Involvement of host and bacterial factors in Agrobacterium-mediated transformation

Shao, S.

Chapter 5
Complete sequence of the tumor-inducing plasmid pTiChry5 from the hypervirulent *Agrobacterium tumefaciens* strain Chry5

Shuai Shao, Xiaorong Zhang, G. Paul H. van Heusden and Paul J. J. Hooykaas

Department of Molecular and Developmental Genetics, Plant Cluster, Institute of Biology, Leiden University, Leiden, 2333 BE, The Netherlands

This work is published as:
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, 1-6.

Abstract
Agrobacterium tumefaciens strain Chry5 is hypervirulent on many plants including soybean that are poorly transformed by other *A. tumefaciens* strains. Therefore, it is considered as a preferred vector for genetic transformation of plants. Here we report the complete nucleotide sequence of its chrysopine-type Ti-plasmid pTiChry5. It is comprised of 197,268 bp with an overall GC content of 54.5%. Two T-DNA regions are present and 219 putative protein-coding sequences could be identified in pTiChry5. Roughly one half of the plasmid is highly similar to the agropine-type Ti plasmid pTiBo542, including the virulence genes with an identical *virG* gene, which is responsible for the supervirulence caused by pTiBo542. The remaining part of pTiChry5 is less related to that of pTiBo542 and embraces the *trb* operon of conjugation genes, genes involved in the catabolism of Amadori opines and the gene for chrysopine synthase, which replaces the gene for agropine synthase in pTiBo542. With the exception of an insertion of IS869, these Ti plasmids differ completely in the set of transposable elements present, reflecting a different evolutionary history from a common ancestor.

Introduction

Agrobacterium tumefaciens, a gram-negative and rod-shaped plant pathogen belonging to the family *Rhizobiaceae*, is the causative agent of crown gall disease. It induces tumor formation in plants by transferring a segment of its tumor-inducing plasmid (Ti-plasmid) to plant cells. This transferred DNA (T-DNA) contains genes involved in the synthesis of plant growth regulators and thus causes uncontrolled cell proliferation at infection sites (For review see: Nester et al., 1984; Gelvin, 2003; Tzfira and Citovsky, 2006; Păcurar et al., 2011; Gordon and Christie, 2014). Under laboratory conditions *Agrobacterium* is also able to transform other eukaryotes such as yeast and fungi (Bundock et al., 1995; De Groot et al., 1998). Hence, *A. tumefaciens* and its Ti plasmid have been extensively used as a vector to create transgenic plants and fungi and *Agrobacterium*-mediated transformation (AMT) has become the preferred method of transformation of these organisms over the past decades.

Many different strains of *Agrobacterium* have been isolated from nature and they may differ in their virulence on specific plants. They are often classified on the basis of the opines, tumor-specific metabolites, that are found in the tumors and serve as an energy source for the bacterium (Dessaux et al., 1993). The *A. tumefaciens* strain Chry5, which was originally isolated from chrysanthemum, turned out to have hypervirulence on soybean, which is poorly transformed by other *A. tumefaciens* strains (Bush and Pueppke, 1991). Tumors induced by Chry5 contain a novel opine called chrysopine (Chilton et al., 1995). So far pTiChry5 has only been characterized by restriction endonuclease analysis and partial sequencing (Kovács and Pueppke, 1994; Palanichelvam et al., 2000). Here, we present the complete sequence of pTiChry5.

Materials and methods

After cultivation of strain Chry5, total DNA was extracted (DNeasy Blood & Tissue Kit, QIAGEN) and two libraries were constructed using the HiSeq SBS kit v4 (Illumina). Paired-end 125 cycles sequence reads were generated using the Illumina HiSeq2500 system by BaseClear (Leiden, the Netherlands). After trimming, 4,193,789 paired-end reads (527,368,966 bases) were retained with an average length of 125 bp. Assembly was performed using the CLC Genomics Workbench software (version 7.0.4) and 31 contigs were obtained with a length ranging from 742 bp to 624 kb. Subsequently the Blast 2.2.31+ algorithm (NCBI) was used to map all contigs to the reference sequences, nopaline-type pTiC58 (AE007871) (Wood et al., 2001) and octopine-type pTiAch5 (CP007228) (Henkel et al., 2014) and contigs with high hits were identified. The relative order and relationship of these contigs was subsequently determined by a series of PCRs and small gaps were filled in by sequencing the PCR-generated fragments. The assembled sequence was annotated using the IGS Prokaryotic Annotation Pipeline (Galens et al., 2011) and the Rapid Annotations using Subsystems Technology (RAST) server (Overbeek et al., 2014) with manual modification. We found the following sequencing data deposited in GenBank under accession numbers AF065246.1 GI: 4883694 (4,168bp *traR* and *virH* genes), U88627.1 GI: 1881616 (1,000 bp 6b gene), AF229156.1 GI: 13377005 (3,797 bp *acs* gene), AF065242.2 GI: 12831440 (38,235 bp T_R DNA and Amadori catabolism genes), AF229155.1 GI: 13377004 (1,375 bp T_L border). We have compared our sequence with the previously established sequences and found single nucleotide differences in specific regions, but not evenly distributed over the entire length of the sequence. Similarly, larger differences

including a larger deletion are present in the regions showing multiple nucleotide differences. Therefore, we have carefully rechecked our reads. Although we cannot exclude that some of the differences may be snp's due to maintenance of the strains over time in different labs, it is more likely that the many differences that are located in specific areas are due to sequencing errors. Oger and Farrand proposed several gene names for the genes located in the area which they sequenced (GenBank AF065242.2 GI: 12831440), which we have mostly adopted and used below.

Results and discussion

General features of plasmid pTiChry5

The complete sequence of pTiChry5 was found to comprise 197,268 bp and the overall GC content was 54.5%. In total, 219 putative open reading frames larger than 100 bp were found with an average size of 724 bp (Table 1 and Figure 1). No genes encoding transfer-RNA (tRNA), or ribosomal RNA (rRNA), or genomic islands (GI) were found on pTiChry5 as is the case for other Ti plasmids.

An average nucleotide identity (ANI) analysis was carried out to describe the genetic relatedness with other Ti-plasmids using the Perl script (Konstantinidis and Tiedje, 2005). We extracted all known complete Ti-plasmid sequences from public databases, and calculated ANI values between pTiChry5 and these plasmids (Table 2). The results showed that pTiChry5 shared the highest similarity (97%) with agropine-type pTiBo542 (NC_010929; Oger et al., 2001) within the conserved regions. Similarity with octopine-type plasmids was less (92%) and with nopaline-type Ti-plasmids was even lower (84%). Global plasmid-wide sequence alignment between pTiChry5 and pTiBo542 using GCview showed that the plasmids share large homologous regions in the same order (Figure 1). Therefore, in the following sections we shall focus on a comparison between these two Ti plasmids. A large segment of pTiChry5 stretching from the replication units, over the virulence region, the *tra* operon, the area involved in agrocinopine catabolism up to the left part of the T_L-DNA is most highly conserved with pTiBo542. The remaining parts covering the genes dealing with degradation of the Amadori opines and with conjugation (the *trb* operon) are less or not conserved and probably have a different origin. The presence of two *repABC* replication units in both pTiChry5 and pTiBo542 also suggests that both these plasmids have originated from recombination reactions between at least two different *repABC* plasmids. In their evolution, plasmids pTiChry5 and pTiBo542 have been interrupted by different transposition events. One of these, IS869, is present at exactly the same position in both plasmids, suggesting that its insertion already occurred in the ancestor of both plasmids. Transposable elements IS292, IS1131, IS1312, and IS1313, which are present at different positions in pTiBo542 (Deng et al., 1995a; Ponsonnet et al., 1995; Wabiko, 1992), are all lacking in pTiChry5. Instead, next to the right border repeat of the T_L-DNA of pTiChry5 (see below) a copy of IS66 (96% identity) was inserted and next to it a copy of IS868 (91% identity). The IS66 element was previously discovered in other Ti plasmids (Machida et al., 1984) and IS868 is known from pTiAB3 (Paulus et al., 1991). Adjacent to the right border repeat of the T_R-DNA (see below) copies of ISRel26 (González et al., 2006) (87% identity) and ISAtu3 (81% identity) were found. ISRel26 was identified previously in *Rhizobium etli* plasmid p42a and ISAtu3 is known from accessory plasmid pAtC58 (Wood et al., 2001). Besides, only two other transposable elements were found, a copy of IS21 in the area

between Rep and the virulence region and IS869 just upstream of the *virB* operon (Paulus et al., 1991).

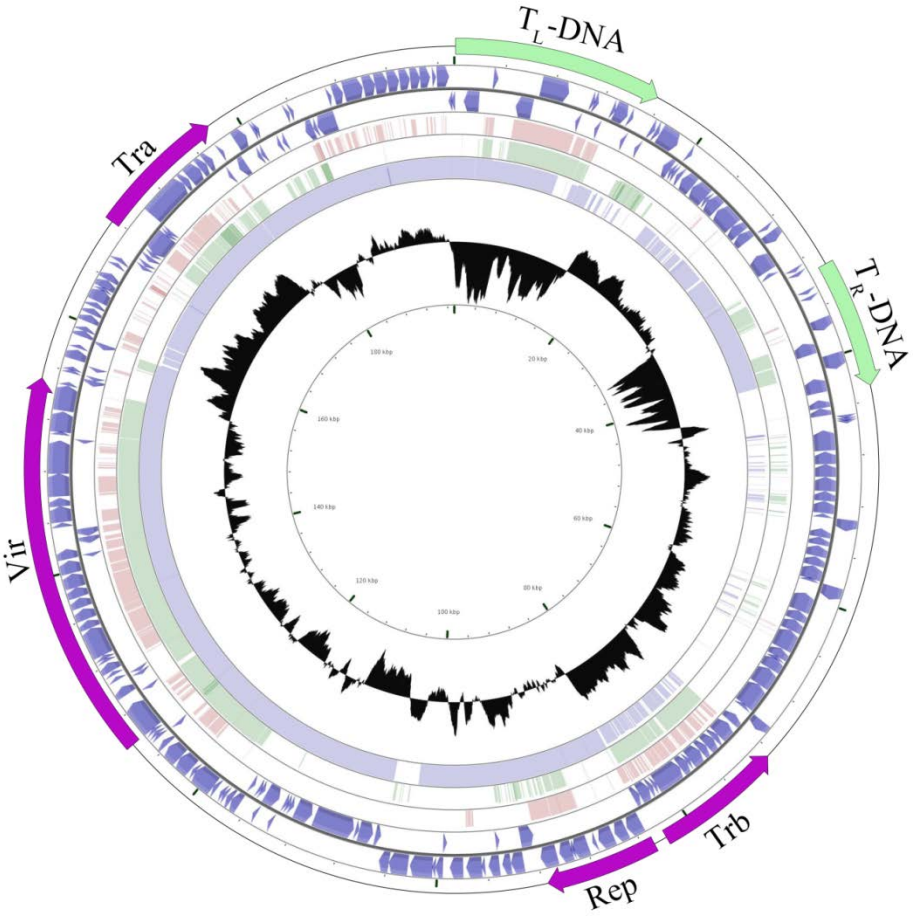


Figure 1. Schematic circular map of Ti-plasmid pTiChry5. Circle ranges from 1 (outer circle) to 6 (inner circle). Circle 1, location of T-DNA, *tra*, *trb*, *vir* and *rep* gene clusters. Circles 2 and 3, predicted open reading frames on the plus and minus strand, respectively. Circles 4, 5 and 6, coordinated mapping of pTiChry5 against Ti-plasmids pTiC58 (red), pTiAch5 (green) and pTiBo542 (light blue) respectively. The single nucleotide based similarities were denoted by color on the circles from blank to fully filled representing 0-100% similarity. Circle 7, GC content percentages. Sequence comparisons and designing the figure was performed using the CGView program (Stothard and Wishart. 2005).

Table 1. General features of Ti-plasmid pTiChry5.

Feature	pTiChry5
Size	197,268 bp
GC Content	54.5%
Protein coding regions	219
Hypothetical	50
Average ORF size	724bp
T-region	2 (15,632bp and 9,626bp)
GC % of T-regions	45.0% and 44.5%
rRNA	-
tRNA	-
IS elements	6

Table 2. Average nucleotide identity (ANI) values of sequenced Ti-plasmids.

ANI (%)	pTiSAKURA	pTiC58	pTiAch5	pTi15955	pTiBO542	pTiChry5
pTiSAKURA	-					
pTiC58	97.99	-				
pTiAch5	82.36	82.14	-			
pTi15955	82.24	82.24	99.82	-		
pTiBo542	84.92	84.32	91.35	91.22	-	
pTiChry5	84.01	84.06	92.04	92.23	97.48	-

T-DNA regions

Like octopine-type and agropine-type Ti-plasmids, pTiChry5 contains two T-DNA regions indicated as T_L-DNA (15,632 bp) and T_R-DNA (9,626 bp). Both T-DNAs are surrounded by border repeats that match the previously determined border sequences described by Palanichelvam et al. (2000). The GC content of the T-DNA regions is 45.0% and 44.5%, respectively, which is significantly lower than that of the rest of the Ti-plasmid. The T_L-DNA containing the onc-genes and the genes for l,l-succinamopine/leucinopine (*les*) and agrocinopine (*acs*) synthesis, is very similar to that of pTiBo542, but the pTiChry5 T_L-DNA harbors gene *6a*, which is not present in the pTiBo542 T_L-DNA (Figure S1). Also within the pTiBo542 T_L-DNA an IS1312-like element is inserted near the left border repeat, but this is not the case in pTiChry5. It has been speculated that this IS1312 insertion containing a "pseudo-border" sequence may cause the transfer of a truncated T-DNA lacking gene5 into plants (Deng et al., 1995b). As gene5 encodes an enzyme which can form the inactive auxin analog indole 3-lactate which may compete with auxin production (Korber et al., 1991), the absence of gene5 in tumors may be the cause of the necrosis seen on tumors induced by Bo542 (Deng et al., 1995b). With regard to T_R-DNA, most genes are well conserved between pTiChry5 and pTiBo542 except the *ags* gene (Figure S2). This gene is located adjacent to the T_R-DNA right border repeat of pTiBo542 and is involved in agropine synthesis (Figure S2). Previously, it was found that in Chry5 tumors chrysopine is produced instead of agropine (Chilton et al., 1995) and that the locus responsible for chrysopine biosynthesis is located adjacent to the right border repeat on the T_R-DNA (Palanichelvam et al., 2000). Instead of the *ags* gene, the T_R-DNA of pTiChry5 harbours the *chsA* gene responsible for chrysopine synthase (Palanichelvam et al., 2000). This gene (AgrTiChry5_50) is distantly related to the *ags* gene (36% identity), which is expected as both encode enzymes carrying out a similar lactonization reaction. It is remarkable that the pTiChry5 T_R-DNA contains not only a gene *chsC* homologous to *mas2*, which mediates formation of deoxy-fructosyl-glutamate (dfg), the direct precursor for lactonizing to chrysopine by chrysopine synthase, but also a gene *chsB* homologous to *mas1*, which encodes the enzyme responsible for the reduction of dfg into mannopine (Ellis et al., 1984). Evidently, tumors induced by Chry5 contain dfg and chrysopine, but no mannopine (Chilton et al., 1995). It is likely therefore that the gene product encoded by pTiChry5 *chsB* can no longer carry out this reduction reaction. Indeed, although the N-terminal part of *chsB* is highly conserved, the C-terminal part, which contains the reductase motifs in *mas1*, is very different from *mas1* (Figure S13).

Plasmid backbone

For further analysis of the evolutionary relationship between pTiChry5 and other Ti-plasmids, we compared their *trb*, *tra*, and *rep* gene clusters at the nucleic acid level by using ProgressiveMauve (Darling et al., 2010) (Figure 2A). With regard to the conjugation genes (Figure S5, S10), we found that the similarity of the *tra* operon with that of pTiBo542 (99% similarity) is much stronger than that of the *trb* operon (82% similarity), reflecting a likely different origin of these two operons in pTiChry5. Indeed, reversely, the similarity of the *tra* operon with that of the octopine Ti plasmid pTi15955 (similarity 81%) is weaker than that of the *trb* operon (88% similarity).

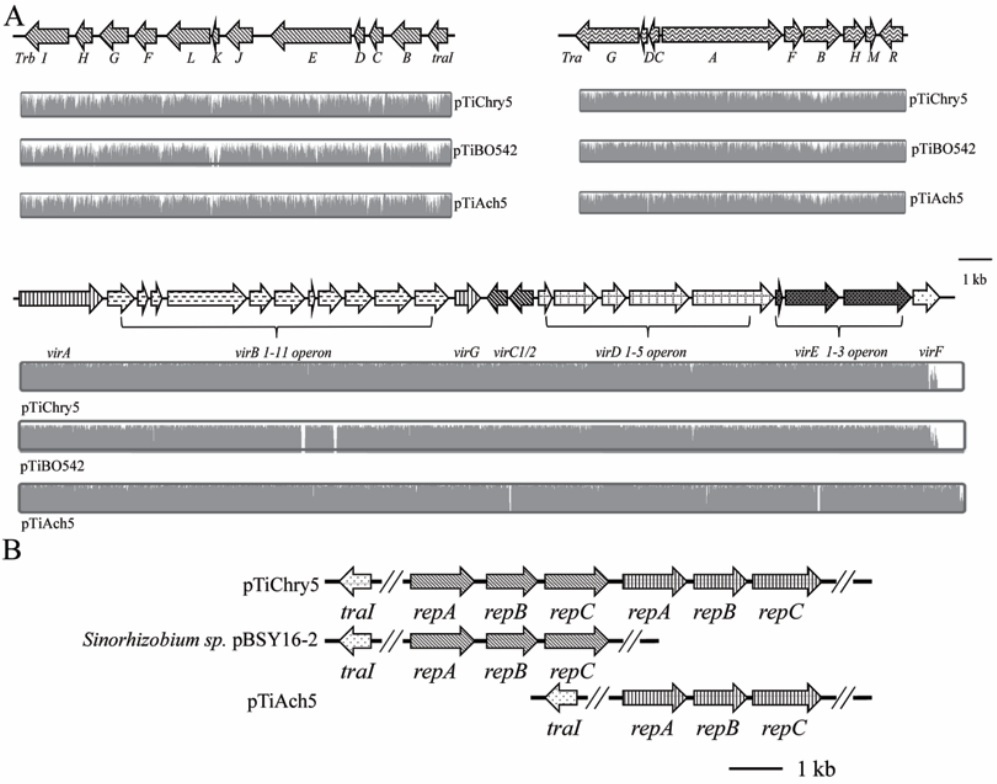


Figure 2. Comparison of the *trb*, *tra*, *vir* and *rep* gene clusters of pTiChry5 with those of other Ti-plasmids. A, Comparison of the *trb*, *tra* and *vir* gene clusters of pTiChry5, pTiBo542 and pTiAch5. The arrows indicate the genes included in these operons. ProgressiveMauve was used to generate pairwise alignments with default parameters and the single nucleotide based similarities are expressed as height of the panels from blank to filled to represent 0-100% similarity. **B,** Fine structure of the *repABC* operons of pTiChry5, pBSY16-2 (CP016451) and pTiAch5. Locations of open reading frames are shown by arrows filled with different patterns and the same pattern represents homologous genes.

Similar to pTiBo542, we found two adjacent copies of the *repABC* operon in pTiChry5 (Figure 2B, S6). The *repABC* operon (*repA* AgrTiChry5_98, *repB* AgrTiChry5_100, *repC* AgrTiChry5_101) is not only almost identical to *repABC*2 of pTiBo542, but also shares high similarity (88% with pTiC58 and 80% with pTiAch5 at the amino acid levels) with *repABC* of incRh1 Ti-plasmids. It remains to be determined whether this operon like that of pTiBo542 (Yamamoto et al., 2017) has lost its role in replication, but has only a function in

incompatibility and maybe plasmid maintenance. The other *repABC* operon, consisting of *repA* (AgrTiChry5_94), *repB* (AgrTiChry5_95) and *repC* (AgrTiChry5_96), is almost identical to *repABC1* of pTiBo542, but is only distantly related to the *repABC* operons of the *incRh1* Ti plasmids (only 40% similarity at the amino acid level). Blasting this rare *repABC* operon with the NCBI non-redundant database revealed that a very similar *repABC* operon (*repA* 94%, *repB* 89% and *repC* 98%) can be found on the accessory plasmid pBSY16-2 from the nitrogen-fixing bacterium *Sinorhizobium* sp. RAC02, (CP016451). A large set of genes of unknown function shared with pTiBo542 is located between the replication region and the virulence region (Figure S6, S7), and between the virulence region and the *tra*-operon (figure S9).

Virulence genes

As to the virulence genes (Figure S8, S10), most of the *vir* genes of pTiChry5 are highly similar to those of pTiBo542 (99% similarity) and to those of pTiAch5 (97% similarity). The *virF, virP, virQ* and *virR* genes are separated from the other *vir* genes by a large set of genes of unknown function and the *tra* genes. The *virF* gene, which enhances the virulence of octopine type strains (Melchers, et al., 1990), is replaced in pTiChry5 by a gene which shares only weak homology with *virF* of pTiAch5. This gene encodes a protein that still has the features of an F-box domain (Schrammeijer et al., 2001) and is also present in pTiBo542 (Figure S12). A previous study has shown that a small fragment containing *virG* is responsible for the enhanced virulence caused by pTiBo542 (Jin et al., 1987). The encoded VirG protein of pTiChry5 is completely identical to that of pTiBo542 and differs from octopine-type pTiAch5 VirG protein in two amino acid substitutions (at residue 7 and 106 in VirG of pTiAch5). Thus it is likely that the hypervirulence of Chry5 is at least partially due to the presence of this *virG* allele.

Opine catabolism genes

A large set of genes of Ti plasmids is involved in the catabolism of opines. We found major differences between these genes present on pTiBo542 and on pTiChry5 in line with the known catabolic properties of the host bacteria (Bush and Pueppke, 1991; Vaudequin-Dransart et al., 1995). The two Ti plasmids share a large area located between the T_L-DNA and the T_R-DNA containing genes involved in the uptake and catabolism of L,L succinamopine and leucinopine (Figure S3). Genes involved in the transport and catabolism of agrocinopines are located adjacent to the left border of the T_L-DNA, separated from this by an IS1131 element in pTiBo542, but not in pTiChry5 (Figure S11). We identified the complete *accR* regulator, which was shown to control not only the induction of the *acc*-genes by agrocinopines C and D, but also the truncated *arc*-operon comprising the *traR* regulator of conjugation (Oger and Farrand, 2001). A major difference between pTiBo542 and pTiChry5 can be seen in the segment adjacent to the T_R-DNA right border containing genes involved in the transport and catabolism of the Amadori opines (Figure S4). The pTiBo542 genes located here are involved in transport and catabolism of mannopine, mannopinic acid, agropine and agropinic acid. Genes with homology to mannopine and mannopinic acid transport and catabolism are not present in pTiChry5, which indeed does not confer catabolism of these compounds on it hosts (Chilton et al., 1995; Vaudequin-Dransart et al., 1995). However, although pTiChry5 neither confers catabolism of agropine and agropinic acid on its hosts, genes 68% similar to the genes involved in agropine transport, agropine catabolism and agropinic acid transport are nevertheless present in pTiChry5. As Chry5 does not catabolize agropine, we suspect that the four “agropine

transport" genes and the three "agropine catabolism" genes in pTiChry5 are in fact involved in the transport and catabolism of chrysopine, which is structurally related. An insertion mutation in *Bam*H1 fragment 18 of pTiChry5 was previously shown to inactivate the capacity to degrade chrysopine (Vaudequin-Dransart et al., 1998). From the sequence we can see that this fragment contains the genes AgrChry5_56 -59 with homology to the agropine transport genes *agtA-agtD*. This indicates that genes AgrChry5_56 -59 determine a chrysopine transport system. However, we find a loss of the stop codon of the *agtC* homolog and also a loss of the start codon of the *agtD* homolog. Apparently, these genes have fused in pTiChry5 and we propose to call these three genes *chtA-chtC*. The gene *chcA* with strong homology to the *agcA* gene, which encodes the agropine delactonase, probably encodes the delactonase which converts chrysopine into deoxyfructosylglutamine (dfg). Dfg in turn can be broken down into a sugar and an amino acid by enzymes encoded by *mocE* and *mocD* (Kim and Farrand, 1996). In pTiChry5 homologous genes are present, which were called *chcE* and *chcD* by Oger and Farrand (Genbank AF065242). A *mocC* gene encoding an oxido-reductase is absent as predicted before (Baek et al., 2005) and explains why Chry5, in contrast to Bo542, cannot catabolize mannopine. A regulator ChcR with high similarity to MocR, is encoded by an adjacent open reading frame. Next are five genes with some similarity to *socA*, *socB* and *socD*, which are known from the pAtC58 plasmid and involved in the uptake and catabolism of dfg, also called santhopine (Baek et al., 2003). Tumors formed by Chry5 contain not only chrysopine, but also the Amadori compounds dfg and deoxyfructosylproline (dfop), which can be catabolized by Chry5 (Chilton et al., 1995; Vaudequin-Dransart et al., 1995). However, transfer of a closely related chrysopine Ti plasmid into C58 cured of both its plasmids did not confer the ability to degrade dfg on the recipients, but this required the co-transfer of a megaplasmid indicating that the "*soc*" -genes present on pTiChry5 may not (all) be functional. In pAtC58 an uptake system for dfg is encoded by only two genes, *socA* and *socB*. While *socA* encodes a periplasmic substrate-binding protein, the *socB* gene encodes the different parts of the membrane transporter itself (Baek et al., 2003). Intriguingly in pTiChry5 the *socA*-like gene is surrounded by three genes that share homology with parts of *socB*. It needs to be tested whether these genes together still encode a functional membrane transporter. These putative "*soc*"- transport genes are followed by a catabolic gene with similarity to *socD* encoding an amadoriase. A *socC* homolog is not present and this may also be a reason that the pTiChry5 plasmid genes are not sufficient for dfg catabolism. The *socD* gene is followed by a regulatory gene, which may control the "*soc*"- genes. On the other side of this regulator a large operon is located consisting of 8 genes. At the 3'end of the operon four genes (AgrChry5_71 –74) are present with high similarity to the agropinic acid transport genes *agaA-agaD*. As Chry5 does not catabolize agropinic acid, these genes, named *dfpA-dfpD* by Oger and Farrand (Genbank AF065242), may encode a transport system for the uptake of deoxyfructosylproline (dfop), which is a lactam opine, which has structural similarity to agropinic acid. The four catabolic genes (AgrChry5_75 -78), which are present at the 5'end of the operon have similarity to *mocE*, *mocD* and to *hyuA* and *hyuB* genes encoding hydantoinases and oxoprolinases. The proteins encoded by the latter two genes may be involved in the opening of the oxoproline ring of dfop, converting it back into dfg, which can be subsequently degraded by enzymes encoded by the *mocD*- and *mocE* -like genes, which we propose to name *dfpH* and *dfpI* following the proposed nomenclature of *dfpF* and *dfpG* for the *hyu* homologs. In octopine Ti plasmids two genes with similarity to *hyuA* and *hyuB* were

previously identified and called *agaF* and *agfG* as they were involved in the catabolism of agropinic acid, which also requires lactam ring opening (Lyi et al., 1999). Some evidence for an involvement of this "dfop-operon" in the catabolism of dfop could be found in Vaudequin-Dransart et al.(1998) who reported that an insertion into the *Bam*HI fragment 8 of pTiChry5 inactivated the ability to degrade dfop. From the DNA sequence we can see that this fragment contains the genes AgrChry5_74 -79. Although this would suggest that pTiChry5 should confer not only growth on dfop, but also on dfg, this is evidently not the case. It may be that the *dfp* operon only becomes active in the presence of dfop and not by the presence of dfg, thus precluding the catabolism of dfg in the absence of dfop. Similarly, the *mocD* and *mocE* genes, which are involved in the catabolism of chrysopine and which should also allow breakdown of dfg, may need induction by chrysopine, precluding its degradation in the absence of chrysopine. It may be beneficial for the bacteria to activate these Ti plasmid-encoded opine degradation systems only in tumors when there is an abundance of opines and to prevent activation in soil, where dfg may be present. Separate dfg degradation genes, the *soc* genes, are present on non-Ti plasmids in strains such as C58 and Chry5 (Baek et al., 2005; Vaudequin-Dransart et al., 1998). These are used for degradation of dfg from soil and possibly also dfg from tumors. In pTiChry5 such genes became integrated into the Ti plasmid, but as explained above it seems that they became (partially) inactive so that also pTiChry5 requires the *soc* genes on an accessory plasmid to scavenge dfg from soil.

Conclusions

We report here the complete nucleotide sequence of pTiChry5 and the detailed alignments between pTiChry5 and pTiBo542 in order to more comprehensively understand their evolutionary relationship. As illustrated in the Figures 1 and S1-13, in general these two plasmids have a similar structure and order of functional areas with different interruptions by transposable elements and losses or gains of small sets of genes. Nevertheless, based on similarity, unexpectedly roughly one half of these plasmids (*rep*, *vir*, *tra*, *acc*, T_L-DNA) is much more similar than the other half (*trb*, Amadori opine catabolism up to the chrysopine synthase gene), suggesting that these plasmids are chimaeric due to recombination with other related plasmids. Presence of two *repABC* units in both these plasmids is in line with this hypothesis. Especially, the development of new opine profiles may have conferred evolutionary advantage on their host bacteria in some specific environments. The complete sequence presented here will be helpful in further detailed analysis of pTiChry5 and its development into a more efficient transfer vector.

Nucleotide sequence accession numbers

The complete sequence of pTiChry5 has been deposited at GenBank under the accession number KX388536 and the strain Chry5 has been deposited in the collection of the Institute of Biology, Leiden University, the Netherlands.

Acknowledgements

We thank prof Stephen Farrand (University of Illinois, Urbana, USA) for providing us with strain Chry5 and prof Phil Oger (INSA, Lyon, France) for sending us data on the sequence of pTiBo542 prior to publication.

References

Baek, C., Farrand, S.K., Lee, K., Park, D., Lee, J.K., Kim, K., 2003. Convergent evolution of Amadori opine catabolic systems in plasmids of *Agrobacterium tumefaciens*. J.Bacteriol. 185, 513-524.

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

Bundock, P., den Dulk-Ras, A., Beijersbergen, A., Hooykaas, P. J.J., 1995. Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. EMBO J. 14, 3206-3214.

Bush, A. L., Pueppke, S. G., 1991. Characterization of an unusual new *Agrobacterium tumefaciens* strain from *Chrysanthemum moriflorum* Ram. Appl. Env. Microbiol. 57, 2468-2472.

Chilton, W. S., Stomp, A. M., Beringue, V., Bouzar, H., Vaudequin-Dransart, V., Petit, A., Dessaux, Y., 1995. The chrysopine family of Amadori-type crown gall opines. Phytochemistry 40, 619-628.

Darling, A. E., Mau, B., Perna, N. T., 2010. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. Plos One 5, e11147.

De Groot, M. J., Bundock, P., Hooykaas, P. J.J., Beijersbergen, A. G., 1998. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. Nature Biotechnol. 16, 839-842.

Deng, W., Gordon, M.P., Nester, E.W., 1995. Sequence and distribution of IS1312: evidence for horizontal DNA transfer from *Rhizobium meliloti* to *Agrobacterium tumefaciens*. J.Bacteriol. 177, 2554-2559.

Deng, W., Pu, X., Goodman, R.N., Gordon, M.P., Nester, E.W., 1995. T-DNA genes responsible for inducing a necrotic response on grape vines. Mol. Plant Microbe Int. 8, 538-548.

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

Ellis, J.G., Ryder, M.H., Tate, M.E., 1984. *Agrobacterium tumefaciens* T_R-DNA encodes a pathway for agropine biosynthesis. Mol.Gen. Genet. 195, 466-473.

Galens, K., Orvis, J., Daugherty, S., Creasy, H. H., Angiuoli, S., White, O., Wortman, J., Mahurkar, A., Giglio, M. G., 2011. The IGS standard operating procedure for automated prokaryotic annotation. Stand. Genomic Sci. 4, 244.

Gelvin, S.B., 2003. Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol. Biol. Rev. 67, 16–37.

González, V., Santamaría, R. I., Bustos, P., Hernández-González, I., Medrano-Soto, A., Moreno-Hagelsieb, G., Janga, S.C., Ramírez, M.A., Jiménez-Jacinto, V., Collado-Vides,J., Dávila, G., 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc.Natl.Acad. Sci. USA. 103, 3834-3839.

Gordon, J.E., Christie, P.J. 2014. The Agrobacterium Ti plasmids. Microbiol Spectr. 2, doi:10.1128 PLAS-0010-2013.

Henkel, C. V., den Dulk-Ras, A., Zhang, X., Hooykaas, P.J.J., 2014. Genome sequence of the octopine-type *Agrobacterium tumefaciens* strain Ach5. Genome Announc. 2, e00225-14.Jin, S. G., Komari, T., Gordon, M. P., Nester, E. W., 1987. Genes responsible for the supervirulence phenotype of *Agrobacterium tumefaciens* A281. J Bacteriol. 169, 4417-4425.

Kim, K., 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. J.Bacteriol. 178, 3275-3284.

Kim, K., Chilton, W.S., Farrand, S.K., 1996. A Ti plasmid-encoded enzyme required for degradation of mannopine is functionally homologous to the T-region-encoded enzyme required for synthesis of this opine in crown gall tumors. J.Bacteriol. 178, 3285-3292.

Konstantinidis, K. T., Tiedje, J. M., 2005. Genomic insights that advance the species definition for prokaryotes. Proc.Natl.Acad. Sci. USA. 102, 2567-2572

Korber, H., Strizhov, N., Feldwisch,J., Olsson, O., Sandberg, G., Palme, K., Schell, J., Koncz, C. 1991. T-DNA gene 5 of Agrobacterium modulates auxin response by autoregulated synthesis of a growth hormone antagonist in plants. EMBO J. 10, 3983-3991.

Kovács, L. G., Pueppke, S. G, 1994. Mapping and genetic organization of pTiChry5, a novel Ti plasmid from a highly virulent *Agrobacterium tumefaciens* strain. Mol. Gen. Genet. 242, 327-336.

Lyi, S.M., Jafri, S., Winans, S.C., 1999. Mannopinic acid and agropinic acid catabolism region of the octopine-type Ti plasmid pTi15955. Mol.Microbiol. 31, 339-347.

Machida, Y., Sakurai, M., Kiyokawa, S., Ubasawa, A., Suzuki,Y., Ikeda, J.-E., 1984. Nucleotide sequence of the insertion sequence found in the T-DNA region of mutant Ti plasmid pTiA66 and distribution of its homologues in octopine Ti plasmid. Proc.Natl.Acad.Sci. USA 81,7495-7499.

Melchers, L. S., Maroney, M. J., den Dulk-Ras, A., Thompson, D. V., van Vuuren, H. A., Schilperoort, R. A., Hooykaas, P. J.J, 1990. Octopine and nopaline strains of *Agrobacterium tumefaciens* differ in virulence; molecular characterization of the *virF* locus. Plant Mol. Biol. 14, 249-259.

Nester, E.W., Gordon, M.P., Amasino, R.M,, Yanofsky, M.F. 1984. Crown gall: a molecular and physiological analysis. Ann. Rev. Plant Physiol. 35, 387-413.

Oger,P., Farrand, S.K., 2001. Co-evolution of the agrocinopine opines and the agrocinopine-mediated control of TraR, the quorum-sensing activator of the Ti plasmid conjugation system. Mol.Microbiol. 41, 1173-1185.

Oger,P., Reich, C., Olsen, G.J., Farrand, S.K., 2001. Complete nucleotide sequence and analysis of pTiBo542. Plasmid 45, 169-170.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, D206-D214.

Păcurar, D. I., Thordal-Christensen, H., Păcurar, M. L., Pamfil, D., Botez, C., Bellini, C., 2011. *Agrobacterium tumefaciens*: From crown gall tumors to genetic transformation. Physiol. Mol Plant Pathol. 76, 76-81.

Palanichelvam, K., Oger, P., Clough, S. J., Cha, C., Bent, A. F., Farrand, S. K., 2000. A second T-region of the soybean-supervirulent chrysopine-type Ti plasmid pTiChry5, and construction of a fully disarmed vir helper plasmid. Mol. Plant Microbe Int. 13, 1081-1091.

Paulus, F., Canaday, J., Otten, L., 1991. Limited host range Ti plasmids: recent origin from wide host range Ti plasmids and involvement. Mol. Plant Microbe Int. 4, 190-197.

Ponsonnet, C., Normand, P., Pilate, G., Nesme, X., 1995. IS292: a novel insertion element from *Agrobacterium*. Microbiology 141, 853-861.

Schrammeijer, B., Risseeuw, E., Pansegrau, W., Regensburg-Tuïnk, T. J., Crosby, W. L., Hooykaas, P.J.J, 2001. Interaction of the virulence protein VirF of *Agrobacterium tumefaciens* with plant homologs of the yeast Skp1 protein. Current Biology 11, 258-262.

Stothard, P., Wishart, D. S., 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21, 537-539.

Tzfira, T., Citovsky, V., 2006. *Agrobacterium*-mediated genetic transformation of plants: biology and biotechnology. Curr. Opin. Biotech. 17, 147-154.

Vaudequin-Dransart,V., Petit,A., Poncet,C., Ponsonnet,C., Nesme,X., Jones, J.B., Bouzar, H., Chilton, W.S., Dessaux, Y., 1995. Novel Ti plasmids in Agrobacterium strains isolated from fig tree and chrysanthemum tumors and their opinelike molecules. Mol. Plant Microbe Int. 8, 311-321

Vaudequin-Dransart,V., Petit,A., Chilton, W.S., Dessaux, Y., 1998. The cryptic plasmid of *Agrobacterium tumefaciens* cointegrates with the Ti plasmid and cooperates for opine degradation. Mol. Plant Microbe Int. 11, 583-591.

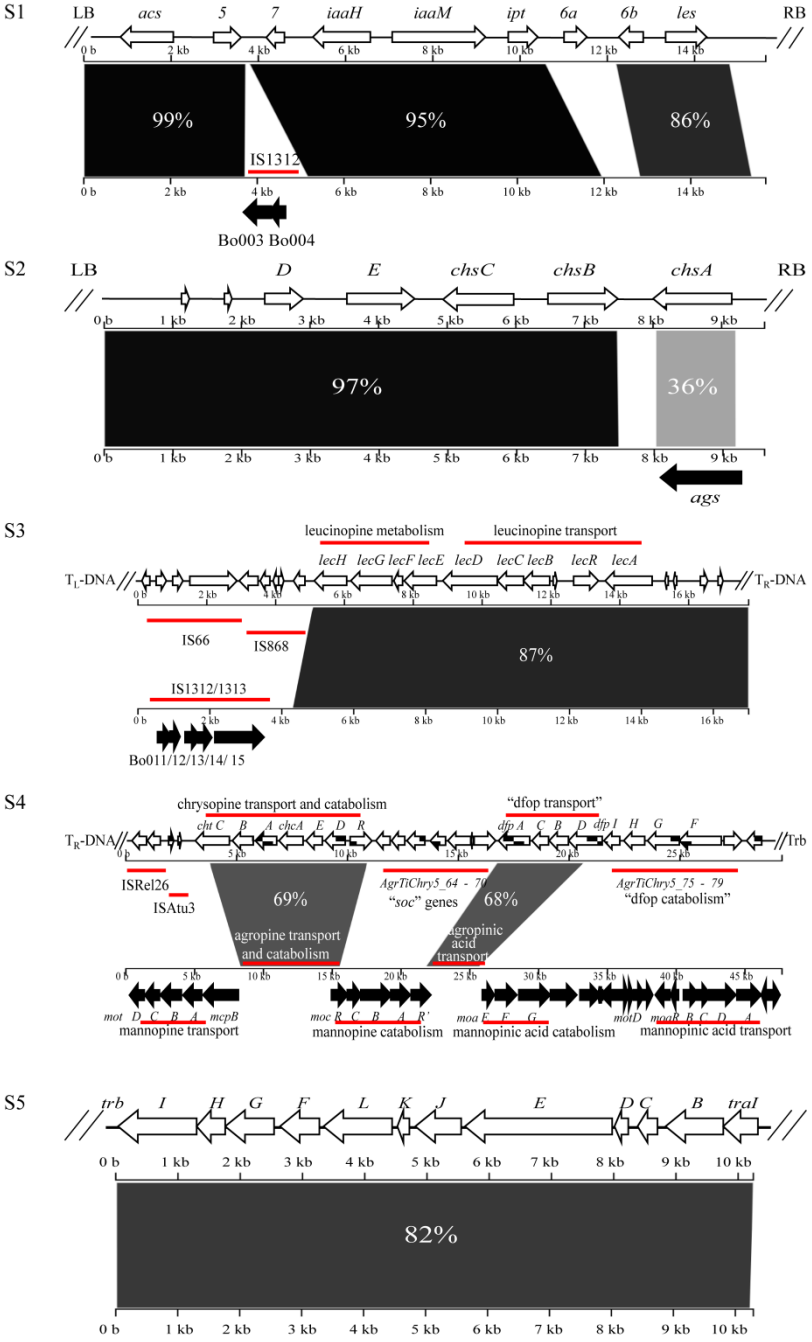
Wabiko, H., 1992. Sequence analysis of an insertion element, IS1131, isolated from the nopaline-type Ti plasmid of *Agrobacterium tumefaciens*. Gene 114, 229-233.

Wintersinger, J. A., Wasmuth, J. D., 2015. Kablammo: an interactive, web-based BLAST results visualizer. Bioinformatics. 31, 1305-1306.

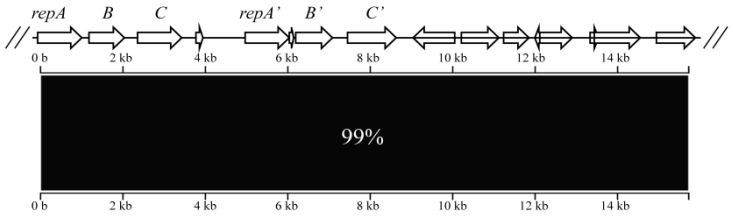
Wood, D. W., Setubal, J. C., Kaul, R., Monks, D. E., Kitajima, J. P., Okura, V. K., Zhou, Y., Chen, L., Wood, G.E., Almeida, N.F., Woo, L., Chen, Y., Paulsen, I.T., Eisen, J.A., Karp, P.D., Bovee, D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Chumley, F., Tingey, S.V., Tomb, J.-F., Gordon, M.P., Olson, M.V., Nester, E.W, 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science 294, 2317-2323.

Yamamoto,S., Agustina, V., Sakai, A., Moriguchi, K., Suzuki, K., 2017. An extra *repABC* locus in the incRh2 Ti plasmid pTiBo542 exerts incompatibility toward an incRh1 plasmid. Plasmid 90, 20-29.

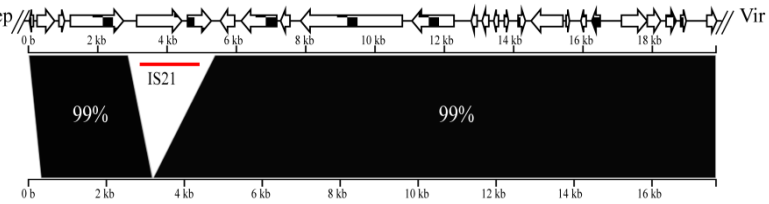
Figure S1-11. Schematic comparison between Ti plasmid pTiChry5 and reference pTiBo542. The alignment was calculated using the Blast 2.2.31+ algorithm (NCBI) and visualized by the web-tool Kablammo (Wintersinger and Wasmuth, 2015). The arrows represent genes on pTiChry5 at the top and the black arrows indicate the different genes on pTiBo542 at the bottom. S1, T_L-DNA; S2, T_R-DNA; S3, the region between T_L-DNA and T_R-DNA; S4, the region between T_R-DNA and *trb* operon; S5, *trb* operon; S6, *repABC* operon and some neighboring genes; S7, the region between the *repABC* operon and the main *vir* genes; S8, *vir* genes; S9, the region between the main *vir* genes and the *tra* operon; S10, *tra* operon; S11, the region between the *tra* operon and T_L-DNA.



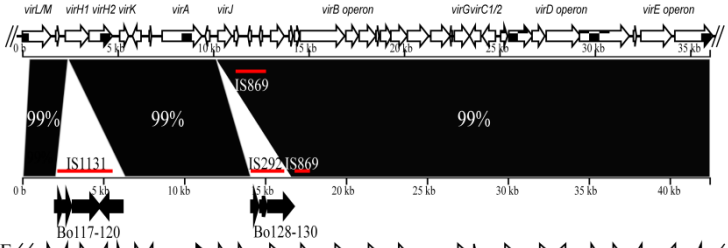
S6



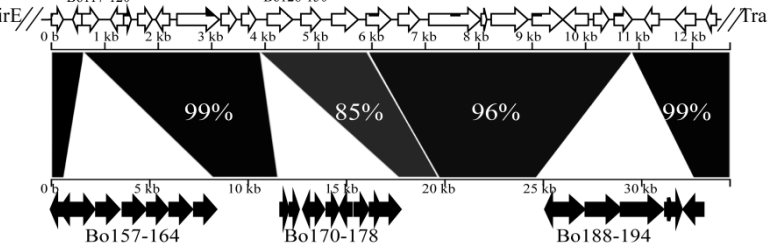
S7



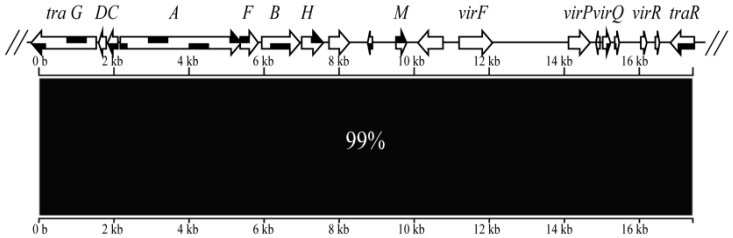
S8



S9



S10



S11

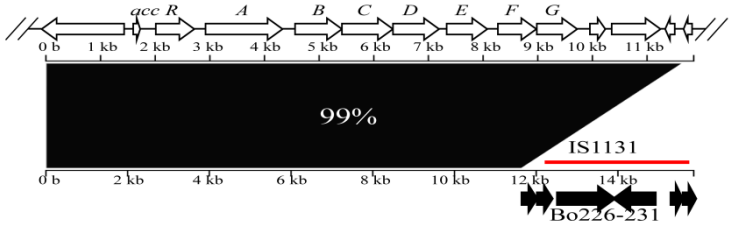


Figure S12. Alignment of the VirF protein from the octopine Ti plasmid pTi15955 in comparison with F-box proteins from nopaline pTiC58, agropine pTiBo542 and pTiChry5. The F-box is indicated in the figure by a red box.

Figure S13. Alignment of the Mas1 proteins from pTiChry5 and pTiBo542; the differences are indicated in the figure by a red box.

S12

```
pTiChry5      MVGSQDFRGVRYMQNIVSNSQTQRQPIEKEYSPFHRVMSDDLISRIVDGMVTNDPVETAR
pTiBo542      ------------MQNIVSNSQTQRQPIEKEYSPFHRVMSDDLISRIVDGMVTNDPVETAR
pTiC58        ---------MEPSQRSLSD-ESEASRTEVHSTSTNRDMPPELLAKVATYIPTQDPVETAR
pTi15955      -----------MRNSSLRD-ASGSNDAQVPMKTELLNLPDHVLTEVAKRLATNNPVESAE
                         :  : :  :  .  :   .     :  .:::.:.  : *::***:*.

pTiChry5      NISNLKLSNKSVKESVERSAGGRFHGQINRLGAASKALYDHAVPTEGFSEHPEFRPALPR
pTiBo542      NISNLKLSNKSVKESVERSAGGRFHGQINRLGAASKALYDHAVPTEGFSEHPEFRPALPR
pTiC58        NLGSLERTGRAGREAVTSDPVGKYHARMKRIGASAKTVFDTVIPGNQLPEWETNRPSPTA
pTi15955      NIANFSKSHRFTRDAVRTEPLEKFSSRLKILSRNAKLLSHAVRHAATLPDGEQLSEAQLS
              *:..:. : :  :::*  .   :: .::: :.  :* : . .     : :      :

pTiChry5      DYALAGERIDAIGGTLKLQTPQRKSALVDH-----ILNIAEPYDQGHALEYVAPHVGEFS
pTiBo542      DYALAGERIDAIGGTLKLQTPQRKSALVDH-----ILNIAEPYDQGHALEYVAPHVGEFS
pTiC58        R---TRAVGPILK----FQSEAGKSRFVTN-----ILNLPESA-QCDAILSVIKHLNDLG
pTi15955      QMRSEVATRPVLGVAYTHQDGQPEERLSGNHLDHKINNIPNL------------------
                        :      *    :. :  :     * *: :

pTiChry5      SEDRRRLVEKMLEHFSTPFTDELPSDENFSGTYALFKALPHMEDSLK----ARMLDTVVN
pTiBo542      SEDRRRLVEKMLEHFSTPFTDELPSDENFSGTYALFKALPHMEDSLK----ARMLDTVVN
pTiC58        EANKTRLIERSIEILRLEP----PLTWNMGQKCPAVDVLVQGEKYLNADQLARIQQERSS
pTi15955      -------------VFNVAE----PIMFNEI---SALEVMAEVR------PIAR-------
                          :        *   *       ...: . .        **

pTiChry5      NPDMAALIQSQKERYDWLRAQDEYEKATWRSTQKPTVDERLADIESSVRSNLSL--GQGS
pTiBo542      NPDMAALIQSQKERYDWLRAQDEYEKATWRSTQKPTVDERLADIESSVRSNLSL--GQGS
pTiC58        RPRLSPLFGRAIADFEVQKTMD-ANPRRYRDAAGPEVRAVDTVIETIERYSASLARGGPN
pTi15955      ------------------------------------------------------------

pTiChry5      DLEKMRSARPLQQTISDTLNLAREELKSRAPDVRGR----
pTiBo542      DLEKMRSARPLQQTISDTLNLAREELKSRAPDVRGR----
pTiC58        AVGLLTTNESFEKNINEVYNRTRAELDASSRDRSRSGLSR
pTi15955      -------------SIKTAHDDARAELMSADRPRSTRGL--
                          .*. . : :* ** :
```

S13

```
                                  20                  40                  60                  80
                                  |                   |                   |                   |
chsB pTiChry5 MPISQAYLGS AEPPSGIKSF YFIRHGTTDL NEKEMVVNGE KHWGVQGSGT NIGLNEKGVR QALLAGNVLR TLPISKVVCS 80
mas1 pTiBO542 MPISQAYLGS AELPSGIKSF YFIRHGTTDL NEKEMVVNGE KHWGVQGSGT NIGLNEKGVR QALLAGNVLR TLPISKVVCS 80
                                 100                 120                 140                 160
                                  |                   |                   |                   |
chsB pTiChry5 PLLRALQTAF IANPGCPTFQ IEDDLQERDF GEHEGGFGPL QMFEDDYPDC EPTEIFSIHV AKALKHARKE NVLLVAHGGV 160
mas1 pTiBO542 PLLRALQTAF IANPGCPTFQ IEDDLQERDF GEHEGGFGPL QMFEDDYPDC EPTEIFSIHV AKALEHARKE NVLLVAHGGV 160
                                 180                 200                 220                 240
                                  |                   |                   |                   |
chsB pTiChry5 LRVVAALLGV AITDEHTDNG RVLHFSLDAG NWSVRVIQSR VVMVTGATRG IGKAIAQELI RHGYRVSLGA RNVQDLEAVF 240
mas1 pTiBO542 LRVVAALLGV AITDEHTANG RVLHFSLDAG NWSVRVIQSP VVMVSGATRG IGKAIAQDLV RHGYRVSLGA RNVQALEAVF 240
                                 260                 280                 300                 320
                                  |                   |                   |                   |
chsB pTiChry5 GDQNEALHYA CFDALDHSSM KDWVSTTVAK FNKIDGLVNN AGCVDHVDLE KEINVEMLQK QWDINCVAPL IMTKLCMPYL 320
mas1 pTiBO542 GDQNEALHYA RFDALDHSSM KDWVSTTFAK FNRIDGLVNN AGCGDHVDLE KEINVELLQK QWDINCVAPL IMTKLCMPYL 320
                                 340                 360                 380                 400
                                  |                   |                   |                   |
chsB pTiChry5 IESGSGRIVN LNSMSRQRGF NSFVGYNVTN HACGSLTQII NHIGLDHRVG AIDICLAVSA TNMSSLTDTV GPEH------ 394
mas1 pTiBO542 IESGSGRIVN LNSMSGQRVA NSLVGYNMTK HGLGGLTKTT QHVGWDHGLR AVDICPGFVA TDMSSWTNLI GPDEMIQPQD 400
                                 420
                                  |
chsB pTiChry5 IAKLIREEIE RPNIAF---- --------N* 412
mas1 pTiBO542 IAKLVREAME RPNRAFVPKN EVLCMKEATR 430
```