Plant Molecular Biology **16**: 601–614, 1991.

Sequence of the *iaa* and *ipt* region of different *Agrobacterium tumefaciens* biotype III octopine strains: reconstruction of octopine Ti plasmid evolution

François Paulus, Jean Canaday, Florence Vincent, Géraldine Bonnard, Christa Kares and Léon Otten*
*Institut de Biologie Moléculaire des Plantes du C.N.R.S., 12, Rue du Général Zimmer, 67084 Strasbourg, France (*author for correspondence)*

Received 9 August 1990; accepted in revised form 29 November 1990

Key words: Agrobacterium tumefaciens, Ti plasmid, evolution, IS elements, host range

Abstract

The TA regions of biotype III octopine/cucumopine (OC) Ti plasmids are closely related to the TL region of the biotype I octopine Ti plasmids pTiAch5 and pTi15955. Sequence analysis shows that the limited and wide host range biotype III OC TA regions are derived from a common ancestor structure which lacked the *6a* gene found in the biotype I octopine TL region. The TA region of the wide host range OC Ti plasmids has conserved most of the original TL-like structure. In most wide host range OC isolates the TA-*iaaH* gene is inactivated by the insertion of an IS*866* element. However, the TA region of the wide host range isolate Hm1 carries an intact TA-*iaaH* gene. This gene encodes a biologically active product, as shown by root induction tests and indole-3-acetic acid measurements.

The limited host range OC Ti plasmids pTiAB3 and pTiAg57 have shorter TA regions which are derived from a wide host range TA region. The AB3 type arose by an IS*868*-mediated, internal TA region deletion which removed the *iaa* genes and part of the *ipt* gene and left a copy of IS*868* at the position of the deleted fragment. The pTiAB3 *iaa/ipt* deletion was followed by insertion of a second IS element, IS*869*, immediately 3′ of the *ipt* gene. pTiAg57 underwent the same *iaa-ipt* deletion as pTiAB3, but lacks the IS*868* and IS*869* elements.

Analysis of the various TA region structures provides a detailed insight into the evolution of the biotype III OC strains.

Introduction

The transformation of plant cells by *Agrobacterium tumefaciens* results from a complex interaction between a procaryotic and a eucaryotic organism. The process of tumour induction is controlled by bacterial virulence (*vir*) genes which are located on a large plasmid (Ti plasmid). Part of the plasmid DNA (the T-region) is nicked at two border sequences and transferred to the plant

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X53945.

nuclei, where the DNA is integrated as T-DNA and expressed (reviewed in [12, 20, 38]). Little is known about the evolutionary origin of this remarkable and unique system. Sequence homology between the *iaa* and *ipt* genes of the T-DNA and the bacterially expressed *iaa* and *ipt* genes found in *Agrobacterium* and *Pseudomonas syringae* ssp. *savastanoi* [2, 20, 24, 35] indicate that the genes carried by the T-regions were originally not transferred to plant cells but expressed in the bacteria. Experiments showing that the T-region borders can be replaced by a bacterial origin of transfer [5] suggest that the DNA transfer mechanism could have arisen by modification of a bacterial conjugation system.

The evolutionary origin of the Ti plasmids and their T-regions might be reconstructed, both on the structural and functional level, if a sufficient number of different Ti plasmids could be isolated and compared. Common features would be primitive, whereas elements specific for a given Ti plasmid type would be more recent.

We have studied a group of Ti plasmids which enables such a comparative analysis [23]. This paper describes several features of the biotype III octopine/cucumopine Ti plasmids which allow the reconstruction of the evolutionary history of this family.

Materials and methods

Bacterial strains
The strains used in this study are listed in Table 1.

Analysis of total bacterial DNA and plasmid DNA
Total *Agrobacterium* DNA and plasmid DNA was isolated and analysed by Southern hybridization as described previously [22, 23].

Introduction of genes in disarmed Ti vector
Transfer of genes from *Escherichia coli* to *Agrobacterium* was carried out by cloning into intermediate vectors carrying a kanamycin resistance gene and mobilization with a helper strain. Cointegrates were selected on minimal medium with neomycin as described [32]. Cointegrate structures were analysed by digestion of total bacterial DNA and Southern hybridization, using appropriate DNA fragments as probes [9].

Cloning, sequencing and sequence analysis
Cloning and sequencing procedures were as de-

Table 1. Strains and plasmids.

Number	Description	Reference
E. coli strains		
BMH71/18	host for pUC derivatives: Δ (lac pro) thi Sup*E*/lacI^{q}Z$^-$ ΔM15 pro$^+$	Müller-Hill, University of Cologne, Germany
GJ23	mobilizing strain, JC2926(pGJ28) (R64drd11), ColE1 Mob$^+$, Tra$^+$	[32]
A. tumefaciens strains		
GV3101(pPM6000)	disarmed pTiAch5 derivative	[3]
GV3101(pPM6000::)		
-pPM1023-39	A-*iaaH* in pPM6000	[13]
-pPM1025-25	TA-*iaaM* in pPM6000	[14]
-pPM1042	HA-*iaaH* in pPM6000	this study
-pPM4000-53	TB-*iaaH* in pPM6000	[13]
Plasmids		
pPM3	4.4 kb *Hin*d III (TA Tm4) in pBR322	this study
pPM1040	5.0 kb *Hin*d III (TA Hm1) in pUC18	this study
pPM1042	5.0 kb *Hin*d III (TA Hm1) in pKC7	this study
T-134	partial *Hin*d III clone TA Tm4	[4]

scribed [25]. Sequences were determined by the Sanger method, using M13 derivatives [26] and analysed with a MicroVAX with the GCG software [8].

IAA measurements

IAA synthesis was determined as described previously [15]. Briefly, *Kalanchoe daigremontiana* stem fragments were inoculated with an *Agrobacterium* suspension, and IAA levels of transformed tissues were determined by extraction, partial purification and quantification by HPLC fluorometry.

Inoculation of plants

Greenhouse plants were inoculated by puncturing stems with a Pasteur pipette and infecting the wounds with 50 μl of an overnight suspension of *Agrobacterium*.

Results

Ipt *region structure of octopine/cucumopine strains*

The *Agrobacterium tumefaciens* biotype III octopine/cucumopine (OC) strains have been divided into limited and wide host range strains [16, 17, 29]. The TA region structure of the wide host range strain Tm4 and its *ipt* region have been described [3, 13, 14]. The Tm4 *ipt* gene [3] is highly homologous to the biotype I Ach5 *ipt* gene [1, 10], but has suffered a deletion in the 3′ region which includes the polyadenylation site. In spite of this, the Tm4 *ipt* gene is biologically active [3]. The wide host range strain Hm1 has the same *ipt* region as Tm4, but differs in the *iaa* region (see below).

The TA region of the limited host range strains AB3 and Ag57 is a deleted version of the Tm4 TA region. The deletion has removed the TA-*iaaH* and TA-*iaaM* genes and part of the T-*ipt* gene. The limited host range strains have retained the *6b* oncogene and the octopine synthase gene. The unique strain Ag57 [11, 16] has a unique TA region structure [22]. The five different types of octopine T-region (15955, Tm4, Hm1, AB3 and Ag57) are shown in Fig. 1. The TB regions of the octopine/cucumopine strains are less well known; they are similar for the five different strain types and carry a set of *iaa* genes [37] which are functional [13].

Earlier comparison of the 3′ coding region of the truncated *ipt* gene of Ag162 and the corresponding *ipt* gene fragment of 15955 [36] had shown both sequences to be homologous. The *ipt* gene sequences of Ag162 and Tm4 [3] are identical (223 bp), indicating a very recent origin for the latter two Ti plasmids. In spite of this, the sequences found 3′ to the coding region of the truncated *ipt* gene of Ag162 did not show any significant homology with Ach5 or Tm4 T-region sequences [36].

We therefore sequenced the AB3 TA region 3′ of the *ipt* gene up to the well-conserved *6b* gene region. The 0.7 kb and the two 2.2 kb *Hin*d III fragments, and the 0.55 and 0.7 kb *Eco* RV fragments (Fig. 1) were sequenced. Position +1 of the sequence shown in Fig. 2a was chosen arbitrarily and corresponds to the fifth nucleotide 5′ to the *ipt* stop codon.

Figure 2a shows the sequence of the region comprised between positions 1 and 863. From position −206 to +8 the sequence is identical to the published Ag162 sequence [36] and the Tm4 sequence [3] and 87% homologous to the 15955 sequence [1]. This region is followed by a sequence not found in either Tm4 or Ach5 and situated between positions 9 and 855. From position 856 to 2452, the AB3 TA region sequence is homologous to the region situated to the right of the *ipt* gene of 15955 (75% homology) and Tm4 (99.7% homology).

Thus, compared to 15955 and Tm4, AB3 carries an additional, 851 bp sequence which is inserted immediately after the stop codon of the *ipt* gene. Further analysis revealed the presence of a 15 bp inverted repeat with two mismatches situated at the borders of this AB3-specific fragment and a small, 4 bp repeat flanking the element and found as a single 4 bp sequence (CTAG) in the *ipt* gene of Tm4. These features show that the sequence corresponds to a 847 bp bacterial insertion element. We propose to call this sequence

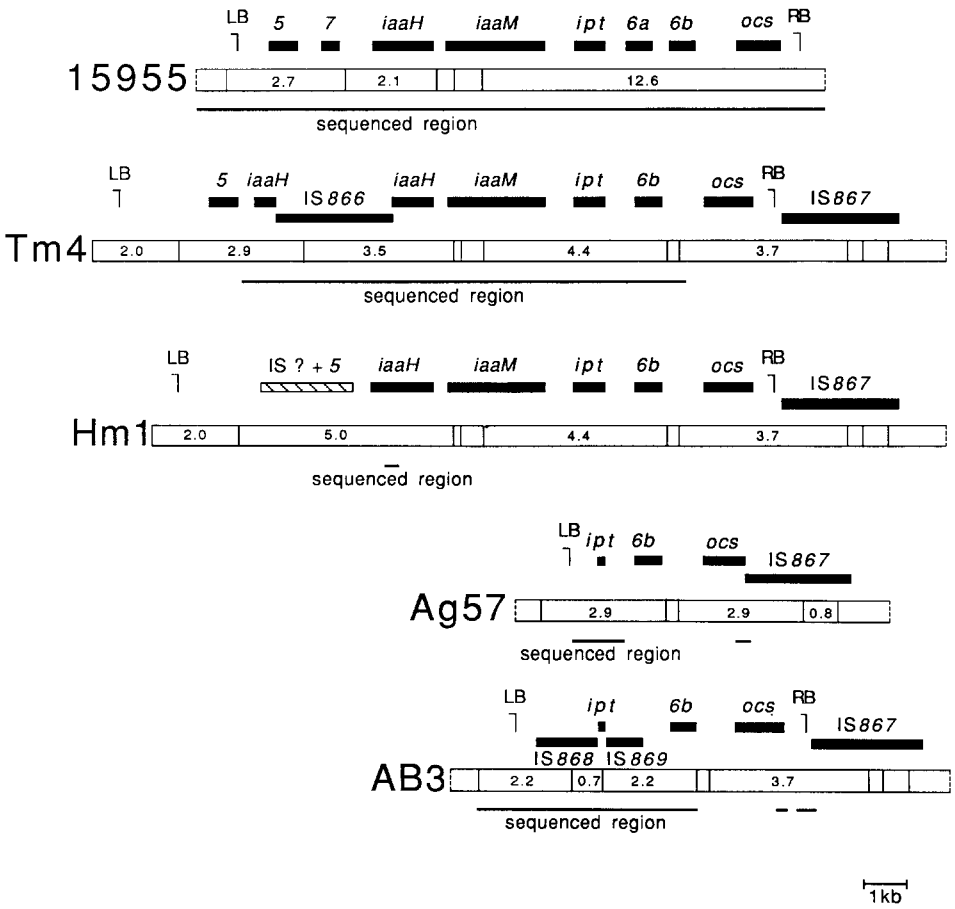

Fig. 1. Maps of the TA regions of pTi15955 [1], pTiTm4 [3, 4 and G. Bonnard, unpublished], pTiHm1, pTiAg57 and pTiAB3. The maps are centred on the stop codon of the *ipt* gene. LB, left border; RB, right border. Vertical lines: *Hin*d III restriction sites. Sizes are indicated in kb. The sequenced regions are indicated by horizontal lines under the restriction maps.

IS*869*. Fig. 2a shows the open reading frames of IS*869* and their deduced amino acid sequences as well as a number of different putative regulatory elements. Comparison with the recently established sequence of the insertion elements IS*427* of *A. tumefaciens* strain T37 [7] showed significant homology (65%) (Fig. 2b). Fig. 2c shows a comparison of the open reading frame organization of IS*869* and IS*427*. ORF 3 and ORF 4 of IS*869* correspond to ORF 4 and ORF 1 of IS*427*. Only the 3′ end of ORF 1 of IS*869* is present in IS*427*. ORF 2 of IS*869* is absent from IS*427*, whereas ORF 2 of IS*427* is lacking in IS*869*. Computer search also revealed homology (82%) between the left end of IS*869* (coordinates 20–371) and a 0.3 kb sequence upstream of the *virB* operon of the biotype I octopine strain A6 (coordinates 6–354 [33, 34], inverted with respect to the orientation of IS*869* in the AB3 TA region and indicated by V in Fig. 2c). Homology starts within the left inverted repeat of IS*869*.

*Ag57 lacks the IS*869 *element*

Previously obtained hybridization data with a pTiAch5 TL probe [22] enabled us to study the presence of IS*869* in 21 limited host range isolates. The presence of the IS*869* element in the TA region generates a characteristic 2.2 kb *Hin*d III

fragment which hybridizes to the Ach5 TL region probe pGV153 because of its homology with the Ach5 *6b* gene [22]. All isolates except Ag57 showed this 2.2 kb *Hin*d III fragment. Since the TA region of pTiAg57 had a different *Hin*d III restriction pattern, it was studied in more detail.

The TA *ipt* region of Ag57 was cloned as a 2.9 *Hin*d III fragment and sequenced using oligonucleotides synthesized on the basis of the earlier obtained AB3 sequence. Sequence analysis (not shown) of a 663 bp sequence situated between the corresponding AB3 positions – 206 and 1308 (*ipt* gene, *6b* gene and 5′ of *6b* gene) and comparison with the Tm4 sequences showed that the Ag57 sequences were 99.7% homologous to the corresponding Tm4 sequences and lack the IS*869* element and the 4 bp target site duplication found in AB3.

Occurrence of IS869-like sequences outside the TA region of AB3-like strains

Insertion sequences have been found in a number of *Agrobacterium* T regions [18, 31]. These studies did not establish whether the strains

a

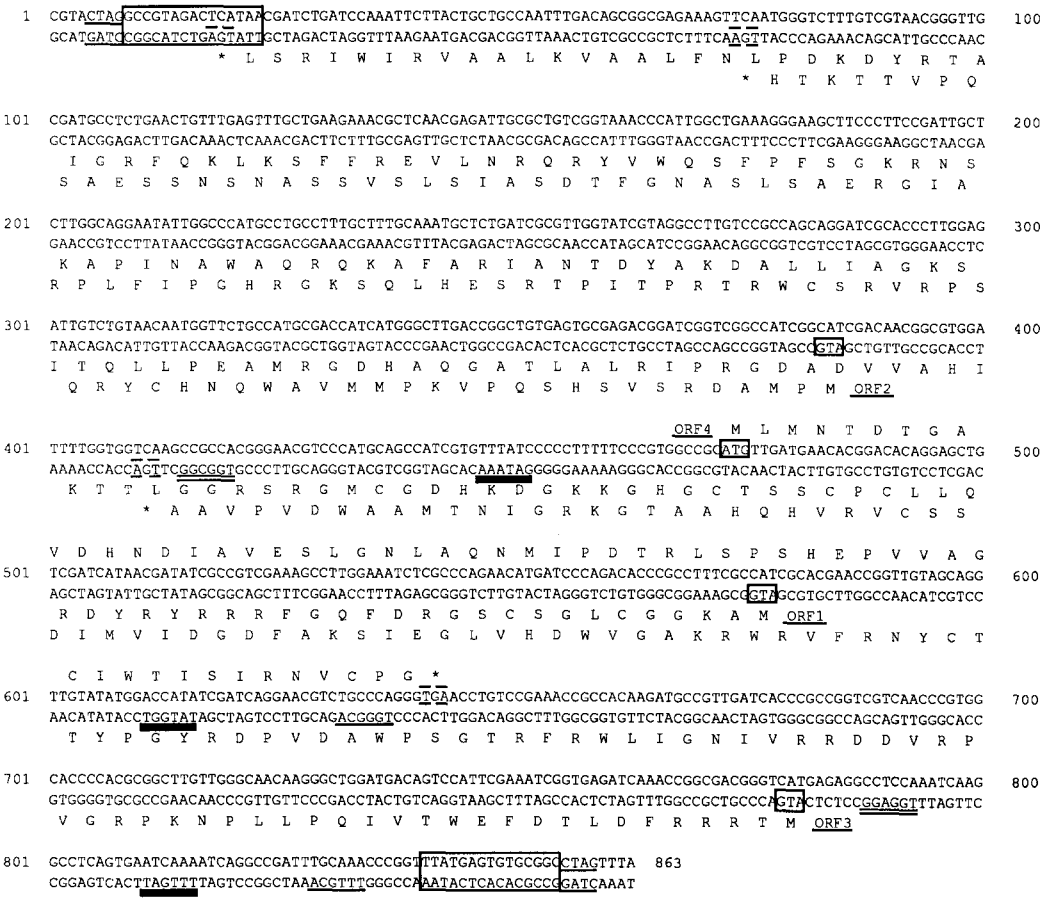

Fig. 2. a. Sequence of IS*869* of pTiAB3. The ORFs are indicated. Boxes for one strand: start codons. Dashed boxes: stop codons. Underlined: – 35 sequences. Heavily underlined: – 10 sequences. Doubly underlined: Shine and Dalgarno sequences. Boxed for both strands: inverted repeats. Underlined for both strands: target site duplications. b. Comparison of nucleotide sequences of IS*869* and IS*427* [7]. Only the homologous regions are shown. Boxed: inverted repeats. c. Comparison of open reading frame organization of IS*869* and IS*427*. V: region of homology with *virB* [33, 34].

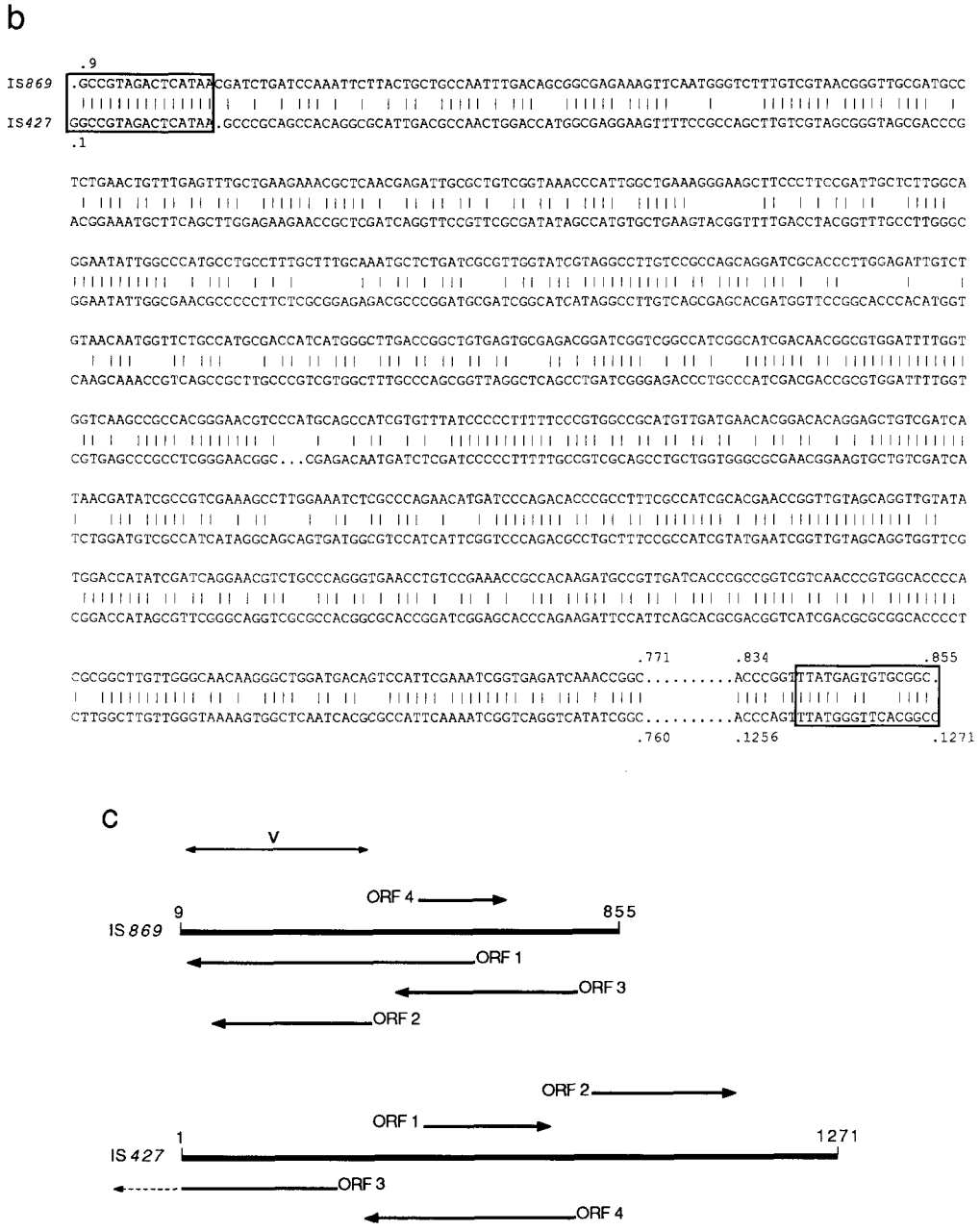

Fig. 2.

containing these elements were rare mutants or represented ecologically significant plasmid types. The occurrence of IS*869* or IS*869*-like elements in different *Agrobacterium* strains was studied by hybridizing Southern blots of total and plasmid DNA with an internal 0.6 kb *Hin*d III-*Hin*f I IS*869* fragment. The results are shown in Table 2. Biotype I and biotype II strains do not contain IS*869*-like elements (with the exception of 2654 and 2655 which carry pTiAB3-like plasmids in a biotype I background, [22]). Within the biotype III strains, all nopaline strains carry

Table 2. Presence of IS*869*-like elements in various *Agrobacterium* strains. Positive strains are indicated in bold face. References for the various strains: [4, 22, 23]. [1] biotype I strains with an AB3-like pTi plasmid [22].

Tested strains	Localization of IS*869* homology
Agrobacterium rubi ATCC13335	
Agrobacterium tumefaciens 1: biotype I. A6, AT3, B6S3, C58C1 (cured C58), **2654[1], 2655[1].**	Two copies on pTi
2: biotype II. AT181, EU6, K599, PPI-1, 281 (pTiBo542 in cured A6), 8196, 15834.	
3: biotype III. a - octopine/cucumopine strains, AB3 type. **AB3, AT6, B10/7, Zw2, 2612, 2613, 2614, 2617, 2644, 2650, 2651, 2653, 2656, 2675, 2676, 2677.**	Two copies on pTi and various copies on chromosome
b - octopine/cucumopine strains, Ag57 type. LBA649 (pTiAg57 in cured C58). c - octopine/cucumopine strains, Tm4 and Hml type. Hml, K305, K308, Tm4, 2615, 2616, 2645, 2646, 2647, 2648, 2649, 2652, 2657, 2678, 2679, 2680, 2686. d - nopaline strains. **AB4, AT1, AT66, EK2, IS1.1, NI.1, 2179, 2609.**	Various copies on chromosome
e - vitopine strains. NW11, NW221, Sz1, Sz2, **S4, 2681.**	Various copies on chromosome

IS*869*-like sequences, which are located on the chromosome. Of the biotype III vitopine strains, S4 and 2681 carry chromosomal IS*869*-like sequences, the other strains do not. Within the octopine/cucumopine groups, Tm4-like strains and C58C9(pTiAg57) i.e. LBA649, do not contain IS*869*-like sequences. The AB3-like strains carry two IS*869* elements on their Ti plasmid (one in the TA region, the other on a unidentified, 6.9 kb *Hin*d III fragment), and a variable number of IS*869*-like elements on their chromosome.

Iaa *region structure of octopine/cucumopine strains*

The wide host range OC strain Tm4 carries an intact, functional TA-*iaaM* gene and an inactive, interrupted TA-*iaaH* gene [4, 14]. The Tm4 *iaa* region was sequenced in order to compare them with biotype I octopine *iaa* genes. Internal restriction fragments of the 8.5 kb partial *Hin*d III insert of T134 [4] and of the 4.4 kb *Hin*d III insert of pPM3 were subcloned into various M13 vectors.

A 10200 bp region was sequenced which corresponds to part of the 2.9 kb *Hin*d III fragment, the entire 3474 bp, 174 bp, 524 bp, 4394 bp and 322 bp *Hin*d III fragments. This sequence will be presented elsewhere. DNA homology between the 15955 TL region [1] and the Tm4 TA region is 60% in the 3′ region of the TA-*iaaH* gene, 89% in the TA-*iaaH* coding region, 84% in the non-coding region between TA-*iaaH* and TA-*iaaM*, 90% in the TA-*iaaM* coding region and 74% in the non-coding region between TA-*iaaM* and T-*ipt*. The TA-*iaaH* gene (468 codons) is interrupted at codon 317 by the insertion sequence IS*866*, the sequence of which has been described [4]. The insertion of IS*866* generated a 8 bp target site duplication (TCTTTTTC), as shown by comparison with the 15955 sequence, which is identical to the Tm4 sequence at this position.

Since the TA region of Hml was found to lack

homology to the IS*866* element [23], the pTiHm1 *iaaH* region was cloned as a 5.0 kb *Hin*d III fragment in pUC18 (pPM1040) and compared to the corresponding Tm4 TA region by restriction mapping. The restriction map of pPM1040 (not shown) indicated that the Hm1 HA-*iaaH* gene lacks most or all of the IS*866* sequence, and carries an additional 1.3 kb sequence to the left of the HA-*iaaH* coding region, not found in Tm4.

To determine whether the HA-*iaaH* gene retained a trace of the IS*866* element or of the 8 bp target site duplication found in Tm4, the relevant part of pPM1040 was sequenced by using two oligonucleotides chosen on the basis of the earlier obtained Tm4 sequences; 326 nucleotides around the IS*866* insertion site were sequenced (results not shown). They were 100% homologous to the Tm4 sequence and showed that the HA-*iaaH* gene was intact and did not contain traces of an insertion event.

The biological activity of HA-*iaaH* was tested in two ways: by measuring its capacity to induce roots on *Kalanchoe tubiflora* and by measuring the amount of IAA produced in coinoculation experiments with the TA-*iaaM* gene of Tm4.

*Biological activity of the Hm1 HA-*iaaH *gene*

The HA-*iaaH* gene was recloned from pPM1040 into the mobilizable, kanamycin-resistant pBR322 derivative pKC7, yielding pPM1042. This intermediate vector was recombined into the disarmed Ti vector pPM6000 carried by GV3101 [3], yielding GV3101 (pPM6000 : : pPM1042). Stems of *Kalanchoe tubiflora* plants were infected with the HA-*iaaH* strain, with the TA-*iaaM* strain GV3101 (pPM6000 : : pPM1025-25) and by a mixture of the two strains.

Plants infected with the TA-*iaaM*/HA-*iaaH* strain mixture showed root formation, the control plants did not (not shown). Thus, the HA-*iaaH* gene is biologically active. Like the well characterized A-*iaaH* gene of Ach5 [27, 30], HA-*iaaH* most probably encodes an indoleacetamide hydrolase which converts the IAM synthesized under control of the TA-*iaaM* gene [14] into IAA.

The HA-*iaaH* gene was compared with the Ach5 A-*iaaH* gene and the Tm4 TB-*iaaH* gene by measuring IAA levels in *Kalanchoe* stem segments infected with the following mixtures: TA-*iaaM*/*HA-iaaH*, TA-*iaaM*/A-*iaaH* and TA-*iaaM*/TB-*iaaH* [13]. On *Kalanchoe* stem fragments, the HA-*iaaH* gene was less efficient than the closely related A-*iaaH* gene, and of about the same efficiency as the TB-*iaaH* gene of Tm4 (Table 3).

Limited and wide host range Ti plasmids have the same 6a *deletion*

Comparison of the Tm4 TA region sequence with the 15955 TL region sequence showed that the Tm4 TA region derives from a TL-like region which underwent a 0.9 kb deletion comprising the 3′ non-translated region of the *ipt* gene (including

Table 3. IAA production by various strain combinations. Values are average values with standard deviations calculated for five independent infections.

Strain(s)	Genes	IAA (pmols/segment)
GV3101(pPM6000)	–	0.36 ± 0.08
GV3101(pPM6000::pPM1025-25)+	TA-*iaaM* +	
GV3101(pPM6000)	–	0.38 ± 0.06
GV3101(pPM6000::pPM1025-25)+	TA-*iaaM* +	
GV3101(pPM6000::pPM1042)	HA-*iaaH*	3.04 ± 0.40
GV3101(pPM6000::pPM1025-25)+	TA-*iaaM* +	
GV3101(pPM6000::pPM1023-39)	A-*iaaH*	13.88 ± 1.08
GV3101(pPM6000::pPM1025-25)+	TA-*iaaM* +	
GV3101(pPM6000::pPM4000-53)	TB-*iaaH*	4.19 ± 0.65

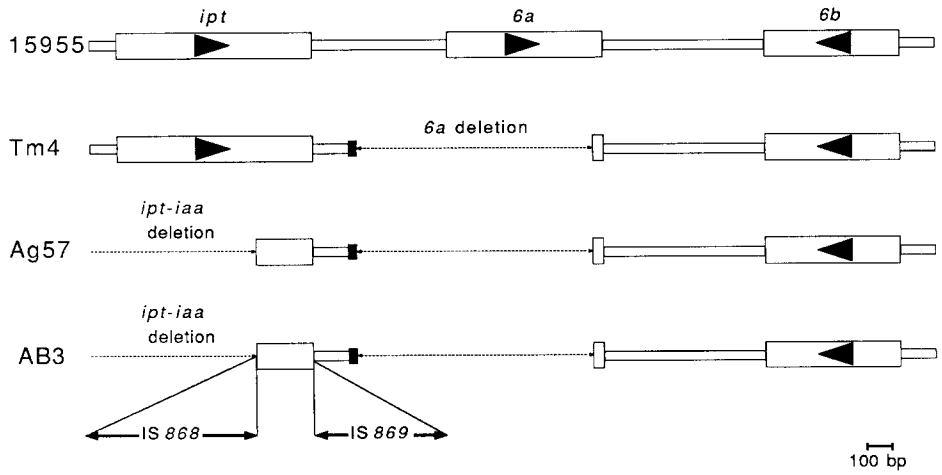

Fig. 3. Organization of the *ipt* region of pTi15955 [1], pTiTm4 [3], pTiAg57 and pTiAB3 (this study). Boxed: coding regions. Arrows indicate direction of transcription. Dotted line: deleted regions. Vertical black bars: 27 bp sequence. The insertion elements IS*868* and IS*869* are not drawn to scale.

its putative polyadenylation site) and the major part of the coding sequence of the *6a* gene [3]. In spite of the modification of its 3′ end, the Tm4 T-*ipt* gene induces shoots on tobacco, but less so than the Ach5 *ipt* gene. To establish whether the *6a* deletion also occurred in the limited host range strains, the corresponding regions of pTiAB3 and pTiAg57 were sequenced and compared with the pTiTm4 sequences. The results (not shown) demonstrate that pTiAB3 and pTiAg57 carry the same *6a* deletion as pTiTm4. Therefore, the deletion which led to the modification of the *ipt* gene and to the loss of the *6a* gene must have occurred before the limited and wide host range Ti plasmids diverged. The different structural features of the *ipt* regions of the different octopine T regions are summarized in Fig. 3.

Discussion

Earlier obtained data and results presented in this paper permit a reconstruction of the major events which led to the present-day octopine Ti plasmid structures. These events are summarized in Fig. 4. The biotype I octopine TL region found in 15955 (Arh5, A6, B6S3) and the TA region of the octopine/cucumopine biotype III strains are derived from an ancestor T-region which carried the *iaaH*, *iaaM*, *ipt*, *6a*, *6b* and *ocs* genes. In biotype I octopine strains, the region to the left of the *iaaH* gene contains genes *5* and *7*. The left part of the biotype III TA region has only low homology to gene *5* (F. Paulus, unpublished).

To the right of the octopine synthase gene the sequences of the biotype I and biotype III strains differ. Interestingly, in both cases an insertion element is situated immediately to the right of the border sequence: the composite TC region element in the case of the biotype I strains [6] and the IS*867* element in the case of the biotype III strains [23]. Further homology studies may indicate how two very similar T regions have come to be surrounded by unrelated sequences. After the separation of the TL and TA regions, the TA region evolved into different directions and generated strains like Tm4, Hm1, AB3 and Ag57. As the common areas of the Tm4, Hm1, AB3 and Ag57 TA regions are nearly 100% identical, the separation between the different strain types has taken place very recently (see below). Previous data and data from this study show that the precursor TA region which gave rise to the present-day TA region structures was composed of a 12.5 kb region, carrying intact TA-*iaaH*, TA-*iaaM*, T-*ipt*, T-*6b* and T-*ocs* genes and a deletion

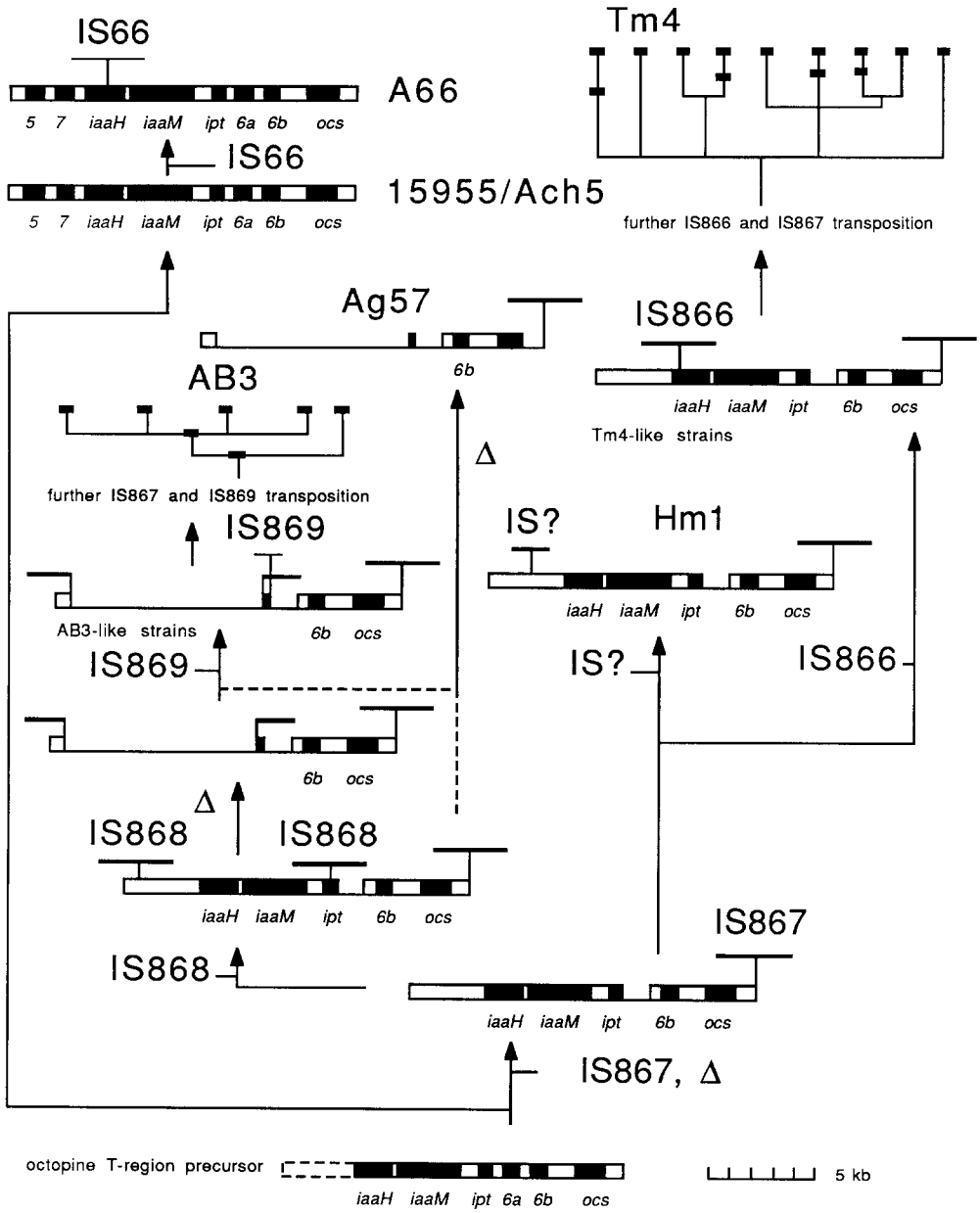

Fig. 4. Evolution of the octopine TL and octopine/cucumopine TA regions. Black areas: coding regions within the T-regions. Dashed lines: alternative pathways. The left parts of TA and TL regions are different; therefore, the left part of the precursor T-region is shown in dashed lines. Black filled-in rectangles: strains with different IS*866* and IS*867* copy numbers, simplified from [23]. For further details, see text.

in the *6a* region which removed the *6a* gene and the polyadenylation site of the *ipt* gene. The *6a* gene has been shown to code for opine secretion [19]; it remains to be tested how the absence of the *6a* gene in OC strains affects opine export from the tumours. We have earlier postulated [3] that the weak shoot-induction properties of the Tm4 *ipt* gene result from the deletion in its 3′ region and could have decreased the tumour-inhibiting effect of an originally strong *ipt* gene on

grapevine. The *6a* deletion must have occurred before Tm4, Hm1, AB3 and Ag57 diverged, since the same deletion is found in all four Ti plasmids.

The precursor TA region evolved along three lines:

1. In the unique Hungarian isolate Hm1 [28], a 1.3 kb DNA fragment, probably an IS element, was inserted in the region to the left of the TA-*iaaH* gene, leaving the *iaaH* gene intact and functional.
2. In a second line, the IS*866* element [4] was inserted into the TA-*iaaH* gene, inactivating the gene, but still allowing IAA production by tumour cells through the activity of the TA-*iaaM*, TB-*iaaM* and TB-*iaaH* genes [14]. This type of strain is common on grapevine and is represented by Tm4. Tm4 and Hm1 both induce undifferentiated tumours on grapevine, *Kalanchoe daigremontiana* and *Kalanchoe tubiflora* stems, but we have noted that Hm1 induces undifferentiated tumours on *K. daigremontiana* leaves, whereas Tm4 induces shooty structures (L. Otten, unpublished). This indicates that Hm1 induces higher levels of IAA in transformed cells than Tm4. Isolates with the Tm4 TA region structure can be further distinguished by the presence of different copy numbers of IS*866* and IS*867* elements, inserted at different positions in the genome. Comparison of the isolates has shown, that the appearance of the additional IS*866* and IS*867* copies in the chromosomal DNA occurred after the insertion of IS*866* into the TA-*iaaH* gene [23].
3. In a third line, the *iaa/ipt* region was deleted. The internal deletion of the TA region was most probably due to the insertion of two IS*868* elements in direct orientation, one to the left of the *iaaH* gene, the other within the *ipt* gene. These elements then recombined, leading to the loss of the internal region and to the presence of a single IS*868* element (F. Paulus, unpublished results). The loss of the *iaa* and *ipt* gene functions changed the host range of the strains; they became limited host range strains. Within the limited host range strains, three subtypes can be distinguished.

a. The most common one (20/23 isolates) is represented by AB3 [23]. As in the case of Tm4-like strains, the AB3-like isolates can be further distinguished by different copy numbers of the IS*867* element. The AB3- and Tm4-like strains do not have common chromosomal IS*867* copies [23]. Thus, chromosomal transposition of IS*867* occurred after the AB3/Tm4 divergence. At the same time, this suggests that the two IS*867* elements located on all OC Ti plasmids (to the right of the TA region and to the right of the TB region) were introduced shortly before the AB3/Tm4 divergence since otherwise, the AB3/Tm4 precursor strain would have been able to accumulate chromosomal IS*867* copies which would then become common elements in the AB3- and Tm4-like strains.

After the *iaa/ipt* deletion, the AB3 TA region was interrupted by another IS element, IS*869*, which is related to IS*427*. All AB3-like strains carry a second copy of IS*869* on their Ti plasmid, and several copies on their chromosomes. IS*869* sequences are not present in Tm4-like strains or on pTiAg57 but have been found in variable numbers on the chromosomes (but not on the Ti plasmids) of all biotype III nopaline strains. The different chromosomal hybridization patterns of biotype III nopaline strains and AB3-like strains may be used to refine the classification of these strains and to reconstruct their evolution as we have done for octopine/cucumopine strains by using the IS*866* and IS*867* patterns [23]. None of the biotype I (except 2654 and 2655) or biotype II strains investigated showed IS*869*-like sequences. The IS*869* element of the AB3 TA region may thus have been acquired from a biotype III nopaline strain. It is interesting to note that one of the two copies was inserted in a region which had probably already lost its function after the *iaa/ipt* gene deletion. It may be asked why the particular strain in which this happened now represents a major type of grapevine strain. It may not be necessary to postulate that the TA-associated IS*869* element brought a selective advantage to AB3-like strains, since this element may have been co-selected with some other, unidentified selective trait.

b. The other type of limited host range OC strain which we have studied in detail, Ag57, lacks IS*868*, as shown by sequencing (F. Paulus, unpublished). pTiAg57 has nearly the same *iaa/ipt* deletion as pTiAB3, which most probably resulted from insertion of two IS*868* copies which subsequently recombined. Therefore, we propose that pTiAg57 originally carried IS*868* elements, which were lost by excision. pTiAg57 also lacks the two IS*869* elements found in pTiAB3, indicating that both IS*869* insertions took place after the *iaa/ipt* deletion. Finally, the Ag57 TA region underwent a deletion of its right TA border, probably through transposition of the TA-associated IS*867* element into the octopine synthase gene and recombination between the two copies (F. Paulus, unpublished). Ag57 is the only isolate of this kind.

c. Two limited host range strains, represented by NW233, lack the *iaa* and *ipt* genes [23]. Their precise TA region structure is unknown and deserves further study.

It is remarkable that the TA regions of the different octopine/cucumopine Ti plasmids have extremely well conserved nucleotide sequences. The differences in TA region structure result from the activity of insertion sequences (IS*866*, IS*867*, IS*868*, IS*869*, an unknown element in Hm1 and a putative IS element which caused the deletion of the *6a* gene and the modification of the *ipt* gene). It remains unknown what caused the rapid and recent divergence. The structural changes may be the result of selection for T-regions with a particular oncogene composition, adapted to different grapevine varieties or to different environmental conditions. However, under greenhouse conditions, the different strain types are equally virulent on grapevine [22]. Alternatively, the various insertion and deletion events result from accidental insertions of IS elements which were selectively neutral. This model predicts that the TA-*iaaH*, TA-*iaaM* and T-*ipt* genes of wide host range strains are not required for tumour induction on grapevine. A study of various Tm4 mutants confirmed that grapevine tumour induction only requires a T-*6b* gene and the *iaaH/iaaM* gene set [14]. Thus, neither the T-*ipt* nor the TA *iaa* genes are required for tumour induction and can be mutated or lost without a change in oncogenic properties. It is tempting to speculate that the recent divergence of the OC Ti plasmids is due to the development of viticulture. The grapevine plant, grown in monocultures at vast quantities, may have become colonized by a strain carrying the OC precursor Ti plasmid in a favourable, biotype III chromosomal background. Non-essential T-region genes would subsequently have started to 'degenerate'. The shortening of the TA region observed for limited host range strains (from 12.5 kb for precursor Ti plasmid to 6.8 kb for the AB3 TA region and about 5.0 kb for the Ag57 TA region) may be thought to represent an advantage. However, the large size of the TA region of the common Tm4 strain type (15.2 kb) indicates that small size does not constitute a strong selective advantage.

The insertion of a bacterial IS sequence in a T-region is unlikely to convey a selective advantage in itself. However, the *iaa/ipt* deletion caused by IS*868* shows that the insertion and survival of an IS element in a non-essential region followed by transposition into a nearby, equally non-essential site and irreversible loss of the internal region through recombination between the two IS copies will gradually lead to the removal of non-essential DNA fragments.

Acknowledgements

We thank C. Demeirsman and J. Vanderleyden for communicating unpublished results. C.K. was supported by a grant from the Deutsche Forschungs Gemeinschaft.

References

1. Barker RF, Idler KB, Thompson DV, Kemp JD: Nucleotide sequence of the T-region from the *Agrobacterium tumefaciens* octopine Ti plasmid pTi15955. Plant Mol Biol 2: 335–350 (1983).
2. Beaty JS, Powell GK, Lica L, Regier DA, Macdonald EMS, Hommes NG, Morris RO: *Tzs*, a nopaline Ti

plasmid gene from *Agrobacterium tumefaciens* associated with *trans*-zeatin biosynthesis. Mol Gen Genet 203: 274–280 (1986).
3. Bonnard G, Tinland B, Paulus F, Szegedi E, Otten L: Nucleotide sequence, evolutionary origin and biological role of a rearranged cytokinin gene from a wide host range biotype III *Agrobacterium* strain. Mol Gen Genet 216: 428–438 (1989).
4. Bonnard G, Vincent F, Otten L: Sequence and distribution of IS*866*, a novel T-region-associated insertion sequence from *Agrobacterium tumefaciens*. Plasmid 22: 70–81 (1989).
5. Buchanan-Wollaston V, Passiatore JE, Cannon F: The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants. Nature 328: 172–175 (1987).
6. De Meirsman C, Desair J, Vanderleyden J, Van Gol J, Jen CG: Similarities between nucleotide sequences of insertion elements of *Agrobacterium tumefaciens* and *Pseudomonas savastanoi* in relation to *Agrobacterium tumefaciens* TC-DNA. Nucl Acids Res 15: 10591 (1987).
7. De Meirsman C, van Soom C, Verreth C, Van Gol J, Vanderleyden J: Nucleotide sequence analysis of IS*427* and its target sites in *Agrobacterium tumefaciens* T37. Plasmid, in press.
8. Devereux J, Haeberli P, Marquess P: The program manual for the sequence analysis software package of the Genetics Computer Group. Nucl Acids Res 12: 387–395 (1987).
9. Dhaese P, De Greve H, Decraemer H, Schell J, van Montagu M: Rapid mapping of transposon insertion and deletion mutations in the large Ti plasmids of *Agrobacterium tumefaciens*. Nucl Acids Res 7: 1837–1849 (1979).
10. Gielen J, de Beuckeleer M, Seurinck J, Deboeck F, De Greve H, Lemmers M, van Montagu M, Schell J: The complete nucleotide sequence of the TL-DNA of the *Agrobacterium tumefaciens* plasmid pTiAch5. EMBO J 3: 835–846 (1984).
11. Hoekema A, de Pater BS, Fellinger AJ, Hooykaas PJJ, Schilperoort RA: The limited host range of an *Agrobacterium tumefaciens* strain extended by a cytokinin gene from a wide host range. EMBO J 3: 3043–3047 (1984).
12. Hooykaas PJJ, Schilperoort RA: The molecular genetics of crown gall tumorigenesis. Adv Genet 22: 209–283 (1984).
13. Huss B, Bonnard G, Otten L: Isolation and functional analysis of a set of auxin genes with low root-inducing activity from an *Agrobacterium tumefaciens* biotype III strain. Plant Mol Biol 12: 271–283 (1989).
14. Huss B, Tinland B, Paulus F, Walter B, Otten L. Functional analysis of a complex oncogene arrangement in biotype III *Agrobacterium tumefaciens* strains. Plant Mol Biol 14: 173–186 (1990).
15. Kares C, Prinsen E, Van Onckelen H, Otten L: IAA synthesis and root induction with *iaa* genes under heat shock promoter control. Plant Mol Biol 15: 225–236 (1990).
16. Knauf VC, Panagopoulos CG, Nester EW: Genetic factors controlling the host range of *Agrobacterium tumefaciens*. Phytopathology 72: 1545–1549 (1982).
17. Knauf VC, Yanofsky M, Montoya A, Nester EW: Physical and functional map of an *Agrobacterium tumefaciens* plasmid that confers a narrow host range. J Bact 160: 564–568 (1984).
18. Machida Y, Sakurai M, Kiyokawa S, Ubuzawa A, Suzuki Y, Ikeda JE: Nucleotide sequence of the insertion sequence found in the T-DNA region of mutant Ti plasmid pTiA66 and distribution of its homologues. Proc Natl Acad Sci USA 81: 7495–7499 (1984).
19. Messens E, Lenaerts A, Van Montagu M, Hedges RW: Genetic basis for opine secretion from crown gall tumour cells. Mol Gen Genet 199: 344–348 (1985).
20. Morris RO: Genes specifying auxin and cytokinin biosynthesis in phytopathogens. Ann Rev Plant Physiol 37: 509–538 (1986).
21. Nester EW, Gordon MP, Amasino RM, Yanofsky MF: Crown gall: a molecular and physiological analysis. Ann Rev Plant Physiol 35: 387–413 (1984).
22. Paulus F, Huss B, Bonnard G, Ridé M, Szegedi E, Tempé J, Petit A, Otten L: Molecular systematics of *Agrobacterium tumefaciens* biotype III Ti plasmids. Mol Plant-Microbe Interact 2: 64–74 (1989).
23. Paulus F, Ridé M, Otten L: Distribution of two *Agrobacterium tumefaciens* insertion elements in natural isolates: evidence for stable association between Ti plasmids and their bacterial hosts. Mol Gen Genet 219: 145–452 (1989).
24. Powell GK, Morris RO: Nucleotide sequence and expression of a *Pseudomonas savastanoi* cytokinin gene: homology with *Agrobacterium tumefaciens tmr* and *tzs* loci. Nucl Acids Res 14: 2555–2565 (1986).
25. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).
26. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
27. Schröder G, Waffenschmidt S, Weiler EW, Schröder J: The T-region of Ti plasmids codes for an enzyme synthesizing indole-3-acetic acid. Eur J Biochem 138: 387–391 (1984).
28. Szegedi E: Host range and specific (L +) tartrate utilization of biotype 3 of *Agrobacterium tumefaciens*. Acta Phytopath Acad Scient Hung 20: 17–22 (1985).
29. Thomashow MF, Knauf VC, Nester EW: Relationship between the limited and wide host range octopine-type Ti plasmids of *Agrobacterium tumefaciens*. J Bact 146: 484–493 (1981).
30. Thomashow LS, Reeves S, Thomashow MF: Crown gall oncogenesis: evidence that a T-DNA gene from the *Agrobacterium* Ti plasmid pTiA6 encodes an enzyme that catalyzes synthesis of indoleacetic acid. Proc Natl Acad Sci USA 8: 5071–5075 (1984).
31. Vanderleyden J, Desair J, De Meirsman C, Michiel K,

Van Gol AP, Chilton M-D, Jen GC: Nucleotide sequence of an insertion sequence (IS) element identified in the T-DNA region of a spontaneous variant of the Ti plasmid pTiT37. Nucl Acids Res 14: 6699–6708 (1986).

32. Van Haute E, Joos H, Maes M, Warren E, van Montagu M, Schell J: Intergeneric transfer and exchange recombination of restriction fragments cloned in pBR322: a novel strategy for the reversed genetics of Ti plasmids of *Agrobacterium tumefaciens*. EMBO J 2: 411–417 (1983).
33. Ward JE, Akiyoshi DE, Regier D, Datta A, Gordon MP, Nester EW: Characterization of the virB operon from an *Agrobacterium tumefaciens* Ti plasmid. J Biol Chem 263: 5804–5814 (1988).
34. Winans SC, Ebert PR, Stachel SE, Gordon MP, Nester EW: A gene essential for *Agrobacterium* virulence is homologous to a family of positive regulatory loci. Proc Natl Acad Sci USA 83: 8278–8282 (1986).
35. Yamada T, Palm CJ, Brooks B, Kosuge T: Nucleotide sequence of the *Pseudomonas savastanoi* indoleacetic acid genes show homology with *Agrobacterium tumefaciens* T-DNA. Proc Natl Acad Sci USA 82: 6522–6526 (1985).
36. Yanofsky M, Lowe B, Montoya A, Rubin R, Krul W, Gordon MP, Nester EW: Molecular and genetic factors controlling host range in *Agrobacterium tumefaciens*. Mol Gen Genet 201: 237–246 (1985).
37. Yanofsky M, Montoya M, Knauf VC, Lowe B, Gordon MP, Nester EW: Limited-host-range plasmid of *Agrobacterium tumefaciens*: molecular and genetic analysis of transferred DNA. J Bact 163: 341–348 (1985).
38. Zambryski P, Tempé J, Schell J: Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants. Cell 56: 193–201 (1989).