cys). Abbreviations of species are described in the legend of Appendix Figure A7.