UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de Biologie Moléculaire des Plantes

UPR CNRS 2357

THÈSE DE DOCTORAT

présentée par:

Ke Chen

soutenue le : **26 février 2016**

pour obtenir le grade de: **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité: Aspects Moléculaires et Cellulaires de la Biologie

Sequencing and functional analysis of cT-DNAs in *Nicotiana*

THÈSE dirigée par :

Prof. Léon OTTEN Professeur, Université de Strasbourg

RAPPORTEURS :

Dr. Paul HOOYKAAS Professeur, Université de Leyde, Pays-Bas

Dr. Denis FAURE Directeur de Recherche, CNRS

AUTRES MEMBRES DU JURY :

Dr. Marie-Claire LETT Professeur, Université de Strasbourg

Sequencing and functional analysis of cT-DNAs in Nicotiana

Summary:

The bacterium *Agrobacterium tumefaciens* is well-known for its utilisation in plant genetic engineering where it serves as a gene vector. This bacterium and the related species *Agrobacterium rhizogenes* are phytopathogens that induce tumors and hairy roots respectively on susceptible plants like grapevine or fruit trees. Their phytopathogenicity is due to horizontal transfer of bacterial genes to the plant host, from a plasmid called the Ti (tumor-inducing) or Ri (root-inducing) plasmid. The subject of my Thesis concerns two particular aspects of this bacterium.

1. Their capacity to stably transform several plant species in nature, thereby yielding naturally transformed plants, especially in the genus Nicotiana. We have shown by deep sequencing of the *Nicotiana tomentosiformis* genome and by analysis of other recently published Nicotiana sequences the presence of five different Agrobacterium-derived sequences (cT-DNAs), totalling 65 kb, some of which carry intact genes. We have shown that two of them (TB-*mas2'* from *N. tabacum* and TE-*6b* from *N. otophora*) have biological activity. A detailed comparative study has allowed us to better understand the evolution of these cT-DNAs (Chen et al., 2014). The *mas2'* gene is well-known, it codes for the synthesis of desoxyfructosyl-glutamine (DFG) in tumors or roots induced by Agrobacterium. Recent work in our group has shown that the TB-*mas2'* gene is highly expressed in some *N. tabacum* cultivars and leads to the accumulation of detectable amounts of DFG. This work is presented as a manuscript to be submitted.

2. A second part of the Thesis describes new properties of the T-*6b* gene, which is part of the DNA transferred by *A. vitis* strain Tm4 and leads to abnormal growth caracterized by the appearance of enations, so far the mode of action of this gene is unknown. The *6b* gene is part of the so called *plast* family (for phenotypic plasticity), with different and often remarkable growth effects on plants. The T-*6b* gene wasearlier placed under control of a dexamethasone-inducible promoter, and tobacco plants transformed with this construct have now been studied in detail, at different times after the start of induction. A large number of changes was analyzed, both at the morphological and anatomical level, these include various unprecedented morphological changes, like for example the appearance of shoot primordia at the base of trichomes, or the appearance of ectopic vascular strands parallel to the normal strands with a regular development leading to complex but predictable structures (Chen and Otten, 2015). The TE-*6b* gene from *N. otophora* was placed under strong and constitutive promoter control and introduced into tobacco, where it was found to cause new types of morphological change, different from those observed for T-*6b*. The latter results are preliminary and will be presented as a complement to the work on T-*6b*. They indicate that the introduction of the TE-*6b* gene in the *N. otophora* ancestor could have caused a change in growth pattern, and might have favored the appearance of a new species.

Keywords: Agrobacterium, *N. tomentosiformis*, cT-DNA, natural transformation, horizontal gene transfer, *6b*, leaf polarity, *plast* genes

Séquençage et analyse fonctionelle d'AND-cT dans Nicotiana

Résumé:

La bactérie *Agrobacterium tumefaciens* est bien connue pour son utilisation en génie génétique végétale où elle sert comme vecteur de gènes. A l'origine, cette bactérie ainsi que l'espèce voisine *Agrobacterium rhizogenes* sont des bactéries phytopathogènes qui induisent respectivement des tumeurs et des racines anormales sur des plantes sensibles telles que la vigne ou des arbres fruitiers. L'action pathogène résulte d'un transfert horizontal de gènes de la bactérie vers l'hôte végétal, à partir d'un plasmide, le pTi (plasmide inducteur de tumeurs) ou pRi (plasmide inducteur de racines). Mon travail de thèse concerne deux aspects particuliers de cette bactérie.

1. Sa capacité à transformer durablement des espèces végétales dans la nature, donnant ainsi naissance à des plantes naturellement transformées, notamment dans le genre Nicotiana. Nous avons pu montrer par séquençage à haut débit du génome de *N. tomentosiformis* et par l'analyse d'autres séquences complètes de Nicotianées publiées récemment l'existence inattendue de 5 séquences venant d'Agrobacterium (cT-DNAs) avec une taille total de 65 kb, dont certaines portent des gènes intacts. Nous avons montré que deux de ces gènes (TB-*mas2'* de *N. tabacum* et TE-*6b* de *N. otophora*) ont une activité biologique. Une étude comparative approfondie a permis de mieux comprendre l'évolution de ces cT-DNAs (Chen et al., 2014). Le gène *mas2'* est bien connu, il code pour une enzyme qui catalyse la synthèse du désoxyfructosyl-glutamine (DFG) dans des tumeurs ou racines induites par Agrobacterium. Des résultats récents dans notre groupe portant sur le gène TB-*mas2'* montrent que ce gène est exprimé de façon très active dans plusieurs cultivars de *N. tabacum*, et y donne naissance à l'apparition de quantités mesurables de DFG. Ce travail est présenté sous forme d'un manuscrit à soumettre.

2. Une deuxième partie de la Thèse concerne les propriétés du gène T-*6b*, qui fait partie de l'ADN transféré par *A. vitis* souche Tm4 et provoque une croissance anormale caractérisée par l'apparition d'énations, sans que l'on connaisse son mode d'action. Le gène *6b* fait partie de la famille des gènes plast (pour plasticité phénotypique), avec des effets différents et souvent remarquables sur la croissance des plantes. Le gène T-*6b* a été mis sous contrôle d'un promoteur inductible par le dexaméthasone, et des plantes de tabac transformées par cette construction ont été étudiées en détail, à différents moments après son induction. Un grand nombre de changements a été décrit incluant des analyses anatomiques montrant des modifications encore jamais décrites chez les plantes, comme par exemple l'apparition de méristèmes foliaires ectopiques à la base de trichomes, ou l'apparition de systèmes vasculaires ectopiques parallèles au système vasculaire normal avec un développement régulier menant à des structures complexes ordonnées (Chen and Otten, 2015). Le gène TE-*6b* de *N. otophora* a été mis sous contrôle d'un promoteur fort constitutif et introduit dans des plantes de tabac, où il provoque des changements de croissance différents de ce qui a été observé pour le gène T-*6b*. Ces derniers résultats préliminaires sont présentés en complément des observations sur le gène T-*6b*. Ils indiquent que le transfert horizontal du gène TE-*6b* vers l'ancêtre de *N. otophora* aurait pu contribuer à une modification de la croissance et ainsi à la création d'une nouvelle espèce.

Mots clefs: Agrobacterium, *N. tomentosiformis*, cT-DNA, naturella transformation, transfert horizontal du gène, *6b*, polarité des feuilles, *plast* gène

Acknowledgements

From the most beginning I would like to thank my supervisor professor Léon Otten for his four years help during my thesis in Strasbourg. He is so patient and precise to explain our topic, related topics, each experiment and its principle. We make efforts together to enhance my scope of knowledge and also vocabulary in plant molecular field. He also encouraged me to learn French in order to get used to a French lab. It is doubtless to say, if there is no help of him, it would be impossible for me to finish my PhD these. I really enjoy the time working with him.

I would like to thank our collaborators Dr. François Dorlhac de Borne and Emilie Julie from Imperial Tobacco of Bergerac. As a coauthor of my first paper, François offered very important data e.g. RNA expression levels of *mas1'* and *mas2'* from different tobacco cultivars. Emilie helped to send lots of tobacco cultivars seeds. It is important to mention that they offered me a chance to present my topic in an international tobacco conference in Paris. I got a valuable experience there.

I am very grateful to another collaborator Dr. Ernö Szegedi from Hungary even I have never met him. He helped to improve the detection method of deoxyfructosyl-glutamine by paper electrophoresis a lot. He works so effcient that as soon as we propose questions, he answers immediately and sends us samples or Agrobacteria we need.

I would like to thank Dr. Patrick Pale and Julie Obszynski from Chemical Institute of Strasbourg University. Since they are experts in chemical field, they gave us quite a lot of advices and help on synthesis and detection of the amadoric compound deoxyfructosyl-glutamine. Only with their help, the *mas2'* gene study can have further progress.

I have to thank my colleagues Patrick Achard, Jean Michel, Lali Achard and Thomas Regnault. Since we work in a same lab, they helped me a lot in many field not only scientific research but also language and culture. We together created a very good research environment.

Many thanks to all people in IBMP, specially greenhouse team and sequencing technician Malek Alioua. Thanks for their hard and prescise work. I am so happy to know all these people and to be friends with them. I like the experience in IBMP. It's a big treasure for my life.

I am very happy and proud to have Dr. Paul Hooykaas, Dr. Denis Faure, Dr. Marie Claire Lett and Dr. Wenhui Shen as my jury members. Thank you so much for your coming and I hope you enjoy my presentation and our discussion.

I have to thank the China Scholarship Council (CSC) for their financial support during four years.

我要感谢我的家人，爸爸，妈妈，奶奶，外婆，其他的家人和朋友。在你们的鼓励下，我才能完成这篇博士论文。尽管这四年来，我身处异国他乡，但感谢你们一直以来的支持和牵挂，感谢你们无条件的付出，你们是我最坚实的后盾。

最后，感谢我的男朋友魏鹏飞。这四年来，我们的工作和学习都很忙。我们的恋爱经过了异国，时差的考验，但是我们从未想过放弃。感谢你一直以来默默的付出和支持，感谢你的体谅和坚持。我相信我们可以克服任何艰难险阻，一起创造美好的明天。

Table of Contents

Abbreviations

AGO	argonaute
ANI	average nucleotide identity
ARF	auxin response factor
ATP	adenosine triphosphate
BA	6-benzyl adenine
BBM	baby boom
BLAST	basic local alignment search tool
bp	base pair
CaMV	cauliflower mosaic virus
cT- DNA	cellular T-DNA
cv.	cultivar
Da	Dalton
DDH	DNA-DNA hybridization
dex	dexamethasone
DNA	deoxyribonucleic acid
EDTA	ethylene diamine tetraacetic acid
g	gram
GFP	green flurescent protein
GUS	β-glucuronidase
h	hours
HD-ZIPIII	homeodomain leucine zipper class III
HGT	horizontal gene transfer
Hyg	hygromycin
IAA	indole-3-acetic acid
IAM	indole-3-acetamide
ipt	isopentenyl transferase
Kan	kanamycin
KAN	kanadi
KNOX	knotted-like homeobox
kb	kilo base
kDa	kilo dalton
LB	Left border
µ	micro
m	milli
M	molar (mole/liter)
min	minutes
miRNA	microRNA

NAA	naphthaleneacetic acid
NCBI	National Center for Biotechnology Information
NtSIP	*Nicotiana tabacum* 6b interacting protein
nm	nanometer
OD_{600}	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PHAN	phantastica
PHB	phabulosa
PHV	phavoluta
pRi	root inducing plasmid
pTi	tumor inducing plasmid
qPCR	quantitative polymerase chain reaction
RB	right border
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
SAM	shoot apical meristem
SDS	sodium dodecyl sulphate
SEM	scanning electron microscope
TBS	Tris buffer saline
T-DNA	transferred-DNA
TE	Tris-EDTA Buffer
TEMED	N,N,N',N'-tetramethyl-ethylene-1,2-diamine
Tris-HCl	Tris (hydroxymethyl) aminomethane hydrochloride
TTBS	triton-tris buffer saline
WUS	wuschel
YAB	yabby
YEB	yeast extract broth

Introduction

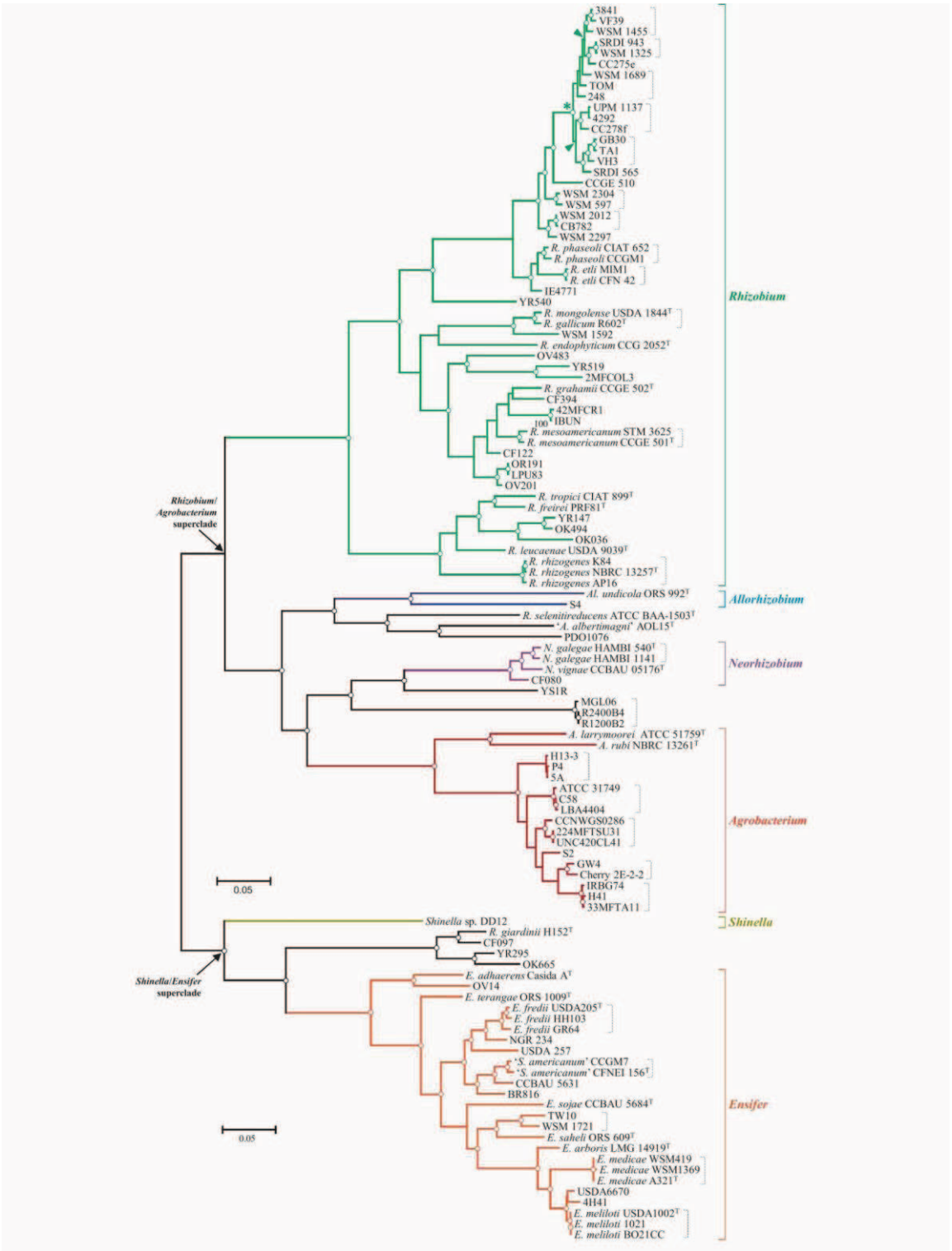


Figure 1. Phylogenetic tree of the Rhizobiaceae family based on sequence analysis
Strains forming species-level clades in different colors based on average nucleotide identity, with a 95% cut-off level. This tree was constructed using a concatenated alignment of the most discriminative amino acid positions of 384 proteins conserved in the chromosomes of all completely sequenced genomes. Abbreviations: R, *Rhizobium*; A, *Agrobacterium*; Al, *Allorhizobium*; S, *Sinorhizobium*; E, *Ensifer*; N, *Neorhizobium*. The scale bar represents the estimated number of amino acid changes per site for a unit of branch length. (Ormeño-Orrillo et al., 2015).

I. Agrobacterium T-DNA transformation

I.1. Agrobacterium

The plant pathogen Agrobacterium was first discovered in 1897 by Fridiano Cavara in Napoli, Italy (Reviewed in Kado et al., 2014). He isolated an unknown bacterium from grapevine crown gall which he named *Bacterium tumefaciens*, the name was later changed to *Phytomonas tumefaciens* and finally to *Agrobacterium tumefaciens*. Agrobacterium is a typical soil bacterium that belongs to the Rhizobiaceae family. Initially, Agrobacteria were classified into three types: *A. tumefaciens*, *A. rhizogenes* and *A. radiobacter* according to their effect on plants: tumorigenic, rhizogenic and non-pathogenic, respectively (Conn, 1942). Later, a biovar (or biotype) classification was used (Kerr and Panagopoulos, 1977) based on chromosomal characteristics. Biovar3 was subsequently named *A. vitis* (Ophel and Kerr, 1990). Apart from these Agrobacterium species, others have been described, like *A. rubi* (Hildebrand, 1940) and *A. larrymoorei* (Bouzart and Jones, 2001). Most Agrobacteria are not host-specific. *A. vitis* is an interesting exception because it has only been found on grapevines (Burr and Otten, 1999).

A new way of classifying Agrobacteria is based on genome sequence analysis. However, care should be used to compare these sequences since some sequences are not stable. During bacterial conjugation, bacteria can gain or lose plasmids. Because of this, plasmid sequences should not be included in phylogenetic studies. In 2001, Young reclassified the Rhizobiaceae family based on 16S rDNA analysis. They proposed to transfer *A. radiobacter*, *A. rhizogenes*, *A. rubi*, *A. undicola* and *A. vitis* to Rhizobium while maintaining *A. tumefaciens*. In 2015, Ormeño-Orrillo classified the Rhizobiaceae family according to *in silico* DNA-DNA hybridization (DDH) and average nucleotide identity (ANI) based on whole genome sequencing (figure 1). The results showed that *A. rhizogenes* and *A. vitis* belong to the Rhizobium and Allorhizobium genus respectively, and should therefore be called *Rhizobium rhizogenes* and *Allorhizobium vitis*. Since the old names *A. rhizogenes* and *A. vitis* are still widely used, I will use them in this thesis.

I.2. Agrobacterium Ti and Ri plasmids

Agrobacteria carry tumor-inducing (Ti) plasmids or root-inducing (Ri) plasmids with T-DNA (transferred-DNA) sequences which are transferred to the plant cell upon infection. However, most natural Agrobacterium isolates lack Ti/Ri plasmids and are non-virulent (Burr and Otten, 1999). Expression of T-DNAs results in a proliferation of transformed and non-transformed cells, thus causing tumors (*A. tumefaciens*, figure 2B) or hairy roots (*A. rhizogenes*, figure 2C) (Binns and Costantino, 1998). The structures of the Ti and Ri plasmids are highly diverse and have been

traditionally classified according to the type of opines found in the transformed tissues. Opines will be discussed in I.4.2.

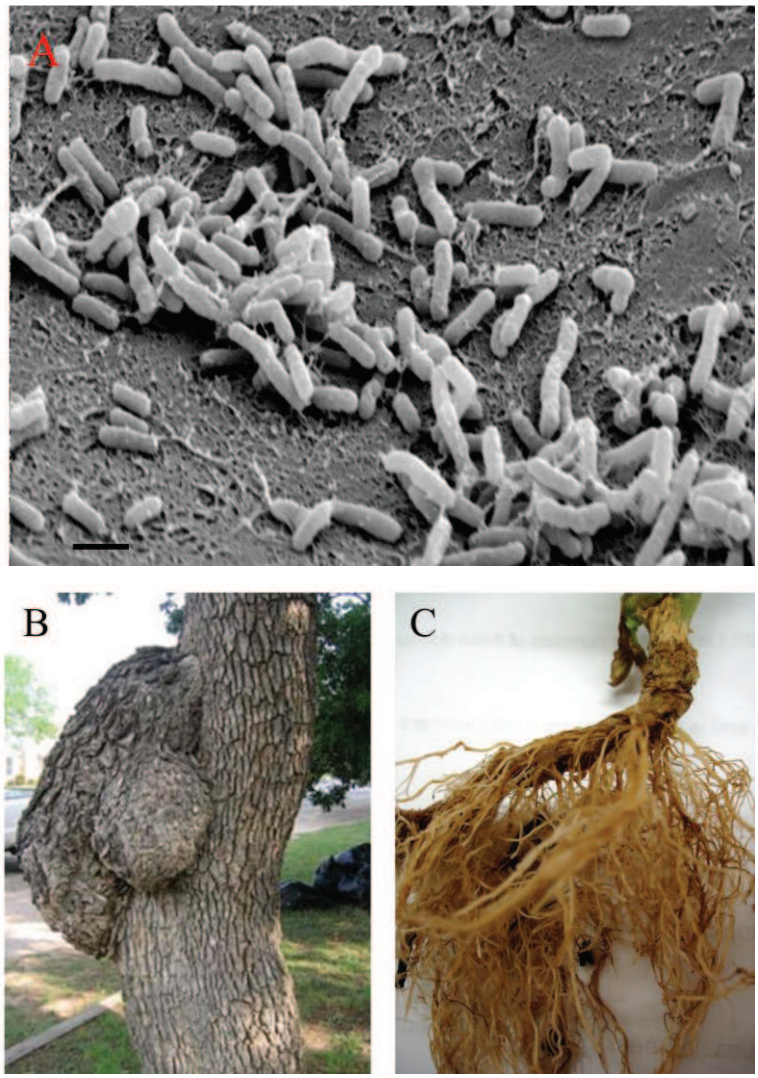


Figure 2. Agrobacterium and associated plant syndrome
(A) Electron microscopy image of *A. tumefaciens* strain C58 (Bar: 1µm) (from http://bacmap.wishartlab.com/organisms/79). (B) Crown gall caused by *A. tumefaciens* (from http://brokenwillow.com/gallery/crown-gall). (C) Hairy roots caused by *A. rhizogenes* (from http://www.cals.ncsu.edu/course/pp728/Rhizobium/Rhizobium_rhizogenes.htm).

A typical Ti or Ri plasmid (figure 3A) includes one or two T-DNAs, a virulence region, an origin of replication, an opine uptake and catabolism region and a conjugative transfer region. On the T-DNA, there are generally genes for the production of indole-3-acetic acid (IAA, an auxin), cytokinins and opines. Genes from the *plast* gene family (see below) are also present on T-DNAs, their functions are not yet clearly defined. After transfer and integration into the plant genome, the T-DNA genes are expressed and lead to local growth. Specific opines will be produced by the transformed plant tissues and can be used by the original bacterial strain which contains specific opine uptake and catabolism genes.

A. tumefaciens Ti plasmid replication and T-DNA transfer involve a cell-to-cell communication system called quorum sensing (QS). A bacterial population can "sense" its cell density by secretion and detection of small signal molecules (Fuqua et al., 1994) and regulate its gene expression accordingly.

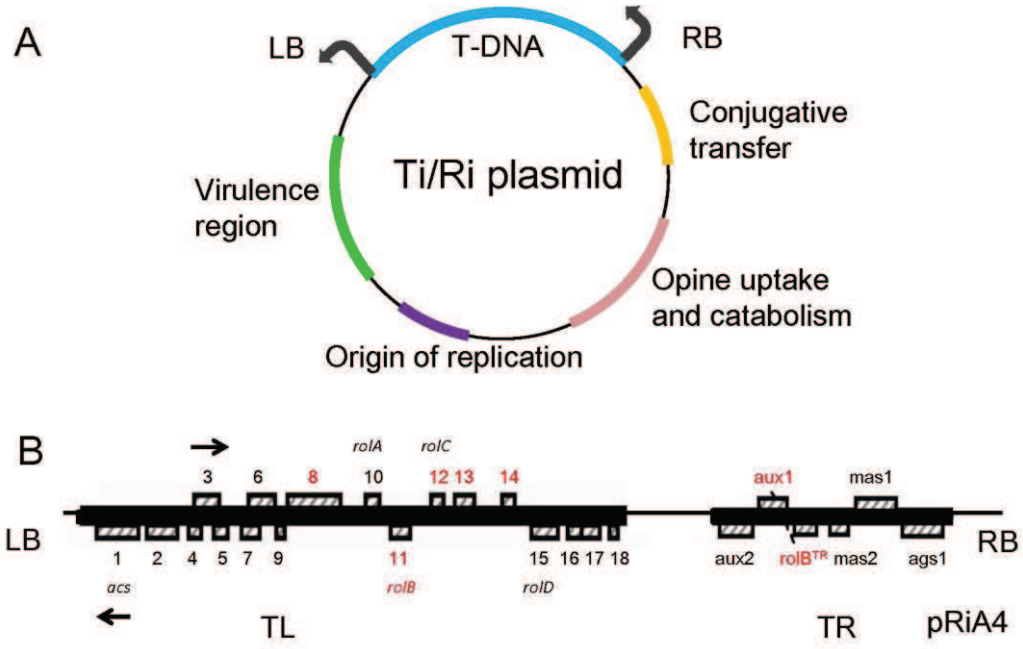


Figure 3. Scheme of Ti/Ri plasmid with an example of the Ri plasmid from *A. rhizogenes* strain A4
(A) A general scheme of a Ti or Ri plasmid from *A. tumefaciens* or *A. rhizogenes*. Arrows indicate the T-DNA transfer direction. (B) Genes present on the two *A. rhizogenes* strain A4 T-DNA regions: TL and TR. Arrows indicate orientation (5' to 3') of ORFs. Genes which belong to the *plast* family are marked in red. LB: left border; RB: right border (adapted from Mohajjel-Shoja 2010, with some modifications).

I.3. Agrobacterium T-DNA transfer mechanism

Agrobacterium T-DNAs can be transferred into plant cells by a special mechanism. This T-DNA transfer system is classified as a bacterial type IV secretion system (figure 4). The ability of Agrobacterium to transfer T-DNAs into plants is widely used in plant genetic engineering and has also been extended to other organisms like animals, fungi and yeast (Kunik et al., 2001; Bulgakov et al., 2006; de Groot et al., 1998; Michielse et al., 2008; Wolterink et al., 2015).

The transformation process consists of different steps.

I.3.1 Attachment of Agrobacterium to plant surfaces

A. tumefaciens binds to many surfaces including those of plants, fungi, soil particles, even polyesters and plastics (Matthysse, 2014). There are two types of binding to these surfaces: unipolar polysaccharide-dependent polar attachment and unipolar polysaccharide-independent

attachment (both polar and lateral). Attachment of *A. tumefaciens* to plants is mediated by the unipolar polysaccharides. Cellulose is another important element for surface attachment. *A. tumefaciens* carries a cellulose synthase gene, *celA*, regulated by *celB*, *celG* and c-di-GMP (Matthysse, 2014). It is known that cellulose fibrils bind tightly to other cellulose fibrils and cellulose synthesis therefore results in a strong attachment of *A. tumefaciens* to cellulose-rich plant surfaces (Matthysse, 2014). VirB on the pTi plasmid may also be involved in this attachment process. Both VirB2 which is part of the shaft of the pilus structure and VirB5 which is located in the pilus tip could help bacterial attachment (Christie et al., 2014).

I.3.2 Virulence gene activation

Most Ti and Ri plasmids carry virulence (*vir*) genes close to the T-DNA region (figure 3A). Their function is to enable T-DNA transfer from Agrobacteria into plant cells. *virA*, *virB*, *virC*, *virD*, *virE*, *virF* and *virG* genes are included in this group. Several “minor” *vir* genes, like *virH* and *virM* have also been described.

The *vir* system can be activated when some components like sugars and phenols reach critical concentrations. Under natural conditions, a wounded plant tissue can lead to an acidic pH and create a high nutrient environment that favors *vir* gene expression (Wolanin et al., 2002). This can be sensed by the VirA protein located on the surface of the bacterium.

VirA protein is a membrane-bound receptor and histidine kinase that can activate VirG by transferring its phosphate to VirG. Phosphorylated VirG is a transcription factor that activates *virB*, *virC*, *virD*, *virE* and *virF* transcription (Brencic and Winans, 2005).

I.3.3 Type IV secretion systems (T4SSs) in *A. tumefaciens*

VirD2 is an endonuclease which cuts a single stranded T-DNA fragment (T-strand) from the Ti/Ri plasmid and covalently links itself with the 5’ end of the T-strand thus forming the T-DNA protein complex (T-complex) (Herrera-Estrella et al., 1988; Ward and Barnes, 1988; Young and Nester, 1988; Durrenberger et al., 1989; Gelvin et al., 2012). The T-complex can be secreted out of Agrobacterium through the T4SS.

The T4SSs are translocation systems that are present in most bacteria and some archaea (Baron, 2005; Alvarez-Martinez et al., 2009; Llosa et al., 2009; Backert and Clyne, 2011; Nagai and Kubori, 2011; Fischer, 2011). These systems are able to translocate DNA and proteins from bacteria to bacteria or to other organisms. Two main groups of T4SSs are the conjugation machines and the effector translocator systems. The first ones enable single strand DNA to translocate within or between bacteria so that antibiotic resistance genes or virulence genes can

disseminate among different bacterial species (Alvarez-Martinez et al., 2009). The effector translocator systems are responsible for transport of proteins or single strand DNA-protein complexes from bacteria into eukaryotic cells. The VirB/VirD4 system that delivers the T-complex from *A. tumefaciens* to plant cells is one of the best well-known effector translocator systems.

A general VirB/VirD4 system contains several Vir proteins which are VirB1 to VirB11, coded by a single *virB* operon and a VirD4 protein coded by a separate *virD* operon (figure 4B). The VirB/VirD4 system consists of different components with different functions (figure 4A). 1. VirD4 is a Type IV coupling protein (T4CP), *i.e.* an ATPase related to the SpoIIIE/FtsK DNA translocases. It binds with DNA and protein substrates and facilitates translocation. 2. VirB3, VirB4, VirB6, VirB8 and VirB11 form the inner membrane complex (IMC), and help the T-complex to transfer across the inner membrane. 3. VirB2, VirB7, VirB8, VirB9 and VirB10 form the envelope-spanning outer membrane complex (OMC) located between inner membrane and periplasm, it helps the T-complex to pass the periplasm and outer membrane. 4. VirB2 and VirB5 form the conjugative pilus, enabling the T-complex to contact with the plant surface (Christie et al., 2014).

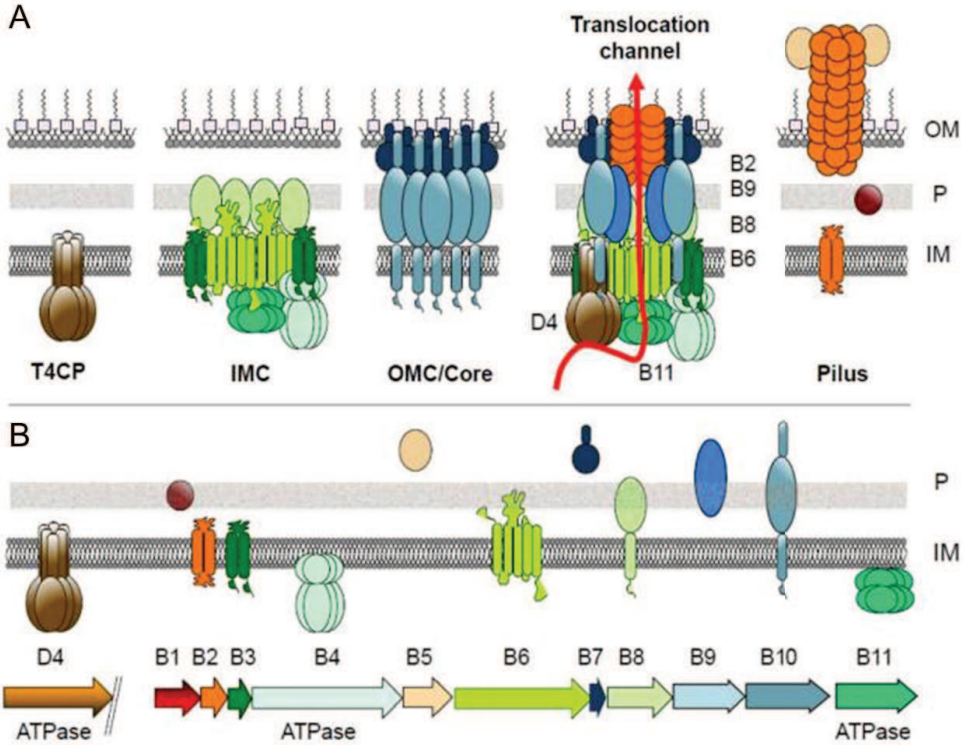


Figure 4. Scheme depicting elements of the *A. tumefaciens* VirB/VirD4 type IV secretion system (T4SS)
(A) Different elements of the T4SS system. T4CP: type IV coupling protein; IMC: inner membrane complex; OMC: outer membrane complex; IM: inner membrane; P: periplasm; OM: outer membrane. (B) Lower: *virD* operon codes for VirD4 protein and *virB* operon codes for 11 VirB proteins. Upper: VirB/VirD4 subunits localized in the inner membrane (IM) or delivered to the periplasm (P). (Christie et al., 2014).

I.3.4 T-complex entry into the plant cell

The mechanism of T-complex transfer through the plant cell membrane is still unknown. In the plant cell, the T-DNA was found as a T-complex bound to VirD2 with the full length of DNA covered with VirE2. One proposal for T-complex transfer is that VirE2 may sit in the plant plasma membrane, forming a channel through which the T-DNA can pass (Duckely and Hohn, 2003).

Once the T-complex enters the plant cell, it needs to travel to the nucleus. VirD2 plays a very important role in this process since the C-terminal of VirD2 contains a nuclear localization signal sequence which enables proteins to target the eucaryotic nucleus (Herrera-Estrella et al., 1990; Howard et al., 1992; Tinland et al., 1992b; Rossi et al., 1993; Citovsky et al., 1994; Mysore et al., 1998). The cytoskeletal structures and the molecular motors of the plant cell may also help T-complex transport (Gelvin, 2012). Some plant proteins like importin α, vir-interacting protein 1 (Vip1) and vir-interacting protein 2 (Vip2) also play a role in helping T-complex transfer (figure 5) (Gelvin, 2012). In the plant cytoplasm, Vip1 directly binds with VirE2 (figure 5). The plant defense signaling mitogen-activated protein kinase is able to activate the phosphorylation of Vip1 allowing phospho-Vip1-T-complex import into the nucleus (Magori and Citovsky, 2012).

In the plant nucleus, once the T-complex is close to the chromosome, it is necessary to uncoat the T-complex. VirF has been demonstrated to interact with plant homologs of yeast Skp1 protein (Schrammeijer et al., 2001). The loss of *virF* reduces virulence of *A. tumefaciens* and can be complemented by expression in plant cells showing that VirF has a function in the plant cell (Otten et al. 1985; Schrammeijer et al., 2001). Later studies showed that VirF is an F-box protein which helps to release single strand T-DNA from the T-complex (Magori and Citovsky, 2012). VirF can be stabilized by VirD5. The Vip1-VirE2-VirD2 protein complex can be polyubiqitinated by VirF thus leading to its degradation. After this process, single strand T-DNA is naked and ready to be inserted into the plant genome.

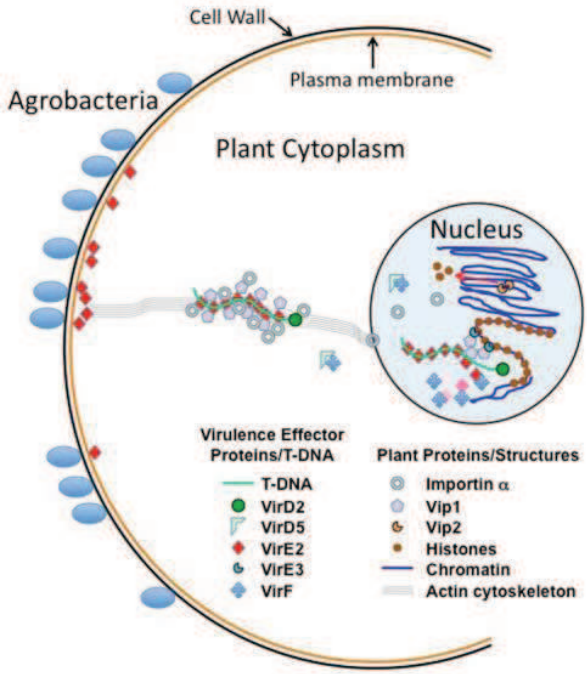


Figure 5. T-complex transfer from Agrobacterium into the plant cell
Virulence proteins from Agrobacteria and plant proteins are both involved in T-complex translocation. The result of this process is the integration of T-DNA into the plant genome (Gelvin, 2012).

I.4. Agrobacterium T-DNA genes

The T-DNA located genes can be divided into different groups: hormone synthesis genes, opine synthesis genes, *rol* genes and *plast* genes. The *rol* genes (*rol* for “root locus”) were initially identified by their effects on hairy root induction (White et al., 1985). The *rolA, rolB* and *rolC* are necessary and sufficient for the induction of hairy roots by *A. rhizogenes*. *rolD* gene encodes an ornithine cyclodeaminase (OCD) enzyme which catalyzes the conversion of ornithine to proline.

I.4.1 Hormone synthesis genes

Several Agrobacterium T-DNAs contain hormone synthesis genes like *iaaM*, *iaaH* and *ipt*. *iaaM* and *iaaH* code for two steps in auxin synthesis. First, the amino acid tryptophan (Trp) is converted to indole-3-acetamide (IAM) by tryptophan-2-monooxygenase (IaaM, coded by *iaaM*), subsequently, indole-3-acetamide hydrolase (IaaH, coded by *iaaH*), converts indole-3-acetamide into indole-3-acetic acid (IAA) (figure 6A).

Cytokinin is formed by isopentenyltransferase (Ipt, coded by *ipt*) which adds a prenyl moiety from dimethylallyl pyrophosphate (DMAPP) or hydroxymethylbutenyl pyrophosphate (HMBPP) to ATP/ADP/AMP to produce isoprene cytokinins (figure 6B).

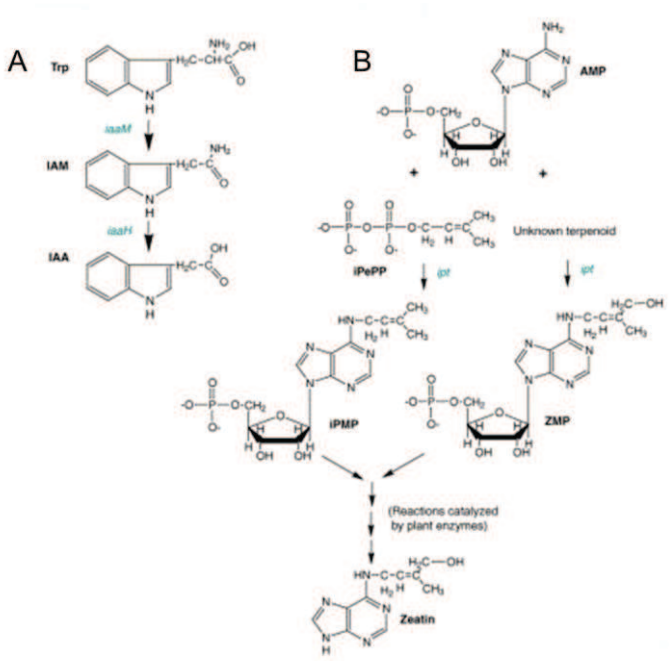


Figure 6. *iaaM*, *iaaH* and *ipt* enzymatic functions
(A) Trp is converted to IAM by tryptophan-2-monooxygenase coded by *iaaM*; IAM is converted to IAA by indole-3-acetamide hydrolase coded by *iaaH.* (B) A prenyl group is added to AMP in the presence of isoprenyl transferase coded by *ipt* to form the cytokinin zeatin.

I.4.2 Opine synthesis genes and opines

Opines are produced by Agrobacterium-infected plants and are found in or around plant tumors or hairy roots. They are low molecular weight molecules which can be separated into two groups: agrocinopines and secondary amine derivatives (Baek et al., 2003). Agrocinopines are sugar-phosphodiesters and can be used by Agrobacterium as carbon and phosphorus sources. The amine derivatives are synthesized by condensation of a sugar or a keto acid with an amino acid. Thus they are a source of carbon and nitrogen. Over 20 types of opines have been characterized (Dessaux et al., 1998). The chemical structures of the major opines are shown in figure 7.

Agrobacterium Ti or Ri plasmids can be classified based on the opines they induce (table 1). Two examples can be cited. On the octopine type plasmid pTiAch5, *ocs* codes for the octopine synthase enzyme Ocs that leads to octopine, octopinic acid, histopine and lysopine synthesis in tumors (Otten et al., 1977; Dessaux et al., 1998; Flores-Mireles et al., 2012). On the agropine type Ri plasmid pRiA4, *mas2'* codes for mannopine synthase 2' (Mas2') that converts glutamine and sugar into DFG. DFG can be reduced to mannopine by the *mas1'* encoded mannopine synthase 1' (Mas1'). Finally, mannopine is cyclized to agropine by agropine cyclase (Ags) encoded by *ags.*

As described before, RolD encodes an ornithine cyclodeaminase which converts ornithine to proline and may therefore be considered as an opine synthesis gene if the proline is used by

Agrobacterium. However, this has not been shown sofar. Unlike the other *rol* genes, *rolD* was only discovered in the agropine Ri plasmid pRiA4 (Christey, 2001). The neighboring *orf17* and *orf18* genes represent two presumably inactive halves of a *rolD* repeat. *rolD*-expressing tobacco plants show inhibition of root formation and early flowering (Trovato et al., 2001).

Interestingly, some opine synthesis genes are not only found in Agrobacterium but also in plants and fungi. Only a few of these genes are intact. These will be described in Chapter III.

Table 1. Different Ti or Ri plasmids based on opines they produce (adapted from Dessaux et al., 1998)

Plasmid type	Relevant opine products
Ti Plasmids	
Octopine	Octopine, octopinic acid, lysopine, histopine, agropine, mannopine, agropinic and mannopinic acid
Nopaline	Nopaline, nopalinic acid, agrocinopine A and B
Agropine	Agropine, mannopine, agropinic and mannopinic acid, agrocinopine C and D, leucinopine, leucinopine lactam, L,L succinamopine
Succinamopine	D,L succinamopine, succinamopine lactam, succinopine
Lippia	Agrocinopine C and D
Chrysopine/succinamopine	Chrysopine, deoxyfructosyl-5-oxoproline (dfop), L,L succinamopine, L,L leucinopine
Chrysopine/nopaline	Chrysopine, deoxyfructosyl-5-oxoproline (dfop), nopaline
Octopine/cucumopine	Octopine, cucumopine
Vitopine	Vitopine
Ri Plasmids	
Agropine	Agropine, mannopine, agropinic and mannopinic acid
Mannopine	Mannopine, agropinic and mannopinic acid
Cucumopine	Cucumopine, cucumopine lactam
Mikimopine	Mikimopine, mikimopine lactam

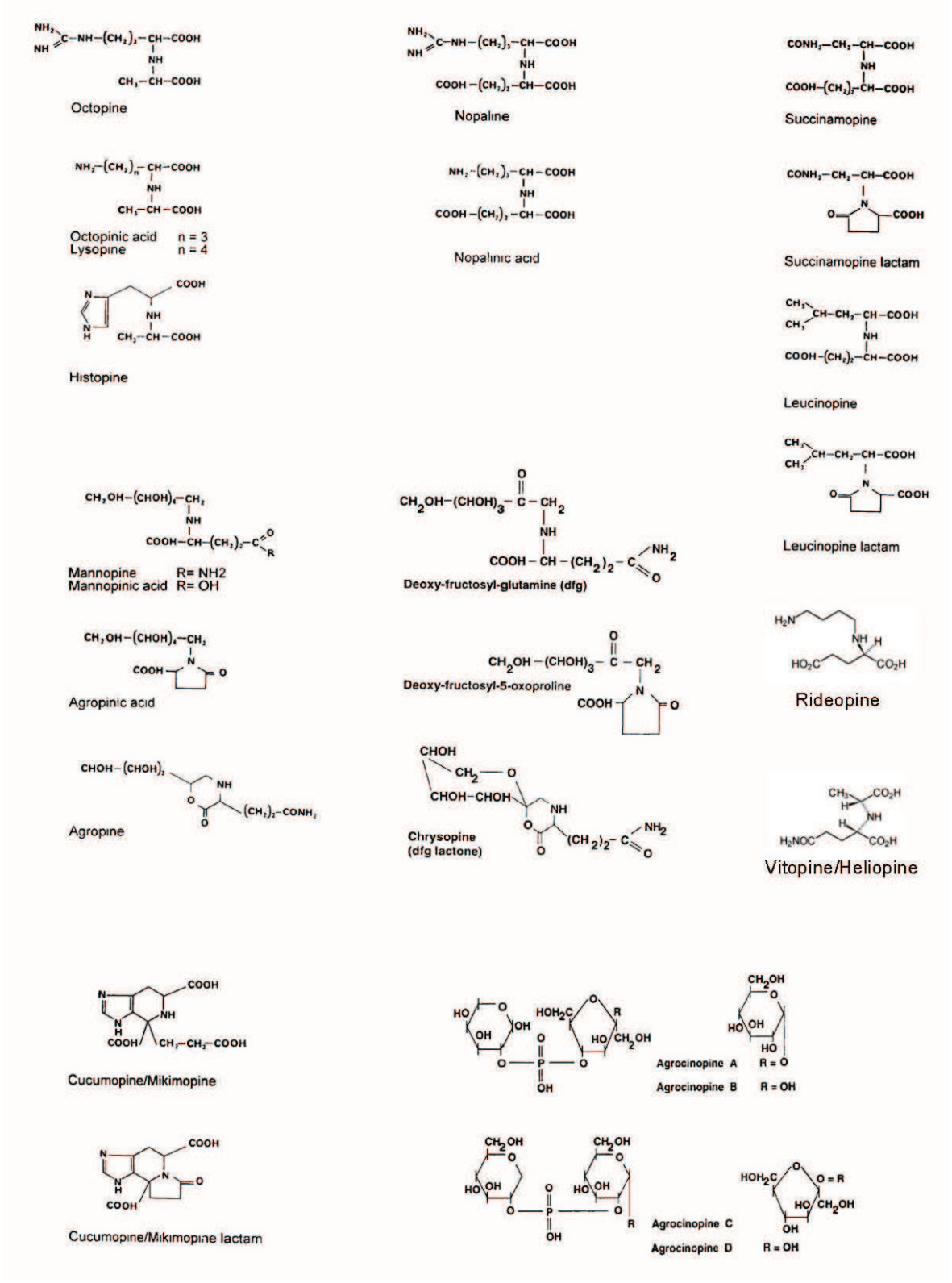


Figure 7. Chemical structures of opines produced by crown galls and hairy roots
Chemical structures of octopine, octopinic acid, lysopine, histopine, nopaline, nopalinic acid, succinamopine, succinamopine lactam, leucinopine, leucinopine lactam, mannopine, mannopinic acid, deoxyfructosyl-glutamine (DFG), deoxyfructosyl-5-oxoproline, agropinic acid, agropine, chrysopine, cucumopine/mikimopine, cucumoplne/miklmopine lactam, agrocinopine A, B, C and D, rideopine, vitopine/heliopine (adapted from Dessaux et al., 1998).

I.4.3 *plast* genes

In the first paper mentioning the *plast* family (Levesque et al., 1988) 11 highly diverged plast proteins were identified. They have sizes of 250-280 amino acids and are encoded by the following genes: 5' end of *orf8* and *iaaM*, *orf11* (*rolB*), *orf12* (*rolC*), *orf13* and *orf14* (on the TL-DNA from *A. rhizogenes* strain A4 strain), gene *5*, *7*, 5' end of *iaaM*, *6a*, *6b*, *orf21* (*3'*) (from the TL and TR-DNA from the *A. tumefaciens* 15955 octopine strain). Later, other T-DNA genes could be added: *rolBTR* (a *rolB* variant on the TR-DNA of *A. rhizogenes* A4, Bouchez and Camilleri, 1990), *e* (from the *A. tumefaciens* C58 nopaline strain, Broer et al., 1995), and *lso* from the *A. tumefaciens* AB2/73 strain (Lippia strain) (Otten and Schmidt, 1998). Although the blastp search method could not confirm inclusion in the *plast* family of gene *7*, *rolC* and the 5' part of *iaaM*, the reiterative psi-blast method did (Otten et al., 1999). These studies also showed that several genes from the largely unexplored left part of the C58 T-DNA were part of the *plast* family: genes *b* (a *5*-like gene), *c'*, *d* (a *3'*-like gene) and *e*. In the same study, weak homology between gene *7* and *orf18* of the *A. tumefaciens* 15955 TR-DNA (also called gene *4'*) was reported. Since 1999, additional plast genes were reported, most being similar to already described ones. The fact that a broad range of Agrobacterium strains had already been investigated suggested that the majority of Agrobacterium *plast* genes might have been identified. In 2005, the *plast* gene family (called the *RolB/RolC* glucosidase family in data banks) was considered to be specific for Agrobacterium species and plants (Studholme et al., 2005). However, *plast* genes from other organisms were reported subsequently. These included genes from *Laccaria bicolor* (Mohajjel-Shoja, 2010) and *Rhizobium mesoamericanum* (Chen et al., 2014). More recent searches (unpublished, this thesis) identified *plast* genes in *Bradyrhizobium* sp., *Mesorhizobium plurifarium*, *Rhizobium leguminosarum*, *Burkholderia* sp., *Pisolithus microcarpus* and *Laccaria amethystina*.

An alignment of the central parts of 50 plast proteins (Helfer et al., 2002) showed only very few conserved residues and two subgroups containing, on the one hand, the proteins encoded by *orf14*, *rolC*, *orf13*, *6a*, *6b*, and on the other hand those encoded by the 5' parts of *iaaM* and *orf8*, *orf18*, *7*, *rolB*, *rolBTR*, *lso*, *c'*, *e*, *5*, *b*, *d* and *3'*. The latest trees (figure 8) confirm the group with *orf14*, *rolC*, *6a* and *6b*, but three (Clustal Omega, MAFFT and MUSCLE) out of four trees do not include *orf13* in this group.

Phylogenetic trees of plast proteins are shown in figure 8. In the following section I will present results obtained for different *plast* genes.

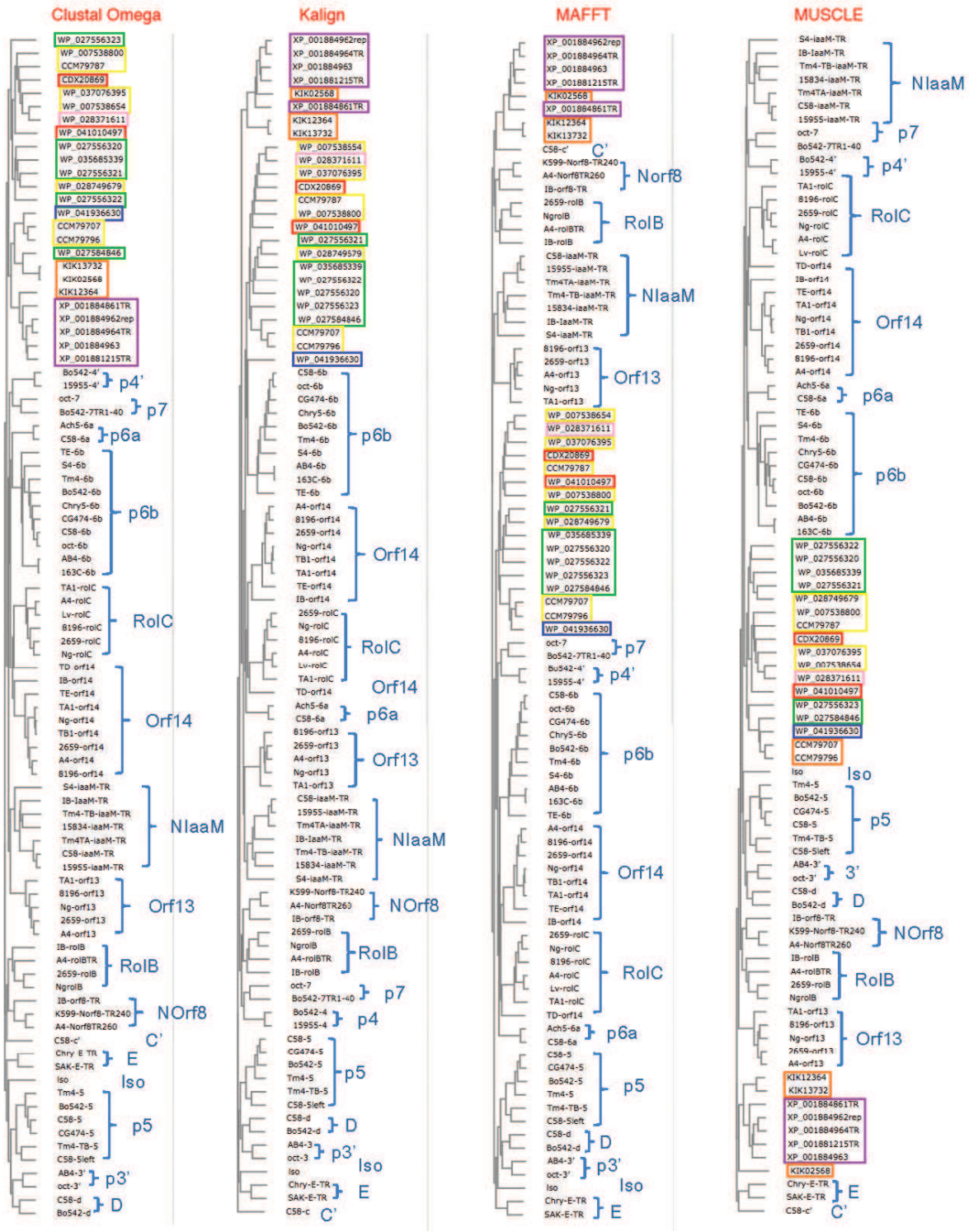


Figure 8. Phylogenetic trees of plast proteins
Proteins were analyzed using four different multiple sequence aligments (MSA): Clustal Omega (suitable for medium to large fragments), Kalign (suitable for large fragments), MAFFT (suitable for medium to large fragments) and MUSCLE (suitable for medium fragments). Boxed sequences are from organisms other than Agrobacterium or from plants naturally transformed by Agrobacterium. The plants are indicated by the following prefixes: t: *Nicotiana tabacum*, Ng: *Nicotiana glauca*, Lv: *Linaria vulgaris*, IB: *Ipomoea batatas*. The codes of the Agrobacterium and plant *plast* genes are given in table 2. Green box: *Bradyrhizobium*; yellow box: *Rhizobium mesoamericanum*; red box: *Mesorhizobium plurifarium*; pink box: *Burkholderia*; blue box: *Rhizobium leguminosarum*; orange box: *Pisolithus microcarpus*; purple box: *Laccaria bicolor*.

Table 2. Accession numbers corresponding to plast proteins in figure 8
Ar: *A. rhizogenes*; At: *A. tumefaciens*; Av: *A.vitis*; Ng: *N. glauca*; Nt: *N. tabacum*; Lv: *Linaria vulgaris*.

	Plast protein	Accession number	Origin		Plast protein	Accession number	Origin
1	C58-5	AAD30487.1	At	34	oct-6b	AAF77126.1	At
2	SAK-5	BAA87806.1	At	35	Chry5-6b	AAB49454.1	At
3	CG474-5	AAB41867.1	Av	36	AB4-6b	CAA54541.1	Av
4	Tm4-5	AAB41873.1	Av	37	Bo542-6b	AAA98501.1	At
5	Bo542-5	AAZ50393.1	At	38	163C-6b	ADC97873.1	At
6	Tm4-TB-b	AAD30490.1	Av	39	S4-6b	AAA25043.1	Av
7	C58-b	AAD30482.1	At	40	A4-orf14	ABI54193.1	Ar
8	Iso	AAC25913.1	At	41	8196-orf14	AAA22099.1	Ar
9	Chry-E-TR	AAK08598.1	At	42	1724-orf14	BAA22339.1	Ar
10	SAK-E-TR	BAA87804.1	At	43	Ng-orf14	BAB85948.1	Ng
11	K599-Norf8-TR240	ABS11822.1	Ar	44	2659-orf14	CAB65899.1	Ar
12	A4-Norf8TR260	ABI54188.1	Ar	45	torf14	CBJ56561.1	At
13	K599-rolB	ABS11824.1	Ar	46	Ngorf14	BAA03991.1	Ng
14	2659-rolB	CAA82552.1	Ar	47	1724-rolC	P49408.1	Ar
15	1724-rolB	CAA45540.1	Ar	48	2659-rolC	CAA82553.1	Ar
16	A4-rolBTR	CAA34077.1	Ar	49	A4-rolC	P20403.1	Ar
17	NgrolB	CAA27161.1	Ng	50	Lv-rolC	ACD81987.1	Lv
18	C58-c’	AAD30484.1	At	51	Ng-rolC	P07051.2	Ng
19	C58-d	AAD30485.1	At	52	8196-rolC	AAA22096.1	Ar
20	Bo542-d	AAZ50418.1	At	53	trolC	CAA62988.1	Nt
21	AB4-3’	CAA54542.1	Av	54	Ach5-6a	P04030.1	At
22	696-3’	CAA52222.1	At	55	C58-6a	AAK90971.1	At
23	oct-3’	CAA25183.1	At	56	oct-7	AAF77121.1	At
24	Ng-orf13R	BAB85946.1	Ng	57	Bo542-7	AAZ50396.1	At
25	8196-orf13	AAA22097.1	Ar	58	15955-4’	CAA25180.1	At
26	2659-orf13	CAB65897.1	Ar	59	Bo542-4’	AAZ50416.1	At
27	A4-orf13	ABI54192.1	Ar	60	C58-iaaM	CAB44640.1	At
28	Ng-orf13	BAA03990.1	Ng	61	15955-iaaM	CAA25167.1	At
29	1724-orf13	BAA22337.1	Ar	62	Tm4TA-iaaM	P25017.1	Av
30	torf13-1	CAA07584.1	Nt	63	Tm4-TB-iaaM	AAD30493.1	Av
31	C58-6b	AAK90972.1	At	64	Ag162-iaaM	AAC77909.1	Av
32	CG474-6b	AAB41871.1	Av	65	15834-iaaM	ABI15642.1	Ar
33	Tm4-6b	CAA39648.1	Av	66	S4-iaaM	AAA98149.1	Av

The morphological effects and possible functions of the different *plast* genes are the following.

I.4.3.1 *rolB*

rolB (*orf11*) genes code for proteins ranging from 28 to 31 kDa in different strains with about 60% similarity among each other (Mohajjel-Shoja, 2010). The *rolB* gene induces roots in leaf explants of many plant species (Cardarelli et al., 1987; Spena et al., 1987; Capone et al., 1989). In transgenic plants it leads to necrosis (Schmülling et al., 1988), *rolB*-expressing roots grow faster than normal roots (Altabella et al., 1995). The *rolB* gene also modifies flower induction from thin cell layers in tobacco (Altamura et al., 1994). Apparently, the *rolB* effects are not graft-transmissible (Hansen et al., 1993). The *rolB* gene is induced by auxin (Maurel et al., 1990; Capone et al., 1991; Capone et al., 1994) and sucrose (Nilsson and Olsson, 1997) and expressed in phloem parenchyma (Altamura et al., 1991). It becomes active at the end of the

globular stage of tobacco zygotic embryos (Chichiricco et al., 1992; Di Cola et al., 1997), and in root pericycle cells during outgrowth of lateral roots (Nilsson et al., 1997).

Further promoter studies identified a promoter region (ACTTTA) that binds the DOF transcription factor NtBBF1 in tobacco (Baumann et al., 1999). Initially it was reported that RolB increases auxin sensitivity in excised organs and protoplasts of tobacco (Cardarelli et al., 1987; Shen et al., 1988, 1989; Barbier-Brygoo et al., 1992; Maurel et al., 1991; Schmülling et al., 1993). In 1991, it was proposed that RolB has glucosidase activity acting on indoxyl-beta-glucoside suggesting it could liberate auxin from conjugated auxin forms (Estruch et al., 1991b), but this has been questioned by others (Nilsson et al., 1993; Schmülling et al., 1993; Delbarre et al., 1994). In 1996 it was proposed that RolB has tyrosine phosphatase activity and is localized in the plasma membrane (Filippini et al., 1996). This was partly based on the presence of a CX5R motif. However, mutation of this motif did not abolish the necrotic effect in leaf patches (Mohajjel-Shoja, 2010). A *rolB*-like gene, *rolBTR*, was found on the TR-DNA of A4, and found to differ from *rolB* by the lack of the CX5R motif and its lack of root induction, transgenic plants show a different phenotype as *rolB* plants (Lemcke and Schmülling, 1998). An extensive RolB study in 2004 (Moriuchi et al., 2004) found an interaction between RolB from *A. rhizogenes* strain 1724 and the tobacco Nt-14-3-3-omega II protein, and localized RolB in the nucleus.

A *rolB*-like gene in *N. glauca* (*NgrolB*) was found to be mutated, but upon re-activation induced several abnormalities that differed from those of *RirolB* genes from *A. rhizogenes* (Aoki and Syono, 1999; Aoki, 2004).

I.4.3.2 *rolC*

rolC genes (also called *orf12*) code for proteins of about 20 kDa with more than 65% similarity among strains (Mohajjel-Shoja, 2010). *rolC* is required for hairy root induction. 35S-*rolC* tobacco and potato plants show dwarfism, reduction of apical dominance, pale green leaves, and male-sterile flowers (Spena et al., 1989; Schmülling et al., 1988; Fladung et al., 1990) and stem fasciation (Nilsson et al., 1996). The *rolC* effect is cell-autonomous (Spena et al., 1989). *rolC* is expressed in phloem cells and induced by sucrose (Yokoyama et al., 1994; Nilsson et al., 1997). It was reported that *rolC* releases cytokinin from conjugates by a glucosidase activity (Estruch et al., 1991a), is located in the cytoplasm (Estruch et al., 1991b) and alters cytokinin and gibberellin metabolism (Nilsson et al., 1993), but the glucosidase activity was strongly questioned in 1996 (Nilsson et al., 1996). Because of the *rolC*-induced phenotype and its expression in the phloem, it has been proposed that *rolC* could create a sink for sucrose (Nilsson and Olsson, 1997, overview). The gene has been expressed in many plant species and may be used for modification of flowers (Casanova et al., 2005). An active *rolC* gene was found in *N. glauca* (Aoki

and Syono, 1999) where it is expressed in the vascular system (Nagata et al., 1996) and in *N. tabacum* (Meyer et al., 1995; Mohajjel-Shoja, 2010).

I.4.3.3 *orf13*

orf13 genes were found on T-DNAs from *A. rhizogenes* and in some plants like *N. glauca, N. tabacum* and the ancestor of tobacco *N. tomentosiformis* (Fründt et al., 1998; Chen et al., 2014). DNA sequence homology among these *orf13* genes is about 70-90% (Fründt et al., 1998). The *orf13* gene does not belong to the *rol* genes involved in hairy root induction but seems to enhance rooting by *rolA*, *B* and *C* (Capone et al., 1989) and the necrotic *rolB* response (Aoki and Syono, 1999a). Plants expressing a 35S-*orf13* gene are dwarfs, have irregular leaves, and curly roots (Lemcke and Schmülling, 1998). The phenotypic effects of the 8196-*orf13* gene are graft-transmissible (Hansen et al., 1993). In tomato *orf13* induced stunted leaves with "spikes" on the abaxial side, changes in phyllotaxis and fasciation, and increased cell division in the SAM (Stieger et al., 2004). The 8196-*orf13* expression is high in roots and induced by wounding (Hansen et al., 1997), A4-*orf13* root expression seems lower (Udagawa et al., 2004) and can be induced by wounding or ageing. It has been proposed that 8196 Orf13 binds to the Retinoblastoma protein with an LxCxE motif. This motif is not found in other plast proteins. An *orf13* gene was found in *N. glauca* and a 35S-*Ngorf13* gene leads to round and dark-green leaves in tobacco (Aoki and Syono, 1999a). An *N. tabacum torf13* gene was also shown to be active, it causes cell proliferation on carrot disks (Fründt et al., 1998).

I.4.3.4 *orf14*

The *orf14* gene of *A. rhizogenes* has also been reported to enhance *rolB*-induced rooting (Capone et al., 1989). A 35S-A4-*orf14* construct did not modify the normal tobacco phenotype (Aoki and Syono 1999a). The *orf14* gene was also found in *N. glauca* and weakly promotes root induction by *rolB* as does the *A. rhizogenes orf14* gene (Aoki and Syono, 1999). *N. tabacum* contains several *orf14* genes (see Chapter I), their possible activities remain to be tested.

I.4.3.5 Gene *6a*

Gene *6a* is part of the *A. tumefaciens* T-DNAs. The octopine and nopaline-type *6a* genes were found to play a role in opine secretion (Messens et al., 1985) and were therefore named *ons*, for octopine/nopaline secretion. Unfortunately, this very interesting result was not followed up by further research on the precise mechanism of transport.

I.4.3.6 Gene *6b*

The *6b* gene has been found on T-DNAs from *A. tumefaciens* and *A. vitis*. This gene causes tumors on certain plant species by an unknown mechanism (Hooykaas et al., 1988). It was recently also discovered in *N. otophora* as part of the TE cT-DNA (Chen et al., 2014). The *6b* gene causes very strong morphological changes when overexpressed in plants. An important part of my thesis concerns the *plast* gene *6b*. An overview of the studies concerning this gene will therefore be presented in Chapter II.

I.4.3.7 Gene *5*

Gene *5* is found on T-DNAs from several *A. tumefaciens* strains like C58 (nopaline strain) and Ach5 (octopine strain). The Ach5 *5* gene was reported to direct synthesis of indole-3-lactate (Körber et al., 1991). Unfortunately, no further research has been done with this gene. It does not lead to a phenotype when expressed under 35S promoter control in tobacco (Otten and Schmidt, 1998).

I.4.3.8 Gene *e*

The protein encoded by gene *e* has high similarity to RolB from *A. rhizogenes* and to p5 from *A. tumefaciens* (Broer et al., 1995). The same authors reported that *e* was involved in tumor formation but this was later questioned since a test of gene *e* on *N. glauca*, *N. rustica* and *N. tabacum* failed to induce tumors (Otten et al., 1999).

I.4.3.9 *lso*

The *lso* gene (for Lippia strain oncogene) was initially found in *A. tumefaciens* strain AB2/73 (Otten and Schmidt, 1998). It is able to induce small tumors on different plant species. Plants over-expressing *lso* have a dwarf phenotype with wrinkled leaves and reduced root growth, the phenotype is not graft-transmissible (Schmidt, 1999).

I.4.3.10 5’ end of *iaaM* and *orf8* genes

The *iaaM* and *orf8* genes have been found in *A. tumefaciens*, *A. rhizogenes* and *A. vitis*. The *iaaM* gene codes for the synthesis of indoleacetamide (IAM) from tryptophan, this activity is well documented (Van Onkelen et al., 1985; Umber et al., 2005). The *iaaM* gene is associated with an *iaaH* gene that codes for the conversion of IAM into IAA. The *orf8* gene is highly similar to the *iaaM* gene, but is not associated with an *iaaH* gene, and does not code for IAM production. The original *plast* paper (Levesque et al., 1988) noticed that the N-terminal part of IaaM (about 250 amino acids) is part of the *plast* family and resembles RolB. *iaaM* genes in other organisms, like Pseudomonas or

Rhizobium, do not contain this *rolB*-like extension (Otten and Helfer, 2001). Expression of the RolB part of A4-Orf8 under 35S promoter control leads to strong accumulation of hexose sugars and starch in tobacco, whereas the equivalent region from Ach5-IaaM has no such activity (Otten and Helfer, 2001; Umber et al., 2002, 2005). 35S-A4-*Norf8* plants retain sucrose in the source leaves and convert it to massive amounts of starch.

I.4.3.11 Gene *3'*

Gene *3'* from *A. tumefaciens* octopine strain Ach5 induces very small tumors on *K. tubiflora* (Otten et al., 1999). In the *A. vitis* strain AB4 (Otten and De Ruffray, 1994) and in *A. tumefaciens* strain 82.139 (Drevet et al., 1994) gene *3'* replaces gene *6a* between the *6b* and *nos* genes. No transgenic plants expressing the *3'* gene have yet been regenerated.

I.4.3.12 Other *plast* genes

Less well-studied *plast* genes include: gene *4'* (from the TR-region of *A. tumefaciens* octopine strains), *7* (from the TL-region of *A. tumefaciens* octopine strains), *b* (found in *A. tumefaciens* C58 and on the TB-region of the *A. vitis* octopine/cucumopine strain Tm4), *c'*, *d* and *e*, the latter three from *A. tumefaciens* nopaline strain C58 (Otten et al., 1999). They do not seem to have a notable role in tumor induction, but it will be interesting to study their capacity to modify plant growth by over-expression in different host plants.

I.4.4 Genes unrelated to the first three groups

I.4.4.1 *rolA*

rolA encodes a small protein of about 11 kDa (Nilsson and Olsson, 1997). *rolA* transgenic tobacco plants have wrinkled leaves, show inhibition of plant growth and abnormal flowers (Trovato and Linhares, 1999). The phenotype of *rolA* is graft transmissible from both rootstocks and scions to the untransformed plant part (Guivarch et al., 1996). *rolA* transgenic *Arabidopsis thaliana* has much lower levels of phytohormones (auxin, cytokinin, gibberellic acid and abscisic acid) compared to WT plants (Dehio et al., 1993). The mRNA of *rolA* has an untranslated intron region. In the 5' end of *rolA*, this transcribed but untranslated intron region is important for *rolA* function since the splicing of this part abolishes the *rolA* phenotype in Arabidopsis (Magrelli et al., 1994). This intron was later found to be a prokaryotic promoter which allows *rolA* expression in *A. rhizogenes* (Pandolfini et al., 2000).

I.4.4.2 Gene *c*

Gene *c* is found on the *A.tumefaciens* pTiC58 T-DNA and *A. vitis* pTiTm4 TB-DNA and code for a 523 amino acid protein (Otten et al., 1999). Recently, homologs have been found in other organisms (Chen et al., 2014) but the function remains unknown.

I.4.4.3 Gene *3*

Gene *3* (also called *ORF3n*) is located on the left part of the T-DNA from *A. rhizogenes* strain HRI (Lemcke and Schmülling, 1998). A 35S::*3* transgenic tobacco is clearly modified. The stem of this transgenic tobacco is about 20-30% shorter than WT tobacco. The leaf tip, base and sepals all have various levels of necrosis. The flowering time of these plants is about 2-3 weeks later compared to the WT plant (Lemcke and Schmülling, 1998).

I.5. Stable horizontal gene transfer

Horizontal gene transfer (HGT) is the abnormal transfer of genes between different species as opposed to vertical transfer, i.e. normal gene transfer to the next generation within the same species. HGT has been shown to play an important role in the evolution of many organisms and occurs between bacteriophages and bacteria, among different types of bacteria, from bacteria to eukaryotes like fungi or plants and within eukaryotes from the chloroplast and mitochondrial genomes to the nuclear genomes.

I.5.1 HGT from Agrobacteria to plants

About 30 years ago, sequences with homology to the T-DNA of the *A. rhizogenes* pRi plasmid were found in the genome of *Nicotiana glauca*; these were named "cellular T-DNA" (cT-DNA) sequences (White et al., 1983, Furner et al., 1986). Southern blot analysis showed that the *N. glauca* cT-DNA consisted of an imperfect repeat. Initially, only part of this cT-DNA was sequenced, to be completed several years later (Furner et al., 1986; Aoki et al., 1994; Suzuki et al., 2002). The *N. glauca* cT-DNA contains the *NgrolB, NgrolC, NgORF13, NgORF14* and *Ngmis* genes (figure 9) (Suzuki et al., 2002). *NgrolB, NgrolC, NgORF13* and *NgORF14* were found to be expressed in genetic tumors formed by hybrids between *N. glauca* and *N. langsdorffii* (Aoki et al., 1994). Later, the *Ngmis* gene was found to encode an enzymatically active mikimopine synthase when expressed in *E. coli*, but no mikimopine was found in *N. glauca* plants (Suzuki et al., 2002).

Additional studies showed that other Nicotiana species also contained cT-DNA sequences. These included cT-DNA fragments from *N. tomentosa, N. kawakamii, N. tomentosiformis* and *N. otophora* (figure 10). Partial cT-DNA sequences were also reported from *N. tabacum* (Meyer et al., 1995, summary in Chapter I).

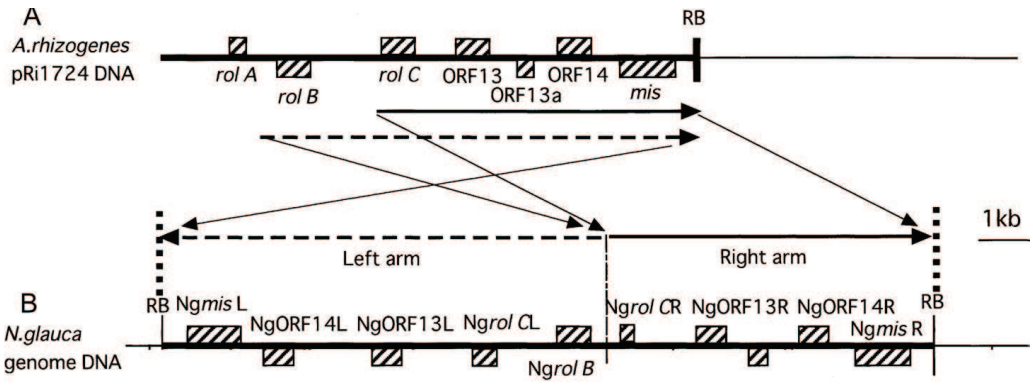


Figure 9. Map of cT-DNA from *N. glauca* compared with the *A. rhizogenes* pRi1724 TL-DNA
(A) Genes located on right part of the *A.rhizogenes* pRi1724 T-DNA and its right border (RB). The arrows indicate regions with homology to cT-DNA in *N. glauca.* (B) cT-DNA in *N. glauca* with indications of open reading frames and imperfect inverted repeats (left and right arm). (Adapted from Suzuki et al., 2002).

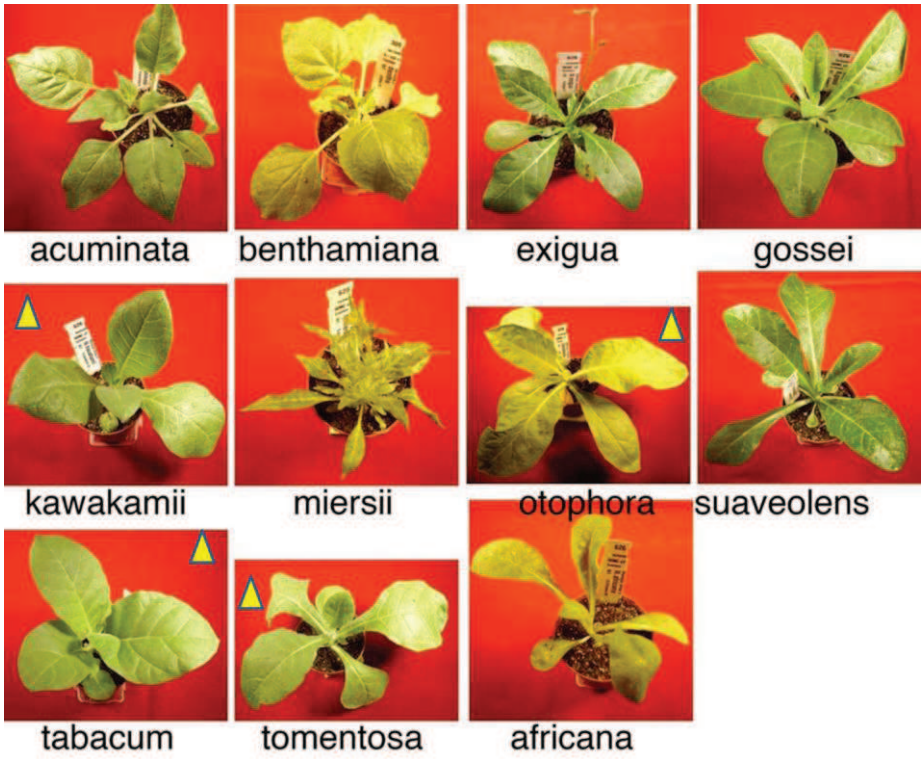


Figure 10. Different Nicotiana species, some of which contain cT-DNA sequences (marked by yellow triangles)

In 2012, after a systematic search in 127 plant species, a cT-DNA with genes *rolB, rolC, ORF13, ORF14* and *mis* was found in *Linaria vulgaris* (Matveeva et al., 2012) (figure 11A). This cT-DNA is organized as a single, imperfect direct tandem repeat. None of the cT-DNA genes seems to be expressed (Matveeva et al., 2012).

More recently, cT-DNAs were found in *Ipomoea batatas* (sweet potato) which is an important food crop (figure 11B). Two different T-DNA regions called *ib*T-DNA1 and *ib*T-DNA2 were described. *ib*T-DNA1 was found to contain four ORFs homologous to the *iaaM, iaaH, C-protein (C-prot),* and

agrocinopine synthase (*acs*) genes. i*b*T-DNA2 contains at least five ORFs with significant homology to the *orf14, orf17n, rolB/rolC, orf13*, and *orf18/orf17n* genes of *A. rhizogenes*. The *ib*T-DNA1 insertion is located in an *Ipomoea* gene which could have been functional before T-DNA insertion. *acs*, *iaaM, iaaH, C-prot, rolB* and *rolC* genes all have very low RNA expression levels in different *Ipomoea* tissues (Kyndta et al., 2015). The different plant species with cT-DNAs are natural GMOs, and studies on their origin, evolution and properties are only just starting.

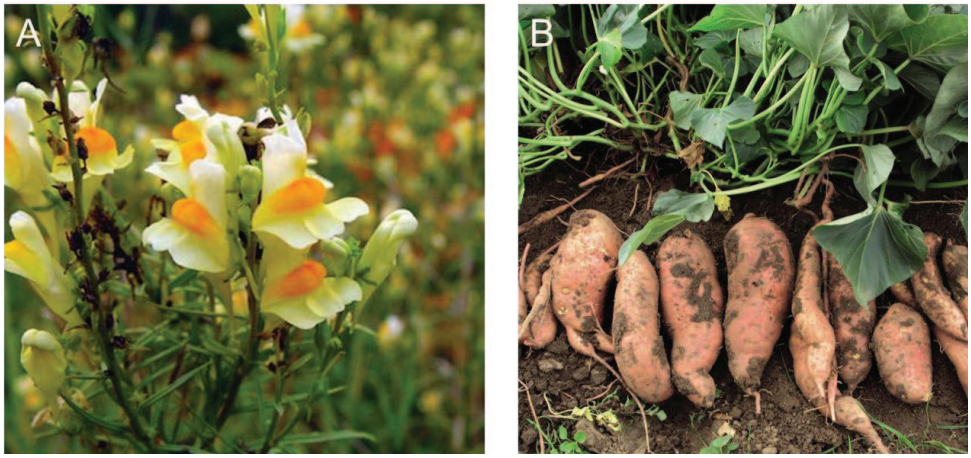


Figure 11. Natural GMOs
(A) *Linaria vulgaris* (from https://gobotany.newenglandwild.org/species/linaria/vulgaris/). (B) *Ipomoea batatas* (from http://www.simplyappalachian.com/article/2015/10/sweet-sweet-potatoes).

I.5.2 HGT among other organisms

HGT is one of the main causes of microrganism genome diversity. Based on deep-sequencing from 116 prokaryotic genomes, it has been found that about 14% of ORFs were introduced by recent HGT (Nakamura et al., 2004). Most of the genes involved in HGT among bacteria code for antibiotic resistance, utilization of different substrates, bacteriocins, immunity to bacteriophages, production of exopolysaccharide and biogenic amines (Rossi et al., 2014). Antibiotic resistant genes have spread from bacteriophages to bacteria through mobile elements and have been strongly selected (Muniesa et al., 2013).

Plasmid-derived HGT from bacteria to fungi has become more and more likely since in experimental conditions, Agrobacterium is able to mediate the transformation of the filamentous fungi (Michielse et al., 2008). Soil bacteria are quite often associated with fungi and some fungal genes were found to have homology to bacterial genes (Haq et al., 2014; Zhang et al., 2014). Tryptophan-dependent IAA biosynthesis in plants might be derived from HGT involving bacteria (Huang et al., 2014). The mitochondrial genome of the angiosperm *Amborella* contains six genome equivalents of foreign mitochondrial DNA, indicating the HGT of entire genomes from one angiosperm to another (Rice et al., 2013). Recently, it was shown that about one in six of

Hypsibius dujardini (water bear or tardigrade) genomic sequences were derived from bacteria, plants, fungi or Archaea according to deep-sequencing data (Boothby et al., 2015). These two examples demonstrate the big potential of HGT between plants and even between animals with other organisms.

In my thesis, I will describe five new cT-DNA sequences from the genus Nicotiana (Publication Chen et al. 2014, Chapter I). We have shown that the evolution of cT-DNA sequences in the genus Nicotiana is much more complex than previously assumed.

II. Leaf formation, and effects related to hormones and sugar

Among the cT-DNA genes that we discovered in *N. otophora* (Chen et al., 2014), TE-*6b* (from the *N. otophora*-specific TE region) has homology to the *6b* gene from *A. tumefaciens*. The Agrobacterium *6b* gene was found to incite tumors on *N. glauca* and *K. tubiflora* and can therefore be considered as an oncogene (Hooykaas et al., 1988). The *6b* genes from different Agrobacterium strains show strong and highly specific morphogenetic effects in transgenic plants, collectively called the "enation syndrome" (Helfer et al., 2003). Enations are abnormal additional leaf blades formed on the lower (abaxial) leaf side. Another remarkable property of *6b*-transformed plants is their capacity to transmit the enation phenotype to normal plants by grafting. In order to compare the *N. otophora* TE-*6b* gene with other *6b* genes, we decided to obtain a more precise description of the morphological changes of the aerial part of the plant using an inducible T-*6b* gene (publication Chen et al., 2015, Chapter II). For a better understanding of the "enation syndrome" effects, I will first present an overview of what is known about shoot apical meristem (SAM) development and leaf initiation.

II.1. Leaf formation

Leaf formation can be considered to occur in three steps: 1. leaf primordium initiation; 2. leaf differentiation; 3. leaf expansion and maturation (Dkhar and Pareek, 2014). Each will be discussed separately.

1. Leaf primordium initiation (figure 12A). PIN-FORMED1 (PIN1) is a transmembrane protein which is responsible for auxin transport. In the epidermis cells surrounding the SAM, the polarized PIN1 is highly expressed. The PIN1 is responsible for auxin gradient formation in the area beside the centrally located SAM where auxin has the highest concentration. A provascular system which is the beginning of a leaf primordium, is formed here. Another very important factor is *KNOX1* which plays a role in stem cell differentiation (figure 12B) (Dkhar, 2014). In the *KNOX1* mutant of maize, SAM development is strongly repressed (Kerstetter et al., 1997).

2. Leaf cell proliferation and differentiation (figure 12B). Proximodistal, adaxial/abaxial, and mediolateral polarities are created once a leaf primordium is formed. In Arabidopsis, the *KAN*, *HD-ZIPIII* gene family and *YAB* are the major contributors. *AS1, AS2, TAS3, ARF3, ARF4,* and miRNAs (miR165, miR166) are also involved in this process. These genes work together and many of them can inhibit expression of the others (Dkhar, 2014). For example, it was shown that in the abaxial domain, miR165/miR166 repress expression of the *HD-ZIPIII* gene family (Zhou et al., 2007; Kim et al., 2005). *KAN* and the *ARF* gene family are very important for abaxial domain formation. Besides, recent research has identified a new participant: the PUF RNA-binding protein family (*APUM23*). It also plays a role in leaf polarity formation. Mutations of *APUM23* show expression level changes of major adaxial/abaxial leaf polarity genes (Huang et al., 2014). The functional relation between the different genes is shown in figure 12B.

3. Leaf expansion and maturation. Transcription coactivator ANGUSTIFOLIA3 (*AN3*) and transcription factor GROWTH-REGULATING FACTOR5 (*GRF5*) are key factors in this stage (Horiguchi et al., 2005).

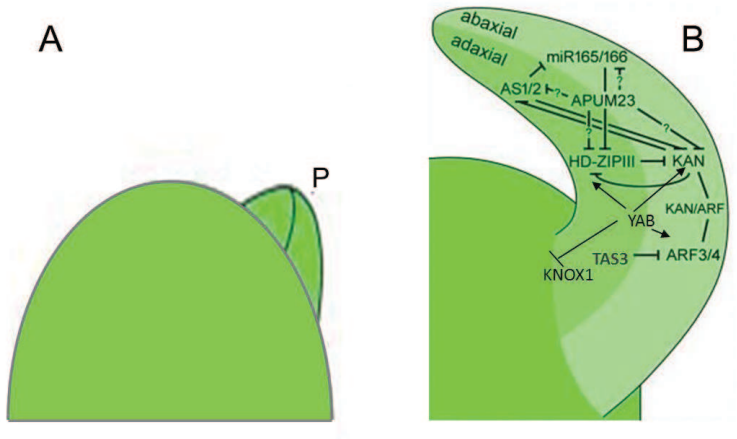


Figure 12. A model of genes involved in leaf polarity development in Arabidopsis
(A) Leaf primordium initiation. P: primodium. (B) Genes involved in leaf maturation. miR165/miR166 and *KAN* inhibit the expression of *HD-ZIPIII*; *AS1/2* and *KAN* downregulate each other; *YAB* promotes expression of *KAN*, *HD-ZIPIII* and ARF3/4 but inhibits *KNOX1*; the *APUM23* function is not yet defined (adapted from Dkhar, 2014).

II.2 Auxin and cytokinin metabolism, transport and signaling

Auxins and cytokinins are two of the most important phytohormones and have been implied in various aspects of plast gene activity.

II.2.1 Auxin

Auxin has a general effect on plant wound response, root growth and development, apical dominance, fruit growth and development, flowering and ethylene biosynthesis. *TAA* codes for a

transaminase which catalyzes the conversion of tryptophan to indole-3-pyruvate (Zhao, 2014). Indole-3-pyruvate is then converted into IAA by a flavin monooxygenase encoded by *YUCCA* (Tao et al., 2008; Stepanova et al., 2008). As mentioned in II.1, transmembrane protein PIN1 determines auxin flux. The *AUX1/LAX* gene is responsible for auxin import (figure 13). The F-box protein TIR1 is an auxin receptor and is part of an SCF complex which leads to the proteolytic degradation of the AUX/IAA repressor protection in the presence of auxin (Salehin et al., 2015). The auxin response factor (*ARF*) can be repressed by *AUX/IAA* when the auxin concentration is low. *ARF* is responsible for activating auxin-regulated gene expression (Ulmasov et al., 1999).

II.2.2 Cytokinin

Cytokinin is an important factor in plant cell division and differentiation. It works together with auxin and other factors like sugar to direct plant growth. Cytokinin biosynthesis involves two steps. First, an N^6-isopentenyladenine (iP) ribotide is formed by adding a prenyl moiety from dimethylallyl diphosphate to ATP/ADP. This step is catalyzed by isopentenyltransferase (IPT). Cytochrome P450 enzyme CYP735A1/2 then hydroxylates the isoprenoid side chain of iP ribotide yielding *trans*-zeatin cytokinin (Sakakibara, 2006). In cytokinin signaling, a histidine kinase (AHK) transfers a phosphate group to a histidine-containing phosphotransfer (AHP) protein which directly regulates the response regulators (RRs) in order to get the cytokinin effect (figure 13).

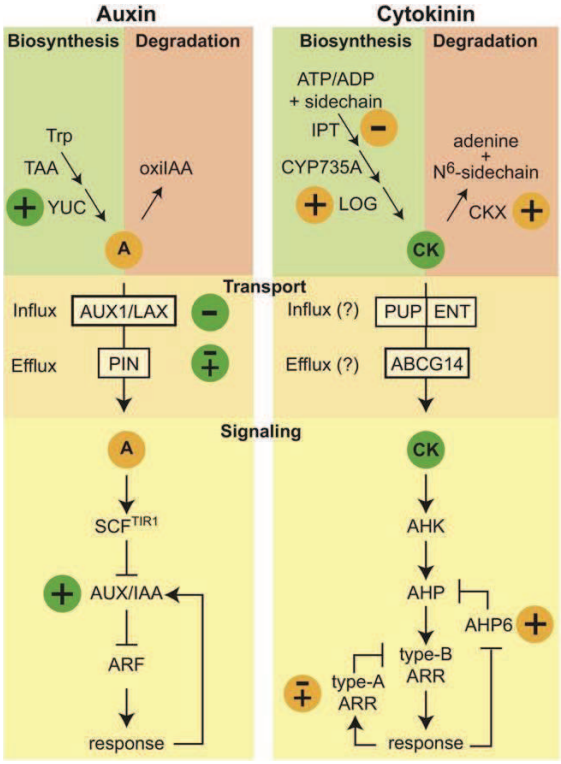


Figure 13. Auxin and cytokinin metabolism, transport and signaling
Different genes and their function involved in the auxin and cytokinin pathway. Orange round box: auxin; green round box: cyctokinin; +: stimulation; -: inhibition. (Schaller et al., 2015).

II.3 The combined role of sucrose and auxin on plant cell division and expansion

Sucrose is one of the most important carbon sources and a source of energy. It is transported from source (photosynthetic leaves) to sink (roots, flowers, seeds and meristem) tissues. Sucrose concentration can strongly influence plant activity. Generally, high sucrose levels inhibit photosynthesis in source tissues and stimulate growth and storage in sink tissues. Conversely, low sucrose levels stimulate photosynthesis in source leaves (Wang and Ruan, 2013). There are two sucrose degradation pathways. One is by hydrolysis into glucose and fructose catalyzed by invertase, the other is by degradation into UDP-glucose and fructose by sucrose synthase.

II.3.1 Sugar and auxin affect plant cell division

Sugar works not only as carbon source but also as a signaling compound which directly regulates plant gene expression (reviewed by Gibson, 2005; Eveland and Jackson, 2012). In plant cells, glucose is able to crosstalk with plant hormones like auxin, cytokinin, abscisic acid and ethylene (Gazzarrini and McCourt, 2001; Hartig and Beck, 2006). Activation of glucose-induced auxin biosynthesis and transport genes has differential effects on auxin receptors (Mishra et al., 2009). Interestingly, exogenous glucose enhances defects in lateral root initiation and root hair elongation in auxin sensitive mutants, indicating glucose may affect root development through auxin-based gene regulation (Mishra et al., 2009; Andrea et al., 2012).

The plant cell cycle is divided into different phases: gap0 (G0), gap1 (G1), DNA synthesis (S), gap2 (G2) and mitosis (M). Sugar and auxin work together to regulate the plant cell cycle (figure 14). In G1 phase, glucose (Glc) and auxin both induce cyclinD (CycD) gene expression. Auxin independently induces the cyclin-dependent kinase A-1 gene (CDKA) and activates the CycD/CDKA complex by phosphorylation and repression of the retinoblastoma related (RBR) protein (Wang and Ruan, 2013). Activated CycD/CDKA can phosphorylate RBR in order to release adenovirus E2 promotor binding factor A/B (E2FA/B) and the dimerization partner A (DPA) (Wang and Ruan, 2013). The E2FA/B-DPA complex initiates S phase gene expression. In S phase, glucose promotes phosphorylation of the CycA3/CDKA complex and leads to expression of S phase genes. Auxin accelerates S phase kinase associated protein 2A (SKP2A) degradation and stablilizes the E2FC (E2 promoter binding factor C)-DPB complex which represses the expression of S phase genes (Wang and Ruan, 2013). Recent data show that G2/M transition is regulated by TPR domain suppressor of stimpy (TSS) which is repressed by glucose (Skylar et al., 2011).

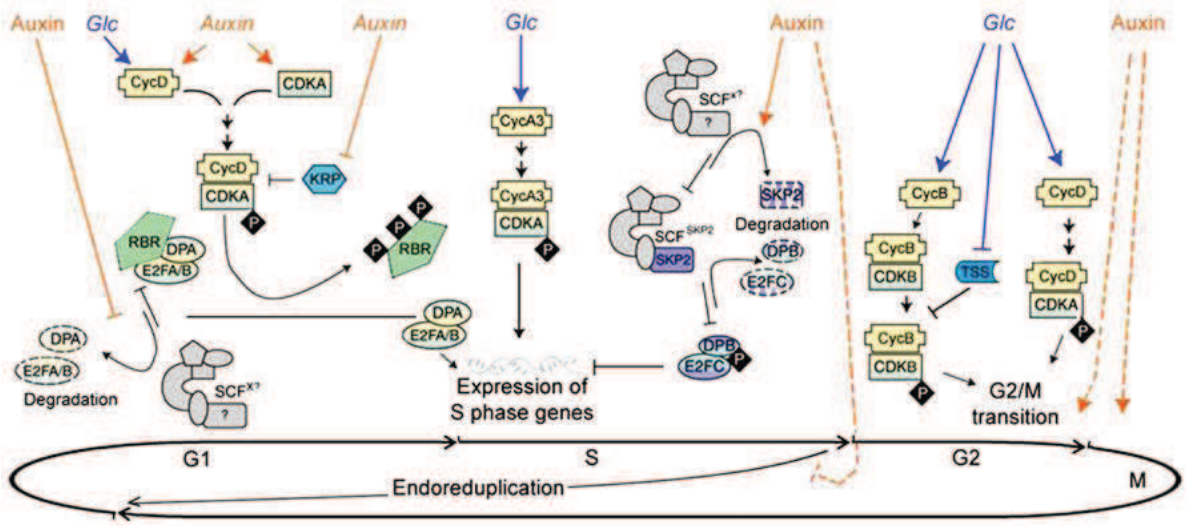


Figure 14. Roles of glucose and auxin in regulating the plant cell cycle
Different genes the expression of which is regulated by auxin and glucose. In G1, auxin induces both CycD and CDKA expression. Glc independently promotes CycD expression. Phosphorylated CycD/CDKA transfers the phosphor groups to RBR so that the DPA/(E2FA/B) complex is released and favors initiation of S phase. In S, Glc stimulates phosphorylation of the CycA3/CDKA complex through release of CycA3 resulting in S phase gene expression. In contrary, auxin inhibits S phase gene expression by inhibiting phosphorylation of DPB/E2FC through repression of the SCF/SKP2 complex. Glc and auxin both induce G2 to M transition. *Glc* and *auxin* show transcriptional regulation, Glc and auxin protein level regulation. (Wang and Ruan, 2013).

II.3.2 Sugar and auxin effect on plant cell expansion

Cell expansion is an important progress which includes the accumulation of osmotically active solutes, water influx into the cell, cell wall loosening and reconstruction (Dkhar and Pareek, 2014; Wang and Ruan, 2013). The final plant cell size is defined by the end of cell expansion.

Sugars are osmotically active solutes and their intercellular increase can lead to increased osmotic pressure. The hydrolysis of sugar into glucose and fructose doubles the osmotic contribution to the osmotic pressure in the vacuole. This in turn leads to cell size increase under the condition of cell wall plasticity. Sugars not only work as osmotically active solutes, but also as signals to modulate cell expansion. One possibility is that sugar influences cell plasticity by downstream regulation of wall associated kinases (Wang and Ruan, 2013).

Results

Chapter Ⅰ: Deep sequencing of *Nicotiana tomentosiformis* reveals multiple T-DNA inserts

I. Publication 1

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2014) **80**, 669–682 doi: 10.1111/tpj.12661

Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*

Ke Chen[1], François Dorlhac de Borne[2], Ernö Szegedi[3] and Léon Otten[1,*]

[1]*Department of Molecular Mechanisms of Phenotypic Plasticity, Institut de Biologie Moléculaire des Plantes du C. N. R. S., Rue du Général Zimmer 12, 67084 Strasbourg, France,*

[2]*Imperial Tobacco Bergerac, La Tour, 24100 Bergerac, France, and*

[3]*National Agricultural Research and Innovation Centre, Research Institute for Viticulture and Enology, Experimental Station of Kecskemét, PO Box 25 Kecskemét, H-6001 Hungary*

Received 22 March 2014; revised 30 July 2014; accepted 29 August 2014; published online 15 September 2014.
*For correspondence: (e-mail: leon.otten@ibmp-cnrs.unistra.fr).

SUMMARY

Nicotiana **species carry cellular T-DNA sequences (cT-DNAs), acquired by Agrobacterium-mediated transformation. We characterized the cT-DNA sequences of the ancestral *Nicotiana tabacum* species *Nicotiana tomentosiformis* by deep sequencing. *N. tomentosiformis* contains four cT-DNA inserts derived from different *Agrobacterium* strains. Each has an incomplete inverted-repeat structure. TA is similar to part of the *Agrobacterium rhizogenes* 1724 mikimopine-type T-DNA, but has unusual *orf14* and *mis* genes. TB carries a 1724 mikimopine-type *orf14-mis* fragment and a mannopine-agropine synthesis region (mas2-mas1-ags). The mas2′ gene codes for an active enzyme. TC is similar to the left part of the *A. rhizogenes* A4 T-DNA, but also carries octopine synthase-like (*ocl*) and *c*-like genes normally found in *A. tumefaciens*. TD shows a complex rearrangement of T-DNA fragments similar to the right end of the A4 TL-DNA, and including an *orf14*-like gene and a gene with unknown function, *orf511*. The TA, TB, TC and TD insertion sites were identified by alignment with *N. tabacum* and *Nicotiana sylvestris* sequences. The divergence values for the TA, TB, TC and TD repeats provide an estimate for their relative introduction times. A large deletion has occurred in the central part of the *N. tabacum* cv. Basma/Xanthi TA region, and another deletion removed the complete TC region in *N. tabacum*. *Nicotiana otophora* lacks TA, TB and TD, but contains TC and another cT-DNA, TE. This analysis, together with that of *Nicotiana glauca* and other *Nicotiana* species, indicates multiple sequential insertions of cT-DNAs during the evolution of the genus *Nicotiana*.**

Keywords: *Agrobacterium rhizogenes*, *Nicotiana tomentosiformis*, cT-DNA, natural transformation, horizontal gene transfer.

INTRODUCTION

Some plant genomes contain transferred DNA (T-DNA) derived from the root-inducing plasmids (pRi) of *Agrobacterium* (White *et al.*, 1983). It has been proposed that such T-DNA sequences were introduced into the host genome during *Agrobacterium rhizogenes* infections (White *et al.*, 1983; Furner *et al.*, 1986; Meyer *et al.*, 1995; Intrieri and Buiatti, 2001; Aoki, 2004). T-DNAs carry genes for plant growth and for the synthesis of specific molecules, called opines, used by the bacterium for its growth. The first cellular T-DNA (cT-DNA) sequence was identified in *Nicotiana glauca* and fully sequenced (Furner *et al.*, 1986; Aoki, 2004). It resembles the T-DNA of *A. rhizogenes* strain 1724, and is an imperfect inverted repeat with two fragments extending from *NgrolB* or *NgrolC* up to a mikimopine synthase (*mis*) gene located at the right T-DNA border. Both *NgrolC* (Aoki and Syono, 1999a) and *Ngorf13* (Aoki and Syono, 1999c) retain morphogenetic activity. Another 1724-like cT-DNA sequence has recently been characterized from *Linaria vulgaris* (Matveeva *et al.*, 2012). It carries an imperfect direct repeat, extending from *orf2* to *mis*. Besides *N. glauca*, other *Nicotiana* species (Furner *et al.*, 1986; Meyer *et al.*, 1995; Fründt *et al.*, 1998a; Intrieri and

Buiatti, 2001) also contain cT-DNAs, although only short fragments have been described. The genus *Nicotiana* has been subdivided in several sections (Goodspeed, 1954; Knapp *et al.*, 2004). Interspecific hybridization in this group has led to a complex phylogenetic pattern (Goodspeed 1954, Knapp *et al.*, 2004; Kelly *et al.*, 2010). The recent *Nicotiana* classification proposed by Knapp *et al.* (2004) comprises 13 sections. Some *Nicotiana* hybrids spontaneously develop genetic tumors that may be related to cT-DNA activity (Ichikawa *et al.*, 1990; Aoki *et al.*, 1994; Intrieri and Buiatti, 2001). The cT-DNA sequences of 42 *Nicotiana* species have been studied by PCR amplification of genomic DNA with *rolC*, *rolB*, *orf13* and *orf14* primers (Intrieri and Buiatti, 2001). The authors detected *rolC*, *orf13* and *orf14* in *Nicotiana cordifolia* (section *Paniculatae*), *rolC*, *orf13* and *orf14* in *Nicotiana tomentosiformis*, *Nicotiana otophora* (section *Tomentosae*) and *Nicotiana tabacum* (section *Nicotiana*), and *rolC* in *Nicotiana debneyi* (section *Suaveolentes*). Weak hybridization signals were detected on Southern blots for *Nicotiana benavidesii* (section *Paniculatae*), *Nicotiana setchellii* (section *Tomentosae*), *Nicotiana arentsii* (section *Undulatae*), *Nicotiana acuminata* and *Nicotiana miersii* (section *Petunioides*), *Nicotiana quadrivalvis* (earlier designated as *Nicotiana bigelovii*, section *Polydicliae*), *Nicotiana gossei*, *Nicotiana suaveolens* and *Nicotiana exigua* (section *Suaveolentes*). *N. tabacum* contains closely linked pairs of *rolC* and *orf13* genes (Fründt *et al.*, 1998a). These data were extended by sequencing two different cT-DNA fragments in *N. tabacum*: TA and TB (Mohajjel-Shoja *et al.*, 2011). Both resemble *A. rhizogenes* T-DNAs. The TA fragment contains *orf8–orf13*, with 70–80% identity with the *N. glauca* cT-DNA (gT). The TB fragment contains *orf14* and *mis* sequences, and shows 96% identity with gT. The TB fragment and gT are inserted in different sites within the *Nicotiana* genome (Suzuki *et al.*, 2002; Mohajjel-Shoja *et al.*, 2011). We therefore proposed (Mohajjel-Shoja *et al.*, 2011) that a single infection of a *glauca–tabacum* ancestor by a 1724-type *A. rhizogenes* strain led to the insertion of two T-DNA fragments, one (gT) being preserved in *N. glauca* and the other (TB) being preserved in *N. tabacum*, whereas TA was introduced during a second transformation event by another *A. rhizogenes* strain. As the regeneration of transformed plant cells and survival of natural transformants is probably rare, successive events of this nature seemed quite unlikely. The *N. tabacum* TA genes *torf13* (Fründt *et al.*, 1998a) and *trolC* (Mohajjel-Shoja *et al.*, 2011) are biologically active. *N. tabacum* is a recent amphidiploid hybrid of *Nicotiana sylvestris* and *N. tomentosiformis* (Murad *et al.*, 2002); the latter is the paternal progenitor and also carries cT-DNA sequences (Meyer *et al.*, 1995). Some authors proposed that *N. tabacum* contains *N. otophora* sequences (Kenton *et al.*, 1993; Riechers and Timko, 1999; Ren and Timko, 2001), but this has been questioned (Murad *et al.*, 2002; Moon *et al.*, 2008). In order to learn more about the nature and origin of cT-DNAs in the genus *Nicotiana*, we deep-sequenced the genome of *N. tomentosiformis*. cT-DNAs were detected by BLAST searches with known T-DNA sequences. The results show a surprisingly complex pattern with four different cT-DNA regions derived from different *Agrobacterium* strains. An additional cT-DNA was found in *N. otophora*.

RESULTS

Nicotiana tomentosiformis contigs with cT-DNA sequences

Total DNA from *Nicotiana tomentosiformis* (ITB646) was sequenced as described in the Experimental procedures. Two sets of paired reads were obtained for 200- and 500-bp fragments, respectively (59x total coverage), for an accepted genome size of 2.682 Gb (Sierro *et al.*, 2013). Reads were assembled into contigs (Experimental procedures). In all, 1.78 G reads (130.96 Gb) were obtained, generating 661 878 contigs (1 Gb). Less complete sequences were also obtained for *N. sylvestris* (ITB626, 35x coverage). For *N. sylvestris* we obtained 1.02-G reads (76.12 Gb) and 1 010 322 contigs (0.98 Gb) for an accepted genome size of 2.636 Gb (Sierro *et al.*, 2013). *N. tomentosiformis* single-copy regions showed no evidence of alleles, showing that the *N. tomentosiformis ITB646* genome is largely homozygous. Reads and contigs were analyzed for similarity to representative T-DNA sequences (reference T-DNAs; Table 1), using BLAST analysis (Altschul *et al.*, 1990;

Table 1 T-DNA and cT-DNA sequences used for detection of cT-DNAs in different *Nicotiana* species

DNA number	Accession number	Organism, strain	Genes, region
1	EF433766.1	Ar, K599	T-DNA
2	AJ271050.1	Ar, 2659	T-DNA
3	M60490.1	Ar, 8196	T-DNA
4	K03313.1	Ar, A4	TL-DNA
5	DQ782955.1	Ar, 15834	TR, *iaaH* and *iaaM*
6	X15952.1	Ar, 15834	TR, *rolBTR*
7	NC_002575.1	Ar, 1724	Ti plasmid
8	AF065242.2	At, Chry5	T-DNA
9	DQ058764.1	At, Bo542	pTiBo542
10	EU735069.2	*Linaria vulgaris*	T-DNA
11	AB071334.1	*Nicotiana glauca*	T-DNA, right arm
12	AJ237588.1	At, C58	pTiC58
13	CP000637.1	Av, S4	pTiS4
14	FN667969.1	*Nicotiana tabacum*	Partial TA
15	FN667970.1	*Nicotiana tabacum*	Partial TB

Ar, At, Av: *Agrobacterium rhizogenes*, *Agrobacterium tumefaciens*, *Agrobacterium vitis*.

Experimental procedures). Contigs containing cT-DNA sequences were assembled into single T-regions, which were checked by PCR amplification and sequencing of selected regions. In this way, four cT-DNAs were identified (TA–TD), all of which formed incomplete inverted repeats, see below.

TA region

The partial TA sequences from *N. tabacum* (FN667969, Mohajjel-Shoja *et al.*, 2011) were also found in *N. tomentosiformis* and could be completed to a cT-DNA region of 31.6 kb with 1531 and 4706 bp of plant sequences on the left and right side, respectively (KJ599826, 31627 bp; Figure 1b). The TA sequences are similar to the TL-DNA from *A. rhizogenes* strain A4 (K03313.1; Figure 1a). A contig from *N. sylvestris* (KJ599825, 5509 bp) has 95% identity with the sequences flanking the TA region (a and b, in Figure 1b), and probably represents the unmodified insertion site. An 8-bp sequence (ATGCAACT) from this site is missing in *N. tomentosiformis* and was probably deleted during T-DNA insertion (Figure 1c). TA contains an inverted repeat consisting of TA-1 and TA-2, with 1.2% divergence. Both align with the middle part of the A4 TL-DNA. Figure 1b shows TA-1 with its left border on the left, and TA-2 thus appears inverted. TA-1 extends 0.4 kb more to the left. TA lacks the A4 *IS870* transposon, as in the T-DNAs of *A. rhizogenes* strains 1724 and 2659. TA-1 starts within the 3′ part of the *orf3* gene, TA-2 just after *orf3*. Beyond *orf13* there are no sequences with similarity to A4-*orf14*, but a translated sequence shows 36% protein identity with Orf14 from *A. rhizogenes* strain 8196. Further to the right lies a region with 65% partial identity with the 1724 *mis* sequence. Similarity between TA–1 and TA-2 breaks off at the right border position. Between TA–1 and TA-2, sequences with similarity to *N. tomentosiformis* repeats occur (KJ599826, 13376–15450) and delimit the TA cT-DNA to 23.3 kb. Because the polymorphic sites in the two copies of the TA repeat could not always be attributed to a particular copy, we designed PCR experiments to obtain and sequence copy-specific fragments (Figure 1b, a1–a4; Table 2; Experimental procedures). In this way, 79 polymorphic sites in the 11.4-kb repeat were attributed to TA-1 and TA-2. The only intact open reading frames (ORFs) in TA-1 and TA-2 are the two *rolC* and *orf13* gene copies.

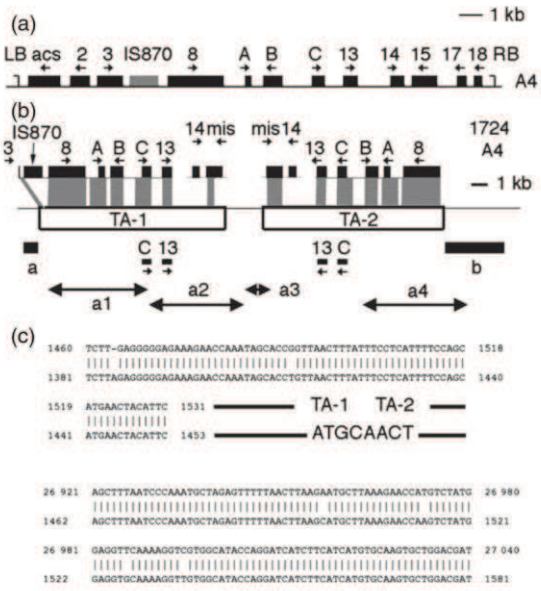


Figure 1. TA region (KJ599826).
(a) TL-DNA of *Agrobacterium rhizogenes* A4 (from K03313.1).
(b) TA region aligned with part of A4 TL-DNA and 1724 T-DNA. The *orf14* gene was detected by PSI-BLAST analysis (Experimental procedures). Regions a and b show similarity to *Nicotiana sylvestris* contig KJ599825; black rectangles, ORFs; gray boxes, regions with sequence similarity; double-headed arrows, fragments used for sequencing (see also Table 2).
(c) Insertion site of TA, only part of the alignment is shown; top, *Nicotiana tomentosiformis*; bottom, *N. sylvestris* contig KJ599825. An 8-bp sequence is missing in *N. tomentosiformis*; nucleotides 1531 and 26 921 delimit TA.

TB region

The TB sequences from *N. tabacum* (FN667970; Mohajjel-Shoja *et al.*, 2011) were also found in *N. tomentosiformis* and completed using several contigs to a 14 404-bp sequence, including cT-DNA and surrounding plant sequences (KJ599827). The TB insert contains two dissimilar regions, TB-1 and TB-2, organized as an incomplete inverted repeat. Figure 2a shows TB-1 (3.3 kb) in the reverse orientation and TB-2 (8.1 kb) in the usual orientation. TB-1 carries *mis* and *orf14* genes, followed by a 122-bp fragment with 86% identity with the A4 TR-DNA-located *iaaM* gene (indicated by an asterisk). TB-2 is highly similar to TB-1, but carries a 4.8-kb extension on the right. This extension contains an agropine synthase (*ags*)-like sequence similar to an *ags*-like sequence from *Nectria haematococca* (XM_003046964.1, 67% identity) and *A. tumefaciens* Chry5 (AF065242.2, 68% partial identity), and another sequence with 85% identity with the mannopine synthase *mas1′* and *mas2′* genes of *A. rhizogenes* 8196 (M60490.1). The mannopine–agropine region is probably derived from a second T-DNA that might also be the source of the *iaaM* fragment. The *mas1′*, *mas2′* and *ags* genes have been identified in *A. rhizogenes* strains A4 (TR-DNA) and 8196 (Hansen *et al.*, 1991), and in *A. tumefaciens* strains 15955 (TR-DNA) and Chry5 (TR-DNA). The *mas2′* gene encodes the synthesis of deoxyfructosylglutamine (DFG), whereas *mas1′* and *ags* are required for mannopine and agropine synthesis, respectively. Sequences flanking TB (a and b in Figure 2a) align with contiguous *N. tabacum* sequences (for example, FH649169.1), with the latter most likely corresponding with the unmodified TB insertion site (Figure 2b). In *N. tomentosiformis* a 69-bp fragment has apparently been lost from this site. A 402-bp fragment to

Table 2 Primers for amplification and sequencing of copy-specific fragments

Region	Primer name	Direction	Sequences (5′→3′)	Position
TA	3017	F	CGCTGCATCAGTATTGCTCA	794
	1725	R	GTACCTCGCATTCGGCAATCGC	7467
	1709	F	ACAGGCAAGCCCAAAATCAGGTC	7067
	1710	F	CAGGTCAAGCTTTTGCTGGATCAT	7084
	2967	R	CAAAGTAATTCCGTCTCAGCT	13312
	3069	F	GAAGCAGCGTGCTTAGCGGAAC	13331
	3071	R	TCACGTGCTTAAGATGCTGATAC	15474
	1711	F	CTGGAACCCACGTTAGGGCTGTCG	21079
	3012	R	GAGTGGCGTGCTGGAGTATC	27569
TB	2327	F	ACTTATTTCAACCCTTTTCG	862
	2347	R	CCGCCCCAACGGCTATGCTC	4184
	2348	F	GGTGCTTTACCGTTGTTCTG	3899
	3081	F	AAACGTGTTCAGAATGGCGA	3734
	1846	F	GTCGCTGGACGCACCATACG	1187
	1917	F	TGAGCATGCCGTCTTGAC	4304
	2370	R	TTCAAGGCGAACACGATTTG	7431
	1841	F	TTGACTACCGCGGTGATCTC	4986
	1843	F	GTATGTGCTGGGAAAAGCATA	5429
	1848	F	GCGATTGTCAGTCTCTTGAA	6083
	1849	F	TCAAGTGCAGCAGAATAGAT	6125
TC	2986	F	TCTGCTACTATCCCGATGCC	10501
	2376	R	GGAGGGTGCTTTCTCTGAAC	15291

The repeated regions of TA, TB and TC were analyzed for polymorphisms by amplification with copy-specific primers and sequencing. For map positions, see Figures 1b (TA), 2a (TB) and 3a (TC).

the right of this is similar to an *N. tabacum* DNA repeat (found in ET728024.1). Sequences a, b and R define TB as an 11.3-kb region. Polymorphic sites in TB-1 and TB-2 were attributed to each of the two TB region arms by PCR amplification and sequencing of TB-1- and TB-2-specific fragments (Figure 2a, b1 and b2; Table 2). The TB repeat divergence is 2.6%. TB-1 and TB-2 contain only three intact ORFs: *orf14* from TB–1 and TB–2 and *mas2′* (TB-2).

TC region

Several *N. tomentosiformis* reads and contigs showed similarity to the TL-DNA sequences from A4, but differed from TA and TB. These sequences could be assembled into a 16 918-bp segment with an inverted repeat structure (KJ599828; Figure 3a). The sequences that delimit the ends of the cT-DNA show similarity to various *N. sylvestris* sequences (ASAF01011933.1 and others), and probably correspond to the unmodified site (Figure 3b). A 1997-bp fragment appears to have been deleted at the TC insertion site. The structure of the TC region includes complex rearrangements involving repeated A4-like regions and additional unique regions. The A4-like region of TC-1 starts in *orf2* and ends shortly after *rolB*. A sequence at the left end of TC-1 translates into a 373-AA octopine/nopaline dehydrogenase-like protein, with 40% identity with EHA20957.1 from *Aspergillus nidulans*, 34% identity with EGG11641.1 from the poplar rust fungus *Melampsora larici-populina* and 28% identity with YP_007513935.1 from *Bradyrhizobium oligotrophicum*. The most related opine synthases from *Agrobacterium* are Ocs from *Agrobacterium tumefaciens* strain A6 and *Agrobacterium vitis* strain Tm4 (both 25% identity). We designate this ORF *ocl* (for octopine synthase-like), and note that it is found at the same position as the agrocinopine synthesis gene (*acs*) from A4 (Figure 1a). TC-2 is a smaller T-DNA fragment, with A4 similarity starting at *orf2*, and ending after *rolB*, and with a large part of *orf8* missing. At the right of TC-2 *orf2*, an ORF translates into a putative 245-AA fragment with 42% identity with the central part of protein C (Otten *et al.*, 1999) from *A. tumefaciens* C58 T-DNA (NP_862657.1, 523 AA), *A. vitis* Tm4 TB-DNA (AAD30491.1, 520 AA, 41% identity), *Agrobacterium larrymoorei* (WP_027676702, 317 AA, 41% identity), *Melampsora larici-populina* (EGG11381.1, 533 AA, 41% identity) and the endophytic fungus *Pestalotiopsis fici* (XP_007840635.1, 540 AA, 30% identity). The TC sequences show 68–75% identity with the corresponding TA sequences, and 69–77% identity with those of A4 (TA/A4: 70–79% identity). The TC repeats show 5.8% divergence, they could be separated by computer analysis and by sequencing of a TC-2-specific PCR fragment (c1; Figure 3a). The only complete ORF in the TC region is *ocl*; this gene was tested for its capacity to encode opine synthesis (see below).

TD region

Several *N. tomentosiformis* reads and contigs also contained sequences similar to the TL-DNA region from

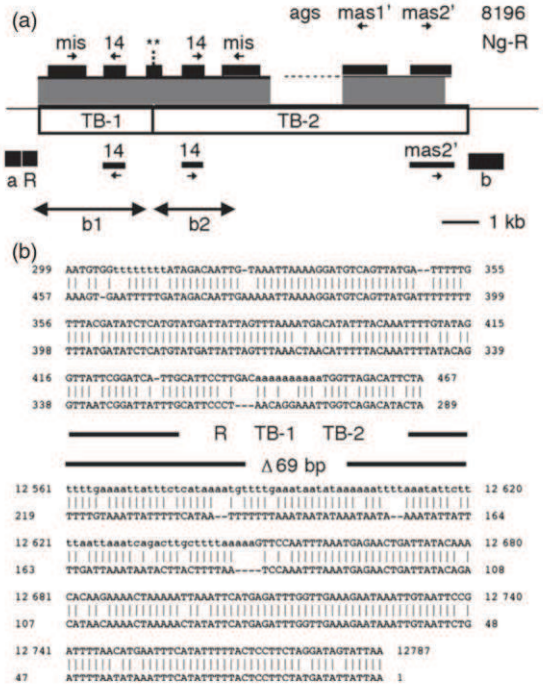


Figure 2. TB region (KJ599827).
(a) TB region aligned with part of *Nicotiana glauca* right arm cT-DNA (Ng-R) and part of 8196 T-DNA; a and b, regions with similarity to contiguous plant sequences; R, repeated plant sequence; dotted vertical line, border between TB-1 and TB-2; dotted horizontal line, region with similarity to *ags* from *Nectria haematococca*; **, 122-bp region with similarity to A4 *TR-iaaM*; black rectangles, open reading frame; grey boxes, similarity regions; double-headed arrows, fragments used for sequencing (see also Table 2).
(b) Insertion site of TB; top, *Nicotiana tomentosiformis* sequence; bottom, *Nicotiana tabacum* sequence FH649169.1. Only part of the alignment is shown. A 69-bp fragment is missing in *N. tomentosiformis*. *N. tomentosiformis* nucleotides 467 and 12561 delimit TB.

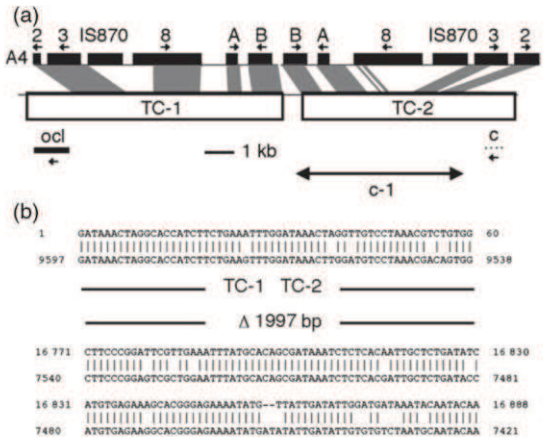


Figure 3. TC region (KJ599828).
(a) TC region aligned with A4 TL-DNA. On the left, TC-1 carries *ocl*, on the right, TC-2 carries a partial gene *c* sequence (dotted line); black rectangles: ORFs; grey boxes, regions with sequence similarity; double-headed arrow, fragment used for sequencing (see also Table 2).
(b) TC insertion site; top, *Nicotiana tomentosiformis*; bottom, ASAF01011933.1 from *Nicotiana sylvestris* (Sierro *et al.*, 2013). Only part of the alignment is shown. A 1997-bp sequence from *N. sylvestris* is missing in *N. tomentosiformis*. *N. tomentosiformis* nucleotides 60 and 16771 delimit TC.

A. rhizogenes A4 which carries the *orf15* (also called *rolD*), *orf17* and *orf18* genes of *Agrobacterium rhizogenes* A4. These were assembled and found to be part of a 22 172-bp TD region (KJ599829; Figure 4a). In A4 (Figure 1a), *orf15* is followed by a second *orf15*-like sequence fragmented into *orf17* and *orf18*. In *N. tomentosiformis* two separate fragments correspond to *orf15*–*orf17* (A4K03313.1 coordinates 17 090–18 782 bp) and *orf18*-right border RB (18 864–19 898 bp). Close to these fragments an inverted repeat is found (double-headed arrows in Figure 4a), and the right arm contains two ORFs: *orf511* can be translated into an unknown protein of 511 AA, *orf14* codes for a protein of 192 AA, with 36% identity with Orf14 (AAA22099.1) from *A. rhizogenes* 8196. To the left of *orf14*, a truncated *orf15*-like ORF (266 instead of 344 AA) is found. TD-Orf14 differs strongly from TA-Orf14 and TB-Orf14. The left arm of the TD repeat contains truncated versions of *orf511* (*orf455*) and *orf14* (*orf14**, potentially encoding a 165-AA protein).

Close inspection of this complex region suggests that TD most probably originated from a T-DNA with an *orf14*-*orf511*-*orf15*-*orf17*-*orf18*-RB structure by ligation of three dissimilar fragments (a, b and c; Figure 4b). The TD-1 and TD-2 repeats show 1.7% divergence, and could be separated by computer analysis. The surrounding sequences (KJ599829, 3-3246 and 17140-17526) are similar to an *N. sylvestris* sequence (ASAF01209345.1), and probably correspond with the TD insertion site. Apparently, a 28-bp sequence was lost during insertion (Figure 4c). TD has a size of 13.9 kb. In all, TA, TB, TC and TD add up to 67.3 kb.

Presence of cT-DNAs in other *Nicotiana* species

We tested whether cT-DNAs were also present in three close *N. tomentosiformis* relatives, *Nicotiana kawakamii*, *Nicotiana tomentosa* and *N. otophora*, using PCR amplification of genomic DNA with different TA, TB, TC and TD primers (Figure 5; Table 3). Whereas *N. kawakamii* carried TA, TB, TC and TD sequences, *N. tomentosa* contained TB, TC and TD, but lacked TA, and *N. otophora* contained only TC (Table 3). As it was reported that *N. otophora* carries *rolC* and *orf13* sequences (Intrieri and Buiatti, 2001) with high similarity to our TA sequences, we investigated the TA insertion site from *N. otophora* by PCR amplification and sequencing, but found a non-modified insertion site without TA sequences. The TA insertion site of *N. tomentosa* was likewise found to be non-modified (Figure S1).

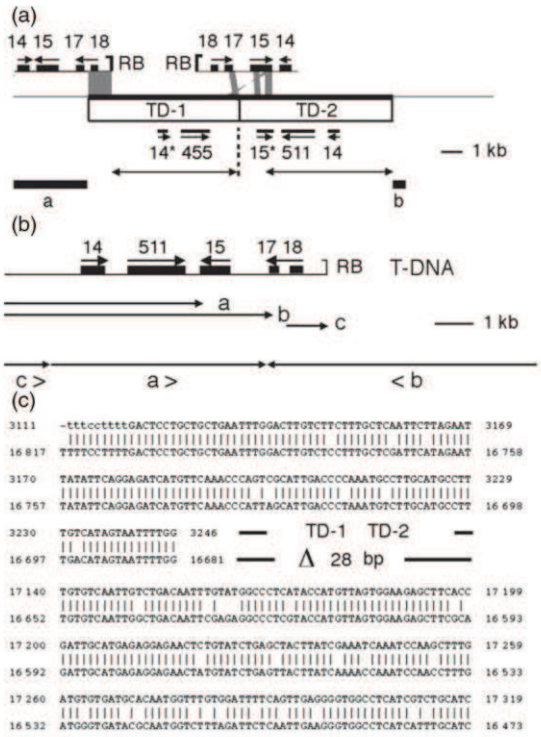


Figure 4. TD region (KJ599829).
(a) TD region aligned with the right part of the A4 TL-DNA (Figure 1a); 511 (*orf511*) and 455 (*orf455*) potentially code for a 511-AA and 455-AA protein, 14 (*orf14*) for an evolutionary distant Orf14 protein; 14* and 15*, truncated *orf14* and *orf15* genes; double-headed arrows, repeat sequence; a and b indicate regions with similarity to ASAF01209345.1 from *Nicotiana sylvestris*; black rectangles, ORFs; RB, right border; grey boxes, regions with sequence similarity.
(b) Reconstruction of the complex TD structure; top, hypothetical structure of the T-DNA that gave rise to TD, carrying *orf14-orf511-orf15-orf17-orf18*-RB; middle, fragments a, b and c that were combined into the TD cT-DNA; bottom, the combination of rearranged fragments as found in TD with their original orientation.
(c) Insertion site of TD. Top, *Nicotiana tomentosiformis* sequence; bottom, ASAF01209345.1 from *Nicotiana sylvestris*. Only part of the alignment is shown. TD is delimited by *N. tomentosiformis* nucleotides 3246 and 17140; a 28-bp sequence has been deleted.

We also searched recently published *Nicotiana* whole genome shotgun (WGS) sequences for cT-DNA sequences by BLAST analysis using the reference T-DNAs (Table 1) and the *N. tomentosiformis* TA, TB, TC and TD regions as query sequences. Contigs from *N. tomentosiformis* TW142 (WGS project ASAG; Sierro *et al.*, 2013) were 99% identical to our sequences. A comparison between the two *N. tomentosiformis* data sets (Figure S2; Table S1) found many small TW142 contigs in the TA region, whereas the TB region sequence detected a single contig, covering TB and neighboring regions. TC and TD are each covered by two large overlapping contigs. Differences between the two

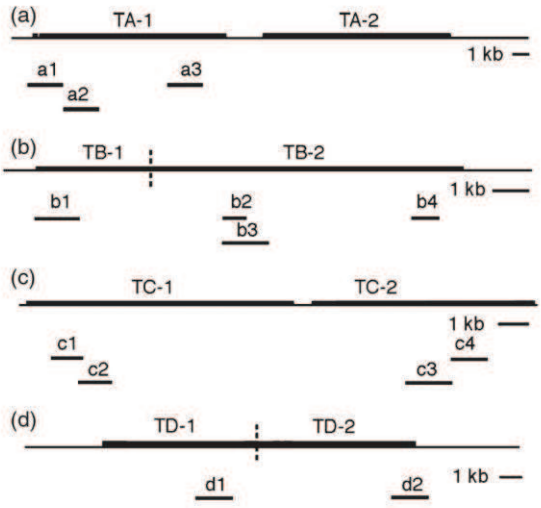


Figure 5. TA, TB, TC and TD fragments amplified in different *Nicotiana* species.
(a–d) TA, TB, TC and TD regions, respectively. Vertical dotted lines: borders between TB-1 and TB-2, and between TD-1 and TD-2. Numbered fragments: PCR fragments. For sequences and coordinates of primers, see Table 3.

sequence sets are listed in Table S1. A similar analysis of genomic contigs of three *N. tabacum* cultivars (cv. K326, cv. TN90 and cv. Basma/Xanthi; WGS projects AWOJ, AYMY and AWOK, respectively; Sierro *et al.*, 2014) detected TA (Figure S3), TB (Figure S4) and TD (Figure S6), and confirmed the lack of TC (Figure S5, see also below). No cT-DNA sequences were found in *N. sylvestris* (TW136, ASAF project; Sierro *et al.*, 2013) or *N. benthamiana* (CBMM WGS project).

cT-DNA sequences in *Nicotiana otophora*

A BLASTN analysis of *N. otophora* genomic DNA contigs (AWOL project; Sierro *et al.*, 2014) using TA, TB, TC and TD as queries (Figures S3d, S4d, S5d and S6d) confirmed the presence of TC and lack of TA, TB and TD, as detected by PCR analysis (Table 3). The *N. otophora* TC region is covered by 25 AWOL contigs that remain to be assembled. Interestingly, seven AWOL contigs showed similarity to the TB region, with an average identity of 90% (Figure S5d, boxed area). A further BLASTN analysis using the reference T-DNA sequences (Table 1) as queries detected 37 AWOL contigs, with sequences that differed from TA, TB, TC and TD. Using these 37 contigs as queries, a further seven AWOL contigs were found. The 44 additional *N. otophora* contigs (listed in Table S2) contain sequences similar to the vitopine synthase (*vis*) gene from *A. vitis* S4, a sequence distantly related to the *6b* genes from *A. tumefaciens* and *A. vitis*, sequences similar to *iaaM*, *iaaH* and *acs* from *A. tumefaciens* C58, and to *rolB*, *rolC*, *orf13*, *orf14*, *mas2′* and *mas1′* from *A. rhizogenes* 8196. They

Table 3 Primers for amplification of the TA, TB, TC and TD regions in different *Nicotiana* species

Region	Frag	TOA	OTO	KAW	Primer name	Dir	Sequences(5′→3′)	Position
TA	a1	–	–	+	2366	F	CACACCAGTCAACATGAACT	1167
					1957	R	ACGCCCGTGCCAAGTAATTCAA	3034
	a2	–	–	nt	1955	F	ATTGAATTACTTGGCACGGGCG	3011
					2004	R	CGTTTGAAATTTGAGCTCCGTT	5098
	a3	–	–	nt	1669	F	AACTGTTAGCCTGCGACGTT	9634
					2334	R	GAGATGTACTCTTCCGCACT	12265
TB	b1	+	–	nt	2327	F	ACTTATTTCAACCCTTTTCG	862
					1849	R	TCAAGTGCAGCAGAATAGAT	1969
	b2	+	–	+	1848	F	GCGATTGTCAGTCTCTTGAA	6083
					1847	R	TTCGCGCTTCATCCAAACGT	6822
	b3	–	–	nt	1849	F	TCAAGTGCAGCAGAATAGAT	6125
					2370	R	TTCAAGGCGAACACGATTTG	7431
	b4	+	–	+	2447	F	CTGAACGCAAGAGAGCACTG	11339
					2452	R	ATGCATAGATTAGGTCCGCC	12079
TC	c1	–	+	+	2372	F	GCCGAGAGATAACCGGCAAG	1143
					2375	R	GCCTTGATCGCTTTCGACAG	2116
	c2	–	+	–	2420	F	CTGTCGAAAGCGATCAAGGC	2096
					2413	R	AGGCGAAGAGGAGTTCAAAG	3093
	c3	+	+	+	2368	F	CCAGGCACCGTTGCTTCTTG	12607
					2420	R	CTGTCGAAAGCGATCAAGGC	14114
	c4	+	+	+	2375	F	GCCTTGATCGCTTTCGACAG	14097
					2376	R	GGAGGGTGCTTTCTCTGAAC	15291
TD	d1	+	–	+	2387	F	TAGGCTCCTCTGGTAGAAGT	4475
					2394	R	TCCGCCCCTAACGAAGCGAA	5498
	d2	–	–	nt	2444	F	GAGATGCAAAGCAGGTGC	9274
					2358	R	GGCCCATTCAATCAGCAAGC	10322

TOA, *Nicotiana tomentosa*; OTO, *Nicotiana otophora*; KAW, *Nicotiana kawakamii*; nt, not tested; Frag, fragment; Dir, direction.
For map coordinates, see Figures 1b (TA), 2a (TB), 3a (TC) and 4a (TD). For PCR fragment positions see Figure 5.

were mapped on the most similar T-DNAs, those of C58 (72% identity; Figure 6a), 8196 (79% identity; Figure 6b) and S4 (71% identity; Figure 6c). AWOL010660148.1 (1–481) and AWOL010282125.1 (1–1413, reverse complement) could be combined into a 1894-bp sequence highly similar to *N. sylvestris* ASAF01131078.1 (40 253–42 182), *N. tomentosiformis* ASAG01005302.1 (9099–10 993), *N. tabacum* AYMY01014259.1 (9796–11 695), AWOK01375210.1 (9113–11 012) and AWOJ01448950.1 (28 939–30 838). These *N. sylvestris, N. tomentosiformis* and *N. tabacum* sequences most probably represent the insertion site of the additional *N. otophora* cT-DNA sequences. An 8-bp sequence (AGGTTGTT) was apparently lost upon insertion of this cT-DNA, which we propose to call the TE cT-DNA (TE; Figure 6d). The TE cT-DNA contigs could not yet be assembled into a single T-region; numerous overlapping contigs with sequence variants and seemingly aberrant contigs indicate that the TE sequences are partially duplicated and rearranged. The unique TE sequences add up to about 18 kb.

Deletion variants of T regions

Earlier PCR analysis showed that the *orf13* genes of *N. tabacum* cv. Basma Drama 2, Samsoun and Xanthi (oriental tobacco cultivars like Basma/Xanthi) are truncated at their 3′ side (leaving 400 bp of the *orf13* gene intact), whereas those of *N. tomentosiformis, N. tabacum* cv. Wisconsin 38 and *N. tabacum* cv. Havana 425 are intact (Mohajjel-Shoja *et al.*, 2011). Using the TA region (KJ599825) as a query for BLASTN analysis, we identified the Basma/Xanthi contig AWOK01769803.1 (414 bp), which carries two differently truncated *orf13* fragments in inverted orientation, and a 42-bp fragment of unknown origin in between; this contig spans a 9684-bp deletion, removing the central part of TA (Figure 7a).

We also investigated the lack of TC sequences in *N. tabacum*. First, the *N. tomentosiformis* TC region was enlarged with ASAG01063776.1 and ASAG01044398.1 (Sierro *et al.*, 2013) to a region of 93 kb. A BLASTN analysis of this sequence against the *N. tabacum* cv. K326 and Basma/Xanthi genomes showed a deletion removing the complete TC insert plus 1097 bp on the left and 1021 bp on the right. In cv. TN90 the deletion on the left is smaller (995 bp), and is followed by a direct repeat of 120 bp (Figure 7b). The most likely explanation for these structures is that the initial deletion was accompanied by a small duplication, and that subsequent recombination between the direct repeats removed the intervening

102 bp (Figure 7c). A more general model showing the order of insertion of the different cT-DNAs with respect to the *Nicotiana* evolutionary tree is presented below (see Discussion).

Functional analysis of open reading frames

Most of the *N. tomentosiformis* cT-DNAs ORFs are interrupted by stop codons, but a few are intact and could be biologically active. In TA these are *rolC* and *orf13*, in TB these are *orf14* and *mas2′*, in TC *ocl* and in TD *orf14*. It is known that the *orf13* (Fründt *et al.*, 1998a) and *rolC* (Mohajjel-Shoja *et al.*, 2011) genes of *N. tabacum* are biologically active, and in view of the high sequence conservation it is likely that this is also the case in *N. tomentosiformis*. The classical *orf14* genes induce little (Costantino *et al.*, 1994; Aoki and Syono, 1999b) or no modification in plant growth (Lemcke and Schmülling, 1998; our own observations). The *mas2′* gene from TB-2 (*TOF-mas2′*) may code for DFG synthesis. *TOF-mas2′* was placed under the control of the constitutive 2x35 promoter and tested by transient expression in *N. benthamiana* leaves, with 2x35S-*A4-mas2′* from A4 used as a positive control (Experimental procedures). DFG synthesis was analyzed by paper electrophoresis and silver nitrate staining. Both *2x35S-TOF-mas2′* and *2x35S-A4-mas2′* constructs induced DFG synthesis (Figure 8a). By the above criteria, the *TOF-mas2′* gene encodes an active Mas2′ enzyme.

The *ocl* gene was analyzed in the same way, with the vitopine synthase (*vis)* gene from *A. vitis* strain S4 serving as a positive control. The *ocl* and *vis* constructs were transiently expressed in *N. benthamiana*. As the *ocl* products are unknown, we chose a negative staining method to detect possible products. Whereas *2x35S-vis* led to a clear white spot at the expected vitopine position, *2x35S-ocl* expression did not lead to any difference with the negative control (Figure 8b). Therefore, *ocl* does not lead to the synthesis of detectable product in this experimental system.

DISCUSSION

The results reported here show that the origin and evolution of cT-DNAs in *Nicotiana* is considerably more complex than previously appreciated. *N. tomentosiformis* contains four different cT-DNAs that total 67.3 kb of horizontally transferred sequences. TA carries *orf3-orf8-rolA-rolB-rolC-orf13-orf14-mis,* TB carries *orf14-mis* and *ags-mas1′-mas2′*, TC carries *orf2-orf3-orf8-rolA-rolB*, *ocl* and *c*, and TD carries *orf15*, *orf511* and *orf14*. These results extend the previously reported cT-DNA sequences by a considerable extent, and provide a much more detailed picture of their structures and insertion sites. Each of the *N. tomentosiformis* cT-DNAs has an incomplete inverted repeat structure. In comparison, the *N. glauca* cT-DNA sequence also carries an inverted repeat (Suzuki *et al.*, 2002), whereas the *Linaria vulgaris* cT-DNA (Matveeva *et al.*, 2012) is a direct repeat.

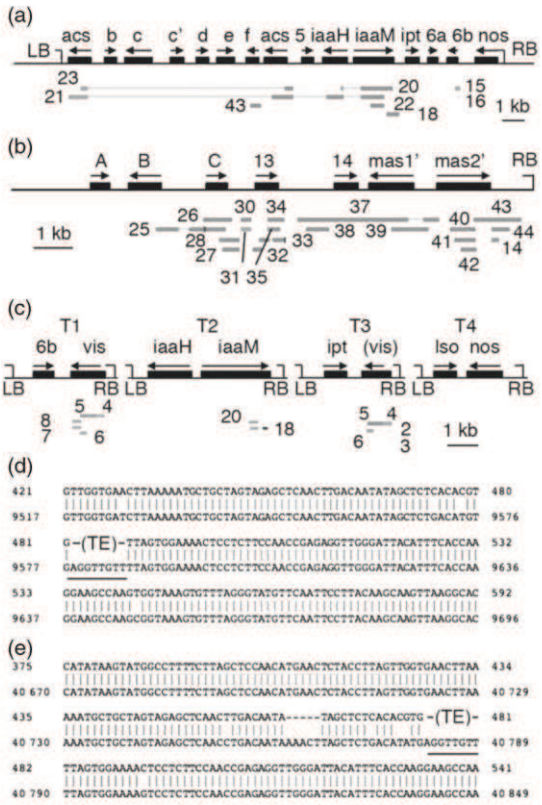


Figure 6. Mapping of TE region contigs from *Nicotiana otophora* on different reference T-DNAs. The numbers refer to the contigs listed in Table S2.
(a) *Agrobacterium tumefaciens* C58 T-DNA (AJ237588.1).
(b) *Agrobacterium rhizogenes* 8196 T-DNA (M60490.1). The left border of this T-DNA is unknown.
(c) *Agrobacterium vitis* T-DNAs T1, T2, T3 and T4 from pTiS4 (CP000637.1). T1–T4 are shown separately; RB, right border; LB, left border.
(d) putative TE insertion site in *Nicotiana tomentosiformis*. The plant sequences from TE contigs AWOL010660148.1 (1–481) and AWOL010282125.1 (1–1413) were combined into a single sequence (top sequence) and aligned with *N. tomentosiformis* contig ASAG01005303.1 (bottom sequence).
(e) Similar alignment as in (d), but with *Nicotiana sylvestris* contig ASAF01131078.1 as bottom sequence. Only part of the alignment is shown. (TE): TE insertion site. An AGGTTGTT sequence (underlined), present in *N. tomentosiformis* and *N. sylvestris* was probably deleted upon insertion of the TE sequences.

T-DNA integration patterns depend on the type of target cells (De Buck *et al.*, 2009). Thus, if spontaneous regeneration of transformed plant cells requires the initial transformation of special plant tissues like roots or stems, this could lead to a selection of specific integration structures. Whatever their origin, these repeats are of great practical interest to reconstruct cT-DNA evolution. At the moment of introduction into the plant DNA the repeats were most probably identical. Provided that no gene conversion occurred and molecular clocks ran at similar rates, repeat

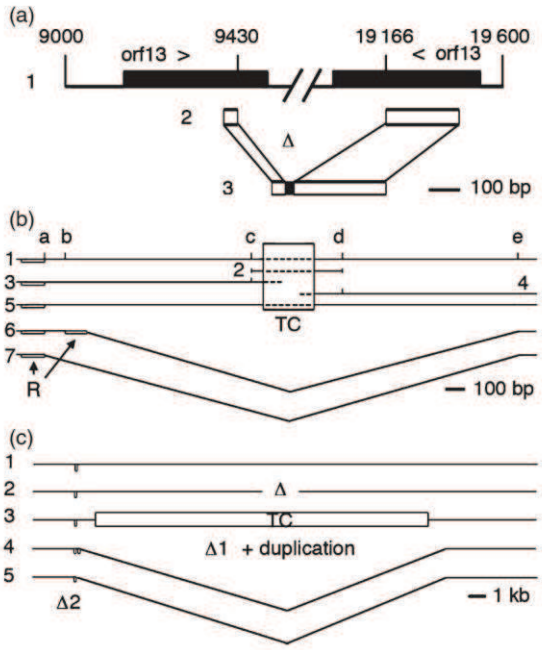


Figure 7. Deletion variants of TA and TC regions.
(a) Deletion of central TA fragment in *Nicotiana tabacum* cv. Basma/Xanthi: 1, partial map of *Nicotiana tomentosiformis* TA region (KJ599826), central part not shown; 2, TA sequences found on the 414-bp *N. tabacum* cv. Basma/Xanthi contig AWOK01769803.1 (Sierro *et al.*, 2014); 3, AWOK01769803.1, unidentified 42-bp sequence connecting the two *orf13* fragments, in black. The deletion removed a 9684-bp fragment between KJ599826 coordinates 9430 and 19166.
(b) TC region deletions in *N. tabacum*: 1, partial map of enlarged TC region with map coordinates (a–e) for different *Nicotiana* sequences (RC, reverse complement) – a, ASAG01044398.1RC (46652), AYMY01287087.1 (6024), AWOJ01378807.1 (9536), AWOK01531567.1 (817); b, ASAG01044398.1RC (46754), AYMY01287087.1 (6144); c, ASAG01044398.1RC (47689), KJ599828 (1); d, ASAG01063776.1RC (5638), KJ599828 (16918); e, ASAG01063776.1RC (6511), AYMY01287087.1 (6262), AWOJ01378807.1 (9536), AWOK01531567.1 (817); 2, *N. tomentosiformis* ITB646 TC region (KJ599827); 3 and 4, ASAG01044398.1RC and ASAG1063776.1RC contigs from *N. tomentosiformis* TW142 (Sierro *et al.*, 2013), used to extend the ITB646 TC region (KJ599828); 5, TC region in *N. tomentosiformis* and *N. otophora*; 6, deleted TC region in *N. tabacum* cv. TN90 (AYMY); 7, deleted TC region in *N. tabacum* cv. K326 (AWOJ) and cv. Basma/Xanthi (AWOK); boxed area, TC (not to scale, dotted lines); R, direct repeat.
(c) Model of TC region evolution: 1, non-modified TC insertion region, found in *N. sylvestris* (Figure 3b); 2, 1997-bp deletion upon insertion of TC; 3, genomes with TC insert; *N. otophora* and *N. tomentosiformis*; 4, *N. tabacum* cv. TN90, deletion of TC plus 995 and 1021 bp on the left and right of TC (Δ1). On the left, a small 126-bp region is duplicated; 5, *N. tabacum* cv. K326 and cv. Basma/Xanthi, additional deletion of 120-bp region between the direct repeats (Δ2). Broken lines indicate connections between the ends of the deleted areas.

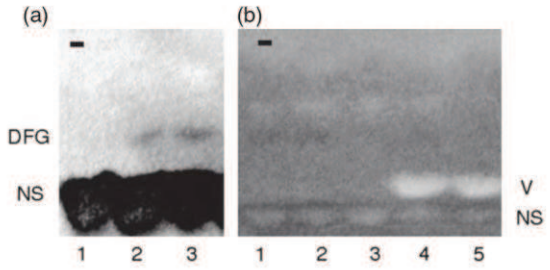


Figure 8. Activity of *Nicotiana tomentosiformis mas2′* and *ocl* genes.
(a) *mas2′* gene. *Nicotiana benthamiana* leaves were transiently transformed with different constructs, and extracts were analyzed by paper electrophoresis and alkaline silver staining: 1, empty vector control; 2, A4 *mas2′* gene; 3, *N. tomentosiformis mas2′* gene; DFG, deoxyfructosylglutamine; NS, neutral sugars.
(b) *ocl* gene from *N. tomentosiformis*, positive control: the vitopine synthase (*vis*) gene from *Agrobacterium vitis* S4. Both were analyzed by transient expression in old and young leaves of *N. benthamiana*, and extracts were analyzed by paper electrophoresis and negative staining. Lanes: 1, negative control; 2–3, *N. tomentosiformis ocl*, old and young leaves; 4–5, S4 *vis*, old and young leaves; V, vitopine; NS, neutral sugars; –, cathode. The *vis* gene yields white spots at the vitopine position; the *ocl* gene does not produce any signal.

divergence values may be used to estimate the relative age of the cT-DNAs. Under these conditions, TC would be the oldest (5.8% divergence), followed by gT (3%), TB (2.6%), TD (1.7%) and TA (1.2%). The value for *Linaria* cT-DNA repeats is about 6% (Matveeva *et al.*, 2012). The cT-DNAs could have been introduced by successive transformation, by hybridization between *Nicotiana* species or subspecies that had separately acquired single cT-DNAs, or by a combination of these routes. Previously, cT-DNAs were reported in 11 additional *Nicotiana* species (Intrieri and Buiatti, 2001). Attempts to obtain cT-DNA sequences from *N. debneyi* and *N. miersii* were unsuccessful in our lab. Intrieri and Buiatti reported *rolC*, *orf13* and *orf14* sequences in *N. otophora* (AF281250.1, AF281243.1 and AF281247.1, respectively), but these do not correspond to the TE region sequences, showing only 72, 72 and 89% identity. *N. cordifolia* was reported to contain *rolB*, *rolC*, *orf13* and *orf14* sequences very similar to those of *N. glauca* (Intrieri and Buiatti, 2001), but an earlier Southern analysis with an *N. glauca*-derived probe (Furner *et al.*, 1986) did not detect such sequences. For future studies it will be essential to sequence the complete cT-DNAs of these additional species, including their insertion sites. Three representative *N. tabacum* cultivars showed a complete TC deletion, including about 1 kb of plant sequences on each side, this deletion could have been present in the *N. tomentosiformis* plant that gave rise to *N. tabacum*, or could have occurred after the *N. tomentosiformis* × *N. sylvestris* hybridization. It has been established that *N. tabacum* lost various parental sequences following the initial hybridization event (Volkov *et al.*, 1999; Lim *et al.*, 2004, 2007; Petit *et al.*, 2007; Renny-Byfield *et al.*, 2011). More data are needed on cT-DNAs of different *N. tomentosiformis* accessions, especially for those most closely related to the *N. tabacum* ancestor (Murad *et al.*, 2002). DNA losses were also observed in synthetic *N. tomentosiformis* × *N. sylvestris* hybrids (Skalicka *et al.*, 2005; Renny-Byfield *et al.*, 2012); it would therefore be interesting to

study the fate of the four T-regions in such hybrids. The central part of the TA region was lost in various members of the oriental tobacco cultivars. The function of the lost cT-DNA sequences may be studied by their re-introduction into the deletion variants, thus the *N. tomentosiformis* TC genes can be studied in *N. tabacum*, whereas the central TA region from *N. tabacum* cv. K326 can be studied in *N. tabacum* cv. Basma/Xanthi. Analysis of the recently released whole-genome shotgun contigs from *N. otophora* (Sierro *et al.*, 2014) revealed yet another cT-DNA, TE. This cT-DNA is composed of 44 partially overlapping contigs and has a size of at least 18 kb. Although its insertion site has been identified it still needs to be assembled. Strikingly, it combines features of *A. vitis*, *A. tumefaciens* and *A. rhizogenes* T-DNAs, and carries *vis*, *6b*, *iaaM*, *iaaH*, *acs*, *rolB*, *rolC*, *orf13*, *orf14*, *mas2′* and *mas1′* sequences. In this study, we found several unusual, highly divergent types of T-DNA genes: *c* and *ocl* on TC, *ags* on TB, *mis* and *orf14* on TA2 and TD, and *6b* on TE. The Orf511 protein potentially encoded by TD-*orf511* has no similarity to other proteins and its function is unknown. The *orf14* and *6b* genes belong to the *plast* gene family, a group of highly divergent T-DNA genes that also includes *orf13*, *rolB* and *rolC* (Levesque *et al.*, 1988). Initially, *plast* genes were considered to be specific for Agrobacterium T-DNAs, but five *plast*-like genes were later identified in the mycorrhizal fungus *Laccaria bicolor* (Mohajjel-Shoja *et al.*, 2011). Recently, we found Plast-like proteins (WP_007538654.1, WP_007538668.1, WP_007538775.1, WP_007538777.1, WP_007538785.1 and WP_007538800.1) in *Rhizobium mesoamericanum* STM3625. Our present data show similarities between TB-*ags* and *Nectria haematococca ags*, TC-*ocl* and *Aspergillus nidulans ocl*, TC-*c* and c-like genes from *Melampsora larici-populina* and *Pestalotiopsis fici*, and strengthen the emerging notion that some Fungi carry T-DNA-like genes. These were possibly acquired by *Agrobacterium*-mediated transformation. Although *Agrobacterium* can transform some fungi under laboratory conditions (De Groot *et al.*, 1998), natural transformation has yet to be demonstrated. The possible transfer of an ancestral *ocl* gene from *Agrobacterium* to *Aspergillus* occurred after this gene had separated from the *ocs* branch, but in the absence of further data it is not possible to say when this happened. Further studies are necessary on the distribution and function of T-DNA-like genes in fungi. The transfer of *ocl* from *Agrobacterium* to *Nicotiana* occurred very recently, as indicated by the low divergence between the *ocl*-associated TB repeats. In general, the high similarities between the repeats of the cT-DNA sequences show that horizontal gene transfers from *Agrobacterium* to plants are very recent. Most *N. tomentosiformis* cT-DNA ORFs are degenerate, and some regions have been deleted. These alterations could have become fixed by negative selection or, in case they were selectively neutral, by progressive gene erosion. Possibly, some cT-DNA genes played a role in the survival of the initial transgenic regenerants, for example by facilitating regeneration or by causing reproductive changes leading to sympatric speciation. The few cT-DNA genes that remained intact may still confer a selective advantage. Reconstruction of *Nicotiana* cT-DNA evolution requires a reliable phylogeny for this group. To establish such a model, we adopted the generally accepted sectional classification of *Nicotiana* (Knapp *et al.*, 2004). Earlier models of cT-DNA evolution (Fründt *et al.*, 1998b; Intrieri and Syono 2001, Suzuki *et al.*, 2002; Mohajjel-Shoja *et al.*, 2011) are presented in Figure 9a, and a model based on the present data is shown in Figure 9b. According to our results, at least six independent transformation events occurred in the *Nicotiana* group (marked 1–6 in Figure 9b): TC (1) was introduced in the ancestor of the *Tomentosae*/*Nicotiana* sections; gT (2) was introduced in a *N. glauca* ancestor from the *Noctiflorae* section. TB (3) and TD (4) were introduced in the ancestor of the *tomentosiformis*/*tabacum*–*kawakamii*–*tomentosa* group, with subsequent deletion of TC in *N. tabacum*. TA (5) was introduced in the ancestor of the *tomentosiformis*/*tabacum*–*kawakamii* group. Finally, TE (6) was introduced in an *N. otophora* ancestor after it had split off from the *tomentosiformis*/*tabacum*–*kawakamii*–*tomentosa* branch. There is some uncertainty about the branching order in the *Tomentosae* section (Clarkson *et al.*, 2004, 2010; Knapp *et al.*, 2004; Murad *et al.*, 2004; Moon *et al.*, 2008; Kelly *et al.*, 2010). The cT-DNA distribution among the members of this group could help to resolve the discrepancies. Our data confirm the tree based on the MADS1/FUL genes from Kelly *et al.*, 2010; showing *N. tomentosa* as a sister to the species pair *N. tomentosiformis*–*N. kawakamii*. It will be interesting to investigate the cT-DNA sequences in *N. setchellii* (Intrieri and Buiatti, 2001) in more detail, as this species belongs to section *Tomentosae* (Knapp *et al.*, 2004). Our present results also provide further insight into the relation between gT from *N. glauca* and the *N. tomentosiformis* cT-DNAs. The TB *orf14-mis* fragment is 96% identical to the corresponding gT sequences, this suggested initially that both resulted from a single transformation event with T-DNA insertions at two different loci (Mohajjel-Shoja *et al.*, 2011); however, the lack of TB in *N. otophora* now argues against the possibility that TB and gT are derived from one transformation event. We postulate therefore that gT and TB were introduced by independent events, but from *Agrobacterium* with very similar *orf14-mis* sequences. Attempts have been made to compare cT-DNA evolution in *Nicotiana* with *Agrobacterium* evolution (Intrieri and Buiatti, 2001); however, frequent horizontal transfer and recombination in *Agrobacterium* make this very difficult. Many intermediate structures that would be essential to understand the origin of the present-day T-DNA structures may have been lost (Otten and De Ruffray, 1994). The

N. tomentosiformis cT-DNAs also show evidence for such recombined T-DNAs. The TC sequences are composed of a large *A. rhizogenes*-like T-DNA fragment and *A. tumefaciens*-like *ocl* and *c* genes, the highly divergent TA *orf14* and *mis* genes are associated with classical *orf8-rolABC-orf13* sequences, classical *mas1′* and *mas2′* genes in TB are linked to an unusual *ags* gene, and a highly divergent *orf14* gene in TD is linked to classical *orf15* sequences. Although TE remains to be assembled, it also shows an unusual mixture of *A. vitis*, *A. tumefaciens* and *A. rhizogenes* T-DNA sequences. The successive accumulation of TC, TB, TD and TA cT-DNA inserts in various *N. tomentosiformis* ancestors is remarkable and unexpected, as each transformation/regeneration event is probably very rare. We speculate that cT-DNA stacking could have been favored by opine synthesis in the transformants. Opines could have attracted other *Agrobacterium* strains, leading to further transformation–regeneration events. As an example, the *tomentosiformis–kawakamii* ancestor carrying TB may have synthesized enough mikimopine to favor the growth of other mikimopine-degrading strains. A TA-carrying strain with *mis* genes (and, most likely, genes for the degradation of mikimopine) could have been among those strains. The two intact *N. tomentosiformis* opine synthesis genes *ocl* and *mas2′* were tested for activity. We could not detect an enzymatic *ocl* product under conditions in which the gene was highly expressed, and in which a control vitopine synthase gene led to strong opine synthesis. Interestingly, *TOF-mas2′* produced similar quantities of DFG as *A4-mas2′* from the A4 TR-DNA. The *N. glauca mis* genes encode mikimopine synthesis, but no mikimopine has so far been found in this *Nicotiana* species (Suzuki *et al.*, 2002). Further studies are required in order to test DFG synthesis in *N. tomentosiformis*. As normal plants can synthesize DFG under aging conditions, it will be necessary to silence *mas2′* to demonstrate its eventual role in DFG synthesis. Functional analysis of cT-DNA genes in *Nicotiana* and their role in the evolution of this group may become a rich area of investigation for the future.

EXPERIMENTAL PROCEDURES

Plant material

The following plant material was used: *N. tomentosiformis* (ITB646), *N. sylvestris* (ITB626), *N. kawakamii* (ITB642), *N. tomentosa* (ITB1015) and *N. otophora* (ITB643). Seeds of these plants can be obtained from the Institut du Tabac de Bergerac (Imperial Tobacco, http://www.imperial-tobacco-bergerac.com), upon request.

DNA preparation, sequencing and contig construction

Nicotiana tomentosiformis (ITB646) and *N. sylvestris* (ITB626) genomic DNAs were prepared from leaves of sterile plantlets following Dellaporta (1983) and purified by CsCl centrifugation. *N. tomentosiformis* DNA was sequenced by the BGI Company (Hong Kong, http://bgitechsolutions.com) as follows: DNA was randomly fragmented by sonication and after the addition of adaptors, fragments of 200 and 500 bp were sequenced by the Illumina technique (Hiseq 2000). The pre-processing of the *N. tomentosiformis* sequence data was performed by the BGI Company as follows: first, the adaptor pollutions in the reads were deleted, and then the reads that contained more than 50% low-quality bases (quality value ≤ 5) were removed. *N. sylvestris* (ITB626) DNA was sequenced on an Illumina Hi-Seq 2000 (2×100 bp) by DNAVision (http://www.dnavision.com). In order to trim sequencing adaptors from reads, the following procedure was applied to each read: trimming the tail of the sequence that matched to PCR primers with a maximum mismatch of 20%; trimming the tail of the sequence that matched to the paired end sequencing adaptors with a maximum mismatch of 20%. Reads with more than 5% unknown nucleotides and low-quality reads with more than 20% bases of quality value ≤ 10 were removed. For both *N. tomentosiformis* and *N. sylvestris* reads, CLC GENOMICS WORKBENCH 6.5.0 (http://www.clcbio.com) was used to assemble reference genomes with the default parameters, excepting: Mapping mode = Map reads back to contigs; Update = Yes; Minimum contig length = 500; Length fraction = 0.9; Similarity fraction = 1.0.

BLAST analysis

BLAST analysis (Altschul *et al.*, 1990) on assembled contigs was performed by BLASTN (NCBI, http://www.ncbi.nlm.nih.gov; or locally on the CLC GENOMICS WORKBENCH 6.5.0, with default parameters) using reference T-DNAs (Table 1) as queries. After assembly of the contigs with cT-DNA sequences into a single T-region, and verification by PCR amplification and sequencing of selected regions, T-regions were used as queries in BLASTN analyses at the NCBI site, using the nucleotide collection (nr/nt), the genomic survey sequences (GSS) and the WGS contigs as databases. «Optimized for» option: somewhat similar sequences (BLASTN). The use of WGS was limited by: chosing WGS projects ASAG (*N. tomentosiformis*), AWOL (*N. otophora*), AWOK (*N. tabacum* cv. Basma/Xanthi), AWOJ (*N. tabacum* cv. K326), AYMY (*N. tabacum* cv. TN90), ASAF (*N. sylvestris*) and CBMM (*N. benthamiana*). Algorithm parameters for BLASTN: max target sequences, 100; short queries, automatically adjust parameters for short input sequences; expect threshold, 10; word size, 11; max matches in a query range, 0; match/mismatch scores; 2, –3, gap costs, existence; 5, extension; 2, filter; low complexity regions; mask, mask for look-up table only. Only alignments with *E* values lower than 0.001 were retained. For protein analysis we used the iterative PSI-BLAST analysis (Altschul *et al.*, 1997), using putative cT-DNA-encoded proteins as queries and the non-redundant protein sequences (nr) as our database, with the following algorithm parameters for PSI-BLAST: max target sequences, 500; short queries; automatically adjust parameters for short input sequences; expect threshold, 10; word size, 3; max matches in a query range, 0; matrix, BLOSUM62; gap costs; existence, 11; extension, 1; compositional adjustments; conditional compositional score matrix adjustment; filter, none; mask, none; PSI-BLAST threshold, 0.005; pseudocount, 0.

Attribution of polymorphisms to repeated regions

Repeated regions of TA, TB and TC could only partially be separated by computer analysis. Copy-specific regions were therefore amplified by PCR and sequenced. The following primers were used (AP, amplification primer; SP, sequencing primer; primer sequences are given in Table 2; for fragment positions on maps see Figures 1b, 2a and 3a): TA, a1; AP, 3017, 1725; SP, 3017, 1725,

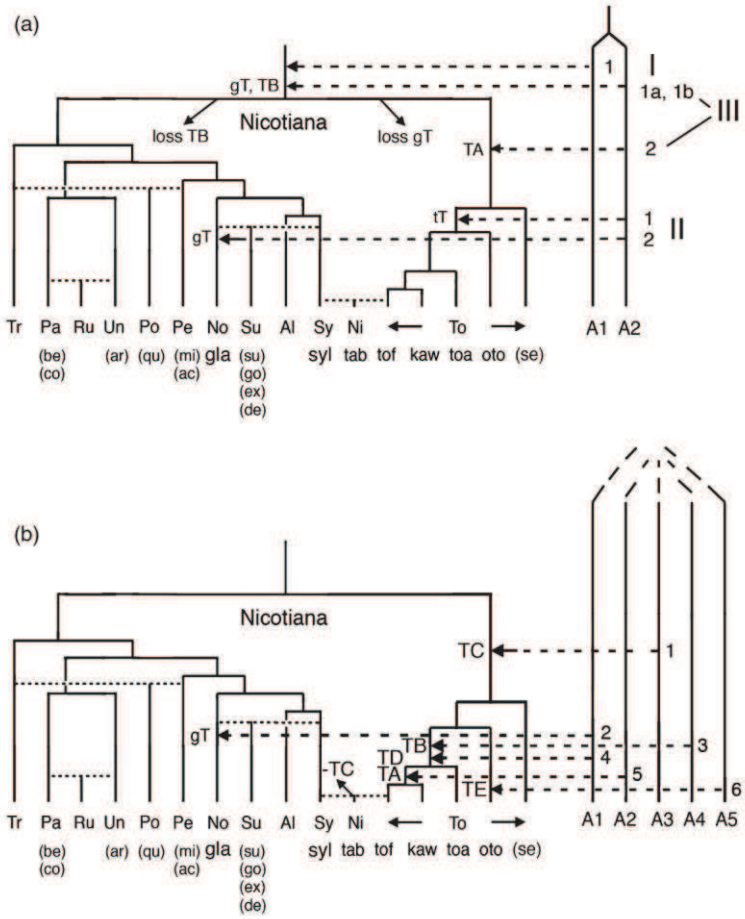


Figure 9. Evolutionary models for cT-DNAs in *Nicotiana*.
(a) Earlier models based on results by Fründt *et al.* (1998a), Intrieri and Buiatti (2001), Suzuki *et al.* (2002) and Mohajjel-Shoja *et al.* (2011). This scheme shows three possibilities: I, insertion of a unique cT-DNA (I-1) that later diverged and became part of *Nicotiana tabacum*, *Nicotiana tomentosiformis* and *Nicotiana glauca*; II, two independent insertions, one (II-1) leading to tT of *N. tabacum* and *N. tomentosiformis*, the other (II-2) to gT of *N. glauca* (Suzuki *et al.*, 2002 favor possibility II); III, two transformation events, one (III-1) leading to two different insertions in the initial transformant (III-1a and III-1b), with segregation of the two inserts leading to gT and TB, the second (III-2) leading to TA (Mohajjel-Shoja *et al.*, 2011).
(b) Model based on the present data. Insertion 1 of TC is followed by insertion 2 of gT, insertion 3 of TB, insertion 4 of TD and insertion 5 of TA. The relative time of insertion 6 (TE) is unknown but occurred after the *tomentosiformis*/*tabacum*-*kawakamii*-*tomentosa* branch separated from the *otophora* branch because *N. sylvestris*, *N. tomentosiformis* and *N. tabacum* have highly similar, unmodified TE insertion sites. Insertions 2 and 4 are from *A. rhizogenes* strains with similar *14-mis* sequences. TC was fully deleted in the *tabacum* branch (–TC), TA was partially deleted in some of the *N. tabacum* cultivars. Abbreviations for sections: Al, *Alatae*; Ni, *Nicotiana*; No, *Noctiflorae*; Pa, *Paniculatae*; Pe, *Petunioides*; Po, *Polydicliae*; Ru, *Rusticae*; Su, *Suaveolentes*; Sy, *Sylvestres*; To, *Tomentosae*; Tr, *Trigonophyllae*; Un, *Undulatae* (section *Repandae* not shown). Species abbreviations: ac, *acuminata*; ar, *arentsii*; be, *benavidesii*; co, *cordifolia*; de, *debneyi*; ex, *exigua*; gla, *glauca*; go, *gossei*; kaw, *kawakamii*; mi, *miersii*; oto, *otophora*; qu, *quadrivalvis*; se, *setchellii*; su, *suaveolens*; syl, *sylvestris*; tab, *tabacum*; toa, *tomentosa*; tof, *tomentosiformis*. Dotted horizontal lines: interspecific hybridization events. Names within brackets indicate species for which only partial data have been reported (Intrieri and Buiatti, 2001). Dotted lines at the top of the *Agrobacterium* tree indicate the unknown relationships between the different strains. A1–A5, different *Agrobacterium* types. Figures are not drawn to scale but indicate relative times of occurrence.

1709, 1710; TA, a2; AP, 1710, 2967; SP, 1710, 2967, 1725; TA, a3; AP, 3069, 3071; SP, 3069, 3071; TA, a4; AP, 1711, 3012; SP, 3012, 1711, 1709, 1710; TB, b1; AP, 2327, 2347; SP, 2327, 2347, 2348, 3081, 1846; TB, b2; AP, 1917, 2370; SP, 1917, 2370, 1841, 1843, 1848, 1849; TC, c1; AP, 2986, 2376; SP, 2986, 2376.

Gene constructs

For constitutive expression, *A4-mas2′*, *TOF-mas2′*, *ocl* and *S4-vis* coding regions were cloned by PCR in a 2x*35S* promoter cassette and subcloned into the pBI121.2 binary vector (Jefferson *et al.*, 1987), as described by Otten and Helfer (2001). All constructs were sequenced before subcloning. Binary vectors with constructs were introduced in *Agrobacterium* helper strain LBA4404 (Hoekema *et al.*, 1983).

Transient expression and stable transformation

Transient expression was carried out by agroinfiltration of *N. benthamiana* leaves (Yang *et al.*, 2000). Samples were extracted for opine detection 5 days after infection.

Opine analysis

Plant material was frozen in liquid nitrogen, ground with mortar and pestle, and extracted with 70% ethanol (1 : 1, w/v). After centrifugation for 5 min at 10 000 ***g***, the supernatant was either dried in a SpeedVac and redissolved in 70% ethanol in one-tenth of the initial volume or directly analyzed. Opines were separated by paper electrophoresis in acidic acid : formic acid : water (60 : 30 : 910, v/v, pH 1.6) at 400 V, and revealed by staining with alkaline silver nitrate (Dessaux *et al.*, 1992) or by reversed silver nitrate staining after dipping the paper in a 0.01% glucose solution in acetone (Szegedi *et al.*, 1988).

ACKNOWLEDGEMENTS

K.C. was supported by a doctoral grant from the Chinese Scholarship Council. E.Sz. was supported by National Scientific Foundation (OTKA) grant no. K-83121. We thank the IBMP gardeners for providing plants and M. Alouia for the sequencing service. We thank Professor David Gubb for his critical reading of the article.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Sequences of empty TA insertion sites in *Nicotiana otophora* and *Nicotiana tomentosa*.

Figure S2. Alignment of cT-DNAs from *Nicotiana tomentosiformis* ITB646, with contig sequences from *N. tomentosiformis* TW142.

Figure S3. TA region sequences in other *Nicotiana* species.

Figure S4. TB region sequences in other *Nicotiana* species.

Figure S5. TC region sequences in other *Nicotiana* species.

Figure S6. TD region sequences in other *Nicotiana* species.

Table S1. Comparison of TA, TB, TC and TD from *Nicotiana tomentosiformis* ITB646 and *N. tomentosiformis* TW142.

Table S2. TE region contigs from *Nicotiana otophora*.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Aoki, S. (2004) Resurrection of an ancestral gene: functional and evolutionary analyses of the *Ngrol* genes transferred from *Agrobacterium* to *Nicotiana*. *J. Plant. Res.* **117**, 329–337.

Aoki, S. and Syono, K. (1999a) Horizontal gene transfer and mutation: *ngrol* genes in the genome of *Nicotiana glauca*. *Proc. Natl Acad. Sci. USA*, **96**, 13229–13234.

Aoki, S. and Syono, K. (1999b) Synergistic function of *rolB, rolC, ORF13* and *ORF14* of TL-DNA of *Agrobacterium rhizogenes* in hairy root induction in *Nicotiana tabacum*. *Plant Cell Physiol.* **40**, 252–256.

Aoki, S. and Syono, K. (1999c) Function of Ngrol genes in the evolution of Nicotiana glauca: conservation of the function of NgORF13 and NgORF14 after ancient infection by an Agrobacterium rhizogenes-like ancestor. *Plant Cell Physiol.* **40**, 222–230.

Aoki, S., Kawaoka, A., Sekine, M., Ichikawa, T., Fujita, T., Shinmyo, A. and Syono, K. (1994) Sequence of the cellular T-DNA in the untransformed genome of *Nicotiana glauca* that is homologous to ORFs 13 and 14 of the Ri plasmid and analysis of its expression in genetic tumours of *N. glauca* x *N. langsdorffii*. *Mol. Gen. Genet.* **243**, 706–710.

Clarkson, J.J., Knapp, S., Garcia, V.F., Olmstead, R.G., Leitch, A.R. and Chase, M.W. (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. *Mol. Phylogenet. Evol.* **33**, 75–90.

Clarkson, J.J., Kelly, L.J., Leitch, A.R., Knapp, S. and Chase, M.W. (2010) Nuclear glutamine synthetase evolution in *Nicotiana*: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. *Mol. Phylogenet. Evol.* **55**, 99–112.

Costantino, P., Capone, I., Cardarelli, M., De Paolis, A., Mauro, M.L. and Trovato, M. (1994) Bacterial plant oncogenes: the *rol* genes' saga. *Genetica*, **94**, 203–211.

De Buck, S., Podevin, N., Nolf, J., Jacobs, A. and Depicker, A. (2009) The T-DNA integration pattern in Arabidopsis transformants is highly determined by the transformed target cell. *Plant J.* **60**, 134–145.

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

Dellaporta, S.J. (1983) A plant minipreparation: version II. *Plant Mol. Biol. Rep.* **1**, 19–21.

Dessaux, Y., Petit, A. and Tempé, J. (1992) Opines in *Agrobacterium* biology. In *Molecular Signals in Plant-Microbe Communications*. (Verma, D.P.S., ed.). Boca Raton, FL: CRC Press, pp. 109–136.

Fründt, C., Meyer, A.D., Ichikawa, T. and Meins, F. (1998a) A tobacco homologue of the Ri-plasmid *orf13* gene causes cell proliferation in carrot root discs. *Mol. Gen. Genet.* **259**, 559–568.

Fründt, C., Meyer, A., Ichikawa, T. and Meins, F. (1998b) Evidence for the ancient transfer of Ri-plasmid T-DNA genes between bacteria and plants. In *Horizontal gene transfer*. (Syvanen, M. and Kado, C., eds). London: Chapman-Hall, pp. 94–106.

Furner, I.J., Huffman, G.A., Amasino, R.M., Garfinkel, D.J., Gordon, M.P. and Nester, E.W. (1986) An *Agrobacterium* transformation in the evolution of the genus *Nicotiana*. *Nature*, **319**, 422–427.

Goodspeed, T.H. (1954) The genus *Nicotiana*. *Chron. Bot.* **16**, 102–135.

Hansen, G., Larribe, M., Vaubert, D., Tempé, J., Biermann, B.J., Montoya, A.L., Chilton, M.-D. and Brevet, J. (1991) *Agrobacterium rhizogenes* pRi8196 T-DNA: mapping and DNA sequencing of functions involved in mannopine synthesis and hairy root differentiation. *Proc. Natl. Acad. Sc. USA*, **88**, 7763–7767.

Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J. and Schilperoort, R.A. (1983) A binary vector system based on separation of vir- and T-region of the *Agrobacterium* Ti plasmid. *Nature*, **303**, 179–180.

Ichikawa, T., Ozeki, Y. and Syono, K. (1990) Evidence for the expression of the *rol* genes of *Nicotiana glauca* in genetic tumors of *N. glauca* x *N. langsdorffii*. *Mol. Gen. Genet.* **220**, 177–180.

Intrieri, M.C. and Buiatti, M. (2001) The horizontal transfer of *Agrobacterium rhizogenes* genes and the evolution of the genus *Nicotiana*. *Mol. Phylogenet. Evol.* **20**, 100–110.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.

Kelly, L.J., Leitch, A.R., Clarkson, J.J., Hunter, R.B., Knapp, S. and Chase, M.W. (2010) Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). *Mol. Biol. Evol.* **27**, 781–799.

Kenton, A., Parokonny, A.S., Gleba, Y.Y. and Bennett, M.D. (1993) Characterization of the *N. tabacum* L. genome by molecular cytogenetics. *Mol. Gen. Genet.* **240**, 159–169.

Knapp, S., Chase, M.W. and Clarkson, J.J. (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). *Taxon*, **53**, 73–82.

Lemcke, K. and Schmülling, T. (1998) Gain of function assays identify non-rol genes from *Agrobacterium rhizogenes* TL-DNA that alter plant morphogenesis or hormone sensitivity. *Plant J.* **15**, 423–433.

Levesque, H., Delepelaire, P., Rouzé, P., Slightom, J. and Tepfer, D. (1988) Common evolutionary origin of the central portion of the Ri TL-DNA of *Agrobacterium rhizogenes* and the Ti T-DNAs of *Agrobacterium tumefaciens*. *Plant Mol. Biol.* **11**, 731–744.

Lim, K.Y., Skalicka, K., Koukalova, B., Volkov, R.A., Matyasek, R., Hemleben, V., Leitch, A.R. and Kovarik, A. (2004) Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. *Genetics*, **166**, 1935–1946.

Lim, K.Y., Kovarik, A., Matyasek, R., Chase, M.W., Clarkson, J.J., Grandbastien, M.A. and Leitch, A.R. (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytol.* **175**, 756–763.

Matveeva, T.V., Bogomaz, D.I., Pavlova, O.A., Nester, E.W. and Lutova, L.A. (2012) Horizontal gene transfer from genus *Agrobacterium* to the plant *Linaria* in nature. *Mol. Plant-Microbe Interact.* **25**, 1542–1551.

Meyer, A.D., Ichikawa, T. and Meins, F. (1995) Horizontal gene transfer: regulated expression of a tobacco homologue of the *Agrobacterium rhizogenes rolC* gene. *Mol. Gen. Genet.* **249**, 265–273.

Mohajjel-Shoja, H., Clément, B., Perot, J., Alioua, M. and Otten, L. (2011) Biological activity of the *Agrobacterium rhizogenes*-derived *trolC* gene of *Nicotiana tabacum* and its functional relation to other *plast* genes. *Mol. Plant-Microbe Interact.* **24**, 44–53.

Moon, H.S., Nicholson, J.S. and Lewis, R.S. (2008) Use of transferable *Nicotiana tabacum* L. microsatellite markers for investigating genetic diversity in the genus *Nicotiana*. *Genome*, **51**, 547–559.

Murad, L., Lim, K.Y., Christopodulou, V., Matyasek, R., Lichtenstein, C., Kovarik, A. and Leitch, A. (2002) The origin of tobacco's T genome is traced to a particular lineage within *Nicotiana tomentosiformis* (Solanaceae). *Am. J. Bot.* **89**, 921–928.

Murad, L., Bielawski, J.P., Matyasek, R., Kovarik, A., Nichols, R.A., Leitch, A.R. and Lichtenstein, C.P. (2004) The origin and evolution of geminivirus-related DNA sequences in *Nicotiana*. *Heredity*, **92**, 352–358.

Otten, L. and De Ruffray, P. (1994) *Agrobacterium vitis* nopaline Ti plasmid pTiAB4: relationship to other Ti plasmids and T-DNA structure. *Mol. Gen. Genet.* **245**, 493–505.

Otten, L. and Helfer, A. (2001) Biological activity of the *rolB*-like 5′ end of the A4-*orf8* gene from the *Agrobacterium rhizogenes* TL-DNA. *Mol. Plant-Microbe Interact.* **14**, 405–411.

Otten, L., Salomone, J.-Y., Helfer, A., Schmidt, J., Hammann, P. and De Ruffray, P. (1999) Sequence and functional analysis of the left-hand part of the T-region from the nopaline-type Ti plasmid, pTiC58. *Plant Mol. Biol.* **41**, 765–776.

Petit, M., Lim, K.Y., Julio, E., Poncet, C., Dorlhac de Borne, F., Kovarik, A., Leitch, A.R., Grandbastien, M.A. and Mhiri, C. (2007) Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). *Mol. Genet. Genomics*, **278**, 1–15.

Ren, N. and Timko, M.P. (2001) AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. *Genome*, **44**, 559–571.

Renny-Byfield, S., Chester, M., Kovařík, A. *et al.* (2011) Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. *Mol. Biol. Evol.* **28**, 2843–2854.

Renny-Byfield, S., Kovalik, A., Chester, M., Nichols, R.A., Macas, J., Novak, P. and Leitch, A.R. (2012) Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. *PLoS One* **7**, e36963.

Riechers, D.E. and Timko, M.P. (1999) Structure and expression of the gene family encoding putrescine N-methyltransferase in *Nicotiana tabacum*: new clues to the evolutionary origin of cultivated tobacco. *Plant Mol. Biol.* **41**, 387–401.

Sierro, N., Battey, J.N.D., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M.C. and Ivanov, N.V. (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* **14**, R60.

Sierro, N., Battey, J.N.D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M.C. and Ivanov, N.V. (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* **5**, 3833.

Skalicka, K., Lim, K.Y., Matyasek, R., Matzke, M., Leitch, A.R. and Kovarik, A. (2005) Preferential elimination of repeated DNA sequences from the paternal *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. *New Phytol.* **166**, 291–303.

Suzuki, K., Yamashita, I. and Tanaka, N. (2002) Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution. *Plant J.* **32**, 775–787.

Szegedi, E., Csako, M., Otten, L. and Koncz, C.S. (1988) Opines in Crown gall tumours induced by biotype 3 isolates of *Agrobacterium tumefaciens*. *Physiol. Mol. Plant Pathol.* **32**, 237–247.

Volkov, R.A., Borisjuk, N.V., Panchuk, I.I., Schweizer, D. and Hemleben, V. (1999) Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. *Mol. Biol. Evol.* **16**, 311–320.

White, F., Garfinkel, D., Huffman, G.A., Gordon, M. and Nester, E.W. (1983) Sequences homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature*, **301**, 348–350.

Yang, Y., Li, R. and Qi, M. (2000) *In vivo* analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *Plant J.* **22**, 543–551.

Table S1. Comparison of TA, TB, TC and TD sequences from *N. tomentosiformis* ITB646 with contig sequences from *N. tomentosiformis* (TW142) reported by Sierro *et al*. 2013. Size of TW142 contigs, number of identities and gaps are indicated. TC-6 belongs to TA.

cT-DNA	Number in Fig. S1	Ntom_contig (TW142)	Length contig (bp)	identities	gaps
TA	1	11442	79974	7484/7499	0
	2	121015	4468	3904/3936	3
	3	61177	13341	2104/2114	9
	4	214639	1335	1333/1335	0
	5	214471	1205	1203/1205	0
	6	214369	1147	1146/1147	0
	7	213549	867	867/867	0
	8	212858	744	743/744	0
	9	212752	731	730/731	0
	10	164652	772	749/772	19
	11	164205	665	665/665	0
	12	210769	551	550/551	0
	13	208989	468	467/468	0
	14	208575	454	454/454	0
	15	208574	454	453/454	0
TB	1	39652	33077	14332/14451	73
	2	214372	1150	1148/1150	0
	3	214268	1099	1096/1105	6
	4	208936	466	462/466	2
	5	208862	463	459/464	1
	6	205336	377	376/377	0
	7	181320	223	222/223	0
	8	181321	223	221/223	0
TC	1	44398	58632	8878/9003	17
	2	63776	31931	5600/5649	12
	3	175608	211	211/211	0
	4	173099	207	206/207	0
	5	173095	207	204/207	0
	6	213549	867	405/531	7
TD	1	37704	24314	14726/14803	21
	2	37703	27578	6757/6794	21
	3	195780	277	277/277	0

Table S2. TE region contigs from *N. otophora*

The T-DNA and cT-DNA sequences from Table 1 (numbers 1-13) were used as queries in a blastn analysis using the *N. otophora* contig collection from the wgs AWOL project as database. Thirty-seven contigs with similarity to T-DNA but differing from the TC sequences were retrieved and used to search for overlapping AWOL sequences, yielding seven additional contigs (1, 9-13 and 24). Contigs 1 and 44 carry plant sequences that correspond to the TE insertion site.

N°	AWOL contig	size (bp)	1	2	3	4	5	6	7	8	9	10	11	12	13	overlap with:
1	010660148.1	507														2,3
2	010173010.1	433													x	4
3	010172804.1	433													x	4
4	011050915.1	264													x	2,3,5
5	010594028.1	722													x	4,6
6	010991983.1	497													x	5,7,8
7	011090788.1	289													x	6,8,9
8	011090789.1	289													x	6,7,9
9	010851895.1	201														7,8,10,11
10	010864716.1	203														9,11,12
11	010864715.1	203														9,10,12
12	010185183.1	483														10,11,13,14
13	010974325.1	232														12,14
14	010609991.1	582				x					x					12,13,15,16,43,44

15	010609990.1	2366									x			x					14,16,17,18,19
16	010392196.1	557									x			x					14,15
17	010770268.1	391																	15,18,19
18	010408703.1	1035					x				x			x	x				15,17,18,19,20
19	010850947.1	200																	17,18
20	010502717.1	2034					x				x			x	x				18,21,22
21	010502718.1	3771				x	x				x			x					20,22,23,24
22	010662377.1	1471					x				x			x					20,21,24
23	010199038.1	579												x					21,24,25,26
24	011101811.1	298																	22,23
25	010654058.1	1861	x	x	x	x		x			x								23,24,26,27,28
26	010544408.1	730	x	x	x	x		x			x								23,24,25,27,28,29
27	010544409.1	422	x	x	x	x	x					x							25,26,28,29,30
28	010654059.1	504	x	x	x		x				x	x							25,26,27,29,30
29	011080209.1	282	x	x				x			x								26,27,28,30
30	010135827.1	349	x	x	x	x		x			x	x							27,28,29,31,32
31	010763171.1	307	x	x	x	x		x			x	x							30,32
32	011107188.1	303	x	x	x	x		x			x	x							30,31,33
33	010452552.1	627	x	x	x	x	x				x	x							32,34,35
34	010261697.1	479		x	x	x	x				x	x							33,35,36
35	011115268.1	310		x	x	x	x				x	x							33,34
36	010950854.1	225										x							37
37	010717398.1	3134	x	x	x	x		x	x	x	x	x							36,38,39
38	010730224.1	542	x	x	x	x		x			x	x							37
39	010717399.1	952			x				x	x									37
40	010796941.1	649			x				x	x									41,42,43
41	010426942.1	519			x				x	x									40,42,43
42	010152082.1	378			x				x	x									40,41,43
43	010282125.1	2589			x				x	x				x					14,40,41,42,44
44	010282126.1	453	x		x					x									14,43

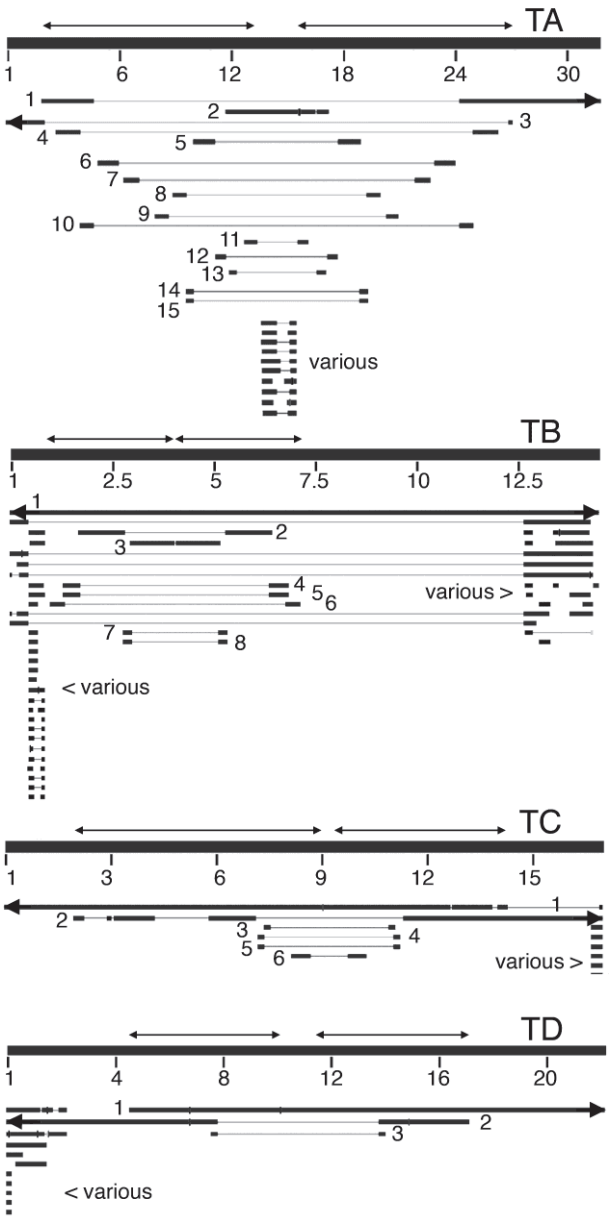


Figure S1. Alignment of cT-DNAs from *N. tomentosiformis* ITB646 with contig sequences from *N. tomentosiformis* TW142 (Sierro *et al*., 2013)

TA to TD were blasted against the TW142 genomic sequences from the whole-genome shotgun contigs (wgs) collection of the NCBI data bank (Sierro *et al*., 2013). The highly similar contigs from repeated regions outside the cT-DNAs were artificially reduced in number and are indicated by "various". cT-DNA repeats are indicated by double-headed arrows. Single arrows indicate extensions of the cT-DNA regions. The TW142 contig numbers indicated on the maps are detailed in Table S1. TA: covered by 15 TW142 contigs. TB: a single TW142 contig covers the complete region, in addition, seven small contigs cover parts of TB. TC: two TW142 contigs overlap, four additional ones cover small regions. TD: two TW142 contigs overlap, an additional small one covers a small part.

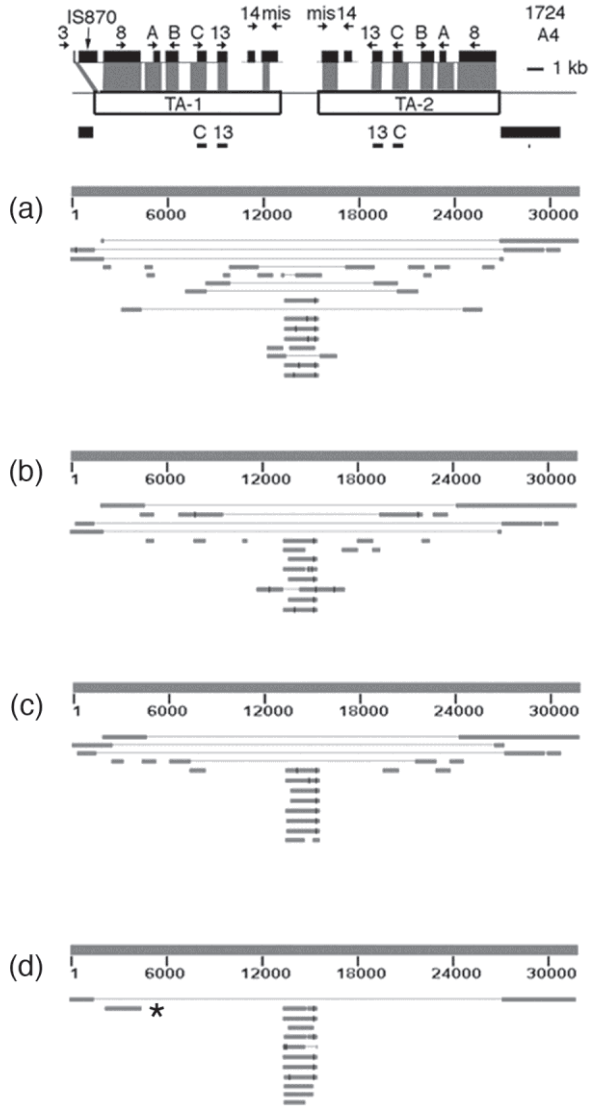


Figure S2. TA region sequences in other *Nicotiana* species
The *N. tomentosiformis* TA region was used as a query against genomic contig collections from different *Nicotiana* species. (a) *N. tabacum* cv. K326. (b) *N. tabacum* cv. TN90. (c) *N. tabacum* cv. Basma/Xanthi. (d) *N. otophora*. Contigs are not detailed. Asterisk: contig from the *N. otophora* TC region with similarity to TA sequences.

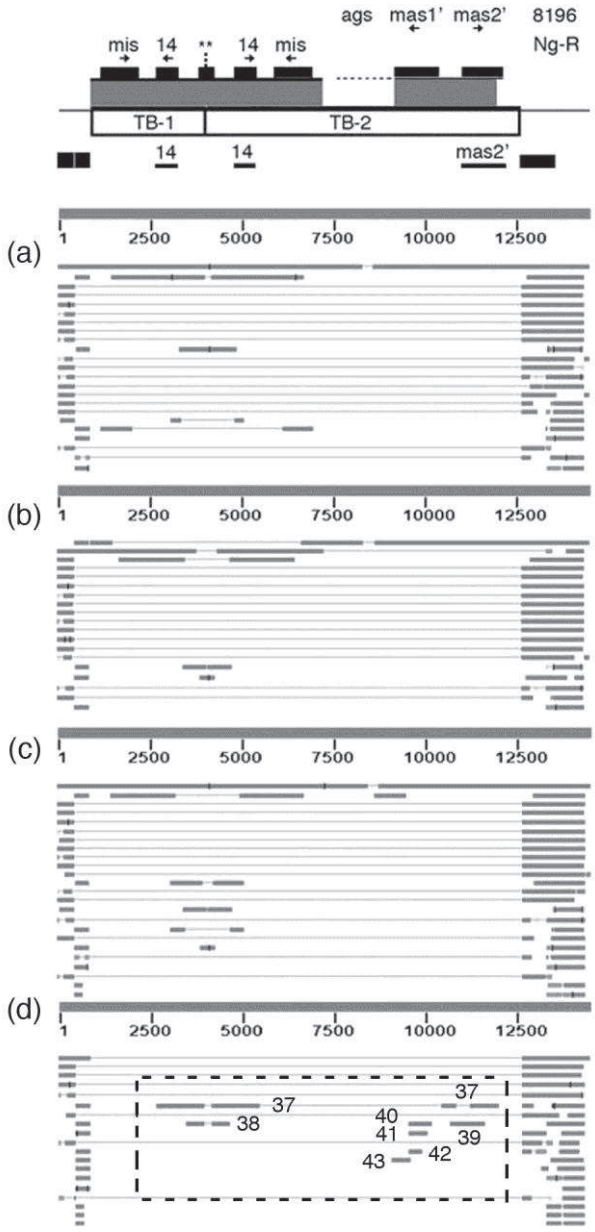


Figure S3. TB region sequences in other *Nicotiana* species
The *N. tomentosiformis* TB region was used as a query against genomic contig collections from different *Nicotiana* species. (a) *N. tabacum* cv. K326. (b) *N. tabacum* cv. TN90. (c) *N. tabacum* cv. Basma/Xanthi. (d) *N. otophora*. Contigs are not detailed. Boxed region in (d): TE contigs with similarity to TB sequences, contig numbers refer to those in Table S2 and Figure 6a.

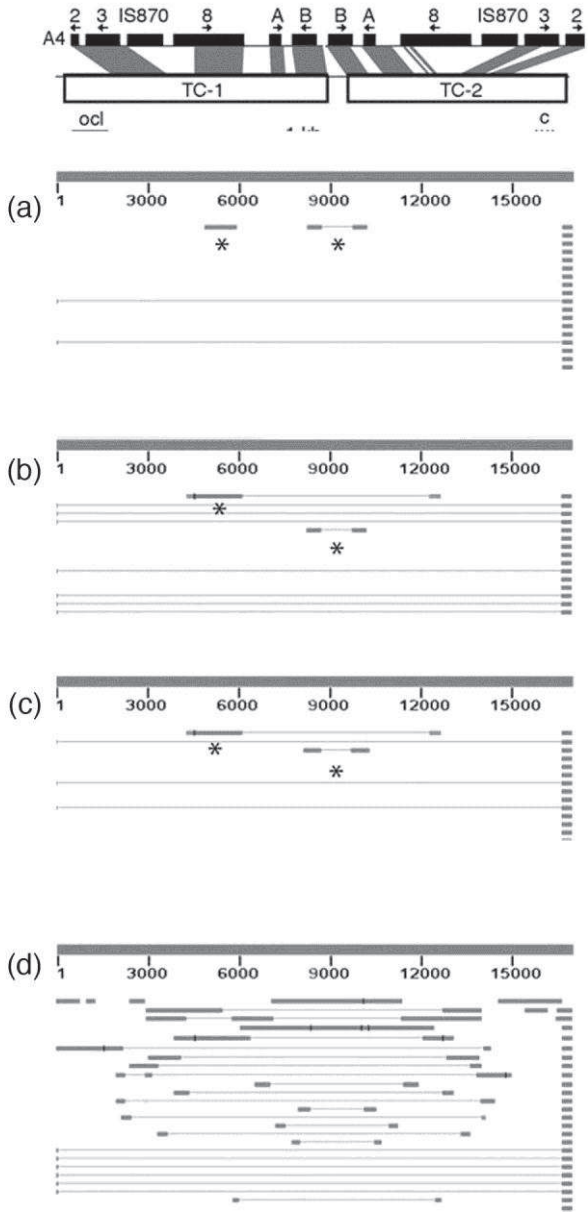


Figure S4. TC region sequences in other *Nicotiana* species
The *N. tomentosiformis* TC region was used as a query against genomic contig collections from different *Nicotiana* species. (a) *N. tabacum* cv. K326. (b) *N. tabacum* cv. TN90. (c) *N. tabacum* cv. Basma/Xanthi. (d) *N. otophora*. Contigs are not detailed. Contigs with asterisks belong to the *N. tabacum* TA region. TC is present in *N. otophora* and absent in *N. tabacum*.

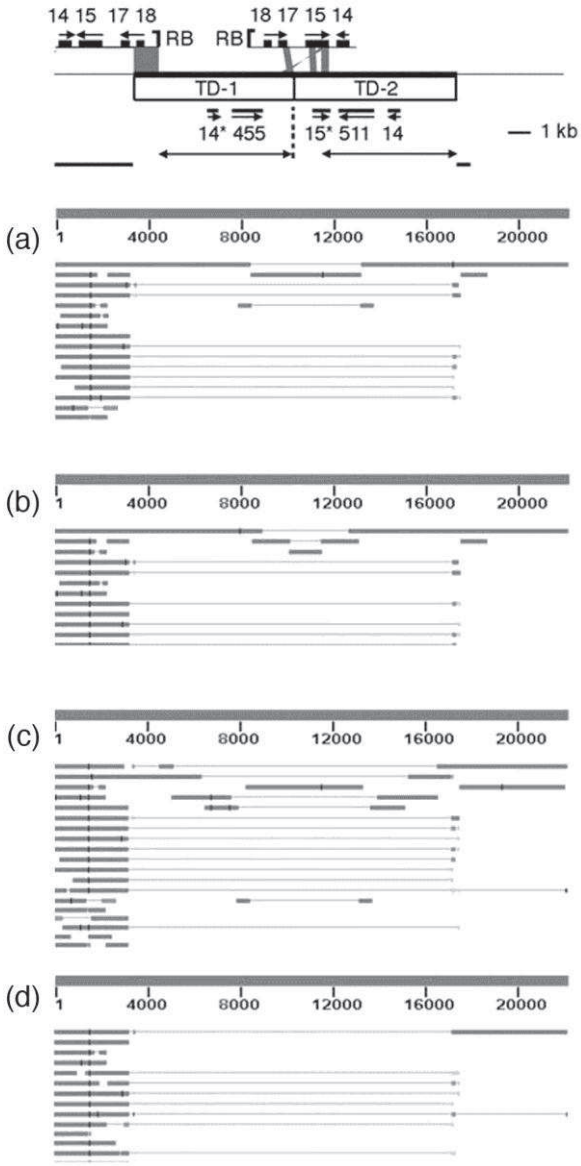


Figure S5. TD region sequences in other *Nicotiana* species
The *N. tomentosiformis* TD region was used as a query against genomic contig collections from different *Nicotiana* species. (a) *N. tabacum* cv. K326. (b) *N. tabacum* cv. TN90. (c) *N. tabacum* cv. Basma/Xanthi. (d) *N. otophora*. Contigs are not detailed. TD is absent in *N. otophora.*

II. Additional data

Since the publication of this work, I obtained additional data on four aspects of the Nicotiana cT-DNAs. 1. The expression levels of all intact ORFs from *N. tomentosiformis* cT-DNAs. 2. A search for cT-DNAs in *N. setchellii.* 3. Functional analysis of the potentially active TE-*6b* gene from *N. otophora*. 4. Further studies on the *mas2'* gene from *N. tomentosiformis* and *N. tabacum*.

II.1 Expression levels of ORFs from *N. tomentosiformis* cT-DNAs

Only a few ORFs from the *N. tomentosiformis* cT-DNAs are intact: TA-*rolC*, TA-*orf13*, TB-*orf14*, TB-*mas2'*, TC-*ocl*, TD-*511* and TD-*orf14*. I tested their expression levels in *N. tomentosiformis* leaves by qPCR. For genes TA-*rolC*, TA-*orf13*, TB-*orf14* and TD-*511*, there are two copies each. In the qPCR experiments, I did not try to distinguish the individual genes. The data are shown in figure 15. Most of these expression levels are very low compared to the reference gene *EF2* from *N. tabacum* (GenBank: AJ299248) (*EF2* value considered as 1). The levels of TA-*orf13*, TB-*mas2'* and TD-*orf14* expression are below the detection level. The TA-*rolC*, TB-*orf14*, TC-*ocl* and TD-*511* genes were expressed with relative values lower than 0.009.

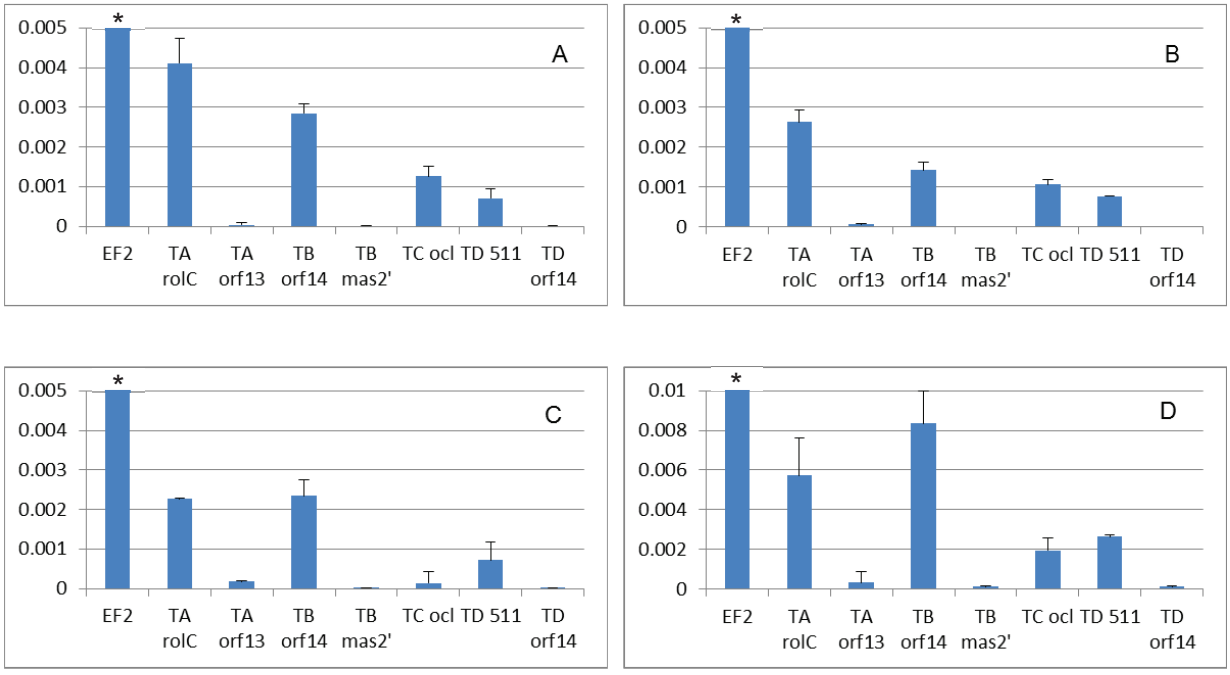


Figure 15. Expression levels of ORFs in *N. tomentosiformis* cT-DNAs detected by qPCR
(A)-(D) Samples from leaf 1 to leaf 4. *: *EF2* (a reference gene from *N. tabacum*) value is set at 1, out of scale in the figures. The expression levels were calculated relative to the *EF2* value. Samples were taken when the *N. tomentosiformis* plant was at the 7 leaf stage.

Table 3. qPCR primers used in figure 15

Region	ORF	Primer (F)	Primer F sequences(5'→3')	Primer (R)	Primer R sequences(5'→3')
TA	*rolC*	3106	CATGAGGTGTCTCGTGGACTC	3107	CGAATAGCATCCTCATAGCTGA
	orf13	3110	GCTGCTTTAAGGGAAATCCA	3111	CATATCTGCTGGTCCCTGGT
TB	*orf14*	3114	GATCACCGCGGTAGTCAAGT	3115	AGTCCTGGGTCCGGATGT
	mas2'	3118	CTGAACGCAAGAGAGCACTG	3119	GAAGTGCCTGCAGAGATGAA
TC	*ocl*	3122	CGATGAGAGGCCCAAAATAC	3123	CGACTTTTCCAACAGATTCGT
TD	*orf511*	3124	AGTTCCCGATCAAACGTCAT	3125	GCTGCAACAGTAGCCGAAAT
	orf14	3126	ACGGATCGCCATAGTTGATT	3127	CCTTCTCTGAGGTATCCAACGTA

II.2 *N. setchellii* cT-DNA

Initially, we tested cT-DNAs in all species of the Tomentosa section species except for *N. setchellii* for which no seeds were available at the Institut du Tabac de Bergerac. Our results showed that it would be very interesting to test this rare species for cT-DNA. According to our evolutionary model for cT-DNAs in Nicotiana (Chen et al., 2014, figure 9) and earlier data on the Nicotiana species evolution (Knapp et al., 2004), *N. setchellii* would not be expected to contain TA, TB, TD and TE but only TC. Seeds were obtained from the US Department of Agricultural Research Service and genomic DNA was tested by PCR using primers flanking the insertion sites of TA, TB, TC, TD and TE. Results showed that these insertion sites were all intact, except that TC could not be amplified (table 4). Therefore, TC was tested in more detail. Two fragments, 2372 to 2375 (c1) and 2420 to 2413 (c2) (Chen et al., 2014, table 3) inside the TC region were tested, but only abnormal amplification fragments were obtained. It could be that the region in which TC was inserted in *N. tomentosiformis* has been deleted in *N. setchellii*, either before or after TC insertion. It seems therefore that *N. setchellii* does not contain any of the known cT-DNA sequences. However, other cT-DNAs might be present. Only a full sequence analysis using genomic sequencing will give a definitive answer.

Table 4. PCR primers and results of TA to TE insertion site amplification in *N. setchellii*
+: PCR fragment with correct size for an intact target region; -: no significant PCR fragment.

Region	Primer (F)	Primer F sequences(5'→3')	Primer (R)	Primer R sequences(5'→3')	Result
TA	3017	CGCTGCATCAGTATTGCTCA	3012	GAGTGGCGTGCTGGAGTATC	+
	3017	CGCTGCATCAGTATTGCTCA	3013	CTCAGGTATTTGCAGGAACAC	+
TB	3136	TCTCCCATGTACTAGCTATC	3144	GAAGATCACCTAACTGAG	+
	3136	TCTCCCATGTACTAGCTATC	3145	CGCTAATGGACTTCGTAC	+
	3137	GAGATATGCATACTATGTTA	3144	GAAGATCACCTAACTGAG	+
	3137	GAGATATGCATACTATGTTA	3145	CGCTAATGGACTTCGTAC	+
TC	2420	CTGTCGAAAGCGATCAAGGC	2368	CCAGGCACCGTTGCTTCTTG	-
	2420	CTGTCGAAAGCGATCAAGGC	2369	GAAACACATACCGTCCATAG	-
	2372	GCCGAGAGATAACCGGCAAG	2375	GCCTTGATCGCTTTCGACAG	-
	2420	CTGTCGAAAGCGATCAAGGC	2413	AGGCGAAGAGGAGTTCAAAG	-
TD	3147	ATTTCACTCCACATCACCTC	3150	CCCCATCACTAGAAATCTCC	+
	3148	AATCCCTGATGAGCGTCTAC	3150	CCCCATCACTAGAAATCTCC	+
TE	3206	AGGAGGAACTGTTGCGTTGT	3203	TGTGAGTTAGTTTAAGTTGGGGC	+
	3207	CACAGATATACCCAAGGTGCCA	3205	TGATCACGGAGAAGGGAAAGAG	+

II.3 TE-*6b*

An intact *6b* gene was earlier found in the *N. otophora* cT-DNA TE region (Chen et al., 2014). It was very surprising to find a *6b* gene since cT-DNAs are normally derived from *A. rhizogenes* strains and the *6b* gene has never been found in such strains. Also *6b* genes are known to produce remarkable growth effects. In order to learn more about the unusual growth induction properties of the *6b* genes, a detailed study of a dex inducible T-*6b* gene was carried out (Chapter II, Chen and Otten, 2015). Additional data on TE-*6b* will therefore be presented at the end of Chapter II.

II.4 *mas2'*

A *mas2'* gene was found in the *N. tomentosiformis* cT-DNA TB region. A 35S::*mas2'* construct led to the production of DFG in a transient expression assay in *N. benthamiana* (Chen et al., 2014, Chapter I). The expression level of *mas2'* in *N. tomentosiformis* is very low (figure 15). However, we discovered by chance that in some *N. tabacum* cultivars, the *mas2'* gene is highly expressed and therefore carried out an extensive study on these cultivars. This work will be presented in Chapter III as a manuscript to be submitted, entitled: "High *mas2'* gene expression and opine synthesis in wild-type *Nicotiana tabacum*".

Chapter II: Morphogenetic properties of *6b* genes from Agrobacterium and *Nicotiana otophora*

I. Introduction

One of the Agrobacterium T-DNA *plast* genes is the *6b* gene. In 1988 it was noticed that this gene induces tumors on *Nicotiana glauca* and *Kalanchoe tubiflora* (Hooykaas et al. 1988). Our group has been interested in the *6b* gene for several years. We found earlier that different *6b* genes have different effects on plant growth. A-*6b* from Ach5, C-*6b* from C58 and T-*6b* from Tm4, when coinfected with the Ach5 cytokinin (*ipt*) gene, induced tumors with shoots, green tumors and tumors with a necrotic surface respectively (Tinland et al., 1989). T-*6b* from Tm4 co-transferred with *rolA*, *B*, and *C* from *A. rhizogenes* caused unorganized calli on carrot roots and very large roots on *N. rustica* (Tinland et al., 1990). In order to study T-*6b* effects on normal plants, this gene was placed under control of a hsp70 heat-shock promoter from *Drosophila melanogaster*. It was found that heat shock induced hsp-T*6b* protoplasts have higher sensitivity to hormones compared to normal protoplasts (Tinland et al., 1992). In 2003, we reported that the AB-*6b* gene induced enations and that the enation syndrome was graft-transmissible (Helfer et al., 2003). Subsequently, the T-*6b* gene was placed under a dexamethasone (dex) inducible promotor to study its effects on normally growing plants (Grémillon et al., 2004). It was found that T-*6b* induced cell expansion like indole-3-acetic acid (IAA), but T-*6b* expression did not increase IAA levels, nor did it induce an IAA responsive gene (Clément et al., 2006). Interestingly, in the T-*6b* tobacco roots, large amounts of sugars and phenolics accumulated (Clément et al., 2007). In *6b* roots, the dominant phenolic compound is chlorogenic acid (CGA). When the synthesis of CGA was blocked by 2-aminoindan-2-phosphonic acid (AIP) in T-*6b* tobacco, CGA accumulation was abolished, but the *6b* phenotype and sugar accumulation remained unmodified (Clément et al., 2007). The contribution of other groups to the study of *6b* genes is presented in the introduction of our recent paper in Planta (Chen and Otten, 2015).

Recently, we found an intact *6b* gene on the TE cT-DNA of *N. otophora*, TE-*6b*. In view of the strong growth-modifying effects of the well-known T-*6b* gene, it would be interesting to investigate the possible role of TE-*6b* in the growth and development of *N. otophora*. We recently studied the properties of the T-*6b* gene in great detail (Chapter II, Chen and Otten, 2015). Subsequently, TE-*6b* was compared with the T-*6b* gene by expressing it under 2x35S promoter control in tobacco (Chapter II, additional data). Remarkably, the 2x35S::TE-*6b* coninduces a phenotype which differs from the enation syndrome, however both may have some basic elements in common.

II. Publication 2

Planta
DOI 10.1007/s00425-015-2387-0

ORIGINAL ARTICLE

Morphological analysis of the *6b* oncogene-induced enation syndrome

Ke Chen[1] · **Léon Otten**[1]

Received: 12 May 2015 / Accepted: 14 August 2015

Abstract

Main conclusion **The T-DNA *6b* oncogene induces complex and partly unprecedented phenotypic changes in tobacco stems and leaves, which result from hypertrophy and hyperplasia with ectopic spot-like, ridge-like and sheet-like meristems.**

The *Agrobacterium* T-DNA oncogene *6b* causes complex growth changes in tobacco including enations; this unusual phenotype has been called "*6b* enation syndrome". A detailed morphological and anatomical analysis of the aerial part *of Nicotiana tabacum* plants transformed with a dexamethasone-inducible dex-T-*6b* gene revealed several striking growth phenomena. Among these were: uniform growth of ectopic photosynthetic cells on the abaxial leaf side, gutter-like petioles with multiple parallel secondary veins, ectopic leaf primordia emerging behind large glandular trichomes, corniculate structures emerging from distal ends of secondary veins, pin-like structures with remarkable branching patterns, ectopic vascular strands in midveins and petioles extending down along the stem, epiascidia and hypoascidia, double enations and complete inhibition of leaf outgrowth. Ectopic stipule-like leaves and inverted leaves were found at the base of the petioles. Epinastic and hyponastic growth of petioles and midveins yielded complex but predictable leaf folding patterns. Detailed anatomical analysis of over sixty different *6b*-induced morphological changes showed that the different modifications are derived from hypertrophy and abaxial hyperplasia, with ectopic photosynthetic cells forming spot-like, ridge-like and sheet-like meristems and ectopic vascular strands forming regular patterns in midveins, petioles and stems. Part of the enation syndrome is due to an unknown phloem-mobile enation factor. Graft experiments showed that the *6b* mRNA is mobile and could be the enation factor. Our work provides a better insight in the basic effects of the *6b* oncogene.

Keywords Hyperplasia · Hypertrophy · Leaf development · Leaf polarity · Plast genes

Abbreviations

dpi	Days post induction
EPC	Extra photosynthetic cell
EVS	Ectopic vascular strand
GST	Glandular secreting trichome

Introduction

Crown gall tumors and hairy roots result from the transfer of T-DNA sequences from Agrobacteria to plant cells. T-DNA structures vary considerably but all contain genes from the plast family (plast for "phenotypic plasticity"), defined on the basis of weak protein similarity (Levesque et al. 1988). Most of the plast genes induce abnormal growth. The *6b* gene also belongs to the plast family and induces tumors on *Nicotiana glauca*, *Kalanchoe tubiflora* (Hooykaas et al. 1988), *N. rustica* (Tinland et al. 1992) and

Electronic supplementary material The online version of this article (doi:10.1007/s00425-015-2387-0) contains supplementary material, which is available to authorized users.

✉ Léon Otten
leon.otten@ibmp-cnrs.unistra.fr

[1] Department of Molecular Mechanisms of Phenotypic Plasticity, Institut de Biologie Moléculaire des Plantes, Rue du Général Zimmer 12, 67084 Strasbourg, France

Published online: 09 September 2015

N. tabacum (Canaday et al. 1992). *6b* genes with different properties have been isolated from different Agrobacterium strains (Helfer et al. 2002), and recently a *6b* gene was identified in a natural transformant, *N. otophora* (Chen et al. 2014). A *N. rustica* transformant with a heat-shock-inducible T-*6b* gene (CAA39648 from Agrobacterium strain Tm4) formed thick roots, tubular leaves and ectopic shoots along the hypocotyls, and its shoot apical meristem was blocked. AK-*6b* (from AKE10) induced thin shoots in tobacco growing from wounded stems, small leaves from the veins on the abaxial side and asymmetrical leaf shapes (Wabiko and Minemura 1996). AB-*6b* (from AB4) induced enations on *N. tabacum* (Helfer et al. 2003). Enations are secondary leaf blades which grow out along the midveins on the abaxial leaf side; they are mirrored with respect to the normal leaf blade (Masters 1869). 2x35S-AB-*6b* tobacco plants had long petioles, leaf wrinkling and dark green spots consisting of ectopic densely packed small photosynthetic cells on the abaxial leaf side. These spots carried an additional vein system, mirrored with respect to the normal veins (Helfer 2001). In addition, 2x35S-AB-*6b* plants had dark green veins, narrow and tubular leaf blades, ectopic tubular structures, corkscrew stems and catacorollas, a special type of double flowers resulting from petal enations (Helfer et al. 2003). T-DNA-induced catacorollas were already reported in 1990 (Komari 1990) but were not attributed to the *6b* gene. The complex set of *6b*-induced modifications was called the "enation syndrome" (Helfer et al. 2003). 35S-AK-*6b* tobacco plants showed leaf blade reduction and extra clusters of vascular-like tissues in petioles (Terakura et al. 2006). 2x35S-AB-*6b* Arabidopsis plants had upward curled leaves with tubes emerging from the abaxial surface (Helfer et al. 2003), yet such tubes were not reported for 35S-AK-*6b* Arabidopsis plants (Terakura et al. 2006). 2x35S-T-*6b* Arabidopsis tissues formed tubes, but did not root. Arabidopsis does not seem to form enations contrary to tobacco, indicating a fundamental difference in morphogenetic potential between the two species. *6b*-expressing tobacco plants produce an "enation factor" that trafficks in the phloem and accumulates in sink tissues where it induces enations (Helfer et al. 2003). Dexamethasone (dex)-inducible dex-T-*6b* plants showed enlarged cotyledons and thick roots, especially on media with sucrose (Grémillon et al. 2004), and isolated dex-T-*6b* root fragments took up external sucrose and expanded after induction (Clément et al. 2007). Four hypotheses have been proposed to explain the *6b* enation syndrome (for recent reviews see Ishibashi et al. 2014; Ito and Machida 2015). (i) *6b* induces the phenylpropanoid pathway with strong accumulation of phenylpropanoids and concomitant inhibition of auxin transport (Gális et al. 2002, 2004; Kakiuchi et al. 2006). However, a phenylammonia lyase inhibitor abolished phenylpropanoid accumulation without changing the *6b* phenotype (Clément et al. 2007). Recent results (Takahashi et al. 2013) reported abnormal auxin and cytokinin accumulation in dex-AK-*6b* seedlings, indicating that these hormones play an important role in the *6b* phenotype. (ii) *6b* stimulates uptake and retention of sucrose. Expansion of induced dex-T-*6b* leaf disks was enhanced by sucrose in the medium (Clément et al. 2006) and induced dex-T-*6b* root fragments took up external sucrose and accumulated it at the site of uptake (Clément et al. 2007). (iii) *6b* binds to three different tobacco proteins: NtSIP1, NtSIP2 and NtSIP3, and could act as a transcription cofactor and histone H3 chaperone (Kitakura et al. 2002, 2008; Terakura et al. 2006, 2007). (iv) The 6b protein crystal structure resembles cholera toxin subunit A (an ADP-ribosylating enzyme, Wang et al. 2011). The 6b protein bound AGO-1 and SE (involved in RNA silencing) and caused changes in miRNA populations. It also exhibited ribosylation activity in vitro when SE and an ADP-ribosylation factor (ARF) from Arabidopsis were added. The authors proposed that 6b ADP-ribosylates AGO-1 and SE, leading to RNA silencing and phenotypic changes.

The four hypotheses (i–iv) mentioned above propose very different mechanisms and it is not yet clear how they can be combined into a single model. The final model should account for every aspect of the *6b* enation syndrome. However, in spite of many molecular and physiological studies, the *6b* phenotype has not yet been explored in much detail. Here, we provide a comprehensive description of the *6b* enation syndrome in tobacco. Our observations can be summarized in a model that reduces the complex enation syndrome to a few basic phenomena.

Materials and methods

Plant material and *6b* gene induction

The transgenic dex-T-*6b Nicotiana tabacum* cv. Samsun nn lines 17.1 and 21.1 (Grémillon et al. 2004; Clément et al. 2006, 2007) were used; both have a single T-DNA locus, are homozygous and inducible to similar degrees. Greenhouse-grown plants were induced by a single spray with an aqueous solution of 3 μM dexamethasone and 0.01 % Tween 20.

Bacterial strains and infection

Agrobacterium tumefaciens LBA4404(pBI121.2::2x35S-T-*6b*) and the empty vector control strain LBA4404(pBI121.2) (Helfer et al. 2002) were used for infection of *N. rustica* stems. The surface of the stem was scratched superficially with a fine needle and bacteria were introduced from an agar plate. Results were analyzed 14 days after infection (dpi).

Microscopy

Plant material was either hand-cut with a razorblade, or prepared for inclusion and cutting with a microtome. Sections of 35 µm were stained with a 0.05 % solution of toluidine blue O. Sections were prepared as described (Tinland et al. 1992).

Scanning electron microscope observations were carried out with a Hitachi TM-1000 Tabletop microscope.

Grafting experiments

Normal tobacco (Samsun nn) was grafted on dex-T-*6b* rootstocks when both were at the 10-leaf stage. One week after grafting, non-induced control leaf disks were removed with a corkborer (diameter 1 cm) for RNA analysis and rootstocks were sprayed with 3 µM dexamethasone and 0.01 % Tween 20. Samples of rootstocks and scions were removed 2 and 5 days later and tested for *6b* mRNA.

PCR experiments

RNA was extracted by grinding leaf disks with Tri-Reagent (Sigma), 200 µl per 1 cm disk, on ice. After grinding, 50 µl chloroform was added. After centrifugation for 5 min at 11000 *g*, the upper phase was removed and precipitated with 1 volume isopropanol. After washing with 70 % ethanol (1 ml), RNAs were taken up in 50 µl RNAse-free H_2O. After treating with RNase-free DNase (Qiagen kit) following the instructions of the provider, cDNA was synthesized with 5 µg total RNA, using Superscript III Reverse Transcription kit (Invitrogen). PCR amplification was done with the Phusion high-fidelity PCR master mix (F-531S) from Thermo Scientific. The components of the PCR reaction were: 2xPhusion Master Mix 5 µl; primer A and B 0.6 µl/each with final concentration 0.5 µM; template DNA 1 µl with final concentration 1–5 ng/µl; H_2O until 10 µl. PCR was performed as follows: one cycle at 98 °C (30 s), 35 cycles at 98 °C (5 s), 60 °C (10 s) and 72 °C (20 s), one cycle at 72 °C (10 min). T-*6b* primers: forward primer: TCACACGCATCCTGAACG, reverse primer: CAAGGTCTCCGAACTGGTAATC (T-*6b* coordinates 32-577, predicted fragment size: 546 bp).

Results

Choice of induction parameters and general observations

In preliminary experiments, greenhouse-grown dex-T-*6b* tobacco plants from the homozygous 17.1 and 21.1 lines (see Materials and methods) were induced at different stages of development. When induced at the emergence of leaf 4 (*L4*, counting from the first real leaf), abnormal leaves and long tubular structures developed and further development became highly variable (Fig. S1a–c, Suppl. Material); in most induced plants apical meristems were blocked, in rare cases normal growth resumed (Fig. S1c, Suppl. Material). Induction at the 6-leaf stage (L6, Fig. 1a) also led to many leaf and stem modifications, but leaves normalized after L16 (Fig. 1b, c). 6b protein levels increased from L1 to L7 (beyond L7 no 6b protein was found), they became undetectable by 8 dpi (results not shown). Induction at the 10-leaf stage resulted in less modifications (not shown). We therefore chose induction at the 6-leaf stage for detailed analysis of *6b*-induced growth modifications. Some plants induced at the 4-leaf stage were also used, this will be specially indicated. Morphological and anatomical observation of over 100 induced plants revealed various new types of stem and leaf changes. In the following paragraphs, we detail these changes, investigate their internal structure and where possible, establish causal connections.

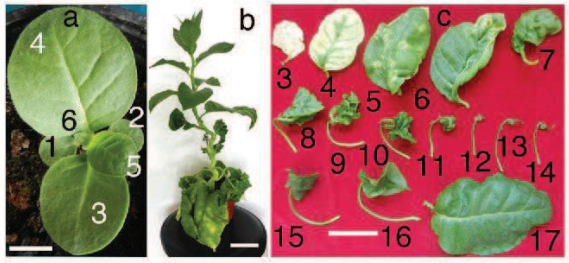


Fig. 1 **a** Plant at 6-leaf stage, non-induced. Leaves are numbered from first real leaf onwards. **b** Two months after induction: the *lower part* is modified, the *upper part* normal. **c** Individual leaves of an induced plant. Leaves *3–16* show various abnormalities, leaf *17* is normal (leaves *1* and *2* were lost). *Bars* 2 cm (**a**), 5 cm (**b**), 10 cm (**c**)

Leaf chlorosis

Induction of dex-T-*6b* cotyledons on dex media (Grémillon et al. 2004) and local induction of dex-T-*6b* leaves (Clément et al. 2006) led to chlorosis. In the present study, plants induced at the 6-leaf stage showed chlorosis in all leaves longer than 5 mm at the time of induction (Fig. 1c) starting at 3 dpi. After 3 weeks, L1 and L2 showed complete chlorosis, L3 intercostal chlorosis retaining chlorophyll around midveins and secondary (II) veins, whereas L4 and L5 showed partial and irregular chlorosis, often at leaf tips and margins (Fig. 1c, Fig. S2, Suppl. Material).

Extra abaxial leaf cells and enations

The L5 surface of induced dex-T-*6b* plants appeared homogeneously dark green (Fig. 2b) compared to normal

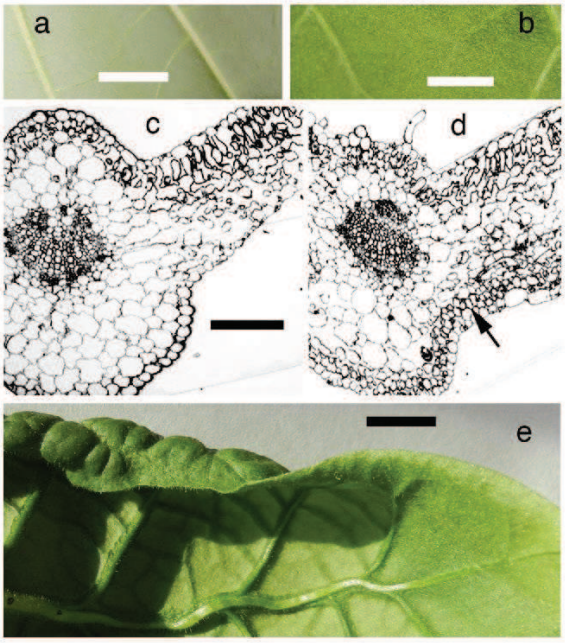
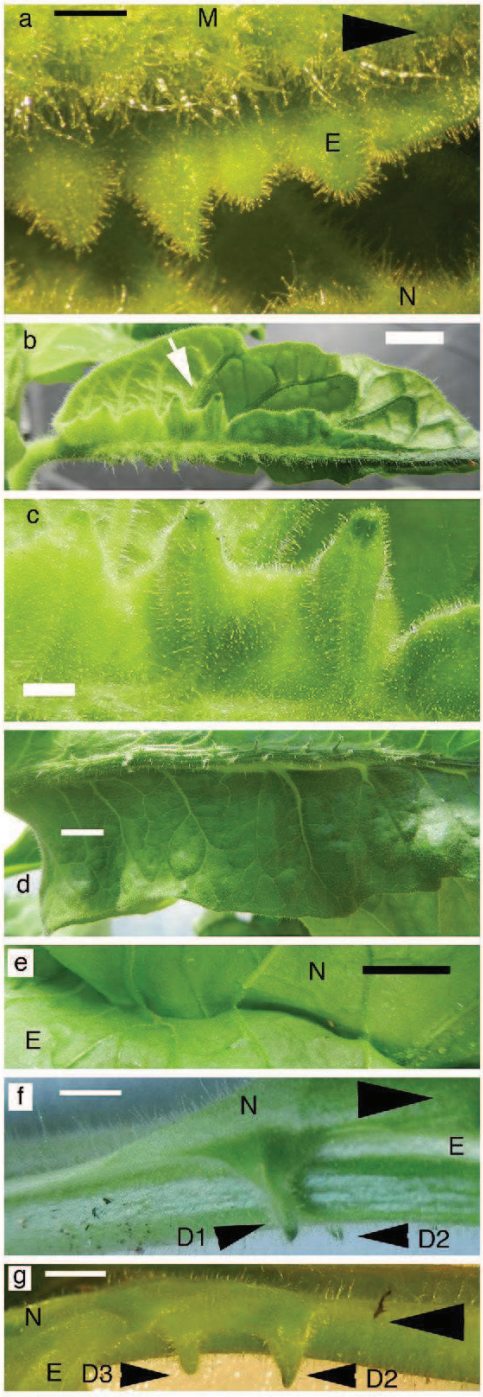


Fig. 2 **a** Normal leaf, abaxial surface. **b** Induced L5 leaf, the abaxial surface is dark green and looks grainy. **c** Transverse section through II vein of normal L5 leaf. **d** Induced L5 leaf: extra cells (*arrow*) cover the abaxial surface. **e** L6 leaf: the apex is uniformly *dark green*, towards the base the *dark green* zones are restricted to the veins. *Bars* 1 cm (**a**, **b**, **e**), 1 mm (**c**, **d**)

leaves (Fig. 2a). Numerous small extra photosynthetic cells (EPCs) were found on the abaxial leaf side (Fig. 2d, arrow) which did not seem to develop any further. No ectopic vascular strands were observed in this area. In L6, the EPCs formed a homogeneous layer in the leaf apex, but in the median and basal part they were concentrated along midveins and smaller veins (Fig. 2e, Fig. S3a–d, Suppl. Material) and formed seemingly random, irregular cell groups (Fig. S3b, inset, Suppl. Material). Since tobacco leaf maturation proceeds in a basipetal fashion, we conclude that *6b* induces vein-associated extra cell division in early stages of leaf development, uniform extra cell division in older leaves and chlorosis in still older ones. At the edges of the narrow EPC zones along midveins and secondary (II) veins, ridges emerged and produced narrow enation blades (Fig. S3c, d, Suppl. Material). Relatively large non-photosynthetic areas formed between the veins and the ridges (Fig. S3d, arrow, Suppl. Material). Enations were accompanied by reduction in leaf blade surface (Fig. 1c) as noted before (Helfer et al. 2003). From L9 to L10 on the proximal and median parts of the leaves, larger enations with veins were formed, growing out from small deltoid outgrowths perpendicular to the midvein (Fig. 3a) each with its own central vein (Fig. 3b, c). Further outgrowth led to relatively straight leaf margins (Fig. 3d). The



positions of the II veins of the enation blades corresponded to those of the normal blade (Fig. 3e). Leaves with a normal and an enation blade will be called "double leaves". In exceptional cases, small isolated outgrowths with a central

◀**Fig. 3** **a** Early stage of outgrowth of large enations: deltoid enations (E) form perpendicular to the midvein (M). N normal leaf blade. *Arrow* points to the apex. **b** Later stage: each deltoid outgrowth has formed a small blade with a central vein, individual outgrowths have started to combine. *Arrow* region shown in **c**. **c** Detail of deltoid outgrowths with *dark green* tip and central vein. **d** Well-developed enation blade. **e** II veins from normal (N) and enation (E) blade are connected at the midvein. **f** Emergence of two symmetrical structures (D1, D2) from the base of an enation-bearing leaf. These structures have a dark green tip and a central vein. They resemble the deltoid structures in **a**, **b** and **c**. Normal leaf (N), enation leaf (E). **g** As **f**, seen from the other side, with an additional deltoid structure (D3). *Arrows* point to the apex. *Bars* 2 mm (**a**), 1 cm (**b**, **d**, **e**), 4 mm (**c**, **f**, **g**)

vein were observed to emerge from the petiole at the base of the leaf blades (D1-3, Fig. 3f, g); they seem to correspond to the deltoid structures of the early enation stage.

Wrinkled leaves

Leaves L8–10 carried II veins with enations at their apical ends, and this region was strongly wrinkled contrary to the basal part (Fig. S4a, b, Suppl. Material). Wrinkling is probably due to restricted outgrowth of the II veins due to the enations, combined with normal outgrowth of the unmodified intercostal lamina.

Composite II veins and hypoascidia

In double leaves, the II veins of enation blades were partially shared with those of the normal blade, thus forming composite II veins. Composite II veins were found at the distal leaf ends; towards the base, the enation blades carried their own II veins (Fig. S4b, Suppl. Material, see also below). However, whereas normal leaves have leaf blade tissue at their base, double leaves generally carried a composite II vein at their base (Fig. S4c, Suppl. Material). Such composite II veins form hypoascidia (pitcher-like structures with adaxial tissue on the outside, Masters 1869), one on each side of the midvein. These hypoascidia lead to ring-like structures in cross sections (Fig. S5a–c, Suppl. Material). Single hypoascidia were found at the apex of highly tubular leaves (see below).

Complex enations and epiascidia

In some cases, several enation leaves developed on the same side of the midvein forming cup-like structures. Plants induced at the 10-leaf stage had a relatively weak *6b* phenotype and developed epiascidia (Fig. S6a, Suppl. Material, see also below). Epiascidia are pitcher or cup-like outgrowths with adaxial tissue on the inside (Masters 1869). They were also observed in wild-type scions grafted on *6b* plants (Grémillon et al. 2004). A unique complex

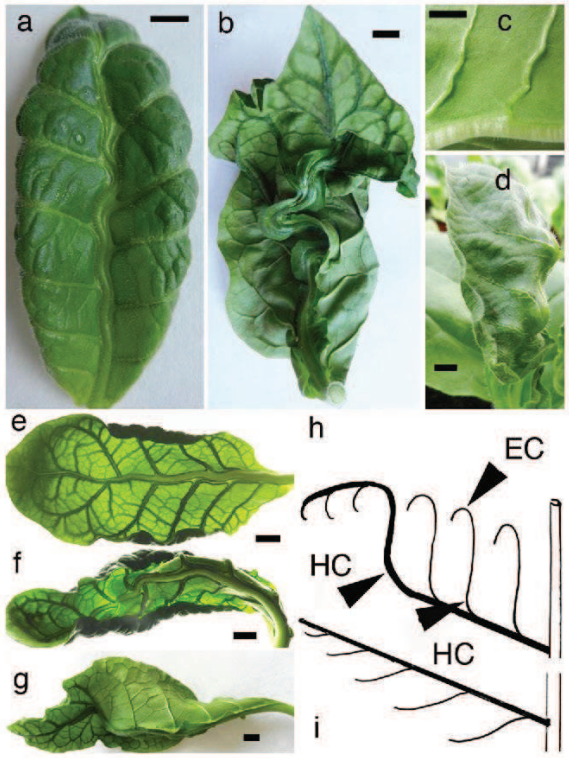


Fig. 4 **a** Early stage of leaf development: the midvein follows the direction of the branching II veins. **b** Later stage, the midvein shows a more pronounced curvature. **c** Curvature of II veins: like midveins, they curve at the position of the branching points with the next-order veins. **d** Up and down hyponastic curvature around midvein-II vein branching points at later stages of leaf development. Prominent bulges appear. **e** Curvature in L6: hyponastic II veins at the leaf base, with epinastic leaf margins. **f** Another L6 leaf, more pronounced. **g** An L6 leaf at a later stage tilted to show the profound dip at the leaf base. **h** Interpretation of structures shown in **e**–**g**. Hyponastic curvature (HC) of basal II veins at proximal ends, epinastic curvature (EC) at distal ends; the median midvein region is hyponastic. This bending pattern resembles early bending of petioles (Fig. S7, Suppl. Material). **i** Normal tobacco leaf for comparison. *Bars* 1 cm (**a**–**g**)

double enation and its interpretation are shown in Fig. S6b, c (Suppl. Material).

Hyponasty, epinasty and undulating growth

6b plants have hyponastic petioles and epinastic midveins (Clément et al. 2006) which appeared as early as 3–4 dpi, from L4 onwards (Fig. S7a, b, Suppl. Material). The epinastic curvature probably results from expansion since we could find no abnormal cell division at the adaxial side. The *6b* gene is known to induce cell expansion in leaves, roots, cotyledons, petioles, midveins and smaller veins (Grémillon et al. 2004; Clément et al. 2006). In L5–L6

leaves, the midveins and II veins showed a left–right zig-zag pattern in the plane of the leaf, at branching points the midveins turned in the same direction as the II veins (Fig. 4a). Possibly, this was due to enhanced expansion on the side opposite the branching point. The effect became stronger with time (Fig. 4b). *L4* leaves had undulating II veins following the direction of the III veins (Fig. 4c). In older plants, midveins not only undulated sideways but also up and down at midvein-II vein branching points (Fig. 4d). Highly characteristic folded structures appeared in L6–L8 (Fig. 4e–h), II veins at the leaf base were hyponastic at the branching points with the midvein and epinastic at their distal ends (leading to epinastic leaf margins), much like petioles in younger leaves (see above). At its base, the lamina carried II veins without enations and although strongly folded upwards seemed to expand normally. Towards the apex, the midvein became hyponastic, the II veins were much less folded and carried enations (Fig. 4g). The various modifications yielded leaves with a narrow and deep central "dip" at their base (Fig. 4f, g). The striking proximal/distal differences in vein curvature, leaf expansion and enation growth might be related to features of basipetal leaf development such as vein maturation and/or sink–source transition.

Gutter-like petiole structures and fan-shaped leaves

Leaves of the tobacco cultivar Samsun have relatively short petioles and large and broad leaves, but *6b* plants formed long petioles and much shorter and narrow leaf blades (L8–L16, Fig. 1b, Fig. S1b, Suppl. Material). In L8–L10, the petioles broadened progressively towards the base of the leaf and folded upwards both transversely and longitudinally, forming hyponastic gutter-like structures (Fig. 5a, b), up to tightly closed structures (Fig. 5c). In the folded area, several II veins were found to branch off from the midvein at small angles without forming leaf blade. Up to 4 parallel II veins were found on each side (Fig. 5d). At more apical positions, the parallel II veins fanned outwards due to the growth of intercalary leaf blade tissue, forming normal and enation blades (Fig. 5a).

Twisted petioles

Petioles/midveins of *6b* leaves around L10 (Fig. 1c) showed complicated folding patterns consisting of hyponastic and epinastic curvature and torsion along the main axis (Fig. S8a–i, Suppl. Material). Figure S9 (Suppl. Material) shows a leaf with 5 different folded regions: its petiole started with a hyponastic region close to the stem (see above, Fig. S7b, Suppl. Material), followed by a long straight region (1), an epinastic part (2) with a dark green ridge on the abaxial side (see below), a hyponastic gutter-shaped region (3), see above (Fig. 5), an epinastic part where the leaf blades grew out (4, see above, Fig. S7b, Suppl. Material) and a short hyponastic region at the apex (5). The leaf shown in Fig. S8d–f (Suppl. Material) had the same overall pattern but in addition showed torsion along the main axis. L12–14 had no hyponastic gutter structure and their long petioles formed epinastic coils (Fig. S8b, c, h, i, Suppl. Material). In some leaves, the double blades were relatively well developed but fully enclosed and rolled up by the coiled epinastic midvein, leading to characteristic butterfly forms (Fig. S10a, b, Suppl. Material). In case the leaf blades started at different positions of the

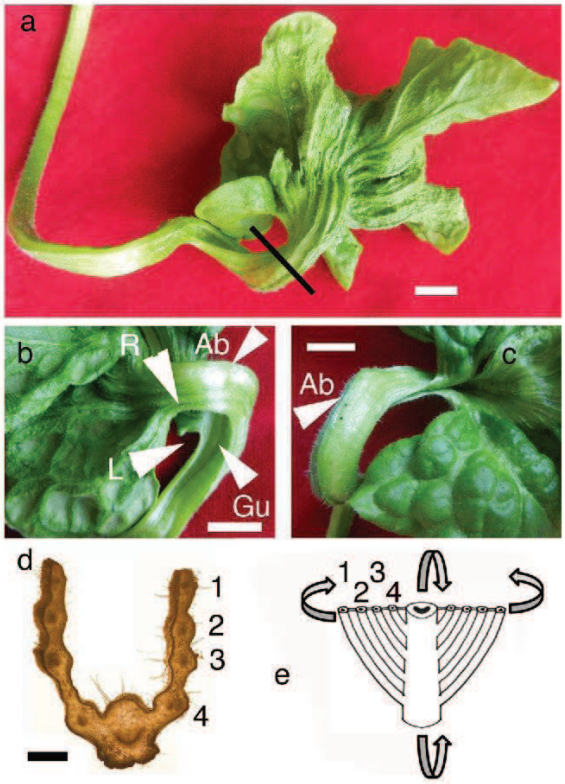


Fig. 5 **a** L8 with long twisted petiole, progressively broadening towards the apex and becoming hyponastic, forming a gutter-like structure. Parallel ridges correspond to II veins, these fan out at a more apical position. *Bar*: position of transverse section shown in **d**. **b** Detail of a gutter (Gu) folding to the *left*, its *right edge* follows the leaf blade emerging on the *left edge* of the gutter (L), the leaf blade emerging more distally on the *right edge* is indicated by (R). Abaxial side (Ab). **c** A fully closed gutter, Ab abaxial side. **d** Cross section of a gutter, at position indicated by *black bar* in **a**. *Numbers* indicate different II veins in order of branching, starting at the most proximal position. No leaf blade has developed between the II veins. **e** Interpretation of the gutter structure: 4 successive II veins (1–4) branch at a sharp angle from the midvein, without leaf blade development. The gutter is shown flattened (seen from the abaxial side) but folds up hyponastically, both in transverse and longitudinal direction (*arrows*). *Bars* 1 cm (**a**, **b**, **c**), 2 mm (**d**)

petiole, the gutter-like structures turned in the direction of the first emerging blade (Fig. 5b). Removal of the lamina showed that the *6b* leaf folding patterns were predominantly determined by the petiole and midvein folding patterns (Fig. S8h, i, Suppl. Material).

Ectopic vascular strands and dark green ridges on petioles and midveins

The vascular system of normal tobacco petioles/midveins is composed of several bundles arranged in an arc. At its lateral ends, it generates the II veins of the petiole wings and leaf blades. In L7 and L8 petioles of *6b* plants, ectopic vascular strands (EVSs) were found at the ends of the normal arc and non-photosynthetic ridges replaced the petiole wings from L9 onwards (Fig. 6a–c), these ridges continued down along the stem (see below). Petioles of later leaves did not form any lateral outgrowth at all. From L9 to L10 on multiple EVSs, a semi-circle close to the abaxial petiole surface was progressively formed (Fig. 6d). At later stages, these EVSs were found to be associated with longitudinal ridges running along the midveins and petioles (Fig. 6e, f). A series of cross sections showed that in apical leaf parts the EVSs were situated inside these ridges close to the abaxial surface. In basal parts, they were found at some distance from the abaxial surface, suggesting that they grow in a basipetal direction and stimulate the growth of the surrounding parenchyma (Fig. 6g, h). Further observations showed that EVSs were always present in the apical parts and extended for various distances towards the basal parts, confirming that they develop in a basipetal fashion like normal vascular strands (not shown). In many cases, the EVSs continued from the petiole down into the stem without connecting to the stem vascular system like normal vascular strands (see below). Later leaves developed an irregular EVS pattern on the inside of the petiole/midvein, with more regularly spaced EVSs close to the abaxial surface (Fig. S11a, Suppl. Material). In plants with tubular leaves, the normal vascular strands and EVSs formed a ring of vascular tissue (Fig. S11c–f, Suppl. Material). Structures of EVS groups along petioles and midveins were quite variable (Fig. S11g, Suppl. Material) and will require further study. For EVS polarity, see below.

Small enations at the base of the midvein

In later leaves, the main vascular system recovered its usual shape and petiole wings and leaf blades reappeared. However, small enations with an adaxial surface could be found along the midveins of some leaves (Fig. S12a, b, Suppl. Material). Cross sections (Fig. S12c, Suppl. Material) showed a few residual EVSs in the central "bottom" part of the abaxial side, the narrow enation blades emerged from their lateral ends. The polarity of these enation blades (adaxial side on the lower leaf side, Fig. S12a, Suppl. Material) indicates that the EVS part closest to the

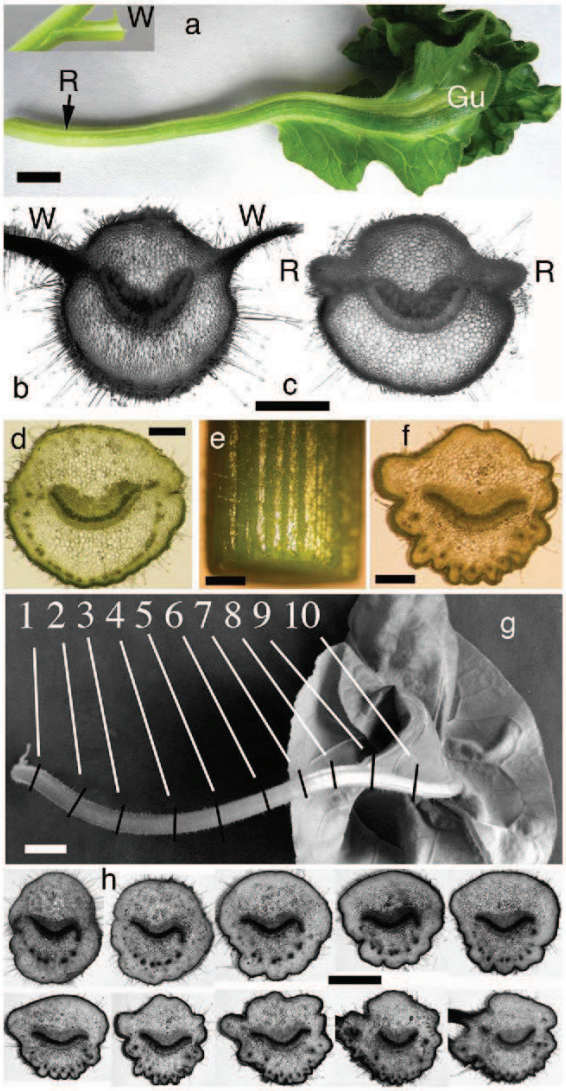


Fig. 6 **a** Leaf with ridges on the abaxial side of the petiole and midvein. Gutter-like structure (Gu). A *light green* ridge (R) runs along each side of the petiole, replacing the petiole wing (*inset*, W). **b** Normal petiole with petiole wing (W), cross section. **c** *6b* petiole, cross section, the petiole wing is replaced by a ridge (R). **d** Cross section of petiole with numerous extra vascular strands (EVSs) close to the lower surface. The petiole wing is fully absent. **e** Ridges on the lower surface of the petiole. **f** Cross section in ridge area: ridges correspond to semicircular outgrowths around the EVSs, compare with **d**. **g** Successive petiole cross sections (*1–10*), section 7 corresponds to **f**. **h** Cross sections of petiole and beginning of midvein (from base to apex). At the base of the petiole, the EVSs are not as close to the lower surface as in the apical region. *Bars* 1 cm (**a**, **g**), 2 mm (**b**, **c**, **h**), 1 mm (**d**, **e**, **f**)

(originally abaxial) petiole surface is its adaxial side. The "bottom" enations are thus different from those originating from the deltoid structures along the midveins (see above). Outgrowth of leaf blade tissue requires juxtaposition of adaxial and abaxial domains of vascular tissues (Waites and Hudson 1995). We therefore hypothesize that the free lateral ends of the small EVS group enabled the outgrowth of these "bottom enation" blades. This was also indicated by the unusual structure of an exceptional asymmetrical leaf (Fig. S13a–g, Suppl. Material). This leaf carried hyponastic II veins on one half, the other half was normal (Fig. S13a, inset, Suppl. Material). Its petiole carried an ectopic leaflet on the modified side. Transverse sections showed a group of EVSs on the modified side, with the leaflet emerging from its free end (Fig. S13e–g, Suppl. Material). The polarity of this leaflet confirmed the polarity of the EVSs indicated above (adaxial side facing the petiole surface). The modified part of the asymmetric leaf also lacked a petiole wing contrary to the normal half, confirming the importance of EVSs in the inhibition of petiole wing formation (see above).

Small tubular structures on petioles and midveins

Small tubular structures appeared on petioles and midveins of young L10–15 leaves (Fig. 3b, d, Fig. S4c, Suppl. Material). They were first noted in 2x35S-AB-*6b* plants (Helfer et al. 2003) and described as "rod-shaped protrusions" in AK-*6b* tobacco plants (Terakura et al. 2006) but not further analyzed. They emerged more or less simultaneously on the dark green abaxial side of midveins and petioles (see above). The tubes had a diameter of about 0.2 mm and generally grew to a length of about 5 mm (some up to 2 cm, Fig. S14a, Suppl. Material). Although they emerged perpendicularly to the leaf surface, their tips later curved in the direction of the leaf apex (Figs. 3d, 7b). At later stages, most tubes were found at the apical end of a longitudinal ridge (Fig. S14b, Suppl. Material). Transverse serial sections showed that the tubes grew directly from the abaxial EPC layer, without connection to the EVSs (Fig. S14c-f, Suppl. Material shows the emergence of three successive tubes). Young tubes did not contain an internal vascular system as shown by serial sections, yet older ones developed a central vascular system (not shown). Remarkably, all tubes in the abaxial EPC areas of petioles and midveins carried a large trichome at their tips (Fig. 7a–c). Among the different types of *N. tabacum* trichomes, the large glandular secreting trichomes (GST) have a strong capacity for secretion and arise very early in leaf development (Wagner et al. 2004). At the time of tube formation, GSTs were already fully developed, suggesting they are essential for tube initiation. Scanning electron microscope analysis of early tubes (Fig. S14g–i, Suppl. Material)

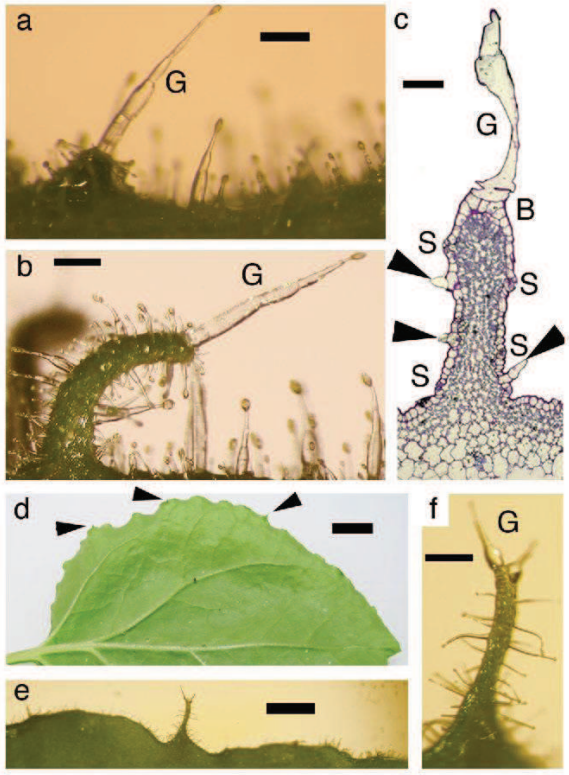


Fig. 7 **a** Induced *6b* tobacco plant: a tube emerges behind a large glandular secreting trichome (GST) on the abaxial side of the midvein. **b** Later stage, the GST tube curves towards the leaf apex (on the *right*). **c** Longitudinal thin section through the center of a GST tube, the inside of the tube contains small meristematic cells. Stomata (s), GST (G), base of the GST (B), small trichomes (*arrows*). **d** *N. benthamiana* leaf (abaxial side) with small GST-associated structures on the margins (*arrows*). **e** Detail of *N. benthamiana* leaf margin with GST tube. **f** Detail of GST tube with two large trichomes at the tip and several smaller trichomes along its length. *Bars* 5 mm (**a**, **b**), 0.2 mm (**c**), 1 cm (**d**), 1 mm (**e**), 0.3 mm (**f**)

showed small epidermal cells close to the GST and elongated epidermal cells at more distal positions. Tubes carried stomata and small trichomes. Their inside consisted of small isodiametric dark green cells, similar to EPCs, as shown by light microscope analysis (Fig. 7c). We tested whether GSTs were required for tube growth by removing them with fine tweezers. Tubes without GSTs continued their growth like intact tubes (Fig. S15a–c, Suppl. Material), showing that GSTs are not required for tube outgrowth. We will call these tubes "GST tubes".

GST-induced outgrowths on *Nicotiana benthamiana* leaves

Since large trichomes play a role in GST tube induction, we investigated their occurrence and nature in different

Nicotiana species (results not shown). Large trichomes were also found on leaves of wild-type *N. benthamiana* plants. Surprisingly, older leaves formed long narrow outgrowths at their margins that always carried large trichomes at their tips (Fig. 7d–f). To the best of our knowledge, these structures have not been reported before and could arise in a similar way as *6b*-induced GST tubes in *N. tabacum*, i.e., by GST-stimulated growth of division-competent cells.

GST tube-derived structures

GST tubes at the edges of EPC areas on petioles and midveins often formed narrow outgrowths on one side of their main axis, somewhat behind the apex, leading to a shallow elongated cup (Fig. 8a, b). Some cups developed into leaflets with midveins and II veins (Fig. 8c, Fig. S16a–d, Suppl. Material) indicating that GST tubes are ectopic leaf primordia. At their base, the leaf-like structures were broadened and flattened in the direction of the petiole/midvein axis. Two adjacent GST tubes were found to form leaflets with opposite polarity: their adaxial sides faced each other (Fig. 8c). GST tube-derived leaflets sometimes formed small enations (Fig. 8d, Fig. S16c, Suppl. Material). GST tubes could be branched in different ways. Branches could be of the same length, each with its own GST (Fig. S16e, Suppl. Material) or dissimilar (Fig. S16f, Suppl. Material). In one particularly interesting case (Fig. 8e), a complex GST tube consisted of three branches, one carrying the GST (most likely the original GST tube) and two others, appearing symmetrically behind the tip of the GST tube and curved upwards and inwards, without GSTs. The two GST-less tubes are most likely tube-like equivalents of the more common leaf blades (Fig. 8f, g). The outgrowth of leaf blades and secondary tubes shows that GST tubes can conserve or acquire dorsiventral polarity. In another exceptional case, a GST tube carried two leaf-like structures at its base (Fig. S16 h, i, Suppl. Material). This could be an extreme form of leaf blades emerging from a GST tube. Sometimes, GST tubes broadened at their base without forming leaflets (Fig. S16g, Suppl. Material). On older plants, fleshy triangular outgrowths without GSTs were observed at the edges of the EPC regions (Fig. 8h).

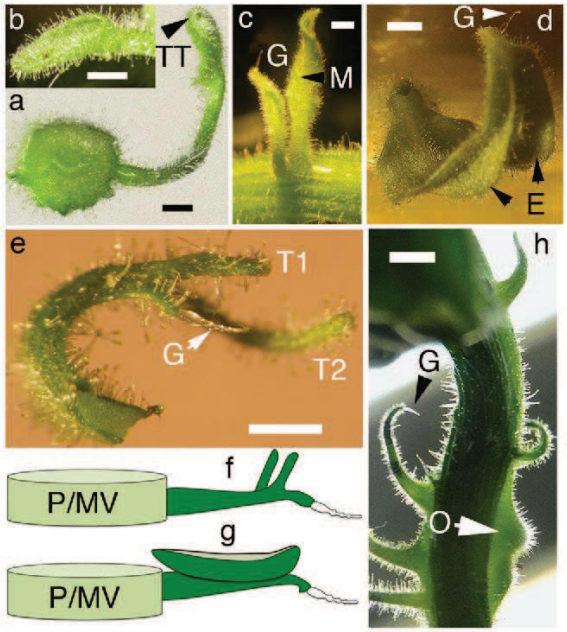


Fig. 8 **a** Petiole cross section at a site where a GST tube grows out and forms a cup-like structure (seen from the lower side) at its apical end. Original tube tip (TT). **b** Another view of same cup-like structure as in (**a**) with margins growing out. **c** Leaflets emerging from two GST tubes. The smaller leaflet is seen from its abaxial side, its GST is marked (G). The bigger leaflet is seen from its adaxial side, it contains a small midvein (M). Both leaflets arise perpendicularly to the main axis on the flanks of the petiole; their bases are flattened parallel to the main axis. **d** GST tube-derived leaflet with enations (E), G: original GST at the apex. **e** Forked GST tube with two symmetrical GST-less blunt-ended tubular branches (T1, T2) arising behind the tip of the GST tube with its GST (G). **f** Schematic drawing of forked GST tube as in **e**. **g** Schematic drawing of GST tube with leaflet as in **b** and **c**, showing possible correspondence with forked GST tube. *P/MV*: petiole/midvein. **h** A fleshy outgrowth (O) on the flank of the petiole, without GST, broadened parallel to the main axis, the other structures are GST tubes (G). *Bars* 2 mm (**a**–**e**), 5 mm (**h**)

Paired tubular structures at the end of II veins of enation leaves

As noted above (Fig. S4, Suppl. Material), double leaves carried composite II veins at the apical half consisting of a normal part and a shorter enation part. At the distal ends of the enation part (Fig. S4b, Suppl. Material), small and paired tubular structures developed (Fig. S17a–g, Suppl. Material). They lacked trichomes and extended outwards from the ends of the enation veins. Because of their paired and curved nature, we propose to call these corniculate (horn-like) tubes.

Single hypoascidia

Leaves of early-induced plants formed long petioles with single hypoascidia (Fig. S18a, b, Suppl. Material). Serial cuts showed how the normal vascular system and the EVSs of the petiole participated in the formation of these structures (Fig. S18c–g, Suppl. Material). The process apparently involves lamina outgrowth between individual well-separated EVSs and the appearance of a cavity lined with an abaxial epidermis between the normal vascular system and the EVSs. These hypoascidia confirm the polarity of

EVSs, i.e., adaxial side facing the surface of the petiole/midvein (see above). We assume that lamina outgrowth in hypoascidia became possible by isolation of a small number of EVSs in well-developed ridges (equivalent to lamina outgrowth from the free ends of the "bottom" EVSs or the asymmetrical EVSs, see above). The ontology of the hypoascidium and its internal abaxial epidermis remain to be studied.

Tubular leaves

Plants induced at the 4-leaf stage formed tubular structures with a blunt end without leaf blade (we will call these "pins"). Some grew to a considerable length (up to 15 cm), others remained small (see also below). Pins carried long and thin trichomes with a small cell at the tip, at the pin's apex the trichomes radiated outwards. Some of the pins developed branches (Fig. 9a–c, Fig. S19a–g, Suppl. Material). Remarkably, one exceptional pin showed five branches: the original pin, a symmetrical pair of small, secondary pins and a symmetrical pair of leaflets (Fig. 9c, d, Fig. S19a–e, Suppl. Material). The secondary pins resembled the secondary tubes on the GST tube discussed above (Fig. 8e). The complex 5-branched structure may be explained by assuming that the main pin conserved or acquired dorsiventral polarity (as in the case of the GST tubes, see above), allowing secondary structures to grow out from the adaxial side somewhat behind the apical end of the original pin. According to their position, the secondary structures could correspond to modified II veins with the proximal ones yielding the secondary pins and the distal ones the leaflets. The internal structure of tubular leaves showed a somewhat flattened structure of central vascular elements, indicating partial conservation of polarity (Fig. S19f, Suppl. Material). Other branched structures were found in which some of the tips were apparently differentiating (Fig. S19g, Suppl. Material). The origin, development and internal structure of simple and branched pins remain to be studied in detail.

Complete inhibition of leaf growth

In several *6b* plants, minute leaf-like structures appeared at the place of normal leaves. They were accompanied by normal buds and were situated just below the normalized leaves (Grémillon et al. 2004). In the present experiment, most early-induced plants also showed such minute leaves (Fig. S20a, b, Suppl. Material), some even showed a complete lack of outgrowth ("empty sites", Fig. S20c–e, Suppl. Material). The corresponding buds were always present. In cases without any outgrowth, the "empty site" was covered with numerous long thin trichomes, giving the surface a lighter color (Fig. S20c–e, Suppl. Material), these trichomes were similar to those found on the pins (see above).

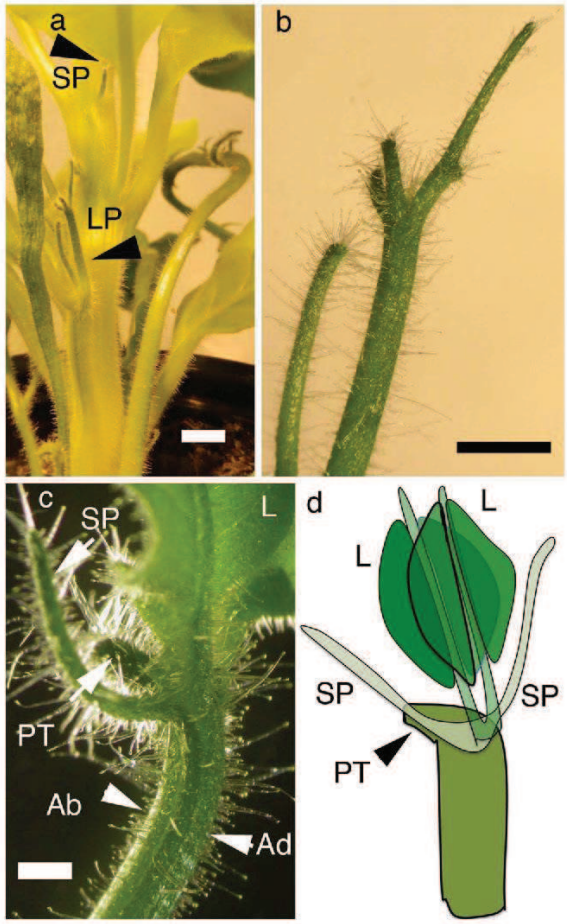


Fig. 9 **a** *6b* plant induced at the 4-leaf stage, 2 months after induction. A short pin (SP) and a long branched pin (LP) grow from the main stem. **b** Detail of branched pin shown in **a**. The pin carries long thin trichomes. No leaf blade has formed; the ends of the pins are flat. **c** A pin with a complex branching pattern. Just behind the pin's tip (PT), four structures grow on one side of the pin (presumably the adaxial side (Ad): two pin-like structures (secondary pins, SP) and two leaflets (L). Abaxial side (Ab). **d** Interpretation of the five-branched structure of **c**. *Symbols* as in **c**. Additional views of this structure in Fig. S19a–e (Suppl. Material). *Bars* 1 cm (**a**), 5 mm (**b**), 2 mm (**c**)

Stem modifications

Stems of induced *6b* plants have corkscrew structures, and they bend towards the petioles at petiole branching positions, as do midveins with respect to II veins (Helfer et al. 2003, see above). In normal tobacco, decurrent petiole wings (Fig. 10a, W) grow down along the stem as adnations, they are extensions of the petiole wings (Smith and Seltmann 1979). In *6b* plants, they were often replaced by

ridges (Fig. 10b, R). In addition to these ridges, *6b* plants carried a single subpetiolar ridge directly below the petiole (Fig. 10b, c, S). A ridge starting from one petiole ended three petioles lower (Fig. 10c). Stem cross sections showed (Fig. 10d, e) that the subpetiolar ridge (S) contained EVSs (E) that ran down from the petioles into the stem; several subpetiolar ridges with EVSs could be seen to descend from successive leaves (Fig. 10e). At the flanks of the subpetiolar ridges, the semicircular series of EVSs turned outwards towards the stem surface. Sometimes narrow leaf blades with the expected polarity (adaxial side facing the subpetiolar ridge) grew out from these free ends (subpetiolar wings, Fig. S21a–c, Suppl. Material) providing a further example of outgrowth of leaf blade from free-ending groups of EVSs (see also above).

Inverted leaf emerging from a petiole

In one exceptional case, a fan-shaped leaf with inverted polarity emerged from the lower part of a petiole (Fig. 10f–i). Based on the EPC spots found in 2x35S-AB-*6b* leaves (Helfer et al. 2003), we interpret this interesting structure as being derived from an EPC spot with inverted polarity on the abaxial side of the petiole. This spot could have generated an ectopic blade with EVSs developing from its margins, allowing it to grow out into a fan-shaped leaf. At the same time, the EVSs continued to grow down into the stem (see above). This hypothesis is strengthened by the appearance in another plant of a very small dark green structure and a shallow ridge below it, with associated EVSs (Fig. S22a–f, Suppl. Material). This less developed structure may have the same origin as the inverted leaf.

Unusual structures at the petiole base

In *6b* plants, the petiole wings (or ridges) formed by the normal vascular system and the subpetiolar wings (or ridges) formed by the EVSs are connected at the petiole base without interruption or change in polarity. At this position, several unusual structures were found. Figure 10j, k shows two small leaflets (L) with a midvein and II veins; they emerge on both sides of a petiole at the connection point between petiole wing (W) and subpetiolar wing (SW) and have the same polarity as the wings. Interestingly, they resemble stipules found in many plants but not in tobacco. A smaller leaflet (similar to the deltoid structures in Fig. 3a–c) developed on the subpetiolar wing (SL, Fig. 10k). *6b*-induced modifications at the petiole base are compared to normal structures in Fig. 10l, m. Larger leaves with long petioles on both sides of the petiole base were also found (Fig. S23a, Suppl. Material). In one asymmetric case, a petiole base showed a well-developed leaflet on one side and a normal petiole wing on the other (Fig. S23b–d, Suppl. Material). Sometimes, the petiole was twisted at its base (Fig. S23e, Suppl. Material). In another case, a small cup had formed on one of the sides (Fig. S23f, Suppl. Material). It is unclear whether the latter two modifications are ontologically related to the stipule-like modifications.

Upward shift in bud positions

In later leaves, buds appeared at a considerable distance from the corresponding petioles (Fig. S24a–d, Suppl. Material). In some cases, an accessory bud was found between the petiole and its normal bud (Fig. S24a, b, Suppl. Material). In rare cases, buds showed a ridge running down to the petiole of the subtending leaf (Fig. S24c, Suppl. Material). At later growth stages, the region above the petiole looked swollen (Fig. S24d, Suppl. Material). All these bud modifications require further investigation.

Phenotypes of plants with leaky dex-T-6b expression

Transformation of tobacco with the dex-T-*6b* construct (Grémillon et al. 2004) led to several transformants with leaky *6b* expression. The striking phenotypes of some of these plants (Fig. S25a–g) apparently represent subsets of the morphological modifications described above (Fig. S25a, b, e–g), another one (Fig. S25c–d) remains unexplained. These leaky transformants demonstrate the extraordinary morphogenetic potential of the *6b* gene.

Movement of *6b* mRNA

The *6b* phenotype is graft-transmissible (Helfer et al. 2003) and led us to postulate a phloem-mobile "enation factor". Earlier we reported that 6b protein moves out of Agrobacterium-infiltrated leaf patches, and postulated that it corresponds to the enation factor (Grémillon et al. 2004). However, subsequent studies showed that such patched plants did not develop enations at distal regions or in wild-type scions (Otten, unpublished). We therefore tested the possible movement of *6b* mRNA by grafting wild-type tobacco on dex-T-*6b* rootstocks. Induction of the *6b* rootstock led to ectopic cell division along III veins of scion leaves with a length of 5 cm or less at the time of induction (Fig. S26a, b, Suppl. Material) as shown before (Grémillon et al. 2004). *6b* mRNA levels in rootstock and scion were measured by semi-quantitative PCR. The results (Fig. S26c, Suppl. Material) showed that the *6b* mRNA can be transported across graft junctions. Additional experiments (not shown) showed that the spatial and temporal *6b* mRNA patterns were variable from plant to plant, probably reflecting the variable quality of the graft junctions.

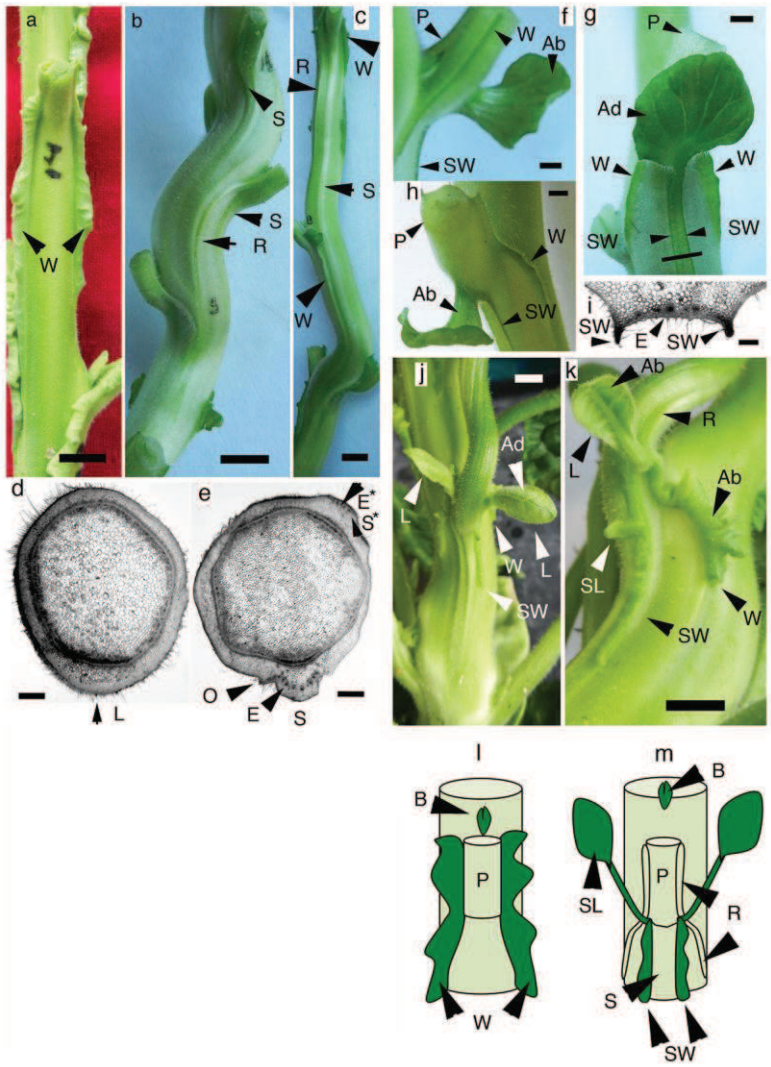


Fig. 10 **a** Normal tobacco stem. Decurrent petiole wings (W) on both sides of the petiole. **b** Stem of a *6b* plant; petiole wings are replaced by ridges (one is seen from the side: R) and another, single ridge appears below the petiole (subpetiolar ridge: S). **c** In this *6b* plant, both petiolar wings (W) and ridges (R) are present, the subpetiolar ridge (S) runs down along the stem for a considerable distance. **d** Cross section of normal stem, somewhat below the site where a leaf emerges (L). **e** Cross section of a *6b* plant, same orientation as **d**. The subpetiolar ridge (S) contains several EVSs (E), on the left these turn towards the surface of the stem and lead to a small outgrowth (O). A subpetiolar ridge (S*) and its EVSs (E*) descend from the leaf above leaf L. See also Fig. S21a–d (Suppl. Material). **f** An exceptional fan-shaped leaflet emerging from the abaxial side of a petiole (P), leaflet viewed from the abaxial side (Ab), narrow petiole wing (W), narrow subpetiolar wing (SW). **g** Leaflet viewed from the adaxial side (Ad). The leaflet shows several radiating veins but no obvious midvein. *Black line* the section shown in **i**. *Symbols* as in **f**. **h** Leaflet base viewed from the abaxial side (Ab). The outer edges of the leaflet continue along the petiole (P), forming subpetiolar wings (SW). Petiole wing (W). **i** Cross section as indicated in **g**; EVSs (E) are present close to the abaxial petiole surface and turn outwards generating subpetiolar wings (SW) , bud (B), petiole ring (W). **j** A plant with two symmetrical leaflets (L) emerging at the base of a twisted petiole at the site where the subpetiolar wing (SW) and petiole wing (W) meet. Leaflets viewed from the adaxial side (Ad). **k** Leaflet viewed from the abaxial side (Ab), a smaller leaflet (SL, seen from its abaxial side) is found on the edge of the subpetiolar wing. The petiole wing is replaced by a ridge (R). Other *symbols* as in **j**. **l** Schematic overview of normal stem. The petiole (P) is cut off somewhat above the point where it connects to the stem. Bud (B), petiole wing (W). **m** Overview of *6b* stem. Bud (B), petiole (P), subpetiolar ridge (S), ridge replacing petiole wing (R), subpetiolar wing (SW). SL, stipulate-like outgrowth emerging at the connection between petiole ridge (R) and subpetiolar wing (SW). *Bars* 1 cm (**a–c**), 2 mm (**d–h**), 0.2 mm (**i**), 5 mm (**j**, **k**)

Induction of cell division in the stem cortex of *Nicotiana rustica*

The induction of ectopic cell divisions (hyperplasia) in leaves, stems, petioles and midveins by *6b* expression is an essential part of the enation syndrome. As the enation phenotype of dex-T-*6b* plants is very complex, it would be interesting to develop a simpler system to study *6b*-induced hyperplasia. Agrobacteria carrying *6b* genes induce large tumors on pierced *N. rustica* stems (Tinland et al. 1992), but such infections involve various cell types, including the vascular cambium. We therefore tested the reaction of superficially wounded *N. rustica* stems to infection with Agrobacteria carrying a 2x35S-T-*6b* gene, an infection system that only involves epidermis and cortex cells. Significant division of cortex cells was observed which did not involve the vascular system (Fig. 11a–d).

Discussion

The various *6b* mechanisms proposed so far (Ishibashi et al. 2014; Ito and Machida 2015, this paper) are not obviously related, and it seems unlikely that the small 6b protein (23 kD) could be multifunctional. While further molecular studies are necessary, we considered it important to improve the morphological analysis of the complex *6b*-induced enation phenotype. We noted over 60 different morphological changes in dex-T-*6b* tobacco plants (summarized in Fig. 12 and Table 1), and they are related in complex ways. Our observations and those of others suggest three basic effects: (i) leaf chlorosis, (ii) ectopic cell division or hyperplasia and (iii) abnormal cell expansion or hypertrophy. These should eventually be traced back to a primary molecular effect, which might be ADP ribosylation of one or more target proteins (Wang et al. 2011).

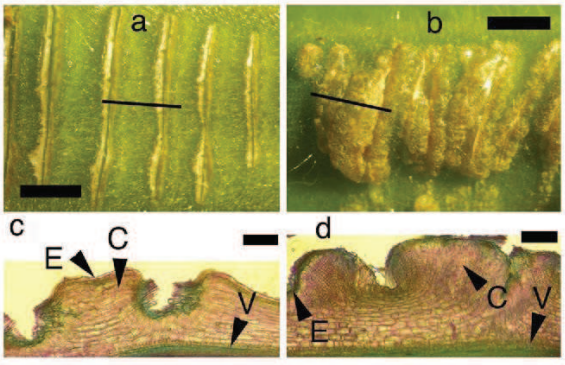


Fig. 11 **a** T-*6b*-induced cell division in the stem cortex of *Nicotiana rustica*. Infection with a control Agrobacterium strain, LBA4404(pBI121.2), 2 weeks after infection. **b** Infection with Agrobacterium strain LBA4404(pBI121.2::2x35S-T-*6b*), 2 weeks after infection. The control infection shows a limited wound reaction, infection with the T-*6b* construct induces callus. **c**, **d** Hand sections corresponding to the thin lines in **a** and **b**. The control infection shows a limited wound reaction, the T-*6b*-induced callus shows many small cells originating from the cortex. Epidermis (E), cortex (C), vascular system (V). *Bars* 2 mm (**a**, **b**), 0.3 mm (**c**, **d**)

(i) *Leaf chlorosis* Chlorosis was only seen at an older leaf stage when *6b* no longer induced hyperplasia. Intact *6b* leaves and leaf disks in vitro accumulate high levels of hexose, known to lead to chlorosis (Clément et al. 2007). So far, ADP ribosylation has not been linked to sucrose accumulation and chlorosis.

(ii) *Hyperplasia* Previously, it was shown that *6b* induced ectopic divisions in the root pericycle and in seedling hypocotyls (Tinland et al. 1989, 1992; Grémillon et al. 2004), wounded stems (Hooykaas et al. 1988; Tinland et al. 1989; Helfer et al. 2002), and leaf veins (Tinland et al. 1990, 1992; Wabiko and Minemura 1996). In tobacco, *6b* induced ectopic photosynthetic cells (EPC) on the abaxial leaf side and ectopic vascular strands (EVS) in leaf blades, petioles and midveins. Extra division could also be induced in superficially wounded *N. rustica* stems, as shown here. Since *6b* induces cell division in different cell types, it probably targets a basic cell division mechanism. However, not all plant species and cell types react to *6b* by cell division suggesting that other factor(s) are required as well. The different results suggest that *6b*-induced hyperplasia is due to abnormal continuation of ongoing (or recently arrested) divisions. Part of the differences in plant and tissue sensitivity might be related to their capacity to re-initiate division. It would be interesting to study why certain cells are more sensitive to *6b* activity than others. EVSs seem to be derived from cortex parenchyma cells whereas the location of EPCs adjacent to the abaxial leaf epidermis suggests that they derive from lower spongy mesophyll cells. The transition from extra cell division to chlorosis in older leaves indicates that these two phenomena have something in common, this could be sucrose accumulation (Clément et al. 2007). In grafting experiments, EPCs could still be induced on leaves of 5–8 mm length well after the formation of the different cell types. Further EPC divisions led to leaf blades with inversed polarity (enations). In normal leaf development, the shoot apical meristem (SAM) determines the adaxial side of the incipient leaf via the epidermis (reviewed in Bar and Ori 2014; Fukushima and Hasebe 2014). It seems likely that the polarity of ectopic *6b*-induced structures is determined by normal structures nearby. Thus, the side of the EPC ridge facing the abaxial side of the normal leaf may acquire abaxial identity whereas EVS polarity may be imposed by nearby normal vascular strands. Studies using molecular markers will be required to establish how EPCs and EVSs are initiated and how polarity appears.

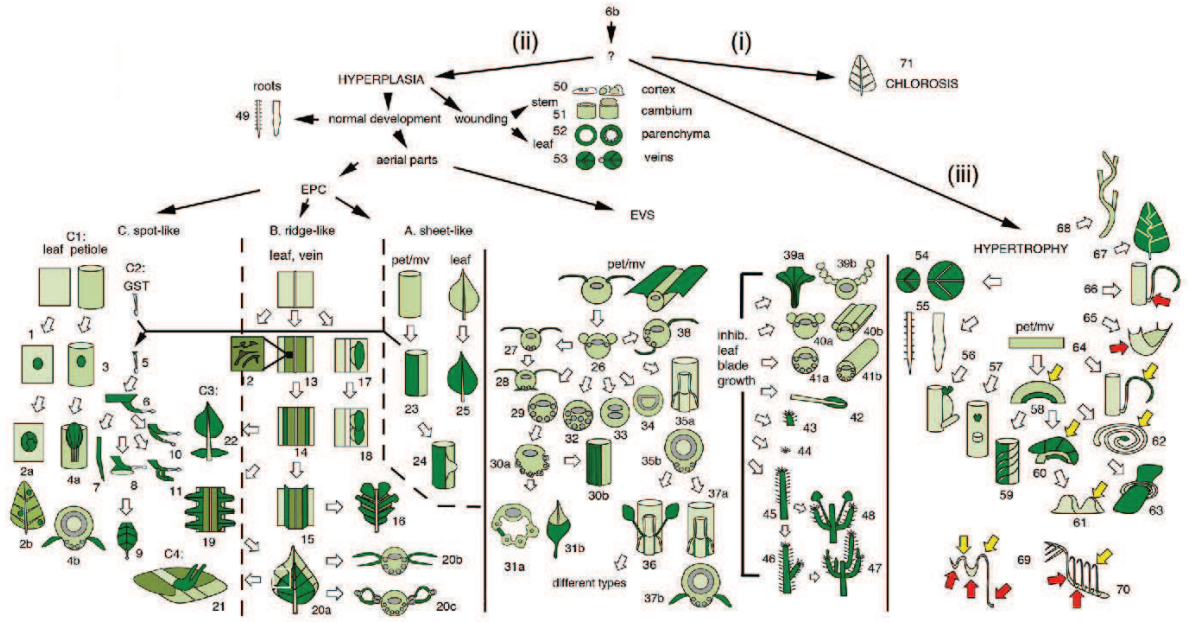


Fig. 12 Schematic overview of different phenotypic changes caused by *6b*. The *6b* gene causes chlorosis (*i*), *71*, hyperplasia (*ii*), *1–53* or hypertrophy (*iii*), *54–70*. Hyperplasia can occur during normal development, in aerial parts (1–48) or in roots (49). In aerial parts, it leads to ectopic photosynthetic cells (EPCs, 1–25) or ectopic vascular strands (EVSs, 26–38). EPC groups form sheet-like (*A*), ridge-like (*B*) and spot-like (*C*) meristems. Spot-like meristems can develop on some sheet-like meristems (*arrow* to 5 from 23), or on ridge-like meristems (19 and 22 from 14, 21 from 20a). Modifications related to inhibition of leaf outgrowth: *39–48*. Hyperplasia induced during wound reactions: *50–53*. Different views of the same structure are indicated by *a*, *b* and *c*. *Yellow arrows* epinastic changes, *red arrows* hyponastic changes. Short descriptions for numbers *1–71* and references to Figures are given in Table 1

EPCs form sheet-like (Fig. 12, A), ridge-like (B) and spot-like groups (C). A. Uniform planar EPC fields were found on the abaxial side of leaf L5-6, they did not further develop. We hypothesize that they are too large and homogeneous to give rise to primordia. Such planar EPC areas were also found on the abaxial side of petioles and midveins (in later leaves), and could develop further under special conditions (see below). B. Linear EPC groups. We propose that vein-associated EPCs are able to develop into linear meristems because there are no sink tissues on the outside of the strips. C. Spot-like EPC groups. In normal leaf development, auxin influx at the primordium tip leads to basipetal vein development and the outgrowth of a tubular primordium, followed by establishment of dorsoventral polarity and emergence of leaf blades from marginal meristems. Auxin convergence points on the leaf margins then lead to II veins that connect with the midvein (Bar and Ori 2014). Four types of ectopic spot-like meristems were found in 6b plants. C1. EPC leaf spots (Helfer et al. 2003; Grémillon et al. 2004). C2. GST tubes. GSTs are strong sinks and as such could induce the outgrowth of underlying EPCs; they are not needed to maintain GST tube growth. Most likely GST tubes rapidly become sinks themselves. GST tubes at the edge of the EPC layer form small leaflets, possibly due to the lack of competing sink tissues outside the EPC layer. Because GST tubes can be induced in a controlled way, appear at predictable positions (behind GSTs) and are easily accessible for observation and manipulation, they could become interesting models for leaf development. C3. Deltoid outgrowths along the midvein. Their spacing seems to be related to the branching points of the II veins of the normal leaf blade. C4. Corniculate tubes. They lack trichomes and might be initiated from EPCs by nutrients accumulating at the free ends of the veins.

Ectopic vascular strands (EVSs) were first reported in EPC spots in 2x35S-AB-*6b* tobacco leaves. These spots contain large amounts of starch, suggesting they are strong sinks (Helfer et al. 2003). In the present investigation, L5–6 contained EPCs but no EVSs. We therefore assume that EPCs precede EVS formation and that EVSs appear when a small group of EPCs acquires local sink activity. Randomly distributed EVSs were observed in 35S-AKE10-*6b* tobacco petioles (Terakura et al. 2006). Dex-AKE10-*6b* cotyledons contained many parallel vascular strands and abaxial outgrowths (Kakiuchi et al. 2006, 2007). According to our findings, these outgrowths may result from EPCs. EVSs seem to constitute a kind of elementary structure without

Table 1 Different morphological modifications found in *6b*-expressing plants

No.	Description	Figures/references
1	Hypothetical EPC without EVS	
2a	Spot with EPC and EVS in 2x35S-AB-*6b* tobacco, detail	Helfer (2001)
2b	As 2a, different spots on one leaf	Helfer et al. (2003)
3	EPC spot on petiole	S22
4a	Inverted leaf on petiole, side view	10f–h
4b	As 4a, cross section of stem	10i
5	Tube starting to grow behind GST	7a–c, S14g, S15a–c
6	Tube growing out and bending towards leaf apex, ridge	7b, S4c, S14b
7	Long tube without differentiation	S14a, S25a
8	Tube with leafy outgrowth	8a, b
9	Tube with leaflet with veins	8c, d, S16a–d
10	Branched tube with GST on each branch	S16e
11	Tube with symmetrical branches without GST	8e
12	Small random cell groups in narrow EPC areas along veins	S3a, b
13	Narrow EPC areas along veins	S3a
14	EPC areas along veins with ridges on outer edges	S3b–d
15	Outgrowth of enations along veins into leaf blades	S3c, d
16	II veins with enations restricting blade growth : wrinkling	S4a, b
17	Epiascidia along midveins	S6a, S25f, g
18	Complex double enations	S6b
19	Deltoid outgrowths	3a–c
20a	Double leaf with common II vein at the base	S4c
20b	As 20a, cross section at apical position	Helfer et al. (2003)
20c	As 20a, cross section at basal position with ring-like structures	S5a–c
21	Corniculate tubes at end of enation part of composite II veins	S17a–g
22	Isolated deltoid outgrowth along the petiole	3f, g, 10 k
23	EPC layer on abaxial side of pet/mv	2d, S11b
24	Fleshy outgrowth on edge EPC layer pet/mv	8 h
25	Uniform abaxial EPC layer on L5-6	2b, d
26	EVS close to normal vascular strands in pet/mv	6c
27	EVS at "bottom" of abaxial part of pet/mv	S12c
28	Outgrowth leaf blade from lateral edges "bottom" EVS	S12a–c
29	EVS along lower pet/mv surface	6d, h
30a	Outgrowth parenchyma around EVS	6f–h, S18e, f
30b	As 30a, longitudinal ridges along lower pet/mv	6e
31a	Hypoascidium, cross section, outgrowth of leaf blade	S18g
31b	Hypoascidium, side view	S18a, b
32	EVS along lower surface and more inside the pet/mv	S11a, g
33	Linear arrangement of EVS	S11g, S18c, d
34	Circular arrangement of EVS combined with normal VS	S11e–g
35a	Stem structure around petiole, petiole ridges	10b, c, m
35b	As 35a, cross section shows EVS	10e, S21a, b
36	Outgrowth leaflets (or other structures) at petiole base	10j, k
37a	Outgrowth subpetiolar blade from EVS at petiole base	10 k
37b	As 37a, cross section	S21c
38	Asymmetrical EVS with outgrowth leaf blade	S13a–g
39a	Parallel II veins at double leaf base	5a–e
39b	As 39a, cross section, hyponastic curvature	5d

Table 1 continued

No.	Description	Figures/references
40a	Semicircular ridges instead of petiole wings, cross section	6c
40b	As 40a, longitudinal view	6a
41a	Complete absence of outgrowth along petiole	6d, h
41b	As 41a, longitudinal view	6g
42	Long tubular leaf with small leaf blade at the tip	1c, S8b, c
43	Small pin without leaf blade, covered with thin trichomes	9a, S25a
44	"Empty site" below normal bud, thin trichomes	S20c–e
45	Long unbranched pin, no leaf blade	9b
46	Branched pin	9b
47	Pin with multiple branches with some differentiation	S19g
48	Pin with symmetrical secondary pins and leaflets	9c, S19a–e
49	Hypoplastic reaction in central root part	Grémillon et al. (2004)
50	Hypoplastic wound reaction in *N. rustica* cortex	11a–d
51	Hypoplastic wound reaction of cambium	Hooykaas et al. (1988)
52	Hypoplastic wound reaction on the edge of leaf holes	Clément et al. (2006)
53	Callus formation on leaf disks in vitro	Wabiko and Minemura (1996)
54	Expansion leaf disks	Clément et al. (2006)
55	Expansion root cells	Grémillon et al. (2004)
56	Increase in diameter of region between petiole and bud	S24d
57	Increase in distance between petiole and bud	S20a–e, S24a–d
58	Epinastic curvature of pet/mv in regions with EPC	S9: region 2
59	Torsion of pet/mv	5a, b, S8c–i, S23e
60	Epinastic curvature in regions with enation blade	S8d–f, S9: region 4
61	Up-down epinastic curvature in region of branching II veins	4d
62	Epinastic curvature of pet/mv leading to coiling	S8a–c
63	Double leaf rolled up by epinastic coiling: butterfly structure	S10a, b
64	Early epinastic curvature of midvein	S7b
65	Hyponastic curvature of gutter-like structure (see also 39b)	5a–e, 6a, S9: region 3
66	Early hyponastic curvature at base of petiole	S7b
67	Left–right curvature of midvein at branching points II veins	4a, b
68	Zig-zag stem curvature at branching points petioles	10b, c, S1b, c
69	Complex curvature with 5 different regions	S8a, d–i, S9
70	Complex curvature of leaves with « dips »	4e–g
71	Full or partial chlorosis	1c, S2

Numbers 1–71 refer to Fig. 12. Each modification is shortly described, full description in the text and figures of the Results section. *pet/mv* petiole/midvein, *EVS* ectopic vascular strands, *EPC* ectopic photosynthetic cells, *GST* glandular secreting trichomes

xylem maturation. Their initiation, growth, anatomical structure, interconnections and transport properties merit further investigation.

The absence of petiole wings and the gutter-like structures at the base of *6b* leaves could result from EVS development. Leaf blade outgrowth requires a normal vascular strand structure with free ends (Waites and Hudson 1995) and EVSs could disturb this structure. However, leaf blades can grow out from the "free ends" of EVS groups, as shown by leaves with "bottom enations", a leaf with asymmetric EVSs and plants with subpetiolar wings. The outgrowth of lamina between EVSs in single hypoascidia might be another case. Whereas epiascidia are relatively frequent, hypoascidia are rare, the remarkable hypoascidiate leaves of *Ficus benghalensis* var. Krishnae ("Krishna's buttercup", de Candolle 1902) are a well-known example. It will be interesting to compare natural and *6b*-induced hypoascidia. Inhibition of lamina outgrowth in *6b* plants is most evident in pins and at "empty sites" which lack any outgrowth. Pins and "empty sites"

are covered by long thin trichomes, their association with pins merits further investigation. Pins can form secondary pins and even leaf tissues, indicating that they retain some dorsoventral polarity. The anatomical structures of secondary pins and their connection to primary pins (venation pattern, branching mechanism) will require more analysis. The subpetiolar leaflets found in this study strongly resemble stipules. They arise at the connection points between normal petiolar and extra subpetiolar wings; the reason for this is unknown.

(iii) *Hypertrophy* Leaves and roots of *6b* plants expand in a sucrose-dependent way (Clément et al. 2006). Unequal expansion leads to different types of curvature. Petiole bends occurred close to the stem, the left–right and up–down bends of midveins and II veins likewise occurred close to branching points, indicating a special *6b* sensitivity for such branching points. Certain leaves showed 5 different curved regions along their main axis. The main curvature of *6b* leaves is epinastic, with preferential growth of the normal blade, this basic pattern is modified by the hyponastic petiole base, the hyponastic gutter-like structure and the hyponastic leaf apex. Torsion of the main axis further increases the three-dimensional complexity of *6b* leaves. The large upward shift of buds could also be related to cell expansion, but needs further analysis. Enations and other types of abnormal plant growth have been found in spontaneous variants and in various diseases (like in enation virus disease). It has long been recognized that the study of growth abnormalities can contribute to a better understanding of normal developmental processes (Masters 1869). The *6b* enation syndrome may be used to test recent models of leaf development (Bar and Ori 2014; Fukushima and Hasebe 2014). *6b* phenotypes are partially determined by the production of a phloem-mobile enation factor (Helfer et al. 2003), and the present study indicates this might be the *6b* mRNA. Several examples of mRNA transport in the phloem have been reported but the transport mechanisms have not yet been elucidated (Spiegelman et al. 2013). It is likely that the *6b* mRNA is loaded in source tissues and unloaded in sinks as shown for green fluorescent protein (GFP) mRNA in tobacco (Imlau et al. 1999). As *6b* induces ectopic sinks, this could restrict or redirect *6b* mRNA movement in complex ways. The development of a non-mobile but otherwise functional *6b* mRNA would be of great interest.

Further molecular studies are needed to understand the cellular effects of the *6b* gene. The homogeneous EPC layers of L5–L6 leaves are a promising system for such studies. *6b* also induces hyperplasia in superficially wounded *N. rustica* stem cortex cells; this simple system has the advantage that it only involves a few cell types and does not require transgenic plants.

Author contributions statement KC and LO conceived and designed research, conducted experiments and analyzed data. LO wrote the manuscript. Both authors read and approved the manuscript.

Acknowledgments K.C. was supported by doctoral Grant 2011679003 from the Chinese Scholarship Council. We thank the IBMP gardeners for providing plants.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Bar M, Ori N (2014) Leaf development and morphogenesis. Development 141:4219–4230

Canaday J, Gérard JC, Crouzet P, Otten L (1992) Organization and functional analysis of three T-DNAs from the vitopine Ti plasmid pTiS4. Mol Gen Genet 235:292–303

Chen K, de Borne Dorlhac, Szegedi E, Otten L (2014) Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*. Plant J 80:669–682

Clément B, Pollmann S, Weiler E, Urbanczyk-Wochniak E, Otten L (2006) The *Agrobacterium vitis* T-6b oncoprotein induces auxin-independent cell expansion in tobacco. Plant J 45:1017–1027

Clément B, Perot J, Geoffroy P, Legrand M, Zon J, Otten L (2007) Abnormal accumulation of sugars and phenolics in tobacco roots expressing the *Agrobacterium* T-*6b* oncogene and the role of these compounds in *6b*-induced growth. Mol Plant Microbe Interact 20:53–62

de Candolle C (1902) Nouvelle étude des hypoascidies de *Ficus*. Bull Herb Boiss 2:753–763

Fukushima K, Hasebe M (2014) Adaxial-abaxial polarity: the developmental basis of leaf shape diversity. Genesis 52:1–18

Gális I, Simek P, Van Onckelen HA, Kakiuchi Y, Wabiko H (2002) Resistance of transgenic tobacco seedlings expressing the *Agrobacterium tumefaciens* C58-*6b* gene to growth-inhibitory levels of cytokinin is associated with elevated IAA levels and activation of phenylpropanoid metabolism. Plant Cell Physiol 43:939–950

Gális I, Kakiuchi Y, Simek P, Wabiko H (2004) *Agrobacterium tumefaciens* AK-*6b* gene modulates phenolic compound metabolism in tobacco. Phytochemistry 65:169–179

Grémillon L, Helfer A, Clément B, Otten L (2004) New plant growth-modifying properties of the *Agrobacterium* T-*6b* oncogene revealed by the use of a dexamethasone-inducible promoter. Plant J 37:218–228

Helfer A (2001) Influence des genes *orf8* et *6b* d'*Agrobacterium* sur la croissance végétale. Dissertation, University of Strasbourg

Helfer A, Pien S, Otten L (2002) Functional diversity and mutational analysis of *Agrobacterium* 6B oncoproteins. Mol Genet Genomics 267:577–586

Helfer A, Clément B, Michler P, Otten L (2003) The *Agrobacterium* oncogene AB-*6b* causes a graft-transmissible enation syndrome in tobacco. Plant Mol Biol 52:483–493

Hooykaas PJJ, den Dulk-Ras H, Schilperoort RA (1988) The *Agrobacterium tumefaciens* T-DNA *6b* is an oncogene. Plant Mol Biol 11:791–794

Imlau A, Truernit E, Sauer N (1999) Cell-to-cell and long-distance trafficking of the green fluorescent protein in the phloem and symplastic unloading into sink tissues. Plant Cell 11:309–322

Ishibashi N, Kitakura S, Terakura S, Machida C, Machida Y (2014) Protein encoded by oncogene *6b* from *Agrobacterium tumefaciens* has a reprogramming potential and histone chaperone-like activity. Front Plant Sci 5:1–7

Ito M, Machida Y (2015) Reprogramming of plant cells induced by 6b oncoproteins from the plant pathogen *Agrobacterium*. J Plant Res 128:423–435

Kakiuchi Y, Gális I, Tamogami S, Wabiko H (2006) Reduction of polar auxin transport in tobacco by the tumorigenic *Agrobacterium tumefaciens* AK-*6b* gene. Planta 223:237–247

Kakiuchi Y, Takahashi S, Wabiko H (2007) Modulation of the venation pattern of cotyledons of transgenic tobacco for the tumorigenic *6b* gene of *Agrobacterium tumefaciens* AKE10. J Plant Res 120:259–268

Kitakura S, Fujita T, Ueno Y, Terakura S, Wabiko H, Machida Y (2002) The protein encoded by oncogene *6b* from *Agrobacterium tumefaciens* interacts with a nuclear protein of tobacco. Plant Cell 14:451–463

Kitakura S, Terakura S, Yoshioka Y, Machida C, Machida Y (2008) Interaction between *Agrobacterium tumefaciens* oncoprotein 6b and a tobacco nucleolar protein that is homologous to TNP1 encoded by a transposable element of *Antirrhinum majus*. J Plant Res 121:425–433

Komari T (1990) Genetic characterization of a double-flowered tobacco plant obtained by a transformation experiment. Theor Appl Genet 80:167–171

Levesque H, Delepelaire P, Rouzé P, Slightom J, Tepfer D (1988) Common evolutionary origin of the central portion of the Ri TL-DNA of *Agrobacterium rhizogenes* and the Ti T-DNAs of *Agrobacterium tumefaciens*. Plant Mol Biol 11:731–744

Masters M (1869) Vegetable teratology. An account of the principal deviations from the usual construction of plants. Ray Society, London

Smith EW, Seltmann H (1979) Anatomy of decurrent extension from winged petiole of *Nicotiana tabacum* L. Bot Gaz 140:324–327

Spiegelman Z, Golan G, Wolf S (2013) Don't kill the messenger: long-distance trafficking of mRNA molecules. Plant Sci 213:1–8

Takahashi S, Sato R, Takahashi M, Hashiba N, Ogawa A, Toyofuku K, Sawata T, Ohsawa Y, Ueda K, Wabiko H (2013) Ectopic localization of auxin and cytokinin in tobacco seedlings by the plant-oncogenic AK-*6b* gene of *Agrobacterium tumefaciens* AKE10. Planta 238:753–770

Terakura S, Kitakura S, Ishikawa M, Ueno Y, Fujita T, Machida C, Wabiko H, Machida Y (2006) Oncogene *6b* from *Agrobacterium tumefaciens* induces abaxial cell division at late stages of leaf development and modifies vascular development in petioles. Plant Cell Physiol 47:664–672

Terakura S, Ueno Y, Tagami H, Kitakura S, Machida C, Wabiko H, Aiba H, Otten L, Tsukagoshi H, Nakamura K, Machida Y (2007) An oncoprotein from the plant pathogen *Agrobacterium* has histone chaperone-like activity. Plant Cell 19:2855–2865

Tinland B, Huss B, Paulus F, Bonnard G, Otten L (1989) *Agrobacterium tumefaciens 6b* genes are strain-specific and affect the activity of auxin as well as cytokinin genes. Mol Gen Genet 219:217–224

Tinland B, Rohfritsch O, Michler P, Otten L (1990) *Agrobacterium tumefaciens* T-DNA gene *6b* stimulates rol-induced root formation, permits growth at high auxin concentrations and increases root size. Mol Gen Genet 223:1–10

Tinland B, Fournier P, Heckel T, Otten L (1992) Expression of a chimaeric heat-shock-inducible *Agrobacterium 6b* oncogene in *Nicotiana rustica*. Plant Mol Biol 18:921–930

Wabiko H, Minemura M (1996) Exogenous phytohormone-independent growth and regeneration of tobacco plants transgenic for the *6b* gene of *Agrobacterium tumefaciens* AKE10. Plant Physiol 112:939–951

Wagner GJ, Wang E, Shepherd W (2004) New approaches for studying and exploiting an old protuberance, the plant trichome. Ann Bot 93:3–11

Waites R, Hudson A (1995) *phantastica*: a gene required for dorsoventrality of leaves in *Antirrhinum majus*. Development 121:2143–2154

Wang M, Soyano T, Machida S, Yang JY, Jung C, Chua NH, Yuan YA (2011) Molecular insights into plant cell proliferation disturbance by *Agrobacterium* protein 6b. Genes Dev 25:64–76

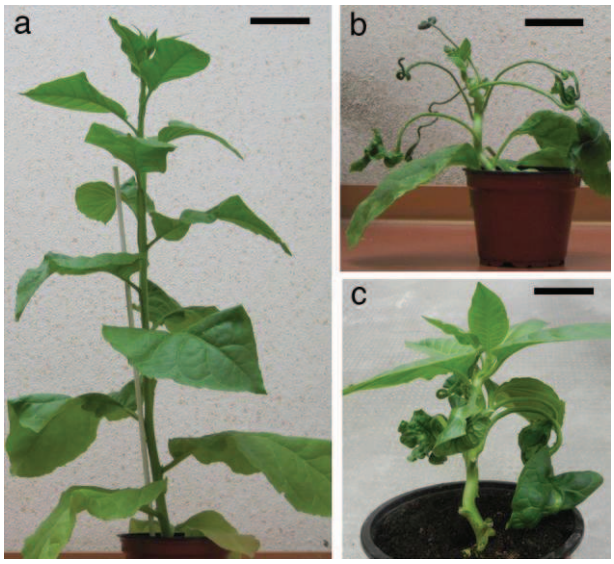


Figure S1. a Modification of 6b plant after induction at 4-leaf stage. Non-induced plant. b Induced plant. One month after induction by a single spray with 3 µM dexamethasone the plant is strongly modified and blocked for further growth. c Rare normalization after induction at the 4-leaf stage. After one month, the lower part of the plant is strongly modified, the upper part has normalized. Bars = 10 cm (a, b), 5 cm (c).

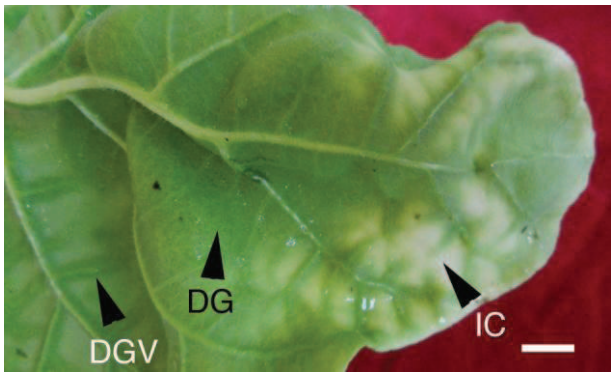


Figure S2. Chlorosis in L4 of a 6b plant, 2 months after induction
From apex to base three types of modification appear: intercostal chlorosis (IC), homogeneous abaxial dark green tissue (DG) and abaxial dark green tissue around II and III veins (DGV). Bar = 1 cm.

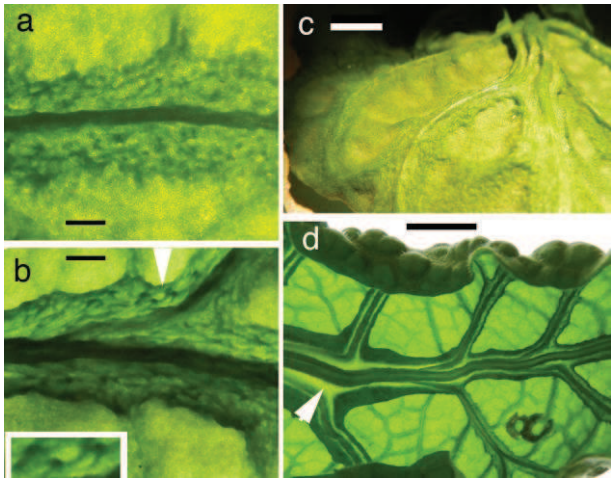


Figure S3. a Early stage of EPCs along veins: even distribution of EPCs in a narrow zone on both sides of a II vein. b Later stage: EPCs are growing preferentially along the outer edges of the narrow zones. Arrow: enlarged area shown in inset. Inset: random small cell groups within EPC area. c Same stage as in (b): ridges with EPCs along midvein and II veins start to emerge from the leaf blade. d Later stage: enation outgrowth along midvein

and II veins. Non-photosynthetic areas appear between veins and enations (arrow). Leaf margins are epinastic. Bars = 1 mm (a, b), 5 mm (c), 1 cm (d).

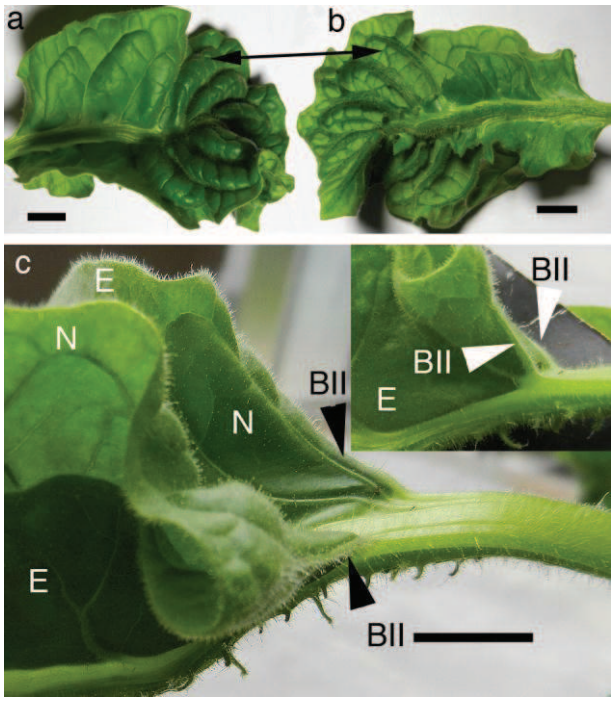


Figure S4. a Double leaf (adaxial side): the leaf is strongly wrinkled at the apex. b Same leaf as in (a), abaxial side; the wrinkled area corresponds to a region with enations along the II veins of the normal leaf, one is indicated by double arrows. Note that the enation leaf has independent II veins at the base, but shares its II veins with those of the normal blade at more apical positions. c Base of a double leaf, the basal II veins (BII) on each side of the midvein are composite veins; they are common to the normal (N) and enation (E) blades. Inset: slightly tilted view. The shared II vein projects upwards from the normal leaf plane. Bars = 1 cm (a, b, c).

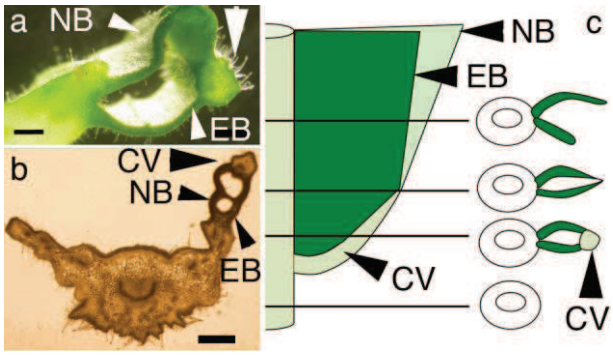


Figure S5. a Ring-like structure in a cross section at the base of a double leaf. Arrow: common II vein. b A similar structure as in (a) but with two rings. c Structure at the base of a double leaf (seen from the lower side) with cross sections at different levels. NB: normal blade, EB: enation blade, CV: common vein at the base of the double leaf. Bars = 5 mm (a), 2 mm (b).

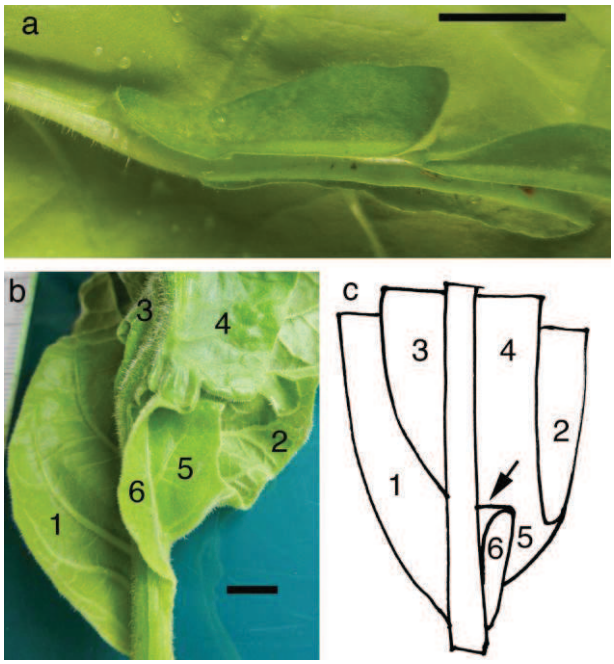


Figure S6. a Epiascidia along the midvein (adaxial surface inside). Three cup-like structures emerge along the midvein: two on one side, one on the other. b A complex double enation. The different parts of this folded structure are numbered (1-6). c Interpretation of the structure in (b): a short hypoascidium with a normal and enation blade (6) emerges on the lower right half of the leaf, its apical end is fused with the enation part (4, 5) of a larger hypoascidium emerging at a more apical position (arrow). Numbers: 1. Normal leaf blade, abaxial side (left). 2. Normal leaf blade, abaxial side (right). 3. Apical enation blade, adaxial side (left). 4. Apical enation blade, adaxial side (right). 5. Same blade as (4), the arrow indicates a transverse fold between part 4 and 5. 6. Adaxial part of the small hypoascidium, folded over to the left. Bars = 1 cm (a, b).

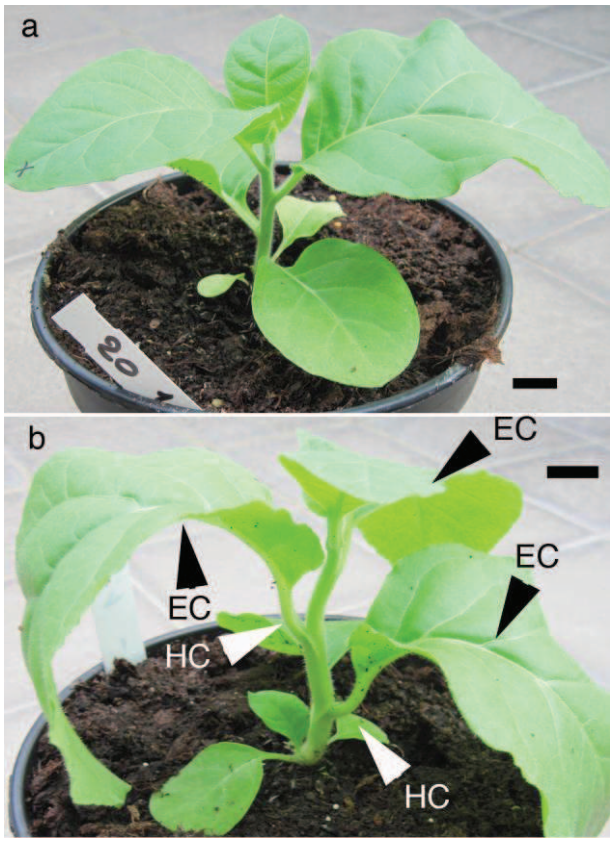


Figure S7. a Curvature of petioles and midveins. Non-induced plant. b Induced plant. Hyponastic curvature (HC) of the petioles and epinastic curvature (EC) of the midvein three days after induction of a 6-leaf 6b plant. Bars = 2 cm (a, b).

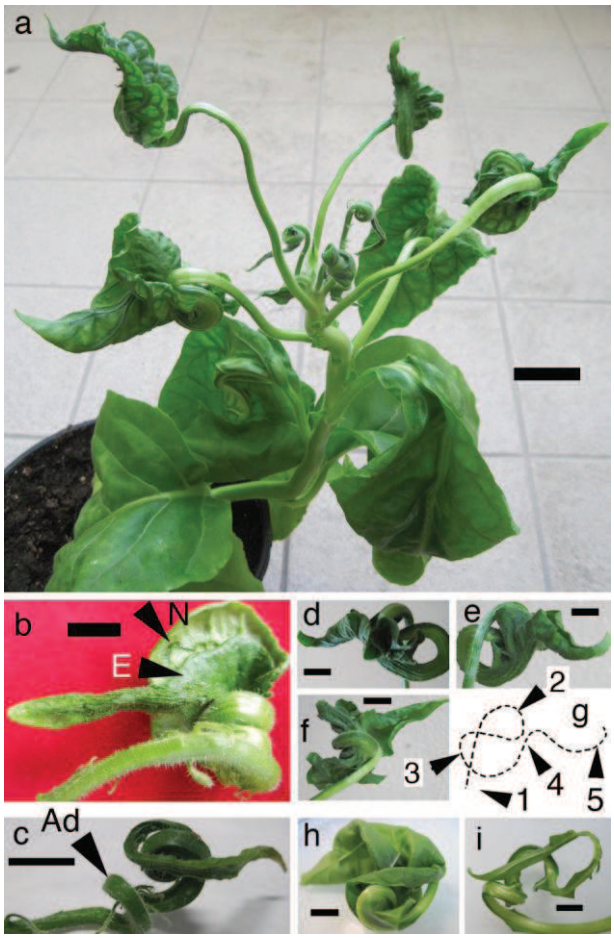


Figure S8. a A 6b plant induced at the 4-leaf stage, one month after induction. Several leaves have long thin petioles and complex curvatures with a consistent pattern (see also figure S9). b Tightly coiled epinastic leaf without gutter-like structure and with a small terminal double blade. N: normal blade, E: enation blade. c Strongly epinastically folded leaf with dark green ridges without a gutter-like structure and a very short leaf blade with a hyponastic tip. Ad: adaxial side. Note the angular petiole structure also seen in figure S25b. d-f Leaf with a gutter-like structure and complex folding pattern, shown from three different angles. g Interpretation of leaf shown in (d-f), oriented as in (e), see also figure S9. 1. Straight part. 2. Epinastic part with dark green ridges. 3. Hyponastic part with gutter-like structure. 4. Epinastic part with double leaf blade. 5. Hyponastic part with leaf apex. Torsion along the longitudinal axis leads to additional complexity. h A creased leaf without gutter-like structure. i Same leaf as in (h), leaf blade removed. The folding is due to an epinastic petiole-midvein structure with torsion along the longitudinal axis. Bars = 2 cm (a), 1 cm (b-f, h, i).

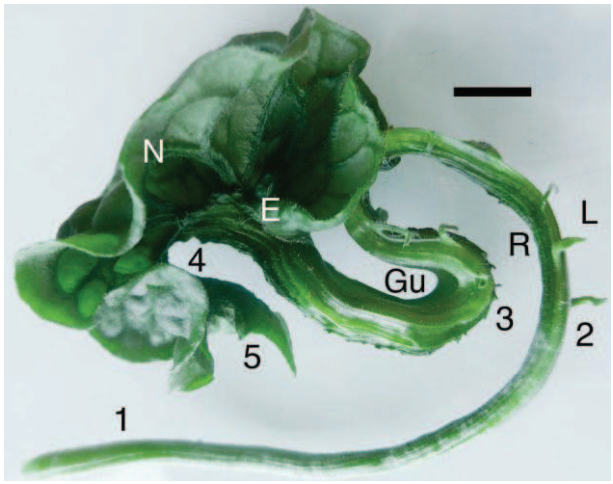


Figure S9. Complex curvature of an L10 leaf, with five different regions: 1. Long straight region growing upwards from a hyponastic bend (not shown) close to the stem. 2. Epinastic region with dark green ridges (R), on the edge small leaflets (L, see figure. 8) emerge. 3. Hyponastic region with gutter-like structure (Gu). 4. Epinastic region with normal (N) and enation (E) leaf blades. 5. Short hyponastic region: apical end of normal leaf blade. Bar = 1 cm.

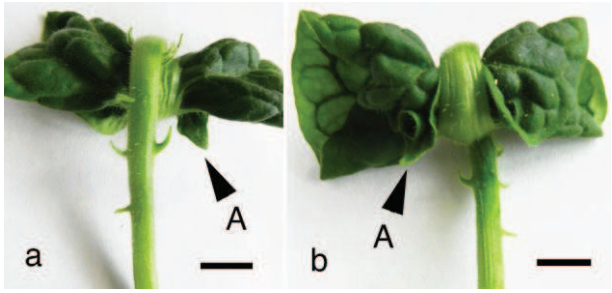


Figure S10. a Leaf with butterfly-like folding pattern. Adaxial view of petiole: strong epinastic curvature of petiole and midvein, the double blade is completely rolled up. b Abaxial view of petiole. A: leaf apex. Bars = 1 cm.

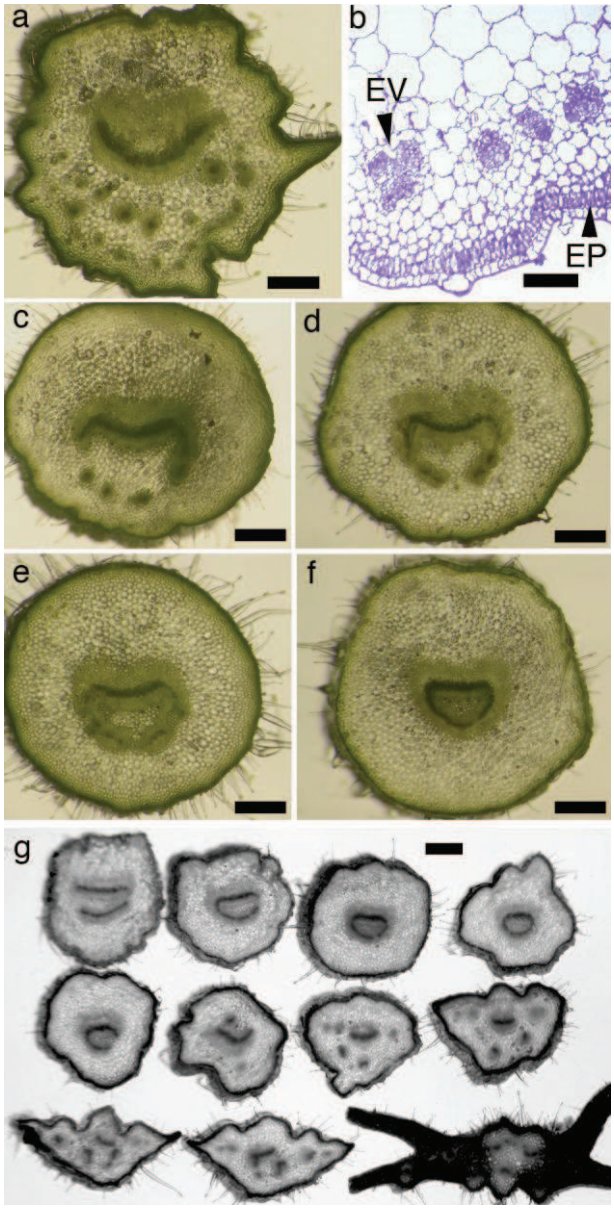


Figure S11. a Cross section of petiole, EVSs are found close to the lower surface and in more interior positions. b Thin section of petiole with EPCs (EP) and EVSs (EV) along the abaxial surface. c-f A series of cross sections

(from proximal to distal, with 1 cm distance between each section) of a petiole with EVSs. From base to apex the EVS pattern changes from individual spaced EVSs to a continuous circular arrangement. g A series of cross sections (proximal to distal) of the petiole and midvein of a late leaf. Various transitions can be seen in the EVS pattern: from an arc to a circular arrangement, which then breaks up in irregular patterns of individual EVSs. Bars = 2 mm (a, c-g), 0.4 mm (b).

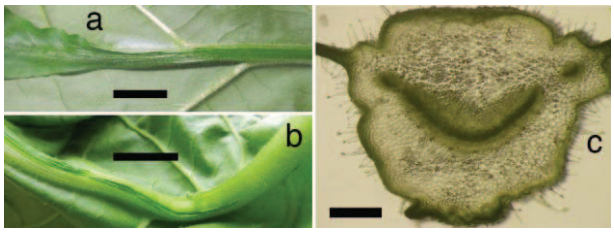


Figure S12. a A late leaf with narrow enations along the midvein, broadening at the leaf apex (on the left). b Another late leaf with partial enations along the midvein. c Cross section through midvein with narrow enations; they arise from a few isolated EVSs at the central “bottom” part of the abaxial side. Bars = 1 mm (a), 5 mm (b), 2 mm (c)

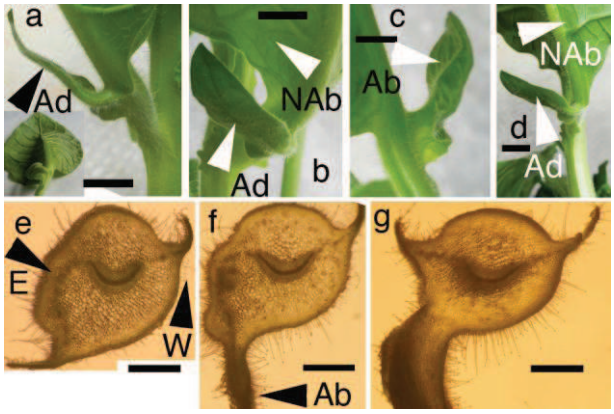


Figure S13. a Lateral view of ectopic leaf, growing on the modified side of a petiole of an asymmetric 6b leaf. Ad: adaxial side of leaflet. Inset: overview of asymmetric leaf: normal veins on the left, hyponastic II veins on the right. b-d Three other views of the same leaf. NAb: abaxial side of normal leaf, Ad: adaxial side leaflet, Ab: abaxial side leaflet. e Cross section of ectopic leaf: on the lower left a leaf blade emerges from the end of a series of unilateral EVSs (E). A petiole wing (W) is present on the normal part, but lacking on the modified side. f, g More basal sections, no EVSs occur below the ectopic leaf, and the left petiole wing reappears. Bars = 1 cm (a-d), 2 mm (e-g).

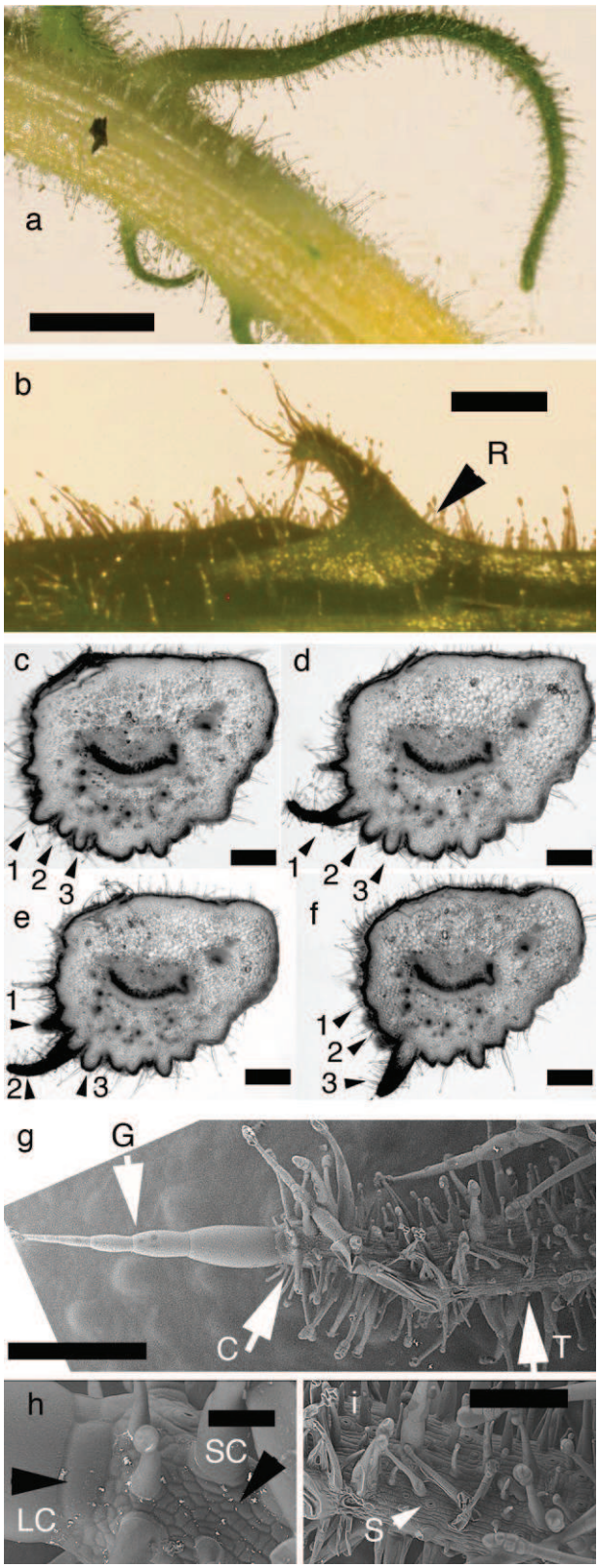


Figure S14. a A long GST tube on an old 6b plant, the petiole has yellowed but GST tubes remain dark green. b A GST tube situated on a longitudinal ridge (R), which ends shortly to the left of the tube. c-f Four consecutive cross sections of a petiole region with three GST tubes, tube 1 emerges in (d), 2 in (e) and 3 in (f). The GST tubes are not connected to the EVSs. g-i Scanning electron microscope pictures of a GST tube. g Overview, with connection (C) between GST (G) and tube (T) indicated by arrows. h Detail of connection between GST and tube with large cells (LC) from the GST base and small cells (SC) from the tube. i Detail of basal part of GST tube. S: stoma. Bars = 5 mm (a), 2 mm (b), 1 mm (c-g), 0.1 mm (h), 0.5 mm (i).

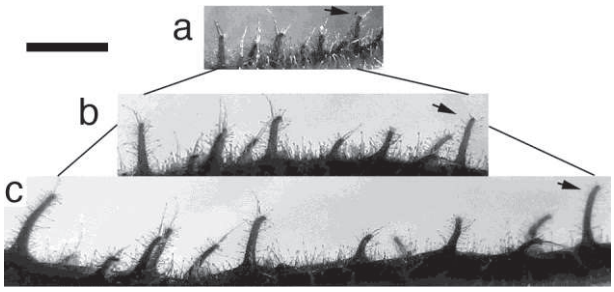


Figure S15. a Importance of GST for outgrowth of GST tubes. At an early stage of GST tube growth, the trichome on one of a group of GST tubes (arrow) was removed and subsequent growth was monitored. b, c Later growth stages, the GST-less tube (arrow) grows as well as the other tubes. Bar = 5 mm.

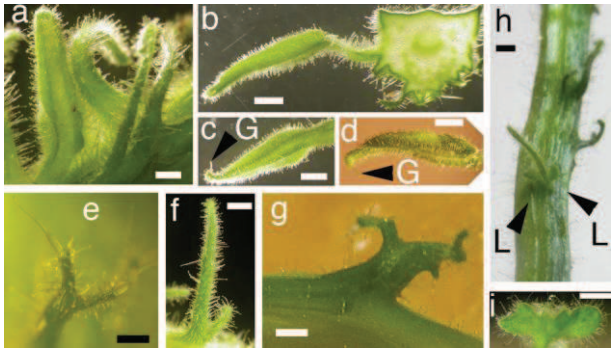


Figure S16. a A group of small leaflets developing from GST tubes. b A well-developed leaflet with II veins, seen from the side. c Same leaf as in (b), abaxial side. Narrow enation outgrowths. G: GST. d Same leaf as in (b), adaxial side. The leaf blade can be seen to emerge as a small ridge along the GST tube. G: GST. e Bifurcated GST tube, both parts are similar in size, each carries a GST. f Bifurcated GST tube, one branch is longer than the other. g Branched structure with a broad base. h GST tube carrying two leaflets (L) at its base. i Detail of isolated leaflets from structure in (h). Bars = 1 mm (a-e, g, i), 2 mm (f, h).

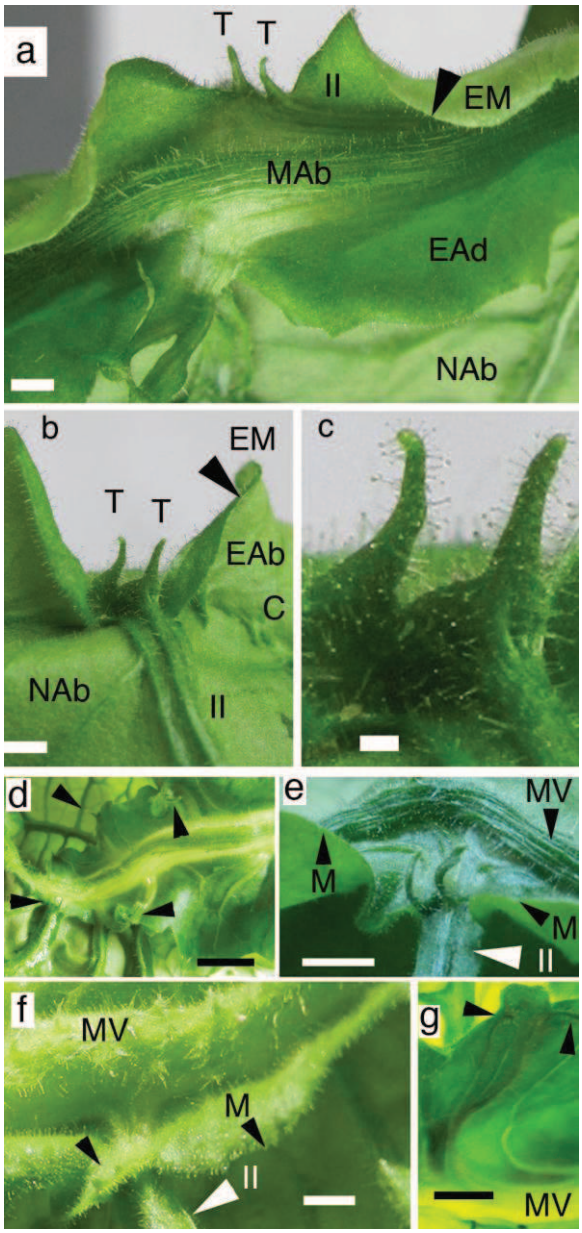


Figure S17. a Part of double leaf, seen from the adaxial side of the enation blade (EAd). Two corniculate tubes (T) emerge from the end of the enation part of the composite II vein (II). EM: enation blade margin. MAb: midvein abaxial side. NAb: normal blade, abaxial side. b Same region, seen from the abaxial side of the normal blade (NAb), with corniculate structures at the end of the enation part of the composite II vein (itself with small enations); the normal part of the II vein (marked "II") continues on the normal blade. T: tubes, EAb: abaxial side enation blade, EM: enation leaf margin, C: connection between enation leaf and normal leaf. c Detail of the corniculate tubes, these carry small trichomes but lack GSTs. d Abaxial side of double leaf. Corniculate tubes (arrows) are seen at the distal ends of the enation part of each composite II vein. e Detail of corniculate tubes at the margin (M) of an enation leaf. MV: midvein, II: secondary vein with small enations continuing on normal leaf blade. f Two small corniculate tubes (arrow) at an early stage of double leaf development. M: margin of enation leaf, MV: midvein, II: secondary vein. g Long corniculate tubes (arrows), MV: midvein. Bars = 1 cm (a, b, d), 2 mm (c, f), 3 mm (e), 5 mm (g).

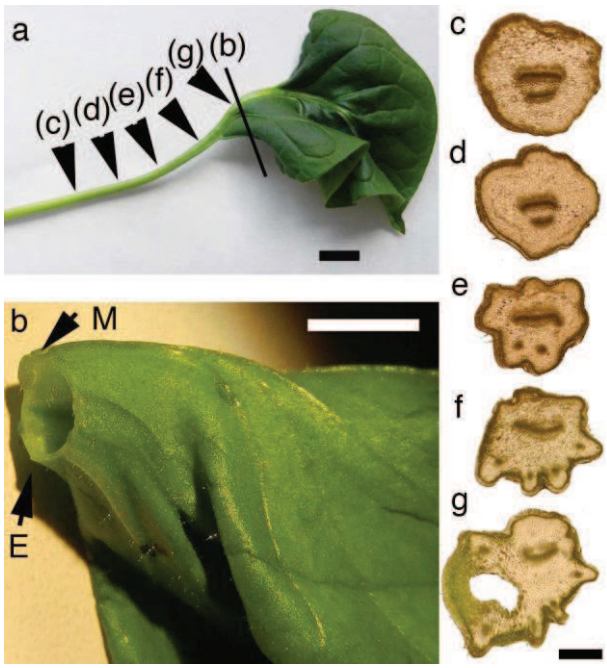


Figure S18. a Leaf with a hypoascidium (adaxial surface on the outside), the line marked (b) indicates the cross section shown in (b), the arrows marked (c) to (g) indicate cross sections shown in c-g. b Cross section along the line shown in (a). M: main vein, E: one of the EVSs. c-g Different cross sections of the petiole shown in (a). At the proximal side (c, d), the vascular system consists of a normal vascular system at the adaxial surface and a group of EVSs below it. Towards the leaf base (e, f), several isolated EVSs appear, surrounded by ridges. g Outgrowth of green lamina between two adjacent EVSs, leading to a cavity and a conical structure. Bars = 1 cm (a, b), 1 mm (c-g).

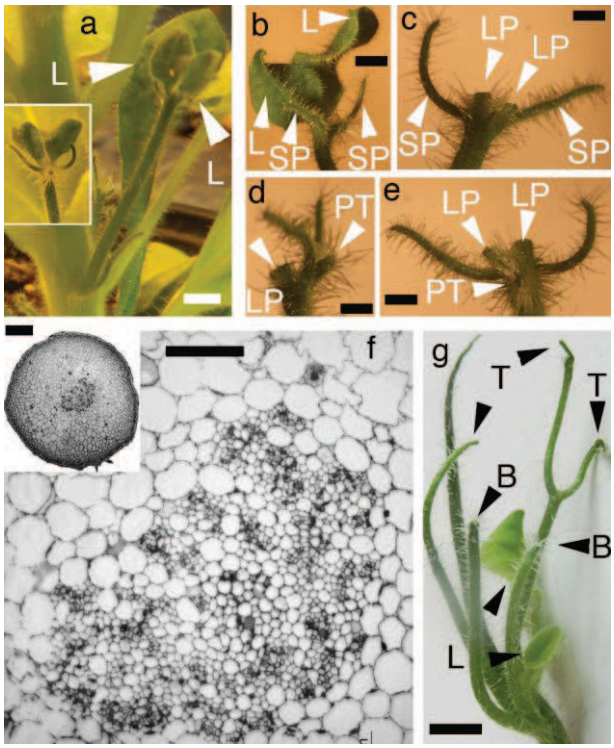


Figure S19. a Five-branched pin on 6b plant induced at 4-leaf stage, two months after induction. L: leaflets. Insert: structure at different angle. b Tip of the branched pin shown in (a), L: leaflets. SP: secondary pin. c-e Structure shown in (b) after removal of the leaflets, seen from three different angles. LP: leaflet petiole. PT: pin tip. SP: secondary pin. f Section through a pin, showing a central vascular system. The vascular system does not contain mature xylem vessels and is flattened on one side, indicating residual polarity (inset: overview). g Branched complex pin, some branches develop thin structures at their tips (T), others remain blunt (B), some carry leaf-like structures (L). Bars = 5 mm (a, g), 2 mm (b-e), 0.1 mm (f), 0.4 mm (inset f). See also figure. 9c, d.

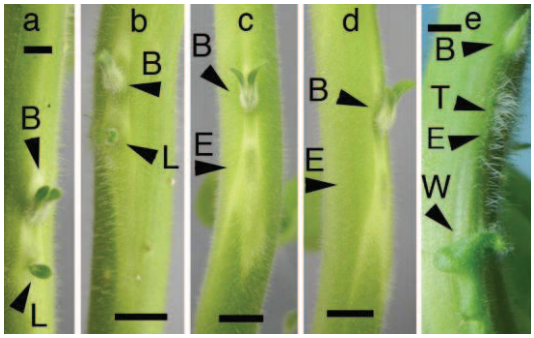


Figure S20. a, b Minute leafy structures (L) at petiole positions below normal-looking buds (B). c, d Two different views of an empty site (E) without a petiole, the bud (B) is normal. The surface structure at the empty site looks different. e Another empty site (E), seen from the side. B: bud, W: abnormal petiole wing. The empty site is covered by long thin trichomes (T) similar to those on pins. Bars = 5 mm (a, e), 1 cm (b-d).

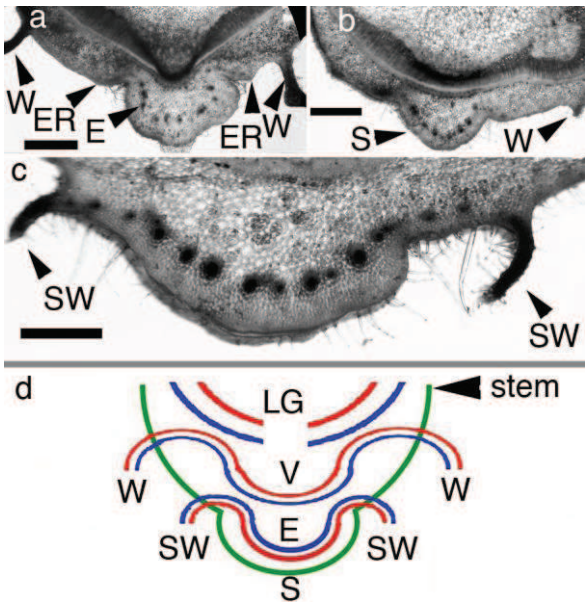


Figure S21. a Transverse stem section at the site of a subpetiolar ridge. On both sides, decurrent petiole wings are present (W). The EVS (E) series turns out towards the stem surface on both sides, leading to shallow EVS ridges (ER). b Another subpetiolar ridge (S), with a very small petiole wing (W), without EVS ridges or subpetiolar wings. c Subpetiolar ridge with outgrowth of narrow laminae (subpetiolar wings: SW). d Scheme showing polarity of subpetiolar structures: blue: abaxial side, red: adaxial side. LG: leaf gap, V: normal vascular system of petiole, E: EVS, W: petiole wing, SW: subpetiolar wing, S: subpetiolar ridge. Bars = 1 mm (a, b), 0.5 mm (c). See also figure. 10 d, e.

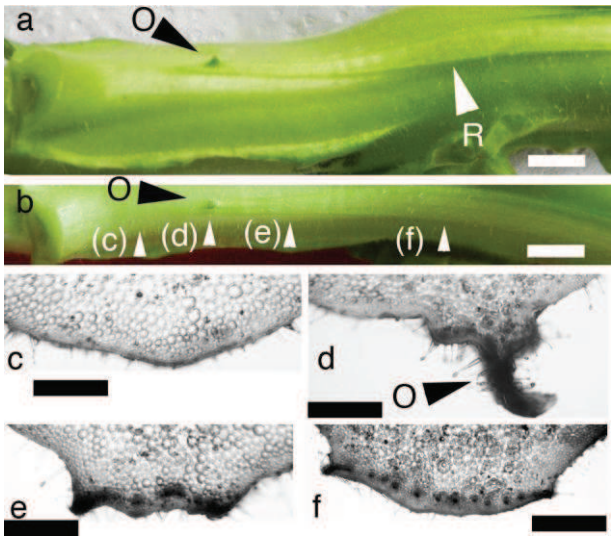


Figure S22. a A very small subpetiolar outgrowth (O) growing out from the stem, situated somewhat below the tip of a structure with small ridges (R) progressively separating towards the base of the stem. Side view. b Viewed from above, (c)-(f): positions of cross sections shown in c-f. c-f Cross sections as indicated in (b), no EVSs above the outgrowth, below it they show a regular pattern. Section in (d) shows the small leafy outgrowth (O). Bars = 5 mm (a, b), 1 mm (c-f)

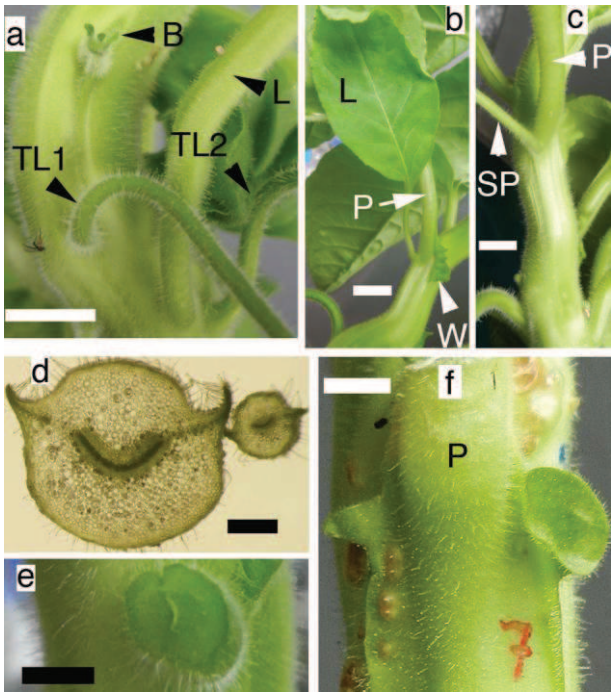


Figure S23. a Outgrowth of long tubular leaves (TL1 and TL2) on both sides of a central tubular leaf (L) identified through the position of the corresponding bud (B). TL1 is shifted upwards with respect to the expected insertion point. b, c Two different views of a plant with a well-developed leaflet (L) and its small petiole (SP) on one side of the normal petiole base (P) and a petiole wing (W) on the other side. d Cross sections of normal petiole and small leaflet petiole shown in b, c. e Petiole of 6b plant twisted counterclockwise by 90° at its base. f Cup-like structure at the petiole base of a 6b plant. Bars = 1cm (a-c, e), 0.4 cm (d), 0.5 cm (f).

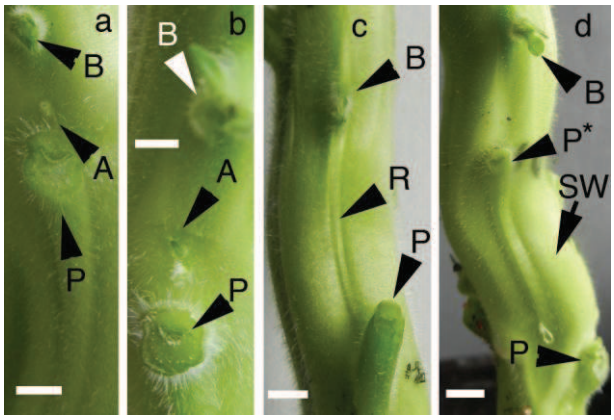


Figure S24. a, b Accessory buds. Two different plants (a and b) show an upward shift of the normal bud (B) with respect to the petiole (P) and an accessory bud (A). c Upward shifted bud (B) with ridge (R) running from the bud down to the petiole (P). d Swelling (SW) above the lower petiole (P). Bud (B) is shifted upward with respect to the upper petiole (P*). Bars = 0.5 cm.

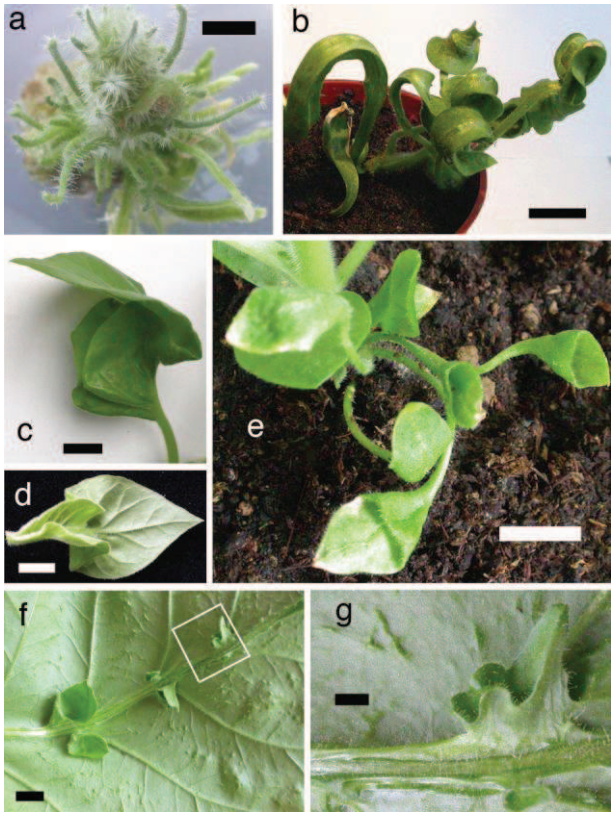


Figure S25. Phenotypes of leaky dex-T-6b tobacco plants
a Rootless in vitro plant, with numerous pins carrying long thin trichomes. b Plant with long ribbon-like "leaves". c, d Two views of a leaf that is strongly folded at the base, but normal at the apex. e Plant with small hypoascidia at its base. f Leaf with small epiascidia along the midvein, and small irregular cell groups within the lamina. Boxed: area shown in (g). g Detail from epiascidium boxed in (f). Bars = 5 mm (a), 2 cm (b-d), 1 cm (e, f), 2 mm (g).

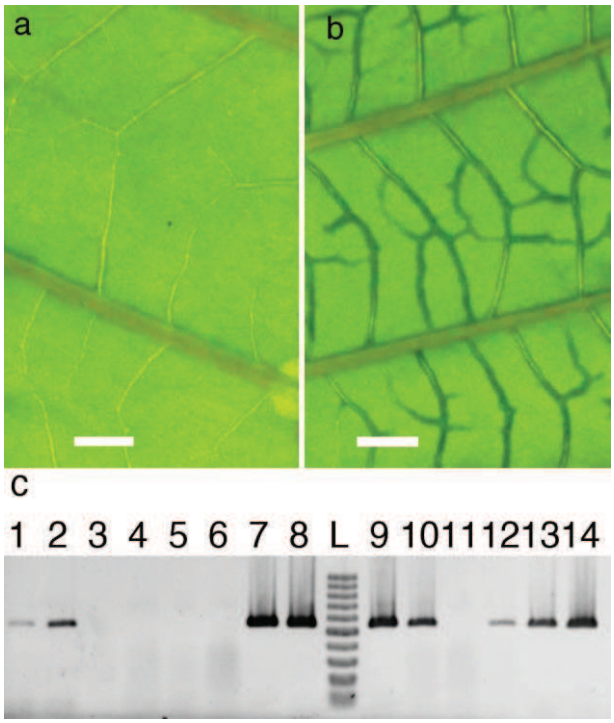


Figure S26. a Leaf of wild-type tobacco scion grafted on dex-T-6b rootstock before induction. b Leaf of wild-type tobacco grafted on dex-T-6b rootstock, one week after induction, with EPCs around III veins. c RT-PCR analysis of RNAs from different leaves of wild-type scion and dex-T-6b rootstock. Part of the 6b coding region (coordinates 32-577) was amplified by PCR. 1-3: 6b rootstock at day 0. Two of three leaves show weak leaky expression. 4-6: three different leaves of wild-type scion at 2 dpi. No amplification products. 7-9: three different leaves of 6b rootstock at 2 dpi. Each shows strong 6b expression. 10-12: Three leaves of wild-type scion at 5 dpi. Two contain 6b mRNA. 13-14: Two 6b rootstock leaves at 5dpi, both contain 6b mRNA. L: 100 bp ladder (from 100 to 1000 bp). The PCR products migrate at the expected size of 546 bp.

III. Additional data

III.1 Introduction of the TE-*6b* gene in *N. tabacum*

Since the publication of this work, we have started to investigate the possible activity of the *6b* gene from *N. otophora*, TE-*6b* (Chapter II). As a first step, the TE-*6b* gene was put under 2x35S promoter control and introduced into *N. tabacum*, which lacks this gene.

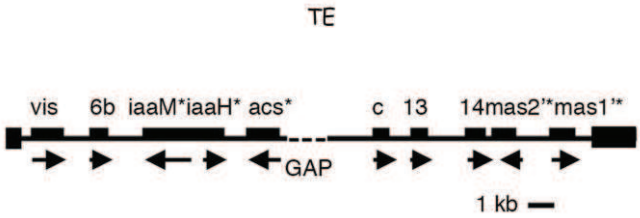


Figure 16. Provisional map of *N. otophora* TE region
This map is based on the manual assembly of various small contigs. Unfortunately, this map is not yet fully assembled. Arrow: direction of transcription. GAP: an unknown region between left and right part. *: truncated gene.

Different accessions of *N. otophora* exist in different seed banks. We chose accession TW95 from the US Department of Agricultural Research Service since it was used to obtain the deep-sequencing data (Sierro et al., 2014). The TE-*6b* gene from *N. otophora* was put under 2x35S promotor control using the pCK GFP S65C and pBI121.2 vectors (materials and methods) and transferred to *N. tabacum* cultivar Samsun nn. About 40 independent TE-*6b* regenerants were cultured *in vitro*, their leaves showed various abnormalities: narrow leaf blades, thicker veins, parallel running secondary veins and ectopic leaves growing on older leaves (figure 17). The TE-*6b* RNA expression levels from 31 independently regenerated R0 plants (TE-*6b*-1 to TE-*6b*-31) were measured by qPCR and found to be variable as expected (figure 17).

The seeds from TE-*6b*-1 to TE-*6b*-31 were harvested and sown on M0222 medium with kanamycin in order to identify single locus plants. 19 were single locus plants: 1, 2, 4, 5, 7, 8, 9, 10, 12, 13, 14, 16, 19, 20, 23, 24, 26, 28, and 31. Kanamycin resistant plants were transferred from the plates to soil in order to visually select for homozygous plants.

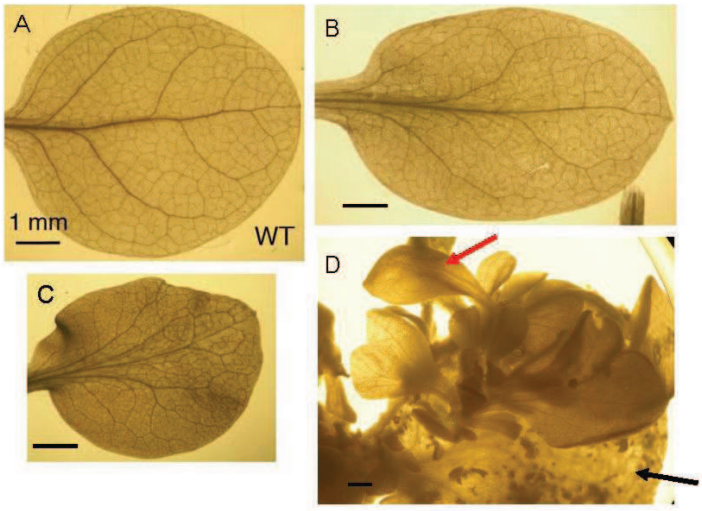


Figure 17. TE-*6b* leaves from *in vitro* plants cleared up by immersion in 70% ethanol
(A) WT *N. tabacum* leaf. (B)-(D) transgenic TE-*6b* *N. tabacum* leaves. (B) A narrow TE-*6b* leaf. The angles between midvein and secondary veins are smaller compared to the WT leaf (B and C). (D) Abnormal leaves growing directly from the leaf surface. Black arrow: normal leaf. Red arrow: abnormal leaf. Bars: 1 mm.

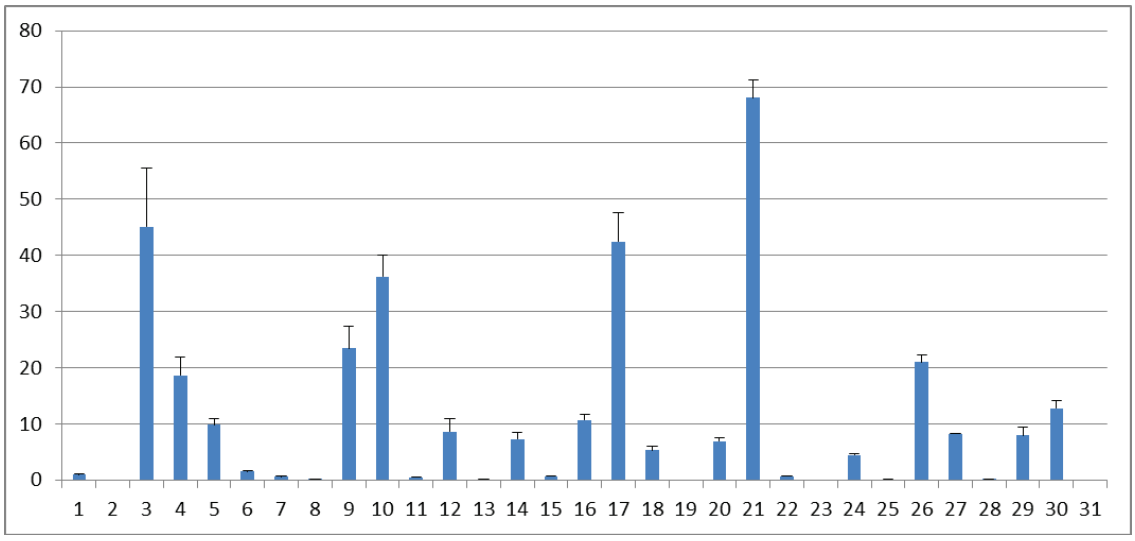


Figure 18. TE-*6b* expression levels in 31 R0 plants measured by qPCR
1 cm diameter leaf disk samples were taken from leaves about 5 cm wide. Values were calculated relative to the one from TE-*6b*-1, chosen as 1.

Typical R0 TE-*6b* tobacco plants are presented in figure 19. A first investigation on the morphology of these plants showed some abnormalities; although they did not appear systematically on all plants, we believe they are nevertheless related to the expression of the TE-*6b* transgene. A striking phenotype is seen on the TE-*6b*-1 plant: two leaves share a common petiole and part of the midvein (figure 19A). Relatively thick and transparent secondary veins were found on the TE-*6b*-5 plant (figure 19B, lower). An abnormal petiole wing with small triangle leaf like structure was found on TE-*6b*-5 plant (figure 19C, right). The TE-*6b*-8 plant had narrow and thicker leaves (figure 19D). Chlorotic leaves were present on the TE-*6b*-13 plant (figure 19E).

Dark green leaves with a wrinkled surface were very common in TE-*6b* plants with high expression. Such plants also showed stunted growth as for example the TE-*6b*-21 (figure 19F) and the TE-*6b*-5 (figure 20) plants.

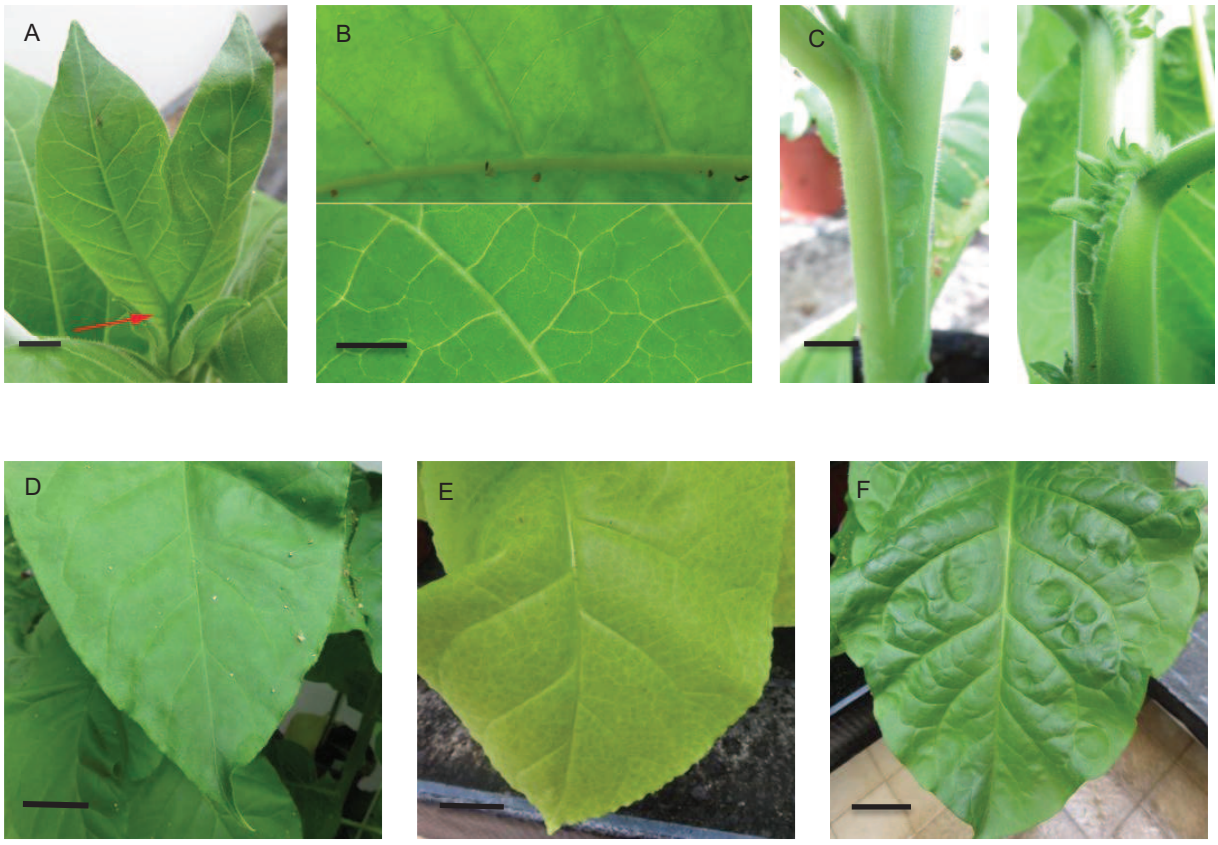


Figure 19. Phenotypes of different R0 TE-*6b* plants
(A) TE-*6b*-1. Two leaves sharing one petiole and part of the midvein. Arrow: divergence point of midvein. (B) Upper: a leaf from WT plant. Lower: a leaf from TE-*6b*-5 plant with transparent secondary veins. (C) Left: a petiole with its petiole wing from WT plant. Right: A petiole with abnormal petiole wing from TE-*6b*-5. (D) TE-*6b*-8. A narrow and thick leaf. (E) TE-*6b*-13. A chlorotic leaf. (F) TE-*6b*-21. A dark green leaf with a wrinkled surface. Bars from A-C: 1 cm. Bars from D-F: 3 cm.

We attempted to select for homozygous single copy plants in the F1 generation by phenotype and qPCR analysis. As an example, the TE-*6b*-5 is chosen. The kanamycin resistant plants of the F1 generation of the single locus TE-*6b*-5 plant with high TE-*6b* expression levels showed two phenotypes (figure 20): small plants (3 out of 9, figure 20 A-C) and larger plants (6 out of 9, figure 20 D-I), both different from a line with very low TE-*6b* expression, TE-*6b*-23. The TE-*6b* expression of these 9 plants was measured by qPCR in order to select homozygous plants (figure 21). The TE-*6b* expression levels of TE-*6b*-5 plants A and B are relatively high and correspond to the stronger plant phenotype but not C (figure 20 and 21). This might be explained by the limited sensitivity of the qPCR method. Plants with highest expression levels were kept for seed production.

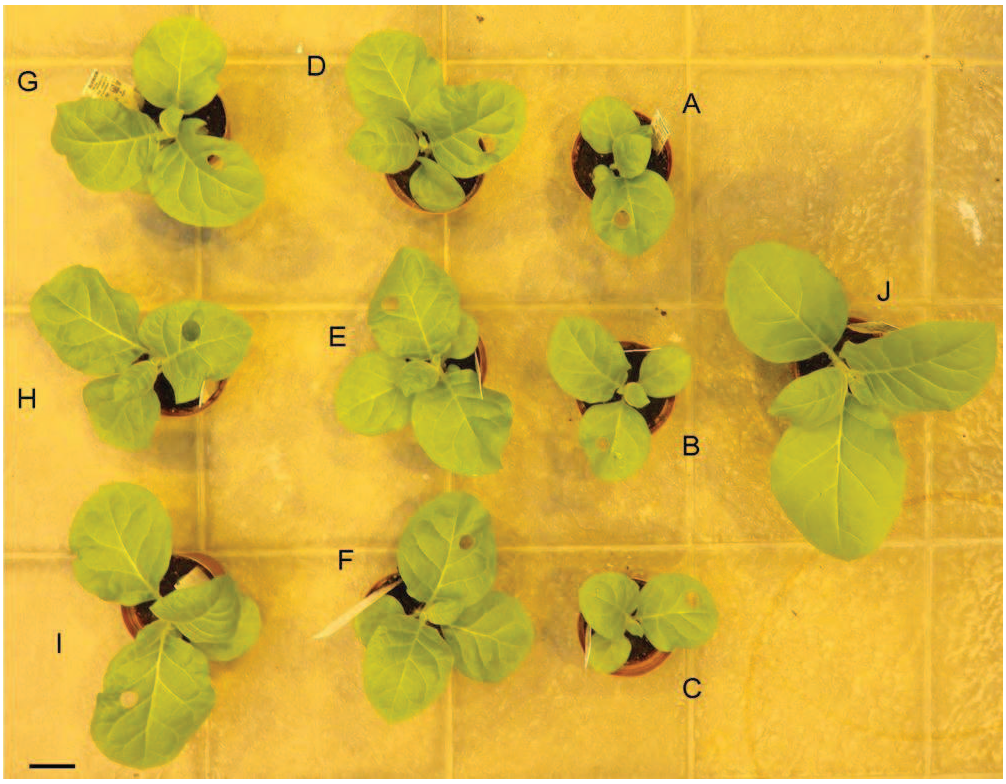


Figure 20. Selection of homozygous TE-*6b*-5 plants
TE-*6b*-5 kan^{R} plants were selected *in vitro.* 9 plants (A-I) were transferred on soil. (A-C) Plants of smaller size with stronger modifications. These are probably homozygous. (D-I) Bigger plants, possibly heterozygous. (J) A control plant from the low expression TE-*6b* line 23. Bar: 5 cm.

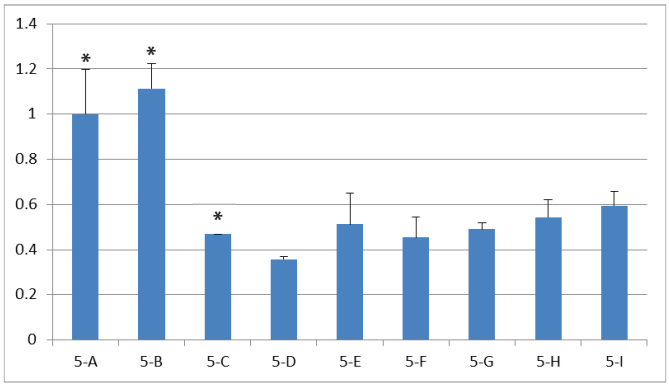


Figure 21. Expression levels of TE-*6b* in 9 TE-*6b*-5 F1 plants tested by qPCR
Plants 5 A-I correspond to A-I plants in figure 20. According to figure 20, A, B and C should be homozygous plants. However, plant C shows low TE-*6b* expression level. Values were calculated relative to the TE-*6b*-5 plant A, chosen as 1. *: plants with a more severe phenotype.

N. rustica stems were scratched with a needle and infected with Agrobacterium carrying an empty vector, a 2X35S::TE-*6b* or a 2X35S::T-*6b* construct. 75 days after infection, pictures were taken of the wounded stems (figure 22). The areas infected by Agrobacterium with the 2X35S::TE-*6b* construct showed the same wound reaction as the one infected by the Agrobacterium with the empty vector. However, the area infected by Agrobacterium with the 2X35S::T-*6b* construct

clearly showed tumor outgrowth. This experiment shows that the *N. rustica* stem is not sensitive to the TE-*6b* gene. Nevertheless, we can not exclude that the TE-*6b* gene could still act as an oncogenic gene on other plant species.

Figure 22. *N. rustica* stem infected by Agrobacterium carrying a 2X35S::TE-*6b* construct
From top to bottom, scratched *N. rustica* stem infected by Agrobacterium with empty vector, 2X35S::TE-*6b* construct (two different areas) and 2X35S::T-*6b* construct. Picture was taken 75 days after infection.

III.2 Conclusion and perspectives

III.2.1 Phenotypic comparison between T-*6b* and TE-*6b* plants

Since the TE-*6b* work is still going on, and homozygous F2 seeds from single copy TE-*6b* plants are not yet available, a phenotypic comparison between 2x35S::TE-*6b* and 2x35S::AB-*6b* plants (Helfer et al., 2003) or dex T-*6b* plants (Grémillon et al., 2004; Clément et al., 2006, 2007; Chen and Otten, 2015) is still in its early stage. However, some modifications of the TE-*6b* plants are similar as those of dex T-*6b* plants, most notably stunted growth, dark-green leaves, leaf wrinkling, chlorotic leaves and parallel running secondary veins, ectopic meristems and modified petiole wings (Chen and Otten, 2015). The precise anatomical structures of these phenotypes need further analysis.

III.2.2 A hypothesis on leaf blade outgrowth from free ends of vascular strands

In the petioles of a normal tobacco plant, petiole wings are formed at the lateral ends of the normal vascular system (figure 23 A and B). In the T-*6b* tobacco plant, several observations

showed the inhibition of petiole wing outgrowth in the presence of EVS (figure 23 C and D). Interestingly, we noticed that leaf blade (petiole wing) formation could still occur, at the free ends of the string-like series of parallel running vascular strands. Four examples illustrate this observation (figure 23 E-L). A handcut cross section was made from an ectopic leaf petiole which grew on an asymmetric T-*6b* leaf (figure 23E). Figure 23E shows that EVS were only formed on the left side of the petiole and the petiole wing grew out exactly at the end of these EVS. Another example shows a T-*6b* stem with a subpetiolar ridge and two narrow laminae growing out from the free ends of the EVS series (figure 23G). In a third case, a small leaf blade was growing on the abaxial side of a midvein and was associated with a few small vascular strands on the abaxial side of the midvein (figure 23I). Finally, a hand section from a cupshaped leaf showed leaf lamina tissue formed between two individual well separated vascular strands (figure 23K). According to these different observations, we propose that the free ends of EVS are necessary for the outgrowth of leaf laminae. However, free ends of vascular strands do not always induce leaf blade outgrowth. In many cases, only EVS were found on T-*6b* petiole or stem without any leaf tissue. The exact mechanism of the relation between free EVS ends and leaf blade outgrowth still remains to be studied. An important question that should also be addressed is the polarity of the EVS.

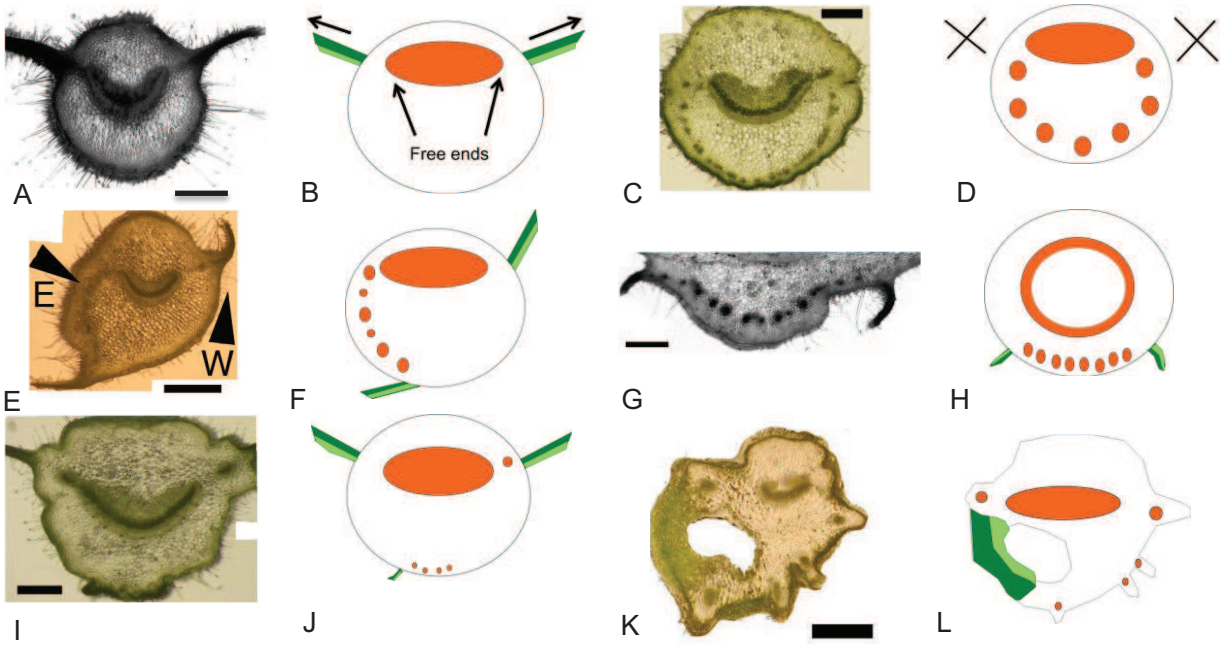


Figure 23. Leaf blade outgrowth induced by free ends of vascular strands
(A) Cross section of a petiole from WT tobacco plant. Publication 2, figure 6b. (B) Scheme of A. Outgrowth of the lamina occurs at the free ends of the vascular system. (C) Cross section of a petiole from a T-*6b* tobacco plant with additional vascular strands (EVS). Petiole without wing. Publication 2, figure 6d. (D) Scheme of C. Crosses indicate loss of petiole wings. (E) Cross section of an abnormal petiole with its asymmetric wing. Publication 2, figure S13e. (F) Scheme of E. (G) Subpetiolar ridge of stem with outgrowth of small laminae. Publication 2, figure S21c. (H) Scheme of G. (I) Cross section through midvein with narrow enations; they arise from EVSs at the central "bottom" part of the abaxial side. Publication 2, figure 12c. (J) Scheme of I. (K) Cross section from the base of a cup shaped leaf from a T-*6b* tobacco plant. Publication 2, figure 18b. (L) Scheme of K. Orange: vascular system; dark green: adaxial leaf side; light green: abaxial leaf blades (B, D, F, H, J and L). Bars = 2 mm (A, I), 1 mm (C), 1 cm (E, K), 0.5 mm (G).

Chapter III: The characteristics of the *Nicotiana tabacum mas2'* gene

I. Introduction

An intact TB-*mas2'* gene which encodes mannopine synthase was found on the TB region in *N. tomentosiformis* (Chen et al., 2014). Apparently its expression level in *N. tomentosiformis* is extremely low (figure 15). In order to study its capacity to encode a functional TB-Mas2' enzyme, it was put under 2x35S promoter control and expressed in *N. benthamiana*. It was found that the Mas2' enzyme from 35S::TB-*mas2'* lead to a product with the same electrophoretic mobility as DFG synthesized by the A4-Mas2' enzyme from *A. rhizogenes* A4 (Chen et al., 2014). Surprisingly, in some *N. tabacum* cultivars, the *mas2'* gene is very highly expressed according to RNAseq data (Publication 3, to be submitted). Such high expression might lead to detectable levels of DFG the production of which might affect normal plant growth by consuming sugars and amino acids. DFG synthesis could also have an effect on soil micro-organisms surrounding these plants and lead to a selection of DFG utilizing species. Such a selection has been shown to occur in *Lotus corniculatus* plants transformed with genes produce mannopine, mannopinic and agropinic acids from *A. rhizogenes* strain 8196 (Guyon et al., 1992). The increased density of octopine-using bacteria was found in the rhizosphere of transgenic *A. thaliana* which is able to synthesize octopine (Mondy et al., 2014).

The following manuscript describes details about *mas2'* gene expression levels in different *N. tabacum* cultivars according to both RNAseq and qPCR data. Possible reasons of the strong variation in expression are studied. A comparison of the DFG produced by *N. tabacum* with chemically synthesized DFG will be presented.

II. Publication 3

High *mas2'* gene expression and opine synthesis in wild-type *Nicotiana tabacum* (to be submitted)

Ke Chen[1], François Dorlhac de Borne[2], Emilie Julio[2], Julie Obszynski[3], Patrick Pale[3], Léon Otten[1]

1. Department of Molecular Mechanisms of Phenotypic Plasticity, Institut de Biologie Moléculaire des Plantes, Rue du Général Zimmer 12, 67084 Strasbourg, France. 2. Imperial Tobacco Bergerac, La Tour, 24100 Bergerac, France. 3. Laboratoire de synthèse, réactivité organiques et catalyse, Institut de Chimie, UMR 7177, Université de Strasbourg, 4 Rue Blaise Pascal, 67070, Strasbourg, France

Author for correspondance: Léon Otten, IBMP, Rue du Général Zimmer 12, 67084 Strasbourg, telephone: 33-3-67155336, fax: 33-3-88614442, e-mail: leon.otten@ibmp-cnrs.unistra.fr

Abstract

Nicotiana tabacum contains several natural T-DNA sequences inherited from its parental ancestor *N. tomentosiformis*. Among these, the TB region carries an intact mannopine synthase (TB-*mas2'*) gene. This gene is related to the *Agrobacterium rhizogenes* A4-*mas2'* gene. It is expressed at a very low level in *N. tomentosiformis* and most *N. tabacum* cultivars, but in a number of cultivars it was expressed to high levels. No differences were found in the TB-*mas2'* promoter regions of low and high expressers. A pTB-mas2'-GUS reporter gene was compared to a pA4-mas2'-GUS gene in stably transformed *N. benthamiana* plants. Both have similar expression patterns and are mainly expressed in the root tips and leaf vasculature. Tobacco cultivars with high TB-*mas2'* expression were found to contain a compound that co-migrated with DFG in a paper electrophoresis assay.

Keywords

Opine synthesis, natural transformant, T-DNA, *Nicotiana tabacum*

Introduction

Agrobacterium is well known for its capacity to transfer part of its DNA (T-DNA) to plant cells, under natural conditions this leads to crown gall tumors and hairy roots that synthesize small molecules used by the bacterium for its growth, the so-called opines (Ian et al., 1979; Vladimirov et al., 2015). Several types of opine synthesis genes have been described. Generally, they code for enzymes that combine two common metabolites into a product that can then be taken up by Agrobacteria and degraded into the constituent parts to serve as carbon, nitrogen and energy sources for the bacterium (Hansen et al., 1991). Among the different types of opines, the Amadori-type opines have been found in both tumors and hairy roots. Synthesis occurs in several steps, in the first one glutamine or glutamate is conjugated with glucose to *N*-1-deoxy-D-fructosyl-L-glutamine (DFG, commonly called santhopine) and *N*-1-deoxy-D-fructosyl-L-glutamate (DFGA) respectively, in the second step DFG is converted into mannopine (MOP) and DFGA into mannopinic acid (MOA), and finally mannopine is cyclised into agropine (AGR) and mannopinic acid into agropinic acid (AGA) (Hong et al., 1994; Kim et al., 2001). Other members of this group are chrysopine (the spiropyranosyl lactone of DFG), and *N*-1-deoxy-D-fructosyl-5-oxo-L-proline (DFOP). Mannopine/agropine synthesis genes have been identified on T-DNAs from different *A. tumefaciens* strains: the octopine/agropine strains A6 and Ach5, the succinamopine strain Bo542, the chrysopine strain Chry5, and on T-DNAs from different *A. rhizogenes* strains: the agropine strains A4 and 1855 and the mannopine strain 8196. The *mas2'* gene codes for DFG synthesis, *mas1'* for MOP synthesis and *ags* for AGR synthesis. Recently, *mas2'*, *mas1'* and *ags* genes

have been detected in the genomes of *N. tomentosiformis* and related species, among which *N. tabacum*. Nicotiana species of the Tomentosa section carry several T-DNAs that result from horizontal gene transfer through ancient *Agrobacterium* infections, these cellular T-DNAs are called cT-DNAs. One of these, called the TB region, carries TB-*mas2'*, TB-*mas1'* and TB-*ags*, organized in the same way as on pTi and pRi plasmids. The *mas1'* and *mas2'* genes are divergently transcribed and controlled by a so-called dual (divergent) promoter region situated between the two initiation codons, *ags* is controlled by its own promoter. In *N. tomentosiformis*, only TB-*mas2'* is intact; it encodes an active enzyme as shown by the appearance of a DFG-like compound upon over-expression in the cT-DNA-less *N. benthamiana* (Chen et al., 2014). Here we investigate the expression of TB-*mas2'* in different tobacco cultivars and show the presence of a DFG-like compound in roots of cultivars with high *mas2'* expression.

Results

DNA and protein sequences of *mas* regions and Mas proteins

The structure and organization of the *mas* region of the *N. tomentosiformis* TB region (KJ599827, Chen et al., 2014) were compared (figure 1) to those of *A. rhizogenes* (A4 and 8196) and *A. tumefaciens* (Ach5 and Bo542).

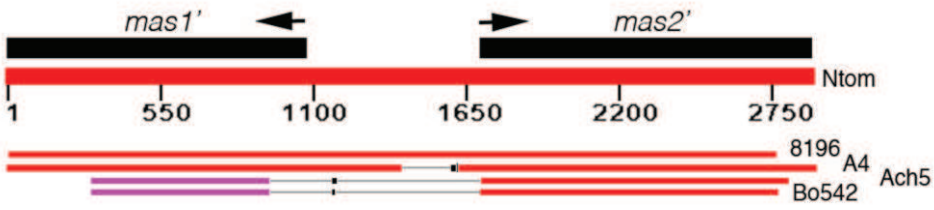


Figure 1. Alignment of *N. tomentosiformis mas* region (delimited by the stop codons of mas1' and mas2') with the corresponding regions from different Agrobacterium strains
Arrows indicate direction of transcription. Note that no similarity is found between the intergenic regions from TB and those of Ach5 and Bo542.

The *N. tomentosiformis* sequence mostly resembles the *A. rhizogenes* A4 sequence (86% identity). The intergenic region is partially similar to those of *A. rhizogenes* strains A4 and 8196, but very different from those of the *A. tumefaciens* strains Ach5 and Bo542 (see also below). The predicted Mas1' and Mas2' protein structures were compared to the published *Agrobacterium* proteins (including protein sequences from *A. larrymoorii*) (figure 2).

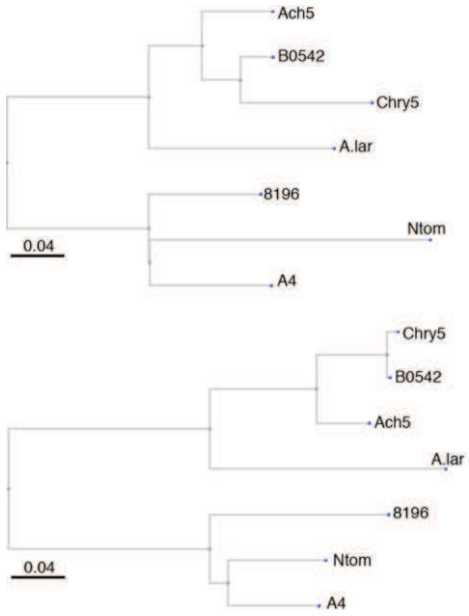


Figure 2. Protein phylogeny for Mas1' (top) and Mas2' (bottom)
The Mas proteins from *N. tomentosiformis* (Ntom) are most similar to the *A. rhizogenes* A4 and 8196 proteins.

The *N. tomentosiformis* Mas proteins cluster with those of *A. rhizogenes* A4 and 8196, and differ from those of *A. tumefaciens* Ach5, Bo542, Chry5 and *A. larrymoorii* (A. lar). Protein structures were analyzed for transmembrane domains with the PHOBIUS software (Stockholm Bioinformatics Centre). TB-Mas1' lacks transmembrane (TM) domains like A4-Mas2' and 8196-Mas2', whereas Ach5-Mas1', Bo542-Mas1' and *A. larrymoorii*-Mas1' contain TM domains (figure 3a). Conversely, TB-Mas2', A4-Mas2' and 8196-Mas2' contain TM domains, whereas Ach5-Mas1', Bo542-Mas1' and *A. larrymoorii*-Mas1' do not (figure 3b). No signal peptides were found. The Ags proteins from A4 and Ach5 do not contain TM domains (not shown). Thus, these data reveal two types of *mas* regions in *Agrobacterium* T-DNAs, with possible functional consequences that remain to be investigated.

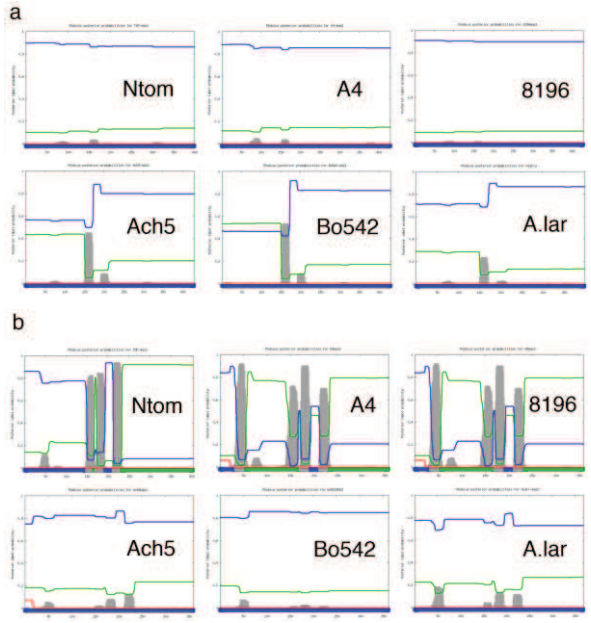


Figure 3. PHOBIUS analysis of Mas proteins. a. Mas1' proteins. b. Mas2' proteins
A. rhizogenes and *N. tomentosiformis* Mas proteins on the one hand, and *A. tumefaciens* Mas proteins on the other hand constitute two groups. A. lar: *A. larrymoorii*. Symbols: Grey: transmembrane, green: cytoplasmic, blue: non-cytoplasmic, red: signal peptide.

mas2' gene expression in different *Nicotiana* species

TB sequences with *mas*/*ags* genes have been found in *N. tomentosiformis* and *N. tabacum* (cultivars K326, Basma/Xanthi and TN90), *N. kawakamii* and *N. tomentosa* but not in the closely related species *N. otophora* (Chen et al. 2014). The latter contains an unrelated cT-DNA (TE) with mutated and presumably inactivated *mas*/*ags* genes. In a preliminary genome-wide RNAseq expression study using a mixture of root and leaf RNAs from *N. tomentosiformis* and 162 cultivars of *N. tabacum* (table 1) we found that most had very low *mas2'* expression levels like the *N. tabacum* parent *N. tomentosiformis* (low expression or LE cultivars).

Table 1. *N. tabacum* cultivars and *Nicotiana* species tested for TB-*mas2'* expression
N. tabacum cultivars and species for which preliminary RNAseq data were obtained. Plants that strongly express TB-*mas2'* according to these data and which were then confirmed by RT-qPCR analysis are marked in bold. nt: high TB-*mas2'* expression in RNAseq experiments but not tested by RT-qPCR.

	Type	Variety		Type	Variety
1	ANCESTRAL	AMBALEMA SR	85	BURLEY	Kentucky 17/BB16A
2	BRUN	ARAPIRA	86	BURLEY	**PULAWSKY 66**
3	BRUN	ARAPIRACA	87	BURLEY	**RUSSIAN BURLEY**
4	BRUN	BAHIA	88	BURLEY	SHIROENSHU
5	BRUN	Big Cuba	89	BURLEY	**SKRONIOWSKI 70**
6	BRUN	BRASIL AND	90	BURLEY	**START**
7	BRUN	BRASIL D	91	BURLEY	TENNESSEE 90
8	BRUN	BRESIL BAHIA	92	BURLEY	TN86
9	BRUN	BRESIL MATTA	93	BURLEY	WHITE BURLEY
10	BRUN	CABOT	94	BURLEY	**ZERLINA**
11	BRUN	CHEBLI	95	EXPERIMENTAL	BEINHART 1000-1

12	BRUN	COMSTOCK SPANISH	96	EXPERIMENTAL	FLORIDA 301
13	BRUN	CORDOBA = (TI 245)	97	EXPERIMENTAL	GAT 4
14	BRUN	COROJO	98	EXPERIMENTAL	KOKUBU
15	BRUN	**COULO (nt)**	99	EXPERIMENTAL	KUO FAN
16	BRUN	CRIOLLO	100	EXPERIMENTAL	TI657
17	BRUN	CRIOLLO CORRENTINO	101	ORIENT	BASMA DRAMA 2
18	BRUN	CRIOLLO ESPECIAL	102	ORIENT	BATROUM 1 (KFAR SIADE)
19	BRUN	DAMAS	103	ORIENT	BULGARE 2 (KFAR ROUMANIE)
20	BRUN	DRAGON VERT	104	ORIENT	**HERCEGOVINE**
21	BRUN	G1=SANTA CRUZ	105	ORIENT	IZMIR
22	BRUN	**GALPAO**	106	ORIENT	KABA KOULAK
23	BRUN	HAVANA CUBANO Q	107	ORIENT	SAMSOUN
24	BRUN	LITTLE CRITTENDEN	108	ORIENT	TOMBACK
25	BRUN	MACHU PICCHU	109	ORIENT	UCHATY KUCZERAVY
26	BRUN	MACOUBA=GRAND MARTINIQUE	110	ORIENT	XANTHI
27	BRUN	MANILLA	111	ORIENT SEMI	BANAT 125
28	BRUN	**MAZINGA (nt)**	112	ORIENT SEMI	BARAGAN 102
29	BRUN	NIJKERK	113	ORIENT SEMI	BOHEME ANDORRE
30	BRUN	NUCA TORCIDA	114	ORIENT SEMI	MULTAN
31	BRUN	ORONOCO RHODESIE	115	ORIENT SEMI	NOVA CRNJA
32	BRUN	PARAGUAY CLARO	116	ORIENT SEMI	ROUM
33	BRUN	PARAGUAY PERIGUEUX 48	117	VIRGINIE	4K78-5
34	BRUN	PARTIDOS	118	VIRGINIE	72C18
35	BRUN	PERALTERO	119	VIRGINIE	CASH
36	BRUN	PERIQUE	120	VIRGINIE	COKER 347
37	BRUN	**PETICO**	121	VIRGINIE	COKER 371 GOLD
38	BRUN	PHILIPPIN	122	VIRGINIE	DELCREST
39	BRUN	PORTO RICO	123	VIRGINIE	Delgold
40	BRUN	RIVET	124	VIRGINIE	DIXIE BRIGHT LEAF 101
41	BRUN	ROBUSTA	125	VIRGINIE	ELKA 245
42	BRUN	**SANTA CRUZ**	126	VIRGINIE	GOLDEN WILT
43	BRUN	SANTA FE	127	VIRGINIE	GOLTA
44	BRUN	SEMOY	128	VIRGINIE	GOSPAYA
45	BRUN	SIMMABA	129	VIRGINIE	HARRISSON SPECIAL
46	BRUN	SKRONIOWSKI CIENNY L.56	130	VIRGINIE	HERCEGOVAC
47	BRUN	SUMATRA	131	VIRGINIE	HICKS BROADLEAF
48	BRUN	TI 448-A	132	VIRGINIE	ITB32
49	BRUN	**TI 567 (nt)**	133	VIRGINIE	K 326
50	BRUN	TRUMPF	134	VIRGINIE	Kutsaga51
51	BRUN	TUA YEPOCA	135	VIRGINIE	LECHIA A
52	BRUN	UNGARISCHER RIESE	136	VIRGINIE	LECHIA LB 838
53	BRUN	**VUELTA ABAJO**	137	VIRGINIE	MAC NAIR 135
54	BRUN	WODECQ	138	VIRGINIE	NC 2326
55	BRUN	ZIMMER SPANISH=OHIO ZIMMER SPANISH	139	VIRGINIE	NC 95
56	BRUN CAPES	Cabaïguan72	140	VIRGINIE	NYIRSEGI 1
57	BRUN KENTUCKY	BLACK MAMMOTH	141	VIRGINIE	OXFORD 1
58	BRUN KENTUCKY	KENTUCKY FLORENCE	142	VIRGINIE	OXFORD 26 /56
59	BRUN KENTUCKY	KENTUCKY G	143	VIRGINIE	POLALTA
60	BRUN KENTUCKY	KENTUCKY PALU	144	VIRGINIE	**STOLAC 17**
61	BRUN KENTUCKY	KENTUCKY PULAWY 1	145	VIRGINIE	VAMOR 48
62	BRUN KENTUCKY	MADOLE	146	VIRGINIE	VESTA 30

63	BRUN KENTUCKY	WALKER'S BROADLEAF	147	VIRGINIE	VIRGINIA 115
64	BRUN KENTUCKY	WESTERN=WESTERN DARK MALAWI	148	VIRGINIE	VIRGINIA 21
65	BRUN KENTUCKY	WHITE STEM ORINOCO	149	VIRGINIE	VIRGINIA SP278
66	BRUN KENTUCKY	YELLOW MAMMOTH	150	VIRGINIE	VIRGINIE 400 FLUE CURED (56)
67	BRUN MARYLAND	**AMARELLO MAURICE 2**	151	VIRGINIE	VIRGINIE GOLD A
68	BRUN MARYLAND	FLORIDA 17	152	VIRGINIE	VIRGINIE PLOVDIV
69	BRUN MARYLAND	**MARYLAND BROAD LEAF**	153	VIRGINIE	VZ 37-1-9
70	BRUN MARYLAND	PAESANA	154	VIRGINIE	WARNE = VIRGINIE OLD BELT
71	BURLEY	Banket A1	155	VIRGINIE	WARTA
72	BURLEY	BB16	156	VIRGINIE	WIKA
73	BURLEY	**BURLEY 21**	157	VIRGINIE	WISLICA
74	BURLEY	**BURLEY 49**	158	VIRGINIE	YELLOW ORONOCO = VIRGINIE NEW BELT
75	BURLEY	**BURLEY 64**	159	VIRGINIE	YELLOW PRIOR
76	BURLEY	**BURLEY AMERICAIN**	160	VIRGINIE	YELLOW SPECIAL 390
77	BURLEY	BURLEY DC (INDIGENE)	161	VIRGINIE	ZAMOJSKA 4
78	BURLEY	BURLEY DL (INDIGENE)	162	VIRGINIE	ZAMOYSKA
79	BURLEY	**BURSAN**	163	NICOTIANA	Nicotiana debneyi
80	BURLEY	**BUS63**	164	NICOTIANA	Nicotiana glutinosa
81	BURLEY	**HARROW VELVET**	165	NICOTIANA	Nicotiana goodspeedi
82	BURLEY	**JAUNE DE BELMONT**	166	NICOTIANA	Nicotiana longiflora
83	BURLEY	**JUDY'S PRIDE BURLEY (56)**	167	NICOTIANA	Nicotiana plumbaginifolia
84	BURLEY	KENTUCKY 14	168	NICOTIANA	Nicotiana tomentosa

However, 25 *N. tabacum* cultivars showed surprisingly high *mas2'* expression levels (high expression or HE cultivars, marked in bold in Table 1). As an example, RNAseq data for the LE cv. Bahia and HE cv. Vuelta Abajo are shown in figure 4a and b respectively.

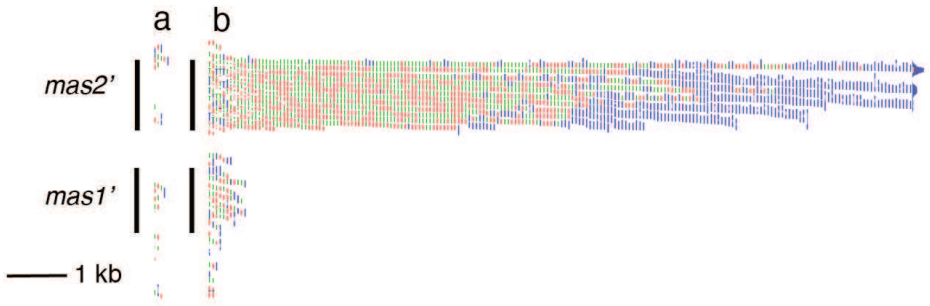


Figure 4. RNAseq analysis of the *mas* region from LE cv. Bahia (a) and HE cv. Vuelta Abajo (b)
100 nt reads mapped on the *mas* region show a very strong increase in *mas2'* expression in cv. Vuelta Abajo compared to cv. Bahia.

Since the proportions of leaf and root tissues in the tobacco extracts used for RNAseq analysis were not precisely controlled, *mas2'* expression was further investigated by RT-qPCR analysis

using different tissues of plants grown under standard conditions *in vitro*. Expression was very low to undetectable for all tissues of cv. NC2326 and detectable in different tissues of cv. Russian Burley, with highest levels in the roots (figure 5, see also below) confirming the RNAseq data.

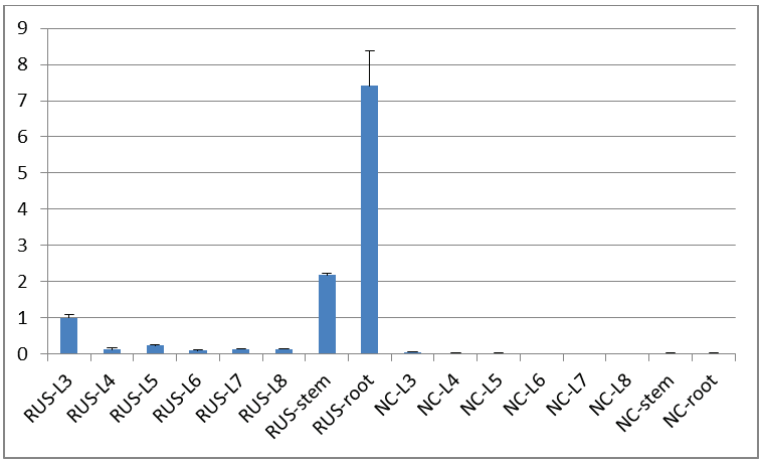


Figure 5. *mas2'* gene expression levels in different tissues of the HE cv
Russian Burley (RUS) and the LE cv. NC2326 (NC) tested by RT-qPCR. L3 to L8: successive leaves starting from oldest leaf. The *mas2'* levels were calculated relative to the RUS leaf 3 value. The highest *mas2'* expression is found in roots of cv. Russian Burley.

The high level of *mas2'* expression in roots as compared to leaves was confirmed for two other HE cultivars, Petico and Burley 49 (results not shown). High expression in roots was then confirmed by RT-qPCR for 23 out of 25 HE cultivars (marked in Table 1). The difference between LE and HE cultivars might be due to their promoters. We therefore investigated the *mas2'* promoter region.

Sequence of *mas* dual promoter region

The TB-*mas2'* promoter is part of a bidirectional (or dual) promoter sequence also found in different *A. tumefaciens* and *A. rhizogenes mas* regions (ref), situated between the two AUG codons of the divergently transcribed *mas1'* and *mas2'* genes (figure. 1a, figure. 6). The dual promoter of the *N. tomentosiformis* TB-region sequence KJ5998267 (coordinates 10224-10847, 624 nt) is identical to the published sequences from another *N. tomentosiformis* accession (ASAG01039652) and three different cultivars of *N. tabacum* (AYMY01187868, AWOK01262755 and AWOJ01062996) and resembles the *A. rhizogenes* A4 and 8196 *mas* promoters (figure 6a, b), but is very different from the *A. tumefaciens* Ach5, Bo542 and Chry5 *mas* promoters; no homology could be detected by blastn analysis even using the "somewhat similar sequences" setting.

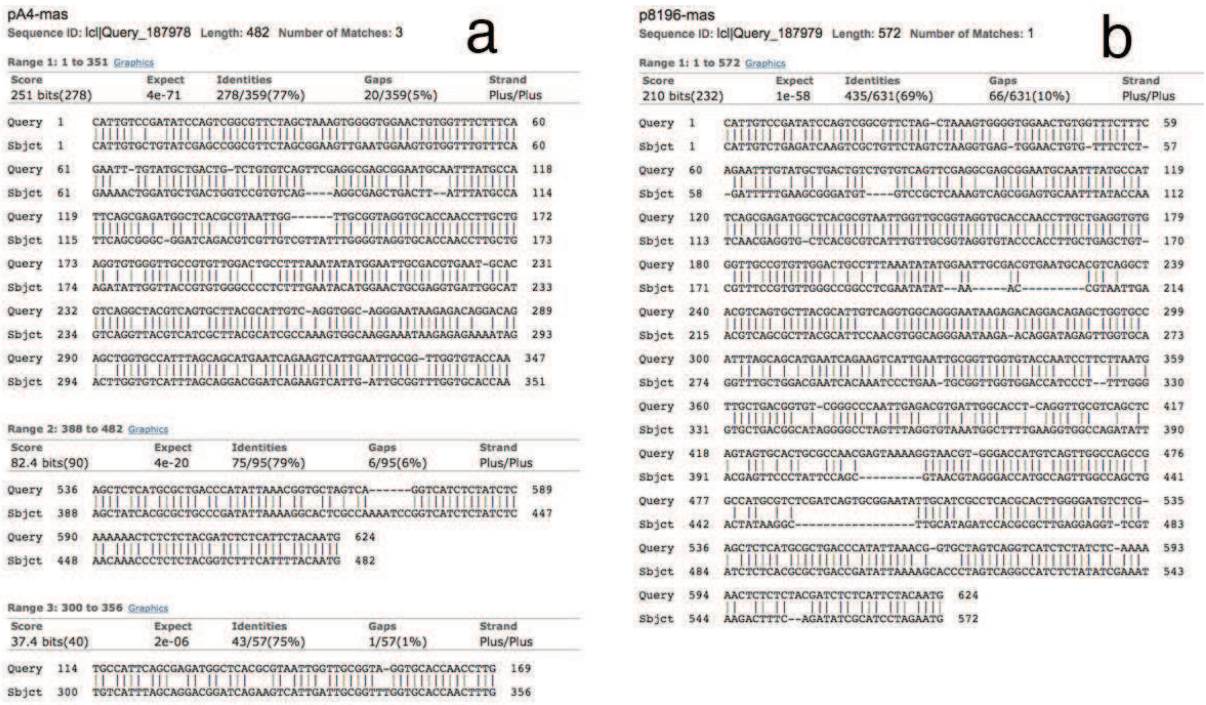


Figure 6. The *mas* dual promoter region from the *N. tomentosiformis* TB-region (top strand) was compared to similar regions (lower strand) from A4 (a) and 8196 (b)
Left end of sequence: start codon *mas2'*, right end: start codon *mas1'*. Identity values TB/ A4: 77% and TB/8196: 69%.

Sequences from *N. tabacum* LE and HE cultivars for which enough data were available showed no differences in the dual promoter region (results not shown). Thus, the differences in *mas2'* gene expression levels in LE and HE cultivars are not due to promoter differences but possibly to sequence differences outside the dual promoter or to epigenetic differences. Since epigenetic differences can lead to non-Mendelian inheritance patterns, this led us it to study the Mendelian genetics of the LE and HE phenotype.

Inheritance of LE and HE phenotypes

In order to study the inheritance of the LE/HE phenotypes, HE cultivars Petico, Burley 49 and Russian Burley were each crossed with LE cv. BB16, each combination in both directions. Levels of *mas2'* expression in roots of F1 seedlings growing *in vitro*, as measured by RT-qPCR, were intermediate between those of the parents (figure 7a).

In F2 seedlings, *mas2'* expression segregated as expected for a single Mendelian factor, results for Russian BurleyxBB16 and BB16xRussian Burley are shown in figure 7b. This result suggests that the difference between LE and HE plants is not due to an epigenetic effects on *mas2'*, but does not fully exclude it.

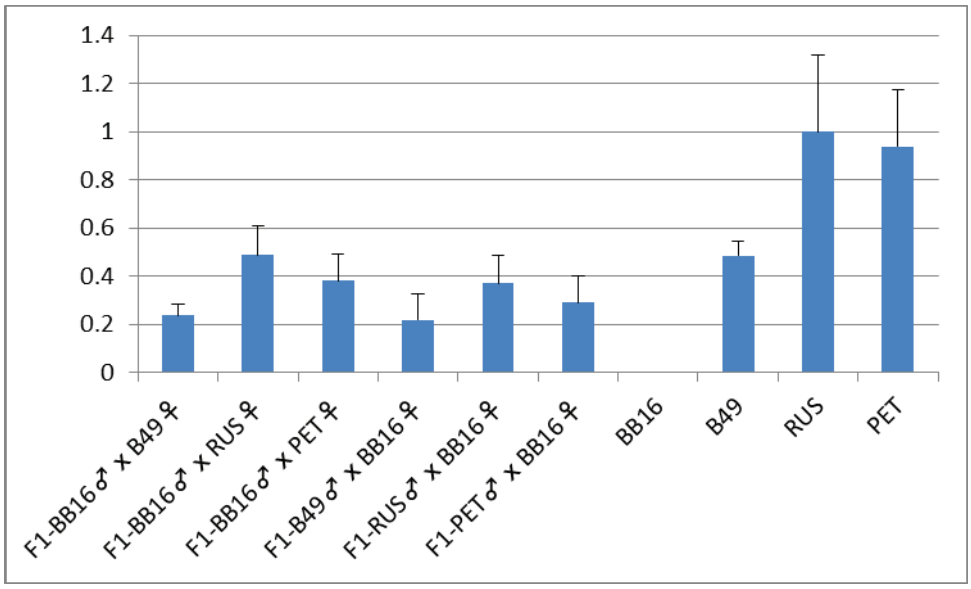


Figure 7a. *mas2'* expression in roots of F1 seedlings of different crosses between LE cv. BB16 and HE cvs
Burley 49 (B49), Russian Burley (RUS) and Petico (PET) measured by RT-qPCR. Parent levels are shown on the right. Results are expressed as fractions of the expression level in cv. Russian Burley.

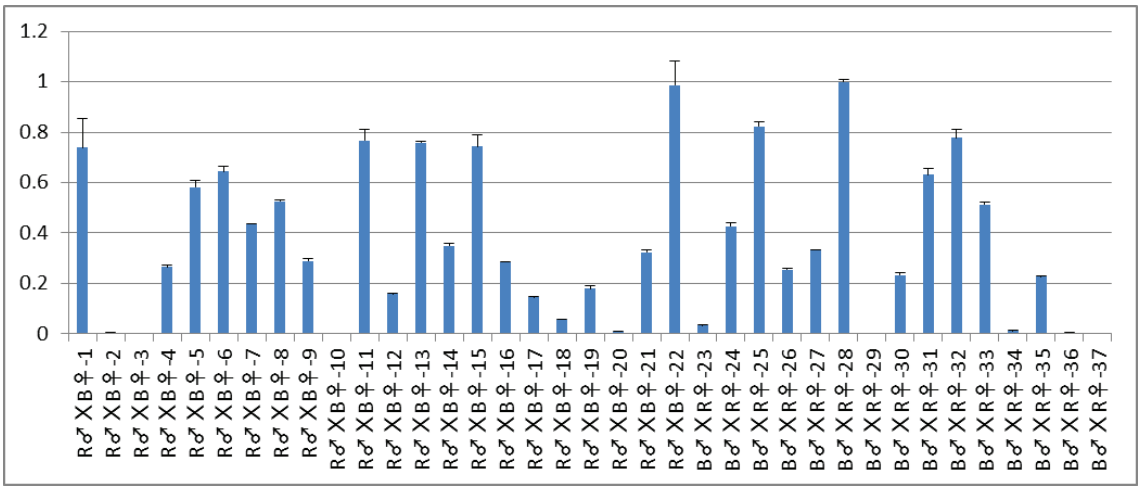


Figure 7b. *mas2'* gene expression in individual F2 seedlings resulting from the Russian BurleyxBB16 (RxB) and the reverse BxR crosses
Values are expressed as fractions of the BxR-28 value. 10 out of 37 F2 plants have very low *mas2'* expression, corresponding to the expected proportion of 1:4.

5-Azacytidine treatment of LE cultivars

Gene methylation can lead to extinction of gene expression, but methylation can be removed by 5-azacytidine treatment leading to reactivation of gene expression, this has also been demonstrated for the Ach5 *mas* region (van Slogteren et al., 1984). An attempt was made to re-activate the TB-*mas2'* gene of the LE cv. BB16 by plating seeds on 0, 1, 2 and 10 µM 5-azacytidine *in vitro*, followed by RT-qPCR analysis of *mas2'* gene expression in roots. At 2 µM 5-

azacytidine seedlings grew slower, at 10 µM growth was fully inhibited. In spite of the clear effect of the 5-azacytidine on growth, no increase in *mas2'* expression was found at any concentration (not shown), strongly indicating that the difference between LE and HE plants is not due to methylation.

Tissue-specific regulation of the TB-*mas2'* promoter

In order to visualize the TB-*mas2'* expression pattern in whole plants, the TB-*mas* dual promoter (figure 1a) was cloned upstream of the promoter-less GUS gene from pBI101 (Materials and Methods) yielding pTB-mas2'-GUS, and transferred to *N. benthamiana* by agro-transformation. The A4-*mas* promoter region linked to the GUS gene (pA4-mas2'-GUS) was used as a control. Several independent transformants were regenerated for each construct. GUS staining of F1 seedlings growing *in vitro* was variable in intensity as expected for independent transformants (figure 8), nevertheless similar with regard to tissue specificity for both constructs. Expression was mostly limited to the sub-apical root area (figure 8u*-y), but in plants with strong expression leaf veins were also stained, in several cases erratic staining along the roots indicated possible induction by wounding, this remains to be investigated. In one exceptional pA4-mas2'-GUS line the apical end of the root was stained (figure 8i, w).

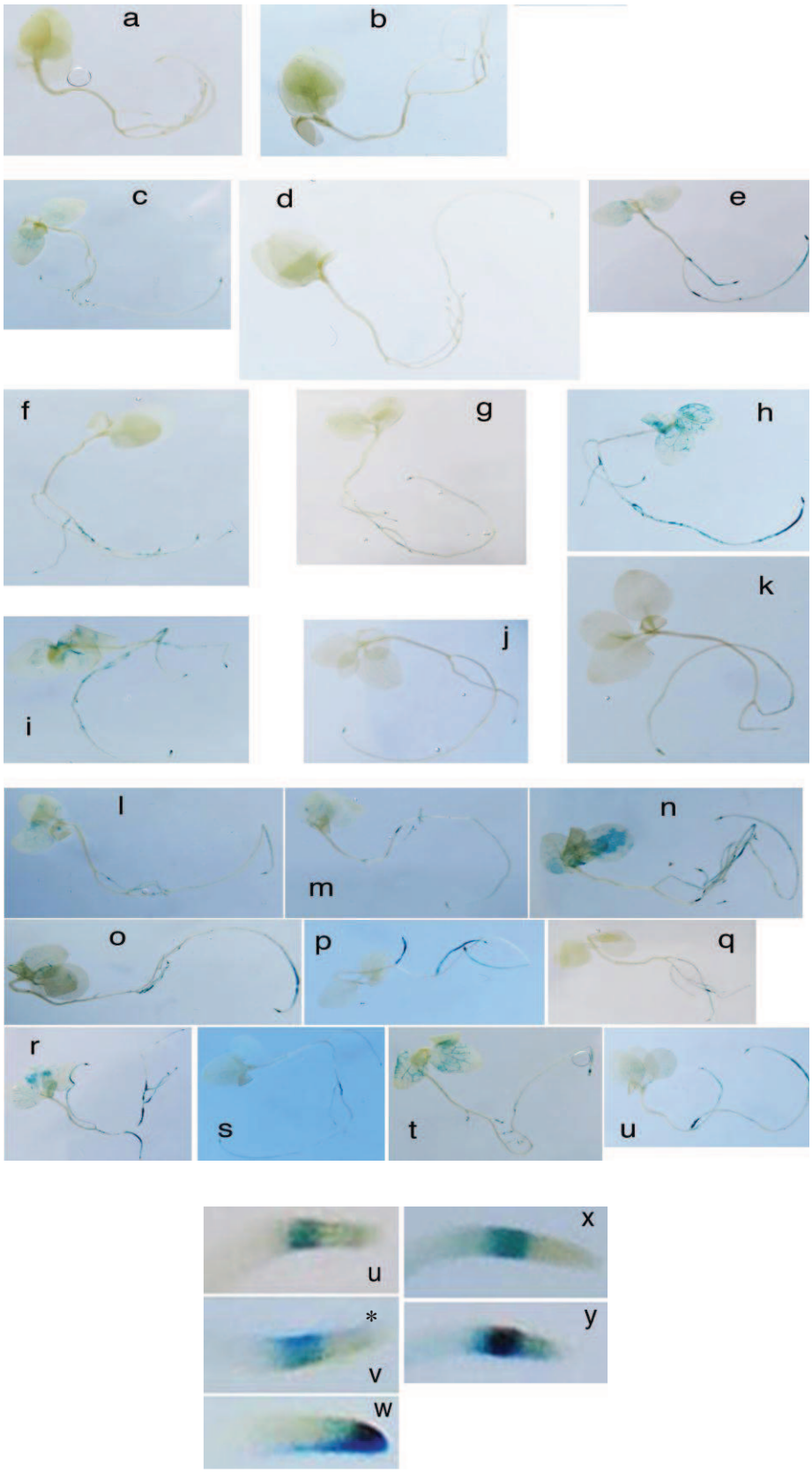


Figure 8. GUS staining of representative F1 seedlings from different transgenic *N. benthamiana* lines expressing the pA4-mas2'-GUS reporter construct (a-k, root tip details: u*-w), or the pTB-mas2'-GUS reporter construct (l-u, root tip details: x-y)
Although expression is variable, all lines show GUS staining in the subapical regions of roots and in some cases in the leaf veins.

These data confirm the root-specific expression found by RT-qPCR (see above) and indicate strongly that the TB-*mas2'* gene promoter has retained its tissue-specific regulation after horizontal transfer from an *A. rhizogenes* strain to the *N. tomentosiformis* ancestor.

DFG presence in different tobacco cultivars

Different LE and HE tobacco cultivars were tested for the presence of DFG in their roots by paper electrophoresis. Since DFG is not commercially available, it was isolated from Chry5 tumors and also prepared by chemical synthesis (Materials and Methods). In HE cultivars, but not in LE cultivars, a spot could be found migrating at the DFG position (figure 9).

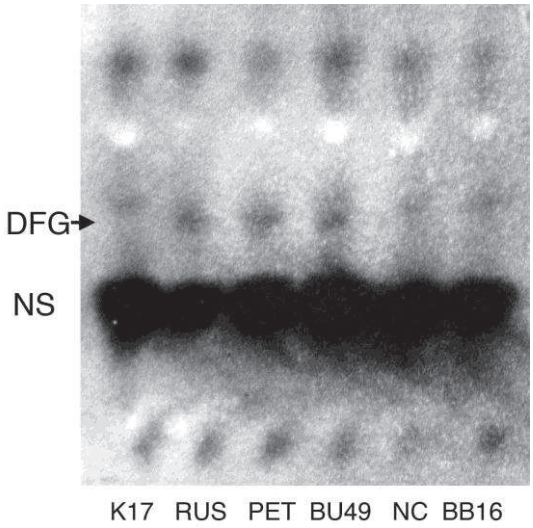


Figure 9. Opine analysis of roots of LE cultivars Kentucky 17 (K17), NC2327 (NC) and BB16 and HE cultivars Russian Burley (RUS), Petico (PET) and Burley 49 (BU49)
Extracts were separated by paper electrophoresis and stained with silver nitrate for reducing compounds. HE cultivars show a spot migrating at the position of the DFG standard. Arrow: DFG position. NS: position of neutral sugars.

Discussion

Earlier it was shown that the *mas2'* gene from the TB cT-DNA of *N. tomentosiformis* placed under control of a 2x35S promoter codes for an active enzyme and leads to a product that co-migrated with DFG in paper electrophoresis (Chen et al., 2014). The TB-*mas* region is most closely related to the *mas* regions from *A. rhizogenes* A4 and 8196, and less so to those of *A. tumefaciens* Ach5, Chry5 and Bo542. The predicted structural properties of the Mas proteins and the absence of DNA similarity for the dual promoter region indicate two different types of *mas* regions, possibly related to functional requirements for opine synthesis in and secretion from tumors and hairy roots. The TB-*mas2'* gene is little or not expressed in *N. tomentosiformis* and several *N. tabacum* cultivars, but in 25 out of 162 tobacco cultivars expression was found to be surprisingly high. Several of these high expressing (HE) cultivars are closely related, like Burley 21, Burley 49 and

Burley 64 or Zerlina, BUS63, Skroniowsky70 and Bursan, as defined by molecular markers (Julio and Dorlhac de Borne, unpublished). Since *N. tomentosiformis* is the paternal ancestor of *N. tabacum* and already contained the TB region (Chen et al., 2014) we assume that TB-*mas2'* has been activated at some point in *N. tabacum* evolution. Alternatively, *N. tabacum* inherited an active *mas2'* gene from a *N. tomentosiformis* ancestor with high activity, and later lost this activity in the ancestor(s) of the LE cultivars. Crosses between HE cultivars and an LE cultivar (in both directions) showed intermediate *mas2'* expression, suggesting that each *mas2'* allele controls its own expression level *in cis*. In the F2 generation, the HE and LE phenotypes segregated as expected for a Mendelian factor. Since the promoter sequences of LE and HE cultivars are identical, the difference is likely due to a sequence outside the promoter region, which remains to be identified. Transformation of *N. benthamiana* with pTB-mas2'-GUS and pA4-mas2'-GUS reporter genes and analysis of different independently transformed lines showed specific *mas2'* expression in the root tisp and leaf veins, suggesting that the promoter properties of the *mas2'* gene have changed little since its stable transfer into *N. tomentosiformis*. It may be interesting to transform an LE cv. and a closely related HE cv. with the pTB-mas2'-GUS construct, we predict this will lead to similar ranges of GUS expression because the results from LExHE crosses indicate that low expression is due to differences acting *in cis*. Studies on the Ach5 TR *mas2'* promoter (Teeri et al., 1989; Leung et al., 1991; Saito et al., 1991) showed highest expression in roots and leaf vascular tissues. Remarkably, although the Ach5 *mas2'* promoter sequence is very different from the A4 and TB sequences, both ensure expression in roots and vascular system. Further studies are needed on the regulation of the TB-*mas2'* and other *mas* dual promoters, these should identify the regions that ensure their tissue-specific expression and the corresponding transcription factors. The high expression of TB-*mas2'* in roots of HE cultivars leads to DFG synthesis as it does in crown gall tumors and hairy roots. This is the first demonstration of opine synthesis in a naturally transformed wild-type species. Although *N. glauca* contains an intact mikimopine synthase (*mis*) gene as shown by expression in *E. coli* (Suzuki et al., 2002) it has so far not been shown to contain mikimopine (Suzuki et al., 2001). *N. tomentosiformis* contains an intact octopine synthase-like gene (*ocl*) on its TC region, and *N. otophora* an intact vitopine synthase-like gene (*vis*) on its TE region (Chen et al., 2014). Possibly, as in the case of DFG, opines synthesized by *mis*, *ocl* and *vis* are synthesized in a restricted number of cells that remain to be identified. The very high *mas2'* expression in some tobacco cultivars is either fortuitous, or could have been selected naturally or by tobacco growers. The production of DFG merits further study, since it could have measurable consequences. One of these could be modification of root metabolism by sequestering glucose and glutamine into DFG. It will be important to establish the transport and turnover of these substances in HE cultivars and the impact on their growth under C and N-limiting conditions. Another consequence could be a modification of the root micro-flora by DFG secretion. DFG can be produced by rotting plant material (Anet and Reynolds, 1957) and is degraded by many micro-organisms (Baek et al.,

2003). Stable introduction of *Agrobacterium mas* genes into plants to create artificial symbioses (Guyon et al., 1993; reviewed in Savka et al., 2002) has shown that mannopine synthesis can lead to changes in bacterial populations in the rhizosphere. Many members of the *Rhizobiaceae* family contain genes for DFG degradation (Baek et al., 2005), and could allow them to utilize DFG released from HE tobacco cultivars. The possible effects of TB-*mas2'* expression on root and plant growth and on root micro-flora might be studied by silencing its expression or by studying pairs of very closely related LE/HE couples like Boheme Andorre (LE)/Vuelta Abajo (HE) or Paraguay Périgueux (LE)/Jaune de Belmont (HE).

Materials and Methods

Sequence analysis

DNA sequences were analyzed using the NCBI blastn program with the following settings: Database: nr/nt, optimized for "somewhat similar sequences". For comparison with *Nicotiana* whole genome sequences the "whole-genome shotgun contigs" (wgs) setting was used with the WGS projects AWOK, AWOJ and AYMY (*N. tabacum*) and ASAG (*N. tomentosiformis*). Protein analysis was done with the NCBI blastp program with the nr database. Trees were constructed by using the NCBI/blastn program, with the "align two or more sequences" setting, then using "distance tree of results" with the following settings: Tree method: Fast Minimum Evolution, Max Seq Difference: 0.85, Distance: Grishin (protein).

Reporter gene construction

The reporter genes pTB-*mas2'*-GUS and pA4-*mas2'*-GUS were constructed by PCR amplification of the mas dual promoter sequences with appropriate primers, using as targets total *N. tomentosiformis* DNA and A4 total DNA. Primers contained a *Hin*dIII site at the promoter proximal end and a *Bam*HI site at the distal end. PCR fragments were cloned into pBI101 (Jefferson et al., 1987). Positive clones were checked by sequencing, and introduced into LBA4404 (Hoekema et al., 1983) by transformation. The resulting strains were used for transformation of *N. benthamiana*.

RT-qPCR analysis

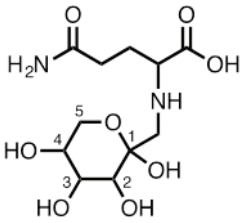
RT-qPCR analysis was done according to Chen et al., 2014.

DFG analysis by paper electrophoresis

DFG analysis was done by extraction of plant tissues with 80% ethanol and separation of neutral sugars and DFG by paper electrophoresis, reducing sugars were revealed by silver nitrate staining (Chen et al. 2014).

DFG synthesis

To a solution of D-glucose (1.35 g, 7.49 mmol, 3.5 equiv.) in AcOH (15 mL) was added L-glutamine (300 mg, 2.05 mmol, 1 equiv.). The reaction mixture was stirred at 95°C. Aliquots (0.5 mL) were collected every 30 sec over 7 min and then every 5 min until 30 min and then frozen at -80°C. Analysis by paper electrophoresis in acidic medium (H2O-AcOH-HCOOH 910-60-30) and silver nitrate staining revealed that the sample at 3 min contained enough DFG for further analysis. This sample was then purified by preparative electrophoresis. Upon elution with the same acid medium and lyophilization, a solid was obtained and analyzed. NMR spectroscopy revealed that the product corresponded to the predicted structure of DFG and that the pyranose was the major form in water solution.



^{1}H NMR (500 MHz, D_2O, 25°C): δ 4.02-3.98 (m, 2H, H_2, H_5), 3.88-3.85 (m, 1H H_4), 3.77-3.71 (m, 3H, H_3, H_5, CH), 3.32-3.25 (m, 2H, CH_2), 2.50-2.48 (m, 2H, CH_2), 2.20-2.13 (m, 2H, CH_2).

^{13}C NMR (125 MHz, D_2O, 25°C): δ 177.8 (C=O), 172.7 (C=O), 95.3 (C-1), 69.8 (C-3), 69.3 (C-4), 68.8 (C-2), 62.6 (CH), 63.8 (C-5), 52.5 (CH_2), 31.0 (CH_2), 24,8 (CH_2).

HRMS (ESI, positive mode): calcd for $C_{11}H_{21}N_2O_8$ ($M+H^+$) 309.1292 found: 309.1312.

Transformation and tissue culture

For transformation, *N. benthamiana* leaf fragments from greenhouse-grown plants were surface-sterilized. *Agrobacterium* cultures were grown overnight in liquid YEB medium, washed twice with 10 mM $MgSO_4$ and diluted to an OD600 of 0.5 in 10 mM $MgSO_4$. Leaf disks were dipped into the *Agrobacterium* suspension, and placed on solid M0237 (Duchefa) medium with 1% sucrose, 0.05 mg/L NAA, 2 mg/L BAP (MH medium), left for 24 hrs and washed in liquid MH with 350 mg/L claforan. Disks were then cultured for one month on MH with 350 mg/L claforan and 150 mg/L kanamycin. Shoots were transferred to M0327 medium without hormones with 1% sucrose, 350 mg/L claforan and 150 mg/L kanamycin for root induction. Plantlets of ten independent transformants were transferred to the greenhouse, and seeds were harvested from selfed plants. For seedling material seeds were germinated *in vitro* on M0237 with 1% sucrose.

GUS staining

GUS staining was done according to Jefferson et al. 1987.

References

Anet, E. F. L. J., and Reynolds, T. M. Chemistry of non-enzymic browning. II. Reactions between amino Acids, organic acids, and sugars in freeze-dried apricots and peaches. *Aust. J. Chem.* **10**: 182-192 (1957).

Baek, C. H., Farrand, S. K., Lee, K. E., Park, D. K., Lee, J. K., and Kim, K. S. Convergent evolution of amadori opine catabolic systems in plasmids of *Agrobacterium tumefaciens*. *J. Bacteriol.* **185**: 513-524 (2003).

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

Chen, K., Dorlhac de Borne, F., Szegedi, E., and Otten, L. Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus Nicotiana. *Plant Journal* **80**: 669-682 (2014).

Guyon, P., Petit, A., Tempe, J., and Dessaux, Y. Transformed plants producing opines specifically promote growth of opine-degrading agrobacteria. *Mol. Plant Microbe* **6**: 92-98 (1993).

Hansen, G., Larribe, M., Vaubert, D., Tempé, J., Biermann, B. J., Montoya, A. L., Chilton, M. D., and Brevet, J. *Agrobacterium rhizogenes* pRi8196 T-DNA: mapping and DNA sequence of functions involved in mannopine synthesis and hairy root differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 7763-7767 (1991).

Hong, S. B., and Farrand, S. K. Functional role of the Ti plasmid-encoded catabolic mannopine cyclase in mannityl opine catabolism by Agrobacterium spp. *J. Bacteriol.* **176**: 3576-3583 (1994).

Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J., and Schilperoort, R. A. A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature* **303**: 179-180 (1983).

Ian, M. S. Opine content of unorganised and teratomatous tobacco crown gall tissues. *Plant Sci. Lett..* **16**: 239-248 (1979).

Jefferson, R. A. Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* **5**: 387-405 (1987).

Kim, K. S., Baek, C. H., Lee, J. K., Yang, J. M., and Farrand, S. K. Intracellular accumulation of mannopine, an opine produced by crown gall tumors, transiently inhibits growth of *Agrobacterium tumefaciens*. *Mol. Plant Microbe Interact.* **14**: 793-803 (2001).

Leung, J., Fukuda, H., Wing, D., Schell, J., and Masterson, R. Functional analysis of cis-elements, auxin response and early developmental profiles of the mannopine synthase bidirectional promoter. *Mol. Gen. Genet.* **230**: 463-474 (1991).

Saito, K., Yamazaki, M., Kaneko, H., Murakoshi, I., Fukuda, Y., and Van Montagu, M. Tissue-specific and stress-enhancing expression of the TR promoter for mannopine synthase in transgenic medicinal plants. *Planta* **184**: 40-46 (1991).

Savka, M. A., Dessaux, Y., Oger, P., and Rossbach, S. Engineering bacterial competitiveness and persistence in the phytosphere. *Mol. Plant Microbe Interact.* **15**: 866-874 (2002).

Teeri, T. H., Lehväslaiho, H., Franck, M., Uotila, J., Heino, P., Palva, E. T., Van Montagu, M., and Herrera-Estrella, L. Gene fusions to lacZ reveal new expression patterns of chimeric genes in transgenic plants. *EMBO J.* **8**: 343-350 (1989).

Van Slogteren, G. M., Hooykaas, P. J. J., Schilperoort, R. A. Silent T-DNA genes in plant lines transformed by *Agrobacterium tumefaciens* are activated by grafting and by 5-azacytidine treatment. *Plant Mol. Biol.* **3**: 333-336 (1984).

Vladimirov, I. A., Matveeva, T. V., Lutova, L. A. Opine biosynthesis and catabolism genes of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *Genetika* **51**: 137-146 (2015).

Discussion and perspectives

During the last few years, we fully sequenced the TA, TB, TC and TD cT-DNAs from *N. tomentosiformis* (Chen et al., 2014). On the TE region of *N. otophora* (Sierro et al., 2014) we identified a TE-*6b plast* gene. This gene is biologically active as shown by expression in *N. tabacum*. A detailed anatomical study was carried out on the *6b* model gene from *A. vitis* Tm4 (Chen and Otten, 2015), in order to serve as a basis for the future analysis of the TE-*6b* gene effect. The *mas2'* gene of the TB region was shown to be highly expressed in roots of some *N. tabacum* cultivars and lead to the synthesis of desoxyfructose-glutamine (Publication 3, to be submitted). In the following part, I discuss these results and propose some possibilities for future studies.

I. Nicotiana genus cT-DNAs

I.1 Assembling the TC and TE regions from *N. otophora*

Genomic contigs covering the *N. otophora* genome have been published (Sierro et al., 2014). The contigs from the TE region were identified on the basis of homology to the T-DNAs from *A. tumefaciens, A. rhizogenes and A. vitis* (Chen et al., 2014, figure 6A-C). Putative TE insertion site sequences were aligned with *N. tomentosiformis* and *N. sylvestris* contigs (Chen et al., 2014, figure 6D-E). However, since the TE region clearly contains repeats, complete assembly will require careful, manual analysis of reads as well as PCR experiments followed by sequencing, as we have done for TA, TB, TC and TD.

The assembly of the TC region from *N. otophora* might be easier since the TC from *N. tomentosiformis* has already been fully assembled (Chen et al., 2014). Nevertheless, the attribution of polymorphisms in the TC repeats still needs to be done and as for TE, will require sequencing of repeat-specific PCR products.

I.2 Nicotiana genus cT-DNA structure

Within the Nicotiana genus, only cT-DNAs from *N. glauca* (Aoki, 2004) and *N. tomentosiformis* (Chen et al., 2014) have been fully assembled. We tested whether TA, TB, TC and TD from *N. tomentosiformis* were present in *N. tomentosa*, *N. otophora* and *N. kawakamii* (Chen et al., 2014, table 3), but the complete cT-DNAs from these species have not yet been sequenced. The deep-sequencing data from *N. benthamiana* and *N. otophora* were published last year (Sohn et al., 2014; Sierro et al., 2014). No cT-DNA was found in *N. benthamiana* (Blast data not shown). *N. otophora* contains TC and TE (Chen et al., 2014). TA, TB and TD are absent from *N. setchellii*, the situation for TC remains unclear because the expected TC insertion site could not be amplified (Chapter I, II.2). In order to show whether TC is absent or not, southern analysis with

TC probes can be used. Apparently cT-DNAs from *N. tabacum* are variable among different cultivars since some of them contain a full TA region whereas others show deletions. It still remains to be investigated how this TA part was lost during evolution. It would be interesting to obtain the assembled cT-DNAs of the other Nicotiana species since a better knowledge of the divergence between the cT-DNA repeats would improve our understanding of Nicotiana species evolution. At the same time, it will be important to obtain more data on the overall evolution of the Nicotiana species. Such data not only concern normal sequence divergence, but will probably also reveal HGT by interspecific hybridization which is common in Nicotiana.

Apart for the cT-DNAs of the Nicotiana genus, cT-DNAs were also found in *L. vulgaris* and *I. batatas*. We expect that additional cT-DNAs will be discovered in other genera as plant genomic sequences continue to accumulate. New natural transformants may reveal new types of T-DNAs like the TE region we found in *N. otophora*. Large scale Agrobacterium sequencing involving hundreds of natural isolates may allow us to find the strains that caused the original transformations leading to the natural transformants. If such strains can be found, they could provide the sequences needed to “repair” the mutated cT-DNA genes, and allow a comparison between the properties of the original T-DNA genes and their present-day plant-borne copies.

I.3 Functional analysis of intact ORFs from cT-DNAs

Several intact ORFs in the cT-DNAs are expressed in *N. tomentosiformis* (Chapter I, figure15) as well as *I. batatas* (Kyndta et al., 2015) and might code for particular functions. In order to learn more about the function of these different intact ORFs, it would be important to knock out their expression, e.g. by RNAi, or to remove them completely by CRISPR, and to study the effects of cT-DNA removal on plant growth or opine synthesis. Reporter gene studies with promoters from active cT-DNA genes and comparison with the promoter properties from the corresponding genes from Agrobacterium would allow us to establish whether the regulation of these genes underwent significant changes after the regeneration of the natural transformants. Cloning these genes under a strong or inducible promoter and expressing them in a plant species without cT-DNA like *N. benthamiana* will be a further essential approach to study their characteristics.

I.4 Functional analysis of cT-DNA insertion sites in plants

One question that we did not address in our cT-DNA study was whether the insertions of cT-DNAs affected the function of original genes located at the insertion sites. The first step would be to investigate the expression of the intact regions in those species that are closely related to plants with a particular cT-DNA but lack this cT-DNA, for example by comparing *N. otophora* which lacks TA with *N. tomentosiformis*. The possible functions of these genes might then be

studied by removal with CRISPR. However, it is quite possible that these functions are redundant or that the effects are too slight to be noticed. One might also try to repair the cT-DNA interrupted regions by CRISPR technology, by removing the cT-DNA and adding back the plant sequences that were deleted during the insertion process.

II. Further study of the *6b* genes

II.1 Mechanism of *6b* gene activity

6b gene is a member of the *plast* gene family whose basic function is still unknown. Although mechanisms have been proposed for several *plast* genes, i.e. gene *5* (Körber et al., 1991), *rolB* (Cardarelli et al., 1987; Spena et al., 1987; Capone et al., 1989; Maurel et al., 1990; Capone et al., 1994), *rolC* (Meyer et al., 1995; Mohajjel-Shoja, 2010) and *orf13* (Fründt et al., 1998; Lemcke and Schmülling, 1998; Stieger et al., 2004), we believe these results need to be confirmed before they can be considered as conclusive. Within the *plast* family, *6b* gene causes the strongest plant modifications, and is therefore a good choice to approach this basic function. The different *6b* variants known sofar have different properties as was shown by tumor induction with six *6b* genes placed in the same expression cassette (Helfer et al., 2003). It would be interesting to obtain plants with inducible versions of each of these variants and to observe the morphogenetic effects at different expression levels. The properties of the other *plast* genes also merit further analysis as they could tell us more about the biological role of these genes. It will be especially important to study their expressions in their natural context, i.e. in tumor or hairy root tissues.

Our anatomical studies on inducible dex-T-*6b* tobacco plants showed that *6b* can induce ectopic photosynthetic cells leading to leaf primordia and ectopic vascular strands. It remains to be investigated where and how these cells start their development. In the case of the ectopic photosynthetic cells, they appear on the abaxial side of young leaves and could result from additional rounds of division at the level of the spongy mesophyll. The first stages of this abnormal development should be studied at the level of the cell cycle, using different cell cycle markers. The extra vascular strands were easily recognized in the petioles and midveins. Their initiation and subsequent development needs more investigation. The remarkable structures observed in T-*6b* tobacco plants may be interest for developmental biology. The study of easily accessible ectopic leaf primodia at large trichome positions may reveal the mechanisms that govern their development and the subsequent initiation of an associated vascular system. It may be possible to induce these ectopic primodia in leaf explants in vitro thereby enhacing the possibilities for experimental approaches.

The C terminal side of T-*6b* amino acid sequence contains a highly acidic region (EEGEDDDNEIGDEGEAGGAE), similar regions were found in other 6b proteins (Helfer et al., 2003). In *R. leguminosarum*, a hypothetical plast protein (locus WP_041936630) of 238 amino acids contains 22 glutamic acids in the beginning of its C terminal region. This acidic region may be of special interest and could be mutated in various ways to determine the importance of the number of acidic residues, their nature, and their position in the protein. The acidic region could interact with basic partners but it is also possible that the acidic amino acids play a role in opine synthesis, if either synthesis of the 6b protein or its degradation could affect the pool of these amino acids.

Basically four mechanisms were proposed for *6b* activity (Chen and Otten, 2015, Introduction). Our group proposed sucrose uptake since large amounts of sucrose, glucose and fructose accumulate in *6b* plants and isolated root fragments were shown to take up sucrose upon *6b* induction (Clément et al., 2007). It is well known that sucrose is able to affect cell cycle and cell expansion (Reviewed in Wang and Ruan, 2013). However, the precise link between *6b* and sucrose uptake is still lacking. Since sucrose is essential for plant growth, it is difficult to obtain sucrose transporter mutants. Nevertheless, it might be possible to engineer plants with an inducible loss of sucrose uptake in order to study the capacity of the 6b protein to complement such transporters. The expression of the *6b* gene in other biological systems like mosses, algae, yeast, bacteria or xenopus oocytes linked with sucrose transport studies might be a promising way to study this gene. It could receal what kind of orgamisms are sensitive to its activity and provide simple experimental systems.

Further studies on the *6b* mechanisms proposed by other groups will be required, e.g. it still has to be confirmed that the 6b protein has ADP-ribosylation activity on AGO1 and SERRATE (SE) proteins in the presence of the ADP-ribosylation factor (Wang et al., 2011). The latter *6b* studies are among the most advanced and should be extended to TE-*6b*, other *6b* genes like AB-*6b*, C-*6b*, S-*6b*, A-*6b*, Bo542-*6b*, AKE10-*6b* and to other *plast* genes.

Recently, plast protein-encoding genes were found in ectomycorrhizal fungi like *Laccaria bicolor* and *Pisolithus microcarpus*, and in nodulating Rhizobium strains like *R. leguminosarum*, *R. mesoamericanum*, *Bradyrhizobium* and *Mesorhizobium plurifarium*. This finding is very striking for two reasons. First of all, the sporadic appearance of *plast* gene in there organisms suggests that their presence results from HGT. Secondly, it indicates that *plast* genes can play roles in two very different types of plant-associated organisms. Studies with knock-out mutants would therefore be very important. In addition, the regulation of these genes merits further study.

II.2 Comparison of TE-*6b* with T-*6b*

2x35S TE-*6b* single copy homozygous tobacco plants derived from different R0 plants have recently been obtained. These plants will be studied in detail and compared to dex-T-*6b* plants (Chen and Otten, 2015). Special emphasis should be given to the anatomical details of leaves and roots. A dex-inducible TE-*6b* version will also be important to test since this allows the induction of TE-*6b* in a normally growing plant and could avoid problems of regeneration of plants with too high TE-*6b* expression levels. An inducible dex-TE-*6b* plant should also be useful to study sucrose uptake and accumulation.

Removal of TE-*6b* from *N. otophora* by CRISPR could show whether the TE-*6b* gene has an effect on the growth of *N. otophora*. The introduction of the gene in the *N. otophora* ancestor could have caused sufficiently strong growth changes to lead to speciation. However, the TE-*6b* gene is maybe not the only TE gene with growth-modifying properties. Other intact genes (*vis*, *rolC*, *orf13* and *orf14*) could also be candidates. To answer this question, it will be necessary to first remove TE by CRISPR and to restore the original insertion region.

II.3 Can the *6b* gene be used as a genetic tool to enhance plant regeneration?

In plant genetic engineering, a common technological problem is poor regeneration efficiency during transgenic plant production. Several genes like *WUS* (Laux et al., 1996; Zheng et al., 2014) and *BBM* (Boutilier et al., 2002; Florez et al., 2015) can be used to increase the regeneration properties of transgenic tissues especially by using inducible promoters. The *6b* gene is able to induce calli (shown by infection of *N. rustica* stems) and more importantly, leaf primordial (Chen and Otten, 2015). Based on these two activities, it seems reasonable to propose the *6b* gene as a tool to transiently increase the regeneration efficiency of transgenic plants.

III. Further studies on the *mas2'* gene

According to *mas2'* gene expression levels, *N. tabacum* cultivars can be separated into two groups (HE and LE) (Publication 3, to be submitted). In order to know more about the function of the *mas2'* gene, it is necessary to knock out this gene in an HE plant and to compare the growth of the mas2'-less plant with that of the original cultivar. It has been shown that the secretion of mannopine, mannopinic and agropinic acid by *L. corniculatus* plants transformed with *A. rhizogenes* strain 8196 can lead to the selection of mannopine-utilizing microorganisms in the rhizosphere (Guyon et al., 1992). It was also found in the rhizopheres of transgenic *A. thaliana* which is able to make octopine, the density of octopine-using bacteria increased (Mondy et al., 2014). It is possible that the high expression of the *mas2'* gene in some *N. tabacum* cultivars

leads to a similar selection pressure. It would therefore be interesting to study this possibility, especially if a *mas2'* knock out version of these HE cultivars will become available that could serve as an isogenic negative control.

In order to study the bacterial use of DFG in HE tobacco cultivars, it will be useful to test Agrobacterium strains with DFG utilizing capacity and their corresponding mutants. The further study of the *N. benthamiana* lines expressing the *mas2'* promoter-gus reporter gene could reveal conditions that lead to even higher *mas2'* expression and DFG synthesis. Furthermore, an inducible system to increase DFG synthesis in an LE line or to decrease DFG synthesis in an HE line would be useful to study the metabolic consequences of DFG synthesis in plants, and especially in roots.

The synthesis of DFG by HE cultivars remains to be placed in the context of amino acid and sugar metabolism and transport in roots. DFG may be produced locally in regions of high *mas2'* expression (the subapical root region) and from there exported or secreted. It may also be converted to other componds. Experiments with labeled precursors will be necessary to study such metabolic effects. Sofar, we do not know which precursors are used by the Mas2' enzyme, and it will be necessary to study its enzymatic properties with the purified enzyme.

Materials and methods

I. Materials

I.1 Plant materials

I.1.1 *N. tabacum*

Cultivar Samsun nn and cultivar Havana 425 were kindly provided by Dr. Fred Meins. The following cultivars were obtained from the Institut du Tabac de Bergerac (ITB).

Type	Variety	ITB N°	Type	Variety	ITB N°
BRUN	DRAGON VERT	652	BRUN	ARAPIRA	261
BRUN	PARAGUAY PERIGUEUX 48	765	BRUN	COULO	291
BRUN	PETICO		BRUN	GALPAO	673
BRUN KENTUCKY	KENTUCKY G	188	BRUN	MAZINGA	732
BURLEY	BURLEY 49	175	BRUN	SANTA CRUZ	817
BURLEY	KENTUCKY 14	211	BRUN	SANTA FE	818
BURLEY	RUSSIAN BURLEY	226	BRUN	VUELTA ABAJO	849
EXPERIMENTAL	KOKUBU	963	BRUN MARYLAND	AMARELLO MAURICE 2	325
ORIENT	BASMA DRAMA 2	339	BRUN MARYLAND	MARYLAND BROAD LEAF	729
ORIENT	KABA KOULAK	908	BURLEY	BURLEY 21	167
ORIENT SEMI	BOHEME ANDORRE	359	BURLEY	BURLEY 64	163
ORIENT SEMI	ROUM	929	BURLEY	BURLEY AMERICAIN	176
VIRGINIE	CASH	19	BURLEY	BURSAN	200
VIRGINIE	NC 2326	77	BURLEY	HARROW VELVET	205
	BB16	450	BURLEY	JAUNE DE BELMONT	207
BURLEY	ZERLINA	241	BURLEY	JUDY'S PRIDE BURLEY (56)	208
ORIENT	HERCEGOVINE	902	BURLEY	PULAWSKY 66	223
VIRGINIE	POLALTA	447	BURLEY	SKRONIOWSKI 70	228
VIRGINIE	STOLAC 17	101	BURLEY	START	100
BURLEY	BUS63	158	BURLEY	WHITE BURLEY	238
EXPERIMENTAL	TI657	978			

I.1.2 *N. otophora*

N. otophora seeds are from the US Department of Agricultural Research Service, National Plant Germplasm System (inventory number PI235553) with accession name TW95.

I.1.3 *N. setchellii*

Seeds (inventory number PI555557 with the accession name TW121) were from the US Department of Agricultural Research Service, National Plant Germplasm System.

I.1.4 *N. benthamiana*

N. benthamiana plants were provided by the IBMP green house.

I.1.5 Other Nicotiana species

Seeds from the Institut du Tabac de Bergerac	
Species	Accession name
Nicotiana tomentosiformis	ITB646
Nicotiana sylvestris	ITB626
Nicotiana kawakamii	ITB642
Nicotiana tomentosa	ITB1015

I.2 Bacteria

I.2.1 *Escherichia coli* Top 10 strain

E. coli Top 10 strain was used for plasmid amplification. Its genotype is: *F-, mcrA, Δ (mrr-hsdRMS-mcrBC), Φ80lacZΔM15, ΔlacX74, recA1, araD139, Δ (ara, leu) 7697, galU, galK, rpsL (StrR), endA1, nupG.*

I.2.2 *Agrobacterium tumefaciens* strain LBA4404

Strain LBA4404 was used for *N. tabacum* transformation and *N. benthamiana* infiltration. This strain contains a disarmed Ti plasmid pAL4404 (Hoekema et al., 1983) which carries the virulence genes, but no T-DNA. It is rifampicin and streptomycine resistance. (Hoekema et al., 1983).

I.3 Vectors

I.3.1 Cloning vector: pCK GFP S65C

The pCK GFP S65C vector contains an ampicillin resistance gene and a *gfp* gene under the 2x35S promoter (figure 24A). Genes of interest were amplified by PCR and used to replace the *gfp* gene of this plasmid. Colonies with the right construct were identified by miniprep DNA analysis, and sequencing. Subsequently, a *Hin*dIII fragment with the target gene was inserted into vector pBI121.2. (Carrington et al., 1991).

I.3.2 Transformation vector: pBI121.2 binary vector

This plasmid contains T-DNA left and right borders (LB and RB) (figure 24B). The T-DNA contains a *nptII* gene, a 2x35S-*gus* gene and a multiple cloning site. A bacterial *nptII* gene outside the T-DNA allows for selection of the plasmid on kanamycin, both in *E. coli* and *A. tumefaciens*. The plasmid can be placed in different Agrobacterium backgrounds that provide the virulence functions. This vector was used for the TB-*mas2'* and TE-*6b* gene clone. (Jefferson et al., 1987).

I.3.3 Reporter gene vector: pBI101 binary vector

Vector pBI101 is similar to pBI121.2 except for the part on the right of the multiple cloning site. Instead of *tNOS*, *gus* and *p35S* from pBI121.2, pBI101 contains only *gus* and *tNOS* (figure 24C). This allows the insertion of promoter sequences in order to test their properties. This vector was used to test the promoter characteristics of TB-*mas2'* and A4-*mas2'*. (Jefferson et al., 1987).

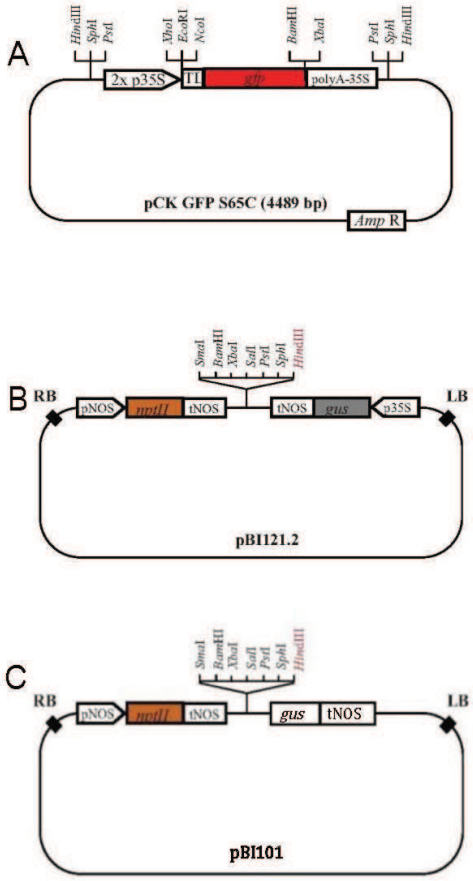


Figure 24 (A) pCK GFP S65C vector (B) pBI121.2 vector (C) pBI101 vector
Multiple cloning sites are shown above the vectors. (A) p35S: CaMV-35S promoter; TL: translational enhancer; *gfp*: green fluorescent protein coding gene; polyA-35S: polyadenylation signal of CaMV 35S gene. (B) pNOS:

promoter of nopaline synthase gene; *nptII*: neomycin phosphotransferase II gene; tNOS: NOS terminator; *gus*: β-glucuronidase gene. (C) Abbreviations are the same as in B. (Adapted from Mohajjel Shoja, 2010)

II. Methods

II.1 Plant techniques

II.1.1 Leaf patch infection

N. benthamiana plants were used for leaf patch infection. *A. tumefaciens* strain LBA4404 carrying the gene of interest was first grown on solid YEB medium with rifampicin (100 mg/L) and kanamycin (25 mg/L) for 2 days at 28°C, then one colony was used to start a liquid culture in 5 mL YEB with the same antibiotics at 28°C. After 48 h, the bacterial culture was centrifuged at 6000 rpm for 5 min at RT. The supernatant was removed and the pellet was washed with $MgSO_4$ (10 mM). The bacterial pellet was suspended in $MgSO_4$ (10 mM) and the bacterial concentration was adjusted to an optical density of 0.5 at 600 nm. *N. benthamiana* leaves were infiltrated with this bacterial suspension using a 1 mL needleless syringe.

II.1.2 Tobacco transformation

II.1.2.1 Preparation of Agrobacteria suspension

The Agrobacteria were cultured in 100 mL YEB medium for one day and centrifuged 10 min at 6000 rpm. The pellet was twice washed by 10 mM $MgSO_4$ and resuspended in the same solution at OD_{600}=1.0.

II.1.2.2 Tobacco leaf transformation

Fresh leaves from green-house grown plants (*N. tabacum* cv. Samsun nn or *N. benthamiana*) were used. 70% ethanol was used to wash the leaves for 10 seconds. Then they were immerged into 200 mL 5% sodium hypochlorite with 500 µL of a 10% Tween 20 solution and shaken gently for 20 min. Sterile water was used to wash the leaves 3 times for 5 min. Finally, the leaves were placed on sterile paper and disks with 2 cm diameter were removed and shortly suspended in the Agrobacteria suspension. These disks were then put on M0222 (Duchefa) medium with 1% sugar and 0.8% agar for 24 h in the culture room. The next day they were washed with liquid M0222 medium with 350 mg/L cefotaxime and transferred on M0222 medium with 350 mg/L cefotaxime, 150 mg/L kanamycin and hormones (0.05 mg/L NAA, 2 mg/L BA).

About 1 month later, shoots had grown on the edge of the leaf disks. These shoots were then transferred on M0222 with corresponding antibiotics without hormones, in order to induce rooting. After 3 weeks, rooted plantlets were transferred on soil.

II.1.3 Nucleic acid analysis

II.1.3.1 Plant DNA extraction

Plant DNAs were prepared from leaves of sterile plantlets or greenhouse plants according to Dellaporta (1983). DNAs were dissolved in water and stored at -20°C or 4°C. DNAs from the following plants were extracted: *N. tomentosiformis* (ITB646), *N. sylvestris* (ITB626), *N. kawakamii* (ITB642), *N. tomentosa* (ITB1015), *N. otophora* (TW95), *N. setchellii* (TW121), and *N. tabacum* cv. Samsun nn.

II.1.3.2 PCR

PCR amplification was done by using the phusion high-fidelity PCR master mix (F-531S) from Thermo. The components of the PCR reaction were as follows:

PCR reaction mixture		
Component	10 µL reaction	Final concentration
H_2O	Add until 10 µL	
2xPhusion Master Mix	5 µL	1x
Primer A	0.6 µL	0.5 µM
Primer B	0.6 µL	0.5 µM
Template DNA	1 µL	1-50 ng

In general, PCR was performed with the following cycles:

Cycle step	Temp.	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	5-10 s	
Annealing	55-65°C	10-30 s	35
Extension	72°C	15-30 s /1 kb	
Final extension	72°C	5-10 min	1
Hold	4°C		1

II.1.3.3 PCR clean-up

PCR products were run on a 0.8% or 1.0% agarose gel. The DNA was stained with ethidium bromide and bands were excised. The clean-up of the products was done according to the user manual of NucleoSpin gel and PCR clean-up from Macherey-Nagel.

II.1.3.4 RNA extraction

300 mg leaf or root tissue was put into a 1.5 mL eppendorf tube with 200 µL TRIzol on ice. The tissue was ground using an adapted drill. Then 50 µL chloroform was added into the tube and mixed well. After centrifuging at 11000 rpm for 5 min, the hyper phase was transferred into a new tube and 140 µL isopropanol was added to precipitate the RNA. The sample was centrifuged again at 11000 rpm for 5 min. The pellet was washed once with 1 mL 80% ethanol and dissolved in 50 µL H_2O.

II.1.3.5 cDNA synthesis

8.5 µL RNA extract, 1 µL 10xDNase buffer and 1.5 µL DNase were added and incubated at 37°C for 60 min to get rid of DNA. 1 µL EDTA was added and incubated at 65°C for 10 min to stop the reaction. 2 µL treated sample, 1 µL of 50 µM oligo dT, 1 µL of 10 mM dNTP and 9 µL of H_2O were added together and heated at 65°C for 5 min. Then the mixture was put on ice for 2 min. cDNA was synthesized by adding 4 µL 5xFS buffer, 1 µL 0.1 M DTT, 1 µL of 40 U/µL RNase out, 1 µL of 200 U/µL Superscript III reverse transcriptase and incubated at 50°C for 60 min. Finally the sample was put into a heating block at 70°C for 15 min to stop the reaction. The sample is then ready to use for PCR amplification.

II.1.3.6 RT-quantitative PCR

This method was used to quantify mRNA levels in plants. 4.5 µL SYBER Green I fluorescein reporter (Roche), 1.8 µL H_2O, 2x0.9 µL primers (0.75 µM) and 1 µL cDNA were added in one well of a 284 well plate. The transcripts of the *EF2* gene from *N. tabacum* (GenBank: AJ299248) were used as internal standard.

II.1.4 Protein analysis

II.1.4.1 Protein extraction from plants

Fresh leaf disks (1.2 cm in diameter) were taken and ground in 100 µL lysis buffer. The sample was centrifuged at 11000 rpm for 3 min and 8 µL supernatant was transferred into a new tube. 8 µL 2xdenaturing buffer was added and the sample was heated at 95°C for 5 min. The sample was put on ice for 1 min and centrifuged at 11000 rpm for 5 min. The supernatant was ready for further analysis.

Protein denaturing buffer (pH 6.8)	
Tris-HCl	0.125 M
β-mercaptoethanol	10%
SDS	4%
bromophenol blue	0.004%
glycerol	20%

II.1.4.2 Western blot

Resolution and stacking gels were prepared as follows:

Resolution gel buffer		Stacking gel buffer	
H_2O	5.28 mL	H_2O	6.7 mL
acrylamide	6.4 mL	acrylamide	1.72 mL
Tris 1.5 M, pH 8.8	4 mL	Tris 1.0 M, pH 6.8	1.3 mL
SDS 10%	0.16 mL	SDS 10%	0.1 mL
APS 10%	0.16 mL	APS 10%	0.1 mL
TEMED	6 µL	TEMED	10 µL

The protein extracts were run on the gel in migration buffer at 100 V for 1 h. The protein in the gel was then transferred onto a nitrocellulose membrane at 500 mA, 80 V for 2 h on ice. The membrane was washed in TBS buffer with 5% degreased milk powder for 5 min and blocked in the same buffer for 1 h at RT. Subsequently it was incubated with primary antibody at 4°C overnight and washed 3 times for 5 min in TBS buffer. A secondary incubation was done by incubating the membrane with secondary antibody for 1 h at RT. Finally the membrane was washed 3 times in TBS and treated with Lumi-Light^{PLUS} Western Blotting kit (Roche) and exposed using an autoradiography film.

Migration buffer (pH 8.3)		Transfer buffer		TBS buffer (pH 7.5)	
Tris	25 mM	Tris	15.12 g	Tris	50 mM
glycine	250 mM	glycine	72.05 g	NaCl	150 mM
SDS	0.1%	ethanol	750 mL	Triton X-100	0.1%
		H_2O	Until 5 L		

II.1.5 Opine analysis

Plant material was extracted and opines were detected by paper electrophoresis as described in Chen et al., 2014.

II.2 Bacterial techniques

II.2.1 Bacterial competent cell preparation

An *E. coli* single colony was inoculated in 5 mL LB medium and incubated overnight at 37°C with moderate agitation (about 250 rpm). 500 mL LB medium was inoculated with 1 mL of this pre-culture and incubated overnight at 37°C with shaking until OD_{595nm}=0.375. The following steps should be done on ice or at 4°C. The medium with *E. coli* cells was transferred into tubes and centrifuged for 10 min at 5000 rpm, the pellet was washed in 500 mL H_2O. This washing step was repeated once. Cells were left on ice for 30 min. The medium was centrifuged again for 10 min at 5000 rpm and the pellet was resuspended in 20 mL of 15% glycerol. This suspension was aliquoted in 100 µL tubes with 50 µL final volume and stored at -80°C.

II.2.2 Plasmid DNA extraction

E. coli strains were harvested after 8 h culture at 37°C. The cells were transferred into 1.5 mL tubes and centrifuged at 11000 rpm for 5 min. The supernatant was removed from the tubes and 100 µL solution I was added for resuspending the pellets by vortexing. After 2 min, 200 µL solution II was added and mixed well. Then, 150 µL solution III was added and the tubes were centrifuged for 5 min at 11000 rpm. 400 µL supernatant was transferred into new tubes and 400 µL phenol/chloroform (v/v=1:1) was added and mixed. The tubes were centrifuged at 11000 rpm for 5 min and 300 µL top phase was transferred into new tubes. 600 µL ethanol was added into tubes at RT, mixed well and spun at 11000 rpm for 5 min. The pellet was resuspended into 50 µL H_2O. The plasmid DNA solution could be used for restriction enzyme test, PCR or sequencing.

Plasmid DNA extraction buffers		
Solution I	Solution II	Solution III
25 mM Tris-HCl pH 8	200 mM NaOH	58.8 g/200 ml potassium acetate
10 mM EDTA	1% SDS	23 ml acetic acid
50 mM glucose		

II.2.3 Transformation of *E. coli* bacteria

1 µL ligation solution was added to 50 µL competent *E. coli* cells and put on ice for 15 min. The mixture was treated by heating at 42°C for 90 s or by electroporation. The mixture was transferred into 1.5 mL LB medium and shaken for 1 h at 37°C. The pre-cultured medium was centrifuged at 8000 rpm for 3 min and plated on antibiotic plates. After incubating at 37°C overnight, single-colonies were used for further analysis.

References

Altabella, T., Angel, E., Biondi, S., Palazón, J., Bagni, N., and Piñol, M. T. Effect of the *rol* genes from *Agrobacterium rhizogenes* on polyamine metabolism in tobacco roots. *Physiol. Plantarum* **95**: 479-485 (1995).

Altamura, M. M., Archilletti, T., Capone, I., and Costantino, P. Histological analysis of the expression of *Agrobacterium rhizogenes rolB*-GUS gene fusion in transgenic tobacco. *New Phytol.* **118**: 69-78 (1991).

Altamura, M. M., Capitani, F., Gazza, L., Capone, I., and Costantino, P. The plant oncogene *rolB* stimulates the formation of flower and root meristemoids in tobacco thin cell layers. *New Phytol.* **126**: 283-293 (1994).

Alvarez-Martinez, C. E., and Christie, P. J. Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* **73**: 775-808 (2009).

Andrea, L. E., and David, P. J. Sugars, signalling, and plant development. *J. Exp. Bot.* **63**: 3367-3377 (2012).

Aoki, S. Resurrection of an ancestral gene: functional and evolutionary analyses of the *Ngrol* genes transferred from Agrobacterium to Nicotiana. *J. Plant Res.* **117**: 329-337 (2004).

Aoki, S., and Syono, K. Synergistic function of *rolB*, *rolC*, *ORF13* and *ORF14* of TL-DNA of *Agrobacterium rhizogenes* in hairy root induction in *Nicotiana tabacum*. *Plant Cell Physiol.* **40**: 252-256 (1999).

Aoki, S., Kawaoka, A., Sekine, M., Ichikawa, T., Fujita, T., Shinmyo, A., and Syono, K. Sequence of the cellular T-DNA in the untransformed genome of *Nicotiana glauca* that is homologous to *ORFs 13* and *14* of the Ri plasmid and analysis of its expression in genetic tumors of *N. glauca* x *N. langsdorffii*. *Mol. Gen. Genet.* **243**: 706-710 (1994).

Backert, S., and Clyne, M. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter.* **16**: 19-25 (2011).

Baek, C. H., Farrand, S. K., Lee, K. E., Park, D. K., Lee, J. K., and Kim, K. S. Convergent evolution of amadori opine catabolic systems in plasmids of *Agrobacterium tumefaciens*. *J. Bacteriol.* **185**: 513-524 (2003).

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

Barbier-Brygoo, H., Ephritikhine, G., Maurel, C., and Guern, J. Perception of the auxin signal at the plasma membrane of tobacco mesophyll protoplasts. *Biochem. Soc. Trans.* **20**: 59-63 (1992).

Baron, C. From bioremediation to biowarfare: on the impact and mechanism of type IV secretion systems. *FEMS Microbiol. Lett.* **253**: 163-170 (2005).

Baumann, K., De Paolis, A., Costantino, P., and Gualberti, G. The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the *rolB* oncogene in plants. *Plant Cell* **11**: 323-334 (1999).

Binns, A. N., and Costantino, P. The Agrobacterium oncogenes. The Rhizobiaceae: molecular biology of model plant-associated bacteria (eds. H. P. Spaink, A. Kondorosi and P. J. Hooykaas), Dordrecht, The Netherlands: Kluwer Academic Publishers. pp 251-266 (1998).

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C. M., van Lammeren, A. A., Miki, B. L., Custers, J. B., van Lookeren Campagne, M. M. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **14**: 1737-1749 (2002).

Brencic, A., and Winans, S. C. Detection of and response to signals involved in host-microbe interactions by plant-associated bacteria. *Microbiol. Mol. Biol. Rev.* **69**: 155-194 (2005).

Broer, I., Dröge-Laser, W., Barker, R.F., Neumann, K., Klipp, W., and Pühler, A. Identification of the *Agrobacterium tumefaciens* C58 T-DNA genes *e* and *f* and their impact on crown gall tumour formation. *Plant Mol. Biol.* **27**: 41-57 (1995).

Boothby, T. C., Tenlen, J. R., Smith, F. W., Wang, J. R., Patanella, K. A., Osborne Nishimura, E., Tintori, S. C., Li, Q., Jones, C. D., Yandell, M., Messina, D. N., Glasscock, J., and Goldstein, B. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc. Natl. Acad. Sci. U.S.A.* Epub ahead of print (2015).

Bouchez, D., and Camilleri, C. Identification of a putative *rolB* gene on the TR-DNA of the *Agrobacterium rhizogenes* A4 Ri plasmid. *Plant Mol. Biol.* **14**: 617-619 (1990).

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C. M., van Lammeren, A. A., Miki, B. L., Custers, J. B., and van Lookeren Campagne, M. M. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **14**: 1737-1749 (2002).

Bouzar, H., Jones, J. B. *Agrobacterium larrymoorei* sp. nov., a pathogen isolated from aerial tumours of *Ficus benjamina*. *Int. J. Syst. Evol. Microbiol.* **51**: 1023-1026 (2001).

Bulgakov, V. P., Kisselev, K. V., Yakovlev, K. V., Zhuravlev, Y. N., Gontcharov, A. A., and Odintsova, N. A. Agrobacterium mediated transformation of sea urchin embryos. *Biotechnol. J.* **1**: 454-461 (2006).

Burr, T. J. and Otten, L. Crown gall of grape: biology and disease management. *Annu. Rev. Phytopathol.* **37**: 53-80 (1999).

Capone, I., Cardarelli, M., Mariotti, D., Pomponi, M., De Paolis, A., and Costantino, P. Different promoter regions control level and tissue specificity of expression of *Agrobacterium rhizogenes rolB* gene in plants. *Plant Mol. Biol.* **16**: 427-436 (1991).

Cardarelli, M., Mariotti, D., Pomponi, M., Spano, L., Capone, I., and Costantino, P. *Agrobacterium rhizogenes* T-DNA genes capable of inducing hairy root phenotype. *Mol. Gen. Genet.* **210**: 111-115 (1987).

Capone, I., Frugis, G., Costantino, P., and Cardarelli, M. Expression in different populations of cells of the root meristem is controlled by different domains of the *rolB* promoter. *Plant Mol. Biol.* **25**: 681-691 (1994).

Capone, I., Spano, L., Cardarelli, M., Bellincampi, D., Petit, A., and Costantino, P. Induction and growth properties of carrot roots with different complements of *Agrobacterium rhizogenes* T-DNA. *Plant Mol. Biol.* **13**: 43-52 (1989).

Carrington, J. C., Freed, D. D., and Leinicke, A. J. Bipartite signal sequence mediates nuclear translocation of the plant potyviral NIa protein. *Plant Cell* **3**: 953-962 (1991).

Casanova, E., Trillas, M. I., Moysset, L., and Vainstein, A. Influence of *rol* genes in floriculture. *Biotechnol. Adv.* **23**: 3-39 (2005).

Chen, K., and Otten, L. Morphological analysis of the *6b* oncogene-induced enation syndrome. *Planta* Epub ahead of print (2015).

Chen, K., Dorlhac de Borne, F., Szegedi, E., and Otten, L. Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus Nicotiana. *Plant Journal* **80**: 669-682 (2014).

Chichiriccò, G., Costantino, P., and Spanò, L. Expression of the *rolB* Oncogene from *Agrobacterium rhizogenes* during zygotic embryogenesis in tobacco. *Plant Cell Physiol.* **33**: 827-832 (1992).

Christey, M. C. Use of Ri-mediated transformation for production of transgenic plants. *In Vitro Cell Dev. Biol.* **37**: 687-700 (2001).

Christie, P. J., Whitaker, N., and González-Rivera, C. Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta.* **1843**: 1578-1591 (2014).

Citovsky, V., Warnick, D., and Zambryski, P. Nuclear import of Agrobacterium VirD2 and VirE2 proteins in maize and tobacco. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 3210-3214 (1994).

Clément, B., Perot, J., Geoffroy, P., Legrand, M., Zon, J., and Otten, L. Abnormal accumulation of sugars and phenolics in tobacco roots expressing the Agrobacterium T-*6b* oncogene and the role of these compounds in *6b*-induced growth. *Mol. Plant Microbe Interact* **20**: 53-62 (2007).

Clément, B., Pollmann, S., Weiler, E., Urbanczyk-Wochniak, E., and Otten, L. The *Agrobacterium vitis* T-*6b* oncoprotein induces auxin-independent cell expansion in tobacco. *Plant Journal* **45**: 1017-1027 (2006).

Conn, H. J. Validity of the genus *Alcaligenes*. *J. Bacteriol.* **44:** 353-360 (1942).

de Groot, M. J. A., Bundock, P., Hooykaas, P. J. J., and Beijersbergen, A. G. M. *Agrobacterium tumefaciens* mediated transformation of filamentous fungi. *Nat. Biotechnol.* **16**: 839-842 (1998).

Dehio, C., and Schell, J. Stable expression of a single-copy *rolA* gene in transgenic *Arabidopsis thaliana* plants allows an exhaustive mutagenic analysis of the transgene-associated phenotype. *Mol. Gen. Genet.* **241**: 359-366 (1993).

Delbarre, A., Muller, P., Imhoff, V., Barbier-Brygoo, H., Maurel, C., Leblanc, N., Perrot-Rechenmann, C., and Guern, J. The *rolB* gene of *Agrobacterium rhizogenes* does not increase the auxin sensitivity of tobacco protoplasts by modifying the intracellular auxin concentration. *Plant Physiol.* **105**: 563-569 (1994).

Dessaux, Y., Petit, A., Farrand, S. K., and Murphy, P. J. Opines and opine-like molecules involved in plant-Rhizobiaceae interactions. The Rhizobiaceae: Molecular Biology of Model Plant-Associated Bacteria. pp 173-197 (1998).

Di Cola, A., Poma, A., and Spano, L. *RolB* expression pattern in the early stages of carrot somatic embryogenesis. *Cell Biol. Int.* **21**: 595-600 (1997).

Dkhar, J., and Pareek, A. What determines a leaf's shape? *Evodevo*. **22**: 5-47 (2014).

Drevet, C., Brasileiro, A. C., and Jouanin, L. Oncogene arrangement in a shooty strain of *Agrobacterium tumefaciens*. *Plant Mol. Biol.* **25**: 83-90 (1994).

Duckely, M., and Hohn, B. The VirE2 protein of *Agrobacterium tumefaciens*: the Yin and Yang of T-DNA transfer. *Fems Microbiol. Let.* **223**: 1-6 (2003).

Durrenberger, F., Crameri, A., Hohn, B., and Koukolikova-Nicola, Z. Covalently bound VirD2 protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 9154-9158 (1989).

Estruch, J. J., Parets-Soler, A., Schmülling, T., and Spena, A. Cytosolic localization in transgenic plants of the *rolC* peptide from *Agrobacterium rhizogenes*. *Plant Mol. Biol.* **17**: 547-550 (1991a).

Estruch, J. J., Schell, J., and Spena, A. The protein encoded by the *rolB* plant oncogene hydrolyses indole glucosides. *EMBO J.* **10**: 3125-3128 (1991b).

Eveland, A. L., and Jackson, D. P. Sugar, signaling, and plant development. *J. Exp. Bot.* **63**: 3367-3377 (2012).

Filippini, F., Rossi, V., Marin, O., Trovato, M., Costantino, P., Downey, P. M., Lo Schiavo, F., and Terzi, M. A plant oncogene as a phosphatase. *Nature* **379**: 499-500 (1996).

Fischer, W. Assembly and molecular mode of action of the *Helicobacter pylori* Cag type IV secretion apparatus. *FEBS J.* **278**: 1203-1212 (2011).

Fladung, M. Transformation of diploid and tetraploid potato clones with the *rolC* gene of *Agrobacterium rhizogenes* and characterization of transgenic plants. *Plant Breeding* **104**: 295-304 (1990).

Florez, S. L., Erwin, R. L., Maximova, S. N., Guiltinan, M. J., and Curtis, W. R. Enhanced somatic embryogenesis in *Theobroma cacao* using the homologous BABY BOOM transcription factor. *BMC Plant Biol.* **15**: 121 (2015).

Flores-Mireles, A. L., Eberhard, A., and Winans, S. C. *Agrobacterium tumefaciens* can obtain sulphur from an opine that is synthesized by octopine synthase using S-methylmethionine as a substrate. *Mol. Microbiol.* **84**: 845-856 (2012).

Fründt, C., Meyer, A. D., Ichikawa, T., and Meins, F. Evidence for the ancient transfer of Ri plasmid T-DNA genes between bacteria and plants. In: Syvanen M, Kado CI (eds.) Horizontal gene transfer. Chapman and Hall, London. pp 94-106 (1998).

Furner, I. J., Huffman, G. A., Amasino, R. M., Garfinkel, D. J., Gordon, M. P., and Nester, E. W. An Agrobacterium transformation in the evolution of the genus Nicotiana. *Nature* **319**: 422-427 (1986).

Fuqua, C., and Winans, S. C. Conserved cis-acting promoter elements are required for density dependent transcription of *Agrobacterium tumefaciens* conjugal transfer genes. *J. Bacteriol.* **178**: 435-40 (1996).

Gazzarrini, S., and McCourt, P. Genetic interaction between ABA, ethylene and sugar signaling pathways. *Curr. Opin. Plant Biol.* **4**: 387-391 (2001).

Gelvin, S. B. Traversing the Cell: Agrobacterium T-DNA's journey to the host genome. *Front. Plant Sci.* **26**: 3-52 (2012).

Gibson, S. I. Control of plant development and gene expression by sugar signaling. *Curr. Opin. Plant Biol.* **8**: 93-102 (2005).

Grémillon, L., Helfer, A., Clément, B., and Otten, L. New plant growth-modifying properties of the Agrobacterium T-*6b* oncogene revealed by the use of a dexamethasone-inducible promoter. *Plant Journal* **37**: 218-228 (2004).

Guivarch, A., Carnerio, M., Vilaine, F., Pautot, V., and Criqui, D. Tissue-specific expression of the *rolA* gene mediates morphological changes in transgenic tobacco. *Plant Mol. Biol.* **30**: 125-134 (1996).

Guyon, P., Petit, A., Tempé, J., and Dessaux, Y. Transformed plants producing opines specifically promote growth of opine-degrading agrobacteria. *Mol. Plant Microbe* **6**: 92-98 (1992).

Hansen, G., Larribe, M., Vaubert, D., Tempé, J., Biermann, B. J., Montoya, A. L., Chilton, M. D., and Brevet, J. *Agrobacterium rhizogenes* pRi8196 T-DNA: mapping and DNA sequence of functions involved in mannopine synthesis and hairy root differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 7763-7767 (1991).

Hansen, G., Vaubert, D., Heron, J. H., Clérot, D., Tempé, J., and Brevet, J. Phenotypic effects of overexpression of *Agrobacterium rhizogenes* T-DNA *ORF13* in transgenic tobacco plants are mediated by diffusible factor(s). *Plant Journal* **4**: 581-585 (1993).

Hansen, G., Vaubert, D., Clérot, D., and Brevet, J. Wound-inducible and organ-specific expression of *ORF13* from *Agrobacterium rhizogenes* 8196 T-DNA in transgenic tobacco plants. *Mol. Gen. Genet.* **254**: 337-343 (1997).

Haq, I. U., Zhang, M., Yang, P., and van Elsas, J. D. The interactions of bacteria with fungi in soil: emerging concepts. *Adv. Appl. Microbiol.* **89**: 185-215 (2014).

Hartig, K., and Beck, E. Crosstalk between auxin, cytokinins, and sugars in the plant cell cycle. *Plant Biol.* **8:** 389-396 (2006).

Helfer, A., Clément, B., Michler, P., and Otten, L. The Agrobacterium oncogene AB-*6b* causes a graft-transmissible enation syndrome in tobacco. *Plant Mol. Biol.* **52**: 483-493 (2003).

Helfer, A., Pien, S., and Otten, L. Functional diversity and mutational analysis of Agrobacterium *6B* oncoproteins. *Mol. Genet. Genomics* **267**: 577-586 (2002).

Herrera-Estrella, A., Chen, Z. M., Van Montagu, M., and Wang, K. VirD proteins of *Agrobacterium tumefaciens* are required for the formation of a covalent DNA protein complex at the 5' terminus of T-strand molecules. *EMBO. J.* **7**: 4055-4062 (1988).

Herrera-Estrella, A., Van Montagu, M., and Wang, K. A bacterial peptide acting as a plant nuclear targeting signal: the amino-terminal portion of Agrobacterium VirD2 protein directs a β-galactosidase fusion protein into tobacco nuclei. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 9534-9537 (1990).

Hildebrand, E. M. Cane gall of brambles caused by *Phytomonas rubi* n. sp. *J. Agric. Res.* **61**: 685-696 (1940).

Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J., and Schilperoort, R. A. A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature* **303**: 179-180 (1983).

Horiguchi, G., Kim, G. T., and Tsukaya, H. The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant Journal* **43**: 68-78 (2005).

Hooykaas, P. J. J., Den Dulk-Ras, H., and Schilperoort, R. A. The *Agrobacterium tumefaciens* T-DNA gene *6b* is an oncogene. *Plant Mol. Biol.* **11**: 791-794 (1988).

Howard, E. A., Zupan, J. R., Citovsky, V., and Zambryski, P. C. The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. *Cell* **68**: 109-118 (1992).

Huang, J., Yue, J., and Hu, X. Origin of plant auxin biosynthesis in charophyte algae: a reply to Wang et al. *Trends Plant Sci.* **19**: 743 (2014).

Huang, T., Kerstetter, R., and Irish, V. F. APUM23, a PUF family protein, functions in leaf development and organ polarity in Arabidopsis. *J. Exp. Bot.* **65**: 1181-1191 (2014).

Ian, M. S. Opine content of unorganised and teratomatous tobacco crown gall tissues. *Plant Sci. Lett.* **16**: 239-248 (1979).

Jefferson, R. A. Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* **5**: 387-405 (1987).

Kado, C. I. Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by *Agrobacterium tumefaciens*. *Front. Microbiol.* **5**: 340 (2014).

Kerr, A., and Panagopoulos, C. G. Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their biological control. *Phytopathol. Z.* **90**: 172-179 (1977).

Kerstetter, R. A., Laudencia-Chingcuanco, D., Smith, L. G., and Hake, S. Loss of function mutations in the maize homeobox gene, *knotted1*, are defective in shoot meristem maintenance. *Development* **124**: 3045-3054 (1997).

Kim, J., Jung, J. H., Reyes, J. L., Kim, Y. S., Kim, S. Y., Chung, K. S., Kim, J. A., Lee, M., Lee, Y., Kim, V. N., Chua, N. H., and Park, C. M. MicroRNA-directed cleavage of ATHB15 mRNA regulates vascular development in Arabidopsis inflorescence stems. *Plant Journal* **42**: 84-94 (2005).

Kim, K. S., Baek, C. H., Lee, J. K., Yang, J. M., and Farrand, S. K. Intracellular accumulation of mannopine, an opine produced by crown gall tumors, transiently inhibits growth of *Agrobacterium tumefaciens*. *Mol. Plant Microbe Interact.* **14**: 793-803 (2001).

Knapp, S., Chase, M. W., and Clarkson, J. J. Nomenclatural changes and a new sectional classification in Nicotiana (Solanaceae). *Taxon* **53**: 73-82 (2004).

Koncz, C., and Schell, J. The promoter of TL-DNA gene *5* controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Mol. Gen. Genet.* **204**: 383-396 (1986).

Körber, H., Strizhov, H., Staiger, D., Feldwisch, J., Olsson, O., Sandberg, G., Palme, K., Schell, J., and Koncz, C. T-DNA gene *5* of Agrobacterium modulates auxin response by autoregulated synthesis of a growth hormone antagonist in plants. *EMBO J.* **10**: 3983-3991 (1991).

Kunik, T., Tzfira, T., Kapulnik, Y., Gafni, Y., Dingwall, C., and Citovsky, V. Genetic transformation of HeLa cells by Agrobacterium. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 1871-1876 (2001).

Kyndt, T., Quispe, D., Zhai, H., Jarret, R., Ghislain, M., Liu, Q., Gheysen, G., and Kreuze, J. F. The genome of cultivated sweet potato contains Agrobacterium T-DNAs with expressed genes: An example of a naturally transgenic food crop. *Proc. Natl. Acad. Sci. U.S.A.* **112**: 5844-5849 (2015).

Laux, T., Mayer, K. F., Berger, J., and Jürgens, G. The *WUSCHEL* gene is required for shoot and floral meristem integrity in Arabidopsis. *Development* **122**: 87-96 (1996).

Lemcke, K., and Schmülling, T. Gain of function assays identify non-*rol* genes from *Agrobacterium rhizogenes* TL-DNA that alter plant morphogenesis or hormone sensitivity. *Plant Journal* **15**: 423-433 (1998).

Lemcke, K., Prinsen, E., van Onckelen, H., and Schmülling, T. The *ORF8* gene product of *Agrobacterium rhizogenes* TL-DNA has tryptophan 2-monooxygenase activity. *Mol. Plant Microbe Interact.* **13**: 787-790 (2000).

Leung, J., Fukuda, H., Wing, D., Schell, J., and Masterson, R. Functional analysis of cis-elements, auxin response and early developmental profiles of the mannopine synthase bidirectional promoter. *Mol. Gen. Genet.* **230**: 463-474 (1991).

Levesque, H., Delépelaire, P., Rouzé, P., Slightom, J., and Tepfer, D. Common evolutionary origin of the central portion of the Ri TL-DNA of *Agrobacterium rhizogenes* and the Ti T-DNAs of *Agrobacterium tumefaciens*. *Plant Mol. Biol.* **11**: 731-744 (1988).

Llosa, M., Roy, C., and Dehio, C. Bacterial type IV secretion systems in human disease. *Mol. Microbiol.* **73**: 141-151 (2009).

Magori, S., and Citovsky, V. The role of the ubiquitin-proteasome system in *Agrobacterium tumefaciens*-mediated genetic transformation of plants. *Plant Physiol.* **160**: 65-71 (2012).

Magrelli, A., Langenkemper, K., Dehio, C., Schell, J., and Spena, A. Splicing of the *rolA* transcript of *Agrobacterium rhizogenes* in Arabidopsis. *Science* **23**: 1986-1988 (1994).

Matthysse, A. G. Attachment of Agrobacterium to plant surfaces. *Front. Plant Sci.* **5**: 252 (2014).

Matveeva, T. V., and Lutova, L. A. Horizontal gene transfer from Agrobacterium to plants. *Front. Plant Sci.* **5**: 326 (2014).

Maurel, C., Barbier-Brygoo, H., Spena, A., Tempé, J., and Guern, J. Single *rol* genes from the *Agrobacterium rhizogenes* T(L)-DNA alter some of the cellular responses to auxin in *Nicotiana tabacum*. *Plant Physiol.* **97**: 212-216 (1991).

Maurel, C., Brevet, J., Barbier-Brygoo, H., Guern, J., and Tempé, J. Auxin regulates the promoter of the root-inducing *rolB* gene of *Agrobacterium rhizogenes* in transgenic tobacco. *Mol. Gen. Genet.* **223**: 58-64 (1990).

Messens, E., Lenaerts, A., Van Montagu, M., and Hedges, R. W. Genetic basis for opine secretion from crown gall tumor cells. *Mol. Gen. Genet.* **199**: 344-348 (1985).

Meyer, A. D., Ichikawa, T., and Meins, F. Horizontal gene transfer: regulated expression of tobacco homologue of the *Agrobacterium rhizogenes rolC* gene. *Mol. Gen. Genet.* **249**: 265-273 (1995).

Michielse, C. B., Hooykaas, P. J. J., van den Hondel, C. A. M. J. J., and Ram, A. F. J. Agrobacterium mediated transformation of the filamentous fungus *Aspergillus awamori*. *Nat. Protoc.* **3**: 1671-1678 (2008).

Mishra, B. S., Singh, M., Aggrawal, P., and Laxmi, A. Glucose and auxin signaling interaction in controlling *Arabidopsis thaliana* seedlings root growth and development. *PLoS One* **4**: e4502 (2009).

Mohajjel-Shoja, H. Contribution to the study of the *Agrobacterium rhizogenes* plast genes, *rol B* and *rol C*, and their homologs in *Nicotiana tabacum*. Thesis of Strasbourg University (2010).

Mondy, S., Lenglet, A., Beury-Cirou, A., Libanga, C., Ratet, P., Faure, D., and Dessaux, Y. An increasing opine carbon bias in artificial exudation systems and genetically modified plant rhizospheres leads to an increasing reshaping of bacterial populations. *Mol. Ecol.* **23**: 4846-4861 (2014).

Moore, L. W., Chilton, W. S, and Canfield, M. L. Diversity of opines and opine-catabolizing bacteria isolated from naturally occurring crown gall tumors. *Appl. Environ. Microbiol.* **63**: 201-207 (1997).

Moriuchi, H., Okamoto, C., Nishihama, R., Yamashita, I., Machida, Y., and Tanaka, N. Nuclear localization and interaction of RolB with plant 14-3-3 proteins correlates with induction of adventitious roots by the oncogene *rolB*. *Plant Journal* **38**: 260-275 (2004).

Muniesa, M., Colomer Lluch, M., and Jofre, J. Potential impact of environmental bacteriophages in spreading antibiotic resistance genes. *Future Microbiol.* **8**: 739-751 (2013).

Mysore, K. S., Bassuner, B., Deng, X. B., Darbinian, N. S., Motchoulski, A., Ream, W., and Gelvin, S. B. Role of the *Agrobacterium tumefaciens* VirD2 protein in T-DNA transfer and integration. *Mol. Plant Microbe Interact.* **11**: 668-683 (1998).

Nagai, H., and Kubori, T. Type IVB secretion systems of Legionella and other Gram-negative bacteria. *Front. Microbiol.* **2**:136 (2011).

Nagata, N., Kosono, S., Sekine, M., Shinmyo, A., and Syono, K. Different expression patterns of the promoters of the *NgrolB* and *NgrolC* genes during the development of tobacco genetic tumors. *Plant Cell Phys.* **37**: 489-498 (1996).

Nakamura, Y., Itoh, T., Matsuda, H., and Gojobori, T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* **36**: 760-766 (2004).

Nilsson, O., and Olsson, O. Getting to the root: the role of the *Agrobacterium rhizogenes rol* genes in the formation of hairy roots. *Physiol. Plant* **100**: 463-473 (1997).

Nilsson, O., Moritz, T., Imbault, N., Sandberg, G., and Olsson, O. Hormonal characterization of transgenic tobacco plants expressing the *rolC* gene of *Agrobacterium rhizogenes* TL-DNA. *Plant Physiol.* **102**: 363-371 (1993).

Nilsson, O., Moritz, T., Sundberg, B., Sandberg, G., and Olsson, O. Expression of the *Agrobacterium rhizogenes rolC* gene in a deciduous forest tree alters growth and development and leads to stem fasciation. *Plant Physiol.* **112**: 493-502 (1996).

Ophel, K., and Kerr, A. *Agrobacterium vitis*-new species for strains of Agrobacterium biovar 3 from grapevine. *Int. J. Syst. Bacteriol.* **40**: 236-241 (1990).

Ormeño-Orrillo, E., Servín-Garcidueñas, L. E., Rogel, M. A., González, V., Peralta, H., Mora, J., Martínez-Romero, J., and Martínez-Romero, E. Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. *Syst. Appl. Microbiol.* **38**: 287-291 (2015).

Otten, L., and Helfer, A. Biological activity of the *rolB*-like 5' end of the A4-*orf8* gene from the *Agrobacterium rhizogenes* TL-DNA. *Mol. Plant Microbe Interact.* **14**: 405-411 (2001).

Otten, L., and De Ruffray, P. *Agrobacterium vitis* nopaline Ti plasmid pTiAB4: relationship to other Ti plasmids and T-DNA structure. *Mol. Gen. Genet.* **245**: 493-505 (1994).

Otten, L., and Schmidt, J. A T-DNA from the Agrobacterium limited-host range strain AB2/73 contains a single oncogene. *Mol. Plant Microbe Interact.* **11**: 335-342 (1998).

Otten, L., Salomone, J. Y., Helfer, A., Schmidt, J., Hammann, P., and De Ruffray, P. Sequence and functional analysis of the left-hand part of the T-region from the nopaline-type Ti plasmid pTiC58. *Plant Mol. Biol.* **41**: 765-776 (1999).

Otten, L., Vreugdenhil, D., and Schilperoort, R. A. Properties of D(+)-lysopine dehydrogenase from crown gall tumour tissue. *Biochim. Biophys. Acta.* **485**: 268-277 (1977).

Pandolfini, T., Storlazzi, A., Calabria, E., Defez, R., and Spena, A. The spliceosomal intron of the *rolA* gene of *Agrobacterium rhizogenes* is a prokaryotic promoter. *Mol. Microbiol.* **35**: 1326-1334 (2000).

Rice, D. W., Alverson, A. J., Richardson, A. O., Young, G. J., Sanchez-Puerta, M. V., Munzinger, J., Barry, K., Boore, J. L., Zhang, Y., dePamphilis, C. W., Knox, E. B., and Palmer, J. D.

Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **342**: 1468-1473 (2013).

Rossi, F., Rizzotti, L., Felis, G. E., and Torriani, S. Horizontal gene transfer among microorganisms in food: current knowledge and future perspectives. *Food Microbiol.* **42**: 232-243 (2014).

Rossi, L., Hohn, B., and Tinland, B. The VirD2 protein of *Agrobacterium tumefaciens* carries nuclear localization signals important for transfer of T-DNA to plant. *Mol. Gen. Genet.* **239**: 345-353 (1993).

Saito, K., Yamazaki, M., Kaneko, H., Murakoshi, I., Fukuda, Y., and Van Montagu, M. Tissue-specific and stress-enhancing expression of the TR promoter for mannopine synthase in transgenic medicinal plants. *Planta* **184**: 40-46 (1991).

Sakakibara, H. Cytokinins: activity, biosynthesis, and translocation. *Annu. Rev. Plant Biol.* **57**: 431-449 (2006).

Salehin, M., Bagchi, R., and Estelle, M. $SCF^{TIR1/AFB}$-based auxin perception: mechanism and role in plant growth and development. *Plant Cell* **27**: 10 (2015).

Savka, M. A., Dessaux, Y., Oger, P., and Rossbach, S. Engineering bacterial competitiveness and persistence in the phytosphere. *Mol. Plant Microbe Interact.* **15**: 866-874 (2002).

Schaller, G. E., Bishopp, A., and Kieber, J. J. The yin-yang of hormones: cytokinin and auxin interactions in plant development. *Plant Cell* **27**: 44-63 (2015).

Schmidt, J. Etude d'un oncogène de la souche AB2/73 d'*Agrobacterium tumefaciens*. Thèse de l'université Louis Pasteur de Strasbourg (1999).

Schmülling, T., Fladung, M., Grossmann, K., and Schell, J. Hormonal content and sensitivity of transgenic tobacco and potato plants expressing single *rol* genes of *Agrobacterium rhizogenes* T-DNA. *Plant Journal* **3**: 371-382 (1993).

Schmülling, T., Schell, J., and Spena, A. Single genes from *Agrobacterium rhizogenes* influence plant development. *EMBO J.* **7**: 2621-2629 (1988).

Schrammeijer, B., Risseeuw, E., Pansegrau, W., Regensburg-Tuïnk, T. J., Crosby, W. L., and Hooykaas, P. J. J. Interaction of the virulence protein VirF of *Agrobacterium tumefaciens* with plant homologs of the yeast Skp1 protein. *Curr. Biol.* **11**: 258-62 (2001).

Sierro, N., Battey, J. N., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M. C., and Ivanov, N. V. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* **5**: 3833 (2014).

Shen, W. H., Petit, A., Guern, J., and Tempé, J. Hairy roots are more sensitive to auxin than normal roots. *Proc. Natl. Acad. Sci. U.S.A.* **85**: 3417-3421 (1988).

Skylar, A., Sung, F., Hong, F., Chory, J., and Wu, X. Metabolic sugar signal promotes Arabidopsis meristematic proliferation via G2. *Dev. Biol.* **351**: 82-89 (2011).

Sohn, S. H., Frost, J., Kim, Y. H., Choi, S. K., Lee, Y., Seo, M. S., Lim, S. H., Choi, Y., Kim, K. H., and Lomonossoff, G. Cell-autonomous-like silencing of GFP-partitioned transgenic *Nicotiana benthamiana*. *J. Exp. Bot.* **65**: 4271-4283 (2014).

Spena, A., Aalen, R. B., and Schulze, S. C. Cell-autonomous behavior of the *rolC* gene of *Agrobacterium rhizogenes* during leaf development: a visual assay for transposon excision in transgenic plants. *Plant Cell* **1**: 1157-1164 (1989).

Spena, A., Schmülling, T., Koncz, C., and Schell, J. Independent and synergistic activity of *rolA, B* and *C* loci in stimulating abnormal growth in plants. *EMBO J.* **6**: 3891-3899 (1987).

Stepanova, A. N., Robertson Hoyt, J., Yun, J., Benavente, L. M., Xie, D. Y., Dolezal, K., Schlereth, A., Jürgens, G., and Alonso, J. M. TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* **133**: 177-191 (2008).

Stieger, P. A., Meyer, A. D., Kathmann, P., Frundt, C., Niederhauser, I., Barone, M., and Kuhlemeier, C. The *orf13* T-DNA gene of *Agrobacterium rhizogenes* confers meristematic competence to differentiated cells. *Plant Physiol.* **135**: 1798-1808 (2004).

Studholme, D. J., Downie, J. A, and Preston, G. M. Protein domains and architectural innovation in plant-associated Proteobacteria. *BMC Genomics* **16**: 6-17 (2005).

Suzuki, K., Yamashita, I., and Tanaka, N. Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution. *Plant Journal* **32**: 775-787 (2002).

Tao, Y., Ferrer, J. L., Ljung, K., Pojer, F., Hong, F., Long, J. A., Li, L., Moreno, J. E., Bowman, M. E., Ivans, L. J., Cheng, Y., Lim, J., Zhao, Y., Ballaré, C. L., Sandberg, G., Noel, J. P., and Chory, J. Rapid synthesis of auxin via a new tryptophan dependent pathway is required for shade avoidance in plants. *Cell* **133**: 164-176 (2008).

Teeri, T. H., Lehväslaiho, H., Franck, M., Uotila, J., Heino, P., Palva, E. T., Van Montagu, M., and Herrera-Estrella, L. Gene fusions to lacZ reveal new expression patterns of chimeric genes in transgenic plants. *EMBO J.* **8**: 343-350 (1989).

Tinland, B., Fournier, P., Heckel, T., and Otten, L. Expression of a chimaeric heat-shock-inducible Agrobacterium *6b* oncogene in *Nicotiana rustica*. *Plant Mol. Biol.* **18**: 921-930 (1992a).

Tinland, B., Huss, B., Paulus, F., Bonnard, G., and Otten, L. *Agrobacterium tumefaciens 6b* genes are strain-specific and affect the activity of auxin as well as cytokinin genes. *Mol. Gen. Genet.* **219**: 217-224 (1989).

Tinland, B., Koukolikova-Nicola, Z., Hall, M. N., and Hohn, B. The T-DNA linked VirD2 protein contains two distinct functional nuclear localization signals. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 7442-7446 (1992b).

Tinland, B., Rohfritsch, O., Michler, P., and Otten, L. *Agrobacterium tumefaciens* T-DNA gene *6b* stimulates *rol*-induced root formation, permits growth at high auxin concentrations and increases root size. *Mol. Gen. Genet.* **223**: 1-10 (1990).

Trovato, M., and Linhares, F. Recent advances on *rol* genes research: a tool to study plant differentiation. *Current Topics Plant Biol.* **1**: 51-62 (1999).

Trovato, M., Maras, B., Linhares, F., and Constantino, P. The plant oncogene *rolD* encodes a functional ornithine cyclodeaminase. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 13449-13453 (2001).

Udagawa, M., Aoki, S., and Syono, K. Expression analysis of the *NgORF13* promoter during the development of tobacco genetic tumors. *Plant Cell Physiol.* **45**: 1023-1031 (2004).

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. Activation and repression of transcription by auxin-response factors. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 5844-5849 (1999).

Umber, M., Clément, B., and Otten, L. The T-DNA oncogene A4-*orf8* from *Agrobacterium rhizogenes* A4 induces abnormal growth in tobacco. *Mol. Plant Microbe Interact.* **18**: 205-211 (2005).

Umber, M., Voll, L., Weber, A., Michler, P., and Otten, L. The *rolB*-like part of the *Agrobacterium rhizogenes orf8* gene inhibits sucrose export in tobacco. *Mol. Plant Microbe Interact.* **15**: 956-962 (2002).

Van Onckelena, H., Rüdelsheima, P., Inzéb, D. Follinb, A., Messensc, E., Horemansa, S., Schellb, J., Van Montagub, M., De Greefa, J. Tobacco plants transformed with the Agrobacterium T-DNA gene 1 contain high amounts of indole-3-acetamide. *FEBS Lett*. **181**: 373-376 (1985).

Van Slogteren, G. M., Hooykaas, P. J. J., Schilperoort, R. A. Silent T-DNA genes in plant lines transformed by *Agrobacterium tumefaciens* are activated by grafting and by 5-azacytidine treatment. *Plant Mol. Biol.* **3**: 333-336 (1984).

Vladimirov, I. A., Matveeva, T. V., and Lutova, L. A. Opine biosynthesis and catabolism genes of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *Genetika.* **51**: 137-146 (2015).

Wang, L., and Ruan, Y. L. Regulation of cell division and expansion by sugar and auxin signaling. *Front. Plant Sci*. **30**: 163 (2013).

Wang, M., Soyano, T., Machida, S., Yang, J. Y., Jung, C., Chua, N. H., and Yuan, Y. A. Molecular insights into plant cell proliferation disturbance by Agrobacterium protein 6b. *Genes Dev.* **25**: 64-76 (2011).

Ward, E. R., and Barnes, W. M. VirD2 protein of *Agrobacterium tumefaciens* very tightly linked to the 5' end of T-strand DNA. *Science* **242**: 927-930 (1988).

White, F. F., Garfinkel, D. J., Huffman, G. A., Gordon, M. P., and Nester, E. W. Sequence homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature* **301**: 348-350 (1983).

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

Wolanin, P. M., Thomason, P. A., and Stock, J. B. Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biol.* **3**: 3013-3018 (2002).

Wolterink-van, Loo, S., Escamilla, Ayala, A. A., Hooykaas, P. J. J., and van Heusden, G. P. Interaction of the *Agrobacterium tumefaciens* virulence protein VirD2 with histones. *Microbiology* **161**: 401-410 (2015).

Yokoyama, R., Hirose, T., Fujii, N., Aspuria, E. T., Kato, A., and Uchimiya, H. The *rolC* promoter of *Agrobacterium rhizogenes* Ri plasmid is activated by sucrose in transgenic tobacco plants. *Mol. Gen. Genet.* **244**: 15-22 (1994).

Young, C., and Nester, E. W. Association of the VirD2 protein with the 5' end of T-strands in *Agrobacterium tumefaciens*. *J. Bacteriol.* **170**: 3367-3374 (1988).

Young, J. M. An overview of bacterial nomenclature with special reference to plant pathogens. *Syst. Appl. Microbiol.* **31**: 405-424 (2008).

Zhang, M., Pereira e Silva Mde, C., Chaib De Mares, M., and van Elsas, J. D. The mycosphere constitutes an arena for horizontal gene transfer with strong evolutionary implications for bacterial-fungal interactions. *FEMS Microbiol. Ecol.* **89**: 516-526 (2014).

Zhao, Y. Auxin biosynthesis. *The Arabidopsis Book* **12**: e0173 (2014).

Zheng, W., Zhang, X., Yang, Z., Wu, J., Li, F., Duan, L., Liu, C., Lu, L., Zhang, C., and Li, F. AtWuschel promotes formation of the embryogenic callus in *Gossypium hirsutum*. *PLoS One* **9**: e87502 (2014).

Zhou, G. K., Kubo, M., Zhong, R., Demura, T., and Ye, Z. H. Overexpression of miR165 affects apical meristem formation, organ polarity establishment and vascular development in Arabidopsis. *Plant Cell Physiol.* **48**: 391-404 (2007).

Résumé de Thèse de Ke CHEN.

Titre: Séquençe et l'analyse fonctionnelle des ADN-T dans Nicotiana

Etat de la question

La bactérie *Agrobacterium tumefaciens* est bien connue pour son utilisation en génie génétique végétale où elle sert comme vecteur de gènes (figure1A). Cette bactérie et les espèces voisines *Agrobacterium rhizogenes* et *Agrobacterium vitis* sont des bactéries phytopathogènes qui induisent respectivement des tumeurs et des racines anormales sur des plantes sensibles telles que la vigne ou les arbres fruitiers (figure1B et C).

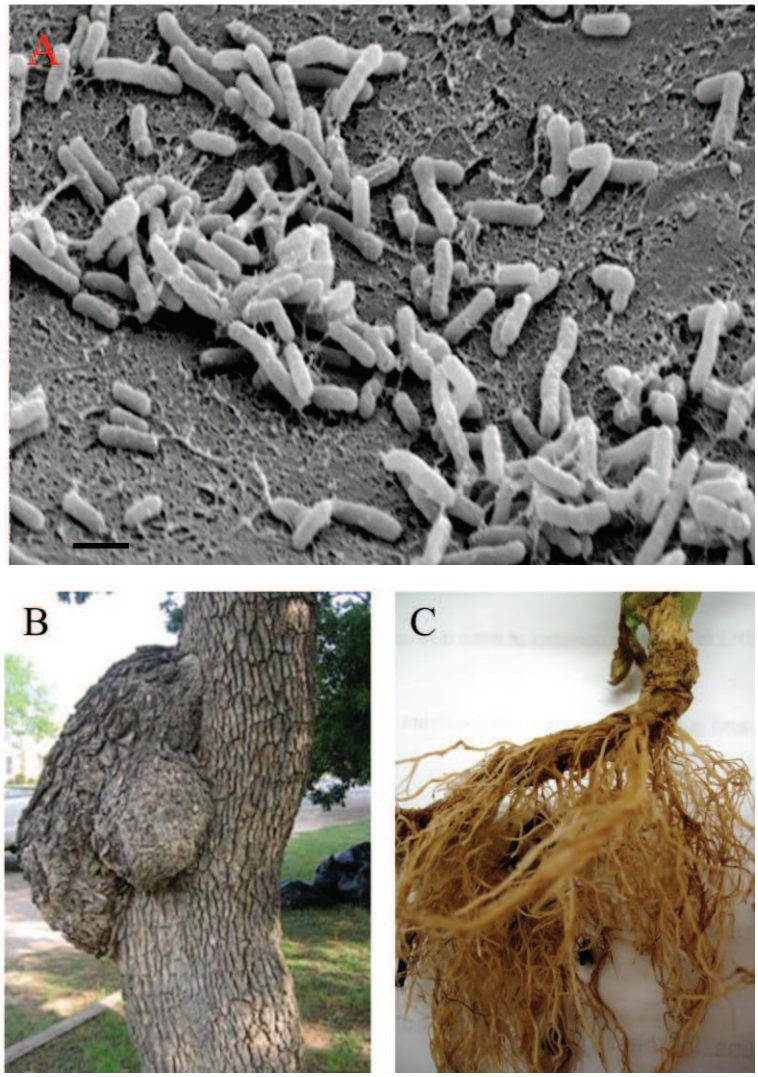


Figure 1. Agrobacterium et le syndrome de la plante associée
(A) l'image de microscopie électronique de *A. tumefaciens* souche C58 (Bar: 1µm) (de http://bacmap.wishartlab.com/organisms/79). (B) La galle du collet causée par *A. tumefaciens* (de http://brokenwillow.com/gallery/crown-gall). (C) Racines poilues causées par *A. rhizogenes* (de http://www.cals.ncsu.edu/course/pp728/Rhizobium/Rhizobium_rhizogenes.htm).

L'action pathogène résulte d'un transfert horizontal d'un fragment d'ADN (T-DNA) de la bactérie vers l'hôte végétal à partir d'un plasmide, le pTi (plasmide inducteur de tumeurs) ou pRi (plasmide inducteur de racines) (figure 2, 3 et 4).

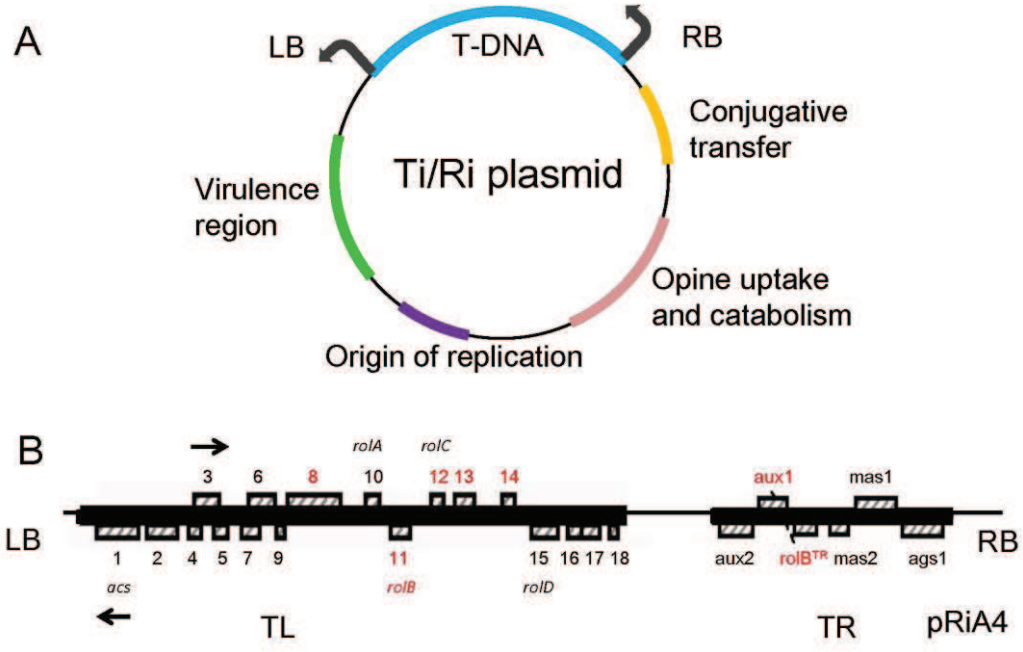


Figure 2. Schéma de plasmide Ti/Ri avec un exemple du plasmide Ri de la souche *A. rhizogenes* A4
(A) Un schéma général d'un plasmide Ti et Ri de *A. tumefaciens* ou *A. rhizogenes*. Les flèches indiquent le sens du transfert d'ADN-T. (B) Les gènes présents sur les deux T-DNAs d'*A. rhizogenes* A4: TL et TR. Les flèches indiquent l'orientation (5 'vers 3') des ORFs. Les gènes qui appartiennent à la famille des gènes plast sont marqués en rouge. LB: bordure gauche; RB: bordure droite (adapté de Mohajjel-Shoja 2010, avec quelques modifications).

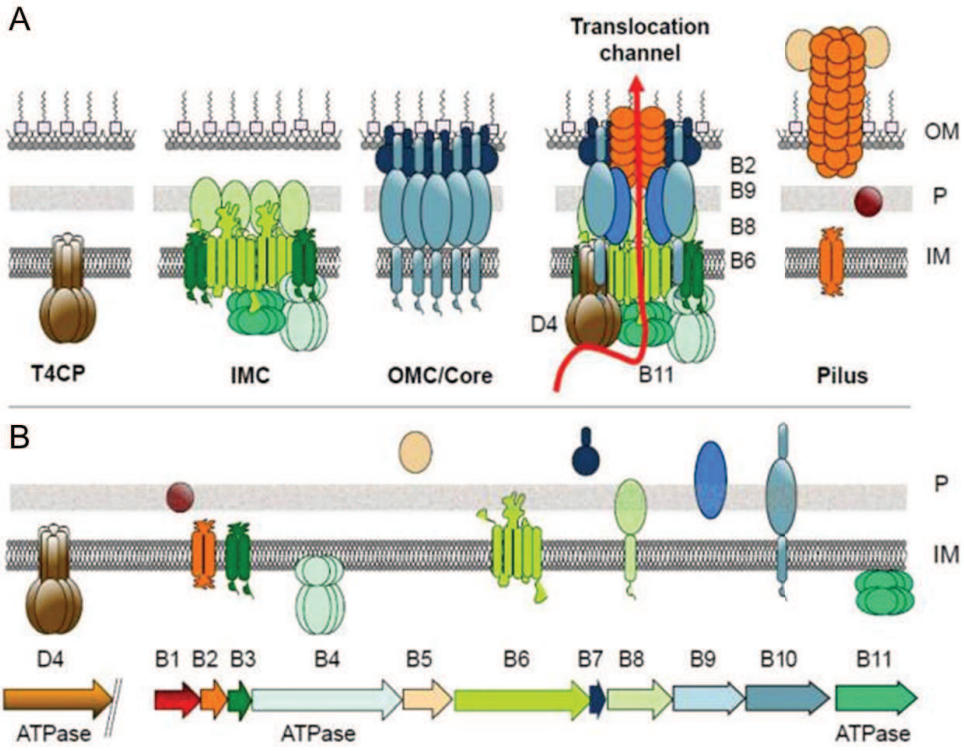


Figure 3. Schéma des éléments de *A. tumefaciens* représentant le système de sécrétion de type IV (T4SS) (Christie et al., 2014)

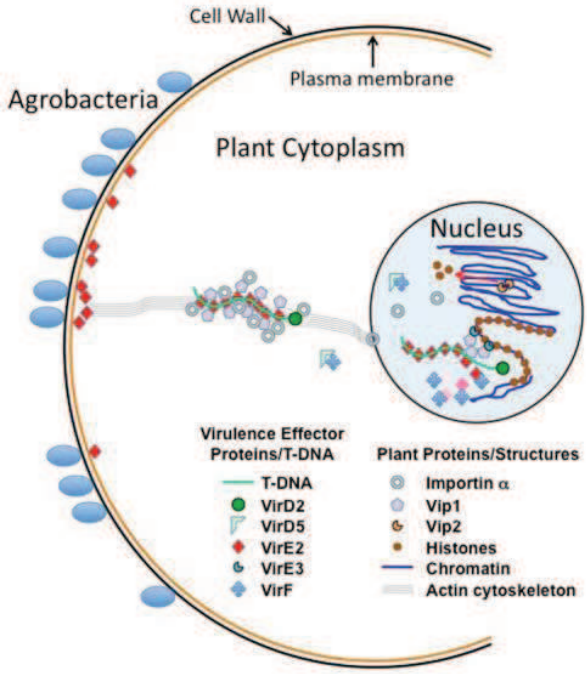


Figure 4. Transfert du complexe T d'Agrobacterium dans la cellule végétale (Gelvin, 2012)

Les gènes des T-DNA sont de quatre types.

1. Les gènes de synthèse d'opines, des petites molécules qui servent comme source de carbone, d'azote et de phosphore pour la croissance de la bactérie (tableau 1 et figure 5).

Tableau 1. Les différents plasmides Ti et Ri fondées sur des opinions qu'ils produisent (d'après from Dessaux et al., 1998)

Type de plasmide	Relevant opine products
Plasmides Ti	
Octopine	Octopine, l'acide octopinique, lysopine, histopine, agropine, mannopine, l'acide agropinic et l'acide mannopinic
Nopaline	Nopaline, l'acide nopalinic, agrocinopine A et B
Agropine	Agropine, mannopine, agropinic et l'acide mannopinic, agrocinopine C et D, leucinopine, leucinopine lactam, L,L succinamopine
Succinamopine	D,L succinamopine, succinamopine lactam, succinopine
Lippia	Agrocinopine C et D
Chrysopine/succinamopine	Chrysopine, deoxyfructosyl-5-oxoproline (dfop), L,L succinamopine, L,L leucinopine
Chrysopine/nopaline	Chrysopine, deoxyfructosyl-5-oxoproline (dfop), nopaline
Octopine/cucumopine	Octopine, cucumopine
Vitopine	Vitopine
Plasmids Ri	
Agropine	Agropine, mannopine, agropinic et l'acide mannopinique
Mannopine	Mannopine, agropinic et l'acide mannopinique
Cucumopine	Cucumopine, cucumopine lactam
Mikimopine	Mikimopine, mikimopine lactam

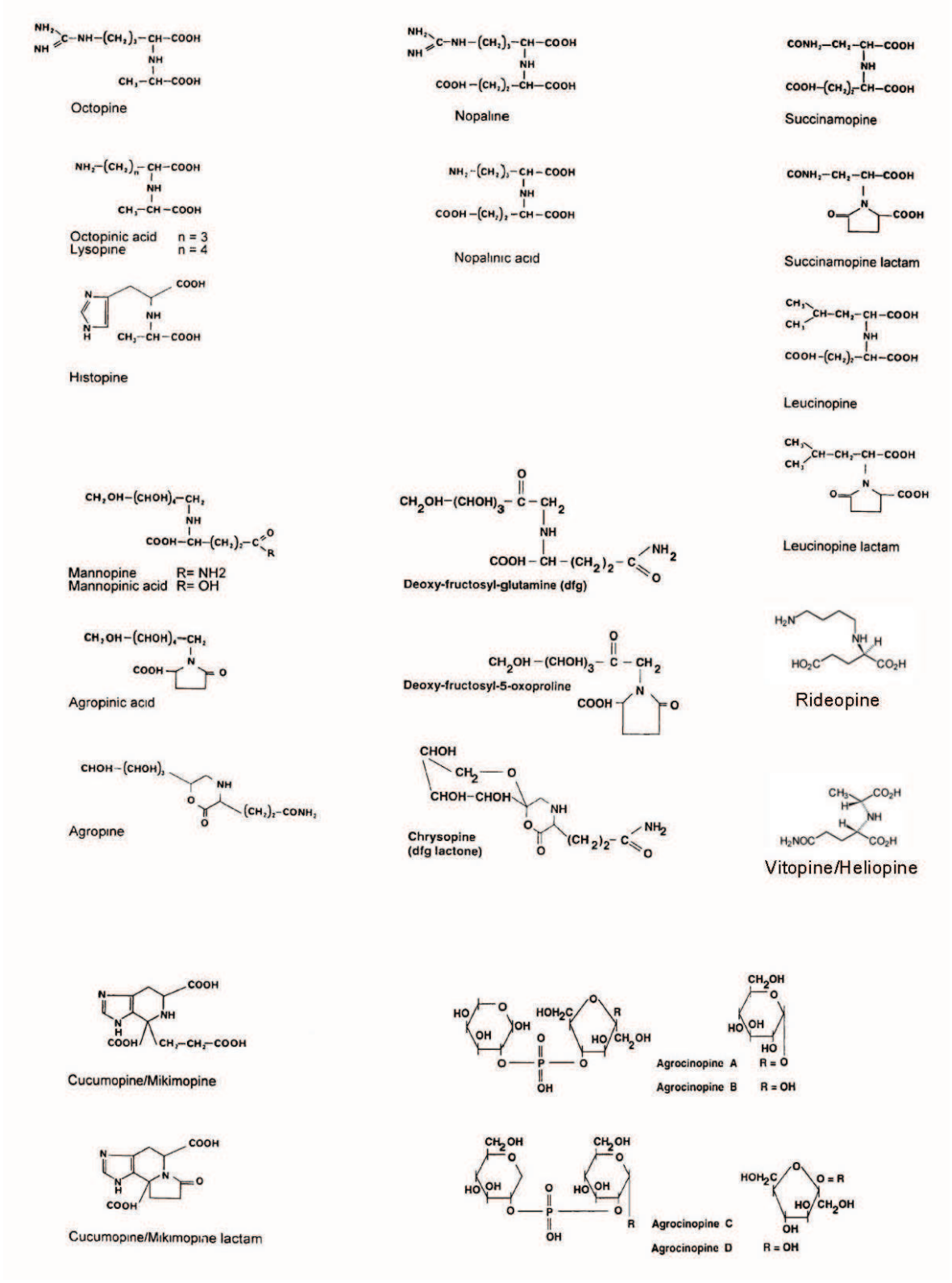


Figure 5. Les structures chimiques des opines produites par galles de la couronne et les chevelus racinaires

Les structures chimiques de octopine, l'acide octopinique, lysopine, histopine, nopaline, acide nopalinic, succinamopine, lactame succinamopine, leucinopine, lactame leucinopine, mannopine, l'acide mannopinique, désoxy fructosyl-glutamine (DFG), désoxy fructosyl-5-oxoproline, l'acide agroponic, agropine, chrysopine, cucumopine/mikimopine, cucumoplne/lactame miklmopine, agrocinopine A, B, C et D, rideopine, vitopine/heliopine (adapté de Dessaux et al., 1998).

2. Les gènes de synthèse d'hormones (des auxines et des cytokinines) qui stimulent la croissance des tissus infectés (figure 6).

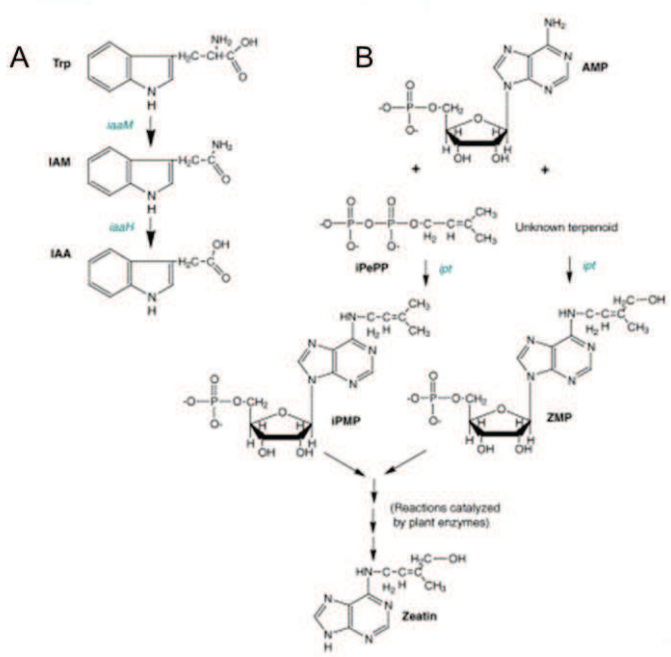


Figure 6. Fonctions enzymatiques des gènes *iaaM*, *iaaH* and *ipt*

3. Des gènes de type «plast» (définis par des homologies faibles au niveau des protéines correspondantes) qui ont une action encore mal connue sur la croissance des cellules végétales (figure 7 et tableau 2). Un de ces gènes plast, le gène *6b*, a été étudié plus particulièrement dans ce travail de Thèse. Ce gène provoque des modifications importantes de la croissance (désigné sous le nom de syndrome d'énation) transmissibles par greffe, avec une mode d'action encore inconnue.

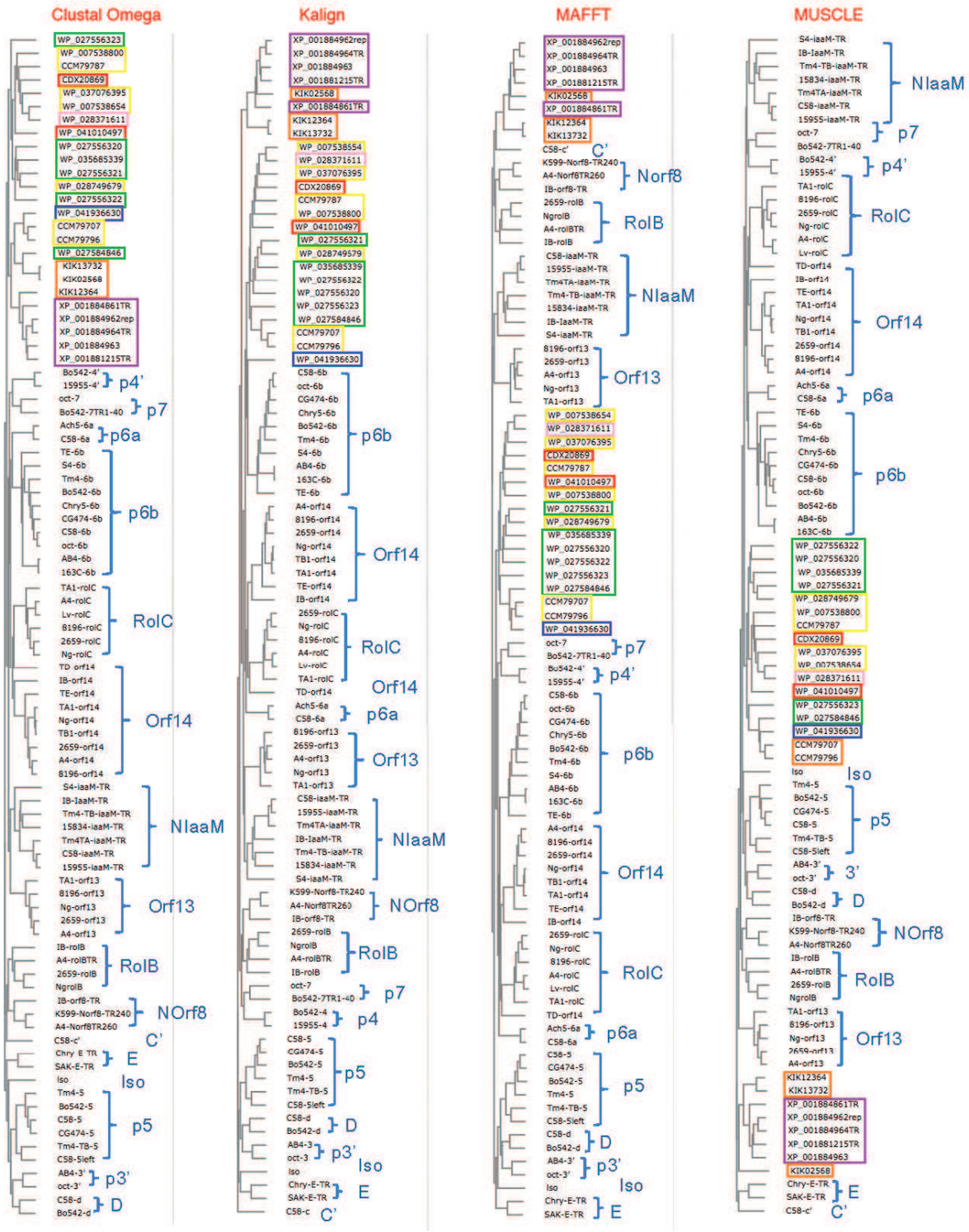


Figure 7. Les arbres phylogénétiques des protéines de type « plast »

Tableau 2. Les numéros d'accès correspondant aux protéines de type « plast » dans la figure 5
Ar: *A. rhizogenes*; At: *A. tumefaciens*; Av: *A.vitis*; Ng: *N. glauca*; Nt: *N. tabacum*; Lv: *Linaria vulgaris*.

	Protéines plasmatiques	Numéro d'accès	Origine		Protéines plasmatiques	Numéro d'accès	Origine
1	C58-5	AAD30487.1	At	34	oct-6b	AAF77126.1	At
2	SAK-5	BAA87806.1	At	35	Chry5-6b	AAB49454.1	At
3	CG474-5	AAB41867.1	Av	36	AB4-6b	CAA54541.1	Av
4	Tm4-5	AAB41873.1	Av	37	Bo542-6b	AAA98501.1	At

5	Bo542-5	AAZ50393.1	At	38	163C-6b	ADC97873.1	At
6	Tm4-TB-b	AAD30490.1	Av	39	S4-6b	AAA25043.1	Av
7	C58-b	AAD30482.1	At	40	A4-orf14	ABI54193.1	Ar
8	Iso	AAC25913.1	At	41	8196-orf14	AAA22099.1	Ar
9	Chry-E-TR	AAK08598.1	At	42	1724-orf14	BAA22339.1	Ar
10	SAK-E-TR	BAA87804.1	At	43	Ng-orf14	BAB85948.1	Ng
11	K599-Norf8-TR240	ABS11822.1	Ar	44	2659-orf14	CAB65899.1	Ar
12	A4-Norf8TR260	ABI54188.1	Ar	45	torf14	CBJ56561.1	At
13	K599-rolB	ABS11824.1	Ar	46	Ngorf14	BAA03991.1	Ng
14	2659-rolB	CAA82552.1	Ar	47	1724-rolC	P49408.1	Ar
15	1724-rolB	CAA45540.1	Ar	48	2659-rolC	CAA82553.1	Ar
16	A4-rolBTR	CAA34077.1	Ar	49	A4-rolC	P20403.1	Ar
17	NgrolB	CAA27161.1	Ng	50	Lv-rolC	ACD81987.1	Lv
18	C58-c’	AAD30484.1	At	51	Ng-rolC	P07051.2	Ng
19	C58-d	AAD30485.1	At	52	8196-rolC	AAA22096.1	Ar
20	Bo542-d	AAZ50418.1	At	53	trolC	CAA62988.1	Nt
21	AB4-3’	CAA54542.1	Av	54	Ach5-6a	P04030.1	At
22	696-3’	CAA52222.1	At	55	C58-6a	AAK90971.1	At
23	oct-3’	CAA25183.1	At	56	oct-7	AAF77121.1	At
24	Ng-orf13R	BAB85946.1	Ng	57	Bo542-7	AAZ50396.1	At
25	8196-orf13	AAA22097.1	Ar	58	15955-4’	CAA25180.1	At
26	2659-orf13	CAB65897.1	Ar	59	Bo542-4’	AAZ50416.1	At
27	A4-orf13	ABI54192.1	Ar	60	C58-iaaM	CAB44640.1	At
28	Ng-orf13	BAA03990.1	Ng	61	15955-iaaM	CAA25167.1	At
29	1724-orf13	BAA22337.1	Ar	62	Tm4TA-iaaM	P25017.1	Av
30	torf13-1	CAA07584.1	Nt	63	Tm4-TB-iaaM	AAD30493.1	Av
31	C58-6b	AAK90972.1	At	64	Ag162-iaaM	AAC77909.1	Av
32	CG474-6b	AAB41871.1	Av	65	15834-iaaM	ABI15642.1	Ar
33	Tm4-6b	CAA39648.1	Av	66	S4-iaaM	AAA98149.1	Av

4. Quelques gènes «orphelins» encore peu étudiés.

Il a été montré qu’Agrobacterium a la capacité de transformer durablement différentes espèces végétales, donnant naissance à des plantes naturellement transgéniques, notamment dans le genre Nicotiana (figure 6). Les séquences d’ADN transférées dans les espèces de ce groupe (appelées T-DNA cellulaires ou cT-DNAs) étaient encore très mal connues au moment où j’ai commencé ma Thèse.

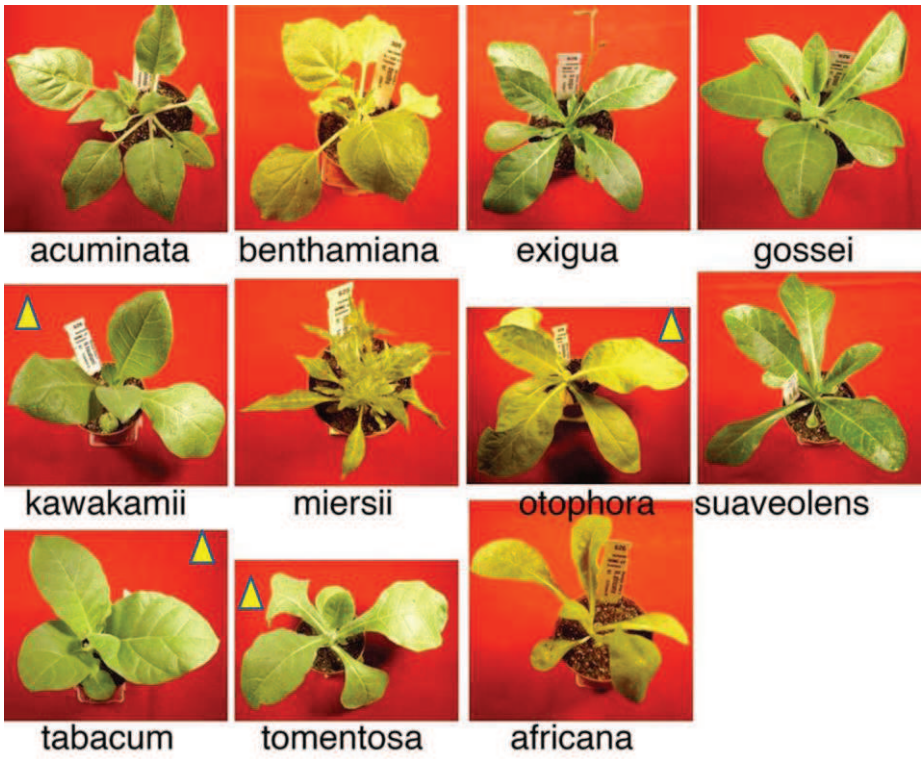


Figure 6. Espèces de Nicotiana, dont certains contiennent des séquences cT-DNA (marqué par des triangles jaunes)

Questions posées.

Nous avons voulu déterminer quelle était l'étendu des séquences cT-DNA transférées dans le genre Nicotiana par Agrobacterium. Si certains des gènes étaient intacts, une analyse de leur fonction éventuelle serait entreprise. Nous avons trouvé au cours de cette étude un gène TE-*6b* intact dans le *Nicotiana otophora*. Nous avons alors décidé d'approfondir l'analyse du phénotype induit par un gène *6b* modèle, le T-*6b* d'*Agrobacterium vitis*, afin de comparer de manière plus précise les activités du TE-*6b* et T-*6b*. Pour cela, nous avons utilisé des plantes de tabac transformées avec un gène T-*6b* sous contrôle d'un promoteur inductible au dexamethasone (dex-T-*6b*).

Approches expérimentales.

L'étude des séquences des cT-DNA dans le genre Nicotiana a été faite par séquençage à haut débit du génome de *Nicotiana tomentosiformis*, assemblage des fragments séquencés en contigs, identification des cT-DNA par bio-informatique, et vérification des cartes, très complexes à cause de la présence de séquences répétées, par PCR avec amorces spécifiques. L'analyse

de la transcription a été faite par PCR quantitative et RNA-seq. Les structures des cT-DNA d'autres espèces de Nicotiana ont été déterminées par PCR ou par analyse de séquences publiées par d'autres groupes, ce qui nous a permis de reconstruire l'évolution de ces séquences dans le groupe des Nicotianées. Les gènes TB-*mas2'* et TC-*ocl* ont été exprimés sous promoteur fort de façon transitoire dans le *N. benthamiana* afin d'étudier leur capacité à coder la production d'opines. Une étude du promoteur du gène TB-*mas2'* a été réalisée à l'aide d'un gène rapporteur GUS, et son introduction stable dans des plantes de *N. benthamiana* qui ont ensuite été analysées.

L'analyse des plantes dex-T-*6b* a été faite par coupes manuelles sériées de différentes parties de ces plantes (plusieurs milliers de coupes en tout, sur plusieurs centaines de plantes), coloration au bleu de toluidine et photographie. La présence de mRNA dans des plantes greffées a été analysée par qPCR.

Résultats

1. Analyse des séquences de type cT-DNA dans le genre Nicotiana. Nous avons identifié cinq cT-DNAs complets (TA à TE), distribuées d'une manière complexe dans différentes espèces de Nicotiana. Les cartes complètes ont été obtenues pour quatre d'entre eux. Nous avons pu établir que ces séquences proviennent de différentes infections et se sont accumulées au cours de l'évolution. Dans certaines lignées, des cT-DNAs ont été partiellement ou complètement délétés par des mécanismes encore inconnus. Nos résultats ont montré que l'évolution des cT-DNAs dans le genre Nicotiana est beaucoup plus complexe qu'attendue. Ces séquences contiennent des gènes intacts. Pour l'un d'entre eux, le gène TB-*mas2'*, nous avons montré qu'il code pour une enzyme active qui produit une molécule très similaire ou identique au désoxyfructosyl-glutamine (DFG), une opine bien connue. Ces résultats ont été publiés dans The Plant Journal (Chen et al., 2014).

2. Des résultats supplémentaires ont montré que le gène TB-*mas2'* est exprimé très fortement dans des racines de certains cultivars de tabac, et que ces racines produisent des quantités détectables d'une molécule de type DFG. C'est la première démonstration d'une modification métabolique dans une espèce végétale naturellement transformée par Agrobacterium (Chen et al., manuscrit en préparation).

3. Dans *N. otophora* nous avons trouvé un cT-DNA appelé région TE (non encore cartographiée). Ce cT-DNA contient un gène de type *6b*. Nous avons alors décidé d'étudier de manière plus détaillée les effets du gène T-*6b* d'*Agrobacterium vitis* déjà décrit, afin de le comparer au nouveau gène TE-*6b*. Le gène T-*6b* avait déjà été introduit

dans le tabac sous promoteur «dex» inductible, ce qui nous a permis d'observer ses effets morphologiques et anatomiques sur la croissance à différents stades après activation par le dexaméthasone. Ce travail a montré des effets inédits sur la croissance végétale ; l'apparition de méristèmes ectopiques (des ébauches foliaires et des systèmes vasculaires), et des développements morphogénétiques réguliers et reproductibles combinant la croissance normale et les méristèmes ectopiques. Plus de soixante modifications ont été identifiées et leurs relations réciproques ont été étudiées. En plus, nous avons pu démontrer que des plantes normales greffées sur des plantes T-*6b* induites contenaient des messagers du gène T-*6b*, ce qui suggère un mécanisme pour la transmission du phénotype *6b* par greffe à travers un transport à longue distance de ces messagers. Ce travail qui comporte un nombre exceptionnel d'illustrations (38 illustrations composites), a été publié dans Planta (Chen et Otten, 2015).

4. Le gène TE-*6b* a été cloné sous promoteur fort 35S et exprimé dans le tabac. Une quarantaine de plantes ont été régénérées et l'expression du gène TE-*6b* a été vérifiée par qPCR. Ces plantes ont un phénotype d'un type nouveau dont l'intensité est directement corrélée avec le taux d'expression du TE-*6b*. Le phénotype diffère de celui du gène T-6b, et affecte principalement les veines des feuilles. Une étude plus détaillée sera entreprise par la suite. Ce résultat est très encourageant, parce qu'il montre que le gène TE-*6b* code pour une fonction affectant la croissance, suggérant que la région TE a pu modifier la croissance de l'espèce ancestrale.

Conclusions et perspectives

Le séquençage complet, la cartographie des cT-DNAs et l'étude de leur expression dans le genre Nicotiana nous ont permis de jeter les bases d'une étude approfondie, s'intéressant aux fonctions de ces séquences. Nous avons déjà pu montrer que les gènes TB-*mas2'* et TE-*6b* possèdent une activité biologique. Les conséquences de ces activités sur les plantes qui les expriment restent à déterminer. Ceci pourrait être fait par leur excision utilisant la technique CRISPR-Cas9, ou par «RNA silencing». Une analyse comparative très détaillée des modifications morphologiques induites par les gènes T-*6b* et TE-*6b* et la construction de gènes hybrides pourrait révéler le mécanisme sous-jacent commun du gène *6b*. Il faudrait tester également si l'effet du gène TE-*6b* est transmissible par greffe comme c'est le cas pour le gène T-*6b*.

Sequencing and functional analysis of cT-DNAs in Nicotiana

Summary:

The bacterium *Agrobacterium tumefaciens* is well-known for its utilisation in plant genetic engineering where it serves as a gene vector. This bacterium and the related species *Agrobacterium rhizogenes* are phytopathogens that induce tumors and hairy roots respectively on susceptible plants like grapevine or fruit trees. Their phytopathogenicity is due to horizontal transfer of bacterial genes to the plant host, from a plasmid called the Ti (tumor-inducing) or Ri (root-inducing) plasmid. The subject of my Thesis concerns two particular aspects of this bacterium.

1. Their capacity to stably transform several plant species in nature, thereby yielding naturally transformed plants, especially in the genus Nicotiana. We have shown by deep sequencing of the *Nicotiana tomentosiformis* genome and by analysis of other recently published Nicotiana sequences the presence of five different Agrobacterium-derived sequences (cT-DNAs), totalling 65 kb, some of which carry intact genes. We have shown that two of them (TB-*mas2'* from *N. tabacum* and TE-*6b* from *N. otophora*) have biological activity. A detailed comparative study has allowed us to better understand the evolution of these cT-DNAs (Chen et al., 2014). The *mas2'* gene is well-known, it codes for the synthesis of desoxyfructosyl-glutamine (DFG) in tumors or roots induced by Agrobacterium. Recent work in our group has shown that the TB-*mas2'* gene is highly expressed in some *N. tabacum* cultivars and leads to the accumulation of detectable amounts of DFG. This work is presented as a manuscript to be submitted.

2. A second part of the Thesis describes new properties of the T-*6b* gene, which is part of the DNA transferred by *A. vitis* strain Tm4 and leads to abnormal growth caracterized by the appearance of enations, so far the mode of action of this gene is unknown. The *6b* gene is part of the so called *plast* family (for phenotypic plasticity), with different and often remarkable growth effects on plants. The T-*6b* gene wasearlier placed under control of a dexamethasone-inducible promoter, and tobacco plants transformed with this construct have now been studied in detail, at different times after the start of induction. A large number of changes was analyzed, both at the morphological and anatomical level, these include various unprecedented morphological changes, like for example the appearance of shoot primordia at the base of trichomes, or the appearance of ectopic vascular strands parallel to the normal strands with a regular development leading to complex but predictable structures (Chen and Otten, 2015). The TE-*6b* gene from *N. otophora* was placed under strong and constitutive promoter control and introduced into tobacco, where it was found to cause new types of morphological change, different from those observed for T-*6b*. The latter results are preliminary and will be presented as a complement to the work on T-*6b*. They indicate that the introduction of the TE-*6b* gene in the *N. otophora* ancestor could have caused a change in growth pattern, and might have favored the appearance of a new species.

Keywords: Agrobacterium, *N. tomentosiformis*, cT-DNA, natural transformation, horizontal gene transfer, *6b*, leaf polarity, *plast* genes