

METABOLIC ENGINEERING OF PLANTS USING A DISARMED POTYVIRUS VECTOR

Eszter Majer

Director: Dr. José Antonio Daròs Arnau
Valencia, July 2016



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Szüleimnek



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

**METABOLIC ENGINEERING OF PLANTS USING A
DISARMED POTYVIRUS VECTOR**

A Thesis presented by

Eszter Majer

to obtain the degree of

Doctor of Philosophy in Biotechnology

Universitat Politècnica de València, July 2016

Director: **Dr. José Antonio Daròs Arnau**



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



El Dr. José Antonio Daròs Arnau, Investigador Científico del Consejo Superior de Investigaciones Científicas (CSIC), en el Instituto de Biología Molecular y Celular de Plantas (IBMCP), centro mixto del CSIC y la Universitat Politècnica de València

CERTIFICA:

Que la Ingeniera en Horticultura ESZTER MAJER ha realizado bajo su dirección en el IBMCP el trabajo que lleva por título “Metabolic engineering of plants using a disarmed potyvirus vector”, y autoriza su presentación para optar al grado de Doctor.

Y para que así conste, expide y firma el presente certificado en Valencia, el 30 de Mayo 2016.

Dr. José Antonio Daròs Arnau

Table of content

Summaries	1
Summary	3
Resumen.....	5
Resum.....	7
Abbreviations	11
Introduction	17
Plant biotechnology and the green factory.....	19
Plant metabolic engineering	30
Reporter genes for tracking viral infection in plants.....	39
Objectives	43
Chapter I	47
Abstract	51
Research article	51
References	60
Supporting information	63
Chapter II	69
Abstract	73
Introduction.....	73
Materials and methods	75
Results	80

Discussion	86
References	91
Supporting information	97
Chapter III	109
Abstract	113
Research article	113
Results	116
Discussion	123
Methods.....	125
References	128
Supporting information	134
General Discussion	143
Conclusions	153
Bibliography	157
Acknowledgements	187

Summaries

Summary

Plant viruses are obligate intracellular parasites which have the ability to make host cells to synthesize high amounts of viral proteins. This capacity was exploited to develop recombinant plant virus vectors to express heterologous proteins in plant biofactories. There are also some examples of viral vectors which were able to modify endogenous metabolic pathways of natural products in plants by silencing certain genes or expressing heterologous transcription factors and metabolic enzymes. The main limitation of many plant virus-based systems is the difficulty to co-express various heterologous proteins in the same cell with proper subcellular localization, which is a crucial question in metabolic engineering, since these pathways are composed of many enzymes located in various subcellular organelles.

This work provides a solution to overcome this problem by using a potyvirus-based vector system. Potyviruses (genus *Potyvirus*, family *Potyviridae*) are plus-strand single-stranded RNA viruses, which have a genome expression strategy that allows the equimolar production of most viral proteins. The genome encodes a main polyprotein, which is processed by viral proteases into about ten mature proteins. cDNAs coding for foreign proteins can be inserted into the viral genome flanked by sequences corresponding to processing sites of the viral proteases, which release them during polyprotein processing. On the basis of an infectious clone of *Tobacco etch virus* (TEV), Bedoya *et al.* (2010) developed an expression system in which the RNA-dependent RNA polymerase (NIb) gene was replaced by an expression cassette, harboring several heterologous proteins. This viral vector was able to express three fluorescent proteins with nucleocytoplasmic localization in equimolar amounts in transgenic tobacco plants in which NIb was supplemented in *trans*.

Despite of the apparent simplicity of potyvirus genome expression strategy, foreign cDNA insertion is a complicated task since inserted sequences can interrupt some viral functions which may influence negatively the virus viability. Thus, our first goal was to analyze the effect of gene insertion on TEV genome stability by relocating the NIb cistron to all possible intergenic positions of the genome and test the viability of chimeric viruses. As a result of this work, a novel insertion position was discovered at the amino-terminal end of the potyvirus polyprotein, which opened the

possibility to explore new questions of recombinant protein expression using the TEV-based vector.

Since metabolic pathways are highly compartmentalized, proper subcellular targeting of enzymes is an essential task. Thus, our second objective centralized on the subcellular targeting of expressed proteins from the TEV-based viral vector. cDNAs coding for the green fluorescent protein (GFP) fused to chloroplast, nucleus and mitochondria targeting signal sequences were inserted into the newly described amino-terminal insertion position or into an internal site, replacing the NIB cistron. Our results showed that for protein delivery to chloroplasts and mitochondria, foreign genes have to be inserted at the amino-terminal site of the viral vector, but for nuclear delivery, both insertion positions are suitable. Our results also demonstrated that the insertion site affects recombinant protein accumulation and virus stability. The amino-terminal site proved to be more suitable to produce higher amounts of targeted recombinant proteins but the internal insertion site showed to be more stable. Based on these findings, we were able to launch two different fluorescent proteins into chloroplasts and nuclei from a single viral vector.

The last objective of this work was to investigate whether the potyvirus-based vector was able to express an entire heterologous multistep biosynthetic pathway in plant cells. For this aim we purposed to produce lycopene, a plant pigment with health promoting properties, which is naturally synthesized inside the chloroplasts and chromoplasts from the isoprenoid precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), through the methylerythriol 4-phosphate (MEP) pathway. However, these precursor molecules are also synthesized outside the plastids through the mevalonate (MVA) pathway. To produce lycopene we inserted cDNAs coding for the enzymes of a three-step metabolic pathway of bacterial origin into the potyvirus-based vector which was supposed to convert cytosolic IPP and DMAPP into lycopene. Infected tobacco plants developed orange symptoms indicating lycopene accumulation, which was confirmed by high-performance liquid chromatography (HPLC) analysis and microscopy observations. Our results also illustrated that the phytoene synthase is the bottleneck of the carotenoid biosynthetic pathway and the sole expression of *Pantoea ananatis* phytoene synthase, crtB, is enough to induce carotenoid accumulation, conferring yellow coloration to the infected tissue.

CrtB was tested as reporter system to visually track viral infection in several plant species of both scientific and agronomical importance. CrtB resulted to be a more general marker than Rosea1, a previously described red marker based on anthocyanin accumulation, since it was functional in plant species where Rosea1 failed to detect virus presence. In addition, crtB and Rosea1 together were able to report two virus populations at the same time with two different, yellow and red, output colors.

Resumen

Los virus de plantas son parásitos intracelulares obligados que tienen la capacidad de forzar las células del huésped a sintetizar cantidades importantes de proteínas virales. Esta capacidad de los virus de plantas ha sido utilizada para desarrollar vectores virales y expresar proteínas heterólogas en plantas biofactoria. Existen algunos ejemplos de vectores virales que han demostrado una capacidad para modificar rutas metabólicas endógenas de productos naturales silenciando ciertos genes o expresando factores de transcripción y enzimas metabólicas. La principal limitación de muchos sistemas basados en virus de plantas es la dificultad de coexpresar diversas proteínas heterólogas en la misma célula con la localización subcelular apropiada, lo cual es una cuestión crucial en ingeniería metabólica, ya que las rutas metabólicas se componen de muchas enzimas localizadas en varios orgánulos subcelulares.

Este trabajo presenta una solución para superar este problema mediante el uso de un vector viral basado en un potyvirus. Los potyvirus (género *Potyvirus*, familia *Potyviridae*) son virus de RNA de cadena positiva simple que tienen una estrategia de expresión génica que permite la producción de la mayoría de las proteínas virales en cantidades equimolares. El genoma codifica una poliproteína principal, que es procesada por las proteasas virales en alrededor de diez proteínas maduras. Los cDNAs de las proteínas heterólogas se pueden insertar en el genoma viral flanqueados de secuencias correspondientes a sitios de procesamiento proteolítico de las proteasas virales, que liberan las proteínas insertadas durante el procesamiento de la poliproteína. Basado en un clon infeccioso del virus del grabado del tabaco (*Tobacco etch virus*, TEV) Bedoya *et al.* (2010) desarrollaron un sistema de expresión en el que el gen de la RNA polimerasa dependiente de RNA (N1b) fue sustituido por un casete de expresión, que albergaba varias proteínas heterólogas. Este vector viral fue capaz de expresar tres

proteínas fluorescentes con localización nucleocitoplásmica en cantidades equimolares en plantas de tabaco transgénicas que complementaban el cistron N1b en *trans*.

A pesar de la aparente simplicidad de la estrategia de expresión génica de los potyvirus, la inserción de un cDNA foráneo es una tarea complicada, ya que estas secuencias pueden interrumpir algunas funciones virales que influyen negativamente en la viabilidad del virus recombinante. Por lo tanto, nuestro primer objetivo fue analizar el efecto de la inserción en la estabilidad del genoma de TEV mediante la reubicación del cistron N1b a todas las posibles posiciones intercistónicas del genoma y ensayar la viabilidad de los virus quiméricos. Como resultado de este trabajo, descubrimos una nueva posición de inserción en el extremo amino-terminal de la poliproteína viral que nos permitió explorar otras cuestiones sobre la expresión de proteínas recombinantes utilizando el vector viral.

Dado que las vías metabólicas son muy compartimentalizadas, la adecuada localización subcelular de enzimas es una tarea esencial en ingeniería metabólica. Por eso, nuestro segundo objetivo se centró en la distribución de las proteínas heterólogas expresadas con el vector viral a diferentes orgánulos subcelulares. cDNAs que codificaban la proteína fluorescente verde (green fluorescent protein, GFP) fusionada a péptidos señal se insertaron en la nueva posición amino-terminal y en un sitio interno, sustituyendo el cistron N1b, para enviarla al cloroplasto, núcleo y a la mitocondria. Nuestros resultados mostraron que para la distribución de proteínas al cloroplasto y mitocondria, los genes foráneos deben ser insertados en el sitio amino-terminal del vector viral, pero para la distribución nuclear, ambas posiciones son adecuadas. Demostramos también que el sitio de inserción utilizado afecta la acumulación de las proteínas recombinantes y la estabilidad del vector viral. El sitio amino-terminal resultó ser más adecuado para producir cantidades más grandes de proteínas recombinantes, pero el sitio de inserción interno demostró ser más estable. En base a estos resultados, hemos sido capaces de distribuir dos proteínas fluorescentes distintas a los cloroplastos y núcleos desde un único vector viral.

El último objetivo de este trabajo fue estudiar si el vector viral basado en potyvirus es capaz de expresar una ruta biosintética de múltiples pasos en células vegetales. Para ello nos propusimos producir licopeno, un pigmento vegetal con propiedades beneficiosas para la salud humana, que se sintetiza en los cloroplastos y

cromoplastos desde los precursores de isoprenoides, el pirofosfato de isopentenilo (isopentenyl diphosphate, IPP) y el pirofosfato de dimetilalilo (dimethylallyl diphosphate, DMAPP) a través de la ruta de metileritriol 4-fosfato (methylerythriol 4-phosphate, MEP). Sin embargo, estos precursores también se sintetizan fuera de los plastos a través de la ruta mevalonato (mevalonate, MVA). Para producir licopeno insertamos un cDNA que codificaba las enzimas de una ruta metabólica de tres pasos de origen bacteriano en el vector viral que debía convertir IPP y DMAPP citosólico en licopeno. Las plantas de tabaco infectadas con el vector viral desarrollaron síntomas de color naranja indicando la acumulación de licopeno, que fue confirmado por análisis de cromatografía líquida de alta eficacia (high-performance liquid chromatography, HPLC) y observaciones de microscopía. Nuestros resultados también ilustraron el papel central del fitoeno sintasa en la ruta biosintética de los carotenoides y se demostró que la sola expresión de la fitoeno sintasa de *Pantonea ananatis*, crtB, es suficiente para inducir la acumulación de carotenoides que confieren una coloración amarilla al tejido infectado.

Ensayamos crtB como sistema reportero visual en varias especies de plantas de importancia tanto científica como agronómica. CrtB resultó ser un marcador visual más general que Rosea1, el marcador rojo basado en la acumulación de antocianinas descrito anteriormente, ya que fue funcional en especies de plantas en las que Rosea1 fue incapaz de detectar la presencia viral. Además, los marcadores crtB y Rosea1 en conjunto fueron utilizados con éxito para seguir con dos colores distintos, amarillo y rojo, dos poblaciones virales al mismo tiempo.

Resum

Els virus de plantes són paràsits intracel·lulars obligats que tenen la capacitat de forçar les cèl·lules de l'hoste a sintetitzar quantitats importants de proteïnes virals. Aquesta capacitat dels virus de plantes ha sigut utilitzada per a desenvolupar vectors virals i expressar proteïnes heteròlogues en plantes biofactoria. Existeixen alguns exemples de vectors virals que han demostrat una capacitat per a modificar rutes metabòliques endògenes de productes naturals silenciament certs gens o expressant factors de transcripció i enzims metabòlics. La principal limitació de molts sistemes basats en virus de plantes és la dificultat de coexpressar diverses proteïnes heteròlogues en la mateixa cèl·lula amb la localització subcel·lular apropiada, cosa que és una qüestió

crucial en enginyeria metabòlica, ja que les rutes metabòliques es componen de molts enzims localitzats en diversos orgànuls subcel·lulars.

Aquest treball presenta una solució per a superar aquest problema mitjançant l'ús d'un vector viral basat en un potyvirus. Els potyvirus (gènere *Potyvirus*, família *Potyviridae*) són virus d'RNA de cadena positiva simple que tenen una estratègia d'expressió gènica que permet la producció de la majoria de les proteïnes virals en quantitats equimolars. El genoma codifica una poliproteïna principal, que és processada per les proteases virals en al voltant de deu proteïnes madures. Els cDNAs de les proteïnes heteròlogues es poden inserir en el genoma viral flanquejats de seqüències corresponents a llocs de processament proteolític de les proteases virals, que alliberen les proteïnes inserides durant el processament de la poliproteïna. Basat en un clon infecciós del virus del gravat del tabac (*Tobacco etch virus*, TEV) Bedoya *et al.* (2010) van desenvolupar un sistema d'expressió en el qual el gen de l'RNA polimerasa dependent d'RNA (NIb) va ser substituït per un casset d'expressió, que albergava diverses proteïnes heteròlogues. Aquest vector viral va ser capaç d'expressar tres proteïnes fluorescents amb localització nucleocitoplàsmica en quantitats equimolars en plantes de tabac transgèniques que complementaven el cistró NIb en *trans*.

Malgrat l'aparent simplicitat de l'estratègia d'expressió gènica dels potyvirus, la inserció d'un cDNA forà és una tasca complicada, ja que aquestes seqüències poden interrompre algunes funcions virals que influeixen negativament en la viabilitat del virus recombinant. Per tant, el nostre primer objectiu va ser analitzar l'efecte de la inserció en l'estabilitat del genoma de TEV mitjançant la reubicació del cistró NIb a totes les possibles posicions intercistròniques del genoma i assajar la viabilitat dels virus quimèrics. Com a resultat d'aquest treball, hem descobert una nova posició d'inserció en l'extrem amino terminal de la poliproteïna viral que ens va permetre explorar altres qüestions sobre l'expressió de proteïnes recombinants utilitzant el vector viral.

Atès que les vies metabòliques són molt compartimentalitzades, l'adequada localització subcel·lular d'enzims és una tasca essencial en enginyeria metabòlica. Per açò, el nostre segon objectiu es va centrar en la distribució de les proteïnes heteròlogues expressades amb el vector viral a diferents orgànuls subcel·lulars. cDNAs que codificaven la proteïna fluorescent verda (green fluorescent protein, GFP) fusionada a pèptids senyal es van inserir en la nova posició amino terminal i en un lloc intern,

substituint el cistró N1b, per a enviar-la al cloroplast, nucli i al mitocondri. Els nostres resultats van mostrar que per a la distribució de proteïnes al cloroplast i mitocondri, els gens forans han de ser inserits en el lloc amino terminal del vector viral, però per a la distribució nuclear, ambdues posicions són adequades. Vam demostrar també que el lloc d'inserció utilitzat afecta l'acumulació de les proteïnes recombinants i l'estabilitat del vector viral. El lloc amino terminal va resultar ser més adequat per a produir quantitats més grans de proteïnes recombinants, però el lloc d'inserció intern va demostrar ser més estable. Sobre la base d'aquests resultats, hem sigut capaços de distribuir dues proteïnes fluorescentes diferents als cloroplasts i nuclis des d'un únic vector viral.

L'últim objectiu d'aquest treball va ser estudiar si el vector viral basat en potyvirus és capaç d'expressar una ruta biosintètica de múltiples passos en cèl·lules vegetals. Per açò ens vam proposar produir licopè, un pigment vegetal amb propietats beneficioses per a la salut humana, que se sintetitza en els cloroplasts i cromoplasts des dels precursors de isoprenoides, el pirofosfat d'isopentil (isopentenyl diphosphate, IPP) i el pirofosfat de dimetilal·lil (dimethylallyl diphosphate, DMAPP) a través de la ruta de metileritriol 4-fosfat (methylerythriol 4-phosphate, MEP). No obstant açò, aquests precursors també es sintetitzen fora dels plastidis a través de la ruta mevalonat (mevalonate, MVA). Per produir licopè inserírem un cDNA que codificava els tres enzims de una ruta metabòlica de tres passos d'origen bacterià en el vector viral que havia de convertir IPP i DMAPP citosòlic en licopè. Les plantes de tabac infectades amb el vector viral van desenvolupar símptomes de color taronja indicant l'acumulació de licopè, que va ser confirmat per anàlisi de cromatografia líquida d'alta eficàcia (high-performance liquid chromatography, HPLC) i observacions de microscòpia. Els nostres resultats també van il·lustrar el paper central del fitoè sintasa en la ruta biosintètica dels carotenoides i es va demostrar que la sola expressió de fitoè sintasa de *Pantonea ananatis*, crtB, és suficient per a induir l'acumulació de carotenoides que confereixen una coloració groga al teixit infectat.

Assajarem crtB com a sistema reporter visual en diverses espècies de plantes d'importància tant científica com agronòmica. CrtB va resultar ser un marcador visual més general que Rosea1, el marcador roig basat en l'acumulació de antocianines descrit anteriorment, ja que va ser funcional en espècies de plantes en les quals Rosea1 va ser incapaç de detectar la presència viral. A més, els marcadors crtB i Rosea1 en conjunt

van ser utilitzats amb èxit per a seguir amb dos colors diferents, groc i roig, dues poblacions virals al mateix temps.

Abbreviations

AMV, *Alfalfa mosaic virus*

BeYDV, *Bean yellow dwarf virus*

bHLH, basic helix-loop-helix

bp, base pairs

CaMV, *Cauliflower mosaic virus*

cDNA, complementary DNA

CLSM, confocal laser scanning microscopy

CP, coat protein

CPMV, *Cowpea mosaic virus*

CRTISO, cis-trans-isomerase

cTP, chloroplast transit peptide

DMAPP, dimethylallyl diphosphate

DNA, deoxyribonucleic acid

dpi, days post-inoculation

GDP, guanosine diphosphate

GFP, green fluorescent protein

GGPP, geranylgeranyl diphosphate

GGPPS, geranylgeranyl diphosphate synthase

GOI, gene of interest

GTP, guanosine triphosphate

GUS, β -glucuronidase

HPLC, high-performance liquid chromatography

Hsp70, heat shock protein-70

Hsp90, heat shock protein-90

HYD, ferredoxin-dependent nonheme diiron hydroxylase

IPP, isopentenyl diphosphate

IRES, internal ribosome entry site

LIR, long intergenic regions in BeYDV genome
LYCb, lycopene β -cyclase
LYCe, lycopene ϵ -cyclase
mAb, monoclonal antibody
MEP, methylerythriol 4-phosphate pathway
MP, movement protein
MPP, mitochondria-processing peptidase
mRNA, messenger RNA
mTP, mitochondrial targeting peptide
MVA, mevalonate pathway
NIaPro, nuclear inclusion *a* protease
NIb, nuclear inclusion *b*
NLS, nuclear localization signal
NPC, nuclear pore complexes
NSY, neoxantin synthase
nt, nucleotide
ORF, open reading frame
PAGE, polyacrylamide gel electrophoresis
PDS, phytoene desaturase
PRE, presequence
PSY, phytoene synthase
PVX, *Potato virus X*
RdRp, RNA-dependent RNA polymerase
RNA, ribonucleic acid
RT, reverse transcription
sgRNA, subgenomic RNA
SIR, short intergenic regions in BeYDV genome

SPP, processing peptidase
ssRNA+, single stranded positive-sense RNA
SV40, Simian virus 40
T-DNA, transferred DNA
TEV, *Tobacco etch virus*
TIC, translocon on the inner chloroplast membrane
TIM, translocase of the inner membrane
TMV, *Tobacco mosaic virus*
TOC, translocon on the outer chloroplast membrane
TOM, translocase of the outer membrane
TP, transit peptide
TVCV, *Turnip vein clearing virus*
UTR, untranslated region
VDE, violaxanthin de-epoxidase
VLPs, virus-like particles
VNPs, viral nanoparticles
VPg, viral protein genome-linked
ZDS, ζ -carotene desaturase
ZEP, zeaxanthin epoxidase
 Δ NIb, deleted nuclear inclusion *b*

Introduction

Plant biotechnology and the green factory

Plants form an integral part of human life, serving as food, feedstock, building material, fuel and source of biologically-active substances for medicinal use. Therefore it is not surprising that over thousands of years humankind tried to improve plants with different desired properties. In the past decades, a variety of techniques commonly known as plant biotechnology was successfully used to develop crops by introducing beneficial traits in plants, focusing on three main objectives: (1) agronomic traits (herbicide tolerance, insect resistance, improvements on additional health benefits of edible parts, abiotic stress tolerance, nutrient use efficiency and yield potential); (2) bioenergy and (3) production of valuable proteins, also known as ‘molecular farming’ (Xu *et al.*, 2012; Chen and Lin, 2013).

Protein expression in living organisms is based on a highly-conserved process: the DNA is transcribed into mRNA which is translated into polypeptides built up from amino acids. The consequence of the universality of this process is that in theory, any organism can be used as platform for recombinant protein production and depending on the characteristics of the expressed protein, the most suitable platform can be selected (Melnik and Stoger, 2013). In the last decades, the high demand of pure, properly folded and biologically active proteins stimulated the development of more productive and suitable platforms. Nowadays, recombinant protein production relies on ‘bio-factories’ of microbial (*Escherichia coli*, *Saccharomyces cerevisiae*, *Pichia pastoris*) or mammalian cell cultures (Chinese hamster ovary cells) and in minor percentage insect cells.

Plants offered an exciting opportunity to develop edible vaccines by introducing genes of bioactive proteins in their genome and combining the benefit of the worldwide established low-cost production of plants with vaccine manufacturing (Mason *et al.*, 2002). Thanks to the intensive research in this field, nowadays plants are used as alternative platforms for heterologous protein production. The advantage of plant cells over prokaryotes is that they possess an endomembrane system and a secretory pathway which helps correct assembly and folding of complex proteins, and they carrying out post-translational modification, very similar to mammalian cells (Vitale and Denecke, 1999). Additionally, plant based platforms remain free from contaminants, like animal pathogenic viruses, prions and bacterial endotoxins (Faye and Gomord, 2010).

One of the most commonly used techniques for heterologous protein expression in plants is based on *Agrobacterium tumefaciens*. This soil-borne bacterium has the capacity to introduce genes into the plant chromosome (Chilton *et al.*, 1977). Replacing the bacterial genes with genes of interest under the control of specific promoters, heterologous protein expression can be achieved (Herrera-Estrella *et al.*, 1983; Zambryski *et al.*, 1983)

The two main methods of heterologous protein expression in plants are stable transformation of nuclear or chloroplast genome, and transient expression. The main difference is that in stable transformation, the foreign gene is integrated into the nuclear or chloroplast genome and therefore, becomes hereditary, while in transient systems the recombinant protein expression is limited to a short period of time and does not affect all plant cells, but in terms of speed and yield, transient systems overcome stable transformation. Additionally, transiently expressed genes avoid transgene flow, the most important inconvenience of stable transgenic plants, because they are not inherited.

The most successful transient expression systems are based on plant viruses which serve as vehicles for heterologous protein production or provide building blocks for combined expression vectors.

Plant virus based vectors for heterologous protein expression

Plant viruses are able to redirect the biosynthetic capacity of the host cell to produce new viruses in a short period of time by abducting host ribosomes and forcing them to synthesize high amount of viral proteins (Nelson and Citovsky, 2005). In recent years, several plant virus based vectors were developed as engineered vehicles for transient expression of heterologous proteins by introducing the gene of interest (GOI) in the viral genome, or using viral genes as building blocks to develop expression vectors, offering a rapid and easy way to produce high quantities of foreign proteins in plant biofactories.

First, viruses with DNA genomes, like those from the family *Caulimoviridae* (Szeto *et al.*, 1977; Hull, 1978) were proposed to be used as viral vectors in plants, but soon new techniques enabling generation of cDNA from an RNA template and *in vitro* transcription of this cDNA to infectious RNA, made possible the genetic manipulation and *in vivo* studies of plant viruses with RNA genome (Ahlquist and Janda, 1984; Ahlquist *et al.*, 1984). Another technical breakthrough was the use of *A. tumefaciens*, as

an alternative way to introduce the viral genome into the plant cell avoiding the inefficient mechanical inoculation of RNA transcripts (Grimsley *et al.*, 1986). This technique, called ‘agroinfection’, consists in the transient expression of the viral genome with *A. tumefaciens*, which results in an autonomously replicating nucleic acid sequence after host cell nuclear transcription and processing, and enables the virus to spread out from the initially infected cells and invade systemic leaves. Thanks to the basic research on plant virus biology, during the past three decades, many expression systems based on plant viruses were designed with several gene expressing strategies, exploiting the possibilities offered by the parental viral genome. Gene expression and peptide-presentation systems, as well as systems to produce empty capsids, were developed (Figure 1).

For gene expression of heterologous proteins with viral vectors, scientists firstly used ‘full virus vectors’ containing the whole viral genome where the heterologous sequence encoding the GOI was inserted (Gleba *et al.*, 2004). These viral vectors behaved like wild-type viruses retaining genes for normal infection and replication, including the capacity to systemic movement and suppression of gene silencing of host plants. The most limiting factor of these systems was the size of the inserted sequence because of the negative correlation between the size of the GOI and vector stability which resulted in lower productivity of larger heterologous proteins (Avesani *et al.*, 2007). To overcome this limitation, plant virologists started to manipulate the viral genome and ‘rebuild’ viral vectors. At this point the main objective in vector design was to maximize heterologous protein yield by maintaining essential elements for protein expression and deleting or providing in *trans* some of the non-essential functions leading to ‘deconstructed’ vectors (Gleba *et al.*, 2007, 2005). These efforts resulted in a wide spectrum of gene expression vectors with robust replication and amplification capacity and with abolished systemic movement which in most cases was replaced by *A. tumefaciens*-mediated replicon delivery (Peyret and Lomonossoff, 2015).

The repetitive geometry of the plant viral capsid made up from several coat-protein molecules stimulated scientists to use capsids as epitope presentation systems and later on as empty vessels for nanotechnology purposes. The capsid of plant viruses serves to enclose genetic material destined to protect and help cell-to-cell and systemic movement through the plant. First, plant viral capsid was decorated with small antigenic peptides, called epitopes, against human or animal diseases, which enhanced the im-

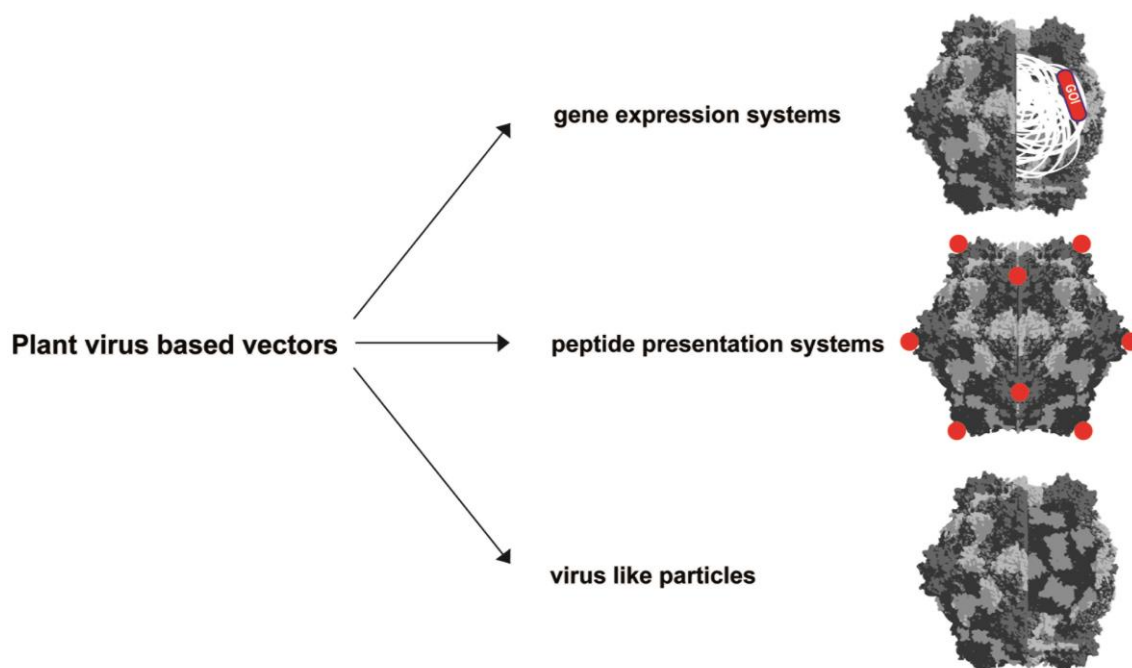


Figure 1. Plant virus based expression vectors. In gene expression systems the GOI is inserted into the viral genome. In peptide presentation systems the viral capsid is engineered to hold several molecules of interest. Virus like particles are empty capsids, which can be decorated on their surface or loaded with molecules of interest.

munogenicity of such peptides (Bachmann and Jennings, 2010). These findings resulted in the generation and use of peptide-presentation systems or viral nanoparticles (VNPs) for vaccine delivery. Nevertheless, these chimeric capsids contained the viral genomic RNA and remained infectious which raised biosafety issues. Further progresses were made when empty capsid formation was achieved in absence of viral genome. Empty capsids, also called virus-like particles (VLPs) provide an ideal basis of targeted drug delivery vehicles or tissue specific imaging reagents since they can be easily modified by genetic or chemical engineering. Furthermore plant VLPs and VNPs are biodegradable and are considered non-hazardous in humans and other animals (Kaiser *et al.*, 2007; Yildiz *et al.*, 2011), and VLPs can also be loaded with inorganic material for chemical or bionanotechnological applications (Evans, 2008; Pokorski and Steinmetz, 2011).

From the many viruses used as backbones to develop virus-based expression systems, here, the most successful plant virus-based vectors are summarized briefly with special attention on gene expression systems (Peyret and Lomonosoff, 2015; Pogue *et al.*, 2002; Gleba *et al.*, 2014; Salazar-González *et al.*, 2015). Probably the most

famous ‘deconstructed’ gene expression vector based on a plant virus is magnICON, a hybrid system with elements from *Tobacco mosaic virus* (TMV) and *Turnip vein clearing virus* (TVCV), both belonging to the genus *Tobamovirus*, family *Virgaviridae*, with rigid rod shaped virions and a single stranded positive-sense RNA (ssRNA+) genome (Figure 2A). The viral genome encodes for 4 proteins. The first two are involved in replication and are translated from the main RNA. The last two are translated from two separated subgenomic RNAs (sgRNAs) and encode the movement protein (MP) and the coat protein (CP). The tobamovirus-based magnICON system is composed by three modules: the 5’ module contains building blocks from two tobamoviruses and a promoter for GOI expression; the 3’ module contains the GOI; and a third module harbors a recombinase (Marillonnet *et al.*, 2004) (Figure 2A). Since the expression vector lacks CP and cannot move systemically, all parts are delivered to the nucleus by *A. tumefaciens*. After agrodelivery, modules are fused by the recombinase forming the magnICON replicon at DNA level which is then transcribed and the recombination site and intron sequences are removed. MagnICON system presented smart solutions to overcome several problems of viral vector systems. Subcellular compartments were efficiently reached by using signal sequences fused to the expressed proteins (Gils *et al.*, 2005), the efficiency of agroinfection was improved by the insertion of silent mutations to enhance the processing of the transcription products of the delivered DNA the nucleus (Marillonnet *et al.*, 2005), two proteins were expressed in the same cell with two non-competing viral vectors (Giritch *et al.*, 2006) and a massive inoculation technique, called ‘Magniffection’ was developed, which helped to raise heterologous protein yields up to 5 g/kg of fresh tissue (Gleba *et al.*, 2004; Gleba *et al.*, 2005). The triple module magnICON system was used to express several proteins in plants, with special attention on monoclonal antibodies (mAbs) (Santi *et al.*, 2006; Webster *et al.*, 2009; Huang *et al.*, 2006). Since TMV capsid is made-up from more than 2000 copies of the CP and both the N- and the C-terminus of the CP are situated on the virion’s surface, magnICON was also successfully used as an epitope presentation system (Turpen *et al.*, 1995; Wigdorovitz *et al.*, 1999; Werner *et al.*, 2006; Noris *et al.*, 2011). Other important gene expression system based on TMV is TRBO which is comparable to magnICON in terms of recombinant protein yields, but simpler in design (Lindbo, 2007). TRBO was constructed by replacing the CP of TMV with a multi cloning site into which the GOI is inserted, resulting in a powerful gene expression vec-

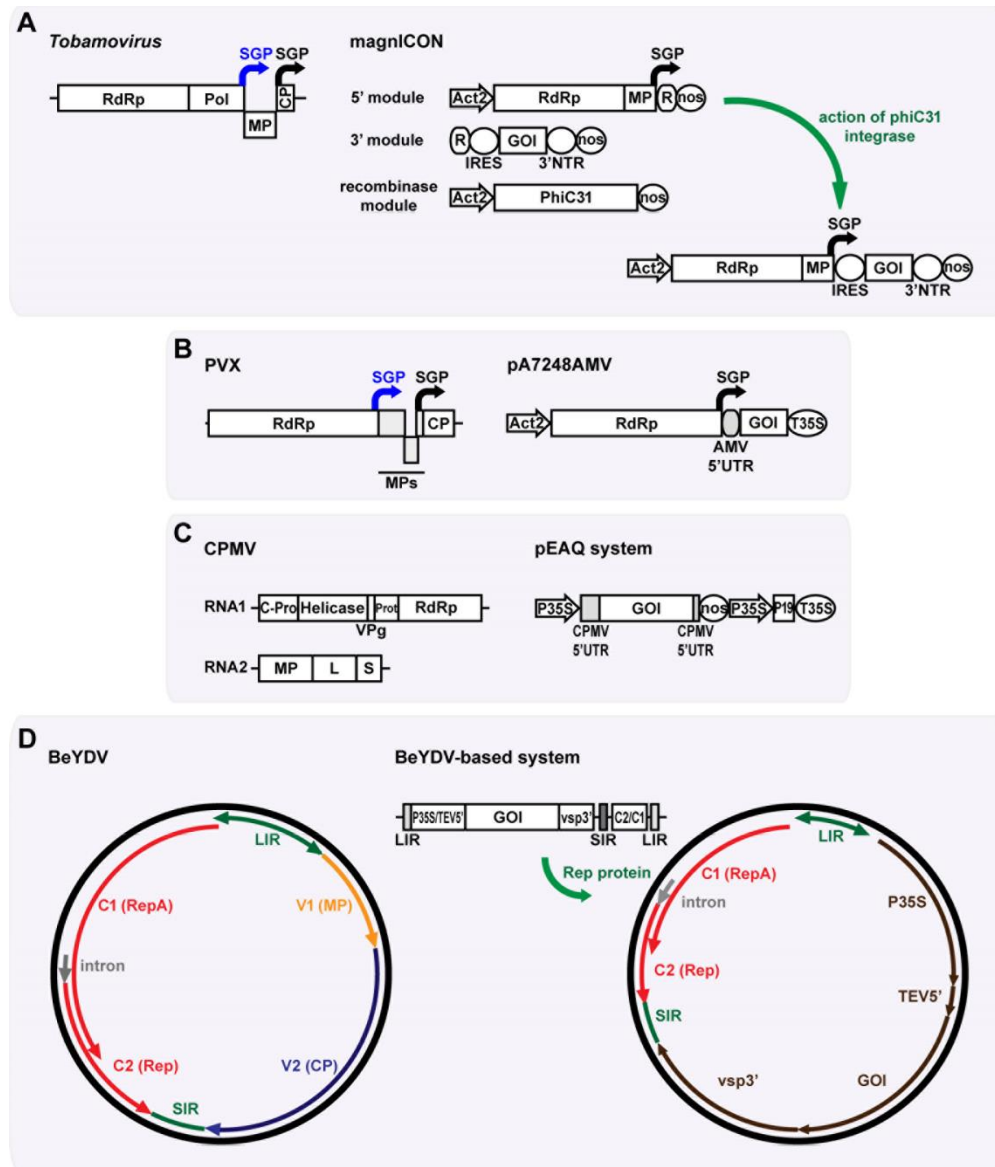


Figure 2. Plant viruses used as gene expression vectors for heterologous protein expression in plants. **A.** The tobamovirus genome: RNA-dependent RNA polymerase (RdRp), polymerase (PoI), MP, CP and two sgRNAs (SGP). The tobamovirus-based magnICON system is composed by three modules which are fused by a recombinase. 5' module: actin promoter (Act2), RdRp, MP, SGP, recombination site (R) and nopaline synthase terminator (nos). 3' module: R site, internal ribosome entry site (IRES), GOI, TMV 3' untranslated region (3'UTR), nos terminator. Recombinase module: Act2, integrase (PhiC31) and nos. **B.** PVX genome: RdRp, three MPs, CP, and two SGPs. In the PVX-based expression vector (pA7248AMV) the MPs and the CP was replaced by the GOI and the 5'UTR of *Alfalfa mosaic virus* RNA4 (AMV 5' UTR) expressed from a SGP under the control of *Cauliflower mosaic virus* (CaMV) promoter and terminator (P35S and T35S). **C.** CPMV genome consists of two RNAs (RNA1 and RNA2). RNA1: processing regulator (C-Pro), helicase, viral protein genome-linked (VPg), proteinase (Prot) and RdRp. RNA2: MP, large and small CPs. In CPMV-based pEAQ system, RNA1 was eliminated and the MP and CP of RNA2 was replaced by the GOI conserving the 5' and 3' UTRs (5'UTR and 3'UTR) of the virus under the control of P35S and a nos. **D.** BeYDV genome: MP, CP, virus replication associated proteins (Rep and RepA) separated by the long and short intergenic regions (LIR and SIR). In the BeYDV-based expression vector, the MP and CP functions were replaced by the GOI between one LIR and SIR element under the control of P35S, soybean vegetative protein 3' (vsp3'), TEV 5'UTR (TEV 5'). In the plant cell, Rep protein helps the release of the inserted DNA-segment, which replicates as an episome. Figure elaborated on the basis of Porta and Lomonosoff (2002) and Pevret and Lomonosoff (2015).

tor able to produce up to 5.5 g heterologous protein per kg of fresh tissues (Lindbo, 2007; Li *et al.*, 2013).

Potato virus X (PVX), the type species of the genus *Potexvirus*, family *Alphaflexiviridae*, with flexuous rod shaped virions and a ssRNA⁺ genome, was used to construct both gene expression and peptide-presentation systems. PVX genome contains five open reading frames (ORFs), coding for five proteins: the RNA-dependent RNA polymerase (RdRp) which is translated directly from the genomic RNA, the triple gene block module, coding for three MPs and the CP, which are expressed from sgRNAs (Figure 2B). PVX genome enables expression of foreign genes either as a separate ORF or as a fusion with the CP. The ‘deconstruction’ of PVX vectors was achieved by the replacement of the triple gene block MPs and the CP with the GOI, expressed from the first native subgenomic promoter resulting in a small replicating vector delivered to the plant genome with *A. tumefaciens* (Komarova *et al.*, 2006). This vector was further improved when the translation enhancer of the *Alfalfa mosaic virus* (AMV) RNA4 was added (Mardanov *et al.*, 2009) (Figure 2B). Beside gene expression systems, PVX VNPs and VLPS were also used for imaging, drug delivery and diagnosis (Lico *et al.*, 2015). Peptides can be fused directly to the CP or by inserting the 2A sequence of the *Foot-and-mouth disease virus* between the capsid protein and the GOI, enabling presentation of larger proteins on the capsid’s surface.

Cowpea mosaic virus (CPMV, genus *Comovirus*, family *Secoviridae*) both served as gene expression, peptide presentation system and VLPs (Sainsbury *et al.*, 2010). CPMV possess a bipartite genome, which encapsidates in isometric particles (Figure 2C). The genome consists of two positive-strand RNAs, named as RNA1 and RNA2 encoding for two polyproteins which are processed by a viral encoded protease. Proteins encoded in RNA1 are involved in the replication while RNA2 encodes for the MP and the large and small CPs. First, the MP and CP functions were substituted by *A. tumefaciens*-delivery (Liu and Lomonosoff, 2002) and the GOI was inserted between the 5’ and 3’ of RNA2 with a suppressor of gene silencing (Liu *et al.*, 2004; Cañizares *et al.*, 2006). This viral vector was further developed on the basis of the observation that RNA1 derived replication was not essential for the expression of the chimeric RNA2. The expressed protein yield was further increased by removing two start codons upstream of the main initiation site which lead to enhanced levels of mRNA translation. The resulting viral vector was named CPMV-hyperstable (CPMV-HT) (Sainsbury and

Lomonosoff, 2008). Subsequently, a series of small and user friendly vectors (pEAQ vectors) were developed, able to produce milligram quantities of recombinant proteins per kilogram of plant tissue (Sainsbury *et al.*, 2009) (Figure 2C). The knowledge of the detailed three-dimensional structure of the CPMV capsid and the availability of infectious cDNA clones let scientist to apply CPMV-based vectors in vaccinology against a number of pathogens (Sainsbury *et al.*, 2010). CPMV VLPs was also used in bionanotechnology and more recently it was demonstrated that the empty CPMV capsid has the ability to suppress metastatic cancer (Steinmetz and Evans, 2007; Steinmetz *et al.*, 2009; Lizotte *et al.*, 2016).

Bean yellow dwarf virus (BeYDV, genus *Mastervirus*, family *Geminiviridae*) is a spherical virus with single stranded circular DNA genome which is composed of four ORFs (V1, V2, C1 and C2) encoding four proteins: the MP, the CP and the virus replication associated proteins Rep and RepA (Liu *et al.*, 1997) (Figure 2D). The V- and C-strand ORFs are separated by long and short intergenic regions (LIR and SIR, respectively). LIR contains a bidirectional promoter and a stem-loop structure which is essential for initiation of the rolling circle replication of the V-strands. SIR is the origin of C-strand synthesis and contains transcription termination and polyadenylation signals. For viral replication, Rep can be supplied in *trans* which led to the design of a two component vector system in which viral functions were separated in two vectors (Mor *et al.*, 2003). The “Rep supply vector” encodes the virus replication associated proteins (Rep/RepA), controlled by inducible promoters to circumvent potential toxicity associated with these proteins; and the “LSL” vector, in which an expression cassette for the GOI was inserted replacing CP and MP. After the integration of the LSL cassette to the plant genome, inserted genes serve as a “master copy” which can be rescued and replicated as an episome in presence of the Rep vector, forming many more copies. Later, BeYDV-based system was further improved by inserting a native C1/C2 coding region under the control of a relatively weak promoter inside the LSL vector (Huang *et al.*, 2009) (Figure 2D). The BeYDV-based vector was able to produce hetero-oligomeric proteins by co-expressing two GOIs from two tandem linked BeYDV replicons stacked in the same vector and the heavy and light chain of a mAb 6D8 against Ebola virus was successfully expressed in *N. benthamiana* (Huang *et al.*, 2010).

A potyvirus-based expression vector for metabolic engineering

Plants are an important source of many valuable secondary metabolites and plant viruses has the ability to alter plant metabolome during the infection causing several physiological and biochemical changes in the chloroplasts (Li *et al.*, 2015). These facts gave rise to the idea of using plant viral vectors to manipulate secondary metabolite profile of plant tissues (Sainsbury and Lomonosoff, 2014). Plant virus based vectors have been modified to silence target genes in various plant species, characterize gene functions and help the elucidation of endogenous biosynthetic pathways (Robertson, 2004). Furthermore, several studies demonstrated the enormous potential of viral vectors for re-engineering of plant metabolism by expressing active enzymes and transcription factors (Kumagai *et al.*, 1995; Kumagai *et al.*, 1998; Bedoya *et al.*, 2010; Mozes-Koch *et al.*, 2012; Zhang *et al.*, 2013; Nielsen *et al.*, 2013; Sainsbury *et al.*, 2012; Geisler *et al.*, 2013). A big disadvantage of viral vectors is that the majority of these systems are unable to co-express various foreign proteins in the same cell but re-engineering of metabolic pathways requires the coordinated expression of several enzymes. To overcome this limitation a prominent solution was developed by using a potyvirus-based vector system (Bedoya *et al.*, 2010).

Potyriviruses (genus *Potyrivirus*, family *Potyriviridae*) has traditionally attracted interest of plant biotechnologists firstly, because of their genome expression strategy that allows production of heterologous proteins in an equimolar amount (Dolja *et al.*, 1992; Carrington *et al.*, 1993). Secondly, because potyvirus genome enables foreign gene insertion in several sites (Masuta *et al.*, 2000; Beauchemin *et al.*, 2005; Kelloniemi *et al.*, 2008) and large amounts of extraneous genetic material can be accommodated thanks to the elongated nature of the virion (Kelloniemi *et al.*, 2008).

Potyrivirus genome consists of a ssRNA⁺ of approximately 10,000 nucleotides (nt) encapsidated in long and flexuous virions (Urcuqui-Inchima *et al.*, 2001; Ivanov *et al.*, 2014) (Figure 3A). The genome is covalently linked in its 5' end to the viral protein genome-linked (VPg) and in its 3' end is polyadenylated. The genome contains a main ORF, which encodes a long polyprotein precursor which is processed co- and post-translationally resulting in ten mature multifunctional proteins. Potyrivirus genome harbors a second small ORF which encodes a single protein, called P3N-PIPO (Chung *et al.*, 2008; Mingot *et al.*, 2016). After host cell entry and uncoating, ssRNA⁺ genome serves as template for translation, executed by host cell ribosomes through a cap-

independent IRES. Subsequently, the polyprotein is processed by the three virus-encoded proteases: P1, HC-Pro and NIaPro. P1 and HC-Pro catalyse their own cleavage at their respective C-termini (Carrington *et al.*, 1989a; Carrington *et al.*, 1989b; Verchot *et al.*, 1991), whereas NIaPro cleaves the remaining sites *in cis* and *in trans*, recognizing a seven-amino acid-long sequence motif (Carrington *et al.*, 1990). Since viral proteins are required in different quantities, potyviruses developed multiple strategies to regulate the amount of viral proteins at each stage of infection (Ivanov *et al.*, 2014). Synthesized potyviral proteins are mainly expressed in the cytoplasm, but some viral proteins accumulate in the nucleus (Riedel *et al.*, 1998).

Several potyviruses have been used as gene expression vectors or peptide displaying VNPs (Masuta *et al.*, 2000; Fernández-Fernández *et al.*, 2001; Bedoya *et al.*, 2010; Sánchez *et al.*, 2013). GOI are usually flanked with sequence-specific recognition sites of a viral protease. After translation, these sites are recognized by the protease, which helps the release of heterologous proteins from the polyprotein (Carrington *et al.*, 1993). There are two frequently used positions for heterologous sequence insertion: between P1 and HC-Pro and between NIb and CP (Fernández-Fernández *et al.*, 2001; Beauchemin *et al.*, 2005; Dietrich and Maiss, 2003). A third insertion position was described which is situated closely to the amino terminus of *Potato virus A* polyprotein (Rajamäki *et al.*, 2005). The use of the right insertion position of foreign cDNAs has a crucial effect on protein accumulation (Choi *et al.*, 2000; Arazi *et al.*, 2001; Beauchemin *et al.*, 2005; Dolja *et al.*, 1997; Guo *et al.*, 1998; German-Retana *et al.*, 2000).

Bedoya *et al.* (2010) developed a *Tobacco etch virus* (TEV)-based gene expression vector by replacing the NIb cistron with a cassette which enables the co-expression of multiple heterologous proteins in tobacco plants (Figure 3B). The system takes advantage of the previously reported fact, that the NIb cistron can be complemented *in trans* (Li and Carrington, 1995) which enables the insertion of a large foreign sequence in place of this gene and which is also advantageous from a biocontainment point of view since it overcomes the risk of vector escape. Simultaneous expression of three fluorescent proteins was achieved by inserting their cDNAs in the expression cassette, with proteolytic processing motifs of the NIaPro on their sides. The TEV-based viral vector allowed simultaneous expression of these fluorescent proteins in equimolar amounts with the same subcellular localization.

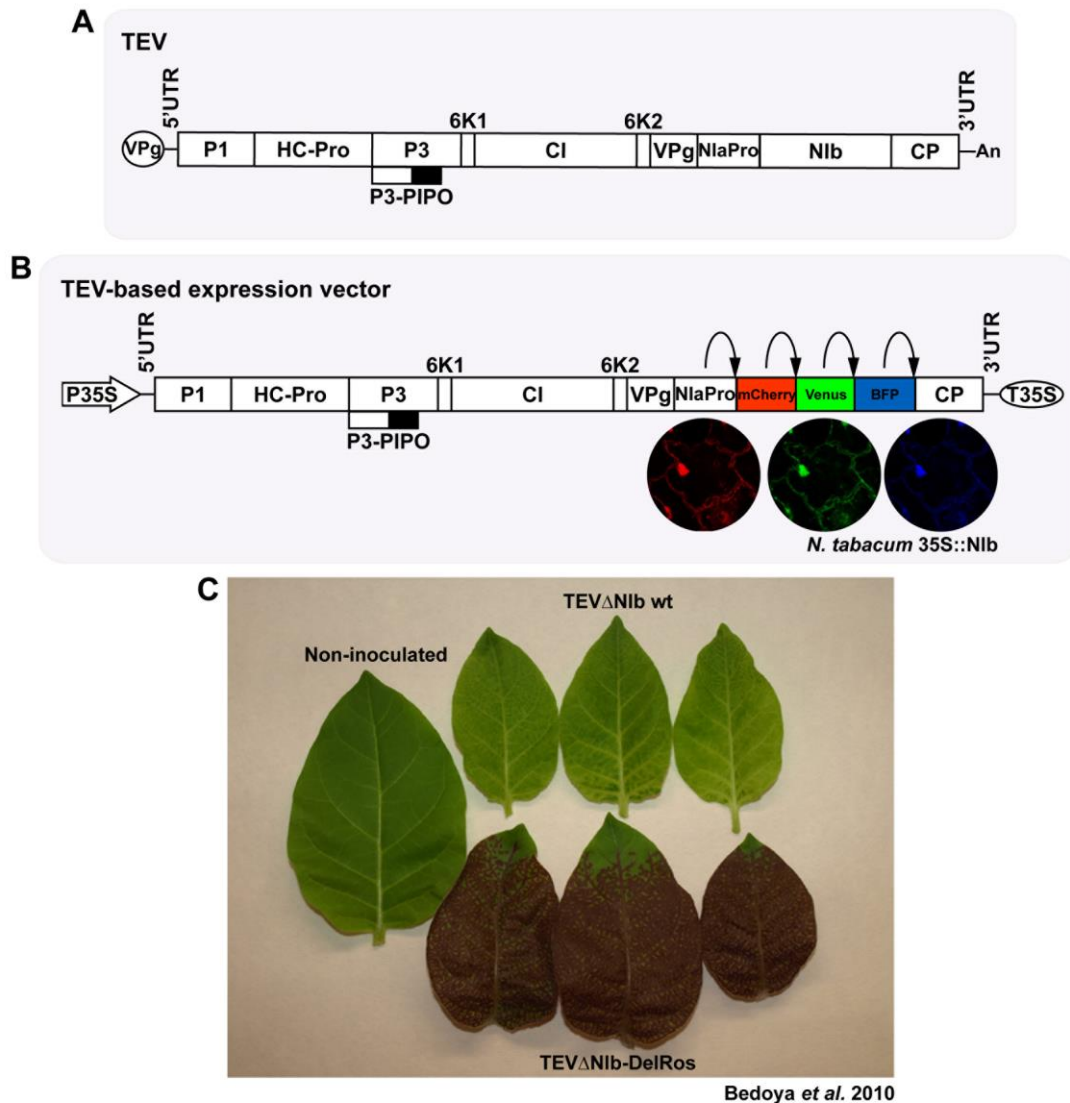


Figure 3. Potyvirus-based vector for protein expression in plants. **A.** Genome organization of parental TEV. VPg is covalently linked to the 5' end of the genome and the 3' end is polyadenylated (An). The genome contains a main ORF which encodes a long polyprotein precursor processed by the three viral proteases resulting in ten mature proteins: P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NlaPro, Nib and CP. The second, small ORF codifies for a single protein, called P3N-PIPO. **B.** TEV-based expression vector system. The TEV genome was inserted in a plasmid of *A. tumefaciens* for agrodelivery under the control of P35S and T35S and the Nib cistron was replaced by an expression cassette harboring three fluorescent proteins (mCherry, Venus and BFP) flanked by proteolytic processing sites of NlaPro which liberates fluorescent proteins (black arrows). These proteins are expressed in equimolar amounts in the same subcellular localization in transgenic *N. tabacum* plants (*N. tabacum* 35S::Nib). **C.** Expression of Delila and Roseal transcription factors with TEV-based viral vector (TEV Δ Nib-DelRos) in transgenic *N. tabacum*. Leaves infected with TEV-based vector lacking Nib cistron (TEV Δ Nib-wt) and noninoculated leaves are also showed. Figure adapted from Bedoya et al. (2010).

Therefore it was purposed to express several enzymes from a metabolic pathway and induce valuable natural products in plants.

To demonstrate the utility of the viral vector in metabolic engineering, two transcription factors, Delila and Rosea1 from *Antirrhinum majus*, were successfully expressed to boost anthocyanin accumulation in infected tobacco tissue. This resulted in the increase of anthocyanin content which conferred deep purple coloration to the leaves (Figure 3C). Subsequently, it was realized that the sole expression of Rosea1 is enough to produce red anthocyanins and Rosea1 was purposed as a candidate of visual marker to track viral infection in plants (Bedoya *et al.*, 2012).

Plant metabolic engineering

Plants have the unique capacity to transform light energy into carbohydrate molecules, like sugars, through photosynthesis, which serves as fuel for the organism. They are also able to defend themselves from pests and predators, participate in complex symbioses and survive extreme environmental conditions (Lau *et al.*, 2014). To do so, plants produce primary and secondary metabolites which are synthesized through a cascade of chemical reactions, called metabolic pathways. While primary metabolites including carbohydrates, lipids, proteins, and nucleic acids, are directly involved in nutrition and essential metabolic processes of plants, secondary metabolites have specific functions and typically mediate interactions between plants and their environment, like the attraction of pollinators or defense against herbivores and pathogens. Thousands of secondary metabolites have been isolated from plants, and many of them have powerful physiological effects in humans, like alkaloids, terpenoids and phenolics. Because of their high value in medicine and human diet, scientists made a strong effort on understanding their biosynthetic pathways and elaborating techniques to manipulate secondary metabolite content. The main objective of these efforts was to boost the accumulation of a target product by increasing the flow of molecules, called carbon flux, through a certain metabolic pathway or to reduce carbon flux through competing pathways (Farré *et al.*, 2014). There are many successful examples of metabolic pathway engineering in plants. The most famous is “Golden Rice”, which was engineered to accumulate vitamin A (β -carotene) in rice endosperm by introducing three enzymes from the carotenoid biosynthetic pathway of other organisms (Ye *et al.*, 2000). Golden rice was destined to poor countries where childhood blindness caused by

vitamin A deficiency is a common problem and rice is the major staple (Potrykus, 2003). These results gave rise to ‘nutritional genomics’, a discipline of modern biology that focuses on the elucidation of biochemical pathways and physiological processes of uptake and mobilization of nutrients which are needed in plants to sustain human health (Beyer, 2010). Several successful examples were documented about nutrient content modification of crops (Wu and Chappell, 2008; Dixon *et al.*, 2013; Farré *et al.*, 2014; Lau *et al.*, 2014). Just to mention some, tomatoes were engineered to accumulate high levels of anthocyanins by the expression of two transcription factors from snapdragon (*A. majus*). The improved content of anthocyanins of transgenic purple tomatoes resulted in significant extension of life span in cancer-susceptible mice feeding on it (Butelli *et al.*, 2008). Maize endosperm was also engineered to express three vitamins, β -carotene, ascorbate and folate, biosynthesized through three distinct metabolic pathways, destined to complete nutritional uptake of people in poverty (Naqvi *et al.*, 2009). Bred wheat potentially suitable for celiac patients and other gluten-intolerant individuals were also developed by RNA interference-mediated gene silencing down-regulating the expression of gluten proteins, responsible for celiac disease (Gil-Humanes *et al.*, 2010). Despite promising early examples, metabolic engineering have to face many challenges resulting from the high complexity of endogenous networks. For this purpose, a major knowledge of endogenous pathways is required, the best intervention points and most suitable genes have to be identified.

The importance of subcellular localization in re-engineering of metabolic pathways

A crucial question in re-engineering of metabolic pathways in plants is to determine the most suitable subcellular environment (Heinig *et al.*, 2013). Several studies revealed that assaying metabolic enzymes in more than one subcellular compartment could result in the discovery of more suitable subcellular environments for metabolite synthesis (Di Fiore *et al.*, 2002; Ohara *et al.*, 2003; Farhi *et al.*, 2011). Furthermore re-targeting enzymes to the most abundant substrate pool can increase production of a desired metabolite (Wu *et al.*, 2006; Aharoni *et al.*, 2003).

In practice, for heterologous enzyme delivery, the protein transport system of the plant cell is exploited. Eukaryotic plant cells contain several organelles such as nucleus, plastids, mitochondria, vacuole, peroxisomes, Golgi and endoplasmic reticulum embedded in the cytosol surrounded by at least one membrane which forms a diffusion

barrier that prevents uncontrolled exchange of molecules. The majority of them have a unique set of enzymes which provide an adequate environment for complex metabolic reactions and form an integrated cellular network evolved in many biosynthetic activities coordinated precisely. Plastids and mitochondria are organelles of endosymbiotic origin with important photosynthetic and metabolic functions in plant cells (Gray, 1992; Jarvis and López-Juez, 2013). Depending on the developmental context, plastids fulfil a diversity of roles in addition to photosynthesis. They are involved in the synthesis of amino acids, fatty acids, purine and pyrimidine bases, terpenoids, various pigments and hormones, and they are related to key functions of nitrogen and sulphur assimilation (Lopez-Juez and Pyke, 2005). Mitochondria are present in virtually all eukaryotic cells and have essential roles in energy metabolism, various metabolic pathways and cell death (Mackenzie and McIntosh, 1999). Despite the existence of an own genome, the genetic information contained in plastids and mitochondria is very low, since the vast majority of genes was transferred to the nucleus of the host cell (Timmis *et al.*, 2004). To maintain the same level of metabolic complexity, several thousands of gene products encoded in the nucleus and translated by ribosomes in the cytosol, must be targeted back to these organelles. This translocation process is facilitated by targeting sequences which are recognized by chaperone transporters and delivered to their target. In molecular biology, genetic and metabolic engineering, organelle targeting of heterologous proteins is achieved by fusing amino terminal signal sequences, called transit peptides (TP) for chloroplast and presequences (PRE) to the foreign cDNAs. The translocation process of proteins with amino terminal targeting peptides for chloroplast and mitochondria shares some general principles (Balsera *et al.*, 2009). First, precursor proteins have to be unfolded or roughly folded in order to be translocated. This is achieved with the help of molecular chaperones which keep targeted proteins in a soluble, not fully folded conformation. The translocation process also proceeds in a similar way. The signal sequences are recognized by the translocation complex and the protein is delivered to the corresponding organelle. After translocation, the preprotein binds to the membrane receptors and it is transferred to the protein import channel. After crossing the lipid bilayer, the signal sequence is processed and the protein acquires its functional conformation (Figure 4).

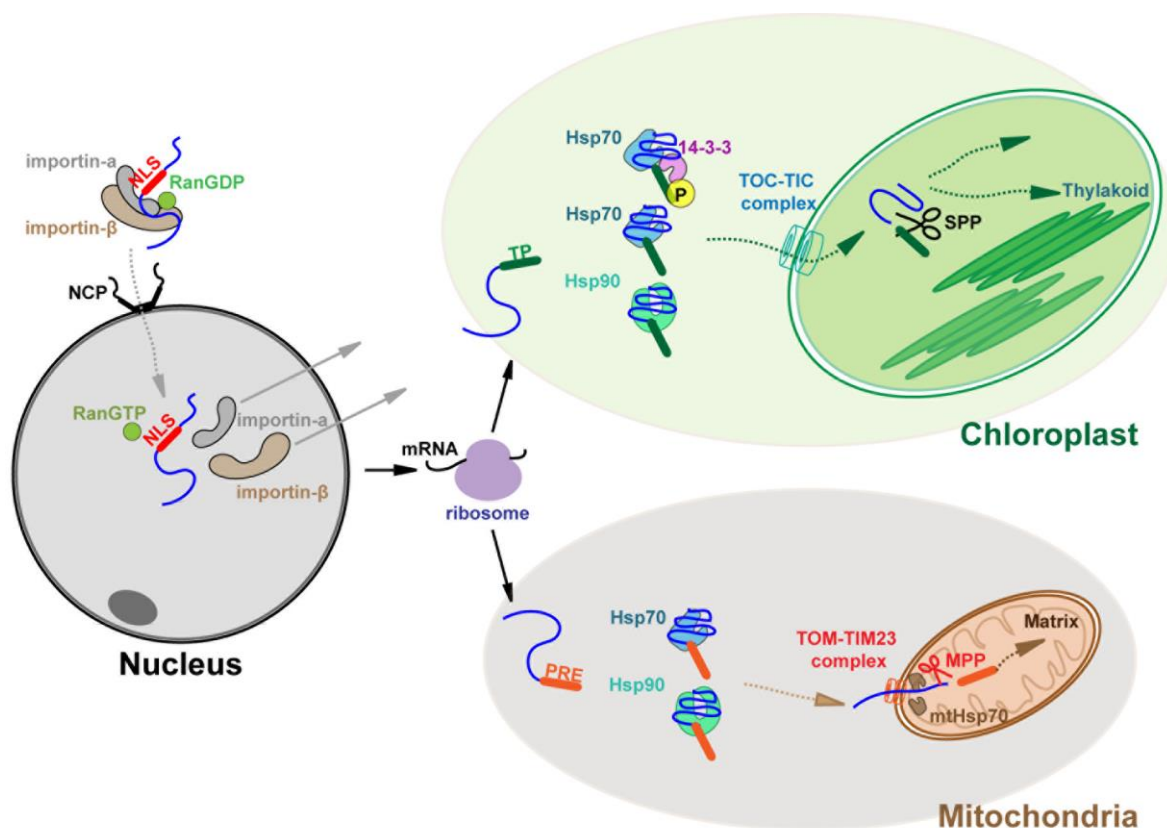


Figure 4. Translocation process of proteins targeted to the chloroplast, mitochondria and nucleus. Proteins encoded in the nucleus are transcribed in the cytosol. Those proteins destined to the chloroplast with amino terminal TP, are recognized by the translocation complex (Hsp70 and 14-3-3 protein or Hsp70 or Hsp90), which drives the unfolded protein to the TIC-TOC complex. After translocation, SPP cleaves the TP; the protein gains its mature conformation and reaches the corresponding localization inside the chloroplast. Mitochondrial targeting proteins with amino terminal PRE are recognized by Hsp70 or Hsp90 and guided to the TOM-TIM23 complex which helps these proteins in the translocation process in coordination with the mitochondrial Hsp70 chaperones (mHsp70) situated in the inner matrix. After translocation, MPP cleaves PRE from the protein, which achieves its mature conformation. Nuclear signal sequences are recognized by the translocation complex formed by importin- α , importin- β and RanGDP which helps these proteins in the translocation through the NCP. After transportation, the import complex dissociates and the importins are exported out of the nucleus.

In case of chloroplast TPs, the nascent protein associates to a guidance complex in the cytosol, composed of a heat shock protein-70 (Hsp70) and a 14-3-3 dimer on phosphorylation (Figure 4). Preproteins associated to this complex are quickly imported into the chloroplast which suggests that phosphorylation might select a subclass of preproteins for preferential import (Waegemann and Soll, 1996; May and Soll, 2000). Alternatively, non-phosphorylated proteins can associate with Hsp70, or with the heat shock protein-90 (Hsp90) (Qbadou *et al.*, 2006). All these complexes bind to receptors in the outer membrane in a GTP-dependent manner. In a joint effort between the translocon complexes at the outer and inner envelope (TOC and TIC, respectively),

preproteins are imported across outer and the inner envelopes of the organelle. After translocation, stromal processing peptidase (SPP) cleaves off the TP, the imported protein assembles in its mature conformation and it is directed to the stroma or the thylakoids.

Most of the mitochondria targeting proteins have a cleavable sequence at their amino terminus, called presequence or matrix targeting sequence, however many of them contain internal targeting signals. The nascent precursor protein is stabilized by cytosolic chaperones Hsp70 and Hsp90. Amino terminal PRE targets proteins to the mitochondrial preprotein translocases: the TOM complex in the outer membrane and the TIM23 complex in the inner membrane (Figure 4). After the preprotein passes through the translocase complex, the PRE reaches the matrix and interacts with mitochondrial Hsp70 (mtHsp70) which helps the translocation of the precursor protein to the matrix. Finally, mtHsp70 release and the PRE is proteolytically removed by the mitochondria-processing peptidase (MPP) (Neupert and Herrmann, 2007; Balsera *et al.*, 2009).

On the other hand, the nuclear import of proteins is one of the key steps in controlling gene expression in eukaryotes. The nucleus of eukaryotic cells is physically separated by a double membrane structure, called nuclear envelop, spanned by nuclear pore complexes (NPC), responsible for mediating bidirectional traffic between cytoplasm and the nucleus (Figure 4). NPCs allow the transport of small molecules and ions by passive diffusion, but for larger molecules, energy dependent active transport is required (Freitas and Cunha, 2009; Tamura and Hara-Nishimura, 2014). Proteins destined to the nucleus hold specific nuclear localization signals (NLS), which are heterologous in sequence and in localization. These sequences are recognized by importin- α proteins, which forms an import complex with importin- β and the Ran-GDPase. The import complex crosses the nuclear pore by the interaction between importin- β and the phenylalanine- glycine (FG) repeat of the NCP complex. Once in the inner face of the NCP, the import complex dissociates after conversion of Ran-GDP to Ran-GTPase and importins are exported back to the cytoplasm.

Carotenoid biosynthetic pathway in plants and progresses in genetic engineering of the pathway

The carotenoid biosynthetic pathway is one of the most studied in plants because of the important role of these molecules in photosynthesis as accessory pigments and in

human diet. In plants, carotenoids are biosynthesized and accumulate in chloroplasts of green leaves and stems, and in chromoplasts of colored flowers, fruits and roots (Li and Yuan, 2013; Jarvis and López-Juez, 2013). Carotenoid content and identity vary depending on the tissue type: in chloroplasts, where carotenoids play an essential role in light harvesting and photoprotection, carotenoid composition provides the optimal performance of photosynthesis (DellaPenna and Pogson, 2006; Domonkos *et al.*, 2013). However, in colored organs, where they serve as attractants for pollination or seed dispersal, carotenoids act like colorants, precursors of plant isoprenoid volatiles and signaling molecules (abscisic acid and strigolactones) (Giuliano, 2014).

In addition to their fundamental roles in plants, carotenoids are also essential for humans. β -carotene, a type of carotenoid is the most effective precursor of vitamin A. In human intestinal mucosa, β -carotene is converted to retinal which is necessary for the production of rhodopsin, the light-absorbing molecule for both low-light and color vision, and for the maintenance of epithelial and immune cells. Carotenoids also have high antioxidative potential and protect cells and tissues from oxidative damage (Stahl and Sies, 2003). Several epidemiological studies revealed that a carotenoid rich diet prevents certain chronic diseases as age related macular degeneration, cardiovascular diseases and various types of cancer (Landrum and Bone, 2001; Sharoni *et al.*, 2003; Ciccone *et al.*, 2013). Beside β -carotene, lycopene with potent antioxidant activity also has some prominent health promoting properties. Higher serum levels of lycopene have been shown to correlate with lower occurrence of prostate cancer (Giovannucci, 2002; Etminan *et al.*, 2004).

These health promoting properties and the effect of vitamin A deficiency intensified the interest on enhancing carotenoid content in crop plants by conventional breeding techniques and metabolic or genetic engineering. These studies also supported fundamental understanding of carotenoid formation and plant metabolism. In plants, carotenoids are synthesized from the precursor molecule isopentenyl diphosphate (IPP) and its double-bond isomer dimethylallyl diphosphate (DMAPP) through the methylerythriol 4-phosphate (MEP) pathway in plastids (Hirschberg, 2001; Sandmann *et al.*, 2006; Moise *et al.*, 2014). Nevertheless these precursor molecules are also synthesized outside the plastids through the mevalonate (MVA) pathway. Not just plants, but some bacteria, like *Pantoea ananatis*, and green algae also biosynthesize carotenoids through the MEP pathway (Rodriguez-Concepcion and Boronat, 2002). The

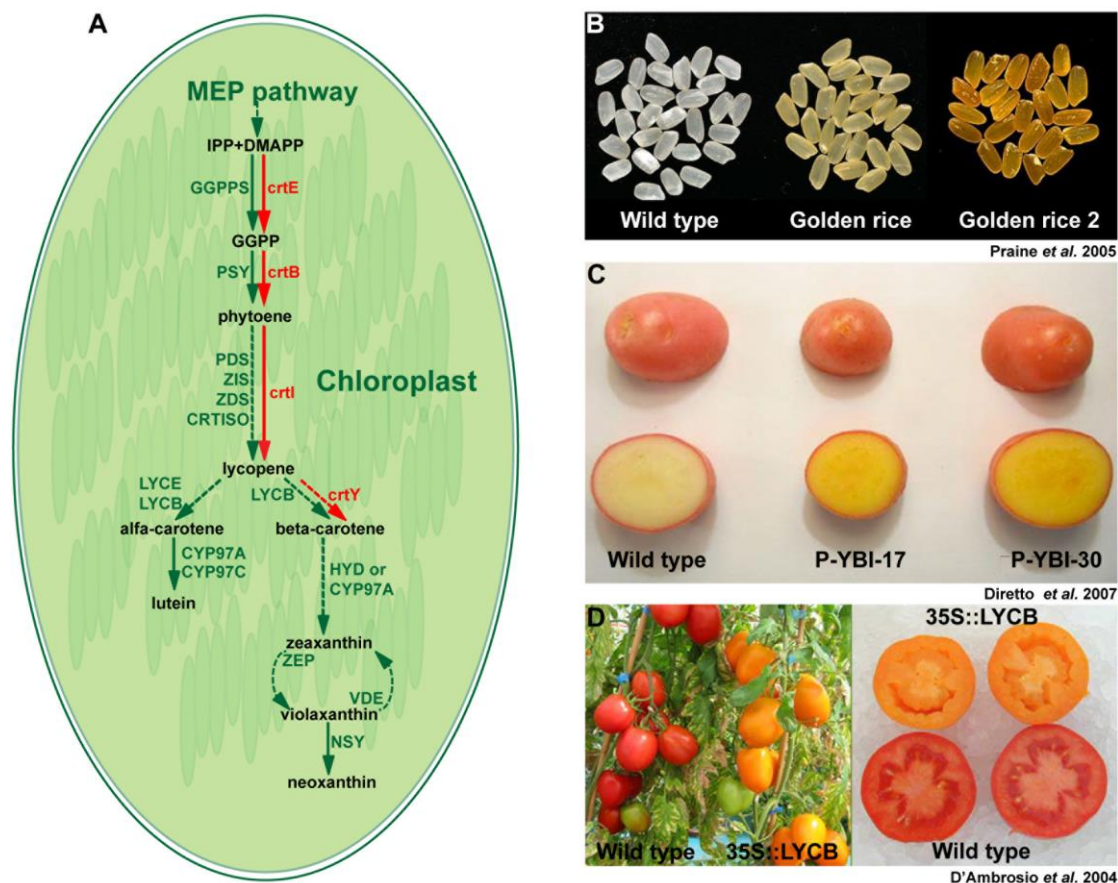


Figure 5. Carotenoid biosynthetic pathway and some successful examples of metabolic engineering of the pathway. **A.** Carotenoid biosynthetic pathway in plants (indicated in green) with some bacterial genes used in re-engineering the pathway (in red). **B-D.** Examples of transgenic crops accumulating high content of β -carotene in rice endosperm (B), potato tubers (C) and tomato fruits (D).

biosynthesis of carotenoids in plants starts with the conversion of three molecules of IPP and one molecule of DMAPP to geranylgeranyl diphosphate (GGPP) by the action of geranylgeranyl diphosphate synthase (GGPPS in plants and *crtE* in *P. ananatis*) (Moise *et al.*, 2014) (Figure 5A). The first committed step in plant carotenoid synthesis is the condensation of two GGPP molecules into the colorless phytoene, catalyzed by the phytoene synthase (PSY in plants and *crtB* in *P. ananatis*). PSY is generally accepted as the most important rate limiting step in the carotenoid biosynthetic pathway, tightly controlled by source and sink metabolites (Cazzonelli and Pogson, 2010). In *Arabidopsis thaliana*, only a single copy of PSY exists, but in many other plants, multiple PSYs control carotenoid metabolic flux in a tissue specific manner (Li *et al.*, 2008; Welsch *et al.*, 2008; Fray and Grierson, 1993; Giorio *et al.*, 2008). Tobacco genome codifies two genes, *psy1* and *psy2*, for phytoene synthase, from which *psy1* is expressed in much higher levels than *psy2* (Busch *et al.*, 2002). Phytoene undergoes a

series of desaturation reactions guided by the phytoene desaturase (PDS) and ζ -carotene desaturase (ZDS) to form poly-cis lycopene, also called pro-lycopene, which is converted to *trans*-lycopene, with a characteristic orange color, through the action of the carotene *cis-trans*-isomerase (CRTISO) (Park *et al.*, 2002). In *P. ananatis* the phytoene desaturase/isomerase, crtI, accomplishes all the above steps and convert phytoene to all-*trans*-lycopene in a one-step reaction (Misawa *et al.*, 1990). Lycopene is a branching point in carotenoid biosynthetic pathway and could undergo different cyclization, hydroxylation and epoxylation reactions resulting in a diverse range of carotenoids. The cyclization of lycopene can be executed by two competing enzymes: ϵ -cyclase (LYCe) or β -cyclase (LYCb). The introduction of two β -rings results in the production of β -carotene and the introduction of both an ϵ - and a β -ring produces the α -carotene. At this point, the LYCe plays a key role in controlling metabolic flux from one branch to another (Pogson and Rissler, 2000; Diretto *et al.*, 2006; Harjes *et al.*, 2008; Bai *et al.*, 2009). The hydroxylation reaction of α - and β -carotene produces lutein and zeaxanthin, respectively. This reaction is performed by two classes of enzymes, namely the ferredoxin-dependent nonheme diiron enzymes (HYD) and enzymes from the cytochrome P450 family, the P450 carotene hydroxylase 97A and 97C (CYP97A y CYP97C). The insertion of two epoxy groups into the zeaxanthin by the zeaxanthin epoxidase (ZEP) yields violaxanthin. This compound can be converted back to zeaxanthin by the violaxanthin de-epoxidase (VDE) under high light conditions. Neoxanthin is formed from violaxanthin by the neoxanthin synthase (NSY).

Foreign gene expression in edible crops led to the successful engineering of carotenoid content and a deeper understanding of the pathway regulation (Giuliano *et al.*, 2008; Fraser *et al.*, 2009). The carotenoid biosynthetic genes of bacterial and algae origin targeted to the chloroplast have proven to be a powerful tool because of their low homology with endogenous genes, which alleviates potential silencing. Furthermore, bacterial crtI was proved to promote phytoene conversion to all *trans*-lycopene in one step, thus overcoming the endogenous four enzyme reaction (Nogueira *et al.*, 2013). Since phytoene synthase is thought to be the most important limiting step in carotenoid biosynthesis, it is not surprising that it was the first target of metabolic engineering studies to increase carotenoid content. Nevertheless, the constitutive overexpression of the phytoene synthase resulted in dwarfism in tomato which was attributed to reduced gibberellin levels since both phytoene synthase and the gibberellin pathway uses GGPP

as substrate (Fray *et al.*, 1995; Busch *et al.*, 2002). The use of constitutive organ specific promoters overcame this problem and resulted in high levels of carotenoids in seeds (Shewmaker *et al.*, 1999; Fujisawa *et al.*, 2008), fruits (Fraser *et al.*, 2002) and tubers (Ducreux *et al.*, 2005).

The co-expression of more bacterial or plant enzymes co-ordinately gave rise to the increment of total carotenoid content or the obtainment of diverse combination of downstream metabolites in edible crops. PSY from *Narcissus pseudonarcissus* and chloroplast-targeted crtI were co-expressed in rice under the control of an endosperm-specific and a constitutive promoter, respectively. Rice endosperm accumulated β -carotene, lutein and zeaxanthin, giving birth to 'Golden rice' (Ye *et al.*, 2000). Carotenoid content was further improved in rice by using PSY from maize resulting in 'Golden rice 2' (Paine *et al.*, 2005) (Figure 5B). The increase of carotenoid levels in maize kernels were also achieved by the expression of *crtB* and *crtI* under the control of an enhanced γ -zein promoter to provide endosperm-specific expression (Aluru *et al.*, 2011). A carotenoid metabolic library was generated by inserting multiple carotenoid biosynthetic genes in a white endosperm maize variety deficient for endosperm carotenoid biosynthesis (Zhu *et al.*, 2008). These results confirmed that in maize endosperm PSY is the key enzyme in carotenoid biosynthesis inducing the increase of total carotenoid content which can be further increase by adding crtI. In potato, the co-expression of crtB, crtI and crtY under tuber specific promoter resulted in the increment of β - and α -carotene and phytoene levels conferring yellow phenotype to potato tubers (Diretto *et al.*, 2007) (Figure 5C). Metabolic engineering efforts in tomato fruits were concentrated both on increasing lycopene and β -carotene content. Bacterial genes did not favoured carotenoid accumulation like in previously mentioned crop plants. Neither the constitutive overexpression of *crtI* nor the fruit specific expression of *crtB* succeeded in obtaining significant quantity of β -carotene in tomato fruit (Römer *et al.*, 2000; Fraser *et al.*, 2002), but the overexpression of a tomato *lycb* gene resulted in the accumulation of β -carotene by the cyclization of almost all lycopene to β -carotene (Rosati *et al.*, 2000; D'Ambrosio *et al.*, 2003). With the combination of several enzymes of different origin, a diverse metabolic profile was reached in edible crops also expressing ketocarotenoids, like astaxanthin, a widely used food supplement, feed additive and natural colorant (Zhu *et al.*, 2008; Fujisawa *et al.*, 2009).

Reporter genes for tracking viral infection in plants

The use of reporter genes in plant virology contributed tremendously to the knowledge about plant viruses by permitting the visualization of viral presence in infected tissue, as well as the quantification and monitoring of viral infection.

The GUS reporter system based on the action of the enzyme β -glucuronidase (GUS) was used in pioneering plant virology studies. GUS is a hydrolase that catalyzes the cleavage of a variety of β -glucuronides, many of which are commercially available as spectrophotometric, fluorometric and histochemical substrates (Figure 6A). This technique, developed almost thirty years ago (Jefferson *et al.*, 1987), consists of the fusion of a GOI with the GUS, and the detection of the GUS enzyme activity. The GUS system enabled the *in vitro* visualization of subcellular localization of viral proteins and the virus extension through the infected plant (Figure 6A) (Restrepo *et al.*, 1990; Chapman *et al.*, 1992; Dolja *et al.*, 1992). Nevertheless, the detection of GUS activity requires a destructive technique which limits studies of plant virus infection dynamics.

The *in vivo* visualization of proteins in cell biology was achieved first thanks to the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* and its many derivatives (Day and Davidson, 2009; Matz *et al.*, 2002). The technique is based on the unique capacity of GFP chromophore to exhibit bright fluorescence when exposed to light with an appropriate wavelength, through a self-catalyzed protein folding mechanism and intramolecular rearrangement without any additional cofactors. The GFP chromophore is buried in the center of a ' β -barrel' protein structure (Ormo *et al.*, 1996) and both amino and carboxy termini are exposed on the surface of the protein, making it ideal as marker gene without affecting structural integrity of the protein and the fluorophore. Fluorescent proteins permitted the observation of virus spreading dynamics through living tissues (Baulcombe *et al.*, 1995; Oparka *et al.*, 1997) and facilitated many discoveries about plant virus proteins (Tilsner and Oparka, 2010) (Figure 6B). The joint use of several fluorescent proteins, available with different excitation and emission wavelengths, made possible the co-visualization of many viral proteins at the same time as well as the co-localization of viral and host factors (Caplan *et al.*, 2008; Wei *et al.*, 2010b; Wei *et al.*, 2010a). Nevertheless this technique requires specialized equipment and dissection of plant organs which imposes important limitations on viral dynamics studies, particularly in large scale or field experiments.

Plants exhibit a great variety of natural colored pigments, as the result of endogenous metabolic reactions. The production of pigments, due to their striking color, also can serve as reporters of viral infection. Plant pigments, except chlorophylls, belong to three major classes: anthocyanins, carotenoids and betalains (Tanaka *et al.*, 2008). Anthocyanins, a class of flavonoids, are water soluble pigments, synthesized in the cytosol, localized in the vacuole and broadly distributed in the plant kingdom. They provide a wide spectrum of colors ranging from red to blue depending on the pigment's structure, metal ions and pH. Colored carotenoids confer yellow-to-red coloration to fruits, flowers and roots. Betalains, also award yellow-to-red colors but these compounds are limited to a number of plants in the order *Caryophyllales*.

Studies on post-transcriptional gene silencing of the phytoene desaturase (PDS), a gene involved in carotenoid biosynthesis, served to visualize unequivocally the presence of virus in carotenoid rich organs (Kumagai *et al.*, 1995; Ruiz *et al.*, 1998; Ratcliff *et al.*, 2001; Liu *et al.*, 2002). Since carotenoids are accessory pigments of photosynthesis and UV protectants of chlorophyll, interfering with carotenoid biosynthesis leads to drastic consequences for the plant. In fact, many herbicides target steps from carotenoid biosynthesis and in particular PDS (Dayan *et al.*, 2014). The result of PDS-inhibition is the loss of carotenoid pigments, and the destruction of chlorophyll by photo-oxidation. The expression of PDS sequences with a viral vector triggers the consequent virus-induced gene silencing against the virus at post-transcriptional level, which in this case leads to the silencing of the endogenous plant PDS in infected tissues causing bleaching phenotype (Figure 6C).

Bedoya *et al.* (2012) developed a visual reporter system for virus infection consisting of red colored plant pigment production in infected tissues (Bedoya *et al.*, 2012) (Figure 6D). This reporter system is based on the expression of a MYB-type transcription factor, *Rosea1*, from *A. majus*. The biosynthesis of anthocyanins, members of the family of flavonoids, is controlled by a triad of transcription factors, an R2R3 MYB, a basic helix-loop-helix (bHLH) and a WD40 repeat protein complex. The anthocyanin biosynthetic pathways are tightly regulated during plant development in a spatio-temporal manner, requiring a specific combination of transcription factors which form a ternary complex (Hichri *et al.*, 2011). In *A. majus* three genes encoding MYB-related transcription factors, namely *Rosea1*, *Rosea2* and *Venosa*, and two genes encoding bHLH proteins, *Delila* and *Mut* interact in a tissue specific manner to regulate

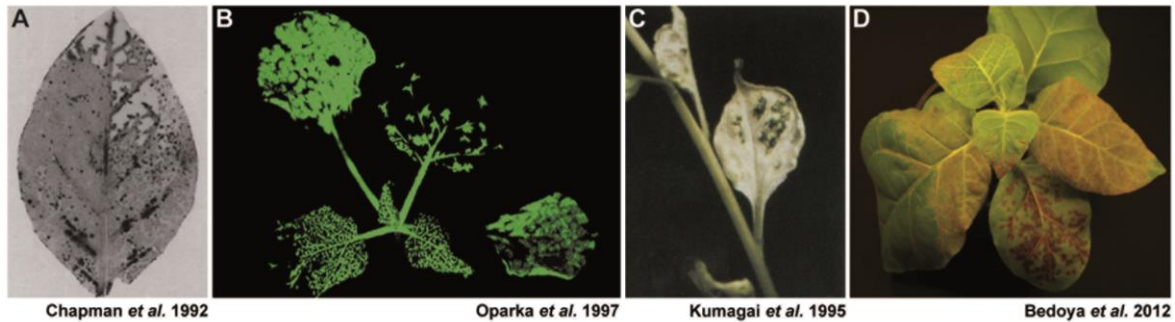


Figure 6. Markers for tracking viral infection in plants. **A.** Histochemical analysis of GUS activity in systemically infected *N. tabacum* leaves with PVX tagged with GUS. **B.** Symptoms of PVX tagged with GFP infecting *N. benthamiana*. **C.** *N. benthamiana* transfected with TMV harboring a PDS cDNA in its genome. **D.** Symptoms of TEV tagged with Roseal visual marker in *N. tabacum*.

anthocyanin biosynthesis (Schwinn *et al.*, 2006). Roseal in combination with Delila were previously used to boost anthocyanin accumulation in tomato fruits conferring purple coloration to these fruits (Butelli *et al.*, 2008). The same transcription factors were used to tag transgenic tomatoes expressing a recombinant human immunoglobulin A, making them easily distinguishable from wild-type tomatoes (Juárez *et al.*, 2012). Based on these findings, Bedoya *et al.* (2010) co-expressed Roseal and Delila in *N. tabacum* plants with the TEV-based expression vector which conferred red coloration to the infected tissue. However, Roseal, a protein whose size is similar to GFP, was enough to induce pigmentation visible to naked eye. Roseal was proposed as a visual marker for tracking viral infection and movement in several host plants (Bedoya *et al.*, 2010; Bedoya *et al.*, 2012) (Figure 6D).

The use of Roseal as visual marker presented several advantages over other reporter systems. First, viral infection and movement can be easily detected by the naked eye without the need of any specific devices enabling quick monitoring of many plants in a nondestructive manner. Second, Roseal is stable in the viral genome during the infectious cycle and through successive passages from plant to plant. Third, Roseal induced anthocyanin accumulation correlates with the viral load, so it can be considered as a quantitative reporter with an easy methanol-based extraction of anthocyanins and a colorimetric measurement of pigment content. These findings make this visual marker ideal for quick monitoring of plants in field experiments. However, each virus-host combination has to be tested, because the capacity of Roseal to induce anthocyanin

biosynthesis depends on the natural set of enzymes in each particular plant species or taxonomic groups.

Objectives

Plants may be excellent biofactories to produce products of interests, including proteins and peptides, but also chemical compounds. Thanks to the simplicity of their genomes and to their remarkable biological properties, plant viruses can be converted into biotechnological tools that may facilitate some of the goals of plant metabolic engineering. The capacity of potyvirus-derived vectors to co-express several heterologous proteins converts this type of plant viruses in promising tools in plant metabolic engineering. The objectives of this work have been:

1. To study potyvirus genome organization analyzing the effect of inserted genes on virus viability.
2. To target heterologous proteins to different subcellular compartments from a potyvirus-based vector.
3. To manipulate carotenoid biosynthesis by expressing several heterologous enzymes with a potyvirus vector and to produce lycopene in the cytosol of tobacco cells.
4. To develop visual markers of virus infection based on the expression of genes which modify endogenous metabolic pathways in plants.

Chapter I

Relocation of the NIb Gene in the Tobacco Etch Potyvirus Genome

Eszter Majer^a, Zaira Salvador^a, Mark P. Zwart^a, Anouk Willemsen^a, Santiago F. Elena^{a,b}, José-Antonio Daròs^a

Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia, Valencia, Spain^a; The Santa Fe Institute, Santa Fe, New Mexico, USA^b

Address correspondence to J.-A. Daròs, jadaros@ibmcp.upv.es

Journal of Virology (2014) 88(8): 4586-4590.

Abstract

Potyvirus express most of their proteins from a long open reading frame that is translated into a large polyprotein processed by three viral proteases. To understand the constraints on potyvirus genome organization, we relocated the viral RNA-dependent RNA polymerase (NIb) cistron to all possible intercistronic positions of the Tobacco etch virus (TEV) polyprotein. Only viruses with NIb at the amino terminus of the polyprotein or in between P1 and HC-Pro were viable in tobacco plants.

Research article

Viral genomes are compact assemblies of genes and regulatory sequences, and plant viruses are no exception (Roossinck, 2011). To compact their genomes (Belshaw *et al.*, 2007; Chirico *et al.*, 2010), plant viruses have evolved mechanisms that include overlapping open reading frames (ORFs), ambisense coding or translational frame-shift and read through (Nguyen and Haenni, 2003). One of the most common strategies, however, coding for a polyprotein that is processed in many different gene products after translation. Potyviruses (genus *Potyvirus*, family *Potyviridae*) are one of the largest groups of plant viruses and take this strategy to an extreme. Potyviral genomes are RNA molecules of positive polarity approximately 10,000 nucleotides (nt) long, consisting of a long ORF flanked by two short untranslated regions (UTRs) (Riechmann *et al.*, 1992). The potyviral ORFs apparently encode for ten mature gene products: P1 proteinase; helper-component proteinase (HC-Pro); P3 protein; 6K1 polypeptide; cylindrical inclusion (CI) protein; 6K2 polypeptide; nuclear inclusion a (NIa) protein, a polyprotein that is further processed to, at least, the viral protein genome-linked (VPg) and the NIa proteinase (NIaPro); nuclear inclusion b (NIb) protein, the viral RNA-dependent RNA polymerase; and the coat protein (CP) (Urcuqui-Inchima *et al.*, 2001). These products are released from the viral polyprotein through the activity of the three viral proteinases: P1, HC-Pro and NIaPro. An additional gene product, P3N-PIPO, results from a translation frame shift in a slippery region of the P3 cistron (Chung *et al.*, 2008).

The aim of this work was to improve our understanding of the functional, structural and evolutionary constraints that rendered the actual genome organization of potyviruses. To achieve this goal, we relocated one of the genes of *Tobacco etch virus*

(TEV) to all possible intercistronic positions in the genome, and assayed the infectivity of the resulting recombinant viruses in tobacco (*Nicotiana tabacum* L. cv. Xanthi nc) plants. To avoid further complexity, P3N-PIPO was excluded from this study. We chose the NIB gene because the infectivity of a TEV deletion mutant lacking the entire NIB cistron can be rescued by NIB expression in *trans* from a transgene (Li and Carrington, 1995) or a compatible viral vector (Bedoya *et al.*, 2010). When relocating the NIB gene to the different intercistronic sites, in some cases we added appropriate sequences to the amino and carboxy NIB termini to create NIAPro proteolytic sites that could mediate NIB release from the viral polyprotein. To facilitate the monitoring of virus infection, spread and accumulation, all recombinant viruses were constructed from a TEV clone carrying the Rosea1 (Ros1) visual marker (Bedoya *et al.*, 2012; Majer *et al.*, 2013). TEV-Ros1 load correlates with anthocyanin accumulation in tobacco tissues (Bedoya *et al.*, 2012). Figure 1 outlines the genome of the parental (TEV-Ros1) and the derived (TEV Δ NIB-Ros1-NIB1 to 9) recombinant viruses. Figure S1 in the supplemental material specifies the exact nucleotide sequences of all recombinant viral clones.

The parental and recombinant TEV clones, constructed from the binary plasmid pGTEV-Ros1 (Bedoya *et al.*, 2012; Majer *et al.*, 2013), were agroinoculated (Bedoya and Daròs, 2010) in two leaves of twenty 3-week-old wild-type plants and 20 transgenic plants constitutively expressing TEV NIB (Li and Carrington, 1995). Plants were grown in a glasshouse at 25 °C and 16 h light, and Ros1 expression was visually monitored during 4 weeks. For each clone, systemic leaves from three wild-type and three transgenic plants were harvested at 15 days post-inoculation (dpi), photographed, and used to estimate viral load by measuring the anthocyanins accumulation induced by the Ros1 marker (Bedoya *et al.*, 2012).

Most of the viral clones tested were not viable in either wild-type or transgenic plants. In other words, systemic tissue 4 weeks after agroinfiltration had no visible anthocyanin accumulation or infection symptoms. Only viruses with NIB relocated to the first two intercistronic positions were viable in wild-type plants. Ros1 activity was detected in systemic leaves of all 20 wild-type plants agroinoculated with TEV Δ NIB-Ros1-NIB1 and TEV Δ NIB-Ros1-NIB2, as well as the parental virus TEV-Ros1 (Figure 2A). There is therefore a statistically significant effect of NIB position on viability (test of equal proportions, $X^2 = 180.00$, 8 d.f., $P < 0.001$). An identical, statistically signifi-

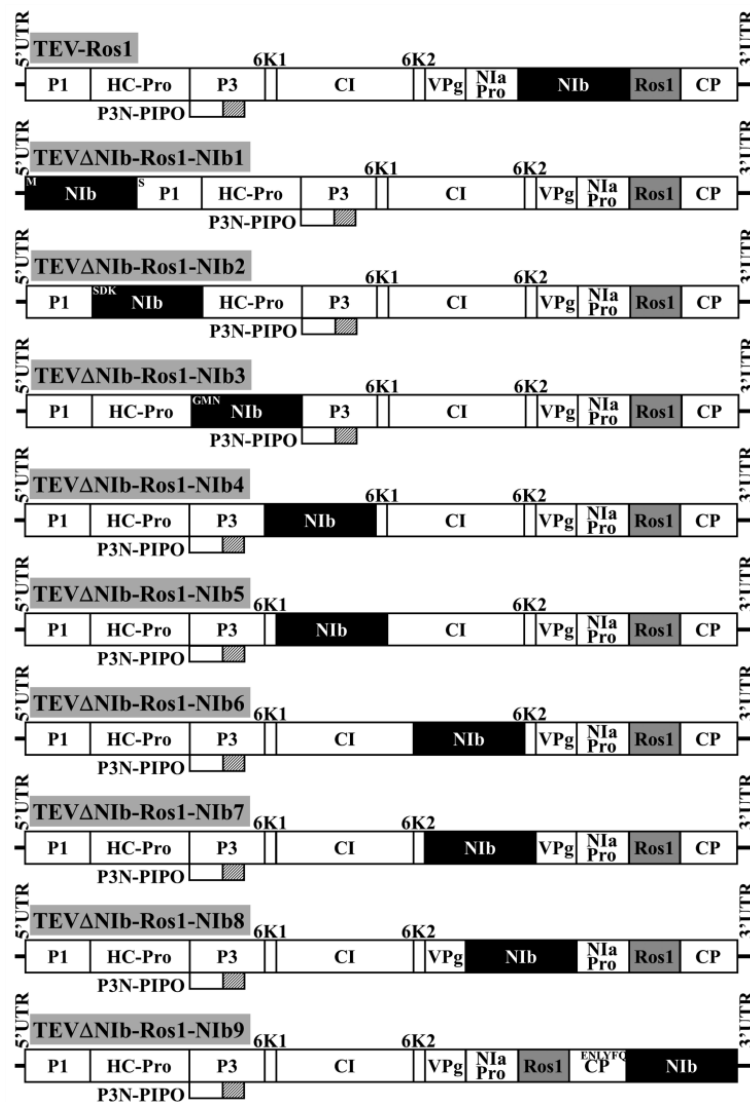


Figure 1. Schematic representation of parental TEV clone (TEV-Ros1) including the visual Ros1 marker (gray rectangle) and the nine recombinant clones in which the NIB gene (black rectangle) was relocated to nine different intergenic positions of the viral polyprotein (TEVΔNib-Ros1-Nib1 to 9). Amino acid sequences that are indicated next to the NIB amino or carboxy terminus in some cases (TEVΔNib-Ros1-Nib1, 2, 3 and 9) were inserted to complement NIB processing from the polyprotein.

cant result was obtained for transgenic plants expressing Nib ($X^2 = 155.077$, 8 d.f., $P < 0.001$), except that approximately half of the plants agroinoculated with the virus carrying NIB in the third intergenic position (TEVΔNib-Ros1-Nib3) also showed systemic marker expression (Figure 2B). The number of Ros1-expressing plants for TEVΔNib-Ros1-Nib3 was intermediate, as it was significantly different from all other clones (pairwise test of equal proportions: $P < 0.001$ for all comparisons).

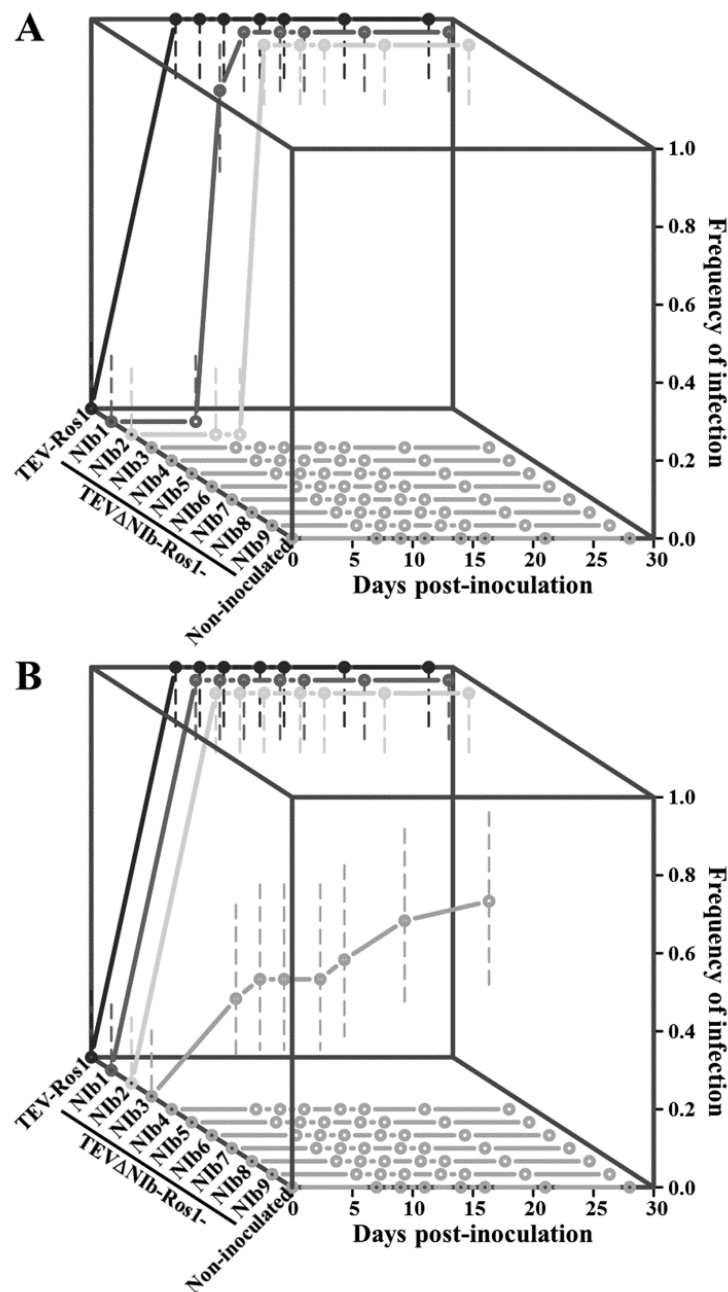


Figure 2. Infectivity of TEV recombinant clones in which the NIB gene was relocated to different intercistronic positions in the viral polyprotein. Plots show the cumulative frequency of infection in wild-type (A) and transgenic (consistently expressing TEV NIB) (B) tobacco plants versus days post-inoculation (dpi) with TEV-Ros1 and TEV Δ NIB-Ros1-NIB1 to 9. Symptoms were screened at 7, 9, 11, 14, 16, 21 and 28 dpi. Error bars represent 95% confidence interval of the estimated frequencies.

Next, we considered at what time after agroinfiltration anthocyanin accumulation was first apparent. In wild-type plants, there were significant differences in median time until visual detection of Ros1 expression for all three viable viruses (Figure 2A) (log-rank test: $P < 0.001$ for all three comparisons between TEV-Ros1, TEV Δ NIB-Ros1-NIB1 and TEV Δ NIB-Ros1-NIB2). So, TEV Δ NIB-Ros1-NIB2 was

significantly slower than TEV Δ NIb-Ros1-NIb1, while TEV Δ NIb-Ros1-NIb1 was significantly slower than TEV-Ros1. In transgenic plants, TEV-Ros1, TEV Δ NIb-Ros1-NIb1 and TEV Δ NIb-Ros1-NIb2 all had the exact same median time until Ros1 expression was observed visually (Figure 2B). For TEV Δ NIb-Ros1-NIb3, anthocyanin accumulation in systemic tissue was first observed significantly later than for any of the other viable viruses (pairwise log-rank test: $X^2 = 11.600$, 1 d.f., $P = 0.001$ for all three comparisons). In summary, TEV Δ NIb-Ros1-NIb1 and TEV Δ NIb-Ros1-NIb2 were viable in both plant genotypes, although infection appears to proceed more slowly than for the ancestral virus in wild-type plants. For TEV Δ NIb-Ros1-NIb3, viral spread was only seen in some transgenic plants, and when it appeared, it was significantly delayed.

We then analyzed the stability of the relocated NIb in the progeny of the viable viruses. For TEV Δ NIb-Ros1-NIb1 and TEV Δ NIb-Ros1-NIb2, reverse transcription polymerase chain reaction (RT-PCR) analysis of the viral progeny 15 dpi confirmed the stability of the relocated NIb gene (Figure 3A and B). For TEV Δ NIb-Ros1-NIb3, which showed delayed anthocyanin accumulation in only some transgenic plants, RT-PCR analysis and sequencing confirmed that the NIb gene was lost (Figure 3C). Anthocyanin accumulation was quantified in extracts from systemic leaves of three plants inoculated per recombinant clone at 15 dpi (Figure 4). The data were analyzed with a generalized linear model with full-factorial design, using a log-link function and gamma distributed error structure. Overall, the recombinant clone ($X^2 = 2383.451$, 10 d.f., $P < 0.001$) and plant genotype ($X^2 = 35.356$, 1 d.f., $P < 0.001$), had significant effects on anthocyanin accumulation, and there was a significant interaction between these two factors as well ($X^2 = 419.276$, 10 d.f., $P < 0.001$). A Tukey *post hoc* test highlights the existence of three non-overlapping groups of recombinant clones (Figure 4A): TEV Δ NIb-Ros1-NIb1 and TEV Δ NIb-Ros1-NIb2 showed the same expression level as the wild-type TEV-Ros1, TEV Δ NIb-Ros1-NIb3 showed an intermediate expression level that depended on the plant genotype (explaining the significant interaction term), and the other six recombinant clones were not significantly different from the mock inoculated plants in either plant genotype.

These results suggest the existence of many restrictions to the organization of the potyviral genome. Even though NIb can be provided in *trans* (Li and Carrington, 1995; Bedoya *et al.*, 2010), it can only be relocated to the amino terminus of the polyprotein or in between P1 and HC-Pro without affecting virus viability.

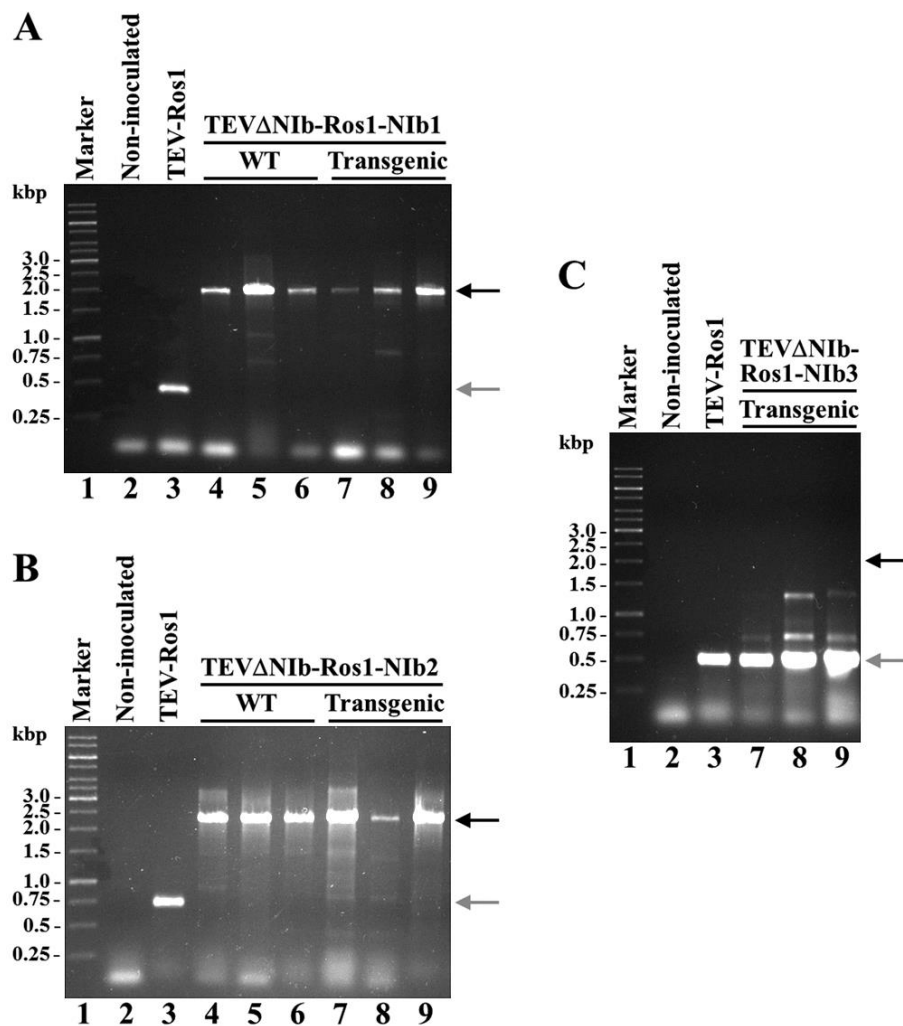


Figure 3. Stability of the relocated NIB gene in the viral progeny at 15 dpi. RNA from systemic leaves of three representative inoculated plants was purified. cDNAs amplified by RT-PCR using primers flanking the new NIB location for TEVΔNIB-Ros1-NIB1 (A), TEVΔNIB-Ros1-NIB2 (B), and TEVΔNIB-Ros1-NIB3 (C), were separated by electrophoresis in 1% agarose gels, and stained with ethidium bromide. Lanes 1, DNA marker with the size of some of the components indicated on the left; lanes 2, noninoculated negative control; lanes 3, parental virus positive control (TEV-Ros1); lanes 4 to 6, three infected wild-type plants; lanes 7 to 9, three infected transgenic plants. Note that in panel C, lanes 4 to 6 (wild-type plants) are missing because none of these plants was infected. The black and gray arrows indicate the positions of the expected RT-PCR products if the relocated NIB is present or absent, respectively. Primers used to amplify the different cDNAs were, for RT reactions, 5'-CTTTACATAGTTTTTTTCCAACATTTTCATG-3', and for PCRs, 5'-AAAATAACAAATCTCAACA-CAACATATAC-3' and 5'-CTCTTGCCATGGGTGAGCGCGAC-3' (A), 5'-CTCAACTCCAAG-AATTTTC-3' and 5'-GTCGCGCTCACCCATGGCAAGAG-3' (B), and 5'-GTGGCCAACAATGCA-AGATGTTGC-3' and 5'-GTAACTTTTGCGCCAAGGCTGAC-3' (C).

The relocation of NIB to the other seven positions renders nonviable viruses, even in a transgenic plant constitutively expressing TEV NIB that can be infected by a TEV mutant with a complete NIB deletion (Li and Carrington, 1995). The relatively late infection of approximately half of the transgenic plants inoculated with TEVΔNIB-

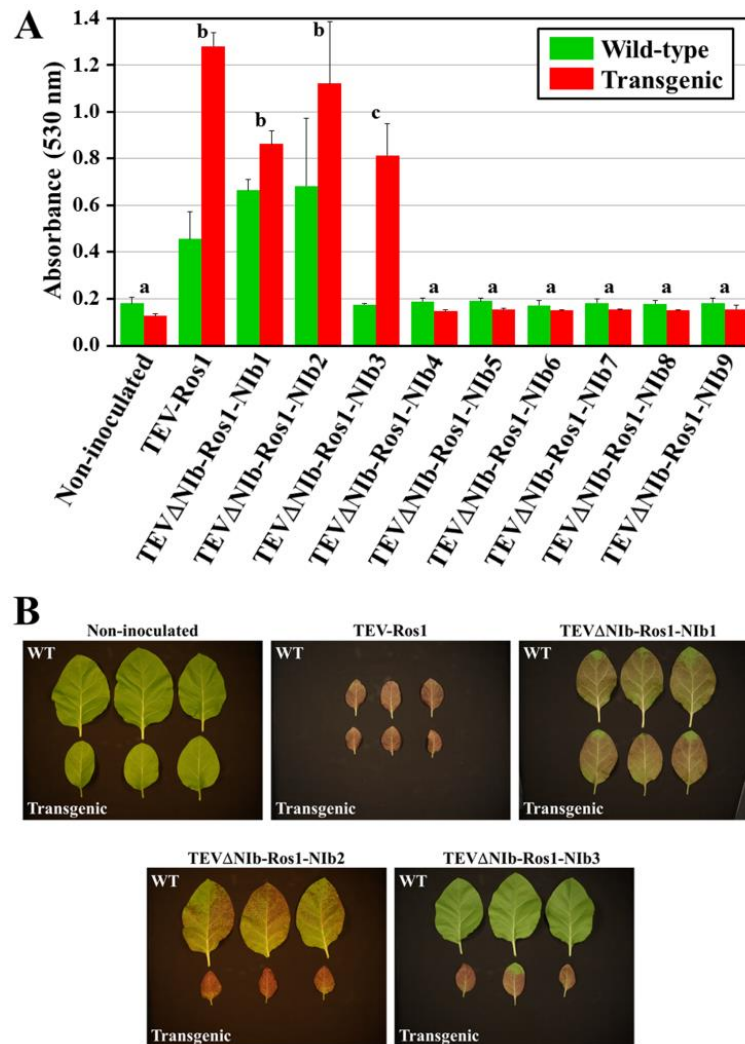


Figure 4. Anthocyanins accumulation at 15 dpi in systemic leaves of tobacco plants inoculated with TEV recombinant clones in which the NIB gene was relocated to different intergenic positions. (A) The absorbance at 530 nm of extracts from systemic leaves from three inoculated wild-type and transgenic plants constitutively expressing TEV NIB was plotted. Error bars represent ± 1 standard deviation among experimental replicates. Letters a, b and c over the bars represent groups of inoculations with homogeneous accumulation of anthocyanins according to a Tukey honestly significant difference test. (B) Pictures of systemic leaves harvested at 15 dpi from three wild-type and three transgenic tobacco plants noninoculated and inoculated with TEV-Ros1 and TEVΔNIB-Ros1-NIB1 to 3 as indicated.

Ros1-NIB3 probably resulted from a sporadic recombination event in which the relocated NIB was deleted. The resulting virus (TEVΔNIB-Ros1) was then able to infect the plant, but only when NIB was provided from a transgene. We amplified by RT-PCR a cDNA fragment corresponding to the HC-Pro/P3 intergenic region from the TEVΔNIB-Ros1-NIB3 progeny arising in the NIB-expressing transgenic plants and clone it. Sequencing of three independent clones showed that recombination cleanly restored the wild-type HC-Pro/P3 junction in all cases. Interestingly, no such

recombination events were observed in the case of any of the recombinant clones. In fact, the inoculated leaf of both wild-type and transgenic plants inoculated with TEV Δ NiB-Ros1-NiB3 showed slight anthocyanin accumulation, suggesting some replication capacity of this chimera.

The relocation of NiB to seven intercistronic positions probably results in non-viable viruses because of the induction of lethal defects in polyprotein processing, disruption of partially processed gene products with distinctive roles during infection, or disruption of regulatory RNA elements in the potyviral genome. A related potyvirus (*Turnip mosaic virus*) was found to support expression of heterologous proteins between HC-Pro and P3, and between 6K1 and CI in *Nicotiana benthamiana* and *Chenopodium quinoa* (Chen *et al.*, 2007). These differences may be due to the use of different virus species and host plants in both studies. On the other hand, in the previous work heterologous genes were inserted in the potyviral genome (Chen *et al.*, 2007), while here a potyviral gene was relocated along the genome. Processing of the potyviral polyprotein seems to be a finely regulated process that produces the right amounts of the different gene products in time and space (Merits *et al.*, 2002). Regulation is mainly based on the specific amino acid sequence recognized by the viral NIaPro, with some processing sites being cleaved faster than others (Carrington *et al.*, 1993; Parks *et al.*, 1995; Kim *et al.*, 1995). Insertion of NiB in some of the intercistronic positions of the polyprotein may have fatal effects on this regulation. Partially processed products from the potyviral polyprotein may have distinctive roles in the infectious cycle, different to those of the final processing products. This seems to be the case for the 6K2/VPg/NIaPro polyprotein that has been suggested to anchor TEV replication complexes to cellular endomembranes while recruiting NiB for replication (Daròs *et al.*, 1999). Insertion of NiB may therefore be lethal because it interrupts functional polyproteins. Finally, the potyviral genome contains regulatory RNA elements overlapping the ORF, including a series of RNA hairpins at the end of the CP cistron and 3' UTR that are involved in TEV replication (Haldeman-Cahill *et al.*, 1998). Disruption of these elements may also have fatal consequences for the virus. Our results highlight the complexity of the potyviral genome organization, suggesting the existence of many more regulatory elements and functional entities than those currently recognized. Our results seem also to reflect the genomic organization of the different *Potyviridae* genera apparently consisting of two genome blocks: a more conserved

block from P3 to the end, and the block including P1 and HC-Pro (in the genus *Potyvirus*) that is more variable between the different genera and even species.

Potyviruses have been used as expression vectors in plants (Dolja *et al.*, 1992; Fernández-Fernández *et al.*, 2001; Sánchez *et al.*, 2013). Their expression strategy, mainly based on production of a large polyprotein, makes them particularly attractive to simultaneously produce equimolar amounts of several heterologous proteins (Masuta *et al.*, 2000; Beauchemin *et al.*, 2005; Kelloniemi *et al.*, 2008; Bedoya *et al.*, 2010). Our results show the potentials and limitations inherent in expressing heterologous proteins from potyviral vectors. According to our results, the only positions where sequences coding heterologous proteins can be inserted without completely compromising viral viability are the amino terminal end of the polyprotein, between P1 and HC-Pro, and between NIb and CP. The P1/HC-Pro and NIb/CP intercistronic positions have been used with great frequency to express heterologous proteins in many potyviruses (Dolja *et al.*, 1997; Whitham *et al.*, 1999; Fernández-Fernández *et al.*, 2001; Dietrich and Maiss, 2003; Beauchemin *et al.*, 2005). However, the outmost amino terminal end of the polyprotein has never been used yet, although the green fluorescent protein has been successfully expressed close to the amino terminus of potato potyvirus A polyprotein (Rajamäki *et al.*, 2005).

Acknowledgements

We thank Verónica Aragonés for excellent technical assistance. This research was supported by grant BIO2011-26741 from the Spanish Ministerio de Economía y Competitividad (MINECO) to J.-A.D, grant PROMETEO/2010/019 from Generalitat Valenciana to S.F.E and J.-A.D., and grants BFU2012-30805 and 22371 from MINECO and the John Templeton Foundation, respectively, to S.F.E. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. E.M. was supported by a predoctoral fellowship (AP2012-3751) from the Spanish Ministerio de Educación, Cultura y Deporte. M.P.Z. was supported by a ‘Juan de la Cierva’ post-doctoral contract (JCI-2011-10379) from MINECO and a Rubicon grant from the Netherlands Organization for Scientific Research (www.nwo.nl).

References

- Beauchemin C, Bougie V, Laliberté JF. 2005. Simultaneous production of two foreign proteins from a polyvirus-based vector. *Virus Res.* 112(1-2): 1-8.
- Bedoya L, Martínez F, Rubio L, Daròs JA. 2010. Simultaneous equimolar expression of multiple proteins in plants from a disarmed potyvirus vector. *J Biotechnol.* 150(2): 268-275.
- Bedoya LC, Daròs JA. 2010. Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* 149(2): 234-240.
- Bedoya LC, Martínez F, Orzáez D, Daròs JA. 2012. Visual tracking of plant virus infection and movement using a reporter MYB transcription factor that activates anthocyanin biosynthesis. *Plant Physiol.* 158(3): 1130-1138.
- Belshaw R, Pybus OG, Rambaut A. 2007. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* 17(10): 1496-1504.
- Carrington JC, Haldeman R, Dolja VV, Restrepo-Hartwig MA. 1993. Internal cleavage and *trans*-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus in vivo. *J Virol.* 67(12): 6995-7000.
- Chen CC, Chen TC, Raja JA, Chang CA, Chen LW, Lin SS, Yeh SD. 2007. Effectiveness and stability of heterologous proteins expressed in plants by *Turnip mosaic virus* vector at five different insertion sites. *Virus Res.* 130(1-2): 210-227.
- Chirico N, Vianelli A, Belshaw R. 2010. Why genes overlap in viruses. *Proc Biol Sci.* 277(1701): 3809-3817.
- Chung BYW, Miller WA, Atkins JF, Firth AE. 2008. An overlapping essential gene in the Potyviridae. *Proc Natl Acad Sci USA.* 105(15): 5897-5902.
- Daròs JA, Schaad MC, Carrington JC. 1999. Functional analysis of the interaction between VPg-proteinase (NIa) and RNA polymerase (NIb) of tobacco etch potyvirus, using conditional and suppressor mutants. *J Virol.* 73(10): 8732-8740.

- Dietrich C, Maiss E. 2003. Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected *Nicotiana benthamiana* plants. *J Gen Virol.* 84(Pt 10): 2871-2876.
- Dolja VV, McBride HJ, Carrington JC. 1992. Tagging of plant potyvirus replication and movement by insertion of b-glucuronidase into the viral polyprotein. *Proc Natl Acad Sci USA.* 89(21): 10208-10212.
- Dolja VV, Hong J, Keller KE, Martin RR, Peremyslov VV. 1997. Suppression of potyvirus infection by coexpressed closterovirus protein. *Virology.* 234(2): 243-252.
- Fernández-Fernández MR, Mouriño M, Rivera J, Rodríguez F, Plana-Durán J, García JA. 2001. Protection of rabbits against rabbit hemorrhagic disease virus by immunization with the VP60 protein expressed in plants with a potyvirus-based vector. *Virology.* 280(2): 283-291.
- Haldeman-Cahill R, Daròs JA, Carrington JC. 1998. Secondary structures in the capsid protein coding sequence and 3' nontranslated region involved in amplification of the tobacco etch virus genome. *J Virol.* 72(5): 4072-4079.
- Kelloniemi J, Mäkinen K, Valkonen JP. 2008. Three heterologous proteins simultaneously expressed from a chimeric potyvirus: infectivity, stability and the correlation of genome and virion lengths. *Virus Res.* 135(2): 282-291.
- Kim DH, Park YS, Kim SS, Lew J, Nam HG, Choi KY. 1995. Expression, purification, and identification of a novel self-cleavage site of the N1a C-terminal 27-kDa protease of turnip mosaic potyvirus C5. *Virology.* 213(2): 517-525.
- Li XH, Carrington JC. 1995. Complementation of tobacco etch potyvirus mutants by active RNA polymerase expressed in transgenic cells. *Proc Natl Acad Sci USA.* 92(2): 457-461.
- Majer E, Daròs JA, Zwart MP. 2013. Stability and fitness impact of the visually discernible Roseal marker in the Tobacco etch virus genome. *Viruses.* 5(9): 2153-2168.

- Masuta C, Yamana T, Tacahashi Y, Uyeda I, Sato M, Ueda S, Matsumura T. 2000. Development of clover yellow vein virus as an efficient, stable gene-expression system for legume species. *Plant J.* 23(4): 539-546.
- Merits A, Rajamäki ML, Lindholm P, Runeberg-Roos P, Kekarainen T, Puustinen P, Mäkeläinen K, Valkonen JP, Saarma M. 2002. Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. *J Gen Virol.* 83(Pt 5): 1211-1221.
- Nguyen M, Haenni AL. 2003. Expression strategies of ambisense viruses. *Virus Res.* 93(2): 141-150.
- Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG. 1995. Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology.* 210(1): 194-201.
- Rajamäki ML, Kelloniemi J, Alminaitė A, Kekarainen T, Rabenstein F, Valkonen JP. 2005. A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. *Virology.* 342(1): 88-101.
- Riechmann JL, Laín S, García JA. 1992. Highlights and prospects of potyvirus molecular biology. *J Gen Virol.* 73(1-16).
- Roossinck MJ. 2011. The big unknown: plant virus biodiversity. *Curr Opin Virol.* 1(1): 63-67.
- Sánchez F, Sáez M, Lunello P, Ponz F. 2013. Plant viral elongated nanoparticles modified for log-increases of foreign peptide immunogenicity and specific antibody detection. *J Biotechnol.* 168(4): 409-415.
- Urcuqui-Inchima S, Haenni AL, Bernardi F. 2001. Potyvirus proteins: a wealth of functions. *Virus Res.* 74(1-2): 157-175.
- Whitham SA, Yamamoto ML, Carrington JC. 1999. Selectable viruses and altered susceptibility mutants in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA.* 96(2): 772-777.

Supporting information

Supplemental figure S1 Nucleotide sequences of TEV-Ros1 and TEV Δ NIb-Ros1 clones. Recombinant clones in which NIb was relocated to different intergenic positions (TEV Δ NIb-Ros1-NIb1 to 9) are like TEV Δ NIb-Ros1 but with the insertions indicated in each case. TEV sequence corresponds to Genbank accession number DQ986288, but includes two silent mutations (G273A and A1119G, in red) to eliminate endogenous Eco31I restriction sites. Cistron borders are highlighted on blue background. NIb sequence is in blue. Antirrhinum majus Ros1 cDNA is in red. In TEV-Ros1, Ros1 is flanked with sequences (in green and underlined) that complement NIaPro proteolytic sites and mediate releasing from the viral polyprotein. Amino and carboxy terminal sequences next to the relocated NIb in some of the clones (TEV Δ NIb-Ros1-NIb1, 2, 3 and 9) to complement processing from the viral polyprotein are in green and underlined.

>TEV-Ros1

```

AAAATAACAAATCTCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGCAGCAATTTAAATCATTTCTTTTAAGCA
AAAGCAATTTTCGAAAATTTTACCATTACGAAACGATAGCCATGGCACTCATCTTTGGCACAGTCAACGCTAACATCCTGAAGGAAGTGTTCGGTGGAG
CTCGTATGGCTTACCGAGCGCACATATGGCTGGAGCGAATGGAAGCATTTGAAGAAGGCAGAGA AACCTCTCGTCAATCATGCACAAACCCAGT
ATCTTCGGAGAAGACTACATTACCGAGGCAGACTTGCCCTTACACCACTCCATTTAGAGGTCGATGCTGAAATGGAGCGGATATATCTTGGTCGTCG
CGCGCTCACCATGGCAAGAGAGCGCAAAGTTTCTGTGAATAACAGAGGAACAGGAGAAGGAAAGTGCCAAAACGTCAGTGGGCGGTGATTCATTGTTG
AGAAGATTGTAGTGCCCCACACCGAGAGAAAGTTGATACCACAGCAGCAGTGGAAAGACATTTGCAATGAAGCTACCCTCAACTTGTGCATAATAGTATG
CCAAAGCGTAAGAAGCAGAAAACCTTTGCCGCCACTTCCACTAAGTAAGCTGTATGCCCAAACCTGGAGCATAGTGCAGCAACCCATATGCAGGTGGA
GATCATTAGCAAGAAGCGTCCGAGCGAGGGTCAAGAGATTTGAGGGCTCGGTGCAATTTGTCGCAAGTGTGCGTCAATGTATGGCGAGAGAAAAGGG
TGGACTTACGTATTGACAACCTGGCAGCAAGAGACACTTCTAGACCTTGCTAAAAGATTTAAGAATGAGAGAGTGGATCAATCGAAGCTCACTTTTGGTTCA
AGTGGCTAGTTTGGAGCAAGGCTCGTACGCACTCGCATTTGGTATCGCATTTGATGTTTACGCGTACGATGTTTACGCGTTCGGTGGATGGATGGTGGATGC
TCGTGCGAAGGTAACGTTGCTGTTGTGCTCACTCAATGACACATTTATAGCGCAAAATCAATCTCTGAGGCATTCTTACATACCATACTCTAAGAAAATCTTGG
AGTTGAGCCAGATGGAATCTCCCATGAGTGTACAAGAGGATATCAGTTGAGCGGTGCGGTGAGGTGGCTGCAATCCTGACACAAGCACTTTTCCACCGTGT
GGTAAGATGCAATGCAAACTGGTATGTTGAACACCTGACATTTGTTGAGGGTGGTTCGGGAGACAGTGTCCACCAACGCTGAAGTCTAGCAATGCT
GAAAGAACAGTATCCAGATTTCCCAATGGCCGAGAAAATCTACTCACAAGGTTTTTGAACAGAAAATCACTAGTAAATACAAAATTTGACAGCCTGCGTGAAGC
TCAAAACACTCATTTGTTGACCGCAAAACAGCTCCATTTCACACAGTACTGGCTGTGAGCGAAAATCTGTTTAAAGGCAATAAACTAACAGGGCCGATCTC
GAAGGCAAGCACACATATGCTTTGAAATAGCAAGGTTCTTTGAACAATCGCACATGAAAATATGCGCATTTGGCCACTTGGTCTTCTTCAGAAATAAAATCTC
ATCGAAGGCCCATGTGAATAACGCACATCATGTGTATAATCACTTGTATCAGAATGGAAATTTTATTTGGGGACTAAGGGGTGCACACGCAAGAGGTTTTT
TTAAAGGATTTTCACTGAGATTGACCAAAATGAAGGATACGATAAAGTATGTTATCAGGAAACATATCAGGGGTAGCAGAAAGCTAGCAATTTGGCAATTTG
ATAATGTCAACTGACTTCCAGACGCTCAGGCAACAAATCAAGGCAGAACTATTGAGCGTAAAGAAAATGGGAATCACTGCAATTTCAATGCGGAATGGTAA
TTACGTTGATCCCATGTTGTTGTGTTACTCTTGAAGATGGTAAGGCTCAATATTCGGATCTAAAGCATCCAACGAGAGACATCTGGTCAATGGCACTCTG
CGGATTCAAAGTACCTAGACCTTCCAGTTCTCAATGAAGGAAAATGTATATAGCTAATGAAGGTTATTTGCTACATGAACATTTTCTTTGCTCTACTAGTG
AATGTCAAGGAAGAGGATGCAAGGACTTCCACCAAGTTTATAAGGGGACACAATTTGTTCCAAAGCTTGGAGCGTGGCCAAACATGCAAGTCTACTTGAAG
ATGCTACTTACTTTCCATTTTACCAGATGTCTGAGTGTGAATTACCAGAATTTTGGTTGATCATGACAACAAAACATGCAATGTTTTGGATTCGTT
ATGGGTTAGAACGACAGGATACCACATGTTGAAAATGAACACAACTCCAGCTAATGAAATTCGTTTCAATTCAGGTTTGGAAATCCGAAATGAAAACCTTAC
AATGTTGGAGGGATGAACCCAGATATGTTCAACAAGGTGCAATTTGATGTTTATCAAGTCCATATACAAACCACATCTCATGAAGCAGTACTTGAAGGA
GGGCCATACATAAATGCTTCGGCAATAGTCTCCCCTTCAATTTTAAATGGCATGTACAACCTTGGAACTTTGAGCAGGCGTTACAAAATGTGGTTGCCAA
ATACAATGAGGTTAGCTAACCTCGCTGCCATCTTGTGAGCCTTGGCGCAAAAGTTAACTTTGGGAGACTTGTTCGTCAGCAGCGTAATTTGATTAATGAG
TATGCGCAGGTAATTTTGGCAATCTGATTGACGGTGTGAGGTTAACCATTCGCTATCCCTAGCAATGGAAAATGTTACTATTAAGCTGGCCACCCAAAGA
GATGGACATGGCGTTGAGGGAAGGTGGCTATGCTGTGACCTCTGAAAAGTGTGATGAAATGTTGGAAAAAACTATGTAAAGGCTTTGAAGGATGCATGGG
ACGAATTAACCTGGTTGGAAAAATTTCCGCAATCAGGCATTCAGAAAAGCTCTTGAATTTGGGCGAAAAGCCTTTAAATCATGAAAAACACCGTAGATTGC
GGCGGACATATAGACTTGTCTGTGAAAATCGCTTTTCAAGTTCCACTTGGAACTCCTGAAGGGAACCATCTCAAGAGCCGTAATGGTGGTCAAGAAAAGGT
AAGATGAGCGAAGATGCCATGACAAAAGGGGTTTTTCTCAAAATCTACAGCATGCTTCCCTGAGCTTACAAGTTTATCACAGTCTCGAGTGCCTTTTCT
TGTGTTGACATCTTATTTCAAATGACTGCATGATAAGGGCACACCGAGAGGCGAAGGTTGCTGCACAGTTCAGAAAAGAGAGCGAGTGGGACAATATC
ATCAAATAGAATTTCCAGTATTTAAGCTTGAATACTTATGGCTATCGCTCTACAGCGGAGGAAAAGACTCCAATCAGAACACCCCGAGGCTTTTCGAGTA
CTACAAGTTTTGCAATGGAAAGGAGACCTCGTTGAACAGGCAAAAACACCGGAGATAGCATACTTTGAAAAGATTAATAGCTTTTATCACACTTGTATTA
TGGCTTTTGGAGCTGAGCGGAGTGTGGAGTGTTCAGATACTCAATAAGTTCAAAGGAATACTGAGCTCAACGGAGAGGGAGATCATCTACACGAGAGT
TTGGATGATTACCTTACACTTTGATGACAATATGACAATCAACCTCGATTTGAAATCTGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA
GAACTCAACTTTTGGTTCATCACCGATTACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
TCATAATTGATGAAATGTCATGTGAATGATGCTTCTGCTATAGCGTTTAGGAATCTACTGTTTGAACATGAAATTTGAAGGAAAAGTCCCAAAGTGTGAGC
ACACCACGGTAGAGAAGTTGAAATTCACAACTTCCCGTTCAGAACTCAAGATAGAGAGAGGCTCTTAGCTTTACAGAAATTTGAAGTTTACAAAGGAC
AGGTGCCAACGCGGATGTGATGATTGTTGGCGCAACATACTAGTATATGTTGCTAGCTACAATGATGTTGATGATGATGATGATGATGATGATGATGATGAT
GATACAAAGTGTGGAAGATTGATGGAAGAACAATGAAGAGTGGAGGAATGAAATAATCACTGAAGGTACTTCAGTAAAAAGCAATTTTATAGTGCAGCACT
AATATTTAGAAATGGTGAAGTTGAAATTCACAACTTCCCGTTCAGAACTTTGGGACTTAAGGTTGATACAGGTTTGGATGAGCAATAAGAGCGGTGCAGTA
CAACAAAATCTGGTGTGATTTATGGGGAGCGCATCCAAGACTCGGTAGAGTTGGGCGCAACAAGGAGGAGTAGCACCTTCAAAATTTGGCCAAACAAATAAAA
CACTGTTGAAATTCAGAAATGGTGGCCACTGAAGTGCCTTTCTATGCTTCATGTACAATTTGCCAGTGCACACAGAGTGTTCACCCACACTGCTG
GAAAATGCCAATTTTACAAAGCTAGAACTTAGCCACAGTTTAGCATATATTTTACACAATTAATTTTGGCGATTGATGGTAGTATGCATCCAGT

```

CATACATGACAAGCTGAAGCGCTTTAAGCTACACACTTGTGAGACATTCCTCAATAAGTTGGCGATCCCAAATAAAGGCTTATCCTCTTGCCCTACGAGTG
GAGATATAAGGACTTGGTTACATAGCAGAGGATGCTGGCATAAAGATCCCAATTCGTGTGCAAGAATAATCCAGACTTCCTGATGAGGAAATTTGGCAC
ATTGTAGTCGCCATAAAGTGACTCGGGTATTGGGAGGCTACTAGCGTACAGGCAGCAAAGGTTGTTTATACCTGCAAAACGGATGTGCACTCAATTGC
GAGGACTCTAGCATGCATCAATGACTATAGCAGACATGAACAATAAGAGAGGACTTCTTTGAGCCGCAACTGGGAGGACATTTCCCTCAAAAATTA
CAATACAAGCAATTTGACACCGCTGAAAGCAAAATTTGCTCAAAAACATAGAAAGAAATATGTCAGTCTTCAGCAGGCAAAATTAATCAATTAAGT
TTTTGAACTTAGCAAGGATCAAGATGTACCGGATATCACTCAAGACTCACTACCTGGAAACTATCTATCTCAATCAGATAGCGAAGTGGCTAAGCA
TCTGAAGCTTAAAGTCACTGGAATAAAGCCAAATCACTAGGGACATCAATAAGCTTGTCTGTGTTAATTGGTGGTGGATGCTTGCACACCTACT
TCAAGGCAAGTTCAATGAACCGTCTATTTCCAAAGGGAAGAAATCAGAAGCACAAGCTTAAGATGAGAGAGGCGCTGGGGCTAGAGGGCAATATGAG
GTTGCAGCGGAGCCAGAGCGCTAGAACATTACTTTGGAAGCGCATATAATAACAAGGAAAGCGCAAGGGCCACCAGAGGAAATGGGTGCAAACTCTCG
GAAATTCATAAAACATGTATGGTTTGATCCAACCTGATTTTTCATACATTAAGTTTGGATTCATTGACAGGTACACTATTTGATGAGTCCACAACCCAC
CTATTGATTTAGTGCAGCATGATTTGGAAAGTTGAAACACGCTGTAAATTGACGATGAGATAGAGCTCAAAGCTTAGCACCCACACCAATCCAT
GCTTATTTGGTGAATAGTGGCACGAAGAAGTCTTAAAGTTGATTTAACACCACACTCGTCCGTACGTGCGAGTGAGAAATCAACAGCAATAATGGGAT
TCCTGAAAGGAGGAATGAATGCTCAAAACCGGATGGCAGTGCAGTGGCTTATGATCAATTTGCCACAAAGAGTGGAGACTTGACGTTTGAAGGAGAAA
GCTTGTTTAAGGGACCACGTGATTACAAACCCGATGGCAGTGCAGTGGCTTATGATCAATTTGCCACAAAGAGTGGAGACTTGACGTTTGAAGGAGAAA
TTTGGTCCCTTCATATTAACAACAGCCTTTTGAAGAAATAATGGAACACTGTTGTTCATCACTACATGGTGTATTCAGGTCGAAGAACCAC
GACTTTGCAACAACACCTTATGATGGGAGGACATGATAAATTTGCGATGCCTAAGGATTTCCACCATTTCCTCAAAAAGTGAATTTAGAGAGCCAC
AAAGGGAAGGCGCATATGCTTTGTGACAAACCAACTTCAAACTAAAGAGATGTCAGCATGGTGTGACACACTAGTTGACACTTCACTGATGGC
ATATTCGGAAGCATTGGATTCAAACCAAGGATGGGCAGTGGCAGTCCATTAGTATCAACTAGAGATGGGTTCATTTGGTGTATCACTCAGCATCGAA
TTTTCAACAACAACAATTTTCAACAAGCGTGGCAGAAACTTCATGGAATTTGTGACAAATCAGGAGGCGCAGCAGTGGGTTAGTGGTGGCGATTAA
ATGCTGACTCAGTATTGTGGGGGGCCATAAAGTTTTCATGAGCAAACTGGAAGAGCCTTTTCCAGCAGTTAAGGAAGCGACTCAACTCACTGATGGC
GTGTACTCGCAAGGGGAGAGGAAATGGTCTGGAAGCACGTGTCAGGGAATTCAGGGCCAGTGGCTGAGTGTCCAGTCACTTAGTCACAAAGCATGT
GGTTAAAGGAAAGTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAAGGAGCATATTTTAAACCGATGATGGGAGCATATAAGCCAAAGTCGAC
TTAATAGAGAGGCGTTCCTCAAGGACATTTAAAATGCTAGTGAATTTGAGATGGAAATGAGATTGGAAATGAGATTGTGACTTCGCAATAGGATTTGCT
ATCACAAGCTCAAGCGTTAGGATTTCCAACTGGAATACATCACTGACCCAGAGGAAATTTTAGTGATTTGAATATGAAAGCAGCTATGGGAGCACT
ATACAAGGCAAGAAAGAGCTCTCAGCGACTCACACTAGATAGGAGGCAAGTCTCAAGCAAGTTGCCGCGACTATACGGGAAAGCTG
GAATTTGAAATGCTCATTTAAAGGACATTTCCGTCATTTGAAAGGTTGAAAACAACAAACCGCAACTTTTACAGCAGCACCATAAGACACTTTCT
GCTGGTAAAGTTTGGCTGGATGATTTCAACAATCAATTTTATGATTCACAATAAAGGCACCAATGGACAGTTGGTATGACTAAGTTTATCAGGGGTGGAA
TGAATTTGATGGAGGTTTACCAAGTGGTGGTGTATTGTGACGCTGATGGTTCGAATTCGACAGTTCCCTGACTCCATCTTAATGCTGTATTG
AAGTCGACTTGCCTTCAAGGAAATGGATTTGGTGGAGAAATGGTGGCAAACTTTGACACTGAGATGTATACACCAACTTCAACCGGATGGT
ACTATCATTAAAGAGCAATAAGGCAACAATAGCGGGCAACCTTCAACAGTGGTGGCAACACACTCATGGTCATTTATGCAATGTTATACACATGTGAGAA
GTGTGAATCAACAAGGAAGAGATTTGTGATTTACGTCAATGGCCATGACCTAAGTGGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGATTTCAAAG
ATACTTTTCGAGAGTTGGGCTGAAATGAAATTTGACTGCACCCAGCACAAGACACAGTTGGTTCATGTACACAGGTTGGAGAGGAGGATGGC
ATGATATACCAAGCTTAGAAGAAAGAGGATTTGTTCTATTTTGAATTTGCAAGCAGATCCAAAGAGCGCTCACATAGGCTTGAAGCCATTCGTGATCAAT
GATCGAAGCATGGGTTGATGACAAGCTGTGAAGAAATCCGCAATTTTGAACATTTGATGCAATGGTTTTGAACAGCGCCGATTTACAGCTTGCAGAGGAA
AGCGCCATATCTGGCTGAGACTGGCTTAAAGTTTGTACACATCTCAGCAGCAACAACCTCTGAGATAGAAGAGTATTTTAAAGTGGTTGATGATTTAC
GATATTTCCACGACTGAGAACTTTATTTTTCAGTGGTCACTGGAAGAAATTTGCTGGAGTGAAGAAAGGTTGGAACCAAGAAAGCACTCT
CTTGAGGCAATGATAGAAGACTTGGTGAAGGAAATGGCATCAAGTTCCACACAGCAGGCTTGAACCGGTTGAGAAAGCTTCAAGCTGAGGTTG
TGAATTTCTGAGGCCAAATATCAAAGAGGTCGGTTTTCGAGAGATGAAGTGGACCTAATTTGTGAGGCTTCATAAGCTGTTGGTAAACAATGGTCCGCTG
ATTGCTGGTGAATTCCTGGAAGGACAGCTTAATGACGTGAAGAAATTTGGAATCTCATGTTGGGGAAGAAATTTAGCGGAGGATGGAGAACGATGCGGGAA
AAATGCTTATGAACAACAACAACCAATTAAGCTGACTAATATCGTAAGACCCGAGCTCCGACCTTCCACCGGATTTGCACGTTACTTGGCCGAGAGGATGGAA
AAACCGATGAAATTTCAAATGTCGGTTTAAACACTGATGAGATTCAGATTGTGAGAAGCAACCGCAATTTTACAATGATGTTGCGTCCGCAACAGATGAA
TTGTAAGACTGCAATTCAGTGGTGGAGTAAAGTTGCTAGAAAACAAGGAGGATGGGAAATTTAGGAAACCTATTCGAGGAGGCCCAAAATTTGGAATCACTAC
AGAGAACCTCTACTTCTCAAAGTGGCACTGTGGTCTGGTGTTCGCTGTAAGAAAGAAAGATCAAAGGATGATAAAGCTGCTGAGCAGCTTCAAGG
ATAGGATGTTAATGCTGGAACCTCAGGAACATTCTCAGTTCCAGCAATAAATGCTATGGCCACAAAATCTCAATATCCAAGGATGAGGGGAGAGGTTGGT
GTAAATCTGAAATCACTTTTAGATACAAGCCACAGCAAAATGATTTGTTCAAAATGCTGCGAGCCACACATGAGCAGTTTGGCCGCTGGGATCAGGCAAGTAT
GACAGCCTATGGAGTGAATGAAGAGCAAAATGAAATATTTGCTAATAGGATTTTGGTGTGGTGCATAGAATAATGGGACTTCCCAAAATTTGAACGGAAT
GGGTTATGATGGATGTTGAGGAGCAAGTTTTCATACCGCTGAAACCAATGGTTGAAAACCGCGAGCCAAACTGAGGCAAAATTAGACACACTTCACTGAC
CTGGCTGAAGCTTATGATGATGAGGAAATAGGGAGCGACCATCACTTAGTATGTTGCTACAGAGAAACATACAGACATGATTTGTCACGCTATGCT
GTTTCGACTTCTAGGCTAATCTCAAACACCTGTTTAGAGCGGAGGCGCATATGCAAAATGAAAGCTGCTGACAGTGAACCAAGCTGAGCAATGAGTTG
TTGGTCTGATGGCAACGTTGGTACTGACAGGAGACACTGAACGGCACACAGCCGACGATGTGAACCGTAACTGCACACACTATTAGGGTCCGCCAG
TGATAGTTTCTCGGTGCTTTGCTTTCCGCTTAAAGCTTATTTGATATATATGAAATAGCTATTCACAGTGGGACTTGGTCTTGTGTTGAATGATATCTT
ATATGTTTAAATATGCTTATTTAGTCTCAATTTCTTAGCGAACGACAAAGTGAAGTCCCTCGGTCATATTTCTCCTATGTAGTGGAGAAAACAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>TEVΔN1b-Ros1

AAAAAACAATCTCAACACAATATACAAAACAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGACAGCAATTTAAATCAATTTCTTTTAAAGCA
AAAGCAATTTCTGAAATTTTACCATTTCAGAACGATAGCCATGGCCTCACTTTTGGCACAGTCAACGCTAACCTCTGGAAGGAAGTGTTCGGTGGAG
CTCGTATGGCTTGGCTTACCAGCGCACATATGGCTGGAGCGAATGGAAGCAATTTTGAAGAGGCGAGAAGA AACCTCTCGTCAATCATGCACAAACAGT
ATCTTCGGAGAAAGACTACATTTCCAGGCGAGACTTGGCTTACACACCACTCAATTTAGAGGTGATGCTGAAATGGAGCGGATGATATATCTTGGTCTGCG
CGCGCTCACCCTAGGCAAGAGAGCAAAAGTTTCTGTGAATAACAAAGAGAAACAGGAGAAGGAAAGTGGCCAAAACGCTACGTTGGGCGGTGATTTCCATTTG
ACAAAGATTTAGTGGCCACACCGAGAGAAGGTTGATACCACAGCAGCAGTGAAGACATTTGCAATGAAAGTACCCTCAACTGTGCAATAATAGTATG
CBAAGCGTAAGAAGACAAAACCTTCTTCCGCCCACTTCACTAAGTAACTGATAGTTCGCCCAAATTTGGAGCATAGTGCACAAACGATGCGAGTGA
GATCATTAGCAAGAGAGGCTCCGAGCGAGGCTCAAGAGATTTGAGGGCTCGTGAATTTGTCGAAGTGTGGCTCACATGATGGCGAGAGGAAAGGG
TGGACTTACGATTGCAACTGGCAGCAAGAGACACTTCTAGACCTTGCTAAAAGATTTAAGAATGAGAGAGTGGATCAATCGAAGCTCACTTTTGGTTCA
AGTGGCTAGTTTTGAGGCAAGCTCGTACGGACCTGCGCATGGTATCGACATGGTATGTTTCAATTTGACCGGCTCGGTGGATGGGATGTTGGTGGATGC
TCGTGGAGAGGTAACGTTTCGCTGTTGCTCACTCAATGACACATTTAGCGACAAATCAATCTCTGAGGCACTTCTTACATCACTACTTAAAGAAATCTTGG
AGTTGAGCCAGATGGAATCTCCCATGAGTGTACAAGAGGAGTATCAGTTGAGCGGTGCGGTGAGGTGGCTGCARTCTGACACAAGCACTTCCACCGT
GGTAAAGATCACATGCAACCGTTGCTGTTGAAACACCTGACATTTGTTGAGGGTGAAGTGGGAGACAGTGTCAACCAACCAAGGTAAAGCTTACGAATGCT
GAAAGAACAGTATCCAGATTTCCCAATGGCCGAGAACTACTACAAGGTTTTTGAACAGAAATCACTAGTAAATACAAATTTGACAGCCTGGGTGAGCG
TCAAAACACTCATTTGGTGCAGCCAAACAGCTCCATTCACACACGACTGGCTGTGAGGAAATTCGTTTAAAGGCAATAACTAACAGGGCCGATCTC
GAAGAGGCAAGCACACATATGCTTGAATAGCAAGGTTCTTGAACAATCGCACTGAAAATATGCGCATTTGCCACCTTGGTTCTTTCAGAAAATAAAATCTC
ATCGAAGGCCATGTAATCAACGACTCATGTTGATAATCACTTGAATGAGTGGGAAATTTTATTTGGGACTAAGGGGTGCACAGCAAGAGGTTTTC
TTAAAGGATTTTCACTGAGATGACCCAAATGAAGGATACGATAAGTATGTTATCAGAAACATATCAGGGGTAGCAGAACTGACAAATTTGGCAATTTG
ATAATGTTCAACTGACTTCCAGCGCTCAGGCAACAAATCAAGGCAAACTATTGAGCGTAAAGAAATTTGGGAAATCACTGCAATTTCAATGCGGAATGGTAA
TTACGTTCCCTCATTTGTTGTTACTCTTGAAGATGTTAAGGCTCAATTTGCGATTTAAAGCACTCAACAGAGACATCTGGTCAATTTGGCAACTCTG
CGATTTCAAAGTCACTGAGCTTCCAGTCTCAATGAAGGAAATGTATATAGCTAATGAGGTTATTGCTACATGAACTTTTCTTGTCTACTAGT
AATGTCAAGGAAAGGAGTGAACAAGGACTTCAACAAGTTTATAAGGGACCAAAATTTGTTCAAAGCTTGGAGCGTGGCCAAACATGCAAGATTTGCAACTGC
ATGCTACTTCTTCAATCTTTACCAGATGCTCCTGAGTGTCAAAATACCAGAAATTTGGTGTATGATGACAAACAAACAAATGCAATGTTTGGATTCGT
ATGGTCTAGAAGCAGGATACACATCTTGAAGATGAACACAACTCCAGCTAATTTGAATTCGTTCACTTACGTTTGAATCCGAAATGAAACATTTAC
AATGTTGGAGGATGAACCGAGATATGGTACACAAGGTCGAATTTGAGATGTTGATCAAGTCCATATACAAACCACTCTCATGAAGCAGTTACTTTGAGGA
GAGGCCATACATAATTTGCTCGCTGCAATAGTCTCCCTTCAATTTTAAATGCAAGTGAACAACCTCTGGAACCTTTTGGAGCCGCTTCAAAATGTTGGTGGCA
ATACATATGAGTTAGCTTAACTGCTGCTTGTGAGCTTGGCCACTGGCCAGGTTAACTTTGGCAGACTTGTGCTGAGCAGGTAATTTGATTTGAGG
TATGCGCAGGTAATTTTGGACAATCTGATTTGACGGTGTGAGGTTAACCATTCGCTATCCCTAGCAATGGAATTTGTTACTATTAAGCTGGCCACCCAAAG
GATGGACATGGCTTGAAGGAAAGGTTGGCTATGCTGTGACCTCTGAAAAGTGCATGAAAATTTGGAAAAAACTGTTAAAGGCTTTGAAAGGCTGGAAGG
ACGAAATTACTTGGTTGGAAAAATTTCCGCAATCAGGACTTCAAGAAAGCTTTGAAATTTGGCCGAAAGCCCTTAAATGATGAAAACCCGATGATTTGC
GGCGGACATATAGACTTGTCTGTGAAATCGCTTTTCAAGTTCCTTGGAACTCTGAAAGGAAACCTCTCAAGAGCCGTAATGGTGGTGAAGAAAGG
AAGAGTAGCGAAGTCCATGACAAAAGGGGTTTTCTCAAATCTACAGCATGCTTCTGACGCTTACAAGTTTATCAGACTTCGAGTCTCTCTTCT

TGGAATCAACAAGGAAGAGATTGTGTATTACGTCAATGGCGATGACCTATTGATTGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGATTCAAAGAAT
 CTTTCGGAGAGTTGGGCCTGAAATATGAAATTTGACTGCACCACAGGGAACAAGACACAGTTGTGGTTCATGTACACAGGCTTTGGAGGGGATGGCATG
 TATATACCAAGCTAGAGAAGAAGAGGATTTGTTCTATTTTGGAAATGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGCATCAATGAT
 CGAAGCATGGGCTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGCATGGGTTTGGAAACAAGCCCGTATTACAGCTTGCAGAGAAGGAAAGG
 CGCCATATCTGGCTGAGACTGGCCTTAAGTTTGTACACATCTCAGCAGGGAACAACCTCTGAGATAGAAGAGTATTTAAAGTGTGTATGATTACGAT
 ATTCCAACGACTGAGAATCTTTATTTTCAG

>TEVANIb-Ros1-Nib3 (NIb insertion between positions 2433-2434 of TEVANIb-Ros1; NIb included the indicated extra nucleotides)

GGGATGAACGGGAGAGAAGGAAATGGTTCGTGGAGCACTGTGAGGCACTGGTGTGAGTGTCCAGTCAGTTAGTCACAAAGCATGTGGT
 TAAAGGAAAGTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAAGGAAGCATATTTAAACCGATGATGGGAGCATATAAGCCAAGTCGACTTA
 ATAGAGAGGCTTCCCTCAAGGAAATCTAAAATATGCTAGTGAATTTGAGATTGGGAATGTGGATTGTGACTTGTCTGGAGCTTGCATAAAGCATGCTCATC
 ACAAGCTCAAGGCGTTAGGATTTCCCACTGTGAATACATCACTGACCCAGAGGAAATTTTGTAGTCATTGAATATGAAAGCAGCTATGGGAGCATATA
 CAAAGCCAGAAGAAGAAGCTCTCAGCGACTCACACTAGATGAGCAGGAGGCAATGTCTCAAAGCAAGTTGCCTGGCAGTGTATACGGGAAAGCTGGGAA
 TTTGGAAATGGCTCATTGAAAGCAGATTGGTCCAAATTTGAGAAGGTTGAAAACAACAAGGCAAGCTTTACACTGAGATAGTGTATACACCAATCTTCACAGGAGCTTGCCTTGTCT
 GGTAAAGTTTGGCTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGCACCATGGACAGTTGGTATGACTAAGTTTATCAGGGGTGGAATGA
 ATTGATGGAGGCTTTACCAAGTGGTGGTGTATTGTGACGCTGATGGTTCGCAATTCGACAGTTCCCTGACTCCATTCCTATTATGTCTGTATTGAAAG
 TCGACTTGGCTTCAAGGAAATGGGATTTGGTGGCAAAATGCTCCGAAATTTGACACTGAGATAGTGTATACACCAATCTTCACAGGAGCTTGCCTTGTCT
 ATCATTAAGAGCATAAAGGCAACAATAGCGGGCAACCTTCAACAGTGGTGGACAAACACTCATGGTCATTATGCAATGTATACACATGTGAGAAGTG
 TGGATCAACAAGGAAGAGATTGTGATTACGTCATGGCGATGACCTATTGATTGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGATTCAAAGAAT
 CTTTCGGAGAGTTGGGCTGAAATATGAAATTTGACTGCACCACAGGGAACAAGACACAGTTGTGGTTCATGTACACAGGCTTTGGAGAGGATGGGATG
 TATATACCAAGCTAGAGAAGAAGAGGATTTGTTCTATTTTGGAAATGGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGCATCAATGAT
 CGAAGCATGGGCTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGCATGGGTTTGGAAACAAGCCCGTATTACAGCTTGCAGAGAAGGAAAGG
 CGCCATATCTGGCTGAGACTGGCCTTAAGTTTGTACACATCTCAGCAGGGAACAACCTCTGAGATAGAAGAGTATTTAAAGTGTGTATGATTACGAT
 ATTCCAACGACTGAGAATCTTTATTTTCAG

>TEVANIb-Ros1-Nib4 (NIb insertion positions 3474-3475 of TEVANIb-Ros1)

GGGAGAGAAGGAAATGGTTCGTGGAGCACTGTGAGGCACTGGTGTGAGTGTCCAGTCAGTTAGTCACAAAGCATGTGGTTAAAGGAAA
 GTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAAGGAAGCATATTTAAACCGATGATGGGAGCATATAAGCCAAGTCGACTTAATAGAGAGG
 CGTTCTCAAGGCAATCTAAAATATGCTAGTGAATTTGAGATTGGGAATGTGGATTGTGACTTGTCTGGAGCTTGCATAAAGCATGCTCATCACAAGCTC
 AAGGCGTTAGGATTTCCCACTGTGAATACATCACTGACCCAGAGGAAATTTTGTAGTCATTGAATATGAAAGCAGCTATGGGAGCACTATACAAAGCCAA
 GAAGAAGAAGCTCTCAGCGAGCTCACACTAGATGAGCAGGAGGCAATGCTCAAAGCAAGTTGCCTGCGACTGTATACGGGAAAGCTGGGAATTTGGAATG
 GCTCATTTGAAGCAGATTGGCTCCAATTTGAGAAGGTTGAAAACAACAAGGCAAGCTTTACAGCAGCACCATAAGCACTCTTCTGTGGTAAAGTT
 TGGCTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGCACCATGGACAGTTGGTATGACTAAGTTTATCAGGGGTGGAATGAATGATGGA
 GGCTTTACCAAGTGGTGGTGTATTGTGACGCTGATGGTTCGCAATTCGACAGTTCCCTGACTCCATTCCTATTATGCTGTATTGAAAGTGGCAGCTTG
 CCTTCATGGAGGAATGGGATTTGGTGGACAAATGCTGGCAAAATTTGTACACTGAGATAGTGTATACACCAATCTCAGCCGATGGTACTATCAATTAAG
 AAGCATAAAGGCAACAATAGCGGCAACCTTCAACAGTGGTGGACAACACTCATGGTCATTATTGCAATGTATACACATGTGAGAAGTGTGGAATCAA
 CAAGGAAGAGATTGTGATTACGTCATGGCGATGACCTATTGATTGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGATTCAAAGAATCTTTTCGGAG
 AGTTGGGCTTGAATATGAAATTTGACTGCACCACAGGGAACAAGACACAGTTGGTTCATGTACACAGGCTTTGGAGAGGATGGCATGTATATACCA
 AAGCTAGAGAAGAAGGATTTGTTCTATTTTGGAAATGGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGCATCAATGATCGAAGCATG
 GGTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGCAATGGGTTTGGAAACAAGCCCGTATTACAGCTTGCAGAGAAGGAAAGGCGCCATATC
 TGCTGAGACTGGCCTTAAGTTTGTACACATCTCAGCAGGGAACAACCTCTGAGATAGAAGAGTATTTAAAGTGTGTATGATTACGATATTCCAACG
 ACTGAGAATCTTTATTTTCAG

>TEVANIb-Ros1-Nib5 (NIb insertion between positions 3633-3634 of TEVANIb-Ros1)

GGGAGAGAAGGAAATGGTTCGTGGAGCACTGTGAGGCACTGGTGTGAGTGTCCAGTCAGTTAGTCACAAAGCATGTGGTTAAAGGAAA
 GTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAAGGAAGCATATTTAAACCGATGATGGGAGCATATAAGCCAAGTCGACTTAATAGAGAGG
 CGTTCTCAAGGCAATCTAAAATATGCTAGTGAATTTGAGATTGGGAATGTGGATTGTGACTTGTCTGGAGCTTGCATAAAGCATGCTCATCACAAGCTC
 AAGGCGTTAGGATTTCCCACTGTGAATACATCACTGACCCAGAGGAAATTTTGTAGTCATTGAATATGAAAGCAGCTATGGGAGCACTATACAAAGCCAA
 GAAGAAGAAGCTCTCAGCGAGCTCACACTAGATGAGCAGGAGGCAATGCTCAAAGCAAGTTGCCTGCGACTGTATACGGGAAAGCTGGGAATTTGGAATG
 GCTCATTTGAAGCAGAGTTGGCTCCAATTTGAGAAGGTTGAAAACAACAAGGCAAGCTTTACAGCAGCACCATAAGCACTCTTCTGTGGTAAAGTT
 TGGCTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGCACCATGGACAGTTGGTATGACTAAGTTTATCAGGGGTGGAATGAATGATGGA
 GGCTTTACCAAGTGGTGGTGTATTGTGACGCTGATGGTTCGCAATTCGACAGTTCCCTGACTCCATTCCTATTATGCTGTATTGAAAGTGGCAGCTTG
 CCTTCATGGAGGAATGGGATTTGGTGGACAAATGCTGGCAAAATTTGTACACTGAGATAGTGTATACACCAATCTCAGCCGATGGTACTATCAATTAAG
 AAGCATAAAGGCAACAATAGCGGCAACCTTCAACAGTGGTGGACAACACTCATGGTCATTATTGCAATGTATACACCAATCTCAGCCGATGGTACTATCA
 CAAGGAAGAGATTGTGATTACGTCATGGCGATGACCTATTGATTGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGATTCAAAGAATCTTTTCGGAG
 AGTTGGGCTTGAATATGAAATTTGACTGCACCACAGGGAACAAGACACAGTTGGTTCATGTACACAGGCTTTGGAGAGGATGGCATGTATATACCA
 AAGCTAGAGAAGAAGGATTTGTTCTATTTTGGAAATGGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGCATCAATGATCGAAGCATG
 GGTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGCAATGGGTTTGGAAACAAGCCCGTATTACAGCTTGCAGAGAAGGAAAGGCGCCATATC
 TGCTGAGACTGGCCTTAAGTTTGTACACATCTCAGCAGGGAACAACCTCTGAGATAGAAGAGTATTTAAAGTGTGTATGATTACGATATTCCAACG
 ACTGAGAATCTTTATTTTCAG

>TEVANIb-Ros1-Nib6 (NIb insertion between positions 5532-5533 of TEVANIb-Ros1)

GGGAGAGAAGGAAATGGTTCGTGGAGCACTGTGAGGCACTGGTGTGAGTGTCCAGTCAGTTAGTCACAAAGCATGTGGTTAAAGGAAA
 GTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAAGGAAGCATATTTAAACCGATGATGGGAGCATATAAGCCAAGTCGACTTAATAGAGAGG
 CGTTCTCAAGGCAATCTAAAATATGCTAGTGAATTTGAGATTGGGAATGTGGATTGTGACTTGTCTGGAGCTTGCATAAAGCATGCTCATCACAAGCTC
 AAGGCGTTAGGATTTCCCACTGTGAATACATCACTGACCCAGAGGAAATTTTGTAGTCATTGAATATGAAAGCAGCTATGGGAGCACTATACAAAGCCAA
 GAAGAAGAAGCTCTCAGCGAGCTCACACTAGATGAGCAGGAGGCAATGCTCAAAGCAAGTTGCCTGCGACTGTATACGGGAAAGCTGGGAATTTGGAATG
 GCTCATTTGAAGCAGAGTTGGCTCCAATTTGAGAAGGTTGAAAACAACAAGGCAAGCTTTACAGCAGCACCATAAGCACTCTTCTGTGGTAAAGTT
 TGGCTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGCACCATGGACAGTTGGTATGACTAAGTTTATCAGGGGTGGAATGAATGATGGA
 GGCTTTACCAAGTGGTGGTGTATTGTGACGCTGATGGTTCGCAATTCGACAGTTCCCTGACTCCATTCCTATTATGCTGTATTGAAAGTGGCAGCTTG
 CCTTCATGGAGGAATGGGATTTGGTGGACAAATGCTGGCAAAATTTGTACACTGAGATAGTGTATACACCAATCTCAGCCGATGGTACTATCAATTAAG
 AAGCATAAAGGCAACAATAGCGGCAACCTTCAACAGTGGTGGACAACACTCATGGTCATTATTGCAATGTATACACCAATCTCAGCCGATGGTACTATCA
 CAAGGAAGAGATTGTGATTACGTCATGGCGATGACCTATTGATTGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGATTCAAAGAATCTTTTCGGAG
 AGTTGGGCTTGAATATGAAATTTGACTGCACCACAGGGAACAAGACACAGTTGGTTCATGTACACAGGCTTTGGAGAGGATGGCATGTATATACCA
 AAGCTAGAGAAGAAGGATTTGTTCTATTTTGGAAATGGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGCATCAATGATCGAAGCATG
 GGTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGCAATGGGTTTGGAAACAAGCCCGTATTACAGCTTGCAGAGAAGGAAAGGCGCCATATC
 TGCTGAGACTGGCCTTAAGTTTGTACACATCTCAGCAGGGAACAACCTCTGAGATAGAAGAGTATTTAAAGTGTGTATGATTACGATATTCCAACG
 ACTGAGAATCTTTATTTTCAG

>TEVANIb-Ros1-Nib7 (NIb insertion between positions 5691-5692 of TEVANIb-Ros1)

GGGAGAGAAGGAAATGGTTCGTGGAGCACTGTGAGGCACTGGTGTGAGTGTCCAGTCAGTTAGTCACAAAGCATGTGGTTAAAGGAAA
 GTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAAGGAAGCATATTTAAACCGATGATGGGAGCATATAAGCCAAGTCGACTTAATAGAGAGG
 CGTTCTCAAGGCAATCTAAAATATGCTAGTGAATTTGAGATTGGGAATGTGGATTGTGACTTGTCTGGAGCTTGCATAAAGCATGCTCATCACAAGCTC
 AAGGCGTTAGGATTTCCCACTGTGAATACATCACTGACCCAGAGGAAATTTTGTAGTCATTGAATATGAAAGCAGCTATGGGAGCACTATACAAAGCCAA
 GAAGAAGAAGCTCTCAGCGAGCTCACACTAGATGAGCAGGAGGCAATGCTCAAAGCAAGTTGCCTGCGACTGTATACGGGAAAGCTGGGAATTTGGAATG
 GCTCATTTGAAGCAGAGTTGGCTCCAATTTGAGAAGGTTGAAAACAACAAGGCAAGCTTTACAGCAGCACCATAAGCACTCTTCTGTGGTAAAGTT
 TGGCTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGCACCATGGACAGTTGGTATGACTAAGTTTATCAGGGGTGGAATGAATGATGGA
 GGCTTTACCAAGTGGTGGTGTATTGTGACGCTGATGGTTCGCAATTCGACAGTTCCCTGACTCCATTCCTATTATGCTGTATTGAAAGTGGCAGCTTG
 CCTTCATGGAGGAATGGGATTTGGTGGACAAATGCTGGCAAAATTTGTACACTGAGATAGTGTATACACCAATCTCAGCCGATGGTACTATCAATTAAG
 AAGCATAAAGGCAACAATAGCGGCAACCTTCAACAGTGGTGGACAACACTCATGGTCATTATTGCAATGTATACACCAATCTCAGCCGATGGTACTATCA
 CAAGGAAGAGATTGTGATTACGTCATGGCGATGACCTATTGATTGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGATTCAAAGAATCTTTTCGGAG
 AGTTGGGCTTGAATATGAAATTTGACTGCACCACAGGGAACAAGACACAGTTGGTTCATGTACACAGGCTTTGGAGAGGATGGCATGTATATACCA
 AAGCTAGAGAAGAAGGATTTGTTCTATTTTGGAAATGGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGCATCAATGATCGAAGCATG
 GGTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGCAATGGGTTTGGAAACAAGCCCGTATTACAGCTTGCAGAGAAGGAAAGGCGCCATATC
 TGCTGAGACTGGCCTTAAGTTTGTACACATCTCAGCAGGGAACAACCTCTGAGATAGAAGAGTATTTAAAGTGTGTATGATTACGATATTCCAACG
 ACTGAGAATCTTTATTTTCAG

GCTCATGAAAGCAGAGTTGCGTCCAATTGAGAAGGTTGAAAAACAACAAAACGCGAACTTTCACAGCAGCACCATAAGACACTCTTCTTGCTGGTAAAGTT
 TGGTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGCACCATGGACAGTTGGTATGACTAAGTTTATCAGGGGTGGAAATGAATTGATGGA
 GGCTTTACCAAGTGGTGGGTGATTTGTGACGCTGATGGTTCGCAATTCGACAGTTCCTTGACTCCATTCTCATTAAATGCTGTATTGAAAGTGGCAGCTTG
 CCTTCATGGAGGAATGGGATATTGGTGGAGCAAAATGCTGCGAAATTTGTACTGAGATAGTGTATACACCAATCCTCACACCGGATGGTACTATCATTAAAG
 AAGCATAAAGGCAACAATAGCGGGCAACCTTCAACAGTGGTGGACAACACACTCATGGTCATTATTGCAATGTTATACACATGTGAGAAGTGGGAATCAA
 CAAGGAAGAGATTGTGATACGTCAATGGCGATGACCTATTGATTGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGATTCAAAGAATCTTTCGGAG
 AGTTGGCCCTGAAATATGAATTTGACTGCACCACCAGGGACAAGACACAGTTGGTTCATGTACACAGGGCTTTGGAGAGGGATGGCATGTATATACCA
 AAGCTAGAAGAAGCAAGGATTTGTTCTATTGGAATGGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGCATCAATGATCGAAGCATG
 GGGTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGATGGTGGTGGAAACAAGCCGCTATTTCACAGCTTGACAGAAAGGAAGGGCCATATC
 TGGCTGAGACTGCGCTTAAAGTTTTGTTACACATCTCAGCAGGAAACAACTCTGAGATAGAAGAGTATTTAAAAGTGGTGTATGATTACGATATTTCCAACG
 ACTGAGAATCTTTATTTTCAG

>TEVANib-Ros1-Nib8 (Nib insertion between positions 6255-6256 of TEVANib-Ros1)

GGGGAGAAGAGGAAATGGGTCGTTGGAAGCACTGTCAGGGAACTTGAGGCCAGTGGCTGAGTGTCCAGTCAAGTGTAGTCACAAAGCATGTGGTTAAAGGAAA
 GTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAAGGAAGCATAATTTAAACCGATGATGGGAGCATATAAGCCAAGTCGACTTAATAGAGAGG
 CGTTCCTCAAGGACATTTCAAAAATATGCTAGTGAATTTGAGATTGGGAATGTGGATTGTGACTTGTCTGGAGCTTGAATAAGCATGCTCATCAAAAGCTC
 AAGCGCTTAGGATTTCCCACTGTGAACATACACTGACCCAGAGGAAATTTTGTAGTGCATTGAATATGAAAGCAGCATATGGGAGCACTATACAAAGGCAA
 GAAGAAGAAGCTTCCAGCGAGCTCACACTAGATGAGCAGGAGGCAATGCTCAAAGCAAGTTGCCCTGCGACTGTATACGGGAAAGCTGGGAATTTGGAATG
 GCTCATGAAAGCAGAGTTGCGTCCAATTGAGAAGGTTGAAAAACAACAAAACGCGAACTTTCACAGCAGCACCATAAGACACTCTTCTTGCTGGTAAAGTT
 TGGTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGCACCATGGACAGTTGGTATGACTAAGTTTATCAGGGGTGGAAATGAATTGATGGA
 GCCTTTACCAAGTGGTGGGTGATTTGACGCTGATGGTTCGCAATTCGACAGTTCCCTTGACTCCATTCTCATTAAATGCTGTATTGAAAGTGGCAGCTTG
 CCTTCATGGAGGAATGGGATATTGGTGGAGCAAAATGCTGCGAAATTTGTACTGAGATAGTGTATACACCAATCCTCACACCGGATGGTACTATCATTAAAG
 AAGCATAAAGGCAACAATAGCGGGCAACCTTCAACAGTGGTGGACAACACACTCATGGTCATTATTGCAATGTTATACACATGTGAGAAGTGGGAATCAA
 CAAGGAAGAGATTGAGTCTACTTGCAGTTGAATCCAGAAAAGGAAGCATAATTTAAACCGATGATGGGAGCATATAAGCCAAGTCGACTTAATAGAGAGG
 AGTTGGCCCTGAAATATGAATTTGACTGCACCACCAGGGACAAGACACAGTTGGTTCATGTACACAGGGCTTTGGAGAGGGATGGCATGTATATACCA
 AAGCTAGAAGAAGAAAGGATTTCTATTTGGAATGGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGCATCAATGATCGAAGCATG
 GGGTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGATGGTGGTGGAAACAAGCCGCTATTTCACAGCTTGACAGAAAGGAAGGGCCATATC
 TGGCTGAGACTGCGCTTAAAGTTTTGTTACACATCTCAGCAGGAAACAACTCTGAGATAGAAGAGTATTTAAAAGTGGTGTATGATTACGATATTTCCAACG
 ACTGAGAATCTTTATTTTCAG

>TEVANib-Ros1-Nib9 (Nib insertion between positions 8463-8464 of TEVANib-Ros1; Nib included the indicated extra nucleotides)

GAGAACTTTATTTTCAGGGGGAGAAGAGGAAATGGGTCGTTGGAAGCACTGTCAGGGAACTTGAGGCCAGTGGCTGAGTGTCCAGTCAAGTGTAGTCACAAA
 GCATGTGGTTAAAGGAAAGTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAAGGAAGCATAATTTAAACCGATGATGGGAGCATATAAGCCAA
 GTCGACTTAATAGAGAGCGTTCCTCAAGGACATTTAAATAATGCTAGTGAATTTGAGATTGGGAATGTGGATTGTGACTTGTCTGGAGCTTGAATAAGC
 ATGCTCATCAAGGCTCAAGGCGTTAGGATTTCCCACTGTGAACATACACTGACCCAGAGGAAATTTTGTAGTGCATTGAATATGAAAGCAGCATGGG
 AGCACTATACAAAGGCAAGAAGAAGAGCTCTCAGCGAGCTCACACTAGATGAGCAGGAGGCAATGCTCAAAGCAAGTTGCCCTGCGACTGTATACGGGAA
 AGCTGGGAATTTGGAATGGCTCATTGAAAGCAGAGTTGCGTCCAATTGAGAAGGTTGAAACAACAACAAAACGCGAACTTTCACAGCAGCACCATAAGCACT
 CTTCTTGCTGGTAAAGTTTGGTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGCACCATGGACAGTTGGTATGACTAAGTTTTATCAGGG
 GTGGAATGAATTGATGGAGGCTTTACCAAGTGGTGGGTGATTTGTGACGCTGATGGTTCGCAATTCGACAGTTCCTTGACTCCATTCTCATTAAATGCTG
 TATTGAAAGTGGCAGTTGCCCTCATGGAGGAATGGGATATTGGTGGAGCAAAATGCTGCGAAATTTGTACTGAGATAGTGTATACACCAATCCTCACACCG
 GATGGTACTATCATTAAGAAGCATAAAGGCAACAATAGCGGGCAACCTTCAACAGTGGTGGACAACACACTCATGGTCAATTATGCAATGTTATACACATG
 TGAGAAGTGGGAATCAACAAGGAAGAGATTGTGATTTACGCTCAATGGCGATGACCTATTGATTGCCATTCACCCAGATAAAGCTGAGAGGTTGAGTGGAT
 TCAAAGAATCTTTCGGAGAGTTGGCCCTGAAATATGAATTTGACTGCACCACCAGGGACAAGACACAGTTGGTTCATGTACACAGGGCTTTGGAGAGG
 GATGGCATGTATATACCAAGCTAGAAGAAGAAAGGATTTGTTCTATTTGGAATGGGACAGATCCAAAGAGCCGTCACATAGGCTTGAAGCCATCTGTGC
 ATCAATGATCGAAGCATGGGGTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGATGGTGGTGGAAACAAGCCGCTATTTCACAGCTTGACAGAA
 AAGGAAGGGCCATATCTGGCTGAGACTGCGCTTAAAGTTTTGTTACACATCTCAGCAGGAAACAACTCTGAGATAGAAGAGTATTTAAAAGTGGTGTAT
 GATTACGATATTTCAAGCACTGAGAATCTTTATTTTCAG

Chapter II

A potyvirus vector efficiently targets recombinant proteins to chloroplasts, mitochondria and nuclei in plant cells when expressed at the amino terminus of the polyprotein

Eszter Majer, José-Antonio Navarro, José-Antonio Daròs

Instituto de Biología Molecular y Celular de Plantas (Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia), Valencia, Spain

Correspondence: Dr. José-Antonio Daròs, IBMCP (CSIC-UPV), Avenida de los Naranjos s/n, 46022 Valencia, Spain. E-mail: jadaros@ibmcp.upv.es

Biotechnology Journal (2015) 10(11): 1792-1802.

Abstract

Plant virus-based expression systems allow quick and efficient production of recombinant proteins in plant biofactories. Among them, a system derived from *Tobacco etch virus* (TEV; genus *Potyvirus*) permits co-expression of equimolar amounts of several recombinant proteins. This work analyzed how to target recombinant proteins to different subcellular localizations in the plant cell using this system. We constructed TEV clones in which green fluorescent protein (GFP), with a chloroplast transit peptide (cTP), a nuclear localization signal (NLS) or a mitochondrial targeting peptide (mTP) was expressed either as the most amino-terminal product or embedded in the viral polyprotein. Results showed that cTP and mTP mediated efficient translocation of GFP to the corresponding organelle only when present at the amino terminus of the viral polyprotein. In contrast, the NLS worked efficiently at both positions. Viruses expressing GFP in the amino terminus of the viral polyprotein produced milder symptoms. Untagged GFPs and cTP and NLS tagged amino-terminal GFPs accumulated to higher amounts in infected tissues. Finally, viral progeny from clones with internal GFPs maintained the extra gene better. These observations will help in the design of potyvirus-based vectors able to co-express several proteins while targeting different subcellular localizations, as required in plant metabolic engineering.

Introduction

Several plant virus-based expression systems are currently available for transiently expressing proteins of interest in plant biofactories. The key to their success is simplicity of manipulation, rapidity and yield, when compared to the difficult and time consuming generation and unpredictability of stably transformed plants. This way, some plant viruses, either as replicating viral vectors or as non-replicating transient expression systems, have been engineered to express antigen vaccines, antibodies, pharmaceutical proteins and peptides, industrial enzymes, as well as components of nanoparticles (Pogue *et al.*, 2002; Sainsbury *et al.*, 2010; Chen *et al.*, 2011; Peyret and Lomonossoff, 2013; Gleba *et al.*, 2014). Attempts have also been made to use plant virus-derived expression systems in metabolic engineering (Bedoya *et al.*, 2010; Zhang *et al.*, 2013; Sainsbury and Lomonossoff, 2014). Nonetheless, goals in plant biotechnology frequently require concurrent expression of several proteins, which is a limitation for most plant viral vectors (Sainsbury *et al.*, 2008). Expression of proteins

involved in multi-step metabolic pathways requires control on stoichiometry and timing of expression. Similarly, expression of protein complexes such as virus-like particles requires several proteins to be expressed in different amounts. Some solutions have been offered to simultaneously express various proteins in plants using viral vectors (Giritch *et al.*, 2006; Kelloniemi *et al.*, 2008; Roy *et al.*, 2011; Wang *et al.*, 2014). To these, we recently added a system derived from *Tobacco etch virus* (TEV; genus *Potyvirus*, family *Potyviridae*) based on replacing a viral gene, nuclear inclusion b (NIb) –coding for the viral RNA-dependent RNA polymerase–, with a cassette to coexpress several heterologous proteins (Bedoya *et al.*, 2010). Two important properties of this system are that the various recombinant proteins are produced embedded in the viral polyprotein which they are then efficiently released from by the activity of the virus-encoded nuclear inclusion a protease (NIaPro), and that the viral vector is only able to replicate and spread in plants in which the viral NIb protein is supplied in trans (Bedoya *et al.*, 2010).

Potyviruses (genus *Potyvirus*) make up one of the two largest groups of plant viruses and infect many economically important crops. The potyviral genome consists of an approximately 10-kb long, plus-polarity, single-stranded RNA that is covalently attached to a viral protein genome-linked (VPg) through its 5' end and contains a poly(A) tail at the 3' end. It is encapsidated in elongated and flexuous virions containing approximately 2000 copies of the coat protein (CP). The potyviral genome encodes a large polyprotein that is apparently processed into ten mature gene products (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIaPro, NIb and CP) by the activity of three viral-encoded proteases (P1, HC-Pro and NIaPro) (Urcuqui-Inchima *et al.*, 2001). P1 and HC-Pro catalyze their own cleavage from the viral polyprotein, whereas NIaPro cleaves in the remaining sites in cis and in trans, recognizing a seven-amino acid-long sequence motif (Carrington *et al.*, 1990). An additional gene product, P3N-PIPO, is produced through a translation frame-shift mechanism in the P3 cistron (Chung *et al.*, 2008).

Using the TEV-derived viral vector, we simultaneously expressed three fluorescent proteins (mCherry, Venus and TagBFP) in transgenic tobacco (*Nicotiana tabacum* L.) plants constitutively expressing NIb, which accumulated in comparable amounts in the same tissues and nucleocytoplasmic location (Bedoya *et al.*, 2010). We also co-expressed two transcription factors, Delila and Rosea1 from *Antirrhinum majus* L. that activate the biosynthesis of anthocyanins in plants (Butelli *et al.*, 2008), and

managed to produce remarkable amounts of red colored anthocyanins in the tobacco tissues invaded by the virus, demonstrating the utility of the TEV-derived vector in plant metabolic engineering (Bedoya *et al.*, 2010). In later works, we showed that expression of Rosea1 was enough to successfully activate anthocyanin biosynthesis to levels readily visible to human eye (Bedoya *et al.*, 2012; Majer *et al.*, 2013). Anthocyanins are a class of flavonoid compounds responsible of many of the vivid colors of flowers and fruits in plants; dietary anthocyanins are considered health promoting compounds (Zhang *et al.*, 2014).

In plant metabolic engineering, an important consideration is how to target the recombinant proteins to specific subcellular compartments, because the complex metabolic networks in plants are highly compartmentalized (Heinig *et al.*, 2013). In addition targeting recombinant proteins to specific subcellular compartments can also be important for proper folding, solubility, stability and ease of purification. The goal of this work was to analyze the possibilities of targeting proteins of interest to different subcellular compartments in plants using a TEV-derived vector. To this end, we constructed recombinant TEV clones lacking N1b (TEV Δ N1b) and harboring the green fluorescent protein (GFP) at two alternative positions in the viral polyprotein: the amino-terminal end or embedded inside the polyprotein, replacing the endogenous viral N1b. Recombinant TEV clones included sequence motifs to target the GFP to the chloroplasts, nuclei or mitochondria. We infected transgenic tobacco plants constitutively expressing N1b and analyzed GFP subcellular location and accumulation, as well as the stability of the recombinant viral vectors. Our results show new possibilities, but also some limitations, in targeting proteins of interest to different subcellular compartments in plants using potyviral vectors.

Materials and methods

Recombinant TEV clones

Recombinant TEV clones TEV Δ N1b-aGFP, TEV Δ N1b-acTPGFP, TEV Δ N1b-aNLSGFP, TEV Δ N1b-amTPGFP, TEV Δ N1b-iGFP, TEV Δ N1b-icTPGFP, TEV Δ N1b-iNLSGFP and TEV Δ N1b-imTPGFP (Figure 1) were constructed using parent plasmid pGTEVa (Bedoya *et al.*, 2012), which contains a TEV infectious cDNA (GenBank accession number DQ986288, including two silent mutations, G273A and A1119G, that were introduced for cloning purposes) flanked by *Cauliflower mosaic virus* (CaMV)

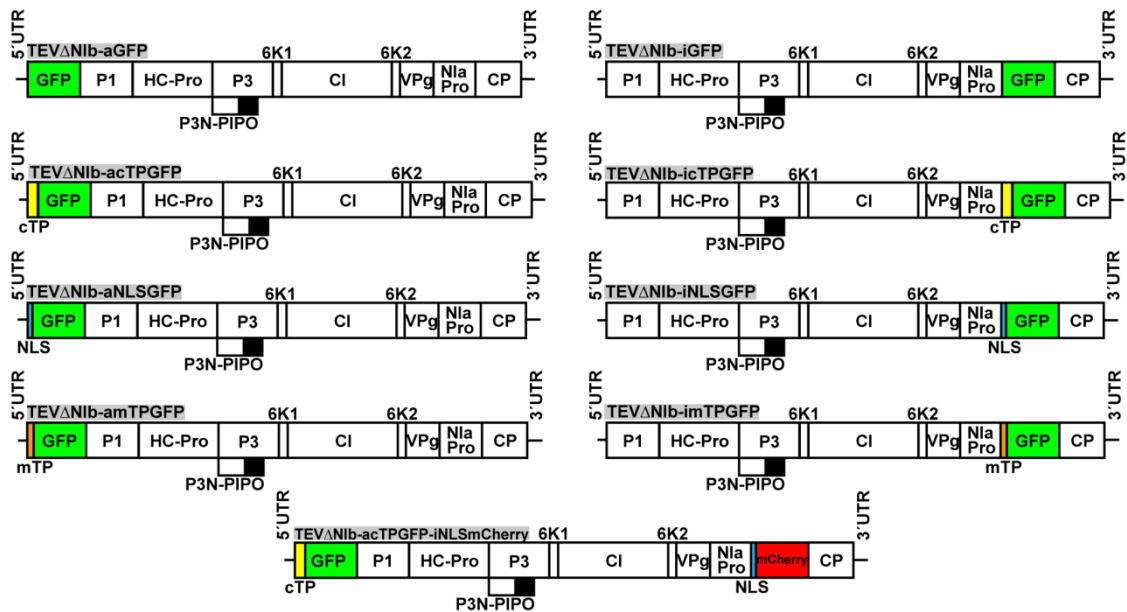


Figure 1. Schematic representation of recombinant TEV clones: TEV Δ NIB-aGFP, TEV Δ NIB-iGFP, TEV Δ NIB-acTPGFP, TEV Δ NIB-icTPGFP, TEV Δ NIB-aNLSGFP, TEV Δ NIB-iNLSGFP, TEV Δ NIB-amTPGFP, TEV Δ NIB-imTPGFP and TEV Δ NIB-acTPGFP-iNLSmCherry, in which cDNAs coding for untagged GFP or GFP and mCherry variants tagged with a chloroplastic transit peptide (cTP), a nuclear localization signal (NLS) or a mitochondrial targeting peptide (mTP) were inserted at the 5' end of the viral ORF (amino-terminal clones, a) or replacing NIB (internal clones, i). TEV cistrons P1, HC-Pro, P3, P3N-PIPO, 6K1, CI, 6K2, VPg, NlaPro and CP and are represented by rectangles and TEV 5' and 3' untranslated regions (UTR) by black lines. GFP and mCherry are represented by green and red rectangles, respectively. cTP, NLS and mTP are represented by yellow, blue and orange rectangles, respectively.

35S promoter and terminator in a binary vector derived from pCLEAN-G181 (Thole *et al.*, 2007). The exact nucleotide sequence of all recombinant TEV clones constructed in this work is shown in Supporting information, Figure S1. All recombinant TEV clones included the insertion of a cDNA encoding the enhanced GFP (GenBank accession number AAB08060) either at the beginning of the viral open reading frame (between positions 144 and 145 of DQ986288; clones TEV Δ NIB-aGFP, TEV Δ NIB-acTPGFP, TEV Δ NIB-aNLSGFP, TEV Δ NIB-amTPGFP; Figure 1) –these clones also contained the deletion of the whole NIB cistron from position 6982 to position 8517 of DQ986288– or replacing most of the NIB cistron (from position 6991 to 8493; clones TEV Δ NIB-iGFP, TEV Δ NIB-icTPGFP, TEV Δ NIB-iNLSGFP and TEV Δ NIB-imTPGFP; Figure 1). The remains of the NIB cistron in these last clones (the first three and last eight codons) served for the NlaPro-mediated proteolytic processing of the different GFP variants. Clones TEV Δ NIB-aGFP, TEV Δ NIB-acTPGFP, TEV Δ NIB-aNLSGFP and TEV Δ NIB-amTPGFP included a sequence (5'-ACGACTGAGAATCTTTATTTTCAG/GGGGAG-

AAG-3') after the GFP cDNA, encoding a typical NIaPro cleavage site (Nt-TTENLYFQ/GEK-Ct). At the nucleotide level, this sequence was not homologous to any fragment of TEV sequence, although the encoded amino acids matched those surrounding the NIb/CP cleavage site in TEV. This sequence was inserted to facilitate the NIaPro-mediated proteolytic release of the different GFP variants from the viral polyprotein. Variant cTPGFP (included in clones TEV Δ NIb-acTPGFP and TEV Δ NIb-icTPGFP; Figure 1) contained the 180 first nucleotides of RuBisCO small subunit from *N. tabacum* (position 1048 to 1227 of GenBank accession number X02353.1), encoding a chloroplastic transit peptide (cTP) (Mazur and Chui, 1985). Variant NLSGFP (included in clones TEV Δ NIb-aNLSGFP and TEV Δ NIb-iNLSGFP; Figure 1) contained a sequence corresponding to two repetitions of the nuclear localization signal (NLS) from Simian virus 40 (SV40) T antigen (5'-CCAAAAAAGAAGAGAAAGGTACCA-AAGAAGAAAAGAAAGGTA-3'; the second repetition contains two silent mutations with respect to the first, underlined nucleotides) (Kalderon *et al.*, 1984; Lassner *et al.*, 1991). The NLS was preceded by an ATG. Finally, variant mTPGFP (included in clones TEV Δ NIb-amTPGFP and TEV Δ NIb-imTPGFP; Figure 1) contained the mitochondrial targeting presequence (mTP) of cytochrome oxidase subunit IV from *Saccharomyces cerevisiae* (5'-ATGTTGTCACCTACGTCAATCTATAAGATTTTTC-AAGCCAGCCACAAGAAGAACTTTGTGTAGCTCTAGATATCTGCTTCAGCAAAAA-CCC-3') (Hurt *et al.*, 1985; Köhler *et al.*, 1997). All these signal sequences (cTP, NLS and mTP) were fused to the 5' end of GFP cDNA.

Recombinant TEV clone TEV Δ NIb-acTPGFP-iNLSmCherry is like TEV Δ NIb-acTPGFP, but also contained the cDNA coding for the red fluorescent protein mCherry (GenBank accession number AY678264) including two repetitions of the SV40 NLS, as described above, replacing most of NIb cistron (from position 6991 to 8493). The exact nucleotide sequence of this recombinant clone is also shown in Supporting information, Figure S1.

Plasmids with the recombinant TEV clones were constructed using standard molecular biology techniques, including PCR amplification of cDNAs with the high-fidelity Phusion DNA polymerase (Thermo Scientific), DNA digestion with the type-IIS restriction enzyme Eco31I (Thermo Scientific) for assembly of DNA fragments (Engler *et al.*, 2009), DNA ligation with T4 DNA ligase (Thermo Scientific) and transformation

of *Escherichia coli* DH5 α by electroporation. The sequences of the resulting plasmids were confirmed by standard DNA sequencing techniques.

Plant inoculation

Transgenic tobacco plants (*Nicotiana tabacum* L. cv. Xanthi nc), constitutively expressing TEV N1b under the control of CaMV 35S promoter (Li and Carrington, 1995), were cultivated in a growth chamber at 25°C with a photoperiod of 12 h day and 12 h night. Four week old plants were infiltrated in the third true leaf with cultures of *Agrobacterium tumefaciens* C58C1 transformed with the helper plasmid pCLEAN-S48 (Thole *et al.*, 2007) and each of the plasmids containing the different recombinant TEV clones described above. Previous to agroinoculation, *A. tumefaciens* cultures were adjusted to an optical density of 0.5 at 600 nm in 10 mM MES-NaOH, pH 5.6, 10 mM MgCl₂, and virulence genes induced with 150 μ M acetosyringone for 2 h at 28°C (Bedoya and Daròs, 2010).

Analysis of fluorescent proteins

Ten days post-inoculation (dpi), symptomatic tissues from upper leaves were collected and analyzed by confocal laser scanning microscopy (CLSM) using a Zeiss LSM 780 microscope with a Plan-Apochromat 63x/1.4 NA oil-immersion lens and 488 (GFP and chlorophyll) and 561 nm (mCherry) excitation lasers. The emission detection windows were 492-532, 590-630 and 700-800 nm for GFP, mCherry and chlorophyll, respectively.

GFP and TEV CP were analyzed in infected tissues by Western blot. For this purpose, frozen aliquots of 0.3 g of symptomatic tissues from upper leaves were homogenized in 0.9 ml of buffer TEW (60 mM Tris-HCl, pH 6.8, 2% (w/v) sodium dodecyl sulfate –SDS–, 100 mM dithiothreitol, 10% (w/v) glycerol and 0.01% (w/v) bromophenol blue). Extracts were incubated for 5 min at 95°C, vortexed and centrifuged for 15 min at 16,000xg. Equivalent aliquots of the clarified supernatants (40 μ l, equivalent to approximately 13 mg of fresh tissue) were separated by discontinuous polyacrylamide gel electrophoresis (PAGE) in 12.5% (w/v) polyacrylamide gels (5% (w/v) polyacrylamide for the stacking gel) containing 0.05% (w/v) SDS. Separated proteins were electroblotted to polyvinylidene fluoride membranes (GE Healthcare). Membranes were blocked for 1 h in 5% (w/v) non-fat milk in buffer WB (10 mM Tris-

HCl, pH 7.5, 154 mM NaCl, 0.1% (w/v) Nonidet P40) and incubated overnight at 4°C with a GFP polyclonal antibody (Sigma-Aldrich) or a TEV CP polyclonal antibody conjugated to alkaline phosphatase (Agdia), both at 1:10,000 dilution in 5% (w/v) non-fat milk in WB. Next, membranes were washed three times with WB and, in the case of GFP, incubated for 3 h at room temperature with an anti-mouse secondary antibody conjugated to horseradish peroxidase (GE Healthcare) at a 1:10,000 dilution in 5% (w/v) non-fat milk in WB. After washing the membranes, horseradish peroxidase and alkaline phosphatase were finally detected using reagents to produce luminescence (ECL Western Blot Detection, Applied Biological Materials; CSPD; Roche). Luminescence was detected and quantified with an image analyzer (LAS-3000, Fujifilm).

Analysis of viral progeny

Ten dpi, viral progeny in symptomatic tissues from upper leaves was analyzed by reverse transcription (RT) using Moloney murine leukemia virus reverse transcriptase (RevertAid, Thermo Scientific) coupled to PCR amplification with *Thermus thermophilus* DNA polymerase (Biotools). RNA from symptomatic tissues of three independent plants infected with each viral clone was purified by chromatography using silica spin-columns (Zymo Research). To analyze the presence of the GFP cDNA in the progeny of TEV Δ NiB-aGFP, TEV Δ NiB-acTPGFP, TEV Δ NiB-aNLSGFP and TEV Δ NiB-amTPGFP, RNA was subjected to RT using primer I (5'-TTGTTT-TGCCTGTTCAACGAGGTC-3') and the reaction products amplified by PCR using primers II (5'-AAAATAACAAATCTCAACACAACATATAC-3') and III (5'-ATAATGTGTCATTGAGTGACAAACAGC-3'). To analyze the presence of the GFP cDNA in the progeny of TEV Δ NiB-iGFP, TEV Δ NiB-icTPGFP, TEV Δ NiB-iNLSGFP and TEV Δ NiB-imTPGFP, RNA was reverse transcribed using primer IV (5'-CTCGCACTACATAGGAGAATTAGAC-3') and the reaction products were amplified by PCR with primers V (5'-TTATTCGCATGCCTAAGGATTTCCC-3') and VI (5'-TACCTAGGCATGTATGGTCGCTCCC-3'). The products of the RT-PCR amplifications were separated by electrophoresis in 1% (w/v) agarose gels in buffer TAE (40 mM Tris, 20 mM sodium acetate, 1 mM ethylenediaminetetraacetic acid -EDTA-, pH 7.0), stained with ethidium bromide and imaged on a UV trans-illuminator.

Results

A TEV Δ NIB-based vector targets recombinant proteins to chloroplast, nucleus and mitochondria

To analyze whether it is possible to target recombinant proteins to different subcellular locations in plants using a viral vector derived from a potyvirus, we constructed a series of recombinant TEV clones to express GFP, with or without a cTP (cTPGFP), a NLS (NLSGFP) and an mTP (mTPGFP). These clones were constructed in the context of our previous TEV-derived viral vector that lacks the NIB cistron (TEV Δ NIB) and is only able to replicate and move systemically in plants in which NIB is expressed in trans (Bedoya *et al.*, 2010). Two series of clones were constructed to express the different GFP variants. The first series contained the GFP variants at the amino-terminal end of the viral polyprotein (TEV Δ NIB-aGFP, TEV Δ NIB-acTPGFP, TEV Δ NIB-aNLSGFP and TEV Δ NIB-amTPGFP; Figure 1). The second series contained the GFP variants embedded inside the viral polyprotein, replacing most of the viral NIB cistron (TEV Δ NIB-iGFP, TEV Δ NIB-icTPGFP, TEV Δ NIB-iNLSGFP and TEV Δ NIB-imTPGFP; Figure 1). All clones were designed to allow release of the GFP variants from the viral polyprotein by NIaPro mediated proteolysis. Clones with the GFP variants at the amino terminus included an artificial NIaPro cleavage site after GFP. Constructs with the GFP variants replacing the NIB cistron maintained the three first and eight last codons from NIB, the GFP thus flanked by two native NIaPro cleavage sites.

Transgenic tobacco plants constitutively expressing NIB were agroinoculated with the eight recombinant TEV clones. Ten dpi, all inoculated plants showed symptoms of infection, although they varied in severity. As a general rule, recombinant viruses with the GFP variants at the amino terminus of the viral polyprotein produced milder symptoms than the counterparts with the GFP variants inside the viral polyprotein. Clones with an internal GFP produced leaf distortion and severe stunting, whereas clones with an amino-terminal GFP only produced vein clearing and mild etching in leaf tissues. To illustrate this difference, Figure 2 shows the symptoms produced in whole plants (Figure 2A) and leaves (Figure 2B) by the two viral clones containing the untagged GFP variants, TEV Δ NIB-aGFP and TEV Δ NIB-iGFP.

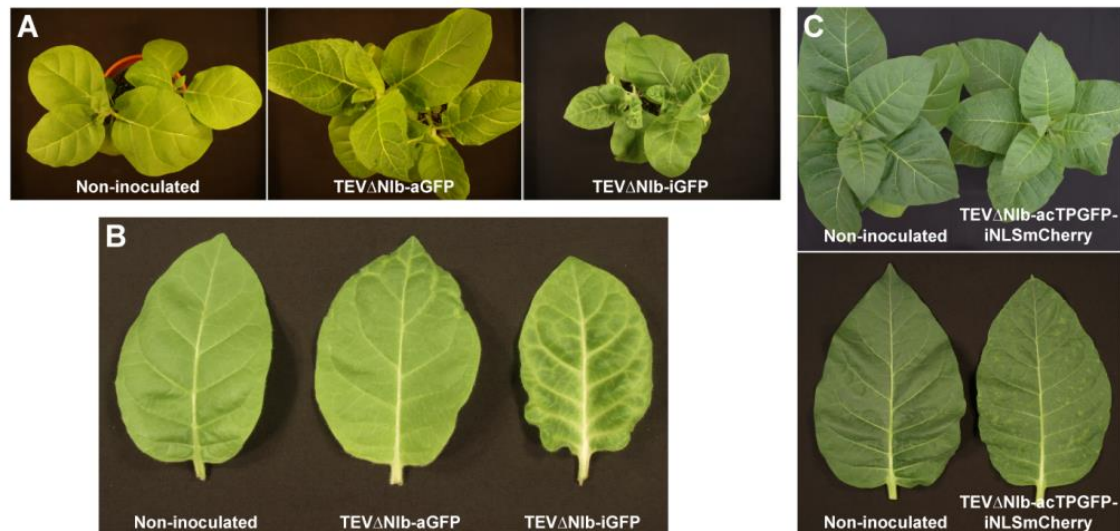


Figure 2. Tobacco plants infected by recombinant TEV clones expressing (A and B) untagged GFP as the most amino-terminal product of the viral polyprotein (TEV Δ NIB-aGFP) or embedded inside the viral polyprotein (TEV Δ NIB-iGFP), and (C) GFP and mCherry targeted to the chloroplast and nucleus, respectively (TEV Δ NIB-acTPGFP-iNLSmCherry). Pictures were taken at 10 (A and B) or 20 dpi (C) and include non-inoculated controls.

Symptomatic tissues from plants infected by the eight TEV recombinant clones were examined by CLSM. Green fluorescence corresponding to GFP emission was detected in tissues from plants infected by all eight recombinant clones (Figure 3A). As expected, tissues from plants infected by TEV Δ NIB-aGFP and TEV Δ NIB-iGFP, expressing untagged GFP, showed similar GFP fluorescence in both the cytoplasm and nucleoplasm of infected cells (Figure 3A, compare images I, V and IX with II, VI and X). Interestingly, tissues from plants infected with TEV Δ NIB-acTPGFP showed GFP fluorescence inside the chloroplast, in contrast to tissues infected with TEV Δ NIB-aGFP (Figure 3A, compare images I and III). Chloroplasts are clearly recognized in all these images thanks to the red fluorescence emission of chlorophylls (Figure 3A). On the contrary, tissues from plants infected with TEV Δ NIB-icTPGFP showed no GFP fluorescence in the chloroplasts (Figure 3A, compare images III and IV). These results indicate that the cTP assayed in this work is effective only when the recombinant protein is expressed as the most amino-terminal product of the viral polyprotein.

In the case of tissues from plants infected with TEV Δ NIB-aNLSGFP and TEV Δ NIB-iNLSGFP, GFP fluorescence was clearly detected in nucleoli, in contrast to what occurred in tissues from plants infected with TEV Δ NIB-aGFP and TEV Δ NIB-iGFP (Figure 3A, compare images V and VI with VII and VIII). These data indicate that

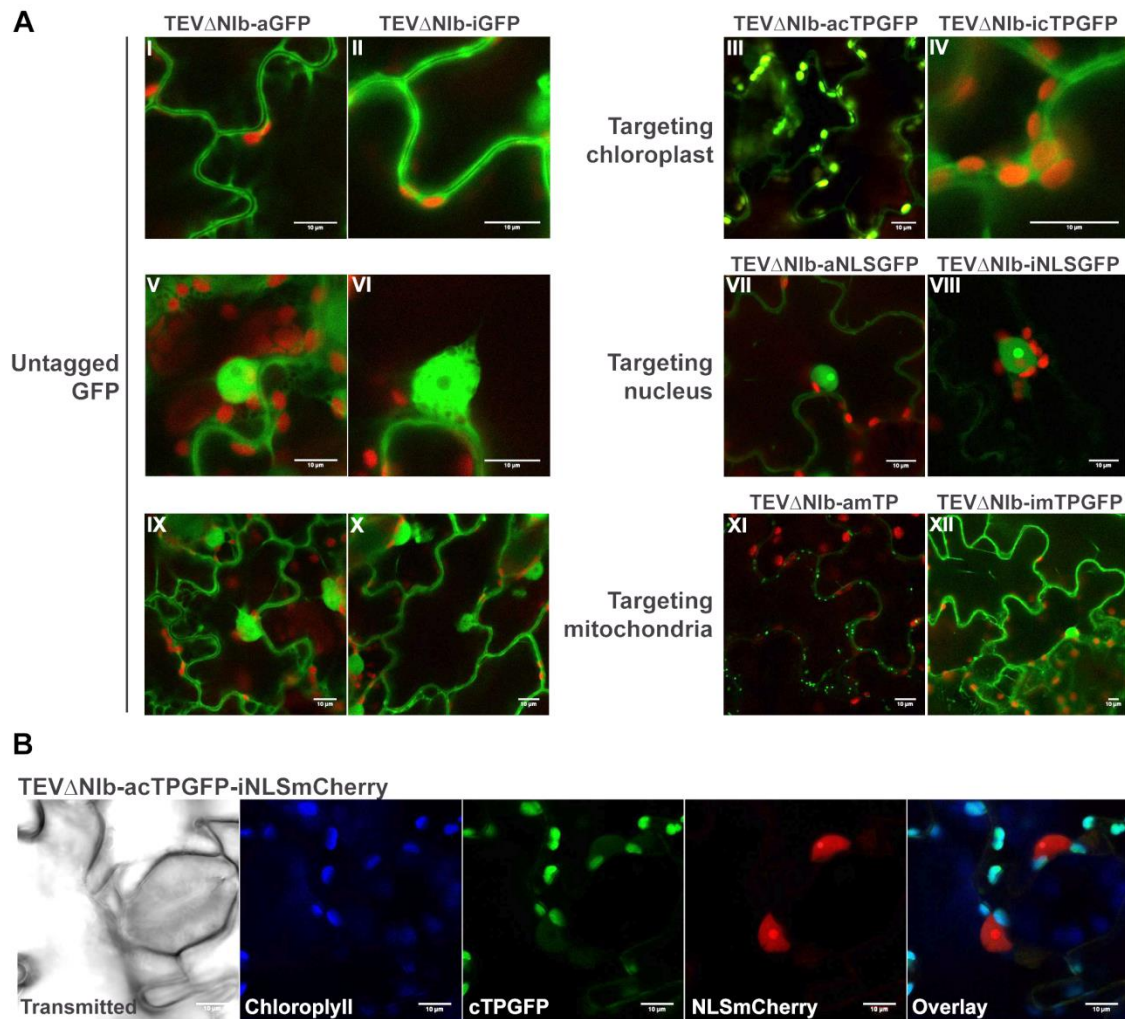


Figure 3. Subcellular localization of GFP and mCherry variants tagged with a chloroplastic transit peptide (cTP), a nuclear localization signal (NLS), or a mitochondrial targeting sequence (mTP) in tobacco tissues infected by different recombinant TEV clones. **(A)** CLSM images of tissues infected with TEVΔNIB-aGFP (images I, V and IX), TEVΔNIB-iGFP (images II, VI and X), TEVΔNIB-acTPGFP (image III), TEVΔNIB-icTPGFP (image IV), TEVΔNIB-aNLSGFP (image VII), TEVΔNIB-iNLSGFP (image VIII), TEVΔNIB-amTPGFP (image XI) and TEVΔNIB-imTPGFP (image XII). Images are overlays of the green and red channels. **(B)** CLSM images of systemically infected tissue by TEVΔNIB-acTPGFP-iNLSmCherry, corresponding to transmitted light, chlorophyll, GFP and mCherry fluorescence, respectively, and the overlay of these three fluorescence images. (A and B) All pictures were taken at 10 dpi. Scale bars in all images represent 10 μ m.

the NLS employed in this work is effective to target proteins to the nucleus and enter the nucleolus at both assayed positions in the viral polyprotein. Finally, tissues from plants infected with TEVΔNIB-amTPGFP showed GFP fluorescence in small subcellular organelles in the cytoplasm that can be undoubtedly assigned to mitochondria based on their number and shape, in contrast to tissues infected by TEVΔNIB-aGFP (Figure 3A, compare images IX and XI). On the contrary, GFP fluorescence in tissues infected by TEVΔNIB-imTPGFP was indistinguishable from that

in tissues infected by TEV Δ NiB-iGFP (Figure 3A, compare images X and XII). These data indicate that the mTP selected in this work to target recombinant proteins to the mitochondria is only functional from the most amino-terminal end of the viral polyprotein. Fluorescence intensities were quantified from CLSM pictures using an image analysis software (Fiji). Supporting information, Figure S2 shows the ratios between the GFP fluorescence in the different organelles and the cytoplasm.

A TEV Δ NiB-based vector is able to target two recombinant proteins to distinct subcellular compartments

Considering the results showed above, we investigated the possibility to build a TEV Δ NiB-based vector able to deliver two recombinant proteins to different subcellular compartments from a single vector backbone. To check this, we constructed TEV Δ NiB-acTPGFP-iNLSmCherry, in which a GFP including a cTP was expressed from the amino terminus of the viral polyprotein followed by an artificial NiAPro cleavage site, and the red fluorescent protein mCherry including a NLS was expressed from inside the viral polyprotein replacing most of NiB cistron, flanked by the viral NiAPro cleavage sites (Figure 1). NiB-expressing tobacco plants, agroinoculated with this viral clone, showed mild symptoms (Figure 2C). Analysis of symptomatic tissue from upper leaves, by CLSM at 10 dpi, demonstrated accumulation of GFP in the chloroplasts and mCherry in nuclei and nucleoli of infected cells (Figure 3B).

Recombinant proteins are efficiently processed from TEV polyprotein

Next, we investigated whether the recombinant proteins expressed from the TEV-derived vectors were properly and efficiently released from the viral polyprotein. To this end, proteins were extracted from systemic symptomatic tissues of plants infected with the different recombinant TEV clones at 10 dpi. Proteins were separated by denaturing PAGE and GFP detected by Western blot using a specific polyclonal antibody (Figure 4A). As a control, we also detected TEV CP in the same protein preparations with an anti-TEV CP polyclonal antibody (Figure 4B). We checked that both antibodies do not produce any cross-reactivity. Theoretical molecular mass of the enhanced GFP used in this study is 26.9 kDa. However, the GFP variants expressed by means of the TEV Δ NiB vector were expected to contain additional amino acids from the different signal peptides and remains from the NiAPro-mediated proteolytic process—

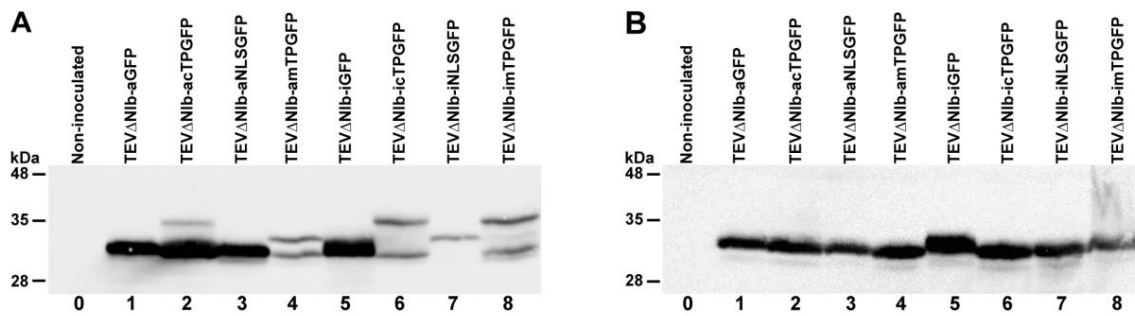


Figure 4. Western blot analysis of (A) GFP variants and (B) TEV CP accumulating in tobacco tissues infected by different recombinant TEV clones at 10 dpi. Proteins were separated by denaturing PAGE, blotted to a membrane and detected with anti-GFP (A) or anti-TEV CP polyclonal antibodies (B). (A and B) Lanes 0, non-inoculated control plant; lanes 1 to 8, plants infected by TEV Δ NiB-aGFP, TEV Δ NiB-acTPGFP, TEV Δ NiB-aNLSGFP, TEV Δ NiB-amTPGFP, TEV Δ NiB-iGFP, TEV Δ NiB-icTPGFP, TEV Δ NiB-iNLSGFP and TEV Δ NiB-imTPGFP, respectively. Positions and molecular mass (in kDa) of marker proteins are indicated on the left side of both panels.

ing that change electrophoretic mobility. Western blot analysis revealed bands exclusively corresponding to GFP forms released from the TEV polyprotein (Figure 4A). No signal arising from a slow migrating species that could be attributed to an unprocessed intermediate from the viral polyprotein was detected. The intensity of the different bands informed about differences in GFP accumulation in infected tissues (Supporting information, Table S1), which were previously unnoticed in our microscopic analyses. Tissues infected with viruses expressing untagged GFP, either at the amino terminus or internally (TEV Δ NiB-aGFP and TEV Δ NiB-iGFP) accumulated comparable amounts of GFP (Figure 4A, lanes 1 and 5; Supporting information, Table S1). Same accumulation was also observed in tissues infected by viruses expressing GFP with a cTP and a NLS, but only when expressed from the amino-terminal end of the viral polyprotein (Figure 4A, compare lanes 2 and 3 with 6 and 7; Supporting information, Table S1). Interestingly, two different bands were detected in tissues infected with viruses expressing GFP variants including cTP (TEV Δ NiB-acTPGFP and TEV Δ NiB-icTPGFP), indicating that GFP forms including the transit peptide are initially produced and, then, the transit peptide is processed in part. Notably, while the unprocessed form is less abundant than the processed GFP in tissues infected by TEV Δ NiB-acTPGFP, in tissues infected by TEV Δ NiB-icTPGFP abundance of both species is similar (Figure 4A, compare lanes 2 and 6; Supporting information, Table S1). Moreover, total GFP accumulation in tissues infected by TEV Δ NiB-icTPGFP was lower than in tissues infected by TEV Δ NiB-acTPGFP (Figure 4A, compare lanes 2 and

6; Supporting information, Table S1). GFP accumulation in tissues infected by TEV Δ NIb-aNLSGFP was also higher than that in tissues infected by the internal counterpart virus TEV Δ NIb-iNLSGFP (Figure 4A, compare lanes 3 and 7; Supporting information, Table S1). Finally, GFP accumulation in tissues infected by both virus with the mTP (TEV Δ NIb-amTPGFP and TEV Δ NIb-imTPGFP) was low compared with the untagged counterparts (Figure 4A, compare lane 1 with 4 and lane 5 with 8; Supporting information, Table S1). In this case, similarly to what occurred with the cTP, two GFP forms were detected, suggesting accumulation of a full-length GFP including the mTP and a processed form without the mTP (Figure 4A, lanes 4 and 8). Some differences were also observed in accumulation of TEV CP in the analyzed tissues (Figure 4B; Supporting information, Table S1).

Cistrons encoding recombinant proteins are more stable inside the TEV genome

Finally, we asked about the stability of the cistrons encoding the different GFP variants in the viral progeny. To analyze this, we collected infected tissue at 10 dpi from upper symptomatic leaves of three different plants per recombinant viral clone. RNA was purified and the presence of the GFP cistron analyzed by RT-PCR amplification followed by electrophoretic separation of the PCR products. In case of the GFP variants at the amino terminus of the polyprotein, RT-PCR products ranging between 1806 and 1986 base pairs (bp) were expected from full-length viral genomes depending on the particular GFP variant. These products were obtained from most of the plants (Figure 5A, white arrowhead). However, one, two, three and one of the plants respectively infected with TEV Δ NIb-aGFP, TEV Δ NIb-acTPGFP, TEV Δ NIb-aNLSGFP and TEV Δ NIb-amTPGFP produced substantial amounts of smaller RT-PCR products, indicating the deletion of the GFP cistron (Figure 5A, gray arrowheads). Conversely, all plants infected with viral clones expressing the GFP variants internally in the polyprotein produced RT-PCR products consistent with full-length viral clones (Figure 5B, white arrowhead; 1768 and 1948 bp depending on the GFP variant). These data indicate that, in the TEV-based vector, recombinant genes replacing NIb in the viral polyprotein are more stable than those inserted as the most amino-terminal product of the viral polyprotein.

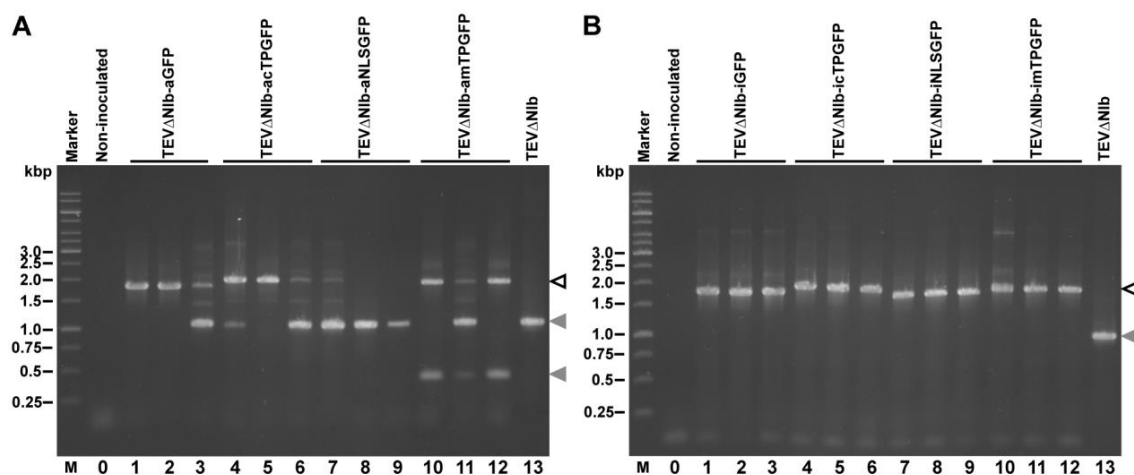


Figure 5. RT-PCR analysis of the viral progeny of the TEV vectors expressing distinct GFP variants from two different positions in the viral polyprotein at 10 dpi. RT-PCR products amplified with primers I to III (**A**) or IV to V (**B**) were separated by agarose electrophoresis and the gel stained with ethidium bromide. (**A** and **B**) Lanes M, marker DNA ladder with the size (in kbp) of some of the components on the left side; lanes 0, RT-PCR control of non-inoculated plant. (**A**) Lanes 1 to 12, RT-PCR products amplified from three plants infected with TEVΔNIB-aGFP (lanes 1 to 3), TEVΔNIB-acTPGFP (lanes 4 to 6), TEVΔNIB-aNLSGFP (lanes 7 to 9) and TEVΔNIB-amTPGFP (lanes 10 to 12). (**B**) Lanes 1 to 12, RT-PCR products amplified from three plants infected with TEVΔNIB-iGFP (lanes 1 to 3), TEVΔNIB-icTPGFP (lanes 4 to 6), TEVΔNIB-iNLSGFP (lanes 7 to 9) and TEVΔNIB-imTPGFP (lanes 10 to 12). (**A** and **B**) Lanes 13, PCR control of TEVΔNIB plasmid. White arrowheads point to RT-PCR products amplified from viral progenies with full-length genomes. Gray arrowhead point to products arising from viral progenies with deletions.

Discussion

Plant virus-derived expression vectors are currently envisioned as powerful platforms to generate products of interest in plant biofactories (Pogue *et al.*, 2002; Sainsbury *et al.*, 2010; Chen *et al.*, 2011; Peyret and Lomonosoff, 2013; Gleba *et al.*, 2014). However, goals approached with this technology frequently require simultaneous expression of more than one recombinant protein in time and precise subcellular localization of the expressed proteins. Potyviruses form one of the largest groups of viruses infecting plants (Gibbs and Ohshima, 2010). Some particularities of potyviruses, specifically their expression strategy of producing a large polyprotein that is efficiently cleaved by viral encoded proteases and their encapsidation in flexuous elongated virions that can accommodate substantial amount of extra genetic material, have made them good starting points to construct vectors to express proteins in plants, and particularly to simultaneously co-express equimolar amounts of several proteins (Masuta *et al.*, 2000; Beauchemin *et al.*, 2005; Kelloniemi *et al.*, 2008; Bedoya *et al.*, 2010). We developed

an expression system derived from TEV (genus *Potyvirus*) that infects solanaceous plants, based on replacing the viral NIB cistron by a cassette containing the cDNAs coding for several recombinant proteins. These cDNAs are flanked by sequences coding for viral and artificial NIaPro cleavage sites. Recombinant proteins are synthesized in equimolar amounts along the viral polyprotein and are efficiently and specifically released by the proteolytic activity of NIaPro. Since this vector lacks the viral RNA-dependent RNA polymerase NIB, it is only able to infect plants in which this protein is supplied in trans, for example through a transgene, which is a very desirable property for biocontainment purposes. Moreover, absence of the approximately 1.5-kb long NIB cistron contributes to increase space to accommodate the recombinant cDNAs in the viral vector. Using this system we have successfully expressed three fluorescent proteins that exquisitely colocalized in the same cells and nucleocytoplasmic localization and two transcription factors that strongly activated the biosynthesis of anthocyanins in plant tissues (Bedoya *et al.*, 2010; Bedoya *et al.*, 2012; Majer *et al.*, 2013). However, to our knowledge, the possibility of targeting proteins to different subcellular compartments using a potyviral vector has never been researched before. One possible reason is that, in potyvirus-derived vectors, recombinant proteins have always been expressed embedded inside the viral polyprotein. Several positions in the potyviral polyprotein have been successfully used to date to express recombinant proteins, particularly the P1/HC-Pro and the NIB/CP intercistronic positions (Dolja *et al.*, 1997; Whitham *et al.*, 1999; Fernández-Fernández *et al.*, 2001; Dietrich and Maiss, 2003; Beauchemin *et al.*, 2005) and inside the P1 cistron (Rajamäki *et al.*, 2005).

Interestingly, in our recent research into relocation of the NIB cistron in the TEV genome, we learned that in addition to the P1/HC-Pro intercistronic position, relocating the NIB cistron to the outermost amino-terminal end of the polyprotein did not affect virus viability, opening the possibility to express recombinant proteins from this position using a TEV-derived vector (Majer *et al.*, 2014). So, in this work we analyzed the possibility to target recombinant proteins to three different subcellular compartments (chloroplast, nucleus and mitochondria) from two substantially different insertion sites in a potyvirus-derived vector, the amino-terminal end or an internal position (replacing the NIB cistron) in the TEV polyprotein.

Recombinant proteins efficiently reach the chloroplasts, nuclei and mitochondria when expressed at the amino terminus of the potyvirus polyprotein

GFP variants including a cTP, an NLS or an mTP, all of them as amino-terminal fusions, efficiently reached the chloroplasts, nucleoli and mitochondria, respectively, of infected cells when expressed from the amino terminus of the TEV polyprotein (Figure 3A). In contrast, only the GFP variant with the amino-terminal NLS reached efficiently the intended subcellular location when expressed from the internal position of the TEV polyprotein (Figure 3A). These results suggest that in the context of a potyvirus vector, a cTP or an mTP, at least those used in this work, are only able to mediate translocation of a recombinant protein to the corresponding organelle when they occupy the amino terminus of the nascent polypeptide. In contrast, they are not functional when the recombinant proteins result from proteolytic processing in an internal position of the viral polyprotein. This observation most probably reflects the necessary coordination between protein synthesis and translocation to these organelles.

When the nascent chains of chloroplastic and mitochondrial preproteins emerge from the ribosome to the cytosol, the amino-terminal presequence is quickly recognized by specific factors that, in coordination with Hsp70 chaperones, form a cytosolic guidance complex. Thus, the precursors remain unfolded until they engage with the corresponding organelle translocon. The interacting proteins are then removed, and the unfolded precursor, amino-terminal presequence first, is inserted into the translocation channel (Soll and Schleiff, 2004; Neupert and Herrmann, 2007). This situation is most likely to occur when recombinant proteins are synthesized as the most amino-terminal product of the TEV polyprotein. In contrast, a recombinant protein resulting from proteolytic processing, despite containing the proper amino-terminal signal peptides may not be efficiently loaded by the translocation machinery. Tertiary structure of the polyprotein prior to its cleavage can mask the targeting information of the inner recombinant protein. In this scenario, the early binding of the organelle guidance factors and the preservation of the extensively unfolded state required for translocation could be compromised. Nonetheless, it should be taken into consideration that in our work, GFP variants with a cTP and an mTP expressed from inside the TEV polyprotein contained three extra amino-terminal amino acids (GEK) resulting from NIaPro cleavage site, which might have interfered with the activities of both signal sequences. It is also worth noting that, in the case of the GFP variants not reaching chloroplasts and mitochondria,

Western blot analysis revealed the presence of two GFP forms (Figure 4A, lanes 6 and 8). Slow forms likely contained the corresponding signal peptides, while fast forms most probably had the signal peptides processed. This may indicate some amounts of these GFP variants being still able to reach the intended subcellular location with the signal peptides being processed or, alternatively, spurious processing of the signal peptides with no relation to traffic. The lack of GFP fluorescence in chloroplasts and mitochondria of tissues infected by viruses expressing these GFP variants (Figure 3A, images IV and XII) supports the second hypothesis.

In any case, our results indicate that, in practice, delivering recombinant proteins to chloroplast and mitochondria from a potyvirus-derived vector must be accomplished from the amino terminus of the viral polyprotein. This rule may also apply to other signal peptides known to efficiently work from the amino terminus of the targeted proteins. In contrast, delivering recombinant proteins to the nucleus can be done either from the amino terminus or from the internal part of the TEV polyprotein. This result could be anticipated from previously reported successful expression of transcription factors, which efficiently activate gene expression in the nucleus of infected cells, embedded inside the polyprotein of a potyvirus vector (Bedoya *et al.*, 2010; Bedoya *et al.*, 2012; Majer *et al.*, 2013). Moreover, nucleolar localization of some TEV proteins such as P1 and NIa has been reported before (Schaad *et al.*, 1996; Martínez and Daròs, 2014).

Successful nuclear targeting of the NLS-tagged proteins independently of their position into the polyprotein most likely results from the fact that protein transport across nuclear envelope differs essentially from the transport of proteins across the membranes of chloroplast and mitochondria, in that it occurs through a large aqueous pore (nuclear pore complex, NPC). Contrary to the membrane-spanning protein transporters, NPC allows the transport of completely folded proteins and, in addition, the precise location of the NLS within the amino acid sequence is not important. NLS are thought to form loops or patches on the protein surface that are recognized by specialized receptors (Freitas and Cunha, 2009).

Together, these observations allowed us to construct a TEV-based vector to target two different recombinant proteins (GFP and mCherry) to two different subcellular locations (chloroplast and nucleus, respectively) from a single viral backbone (Figure 3B).

The type of signal peptide and the insertion position in the viral polyprotein affect recombinant protein accumulation

Western blot analysis of GFP variants in infected tissues showed differences in recombinant protein accumulation depending on the signal peptide and the insertion position in the viral polyprotein (Figure 4A; Supporting information, Table S1). The type of signal peptide and insertion position of GFP also affect TEV load (Figure 4B; Supporting information, Table S1). When GFP variants were expressed from the amino terminus of the polyprotein, good accumulations were obtained for the untagged GFP and for the GFP variants tagged with the cTP and NLS. Only the GFP tagged with the mTP accumulated poorly (Figure 4A; Supporting information, Table S1). In contrast all tagged GFP variants expressed from the internal part of the TEV polyprotein accumulated in lower amount compared with the untagged GFP (Figure 4A; Supporting information, Table S1). These results suggest that, in some cases, signal peptides negatively influence recombinant protein accumulation. The GFP variants with cTP and NLS expressed as the amino-terminal products of the polyprotein are exceptions. Higher accumulation in these two cases (Figure 4A, lanes 2 and 3; Supporting information, Table S1) may result from successful subcellular localization that, in the case of the chloroplastic variant, is apparently associated to efficient cTP processing. Subcellular targeting has been shown to play an important role in determining the yield of recombinant protein (Fischer *et al.*, 2004).

Position of extra genes in the potyvirus genome affects the stability of recombinant viruses

Our results also indicate that cDNAs coding for recombinant proteins are less stable in the TEV genome when inserted in the 5' end of the viral open reading frame than when replacing the NIb cistron (Figure 5, compare A and B). The insertion site, the nature of the recombinant protein and the duration of viral infection have been previously shown to influence stability of heterologous sequences in potyviral genomes (Dolja *et al.*, 1992; Choi *et al.*, 2000; Arazi *et al.*, 2001; Beauchemin *et al.*, 2005; Zwart *et al.*, 2014). This observation indicates that future work is needed to analyze why the inserted cDNA is less stable in this position of the TEV genome and what sequence manipulation can be done to increase stability, for example by breaking homology of sequence regions that could mediate recombination during virus RNA replication. What

sequence is inserted and where in the virus genome have also been shown to affect symptoms of infection (Dolja *et al.*, 1997; Guo *et al.*, 1998; German-Retana *et al.*, 2000). In our work, insertion of the cDNAs coding for all assayed GFP variants at the 5' end of the TEV open reading frame led to milder symptoms than insertion inside the viral open reading frame (Figure 2). In the case of untagged GFP and GFP tagged with a cTP and an NLS, milder symptoms were also associated with higher accumulation of recombinant proteins (Figure 4A, lanes 1 to 3). High accumulation of recombinant proteins and mild symptoms of infection are usually desirable properties from a biotechnological view.

Conclusively, results of this work will help to design potyvirus-based vectors to co-express several proteins in plant biofactories, particularly when it is desired recombinant proteins reach different subcellular localizations.

Acknowledgement

We thank Verónica Aragonés for excellent technical assistance. We also thank Ash N. Watson (Smurfit Institute of Genetics, Trinity College Dublin, Ireland) for English edition. This research was supported by grants BIO2011-26741 and BIO2014-54269-R from Ministerio de Economía y Competitividad (MINECO, Spain) to J.-A.D. E.M. was the recipient of a predoctoral fellowship (AP2012-3751) from Ministerio de Educación, Cultura y Deporte (Spain). J.-A.N was supported by grant BIO2011-25018 (MINECO).

References

- Arazi T, Slutsky SG, Shibolet Y, Wang Y, Rubinstein M, Barak S, Yang J, Gal-On A. 2001. Engineering zucchini yellow mosaic potyvirus as a non-pathogenic vector for expression of heterologous proteins in cucurbits. *J Biotechnol.* 87(1): 67-82.
- Beauchemin C, Bougie V, Laliberté JF. 2005. Simultaneous production of two foreign proteins from a polyvirus-based vector. *Virus Res.* 112(1-2): 1-8.
- Bedoya L, Martínez F, Rubio L, Daròs JA. 2010. Simultaneous equimolar expression of multiple proteins in plants from a disarmed potyvirus vector. *J Biotechnol.* 150(2): 268-275.

Bedoya LC, Daròs JA. 2010. Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* 149(2): 234-240.

Bedoya LC, Martínez F, Orzáez D, Daròs JA. 2012. Visual tracking of plant virus infection and movement using a reporter MYB transcription factor that activates anthocyanin biosynthesis. *Plant Physiol.* 158(3): 1130-1138.

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J, Martin C. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat Biotechnol.* 26(11): 1301-1308.

Carrington JC, Freed DD, Oh CS. 1990. Expression of potyviral polyproteins in transgenic plants reveals three proteolytic activities required for complete processing. *EMBO J.* 9(5): 1347-1353.

Chen Q, He J, Phoolcharoen W, Mason HS. 2011. Geminiviral vectors based on bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in plants. *Hum Vaccin.* 7(3): 331-338.

Choi IR, Stenger DC, Morris TJ, French R. 2000. A plant virus vector for systemic expression of foreign genes in cereals. *Plant J.* 23(4): 547-555.

Chung BYW, Miller WA, Atkins JF, Firth AE. 2008. An overlapping essential gene in the Potyviridae. *Proc Natl Acad Sci USA.* 105(15): 5897-5902.

Dietrich C, Maiss E. 2003. Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected *Nicotiana benthamiana* plants. *J Gen Virol.* 84(Pt 10): 2871-2876.

Dolja VV, McBride HJ, Carrington JC. 1992. Tagging of plant potyvirus replication and movement by insertion of b-glucuronidase into the viral polyprotein. *Proc Natl Acad Sci USA.* 89(21): 10208-10212.

Dolja VV, Hong J, Keller KE, Martin RR, Peremyslov VV. 1997. Suppression of potyvirus infection by coexpressed closterovirus protein. *Virology.* 234(2): 243-252.

- Engler C, Gruetzner R, Kandzia R, Marillonnet S. 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One*. 4(5): e5553.
- Fernández-Fernández MR, Mouriño M, Rivera J, Rodríguez F, Plana-Durán J, García JA. 2001. Protection of rabbits against rabbit hemorrhagic disease virus by immunization with the VP60 protein expressed in plants with a potyvirus-based vector. *Virology*. 280(2): 283-291.
- Fischer R, Stoger E, Schillberg S, Christou P, Twyman RM. 2004. Plant-based production of biopharmaceuticals. *Curr Opin Plant Bio*. 7(2): 152-158.
- Freitas N, Cunha C. 2009. Mechanisms and signals for the nuclear import of proteins. *Curr Genomics*. 10(8): 550-557.
- German-Retana S, Candresse T, Alias E, Delbos RP, Le Gall O. 2000. Effects of green fluorescent protein or beta-glucuronidase tagging on the accumulation and pathogenicity of a resistance-breaking *Lettuce mosaic virus* isolate in susceptible and resistant lettuce cultivars. *Mol Plant Microbe Interact*. 13(3): 316-324.
- Gibbs A, Ohshima K. 2010. Potyviruses and the digital revolution. *Annu Rev Phytopathol*. 48: 205-223.
- Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, Gleba Y. 2006. Rapid high-yield expression of full-size IgG antibodies in plants coinfecting with noncompeting viral vectors. *Proc Natl Acad Sci USA*. 103(40): 14701-14706.
- Gleba YY, Tusé D, Giritch A. 2014. Plant viral vectors for delivery by *Agrobacterium*. *Curr Top Microbiol Immunol*. 375: 155-192.
- Guo HS, López-Moya JJ, García JA. 1998. Susceptibility to recombination rearrangements of a chimeric plum pox potyvirus genome after insertion of a foreign gene. *Virus Res*. 57(2): 183-195.
- Heinig U, Gutensohn M, Dudareva N, Aharoni A. 2013. The challenges of cellular compartmentalization in plant metabolic engineering. *Curr Opin Biotechnol*. 24(2): 239-246.

Hurt EC, Müller U, Schatz G. 1985. The first twelve amino acids of a yeast mitochondrial outer membrane protein can direct a nuclear-coded cytochrome oxidase subunit to the mitochondrial inner membrane. *EMBO J.* 4(13A): 3509-3518.

Kalderon D, Roberts BL, Richardson WD, Smith AE. 1984. A short amino acid sequence able to specify nuclear location. *Cell.* 39(3 Pt 2): 499-509.

Kelloniemi J, Mäkinen K, Valkonen JP. 2008. Three heterologous proteins simultaneously expressed from a chimeric potyvirus: infectivity, stability and the correlation of genome and virion lengths. *Virus Res.* 135(2): 282-291.

Köhler RH, Zipfel WR, Webb WW, Hanson MR. 1997. The green fluorescent protein as a marker to visualize plant mitochondria *in vivo*. *Plant J.* 11(3): 613-621.

Lassner MW, Jones A, Daubert S, Comai L. 1991. Targeting of T7 RNA polymerase to tobacco nuclei mediated by an SV40 nuclear location signal. *Plant Mol Biol.* 17(2): 229-234.

Li XH, Carrington JC. 1995. Complementation of tobacco etch potyvirus mutants by active RNA polymerase expressed in transgenic cells. *Proc Natl Acad Sci USA.* 92(2): 457-461.

Majer E, Daròs JA, Zwart MP. 2013. Stability and fitness impact of the visually discernible Roseal marker in the *Tobacco etch virus* genome. *Viruses.* 5(9): 2153-2168.

Majer E, Salvador Z, Zwart MP, Willemsen A, Elena SF, Daròs JA. 2014. Relocation of the NIb gene in the tobacco etch potyvirus genome. *J Virol.* 88(8): 4586-4590.

Martínez F, Daròs JA. 2014. *Tobacco etch virus* protein P1 traffics to the nucleolus and associates with the host 60S ribosomal subunits during infection. *J Virol.* 88(18): 10725-10737.

Masuta C, Yamana T, Tacahashi Y, Uyeda I, Sato M, Ueda S, Matsumura T. 2000. Development of clover yellow vein virus as an efficient, stable gene-expression system for legume species. *Plant J.* 23(4): 539-546.

- Mazur BJ, Chui CF. 1985. Sequence of a genomic DNA clone for the small subunit of ribulose bis-phosphate carboxylase-oxygenase from tobacco. *Nucleic Acids Res.* 13(7): 2373-2386.
- Neupert W, Herrmann JM. 2007. Translocation of proteins into mitochondria. *Annu Rev Biochem.* 76: 723-749.
- Peyret H, Lomonossoff GP. 2013. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mo. Biol.* 83(1-2): 51-58.
- Pogue GP, Lindbo JA, Garger SJ, Fitzmaurice WP. 2002. Making an ally from an enemy: plant virology and the new agriculture. *Annu Rev Phytopathol.* 40: 45-74.
- Rajamäki ML, Kelloniemi J, Alminaitė A, Kekarainen T, Rabenstein F, Valkonen JP. 2005. A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. *Virology.* 342(1): 88-101.
- Roy G, Weisburg S, Foy K, Rabindran S, Mett V, Yusibov V. 2011. Co-expression of multiple target proteins in plants from a tobacco mosaic virus vector using a combination of homologous and heterologous subgenomic promoters. *Arch Virol.* 156(11): 2057-2061.
- Sainsbury F, Lavoie PO, D'Aoust MA, Vézina LP, Lomonossoff GP. 2008. Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. *Plant Biotechnol J.* 6(1): 82-92.
- Sainsbury F, Cañizares MC, Lomonossoff GP. 2010. *Cowpea mosaic virus*: the plant virus-based biotechnology workhorse. *Annu Rev Phytopathol.* 48: 437-55.
- Sainsbury F, Lomonossoff GP. 2014. Transient expressions of synthetic biology in plants. *Curr Opin.Plant Biol.* 19: 1-7.
- Schaad MC, Haldeman-Cahill R, Cronin S, Carrington JC. 1996. Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *J Virol.* 70(10): 7039-7048.
- Soll J, Schleiff E. 2004. Protein import into chloroplasts. *Nat Rev Mol Cell Biol.* 5(3): 198-208.

Thole V, Worland B, Snape JW, Vain P. 2007. The pCLEAN dual binary vector system for *Agrobacterium*-mediated plant transformation. *Plant Physiol.* 145(4): 1211-1219.

Urcuqui-Inchima S, Haenni AL, Bernardi F. 2001. Potyvirus proteins: a wealth of functions. *Virus Res.* 74(1-2): 157-175.

Wang Y, Cong QQ, Lan YF, Geng C, Li XD, Liang YC, Yang ZY, Zhu XP. 2014. Development of new potato virus X-based vectors for gene over-expression and gene silencing assay. *Virus Res.* 191: 62-69.

Whitham SA, Yamamoto ML, Carrington JC. 1999. Selectable viruses and altered susceptibility mutants in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA.* 96(2): 772-777.

Zhang H, Wang L, Hunter D, Voogd C, Joyce N, Davies K. 2013. A *Narcissus mosaic* viral vector system for protein expression and flavonoid production. *Plant Methods.* 9(1): 28.

Zhang Y, Butelli E, Martin C. 2014. Engineering anthocyanin biosynthesis in plants. *Curr Opin Plant Biol.* 19: 81-90.

Zwart MP, Willemsen A, Daròs JA, Elena SF. 2014. Experimental evolution of pseudogenization and gene loss in a plant RNA virus. *Mol Biol Evol.* 31(1): 121-134.

Supporting information

Table S1. Accumulation of GFP variants and TEV CP in tissues infected by the different recombinant TEV vectors. GFP and TEV CP were detected by Western blot analysis using specific antibodies coupled to luminescent reactions.

Recombinant virus	GFP form	GFP amount*	TEV CP amount*
TEVΔNIb-aGFP		0.97	0.55
TEVΔNIb-acTPGFP	slow	0.13	0.64
	fast	1.15	
TEVΔNIb-aNLSGFP		0.90	0.45
TEVΔNIb-amTPGFP	slow	0.18	0.71
	fast	0.22	
TEVΔNIb-iGFP		1.00	1.00
TEVΔNIb-icTPGFP	slow	0.24	0.87
	fast	0.25	
TEVΔNIb-iNLSGFP		0.10	0.60
TEVΔNIb-imTPGFP	slow	0.25	0.41
	fast	0.25	

*Protein amounts are normalized to the GFP and TEV CP amounts produced by TEVΔNIb-iGFP.

Figure S1. Complete sequence of recombinant TEV clones TEVΔNIb-aGFP, TEVΔNIb-acTPGFP, TEVΔNIb-aNLSGFP, TEVΔNIb-amTPGFP, TEVΔNIb-iGFP, TEVΔNIb-icTPGFP, TEVΔNIb-iNLSGFP, TEVΔNIb-imTPGFP and TEVΔNIb-acTPGFP-iNLSmCherry. cDNAs corresponding to GFP, mCherry and signal peptides are in green, red and blue, respectively. cDNAs corresponding to TEV NIaPro cleavage sites downstream or flanking recombinant protein cDNAs are on gray background. TEV sequence corresponds to Genbank accession number DQ986288 including two silent mutations (G273A and A1119G, in red). Limits between TEV cistrons are marked on blue background.

>TEVΔNIb-aGFP

```

AAAATAACAATCTCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGTCAGCAATTTAAATCATTTCTTTTAAAGCA
AAAGCAATTTTCTGAAAATTTTCACCATTTACGAAACGATAGCCATGGTGAGCAAGGGCGAGGAGCTGTTCCACGGGGTGGTGCCCATCCTGGTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAAGTTCATCTGCACCACCGGCAAG
CTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGACCTACGGCGTGCAGTGTCTCAGCCGCTACCCCGACCCATGAAGCAGCACGACTTCTTCAAGTC
CGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG
TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAAGCTGGAGTACAACCTACAACGCCACAACGCTATATCATG
GCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACCTACCAGCAGAACACCCCC
CATCGGGCAGCGCCCGTGTGCTGCTGCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTCTGC
TGGAGTTCGTGACCCGCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGACGACTGAGAATCTTTATTTTCAAGGGGAGAAGATGGCACTCATCTTT
GGCAGCTCAACGCTAACATCTCTGAAGGAAGTGTTCGGTGGAGCTCGTATGGCTTGCCTTACCAGCCACATATGGCTGGAGCGAATGGAAGCATTTTGAA
GAAGGCAGAAGAACCCTCTCGTGAATCATGCACAAACCAAGTATCTTCGGAGAAGACTACATATTACCGAGGCAGACTTGCCTTACACACCACCTCCATTTAG
AGTTCGATGCTGAAATGGAGCGGATGTATTATCTTGGTCTCGCGCGCTCACCCATGGCAAGAGACGCAAAGTTTCTGTGAATAACAAGAGGAACAGGAGA
AGGAAAGTGGCCAAAACGTCAGTGGGGCGTGATTCATTGTTGAGAAGATTGTAGTGCCCCACACCGAGAGAAAGGTTGATACCACAGCAGCAGTGGGAAGA

```

CATTGCAATGAGCTACCTCAACTTGTGCATAATAGTATGCCAAAGCAGGAAAGCAGAAAACTTCTTGCCCGCCACTTCACTAAGTAACTGCTATG
CCAAACTTGGAGCATAGTGCAGCAACCGCATATGCAAGTGGAGATCATAGCAAGAGAGCGTCCGAGCGAGGGTCAAGAGATTGGAGCTCGGTGCAA
TTGTTCCGAAAGTGTGCGTACATGTATGGCGAGAGGAAAAGGGTGGACTTACGTATTGCAACTGGCAGCAAGAGACACTTCTAGACCTTGTCTAAAAGATT
TAAGAAATGAGAGATGGATCAATCGAAGCTCACTTTTGGTTCAAGTGGCTTGGTTTGGAGGCAAGGCTCGTACGGACCTCGGCATTTGGTATGCAATGGTA
TGTCTATTGTCACCGCTCGTGGATGGATGGT
ATCTCTGAGGCATTTCTCATACTACTCTAAGAAATCTTGGAGTTGAGGCCAGATGGAATCTCCCATGAGTGTACAAGAGGAGTATCAGTTGAGCGGTG
CGGTGAGGTGGCTGCAATCTGCACAAAGCACTTCCACCGTGTGGTAAAGTACATGCAAAACCTTGCATGGTTGAAACACCTGACATTTGGTGGAGGTGAGT
CGGGAGACAGTGTCCCAACCAAGGTAAAGCTCTAGCAATGCTGAAAGACAGTATCCAGATTTCCCAATGGCCGAGAACTACTCACAAGTTTTCACAA
CAGAAATCACTAGTAATAACAAATTTGACAGCCTGCGTGGAGCTCAAAACAACCTCATTGGTGACCACAAACAGCTCCATTACACACGCTACTGGCTGT
CGAAATTTCTTTAAAGGCAATAAACTAACAGGGGGCGATCTCGAAGAGGCAAGCACACATATGCTTGAATAGCAAGGTTCTTGAACAACTCGCATGAA
ATATGCGCATTTGGCCACTTTGGTCTTTTTCAGAAATAAAATCTCATCGAAGGCCATGTGAATAACGCATCATGTGTGATAAACAATTTGATCAGAAATGGG
AATTTTATTTGGGGACTAAGGGTGCACAGCAAGAGGTTTCTTAAAGGATTTTCACTGAGATTGACCCAAATGAAGGATACGATAAGTATGTTATCAG
GAAACATATCAGGGGTAGCAGAAAGCTAGCAATTTGGCAATTTGATAATGTCACTGACTTCCAGACGCTCAGGCAACAAATTTCAAGGCGAAACTATTGAGC
TGAAGAAATTTGGGAATTTCACTCAATGCGGAATGGTAATTTCACTTCCATGACTTCCAGGCTCAGGCAACAAATTTCAAGGCGAAACTATTGAGC
CTAAGCATCCAACGAAGAGACATCTGGTCAATTTGGCACTTGGCGATTTCAAGTACCTAGACCTTCCAGTCTCAATGAAGAGAAAATGTATATAGCTAA
TGAAGGTTATTGCTACATGAACATTTTCTTGTCTACTAGTGAATCTCAAGGAAGAGGATGCAAAAGGACTTCCACAACTTTATAAGGGACAACAATTTGTT
CAAGCTTTGAGCGTGGCCACCTTGGTCTTTCAGAAATAAAATCTCATCGAAGGCCATGTGAATAACGCATCATGTGTGATAAACAATTTGATCAGAAATGGG
TTGGTTGATCATGACAAACAAATGCAATGTTTGGATTGATGGTCTTAGAACACAGGATACCAATGTTGAAATGAACACAACATCCAGCTTAA
TGAATTCGTTCACTCAGGTTTGGAAATCCGAAATGAAACTTACAAATGTTGAGGGATGAACCGAGATATGGTCAACAAAGGTGCAATTTGAGATGTTGATCA
AGTCCATATACAAACCACTCTCATGAAGCACTTCTGAGGAGGAGCCATACATAATTTGCTTGGCAATAGTCTCCCTTCAAAATTTAATGGCATGTAC
AATCTGGAACTTTTGGAGCGGTTACAATGTTGTTGCCAAATACAATGAGGTTAGCTAACCTCGCTGCCATCTTGTGAGCCTTGGCGCAAAAGTTAAC
TTTTGGCAGACTTGTTCGTCAGCAGCGTAATTTGATTAATGAGTATGCGCAGGTAATTTTGGCAACTGATTGACGGTGTGAGGTTAACCATTCGCTAT
CCTTAGCAATTTGGAAATTTGTTACTAATTAAGCTGGCCACCAAGAGATGTTGCAACTGCACTACTTCTTCCATTTTACCAGATGCTGAGTCAAGCTAACGAAAT
ATGTTGGAAAAAATATGTAAGGCTTTGAAGGATGCAATGGGACGAAATTAATTTGGTTGGAAAAATTTCTCCGCAATCAGGCAATCAAGAAAGCTTTGAA
ATTTGGGCAAAAGCTTTAATCATGAAAAACCCGTAGATTGCGGGCAGATATAGACTTGTCTGTGAAATCGCTTTTCAAGTCCACTTGGAACTCTCTGA
AGGAACCACTCAAGAGCGTAAATGGTGGTGCAGAAAGGTAAAGATAGCAAGAAATGCCATGCAAAAAGGGGTTTCTTCAAAATCTACAGCATGCTT
CCTGAGCTTACAAAGTTTATACAGTCTCGAGTGTCTTCTTGTGTTGACATTTCTAATTTCAAATTTGACTGATGATAAGGGCACACCAGAGGGCGAA
GGTGTGTCACAGTTGCGAAGAGAGAGCGAGTGGGCAATATCATCAATAGAATTTCCAGTATTTAAGCTTGAATACTTATGCTATCGCTCTACAG
CGAGGAAAGCTTCAATCAGACACCCCGAGGCTTTCAGTACTCAAGTTTGGATTTGGAAAGGAAAGACCTCGTTGAAACAGGCAAAACCAACCGGAGATA
GCATACTTTGAAAGATATAGCTTTCATCACTTGTATTAATGGCTTTTGGCGCTGAGCGGAGTATGGAGTGTTCAGATACTCAATAAGTTCAAAGG
AATACTGAGCTCAACGAGAGGGAGATCATCTACAGCGCACTTTGGATGATTACCTTACAACTTTGATGACAAATGCAATCAACCTCGATTTGAAAT
TGATGAACTCCACAGACAGGCTTCTCGGATCTCTTAAAGCAATTTGAAAGATGTTGAAACCAACCAACTCAGCCGAGGCAACTGAACTTATAGAACTGAG
GGGCACTTTCAGGATTTTACAGAGATCTGCGGCATCGGTTGCCAGCGAGATATCACACTCACCAGCAAGAGATTTCTTGTGAGAGGTTGCTGTTGGAT
TGGAAATCCCAAGGACTTCCATACCATTTATCAAGAGAGGGAGATTGTAATGCTTACAGCTTACCAGCACCCTCACAGATACGTTGCGACAGCACTGA
GAGTGAACCATTTTAACTGCTTCCAACTTTGAGGATGAGAGGAAAGTCACTTTTGGTGTATCACCGATTACAGTACAGTACTGATGATGCTTTTACAC
CATTTTGCGCAAAACATGACTGAGGTAAAAACATACGATTTTGCATTAATTTGATGAAATGCTGATGATGATGCTTCTGCTATAGCTTTAGGAATCTACT
GTTTGAACATGAAATTTGAGGAAAGTCTTCAAAGTGTGACAGCACCCAGTCCAGGATTTGAAATTTCAAACTGAACTTCCGCTGAACTCAAGATAG
AAGAGGCTCTTAGCTTTTCAAGAAATTTGTAAGTTTACAAAGGACAGGTGCCAACCCGATGTGATTAGTTTGGCGACAACTACTAGTAAATGTTGCTAGC
TACAATGATTTGATAGCTTGGCAAGCTCCTTGTGCAAAAGGATACAAAGTGTGCAAGATTTGAGGAAACAAATGAAGAGTGGAGGAACTGAAATTAAT
CACTGAAAGTACTTTCAGTGAAAAAGCATTTTCACTAGTCCAACTAATTAATTTGAAATGTTGTAACCAATTTGACATTTGATTTGTTGGATTGAACTA
AGGTTGTACCAGTTTGGATGTGGCAATAGAGCGGTGCAGTACAACAAAATGCTGGTGGTATTTGGGAGCGCATCCAAAGACTCGGTAGAGTTGGGCGA
CACAAAGGAGGATAGCACCTCGAATTTGGCCAAACAAATAAAACACTGGTTGCAAAATTTCCAGAAATGTTGCCACTGAACTGCTTCTATGCTTAAAGT
CAATTTGCCAGTGACAAAGCTTTCAGACCAAGCATTTTCAACCACACTGCTGGAAATTTGCAACTTATTAACAAGTGAACATTTGACATTTGACATTTT
ACACAATTAATTTTGTGCGATTGATGGTATGATGCATCCAGTACATACAGCTGAAGCGCTTTAAGCTACACACTTGTGAGACATTTCTCAATAAG
TTGGCGATCCAAATAAAGGCTTATCCTCTTGGCTTACAGATGAGAGTATAAGCGACTGGTTACATAGCAGAGGATGCTGGCATAGAAATCCCATTTGCT
GTCCAAAGAAATTTCCAGACTTCTGATGAGAAATTTGGCACATTTGTAGTTCGCCATAAAGGTTGACTCGGGTATTTGGGAGCTCACTAGCGTACAGCGA
CAAAGGTTGTTTATCTGCAAAAGGATGTGCATCAATTTGGAGGACTCTAGCATGCATCAATAGACTCATAGCACATGAACAAATGAAGCAGAGTCT
TTTTGAAGCCCACTGGGAGAGCATTTTCTTCAAAATTTACTCAATACAAAGATATTTGACACGCTGAAAGCAAAATTTGCTTCAAAGCATGCAAGAA
AAATTTGCAAGTCTTCCAGCAGGCAAAAGATCAATTTGATAGGTTTTCGAACTTAGCAAAGGATCAAGATGTACAGGATCAAGCTTCAATCAAC
TGGAAACTATCTATCTCAATCAGATAGCGAAGTGGCTAAGCATCTGAGCTTAAAAGTCACTGGAATAAAAGCCAAATCACTAGGGACATCAATAGCT
TTGCTGTGTTAATTTGGTGGTGGATGCTTGCACAGTACTTCAAGGCAAGTCAATGAAACAGTCTATTTCCAGGGAAAGAAATCAGAACAACAA
GCTTAAAGATGAGAGAGCGGCTGGGGCTAGAGGGCAATATGAGGTTGCGAGCGAGCCAGAGGGCGTAGAACATTTCTTGGAAAGCATATAAACAAG
GAAAGCGCAAGGGCACACAGAGAGGAAATGGGTGCAAAAGTCTCGGAAATTCATAAACATGATGTTGGTTGATCCAACTGATTTTCATACATAGGTTTGTG
GATCCATTTGACAGGTCACACTATTTGATGAGTCCACAAACCGCACTTATGATTTAGTGCAGATGAGTTTGGAAAGGTTAGAACCGCATTTTAAATTTGACGA
TGAGATAGGCTTCAAAGCTTTCAGACCCACACCAATCATGCTTTTGGTCTTCACTTACAAACAGCACTTGTGTAAGGTTGATTTAAAGTTTAACTCAAC
CGTCTGCTGCGAGTGGAAATCAACAGCAATAATGGGATTTCTTGAAGGGAGAAATGAAATGCGTCAAACCGCATGGCAGTGGCAGTGGCTTATGAT
CAATTTGCCAAAGAGTGGAGCTTGCAGCTTTGAGGAGAAAGCTTTGTTTAAAGGACACCGTATTAACAACCGCATTCAGAGCAACTTTGCTACTTGC
CATGGTGTGACAGACTGTTGCAACTTCCCTTCACTGATGAGGCTTGTGGAAGACTTTGGAAGACTTTGGAACCAAGGATGGGAGTGGCAGTCCATAGTAT
CACTAGAGATGGTTTCAATGTTGGTATCACTCAGCATCGAATTTCAACCAACAAACAATTTTCAACAGCGTCCGAAAACTTCAATGGAATTTG
ACAAATCAGGAGGCGAGCAGTGGGTTAGTGGTGGCGATTAATGCTGACTCAGTATTTGGGGGGCCATAAAGTTTTCATGAGCAAACTGAAAGAGC
TTTTCAGCGATTTAAGGAGCGACTCAACTCAGTGAATTTGGTGTACTCGCAAGTGGGCACCTTGGCAGCTGTGGGTGCTGGTGTGACGCTGGTAAAGAAAGATC
AAAAGGATGATAAAGTCTGAGCAGGCTTCAAAGGATAGGATGTTAATGCTGGAATCTCAGGAACTTCTCAGTCCAGCAATAATGCTATGGCCACA
AAACTTCAATTTCAAGGATGAGGGAGAGGTTGTTGAAACTTGAATCACTTTTAGGATACAAGCCACAGCAAAATGATTTGTCAAATGCTCGAGCCAC
ACATGAGCATTTTGGCCGCTGGCATCAGGCAAGTGTGACAGCCTTATGAGTGAATGAAGCAAAATGAAATAATTTGCTAAATGGATTTATGGTGTGGTGC
TAGAAATGGGACTTCCCAAAATTTGAAAGCACTTGGTATGATGGATGGTGGAGGCAAGTTTCAACCCGCTGAAACCAATGTTGAAAGCGCCGAG
CCAACTGAGGCAAAATGACACACTTCACTGACCTGGCTGAAGCGTATATTTGAGATGAGGAAATAGGGAGCGACCATACATGCTAGGTATGGTCTACA
GAGAAACTTACAGACATGATTTGTCACGCTATGCGTTGCACTTCTATGAGCTTAACTTCAAAAACACCTGTTAGAGCGGAGGAGGCGCATATGCAAAATGA
AAGCTGCTGAGTACGAAACAGTGAAGTATGTTTGGTCTTGAATGGCACTGGTACTGCAAGGAAAGCACTGAAAGCGCACAGCGCAGATGTTG
AACCGTAACTGCACACTATTTAGGGTCCGCCAGTATAGTTTCTGCGTGTCTTGGCTTTCCGCTTTTAAAGCTTATTTGTAATATATGAAATAGCTTAT
CAGTGGGACTTGGCTTGTGTTGAAATGATCTTATATGTTTAAATGCTTATTTAGTCTCACTTACGCAAGCAAGAGTGGTCAAGCTTCCGCT
CTAATTTCTCTATGATGCGAGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGG

>TEVDNIb-acTPGF

AAAAACAATCTCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCATTTCACTTCTATTGCAAGAAATTTAAATCAATTTCTTTAAAGCA
AAAGCAATTTTCTGAAATTTTCAACATTTACGAACGATAGCCATGGCTTCTCAGTCTTCTTCTGCGAGCAGTGGCCACCGCAGCAATGTTGCTCAAG
CTAATCGTTGTCACCTTTTCACTGGCTTAACTCAGCTGCTCCTACTCCCTGTTTCAAGGAGCAAAAACCTTGCATCACTTCCATTTGACCAACCGCGGA
AGAGTGCATTTGTCAGAGTGGTGGAGCAAGGGCGAGGAGCTTTCACCGGGTGGTGGCCATCTTGGTGGAGTGGCCAGCGGCACTTCAAGCGCCACA
GTTTCCGCTGTCGCGGAGGGGAGGGCGATGCCACTACGGCAAGCTGACCTTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGCTGCCCCACCC
TCGCTGACCCCTTGCATTCAGGCGTGCAGTGTCTCAGCCGCTACCCCGCACCATGAAAGCAGCAGCTTCTTCAAGTCCGCCATTTCCCGAAAGGCTTACGCT
CAGGAGCGCACCACTTCTTCAAGGACGACGGCACTACAAAGACCCCGCCAGGTTGAAGTTCAGGGGCGACCCCTGGTGAACCGCTTCAAGCTTCAAGG
CATCGACTTCAAGGAGGACGGCAACTCTTGGGCGAAGCTGGAGTCAACTACAACAGCCACACGCTCTATATCATGGCCGACAGCAGAAAGCGGCA
CTAAGTGAATTTCAAGTCCGCGCAACATCAGGAGCGGCGAGCTGCGGCACTTCCGCAACCACTTCCAGCAGAAACCCCTTCCGCGCAGCGCCGCTGCT
CTCCCGCACAACTCACTTGCAGCACCGCTGCGCCCTGAGCAAAAGACCCCAAGCAAGCGGATCACATGGTCTGCGGATTTCTGAGGCTGCTGCGCCGCG
GATCACTCTCGCATGGACGAGCTGTACAAGCAGACTGAGAACTTTATTTTCAAGGGAGAAGTGGCACTCATCTTTGGCACAGTCAACGCTAACATCC

TGAAGGAAGTGTTCGGTGGAGTTCGTATGGCTTGCCTTACCAGCGCACATATGGCTGGAGCGAATGGAAGCATTGTTGAAGAAAGCAGAAGAACCTCTCGT
GCAATCATGACAAACAGTGTACTTCGGAGAAGACACATACCTCCAGGCGAGACTTGCCTTACACACCCTCCATTAGAGGTCGATGAAATGGAGCG
GATGTATTATCTTGGTCTCGCGCTCACCCATGGCAAGAGACGCAAAAGTTCCTGTGAATAACAAGAGGAACAGGAGAAGGAAAGTGGCCAAACGTCAG
TGGGGCTGTATCCATTGTTGAAGATTTAGTGTCCCAACCGAGAAAGGTTGATACCACAGCAGCAGTGGAAAGACATTTGCAATGAAGCTTACCATT
CAACTTGGTCAATAAGTATGCAAAAGCGTAAGAAAGCAAAAACCTTCTCCGCGCACCTTCACTAAGTAACGTTGATAGCCCAAAAGTGGAGCATAGTGGC
CAAAAGCCATATGCAGTGGAGATCATTAGCAAGAGAGCGTCCGAGCGAGGGTCAAGAGATTTGAGGGCTCGGTGCAATTTGTCGAAAGTGGCGTCA
TGATGGCGAGAGAAAGGGTGGACTTACGTATTGCAACTGGCAGCAAGAGACATTTAGACCTTGTAAAAGATTTAAGAAATGAGAGAGTGGATCAA
TCGAAGCTCACTTTTGGTTCAGTGGCTAGTTTGGAGCAAGGCTCGTAGCGACCTTGGCATTGGTATCGACATGGTATGTTCACGCGTGGCTCGGT
GGATGGATGTTGGTGGATGCTCGTGGCAAGGTACGTTCCGTGTTTGTCACTCAATGACACATTAAGCGACAATCAATCTCTGAGGCATCTTCATAC
CATACTTAAGAAATCTTGGAGTTGAGCCAGATGGAACTCCCACTGAGTGTACAAGAGGATCAGTTGAGCGGTGGCGGTGAGTGGCTGCAATCTCG
ACACAAGCACTTCCACCGTGGTAAGATCACATGCAAAAGCTTGCATGTTGAAACACCTGACATTTGAGGGTGGTGGAGACAGTCCCAACCA
AGTAAGCTCCTAGCAATGCTGAAGAACAGTATCCAGATTTCCCAATGGCCGAGAACTACTCAAGAGTTTTGCACAGAAATCACTAGTAATAACAA
ATTTGACAGCTGCGTGGAGCTCAAAACACTTATGGTGACCCGAAACAGCTCCATTCACACAGCTACTGGCTGTCAGCGAAATTTGTTAAAGGCAAT
AACTCGAAGGCGCCAGTTCGAAGAGGCAAGCACATATGCTTGAACATCGCACTGAAATCGCACTGAAATAGCCGATGGCCACTTGGCCACTTGG
TTCTTTCAGAAATAAATCTCATCGAAGGCCATGTGAATAACGCATCATGTTGTGATAAACAATGATCAGAAATGGGAATTTTATTTGGGGACTAAGGG
GTGCACACGGCAAGAGGTTTCTTAAAGGATTTTCACTGAGATGACCCAAATGAAGGATACGATAAGTATGTTATCAGGAAACATATCAGGGGTAGCAGA
AAGCTAGCAATTTGGCAATTTGATAAGTGTCAATGACTGACTTCCAGAGCTCCAGCAACAAATTCAGAGCGAAACTATGAGCGAAAGCTTGGAACTCACTG
CAITTCATCGGAATGGTAATACGTGTACCCATGTTGTGTGTACTCTTGAAGATGGTAAGGCTCAATATTCGGATCTAAGCAATCCACAGAGAGAC
ATCTGGTCAATGGCACTCTGGCGATTCAAAGTACCAGACTTCCAGTTCTCAATGAAGAGAAAATGATATAGTAAATGAAGTATTTGCTACATGAACT
AATTCCTTTGCTCTACTAGTGAATGTCAAGGAAGAGGATGCAAAAGGACTTACCAGGTTTATAAGGGACACAAATTTGTCGAAAGCTGGCCGCTGGCCAC
AATGCAAGATGTTGCAACTGCATGCTACTTCTTCCATTCTTACCAGATGTCCTGAGTGTGAATACCCAGAAATTTGGTGTGATGACAAACAAA
CAATGCATGTTTGGATTCGTATGGTCTAGAACGACAGGATACCACATTTGAAAATGAACACAACATCCAGCTAATTTGAATTCGTTCACTCAGGTTGG
GAATCGGAAATGAAACATTAACAATGTTGGAGGATGAACCGAGATATGTTGCAACATCGAGGTGAACATTCGCTATCCCTAGCAATGGAAATTTGTTAC
TATTAGGCTGGCCACCAAGAGATGGACATGGCGTGGAGGAAAGGTTGATGTGTGACCTCTGAAAAGGTGCATGAAATGTTGAAAAGAACTATGTAA
AGGCTTTGAAGGATCATGGGCAATTAATCTGGTGGAAAATTTCCGCAATCAGGCATTCAGAAAGCTCTGAAATTTGGGCGAAAGCTTTTAATC
ATGAAAACCTAGATGATTCGCGCGGACATATAGACTTGTCTGTGAATTCGCTTCAAGTTCCATTTGGAACCTTGGAACTGAAAGGAACTCTCAAGAGCGT
AAATGGTGGTGAAGAAAGGTAGAGTAGCGAAGAAATGCCATGACAAAAGGGGTTTTCTCAAATCTACAGCATGCTTCTGAGCTTACAAGTTTATCA
CAGTCTCGAGTTCCTTTCTTGTGTGACATTTCTTCAAATGACTGCATGATAAGGGCACCCGAGAGGGCAAGGTGCTGCACAGTTGCAGAA
GAGAGCGAGTGGGACCAATTCGAATGAACTTTCCAGTATTTAGCTTGAAGTTCCTATGGCTATCGCTCTACAGCGGAGAAAGATCCCAATCAGA
ACACCCCGAGGCTTTCGAGTACTACAAGTTTTCGATTGAAAAGGAAAGCACTGTTGAAACGGCAAAACCAAGGAGATAGCATCTTTGAAAAGATTTATG
CTTTCACTACATGATTAATGATTTGAGCTGAGCGGATGATGGAGTTCAGATCTCAATAAGTTCAGAAATGATGACTGAGCTCAGCTCAGCGGAGG
GAGATCACTTACAGCAGATTTTGGATGATTACGTTTACACCTTTGATGACAAATGACAAATCAACCTCGAGTTGAATATGGATGACTCCACAGAGCAG
CCTTCTCGGAGTCACTTTTAAAGCAATGGTGGAAACAAACAAATCAGCCGAGCAAGCTGAAAGCCATTTATAGAACTGAGGGCACTTCAAGGATTTACCA
GAGATCTCGGCACTCGTTGCAAGGAGATACACACTCCCGCAAGAGATTTTCTGTAAGATGCTTCTGCTATAGCGTTTGAAGTCTACTGTTGAATTTGAAG
TACCATTTATCAAAGAGAGGGAGAGTGTAAATGCTTGGCCTACAGACCACTCACAGATAACGTGCACAAGCAACTGAGAAGTGAACCAATTTAAGTCTT
CCCACTTTGAGGATGAGGGGAAGTCAACTTTTGGTCTATCCAGATACAGCTAGCTAGCTAGTGGATTCGCTTACACCAATTTTGCACAGAAATCAAGCTG
AGTAAAACATACAGTATTTGTCATAATGTGATGATGCTCATGTGAATGATGCTTCTGCTATAGCGTTTGAAGTCTACTGTTGAATTTGAAGTGAAG
AAAGTCTCAAAGTGTGAGCCACACCAGGTAGAGAAGTTGAATTCACAACCTCAGTTTCCCGTGAACCTCAAGATAGAAGAGGCTCTTAGCTTTCAGGA
ATTTGTAAAGTTTACAAGGGCAAGTGGCAAGCCGATGTGATAGTTTGGGCGACACATACTAGTATATGCTGCTACTCAATGATGTTGATGATCTTGT
GCAAGCTCCTTGTGCAAAAGGGATACAAGTGTGCAAGATTTGATGAAAGCAATGAAAGTGGAGGAACTGAAATAATCACTGAAATGATTTGATGAA
AAGCAATTCATAGTGCACAAATAATATTGAGAATGGTGAACCAATGACATTTGATGATGATGTTGGATTTTGGGACTAAGGTTTACCAGTTTTGGATG
GGACAATGAGCCGGTGCAGTACAACAACTGTGGTGAATTTGGGAGGCGCATCAAAAGACTCGGTAGAGTTGGGCGACACAAGGAAGGATGAGCATCTC
GAATTTGGCAACAAATAAAGCACTGTTGAAATTCAGAAATTTGGCTGAACTGCTTCTATGCTTCTATGCTTCAATGTAATTTGGCAATTTGCAACACAG
AGTGTTTCAACCACTGCTGAAAATGCCACATATTACAAGCTAGAATGATGACAGTTTGAGCTATCATATTTTACAATAAATTTTGTGCGATT
TGATGTGAT
TAFTCTTGGCTTACGAGTGGAGAGTATAAGCGACTTGGTATACATAGCAGGAGGATGCTGGCATAGAAGTCCCAATTCGTTGCAAAAGAAATTTCAAGCTC
TTGCATGAGGAAATTTGGCATTGTAGTGCACATAAAGGTGACTCGGTATTGGGAGGCTCACTAGCGTACAGGCAGAAAGGTTGTTTATCTCTGCA
AACGGATGTGCACTCAATTCGAGGACTCTAGCATGATCAATAGACTATAGACATGAACAAATGAAGCAGAGTCAATTTGAGCGCAACTGGGAGG
CAITTTCTTCAAATAATTAATACAAGCAATTTGACACGCTGAAAGAAATTTGCTACAAGCATACGAAAGAAATTTGAGGAGTCTTGCAGCAG
GCAAAAGATCAATGCTAGAGTTTTCGAACCTAGCAAAAGGATCAAGATGTCAGGGTATCAATCAAGACTTCAATCACTGGAACACTATCTATCTCAA
AGATAGCGAAGTGGTGAAGTATGAAAGTCACTGGAATAAAGCCAAATCACTAGGGACATCAATAAGCTTTGTCTGTGTTAAATGGTGGT
GATGATGCTTGCACAGTCACTTCAAGGCAAGTCAATGAACCTGATTTTCCAAGGAAAGAAATCAGAAGCACAAGCTTAAGATGAGAGAGGCGCGT
GGGCTAGAGGGCAATATGAGTTGCGAGCGGAGCCAGAGGCGTAGAACATTAATTTGAAAGCGCATATAATAACAAGGAAAGGCGCAAGGGCACCCAG
AGGAAATGGGTGCAAGTCTCGGAAATTCATAACATGATGGGTTTGAATTCGCAACTGATTTTTCATACATAGTTTGTGGATCCATGACAGCTGACACTA
TTGATGAGTCCCAACAGCACCATTGATTTAGTGCAGCATGATTTGAAAGCTGAAACACGCAATGAAACACGCATGTTAATGACGATGAGATAGAGCTCAA
AGTCAACCCACACCAATCAATGCTTATTTGGTGAATAGTGGCAGAAAGTTCCTTAAAGTGTATTTAACCCACACTCGTTCGCTACGTCGAGTGAGAA
ATCAACAGTAAATAGGGATTTCTGAAAGGGGAGAAATGAATGGCTCAAACCGGCAATGGCAGTGGCAGTGGCTTATGATCAATGGCAGAAAGAGTGG
ACTTGAAGCTTGAAGGAAAGCTTGTAAAGGACCACTGATTAACAACCGATGCTGACACCACTTTGTCATGCAAGAAATCTGATGGGCAACACA
ACATCGTGTATGATGATGATTTGGTCCCTTCAATCAATCAAAACAGCACTTGTGTAAGAAATAATGGAACACTGTTGGTCCAACTCACTACATGTTGT
ATTTCAAGGTCAAGAACACCACTTTGCAACCAACCTCATTGATGGGAGGCAATGATAATTTTCGCATGCTTAAAGGATTTCCCACTTTCTCTCAA
AGCTGAAATTTAGAGAGCCCAAGGAGGCAATGATGCTTGTGACAAAGGCAATTTCCAACTAAGAGCATGCTAGCATGCTGTCAGACTGTTGTCAG
ACATTCCTTCACTGATGACATTTTGGAGCATTGGATTCBAACCAAGGATGGGCGTGGCAGTCCATTAGTATCAACTAGAGATGGGTTCAATGTT
TGGTATACACTCAGCATGCAATTTCCAAACACAAACAAATTTTCCAAAGCGTGGCCAAAACCTTCAATGGAATTTGTGCAAAATCAGGAGGCGCAGT
GGGTAGTGGTGGCGATTAATGCTGACTCAGTATTGGTGGGGGGGCAATTTTTCATGACAAACCTGAAGAGCTTTTTCAGGATTTGAGGAAAGCG
ACTCAACTCATGAGTGAATTTGGTGTACTCGCAAGTGGCAGTGTGGTGTGGTGTGACGCTGGTAAGAAAGAAATCAAAGGATGATAAAGTCCGCTGA
GCAGGCTTCAAAGGATAGGATGTTAATGCTGGAATTCAGGAACATTTCACTGTTCCAGCAATAAATGCTATGGCCAAAACTTCAATATCCAGGATGA
GGGAGAGGTTGTTAAATGCTGACTCACTTTTAGGATACAAGCCACAGCAAAATTTGATGCAAAATGCTCGAGCCACACTGAGCAGTGTGGCAGGTTG
CATCAGGCAGTGTAGCAGCCTATGGAGTGAATGAAGAGCAAAATGAAATATTGCTAAATGGATTTATGGTGGTGGATAGAAATGGGACTTCCCAA
TTTGAACGGAATTTGGTATTGATGGATGGTGGAGCAAGTTTCAATCCCGCTGAAACCAATGGTGAACAGCGGCAACACTGAGCAAAATTTATGA
CACATTTCACTGACCTGGCTGAAGCGTATAATGAGATGAGGAAATGAGGAGCGCAATACATGCTAGTATGGTCTACAGAGAAACATTAACAGCATGAT
TTGTGACGCTATGCGTTCGACTTCTATGAGTCACTTCAAACCAACCTGTTAGAGAGGAGGCGCATATGCAAAATGAAGCTGCTGAGTACGAAACAG
TGAAGCTAGGTTAATTTGGTCTGATGGCAAGTGGTACTGCAGAGGAGCACTGAACGGCACACAGCGCAGATGAAACCTGAACTGACACACTAT
TAGGGTTCGCGCAAGTATGCTGCTGCTTTGCTTCCGCTTTTAAAGCTTTTAAAGCTTTGTAATAATATGAAATGCTTACAGTGGGACTTGGTCTGTG
TTGAAATGGTATCTTATATGTTTAAATATGCTTATTAGTCTACTTAGCGCAAGCAAGGTGAGGTCACCTCGCTCAATCTCTTATGATGAGTGG
GAA

>TEVANIB-aNLSGF

AAAAAACAATCTCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCAATCTACTTCTATTGACAGCAATTTAAATCAATTTCTTTAAAGCA
AAAGCAATTTCTGAAAATTTTCACTTTTACGAAACGATAGCCATGCCAAAAGAGAGAGAAAGGTACCAAAGAAAGAAAAGAGGTAATGGTGAGCAAG
GCCAGGAGTTTTCACCGGGTGGTGGCCATCTGTTGG
ACCTACGGCAAGCTGACCTTGAAGTTCACTGACACCAGGCAAGCTGGCCGTCCTTGGCCACCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG
CAGCCGCTACCCGACCAATGAAGCAGCAGCTCTTCAAGTCCGCCATGCCGAAAGGCTACGCTCCAGGAGCGCACTTCTTCAAGGACAGCGCAAGGCA
ACTCAAGAGCTCCGCGCGAGGTAAGTTTCGAGGCGACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGCGCAACTCTGCGGGG
CACAGCTGGAGTCAACTCAACAGCCACAGCTCTATATCTGCGCCAGCAGAGAGCGCATCAAGGTGAATTCAGATCCGCCACAACATCGA

GGAGCCATACATAAATGTCTGGCAATAGTCTCCCTTCAATTTAATTTAATGGCCATGTACAACTCTGGAACCTTTTGAGCAGCGGTTACAAATGTGGTTGCCAA
ATACCAATGAGGTTAGCTAACTGCTGCCATCTTGTGAGCCTTGGCCAAAAGTTAATCTTTGGCAGACTTGTTCGTCCAGCAGCGTAAATTTGATTAATGAG
TATGGCAGGTAATTTTGGACAATCTGATTTGACGGTGTGAGGTTAACCATTGCTATCCCTAGCAATGAAAATTTGTTACTATTAGCTGGCCACCCAAAGA
GATGGACATGGCTTGAGGAAAGGTGCTATGCTGTGACCTCTGAAAAGGTGCAATGAAATTTGGAAAAAACTATGTAAGGCTTTGAAAGGCTGATGATGG
ACGAAATTAATCTGGTTGGAAAAATTTCCGCAATCAGGCATTCAAAGAGCTCTGAAAATTTGGCCGAAAAGCTTTAATCATGAAAACACCGTGAATTTGC
GGCGGACATATAGACTTGTCTGTGAAATCGCTTTTCAAGTTCACCTTGGAACTCTGAAAGGAAACCTCTCAAGAGCCGTAATGGTGGTGAAGAAAGGT
AAGAGTAGCGAAAGTCCCATGACAAAAGGGGTTTTTCTCAAAATCTACAGCATGCTTCTGACGCTTACAAGTTTATCAGAGCTCGAGTGTCTTTCTCT
TGTGTTGACATTTCTATTTCAAATTTGACTGCTATGATAAGGGCACACCGGAGGCAAGGTTGCTGCACAGTTGCAGAAAGAGGCGAGTGGGCAATATC
ATCAATAGAATTTCCAGTATTTAGCTTGAATCTTATTTGGCTATCGCTCTACAGCGGAGGAAAGACTCCAATCAGAACACCCCGAGGCTTTCCAGTA
CTACAAGTTTGTGATTGAAAGGAAAGACTCGTTGAAACAAGCAAAACACCGGAGTAGCATACTTTGAAAAGATTATAGCTTTTCCATCACCTGTTATTA
TGCTTTTTCAGCTGAGCGGAGTGTGAGGAGTTCGAAAGTACTCAATTAAGTTCAAAGGAACTACTGAGCTCAACGGAGAGGAGATCATCTACACGCAAGT
TTGGATGATTACGTTTACAACCTTTGATGACAATATGACAATCAACCTCGAGTTGAATATGGATGAATCCACAAGACGAGCTTCTTGGAGTCACTTTTAA
GCAATGGTGGAAACACCAATCAGCCGAGGCAACGTTGAAGCCAAATTTAGAACTGAGGGGCACTTTCATGGAGTTTACCAGAGATACCTGGGCACTCGGTT
CCAGCGAGATATACACTACCCGCAAGAGATTTCTTGTGAGAGGTTGTTGGATTTGAGTCTGAAAATCCACAGGACTTCCATACCAATGAAAGAGG
AGAGTGTAAATGCTTTAGCTTACAGACTCAGATTAACGTTGCACAAGCAACTGAGAAGTGAACATTTAAGTGTCCCAACTTTGAGGATGAGAGG
GAAGTCAACTTTGGGTCATCACCGATTACAGTCACTAGTGGATTCGCTTTACACCATTTTGCAGAAAACATAGCTGAGGTAAAAACATAGCTTTTGG
TCATAATTTGATGAATGCTGTAATGCTTCTGCTATAGCCTTTAGAGCTTACTGTTTGAACATGAATTTGAAAGAAAAGATTTCAAAGTCTCAGCC
ACACCACCGGATAGAGAAAGTTGAATTTCAACTCAGTTTCCCGTGAAGACTCAAGATAGAAGAGCTTCTAGCTTTCCAGAAATTTGTAAGTTTACAGGGAC
AGGTGCCAACCGGATGTGATTAGTTGTGGCGACAACATACTAGTATATGTTGCTAGCTACAATGATGTTGATAGCTTTGGCAAGCTCCTTTGTGCAAAAGG
GATACAAAAGTGTGCGAAAGTGTGGAAGAACAATGAAGAGTGGGAAGCACTGAAATAATCACTGAAGGACTTCACTGAGAGGATTTCAATGATGCGCAAT
AATATTTATGAGAAATGTTGTAACATTTGACATTTGATGATGTTGTTGTTGGGACTTAAAGTTTACAGTTTGGATTTGGATTTGGACAAATAGAGCGGTGCAGTA
CAACAAAATCTGGTGGTATGGGGAGCGCATCAAAGACTCGGTAGAGTTGGGGGACACAAGGAAGGAGTAGCACCTTCAATTTGGCCAAAACAAATAAAA
CACTGGTTGAAAATCCACAAATGGTTGCCACTGAAGCTGCTTTCTATGCTTCAATGTTACAAATTTGCCAGTGACAACACAGAGTGTTCACACACTGCTG
GAAAATGCCACATTTATACAGCTAGAACTATGGCAGTTTGGCTTACATATTTTACACAATTAATTTTGTGCGATTTGATGGTAGTATGATCCAGT
CATACTGACAAGCTGAAGCGCTTAAAGCTACACACTTGTGAGACATTTCTCAATAAGTTGGCGATCCCAATAAAGGCTTATCTCTTTGGCTTACGAGT
GAGAGTAAAGGCACTTTGGTTACATAGCAGAGGATGCTGGCATAAAGTCCAAATTTGCTGTGCAAAAGAAATTTCCAGACTTCTGATGAGGAAATTTGGCAC
ATTGTAGTCCGCAATAAAGTGTACTCGGGTATTGGGAGGCTCACTAGCGTACAGGCAAGAGGTTTGTATATCTGCAACCGGATGTGCACTCAATTTGC
GAGGACTTAGCATGCAATAGACTCATAGCAGATGAACAAATGAAGCAGAGTCAATTTGAAGCCGAACTGGGAGAGCAATTTCTTCAAAAATTTACT
CAATCAAAAACATTTTGCACGCTGAAAGCAAAATTTGCTACAAAACATCGAAAGAAAATATTGCAAGTCTTCAAGGCAAAAGATCAATTTGCTAGAG
TTTTCGAACCTAGCAAGGATCAAGATGTCAGGGTATCACTCAAGACTTCAATCACTGGAAACTATCTATCTCCAATCAGATAGCGAAGTGGCTAAGCA
TCTGAAGCTTAAAAGTCACTGGAATAAAAGCAAACTCACTAGGACATCAATAAGCTTTGCTGTGTTAATGGTGGTGGATGGATGCTTGAACCTACT
TCAAGCAAAAGTTCCAAAGTGAACCTTATTTTCAAGGAAAGAAATCAGAAAGCAAGCTTAAAGTAGAGAGGCGGCTGGGCTAGAGGCAATATGAG
GTTGACGCGGAGGAGGCGCTAGAACATTTACTTTGGAAGCGCATATAAACAAGGAAAGCGCAAGGGCACCACGAGAGGATGGGTGCAAAAGTCTCG
GAAATTCATAAACATGATGGTTGATCCAACTGATTTTTCATACATAGTTTGGTGGATCCATTGACAGGTCACACTATTGATGATCCACAAACGCAAC
CTATTGATTTAGTGCAGATTTGGTTTGAAGAGTTGAAACACGATGTTAAATGACATGAGATAGAGCTTCAAAGCTTACAGCCACCAACATCCAT
GCTTATTTGGTGAATGTGCAGAGAAAGTCTTAAAGTTGATTTAAACACCACTCGTTCGCTACGTGCGAGTGAAGAAATCAACGCAATATGGGAT
TCTGAAAAGGCAATGAAATTTGCTCAAAACCGGATGGCAGTGGCTTATGATCAATTTGCCCAAAAGTGGAGTACTGACGTTTGAAGGAGAAA
GCTTTGTTAAGGGACCGTATTACAAACCGATATCGAGCACCATTTGTCATTTGACGAATGAATCTGATGGGCACACAACATCGTTGATGTTATGGTA
TTTGGTCCCTTCACTATTAACAACAGCACTTGTTTAGAAGAAATAATGAACACTGTTGGTCCAATCACTACATGTTGATTTCAAGCTCAAGAACCCAC
GATTTGCAACAACCTCTATTGATGGGAGGACATGATAAATTTCCGATGCTCAAGGATTTCCCAACTTCTCTCAAAAGCTGAAATTTAGAGACCC
AAAGGGAGAGCGCATATGCTTTGTGACAACCACTTCCAAACTAAGAGCATGCTAGCATGGTGTGACAGACTAGTTGCACATTTCCCTTCACTGATGGC
ATATTTCTGGAAAGTATGATTTCAAAACCAAGGATGGGCAAGTGTGCAGTTCGATTAATGATCAACTAGAGATGGGTTACTGTTGGTATAGCTCAGCATCGAA
TTTCCACCAACAACATTTATTTCAAAAGCGTCCGAAAACCTTCAAGGATTTGACAAAATCAGGAGGCGCAGCTAGGTTATAGGTTTGGGATTTAA
ATGCTGACTCAGTATTGTTGGGGGGCCATAAAGTTTTCATGAGCAAACCTGAAGAGCCTTTTTCAGCCAGTTAAGGAAGGAGCTCAACTCATGATGAAATG
GTGATCTCGCAAGGGAGAAAGTGGTGAGCAAGGGGAGGAGGCTTTCACCGGGGTTGGCCATCTTGGTTCGAGCTGGAGCGGCGACGTTAAACCGCCACAA
GTTTCAGGCTTCCGGGAGGCGGAGGCGATGCCACTACGGCAAGCTGGAGTAACTCAACAGCAGCCAAACGCTTATATCATGAGGCGCAAGAAAGGCGCA
TCAAGGTAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAGCAGAACACCCCATCGGCGAGGCGCCCGTGTG
CTGCCCGCAACCACTACTTGGAGCAGCCAGTCCGCCCTGAGCAAGAACCCCAAGCAAGCGGATCAATGTTGCTGCTGGATTTCTGTAAGGCGCCGCG
GATCACTCTCGGCAAGGAGCTGATCAAGACTGAGAATCTTATTTTCAAGTGGGCACTGTTGGTGTGCTGGTGTGTTGACCGCTGAAAGGAAAGATC
AAAAGGATGATAAAGTCTGAGCAGGCTTCAAAGGATAGGATGTTAATGCTGGAATTCAGGAACATTTCTAGTTCACGATAAATGCTATGGCCACA
AACTTCAATATCAAGGATGGAGGAGAGGTTGTTGTAACCTTGAATCACTTTTAGGATACAGCCACAGCAAAATGATTTGTAATGCTCGAGCCAC
ACTAGAGCTTTGGCGGCTGGGATCGAGCTGACAGCTTACAGCTTATGAGTGAATGAAGAGCAAAATGAAATATTGCTAAATGATTTGCTGGTGGTGA
TAGAAAATGGGACTTCCCAAAATTTGAACGGAACCTGGTATTGATGGATGGTGAAGGCAAGTTTCAATCCCGCTGAAACCAATGGTTGAAAACGGCGAG
CCAACACTGAGGCAAAATTTGACACACTTCACTGACCTGGCTGAAGCGTATATTGAGATGAGGAAATAGGGAGGACCATACATGCTTGGTATGGTCTGA
GACAACATACAGACATGAGTTTGTCAAGCTATGCGTTCGACTTCTATAGCTTACTTCAAAAACCTGTTAGAGCGAGGAGGCGATGATGCAAAATGA
AAGCTGCTGAGTACGAAACAGTGAAGTATTTGGTCTTGTGATGCAAGCTGGGTACTGCAGAGGAAGCACTGAACGCCACACAGCGCACGATGTA
ACCGTAAACATGCAACACTTATAGGGGTCGCGCAAGTATAGTTTCTGCGTGTCTTTGCTTTCCGCTTTTAAAGCTTATTTGTAATATATGAAATGCTATT
CAGTGGGACTTTGCTGTTGTTGAAATGATCTTATATGTTTAAATATGTTTAAATGCTTATTTAGTCTCAATTAAGGCAAGCACAAGTGAAGTCACTCGG
CTAATTTCTCTATGTAGTGCAGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>TEVANIB-icTPGFP

AAAAAACAABCTCAACACAATATACAAAACAAACGAATCTCAAGCAATCAAGCATTTACTTCTATTGCAGCAATTTAAATCAATTTCTTTAAAGCA
AAAGCAATTTTCTGAAAATTTTACCATTTCAGAACGATAGCCATGGCACTCATCTTTGGCACAGTCAACGCTAACCTCCTGAAAGAAAGTGTTCGGTGGAG
CTCGTATGGCTTGCCTTACAGCGCACATATGGCTGGAGCGAATGGAAGCAATTTGAAAGAGGCAAGAAACCTCTCGTCCAATCAATGCACAAACCAAGT
ATCTCGGAGAAAGACTACATTACCGAGGCAAGCTTTGCCCTTACACCACTCCATTTAGAGGTCGATGCTGAAATGGAGGATGTTATTTCTTTGCTGCTG
CGCGCTCCCAATGGCAAGAGCGCAAAAGTTTCTGTGAATAACAGAGGAAACAGGAAAGGAAAGTGGCCAAAACGTAAGTGGGCGGTATTTCCATTTGTTG
AGAAGATTTAGTGCACCAACAGGAGAAAGTTGATACCCACAGCAGTGAAGGAAAGCAATTTGCAATGAAGCTTACCCTCAACTTGTGCAATATAGTATG
CCAAAGCGTCAAGAAAGCAAAACTTCTTGGCCCACTTCCATTTCACTTGAAGCTGATGCCCCAAACTTGGAGCATATGCGCAACCGCAATGCAAGTGTGA
GATCATTAGCAAGAAAGCGTCCGAGCGAGGCTCAAGAGATTTGAGGGCTCGGTGCAATTTGTTGCAAGTGTGCTGCACATGATGGCGAGAGGAAAAGGG
TGGACTTACGATTTGACCACTGGCAGCAAGAGACACTTCTAGACCTTTGCTAAAAGATTTAAGAAATGAGAGAGTGGATCAATCGAAGCTCACTTTTGGTTC
AGTGGCCATGTTTGGCAAGGCTCGTAGCCACTCGGCATTTGGTATGCAATGTTGATGTTTCAATTTGTTACCGGCTCGGATCGGATGGATTTGTTGGTATG
TCGTGCGAAGGTTAAGTTCGCTGTTTGTACTCAATGACACATTTAGCGCAAACTCACTCTGAGGCACTTCTTCAATCACTACTTAAGAAAATCTTGG
AGTTGAGCCAGATGGAATTTCCATGAGTGTGAACAAGAGGATCACTGTTGAGGCGTGGGTTGAGGTTGCAATCTGACACAAGCACTTTTCCAGCTGT
GTTAAGATCACATGCAAAACGTTGATGGTTGAACAGGACTGATTTGAGGTTGAGTGGGAGCAGTGTCAACCAACAGTCACTTTCAATGCTTGTGATG
GAAAGAACAGTATCCAGATTTCCCAATGGCCGAGAACTACTCACAAGGTTTTTGAACAGAAATCACTAGTAAATACAAATTTGACAGCTGCGTGAAGC
TCAAAACCACTCATTTGTTGACCGCAAAAGCTCCATTTCAACACGCTACTGGCTGTCAGCGAAAATCTGTTTAAAGGCAATAAACTAAAGGGGCGATCT
GAAGCAACAGCACATAATTTGAAATAGCAAGGTTCTTTGAACATCGCACTGAAAATATGCGCAATTTGGCCACTTTGCTTCTTCAAGAAATAAAATCTC
ATCGAAGGCCCATGTGAATAACGCCTCATGTGTGATAATCACTTGTATGAAATGGAAATTTTATTTGGGACTAAGGGGTGCACACGAAAGAGGTTTC
TTAAAGGATTTTTCATGATTTGACCAAAATGAAGGATACGATAAGTATGTTATCAGGAAACATATCAGGGGTAGCAGAAAGCTAGCAAATTTGGCAATTTG
ATAATGTTCAACTGACTTCTCAGAGCTCAGGCAACAAATTCAGGCGAACTTATGAGCGTAAAGAAATTTGGGAATTTGGAAATGTTGATGTTGATG
TTACGTTGACCCATGTTGTTGTTTACTTGTGAAGTGGTAAAGCTCAATATTCGGATCAAAGCATCAAAGAGAGACATCTGGTCAATGGCAACTCTG
CGCATTCAAAGTACTAGACCTTCCAGTCTCAATGAAGGAAATGATATAGCTCAATGAAGGTTATTTGCTACATGAACATTTTCTTGTCTACTAGT
AATGTCAGGAAAGAGGATCAAAGGACTTCCACAAAGTTTAAAGGAAACCAATTTGTTCCAAAGCTTGGAGCGTGGCCACAAGTCAAGTGTGCAACTG
ATGCTACTTACTTTCCATTTTACCAGATGCTGAGTGTGATTTACCAGAAATTTGGTGTATGATGACAACAAACAAATGCAATGTTTGGATTTGCT

ATGGGCTAGAACGACGAGTACCACATGTTGAAAATGAACACAACTCCAGCTAATGAAATCGTTCATTACAGGTTTGGAAATCCGAAAATGAAAATCTAC
AAATGTTGGAGGATGAACGAGATATGTTGACACAAGGTCGAAATGAGATGTTGATCAAGTCCATATACAAAACCAATCTCATGAAGCAGTCTACTTGAGGA
GGAGCCATACATAAATGCTCTGGCAATAGTCTCCCTTCAATTTTAAATGCCATGTACAACCTCTGGAACCTTTGAGCAGGCGTTACAAATGTGGTTGCCAA
ATACAAATGAGTGTAGCTAACCTCGCTGCCATTTGTGAGCCTGGCCGAAAAGTTAACTTTGGGACAGCTGTTGCTGAGCAGGCAATTTGATTAATAAG
TATGGCCAGGTAATTTGGAACAATCTGATTTGACGGTGTCAAGGTTTAACTTTCCCTAGCAATGGAAAATGTTACTATTAAGCTGGCCACCAAGAA
GATGGACATGGCGTTGAGGGAAGGTGGCTATGCTGTGACCTCTGAAAAGGTGCATGAAATGTTGAAAAGAACTATGTAAGGCTTTGAGGATGCATGGG
ACGAAATTAACCTGGTTGGAAAATTTCTCCGCAATCAGGCATTCAGAAAAGCTCTTGAATTTGGGCGAAAAGCTTTAATCATGAAAACACCCGTAGATTTGC
GGCGACATATAGACTTTGCTGTGAAAATCGCTTTTCAAGTCCACTTTGGAATCTCTGAAGGAAACCATCTCAAGAGCCGTAAATGGTGGTCAAGAAAGTT
AAGAGTAGCGAAGATGCCATGACAAAAGGGGTTTTTCTCAAATCTACAGCATGCTTCTGACGCTACAAAGTTTATCACAGTCTCGAGTGCCTTTCTCT
TGTGTTGACATTTCTTAAATTTGACTGCATGATAAGGGCACCCAGGAGGCGAAGGTTGCTGCACAGTTGCAGAAAAGAGCGAGTGGGCAAAATATC
ATCAATAGAACTTTCCAGTATTTCAAGCTTGAATACTTATGGCTATCGCTCTACAGCGGAGGAAAGACTCCAATCAGAACACCCCGAGGCTTTFCGAGTA
CTACAAGTTTGTGATTTGAAAAGGAGACCTCGTTGAACAGCAAAACACCGGAGATAGCATACTTTGAAAAGATTATAGCTTTTATCACACTTGTATTA
TGGCTTTTGGCGCTGAGCGGAGTTGATGGAGTGTTCAGAGTACTCAATAAGTTCAAAGGAAATCTGAGCTCAACGGAGAGGAGATCATCTACAGCAAGT
TTGGATGATTACCTTCAACCTTTGATGACAATATGACAATCAACCTCGATTTCCGCTGAAAATTAAGTTGATGAAATCCACAAGACGAGCTTTCTGGATGCA
GCAATGGTGAACACCAATTCAGCCGAGGCAACGCTGAAGCCATTTATAGAATGAGGGGCACTTATGGAGTTTACCAGAGATACCTGCGGCACTCGGTTG
CCAGCGAGATATCACACTACCCGCAAGAGATTTTCTGTGAGAGGCTGTGTTGATCTGAAAATCCACAGGACTTCCATACCAATTTATCAAGAGAGGG
AGAGTTTAAATGCTTGAAGCTACCCAGCACTCAGCAATAACCGTGCAGCAACCACTGAGAAGTGAACCAATTTAACTGCTTCCAGAAATTTGAGGATGAGAG
GAAGTCAACTTTTGGGTCATCACGATTTACAGTCTAGTGGATTCGCTTTACACCAATTTGCAAGAAAATAGCTGAGGTAAGAAACATACGATTTTG
TCATAATGATGAAATGTCATGTGAATGATGCTTCTGCTATAGCGTTTAGGAATCTACTGTTTGAACATGAAATTTGAAGGAAAAGTCTCAAAGTGTGAGC
ACACCACGGTAGAGAAGTTGAAATTCACAACCTCAGTTTCCGCTGAAAATCAAGATAGAAGAGGCTCTTAGCTTTCAGGAAATTTGTAAGTTTACAAAGGAC
AGTGGCCAAACGCGATGTGATTTAGTTGTGGCGCAACATACATAGTATATGTTGCTAGCTACAATGATGTTGATAGTCTTGGCAAGCTCCTTGTGCAAAAG
GATACAAAGTGTGGAAGATTGATGGAAGAACAATGAAGAGTGGAGAACTGAAATAACTACTGAAGGTAATTCAGTGAAGAAAGCAATTTTATAGTGTGCAACT
AAATATTTGAGAAATGGTGTAACTTGAACATTGATGATGATGTTGGATTTTGGGCAATGAAATTTGACAGTTTGTACAGTTTGTAAAGTTTACAAAGGAC
CAACAAAATCTGGTGTAGTTATGGGAGCGCATCAAAGACTCGGTAGAGTTGGGCGACAAAGGAAGGATGAGCATTCCGAAATTTGGCAACAAATAAAA
CACTGGTTGAAATTCAGAAAATGGTTGCCACTGAAGTGCCTTTCTATGCTTCATGTACAATTTGCCAGTGAACACAGAGTGTTCACACCACACTGCTG
GAAAATGCCACATTTATACAAAGCTAGAATATGGCACAGTTTAGCTATACATATTTTACACAATTAATTTTGTGCGATTTGATGGTAGTATGCCATCGAGT
CATACATGACAAGCTGAAGGCTTTAAGCTACACACTTTGTGAGACATTTCTCAATAAGTTGGCGATCCAAATAAAGGCTTATCTCTTTGGCTTACGAGTG
GAGAGTATAAGCGACTTGGTTACATAGCAGAGGATGCTGGCATAAAGAAATCCCATTCGTTGTGCAAAAGAAATTCAGACTCTTGGATGAGGAAATTTGGCAC
ATTTAGTGTGCCATTAAGCTCAGCTCGGTTATGGGAGGCTCCTACGCTACAGCGCAAGAGGTTGTTTATACCTCAACCGGATGCAACTCAATTTG
GAGGACTCTAGCATGATCAATAGACTCATAGCACATGAACAAATGAAGCAGAGTCAATTTGAAGCCGCAACTGGGAGGAGATTTCTCTCACAAATTA
CAATACAAAGCATATTTAGCAGCTGAAAAGCAATTTATGCTACAAAGCATACGAAAAGAAATTTGCAAGTCTTCCAGCGCAAGAAATGTTGCTAGAG
TTTTCCGAACTGAAAAGATCAAGATGTCAGCGGATCATCCGAACTCAAACTGGAAACTATCTATCTCCAGTACAGTAAGCAATGCAATGCAATGCA
TCTGAAGCTTAAAAGTCACTGGAATAAAGCCAAATCACTAGGACATCAATAAGCTTTGCTGTGTTAATGGTGGTGGATGGATGCTTGAACAGTACT
TCAAGGACAAGTCAATGAGCGCTATTTCCAAAGGAAGAATCAGAGCAACAGCTTAAAGTGAAGAGGGCGGCTAGAGGGCAATATGAG
TTTTGACCGGAGGACGAGGCTTAGAACATTTACTTTGGAAGCGCATATAAACAAGGAAAGCGCAAGGCAACCCAGGCAACCCAGGAAATGGTGTGCAAGTCTCG
GAAATTCATAAACAATGATGGTTTGTATCCAACTGATTTTTCATACATATAGGTTTGTGGAATCCATGACAGGTCACACTATTTGATGAGTCCCAACACG
CTATTTGATTTAGTGTGACATGTTTGGAAAGTTTGAACACCGCATGTTAATGACGATGAGATAGAGCTCAAAGTCTTACAGCCACCAACCAATCCAT
GCTTATTTGGTGAATAGTGGCACGAAGAAGTTCTTAAAGTTGATTTAACACCACCTGTCGCTACGTCGAGTGAGAAATCAACAGCAATATGGGAT
TCTCTGAAAGCGGAAATGAAATTCGCTCAAAACCGGATGGCAGTCCGATGGCTTATGATCAATGCCCACAAAGAGTGGAGCTGAGCTTTGGAAGGAGAA
GCTTTGTTAAGGACCACTGATTTACAAACCGCATATCGAGCACCATTTTGCATTTAGCAATGAATCTGATGGGCAACCAATCTGTTGATGTTGGA
TTTTGGTCCCTTCACTATTAACAACAAGCACTTGTTTAGAAGAATAATGGAACACTGTTGGTCCATCACTACATGTTGATTTCAAGTCAAGAACACCAC
GACTTTGCAACCAACCTTATGATGGGAGGACATGATAATTTTCGATGCTTAAGGATTTCCACCAATTTCTCAAAGCTGAAATTTAGAGAGCCAC
AAAGGAAAGGCTGATGTTTGTGACAACCAACTTCCAACTAAAGCATGTTAGCATGTTGCTCAGACATAGTTGCAACATTTCCCTTCACTGATGGTGGC
ATATTTGGAAGCATTGGATTTAAACCAAGGATGGGCGTGTGGCAGTCCATTAGTATCAACTAGAGATGGGTTTCAATGTTGGTATACACTCAGCATCGAA
TTTTCAACCAACAAATATTTTCAAAAGCGTCCGCAAAACTTCAATGAAATGTTTGCAGAAAATCAGGAGGCGCAGCAGGTTAGTGGTTGGCGATTA
ATGCTGACTGATTTGTTGGGGGGCCATAAAGTTTTCATGAGCAAACTGAAAGGCTTTTTCAGCCAGTTAAGGAAAGCTACTCACTGAGTGAATTTG
GTGTACTCGCAAGGGGAAAGTGGCTTCTCAGTCTTTTCTCTGCAAGCTTGGCCACCGCAGCAATGTTGCTCAAGCTACATGTTGCACTTTTCC
TGCCCTTAACTCAGCTGCTTCAATTTCCCTGTTTCAAGGAAAGCAAACTTGACATCACTTCCATTTGCCAGCAACCGGCGAAGAGTCCAATGCTGAGCTGA
TGGTGAAGAGGGCGAGGAGCTGTTTCCCGGGTGGTGGCCATCTCTGGTGGAGCTGACAGCGGCAAGTAAAGCGCCCAAGTTTTCAGCTTTCAGGCGG
GAGGGCGATGCCACTCAGGCAAGCTGACCTGAAGTTTCACTGCAACCAGGCAAGCTGCCCCGCTGCCCCTGGCCACCTCTGTGACCACCTGACCTACGG
CGTGCAGTGTCTCAGCCGCTTACCCGACCACTGAAGCAGCAGCACTTTTCAAGTCCGCTTCCCGGAGGCTTACGTTCCAGGAGCGCACCATCTTCTTCA
AGGACGAGGCACTCAAGACCCGCGGAGGTGAAGTTTCAAGGCGCAGCACTGGTGTGAACCGCATCGAGCTGAAGGGCATCACTCTCAAGGAGGACG
AACATCTGGGCAACAAGCTGGAGTACAACACAGCCACAACCTCTATATCATGGCCGCAAGCAGAAGACGGCATCAAGGTGAATTTCAAGTCCG
CCACAACATCGAGGAGGCGAGCTGCGAGCTGCGCGACCACTACAGCAGAACACCCCATCGCGGAGGCGGCTGCTGCTGCCCAGCAACCACTACCTGA
GCACCACTGAGCCCTGAGCAAGCCGCGGAGGATGCAATGCTGCTGGAGTTCTGTCAGCCGCGGCGGATCACTCTCGGAGGATGAGGAGGAGG
CTGTACAGCAGCTGAGAACTTTATTTTCAAGTGGCAGCTGTGGTGTGGTGTGACGCTGGTAAAGAAAGATCAAAGGATGATAAAGTCCGCTGA
GCAGGCTTCAAAGGATAGGATGTTAATGCTGGAATTTCAAGAACTTCTCAGTTCCAGCAATAAATGCTATGGCCACAAACTTCAATATCCAAGGATGA
GGGAGAGTGGTGTGAACTTTGAACTTTGAACTACCTTTTAGGATACAGCCACAGCAAACTGATTTGTCAAATGCTCGAGCAACTCAGGAGGATTTGCCGCTGG
CATCAGGAGTATGACAGCCTATGGAGTGAATGAAGACAAATGAAATATTTGCTAAATGGATTTATGGTGTGGTGCATAGAAAATGGGACTTCCCAA
TTTGAACGGAACTTGGGTTATGATGGATGGTGGAGGCAAGTTTCAATCCCGCTGAAACCAATGGTTGAAAACCGCGCAGCAACACTGAGGCAAAATATGA
CACACTTCAGTGAACCTGGCTGAAGCGTATATGAGATGAGGAAATGAGGAGGACCACTATACATGCGCTAGGATGTTGCTACAGAGAAACATACAGACATGAT
TTGTCAGCTTATCGGTTGCACTTCTATGAGCTAATTTCAAACAACCTGTTAGAGCGAGGAGGCGCATATGCAAAATGAAAGCTGCTGCACTACGAAAC
TGGAACTAGGTTATTTGGTCTTGTGGAACCTGGTACTGCAAGGAAAGCACTGAACGGCACACAGCGCAGGATGTAACCGTAAACCTGACACACTAT
TAGGGTCCGCAAGTATGATGTTTCTGCTGTCTTTCCGCTTTTAAAGCTTTTAAAGTATGATAATATATGAAATAGCTATTACAGTGGGACTTGCCTTTGTG
TTGAAATGGTATCTTATATGTTTAAATATGCTTATTAAGTCTACTTATAGCGAACGCAAAAGTGAAGTCACTCGGCTAATTTCTCTATGATGAGCGA
GAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>TEVAN1b-iNLSGFp

AAAATAACAAATCTCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCATTTACTTCTATTGCAAGCAATTTAAATCAATTTCTTTTAAAGCA
AAAGCAATTTCTGAAAATTTTCAACATTTACGAACGATAGCCATGGCACTCACTTTTGGCAGAGTCAACGCTAAACATCTGCGTGAAGGAAGTGTTCGGTGGAG
CTCGTATGGTGTGTTTACCAGCCACATATGGCTGGAGCGAATGGAAGCAATTTTGAAGAAGCGAAGAAGAACCTCTCTGCAACATCACTCAACAAACAGT
ATCTTCGGAAGACTACATTTACCGAGGCAAGTTCGCTTACACACCCTCCATTTAGAGGTCGATGCTGAAAATGGAGCGGATGATTTATCTGGTCTGCTG
CGCGCTCCCAATGGCAAGAGCGCAAAAGTTCTGTGAATAACAAAGAGGAAAGGAAAGGAAAGTGGCCAAAACGCTACGTTGGGGCGTGAATTTCCATTTGTG
AGAAGTTGTAGTCCGCCCCACAGGAGAAAGTTGATACAGCAGCAGCAGTGGAGAACATTTGCAATGAAGCTACCCTCAACTTGTGCAATAATGATG
CCAAAGCGTAAAGAGCAAAAACTTCTTCCCGCCACTTCACTAAGTAACGTGTATGCCAAAATTTGAGCATAGTGGCAAAACGCCATATGCAGGTGGA
GATCATTAGCAAGAGCGCTCGAGCGAGGCTCAAGAGATTTGAGGCGCTGGTCAATTTGTCGCAAGTGTGCGTCAACTGTATGGGCGAGGAGGAAAGGG
TGGCTTACGATTTGCAACTGGCAGCAAGGACACTTTAGACTTGTCAAGAAATTTAAGAATGAGAGAGTGGATCACTGCAACTCACTTTTGGTTTCA
AGTGGCTAGTTTTGAGGCAAGGCTCGTACGGACCTGCGCATGGTATCGACATGGTATGTTTCAATTTGACGGGCTGGTGGATGGATGTTGGTGGATG
TCGTGCGAAGGTAACGTTTCGCTTTGTTGCTCACTCAATGACACATTAAGTGGCAACCAATCACTCTGAGGCAATTTCTATACATACTTAAAGAAATTTCTGG
AGTTGAGCGCAGATGGAATCTCCATGAGTGTACAAGAGGATCACTGAGGAGGCTGGTGGAGTGGCTGCAATCTGCAACCAAGCACTTTCCCGGCTGT
GGTAAAGTCACTGCAACGTTGCAATGGTTGAACACCTGACATTTGTTGAGGTTGAGTGGGAGCAGTGTACCAACCAAGGTAAGTCTCTAGCAATGCT
GAAAGAACAGTATTCGATTTCCCAATGGCCGAGAACTACTCAACAGTTTGTGCAACAGAAATCACTAGTAAATCAAAATTTGACAGCTCGCTGAGCG
TCAAAACAACATTTGGTGGCCGCAACCAAGCTCCATTCACACACTGACTGCTGAGCGAAATTTCTGTTTAAAGGCAATAAATACAGGGGGGCGCTC
GAAGGCAAGCACATATGCTTGAATAGCAAGTTCTTGAACAATGCACTGAAAATATGCGCATTTGGCCACCTTGGTCTTTCAGAAAATAAATCTC
ATCGAAGCCCAATGTAATACCGCACTCATGTGTGATAAACAATGATAGTAAGTGGGAATTTAATTTGGGCACTAAGGGGTGCAACGAAAGGTTT
TTAAAGGATTTTCTGATGATTTGACCAAAATGAAGGATACGATAAGTATGATTAATCAGGAAACATATCAGGGGTAGCAGAAGCTAGCAATTTGGCAATTTG
ATAATGCACTGACTTCCAGCGCTCAGGCAACAAATTCAGGCGAACTATTTAGCGTAAAGAAATGGGAATCACTGCATTTCAATCGGGAATGGTAA

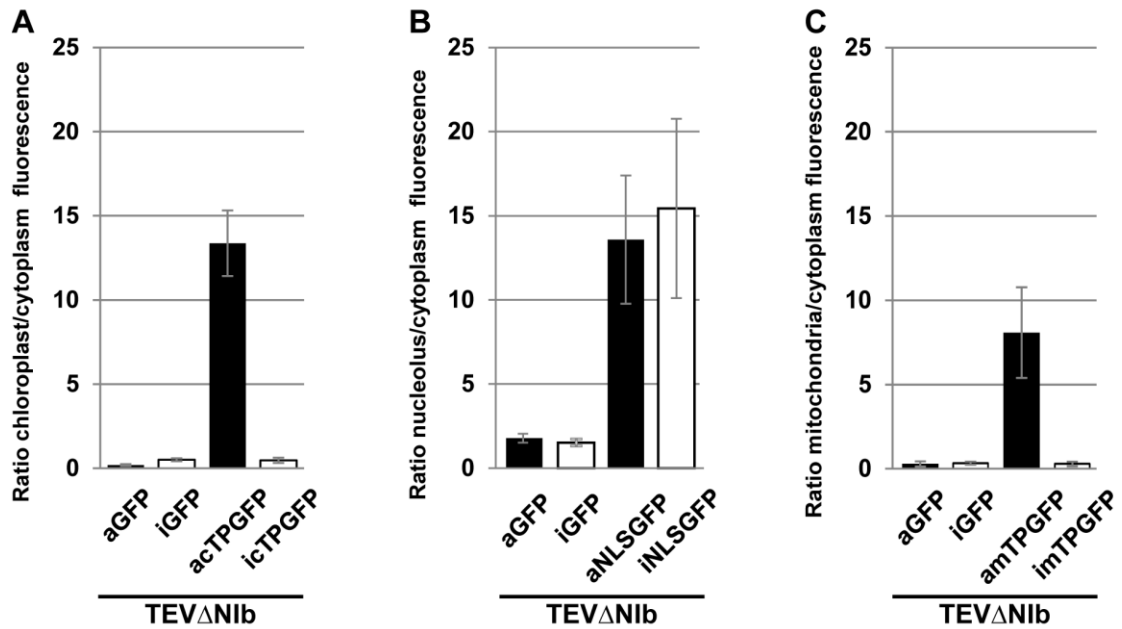
TTACGTTGACCACCTGTTGTTGTTACTCTTGAAGATGGTAAAGCTCAATATTCGGATCTAAAGCATCCAACGAAAGACATCTGGTCATTGGCAACTCTG
 GCATTCAAAAGTACCTGACCTTCAGTTCCTCAATGAAGAGAAATGTATATAGCTAATGAAGGTTATGCTACATGAACATTTCTTTGCTCTACTAGTG
 AATGTCAAGGAAGAGGATGCAAAGGACTTCACCAAGTTTATAAGGGACACAATTTGTTCCAAAGCTTGGAGCGTGGCCAAACATGCAAGATGTTGCAACTGC
 ATGCTACTTACTTTCCATTTCCACAGATGTCCTGAGTGTAAATCCCCAGAAATTTGGTTGATCATGACAACAAAACATGCAATGTTTGGATTCGT
 ATGGGTCAGAAACGACAGGATACCCACATGTTGAAAATGAACACAACATCCCCAGTAAATTTGAAATTCGTTCAATTCAGGTTTGAATCCGAAATTTAC
 AATGTTGGACGGATGAACCGAGATATGGTCACACAAGGTGCAATTGAGATGTTGATCAAGTCCATATACAAACCACATCTCATGAAGCAGTTACTTGGAGG
 GGAGCCATACATAATGTCCTGGCAATAGTCTCCCTTCAATTTAATTTGCAATGTACAACCTCTGGAACCTTTTGGAGCAGCGTTACAATGTTGGTTGCCAA
 ATACAATGAGGTTAGCTTAACCTCGCTGCCATCTTGTGACGCTTGGCGCAAAGTTAACTTTGGCAGACTTGTTCGTCAGCAGCGTAATTTGATTAATGAG
 TATGGCAGGTAATTTGGCAATCTGATGACGGTGTGAGGTTAACCATTCGCTATCCCTAGCAATGGAATTTGTTACTATTAGCTGGCCACCCAAAGA
 GATGGACATGGCGTTGAGGGAAAGTGGCTATGCTGTGACCTCTGAAAAGTGTGATGAAATTTGGAAAAAACATGTAAGGCTTTGAAAGATGCAATGGG
 ACCAATTAACCTGGTTGAAAAAATCTCCGCAATCAGGCATTCAGAAAGCTCTTGAATTTGGCGAAAGCTTTAATCATGAAAAACCCGTAGATTTGC
 GCGGACATATAGACTTGTCTGTGAAATCGCTTTTCAAGTTCACCTTGGAACTCCTGAAGGGAACCATCTCAAGAGCGTAAATGGTGGTCAAGAAAGGT
 AAGAGTAGCGAAGATGCCATGACAAAAGGGGTTTTTCTCAAAATCTACAGCATGCTTCCTGACGCTTACAAGTTTATCACAGCTCGAGTGTCTTTCTT
 TGTGTTGATGAACTTCTTGAATGACTGCTATGATAAGGCCACACCGAGGCGAAGGTTGCTGCACAGTTGCAGAAAGAGCTGACCTGGGCAATATC
 ATCAATAGAATTTCCAGTATTTAAGCTTGAATCCTATTTGGCTATCGCTCTACAGCGGAGGAAAGACTCCAATCAGAACACCCGAGGCTTTTCAGTA
 CTCACAGTTTGGATGGAAGAGGAACTCGTTGAAACAAGCAAAACACCGGAGATAGCATACTTTGAAAAGATTTAGCTTTTCCATCACACTGTATTAAT
 TGCTTTTGGAGCTGAGCGAGTGTGAGGATGTTCAAGATCTCAAATAAGTTTCAAAGGAAATCTGAGCTCAACCGGAGGAGGATCTCATACACGAAGT
 TTTGGATGATTACCTTACAACCTTTGATGACAATATGACAATCAACCTCGAGTTGAAATATGGATGAACTCCACAAGACGAGCTTCTGGATCATTTTAA
 GCAATGGTGAACAACCAATCAGCCGAGGCAACGTGAAGCCACATTTAGAACTGAGGGGCACTTTCATGGAGTTTACCAGAGATCTGCGGCATCGGTTG
 CCAGCGAGATATCACACTCACCCGCAAGAGATTTCTTGTGAGAGGTTGAGGAACTTTGGATCTGGAAATCCACAGGACTTCCATACCATTATCAAGAGAGG
 AGAGTGTAAATGCTTGAACCTACAGACTCACAGATAACGTGCACAAGCACTGAGAAGTGAACATTTAAGTGTCCAACTTTGGAGTGAAGG
 GAAGTCAACTTTTGGGTCATCACCGATTTACAGTCTGACTAGTGGATTCGCTTTACACCATTTTGCAGCAAAACATAGCTGAGGTAACCAATACGATTTTG
 TCATAATTTGATGAAATCTATGTAATGATGCTTCTGCTATAGCCTTTAGGAATCTACTGTTGAAACATGAAATTTGAAGAAAGCTCCAAAGTGTCAAGC
 ACACCACCGGTTAGAGAAGTTGAATTCACAACTCAGTTCCCGTGAACCTCAAGATAGAAGAGCTTATAGCTTTCAGGAATTTGTAAGTTTACAAAGGGC
 AGGTGCCAACCGCGATGTGATTTAGTTGTGGCGACAACATACTAGTATATGTTGCTAGCTACAATGATGTTGATAGCTTGGCAAGCTCCTTGTGCCAAAGG
 GATACAAAGTGTGCAAGATTTGATGGAAGAACAATGAAGAGTGGAGGAACTGAAATAATCACTGAAGGTTACTGAGTGAAGAAAGCTTCAATGCTCGAAC
 AATATTATTGAGAAATGTTAACCTTGAATGATGATGATGTTGTTGTTGGGACTTAAAGTGTGTTACAGTTTGGATGTTGGCAATAGAGCGGTGCAAGTA
 CAACAAACTGTTGGTGAATTTGGGGAGCGCATCAAAGACTCGGTAGAGTGGGGGACACAAGGAAGGAGTAGCATTGCAATTTGGCCAAACCAATAAAA
 CACTGTTGAAATTTCCAGAAATGTTGCCACTGAAGCTGCTTCTATGCTTCAATTTGCAATTTTGGCAGTGAACAACAGAGTGTTCATCAACCACTGCTG
 GAAAATGCCACATTTATACAAAGCTAGAACTATGGCAGCTTTGAGCTATCATATTTTACACAATTAATTTTGTGCGATTTGATGGTATGATGATCCAGT
 CATACTGACAAGCTGAAGCGCTTAAAGCTACACACTTGTGAGACCTTCTCAATAAGTTGGCGATCCCAAAATAAAGGCTTATCTCTTGGCTTACGAGT
 GAGGATGAAGTTGGTTTGGTTACATAGCAGAGGATGCTGGCATAGAATCCCAATCTCTGTGCAAAAGAAATCCAGACTTCTTCATGAGGAAATTTGGCAC
 ATTTGATGCTCCCAATAAGGCTCGGGTATTGGGAGGCTCACTAGCGTACAGCGAGCAAGGTTGTTTATACTCTGCAAAACCGGATGTGCACATTAATGGC
 GAGGACTCTAGCATGATCAATAGACTCATAGCACATGAACAAATGAAGCAGAGTCAATTTGAAGCCGCAACTGGGAGGACTTTTCTTCAAAATTTACT
 CAATACAAAGCTTGTGACAGCTGAAAGCAAAATTTGCTCAAAAGCTACGAAAGAAATAATGCAAGTCTTACAGCAGCAAAAGATCAATTTGCTAGAG
 TTTTCAAGCTTGAAGAGTCAAGATGTCACCGGATCATCAAGACTTCAACTCACTGAAAGCAATCTATCTCCATAGAGGAGGAAATGGGTCAGAAATG
 TCTGAAGCTTAAAGCACTGGAATAAAAGCCAACTACTAGGACATCATATAAGTCTTGTCTGTGTTAATTTGGTGGTGGATGGATGTTGCAACGCTGAT
 TCAAGGACAAGTTCAATGAACAGTCTATTTCACAAGGGAAGAATCAGAAGCACAAGCTTAAAGATGAGAGAGGCGGTTGGGCTAGAGGGCAATATGAG
 GTTGCAGCGGAGCAGAGGCGTGAAGCAATTTGTTGGAAGCGCATATAATAAACAAGGAAAGCGCAAGGCGCACCAGAGAGGAAATGGGTCAGAAATGCTC
 GAAATTTCAAAACACTGTTGATGGTTGATCCAACTGATTTTTCATATATATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
 CTATTGATTTAGTGCAGCATGAGTTTGGAAAGGTTAGAACACGCTATTAATTGACGATGAGATAGAGCTCAAAGTCTTAGCACCCACACCAATCCAT
 GCTTATTGTTGATGATGAGTGCAGCAAGAAAGTTCTTAAGTGTGATTTAAACACCACTCTGCTGCTACGTCAGGATGAGAAATCAACGAATTAATGGGAT
 TCTCTGAAAGGGCAATGAAATTTGGTTCAAACCGGCTGCGAAAGCACTTCAATGCAATTTGTTGCAAAATCAGGAGGCGAGGACTGGGTTAGTGGTTGAAGG
 GCTTGTTTAAGGGACCGTGTATTACAACCCGATATCGAGCACCATTGTCACCTTACGAAATGAAATCTGATGGGCACACAACATCGTTGATGTTGATTGGA
 TTTGGTCCCTTCAATATACAACCAAGCACTTGTTTAGAAGAAATAATGAACACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
 GACTTTGCAACCAACCTCATTTGATGGGAGGACATGATAATTTTCGCATGCTCAAGGATTTCCCAACTTTTCTCAAAGCTGAAATTTAGAGGCGCAC
 AAAGGGAAGAGCGCATATGCTTGTGACAACCAACTTCAAACCTAGAGCATGCTAGCATGTTGTCAGACACTAGTTGCACATTTCCCTTCACTGATGCGC
 ATATTCTGGAAGCATGTTGATTTCAAACCAAGGATGGGCAAGTGTGCACTCCATAGTATCAACTAGAGATGGGTTCAATGTTGTTGATACACTCAGCATCGAA
 TTTCCACCAACACAACAAATTTATTTCAACGCGTGCAGAAAGCACTTCAATGCAATTTGTTGCAAAATCAGGAGGCGAGGACTGGGTTAGTGGTTGGCGATTA
 ATGCTGACTCAGTATTGTTGGGGGGCCATAAAGTTTTCATGAGCAACCTGAAGAGCTTTTACGCCAGTTAAGGAGGCGACTCAACTCATGATGAAATG
 GTGTACTCGCAAGGGGAGAGTTCGCAAAAAGAGAGAAAGTTACAAAGAGGAAAGAAAGGTTAATGTTGAGCAAGGGCGAGGAGCTGTTTACCAGGCGG
 GGTCCCATCTGCTGAGCTGAGCTTAAACCGGCCACAAGTTGCAAGGCTGCTCCGGCGAGGGCGAGGCGATGCCACTTCAAGCTTCAAGCTTCAAGCTGA
 AGTTCATCTGCACCACCGGCAAGCTGCCCCGCTGCCCCGCTGCCCCCCTGTCGACCCTTACGCTACGGCGTGCAGTGTTCAGCCGCTACCCCGACCATG
 AAGCAGCAGCACTTTCTCAAGTCCGCTACCCGGAAGGCTTACGTCAGGAGCGGCACTTCTTCTCAAGGACGACGGCACTACAAGACCCCGCGCGAGGT
 GAAGTTGCAAGGGCGCACCTCTGTGAAACCGCATCGAGCTGAAGGGCATGCACTTCAAGGAGGCGGCAACATCTCGTGGTGGTGGTGGTGGTGGTGGTGGT
 ACAGCCACAACCTTATATCATGCGGCAAGCAGAAGAACCGCATCAGGTGAACTTCAAGATCCGCGCACACATCGAGGAGCGGAGCGTGCAGCTCGCC
 GACCCTTACAGCAGAAACCCCTTCCGCGAGCGGCCGCTGCTGCTGCCCCGACAACTACTTACGACCCAGTCCGCTTGAAGCAAGGAGGCAAGGCAAGCA
 GAAGCGGATTCATGCTCTGCTGGAGTTCTGAGGTTCTGAGCCGCGCGGGGATCTGCTGCGGATGGACGAGCTGTACAAGACGATAGGAATTTTATTTCAGA
 GTGGCACTGTGGTGTGTTGACGCTGTTAAGAAGAAAGATCAAAGGATGATAAAGTCTGCTGAGCAGGCTTCAAAGGATAGGATGTTAATGCTGGA
 ACTTCAGGAACATTTCTCAGTTCCAGGAATAAATGCTATGGCCACAACCTTCAATTTCAAAGGATGAGGGGAGAGGTTGGTTGAAACTTGAATCACTTTT
 AGGATACAGCCACAGCAAAATTTGTCAAATGCTCGAGCCACAATGACAGGAGTTTCCGCGTGGCATCAGGCAAGTGAACAGCCTTGAATGAGTGAATG
 AAGAGCAATGAAATATTGCTAAATGGATTTATGGTGTGGTGTATGAAATGGGACTTCCCAAAATTTGAACGGAATTTGGTATTGATGATGTTGTTGAG
 GAGCAAGTTTATACCCGCTGAAACCAATGGTTGAAAACCGCGCAACCAACTGAGGCAAAATTTAGACACTTCACTGACCTGGCTGAAGCGTATATGTTG
 GATGAGGAAATAGGGAGCGCACTATACATGCTAGGATGTTGCTACAGAGAACAATTTACAGACATGAGTTTGTACGCTATGCTGCTTCAATGATGATTA
 CTTCAAACCACTGTTAGAGCGGAGGGAGCGCATATGCAAAATGAAAGCTGCTGCACTACGAAACAGTGAAGTATGTTGCTTTGATGTTGTTGTTGTTG
 GGTACTGCGAGGAGACACTGAACGGCACACAGCGCAGCATGTAACCGTAACTGACACACTATTAGGGGTCGCGCAAGTATGTTCTGCTGCTGCTT
 TGCTTTCGCTTTTAAAGCTTATGTAATATATATGAAATGCTATTTCAAGTGGACTTGGTCTTGTGTTGAAATGGTATCTTATATGTTTAAATGCTTTA
 TTAGTCTACTTACTTAGCGCAACGACAAAGTGAAGTCACTCGCTCAATTTCTCTATGTTAGTGGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA
 AAAAAA

>TEVANIb-imTPGF

AAAATAACAAATCTCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCATCTACTTCTATTGACGAAATTTAAATCAATTTCTTTAAAGCA
 AAAGCAATTTCTGAAAATTTTACCATTATTCAGCAAGATAGCCATGGCACTCACTTTTGGCAGACTCAACGCTAACCTTGAAGGAAAGTGTTCGGTGGAG
 CTCGATGTTGGCTTCGCTTACCAGCGCACATATGGCTGGAGCGAATGGAAGCAATTTGAAAGAAAGGCAAGAAACCTCTCGTGCACAAACAGGATG
 ATCTTCGGAGAAGACTACATTTACCGAGGAGACTTGCCTTACACACCCTCCATTTAGAGGTCGATGCTGAAATGGAGCGGATGTTATCTTGGTCTGCTG
 CGCGCTACCCATGGCAAGAGAGCGCAAGTTCCTGTGAAATCAACAGAGGACAGGAAAGAAAGGATGGCCAAAACCTACGTTGGGGCGTGAATCCATTTGTT
 AGAAGATTTGAGTGGCCCAACCGGAGAGAAAGTTGATCCACAGCAGGAGGAGGAAACATTTGAAAGAAAGGCAAGCAATTTGAAGTGAAGCTACCCTCAACT
 CCAAGCGTAAAGAGCAAAAACTTCTTGGCCGCACTTCACTAAGTAACTGTATGCCAAAATTTGAGCATAGTGGCAAAACCGCATATGCAAGTGGGA
 GATCATTTAGCAAGAGAGCTCCGAGCGAGGGTCAAGAGATTTGAGGCGTGGTGAATTTGTCGCAAGTGTGCGCTCAACTGTATGGCGAGGAAAGG
 TGGCTTACGATTTGCAACTTGGCAGCAAGGAGACTTCTAGACTTGTGCAATTTGAAAGTAAAGAAATGAGAGAGTGGATCACTCGAAGCTTCTTTGGTTCA
 AGTGGCTAGTTTGGGCAAGGCTGACGCACTGCGCATTTGGTATCGACATGTTATGTTTACGCGGTGGTGGATGGATGTTGGTGGATGCT
 TCGTGGCAAGGTAAGTTCGCTTGTGCTCACTCAATGACACATTAAGGCAAAATCAACTCTGAGGCACTTCTTCAACATACACTTAAAGAAATTTCTTTG
 AGTTAGACGAGTGAATGCACTTCCATGAGTGTACAAGAGGATCATCTGAGCGGTGGGTTGGATGGGTTGGTGGGTTGGTGGGTTGGTGGGTTGGTGGG
 GGTAAAGACGATTTCCGATTTCCAAATGGCGGAGAACTACTCAAGGTTTGGCAAGAAATCACTAGTAAATAAATTTGACAGCTCGCTGGTGGGCT
 TCAACCAACTCATTTGGTGGCGCAACAGCTCCATTCACACAGCTAGCTTGGCTGAGGAAATTTCTGTTTAAAGGCAATAAATTAACAGGGGCGGCTC
 GAAGGCAAGCACACATATGCTTGAATAGCAAGTTCTTGAACAACTGCACGTGAAAATATGCGCATTTGGCCACTTGGTCTTTTCAGAAATAAATCTC

GATGTATTATTTGGTCTGCGCGCTCACCCATGGCAAGAGACGCAAAAGTTCTGTGAATAACAAGAGGAACAGGAGAAGGAAAGTGGCCAAAACCGTACG
 TGGGGCGTATTCATTTGTTGAAGAAGTTGTAGTGGCCACACGAGAGAAAGTTGATACCACAGCAGCAGTGGAGACATTTGCAATGAGCTTACCATT
 CAACTTGTGCATAATAGTATGCCAAAGCGTAAGAAGCAGAAAACCTTCTGCCCCCACTTCACTAAGTAACGTGTATGCCCAAACCTTGGAGCATAGTGGC
 CAAACGCCATATGCAGGTGGAGATCATTAGCAAGAGAGCGTCCGACGCGGGTCAAGAGATTTGAGGGCTCGGTGCAATTTGTTCCGAAAGTGTGCGTCA
 TGTATGGCGAGAGAAAGGGTGGACTTACGTATTGACAACTGGCAGCAAGAGACATTTAGACCTTGGCTAAAAGATTTAAGATGGAGAGTGGATCAA
 TCGAAGCTCACTTTGGTTCAAGTGGCTAGTTTGGAGCAAGGCTCGTACGGACCTGGCCATTGGTATCGACATGGTATGTTTACGCGGTGGTCCG
 GGATGGATGTTGGTGGATGCTCGTGCAGGAACTCGCTCGCTGTTTGTCACTCAATGACACATTAAGCGACAATCAATCTCTGAGGCATTTCTCATAC
 CATACTCTAAGAAATCTTGGAGTTGAGCCAGATGGAACTCCCATGAGTGTACAAAGAGGATATCAGTTGAGCGGTGGGTGAGTGGCTGCAATCCTG
 ACACAAGCACTTCCACGTTGGTAAGATCACATGCAAACTGTCATGGTTGAAACACCTGACATTTGTGAGGGTGGTGGAGACAGTGTCCCAACCA
 AGTAAAGCTCTAGCAATGCTGAAAGAACAGTATCCAGATTTCCCAATGGCCGAGAACTACTCACAGGTTTTTGCACAGAAATCACTAGTAATACAA
 ATTTGACAGCTCGGTGAGCGTCAAAACACTCATTTGGTACCCGCAAAACAGCTCCATTCACACACGTAAGTGGTGTGAGCGAAATCTGTTTAAAGCA
 AAATAACAGGGCCGATCTCGAAGAGGCAAGCACAATATGCTTGAATAGCAAGGTTCTTGAACAATCGCAATGAAATATGCCATTTGGCCACCTTGG
 TTCTTTCAGAAATAAAATCTCATCGAAGGCCATGTGAATAACGCACTCATGTGTGAATAAACAATTTGATCAGAATGGAAATTTTATTTGGGGACTAAGGG
 GTGCACACCGCAAGAGGTTCTTAAAGGATTTTTCATGAGATGGACCAATGTGAAAGATACGATAAGTATGTTATCAGGAAACATACGCGGTTAGCAGA
 AAGCTAGCAATTTGGCAATTTGATATGTCAACTGACTTCCAGAGCTTCCAGCAACAATTAAGCGCAAACTATTGAGCGTAAGAAATTTGGGAATCACTG
 CATTTCATCGGGAATGGTAAATACGTGTACCCATGTTGTTGTACTCTTGAAGATGGTAAGGCTCAATATTCCGATCTAAAGCATCCAACGAGAGAC
 ATCTGGTCATTGGCACTTGGCGATTTCAAAGTACCTAGACCTTCCAGTTCTCAATGAAGAGAAATGTATATAGCTAATGAAGGTTTCTGCTACATGAAC
 ATTTCTTTGCTCTACTAGTGAAATGTCAAGGAGAGGATGCAAGGACTTCCACAAGTTTATAAGGGACACAATTTGTTCCAAAGCTTGGAGCGTGGCCAAC
 AATGCAAGATGTTGCACTGCATGCTACTTACCTTCCATTTTACCAGATGTCCTGAGTGTGAATTTACCAGAAATTTTGGTTGATCATGACAAACAAA
 CAATGCAATGTTGGATTTGGATTTGATTTGAGGCTAGAACGACAGGATACCAATGTGAAATGAACCAACATCCCAGTAAATGTTTCCATTTCAATGGTTG
 GAATCCGAAATGAAACTTCAATGTTGGAGGATGAACCGAGATATGGTCAACAAGGTCATTTGAGTGTGATCAAGTCCATATACAACACCATCT
 CATGAAGCAGTTACTTGGAGGAGCCATACATAATGTCTTGGCAATGATCTCCCCTTCAATTTTAAATGGCATGTACAACCTTGGAACTTTTGGAGCAGG
 CGTTACAAAATGTTGGTTGCCAAAATACAATGAGGTTAGCTAACCTCGCTGCACTTTGTCAGCTTTCAGAGCTTGGCCAAAAGTTAACTTTGCTGCTCCAG
 CAGCGTAATTTGATTAATGAGTATGCGCAGGTAATTTGGACAATCTGATTGAGCGGTGTCAGGTTAACCATTCGCTATCCCTAGCAATGAAATTTGTTAC
 TATTAAGTGGCCCAAGAGATGGACATGGCGTTGAGGGAAGGTGGTATGCTGTGACCTCTGAAAGGTGCAATGAAATGTTGGAAAAAACAATATGTA
 AGCTTTGAAGGATGCATGGGACGAAATTAACCTGGTTGGAAAAATTTCCGCAATCAGGCATTAAGAAAGCTTCAAGAAAGCTTGGGCAAAAGCTTTAATC
 ATGAAACACCCCTAGATTGGCGGGACATATAGACTTGTCTGTGAATCGCTTTTCAAGTTCACCTTGGAACTCCTGAAAGGAACTCTCAAGAGCCGT
 AAATGTTGGTGAAGAAAGTAAAGTAGCGAAGAAATGCCATGACAAAAGGGTTTTTCTCAAAATCTACAGCATGCTTCCGCTCTACAGTTTATCA
 CAGCTCTCGAGTCTTTCTTTCTGTTGACTTTTCAATTTCAATTTGATGATGATAAGGGCACCCGAGAGGCGCACTGAGCTGCTGACAGGAA
 GAGAGCGAGTGGGCAATATCATCAATAGAACTTTCCAGTATTCTAAGCTTGAATACTTATGGCTATCGCTCTACAGCGGAGGAAAGACTCCAATCAGA
 ACCTCCGAGGCTTTTCGAGTACTACAAGTTTGCATTTGAAAGGAAGACTCGTTGAACAAGCAAAACCCGAGATAGCATACTTTGAAAGATATATAG
 CTTTCACTACACTTTGATTTAATGGCTTTTTCAGCTGAGCGGATGAGGATGTTCAAGTACTCAATAAGTTCAAAGGAAATCTGAGCTCAACGGAGG
 GAGATCATCTACACGCAAGTTTGGATGATTACGTTTCAACCTTTGATGACAATATGACAATCAACCTCGAGTTGAAATATGGATGAATCCCAAGACGAG
 CCTTCTGGGACTCTTTTAAAGCAATGGTGGAAACCAAACTCAGCCGAGGCACTGAGCCACATTTAGAACCTGAGGGGCACTTCAAGGATTTACCA
 GAGATCTCGGGCATCGTTTGGCAGGAGATATCACACTACCCGCAAGAGATTTTCTGTGAGAGGTGCTGTTGGATTTGGCAAAATCCACAGGACTTCCA
 TACCATTATCAAGAGAGGAGGAGTGTAAATGCTTGAAGCTACAGCACTCACAGATAACGTGCACAAGCAACTGAGAAGTGAACCAATTTAATCTT
 CCAACTTTGAGGATGAGAGGGAAGTCAACTTTTGGGTCTACCCGATACAGTACAGTACTGAGTGGATTTCGCTTTACACCAATTTTGCACGAAACATAGCTT
 AGGTAACAAACATAGATTTTGTCAATTTGATGAATGTCATGTGAATGATGCTTCTGCTATAGCGTTTAGGAATCTACTGTTTGAACATGAATTTGAAGGA
 AAAGTCTCAAAAGTGTGAGGCAACCCAGGATGAGAAAGTGAATTTGACCAACTGATTTCCCGTGAACCTCAAGATGAGAAGGCTTTCAGTTTTCAGGA
 ATTTGTAAAGTTTAAAGCAAGGTCGCAACCGCGATGTGATTTGTTGGCGCAACAATACTAGTATATGTTGCTACATATGATGTAGTCTTGT
 GCAAGCTCTTGTGCAAAAGGGATACAAGTGTGCAAGATTTGATGGAAGAAACAATGAAGAGTGGAGGAACGAAATAATCACTGAAGTACTTCAAGTAAA
 AAGCATTTCAATGCTCGCACTAATATATTAGAAATGGTGTAAACATTTGACATGATGTTGGATTTGGGACTAAGGTTAGCACGATTTTGGATGT
 GGCAATGAGCGGTGAGTACAACAACTGTGGTGAATTTGGGAGGCGCATTTCAAGACTCGGTAGGTTGGGCAACACAGGAAGGATGACATCTT
 GAATTTGGCCAAAACAAATAAAACACTGGTTGAAATTTCCAGAAATGGTTGCCACTGAAGCTGCCTTTCTATGCTTCATGTACAATTTGCCAGTGACAACACAG
 AGTGTTTCAACCACTGCTGGAAATTTGCCACATTTATCAAGCTAGAATTTGGCAGCTTTGAGCTATCATATTTTACACAATTAATTTTGGGCTAT
 TGATGTTGATGATGATCGCTCACTATACATGACAAGCTGAAGCGCTTTGAGCTACACATTTGTGAGACATTTCTCAATAAGTTGGGCTTTCCAAATAGGCT
 TATCTCTTGGCTTACGAGTGGAGATTAAGCGACTTGGTTACATAGCAGAGGATGCTGGCATAAGAATCCCATTCGTTGTGCAAAAGAAATCCAGACTCC
 TTGATGAGGAAATTTGGCCACATTTGATGCTCGCCATAAAGGTCAGCTCGGTTATGGGAGGCTCACTAGCGTACAGGCAAGAAAGTTGTTATATCTTGC
 AACGATGAGTGGCTCAATTTAGCGAGACTCTAGCATGATCAATAGACTATAGCACATGAACAATGAAGCAGAGTCTTTGAGGCTGCTTGAAGGCTGGGAGAG
 CATTTTCTTCACAAAATTAATCAAAAGCATATTTGACACGCTGAAAGCAAATTAATGCTACAAGCATAAGAAAGAAATATGCAAGTGTCTCAGCAG
 GCAAAAGCATCAATTTGCTAGAGTTTTCGAACCTAGCAAAAGGATCAAGATGTCACGGGTATCATCAAGACTTCACTCGGAACTTATCTATCTTCA
 AGATAGCAAGTGGCTGAAAGCTTGAAGCTTAAAGTCACTGGAATAAAGGCAAACTCACTAGGGACATCAATAATAGTCTTGTCTGTTAATTTGGTGGT
 GATGGATGCTTGCACGTAATTTCAAGGACAAGTTCAATGAACAGCTATTTCCAAGGGAAGAATCAGAAGCACAAGCTTAAGATGAGAGAGGCGCGT
 GGGGCTAGAGGGCAATGAGGTTGCAAGCGGAGCAGAGGCGCTAGAACAATTACTTTGGAAGGCAATATAATAACAAGGAAAGCGCAAGGCGCACAGG
 AGGAAATGGTTCGAAAGCTCTCGAAAGTTCAAAACATGATGTTGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
 TTGATGAGTCCACAACCGCACTTATTGATTTAGTGCAGCATGAGTTTGAAGGTTTAGAACACGCTATTTAATTTGACGATGAGATGAGGCTCAAAGTCTT
 AGCACCCACACCAATTCATGCTTATTTGGTGAATAGTGGCAGGAAGAAAGTTTAAAGGTTGATTTAAACACCACTCGTCCGCTACGTCAGGAGGAGG
 ATCAAGCTCAAGAACACCACTTGGCAACACACTCATTTGATGGGAGGACATGATAATTTATCGCATGCTCAAGGATTTCCACCAATTTCCCTCAA
 AGCTGAATTTAGAGAGCCAAAGGGAAGAGCGCATATGCTTGTGACAACCAACTTCCAAACTAAGAGCATGCTTAGCATGTTGTGAGACACTAGTGTG
 ACATTTCCCTTCACTGATGGCATATTTGGAAGCATTGGATTTCAAACAAGGATGGGCGAGTGGCAGTCCATTTAGTATCAACTAGAGATGGGTTCAATGT
 TGGTATACACTCAGCATCGAATTTACCAACACAAACAATTTATTTCAACAGGCTGCCGAAAAAATTCATGGAAATGTTGACAAAATCAGGAGGCGCAGATG
 GGTGATGTTGGTGGCGATTAATGCTGACTGATTTTGTGGGGGGCCATAAAGTTTTCATGAGCAAACTTGAAGAGCTTTTTCAGCCAGTTAAGGAAAGCG
 ACTCAACTCATGAGTGAATTTGGTGTACTCGCAAGGGGAGAAGTGGCAAAAAGAGAGAAAGGTTACCAAGAAGAAGAAGAAAGGTAATGTTAGCAAAAG
 CGAGGAGGATAAACATGGCCATCATCAAGGATTCATCGCTTCAAGGTGCAATGGAAGGCTCGGTGAACGGCCACAGTTTCAGATCGAGGGCGAGGGCG
 AGGGCCGCCCTACGAGGGCACCCAGACCCCAAGCTGAAGGTTGACCAAGGTTGGCCCTTGCCTTCGCTGGGACATCTGTCCTCAGTTTCACTGATC
 GGCTCCAAGGCTACGTGAAGCACCCCGCGCATCTCCCGACTACTTGAAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGTGAATTTGAGGA
 CGCGCGCTGGTGAACGTTACCCAGGACTCTCCCTGACAGGAGCGGAGTTTCACTACAAGGTTGAAGCTGCGCGGCAACAATCTCCCTCCGAGCGCCCG
 TAATGCAAGAAAGACCATGGCTGGGAGGCTCCTCCGAGCGGATGTACCCGAGGACCGGCGCTTGAAGGGCGAGATCAAGCAGAGGCTTGAAGCTGAAG
 CAGCGCGCCACTACGACGCTGAGTCAAGACCACTCAAAAGGCAAGAGCGGTCAGCTGCCCGGCGCTACAACGCTCAACATCAAGTTGGACATCAC
 CTCACCAACGAGGATCACTCACTCTGTGGAACAGTACGAACCGCCGAGGCGCCACTCCACCGCGGCTGGAATGAGCTGTAAGACCACTGAGAATC
 TTTATTTTCAAGTGGCACTGTGGGTTGCTGGTGTGACGCTGGTAAAGAAAGAAATCAAAAGGATGATAAAGTCTGAGCAGGCTTCAAAGGATAGGAGT
 GTTAATGCTGGAATTCAGAACATTTCTCAATTTCCAGCAATAAATGCTATGGCCAAAAACTTCAATATCAAGGATGAGGGGAGGAGGTTGTTAACTT
 GAATCACCTTTTGGATACAGCCACAGCAATTTGATTTGCAATCTGAGGACACATGAGCAGTTTGGCCGCTGGCATCAGGCGAGTATGACAGCCCT
 ATGGATGAATGAGAGCAATGAAATAATTTGCTAAATGGATTTTGGTGTGGTGCATAGAAAATGGGACTTCCCAAAATTTGAACGGAATTTGGGTTG
 ATGGATGTTGAGGCAAGTTTCAATCCCGCTGAAACCAATGTTGAAAACCGCGCAACACTGAGGCAAAATTTAGACACACTTCAAGTGTGAGTGTG
 AGCTATATTAGATGAGAAATAGGGAGGACCATATGCTTCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG
 TCTATGAGCTTAACTTCAAAACACTGTTAGAGCGAGGGAGGCAATGACAAATGAAGCTGCTGAGTACGAAACAGTGGAACTAGGTTTGGCTT
 GATGGCACTGGGTACTGCAAGAGACACTGAACGCGCACAGCGCATGTTGAAACCGTAACTTGAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG
 TCTGCTGCTTTTGGCTTTTCCGCTTTTAAAGCTTATTTGTAATATATGAAATAGCTATTTACAGTGGGACTTGGTCTTGTGTTGAATGGTATCTTATATGTT
 TAAATGCTTATTTAGTCTCACTACTTAGCGCAAGCAGCAAAAGTGAAGTCACTCGCTTAAATTTCTCTATGTAGTGCAGAAAAAATAAAAAAAAAA
 AAAAAAAAAAAAAAAAAAAAAA

Figure S2. Ratios of GFP fluorescence in the (A) chloroplast, (B) nucleolus and (C) mitochondria versus cytoplasm in tissues infected by the different recombinant viruses, as indicated. Error bars indicate standard deviation of three independent measures.



Chapter III

Virus-enabled rewiring of carotenoid biosynthesis for extraplastidial accumulation and visual tracking of infection dynamics in plants

Eszter Majer^{1,*}, Briardo Llorente^{2,*}, Manuel Rodríguez-Concepción², and José-Antonio Daròs¹

¹Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia, 46022 Valencia, Spain.

²Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB Bellaterra, 08193 Barcelona, Spain. *These two authors contributed equally to this work.

Correspondence and requests for materials should be addressed to J.-A.D. (email: jadaros@ibmcp.upv.es).

Abstract

Plants can be engineered to sustainably produce compounds of nutritional, industrial, or pharmaceutical relevance. This is, however, a challenging task as complexity of biosynthetic pathways hampers major metabolic changes. Here we describe the use of a single engineered RNA virus to produce carotenoids, plastidial pigments with health-promoting properties and industrial interest, outside plastids. We rewired metabolic circuits by connecting the endogenous plant cytosolic isoprenoid pathway to a biosynthetic pathway producing the anticancer carotenoid lycopene. We also showed that plant viruses expressing a bacterial phytoene synthase triggered a reduction in chlorophylls and an accumulation of carotenoids, hence resulting in a bright yellow pigmentation of infected tissues in various host-virus combinations. So, besides illustrating the potential of virus-derived vectors for engineering complex metabolic pathways and demonstrating the feasibility of producing carotenoids outside plastids, we also show a carotenoid-based visual reporter that can be used to visually track infection dynamics of plant viruses.

Research article

Given their photoautotrophic nature and stunning biosynthetic capacity, plants represent exceptional systems from which to sustainably obtain valuable compounds. Much interest has been shown in engineering plants to improve their nutritional quality, to increase the yield of natural compounds, or to produce novel metabolites of industrial and pharmaceutical interest (O'Connor, 2015). Compounds such as C40 carotenoids are valuable for both health-promoting and economically-relevant reasons. Generally, animals cannot synthesize carotenoids but need to take them in their diets as a source of pigments and essential retinoids (including vitamin A, mainly produced from β -carotene). In humans, dietary carotenoids such as phytoene, lycopene and lutein have also been shown to act as health-promoting phytonutrients (Fraser and Bramley, 2004; Meléndez-Martínez *et al.*, 2015). Carotenoids are widely used in the chemical, pharmaceutical, cosmetic, and feed and food industries, and thus represent really interesting targets for biotechnological production.

All photosynthetic organisms and some non-photosynthetic bacteria and fungi synthesize carotenoids. In plants, carotenoids are produced and stored in plastids

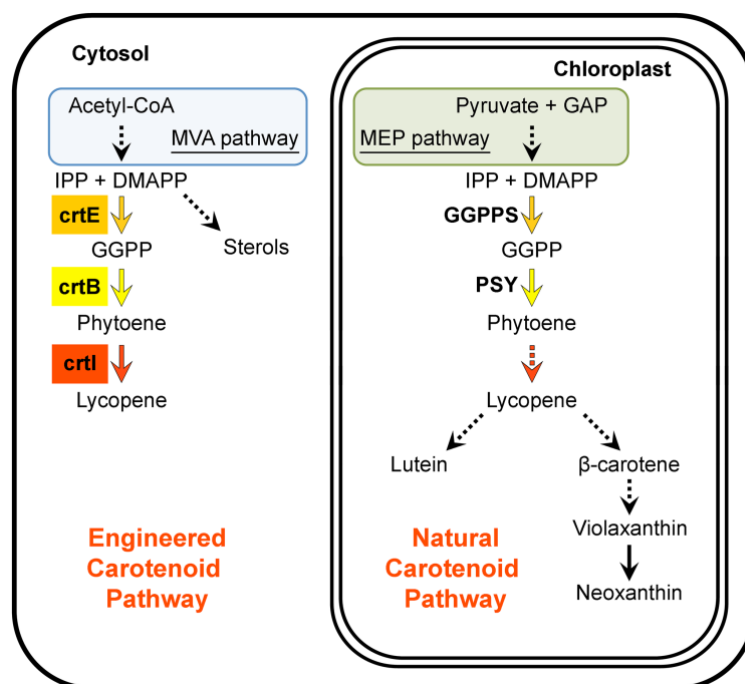


Figure 1. Schematic overview of the endogenous plant isoprenoid pathways and the virus-enabled metabolic engineering approach implemented in this study. The endogenous MEP and carotenoid pathways are localized in the chloroplast. The natural MVA pathway and the engineered carotenoid biosynthetic pathway are localized in the cytosol. Solid and dashed arrows represent single or multiple enzymatic steps, respectively. The steps catalyzed by *Pantoea ananatis* crtE, crtB and crtI are indicated. GAP, glyceraldehyde-3-phosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GGPP, geranylgeranyl diphosphate; crtE/GGPPS, GGPP synthase; crtB/PSY, phytoene synthase; crtI, phytoene desaturase.

(Figure 1). As all isoprenoids, carotenoids derive from the basic C5 building units isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In the cytosol, these units are produced by the mevalonate (MVA) pathway and are then used mainly for the biosynthesis of sterols (Figure 1). In plastids, however, IPP and DMAPP are produced by the completely unrelated methylerythritol 4-phosphate (MEP) pathway and used for the production of carotenoids and other photosynthesis-related compounds (Rodriguez-Concepcion and Boronat, 2002). The condensation of three IPP molecules with one DMAPP acceptor generates C20 geranylgeranyl diphosphate (GGPP) in a reaction catalyzed by GGPP synthase (GGPPS). The first committed step of the carotenoid pathway, catalyzed by the enzyme phytoene synthase (PSY), is phytoene production from two GGPP molecules. The desaturation and isomerization reactions catalyzed by at least four enzymes in plants eventually transform the non-colored phytoene molecule into lycopene, a red carotenoid. Further modifications of lycopene

subsequently generate β -carotene and derived xanthophylls in one branch of the pathway and lutein in the other branch (Figure 1).

Carotenoid metabolic engineering in plants has focused on potentiating, modifying, removing or adding components to the endogenous plastid-localized MEP and carotenoid pathways (Giuliano, 2014). Regrettably, engineering plant pathways toward high-level carotenoid biosynthesis has proven extremely challenging given the extensive multilevel regulatory constraints (Cazzonelli and Pogson, 2010; Ruiz-Sola and Rodríguez-Concepción, 2012; Nisar *et al.*, 2015). Furthermore, standard plant metabolic engineering typically relies on the generation of stable transgenic lines, which is a time-consuming and resource-intensive process. Development of novel metabolic engineering approaches, which are easy to implement and operate beyond the constraints set by evolution should therefore be explored to overcome current limitations and to optimize production yields. In this context, virus-derived systems circumvent the need for stable plant transformation and enable the efficient simultaneous expression of heterologous genes on a timescale of a few days (Sainsbury and Lomonosoff, 2014).

In our previous work, we induced the biosynthesis of large amounts of anthocyanins in tobacco (*Nicotiana tabacum* L.) leaves with a vector derived from *Tobacco etch virus* (TEV) that harbors *Antirrhinum majus* transcription factors Roseal and Delila (Bedoya *et al.*, 2010; Bedoya *et al.*, 2012; Majer *et al.*, 2013). TEV is a cytoplasmic replicating plus-strand RNA virus that belongs to the genus *Potyvirus* in the family *Potyviridae*. Potyviruses are useful expression systems in plants for two main reasons. First, they can accommodate large amounts of extra genetic material in their elongated and flexuous capsids without losing infectivity (Kelloniemi *et al.*, 2008). Second, their expression strategy involves the release of different proteins from a large viral polyprotein by viral-encoded proteases. Insertion into the potyviral genome of heterologous sequences separated by the sequence motifs targeted by these proteases allows the production of encoded proteins in equimolar amounts (Carrington *et al.*, 1993). In the TEV-derived expression system that we developed, the ca. 1.5 kb viral cistron that codes for the RNA-dependent RNA polymerase nuclear inclusion b (NIb) was replaced with a cassette that contains the heterologous sequences flanked by the native cleavage sites of the viral nuclear inclusion a protease (NIaPro). This strategy increases the space to accommodate extra genes while provided advantages from a

biosafety viewpoint as the TEV-derived vectors can infect only plants in which NIB is supplied in *trans* (Li and Carrington, 1995; Bedoya *et al.*, 2010).

In the present work we dramatically extend the use of viral vectors in plants to reconstruct an entire heterologous multistep pathway that rewires metabolic circuits to synthesize carotenoids from cytosolic isoprenoid precursors. We also describe a carotenoid-based metabolic reporter, which can be very useful as a color-based system to track the infection dynamics of plant viruses. We demonstrate the excellent performance of this visual marker in a series of plant-virus combinations, including co-infection with a previously described anthocyanin-based reporter (Bedoya *et al.*, 2012) to simultaneously follow the distribution of two different viral infections based on their clearly distinguishable associated colors.

Results

Redirecting cytosolic isoprenoid biosynthesis to produce extraplastidic lycopene in tobacco leaves infected with engineered viral vectors

Using the TEV-based system that we previously developed (Bedoya *et al.*, 2010; Bedoya *et al.*, 2012; Majer *et al.*, 2013), we aimed to explore whether viral vectors could be used to incorporate completely new metabolic pathways into plant cells. We chose to engineer a three-enzyme pathway from the soil bacterium *Pantoea ananatis* (Misawa *et al.*, 1990) to synthesize the health-promoting carotenoid lycopene in the cytosol of tobacco leaf cells. Lycopene is a red carotenoid that is easy to track visually if produced and accumulated at high enough levels (e.g. it provides ripe tomato and watermelon fruits with their characteristic red color). As MVA-derived IPP and DMAPP are used mostly to synthesize C15 farnesyl diphosphate for the production of downstream isoprenoids such as sterols, we hypothesized that enhanced GGPP supply might be needed for efficient carotenoid biosynthesis in the cytosol. We therefore incorporated the *P. ananatis* *crtE* gene that encodes GGPPS as the first step of the engineered pathway. The second gene of the pathway was *crtB*, which encodes PSY. The third selected gene, *crtI*, encodes the bacterial enzyme that directly converts phytoene into lycopene (Figure 1).

A cassette that contains the coding regions of *P. ananatis* genes *crtB*, *crtI*, and *crtE* separated by artificial sequences that code for a peptide efficiently cleaved by

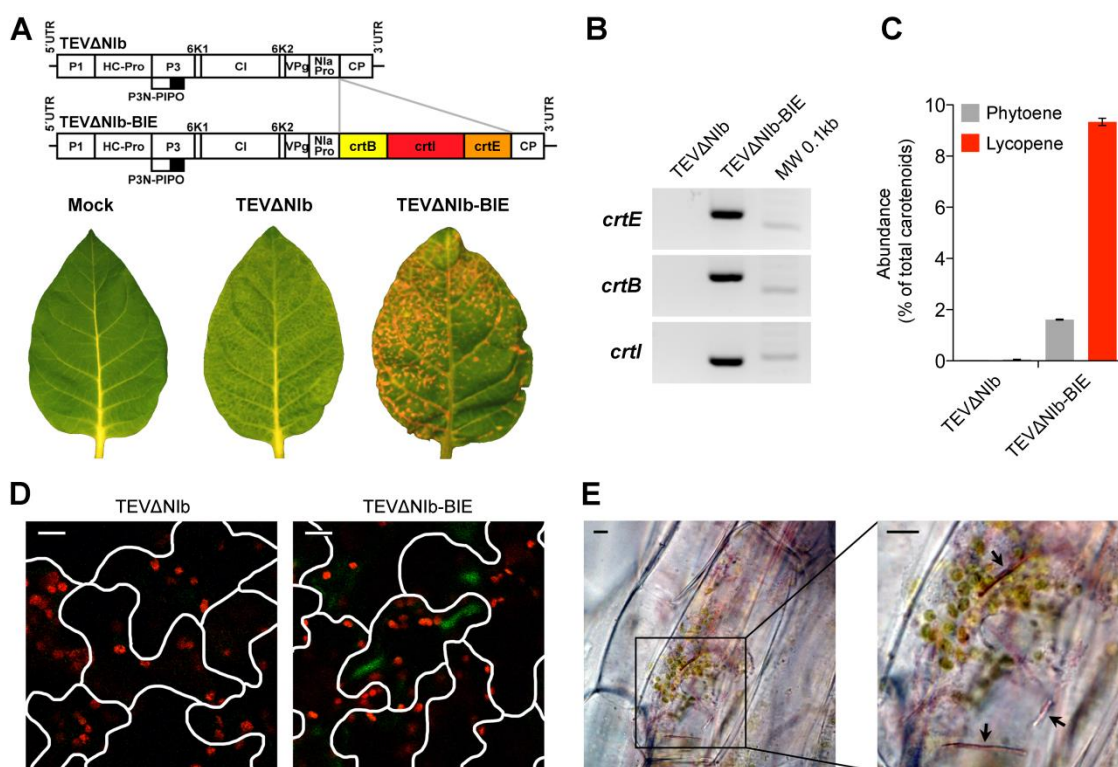


Figure 2. Production of lycopene in the cytosol of the virus-infected plant cells. **(A)** Schematic representation of TEV clones TEVΔNIB and TEVΔNIB-BIE. TEV cistrons P1, HC-Pro, P3, P3N-PIPO, 6K1, CI, 6K2, VPg, NlaPro and CP are represented by rectangles and the 5' and 3' untranslated regions (UTR) by black lines. The NIB cistron is deleted. *P. ananatis* crtB, crtI and crtE are represented by yellow, red and orange rectangles, respectively. Lower panel images correspond to representative systemic leaves of the tobacco plants mock-inoculated or infected with TEVΔNIB or TEVΔNIB-BIE. Pictures were taken at 10 dpi. **(B)** Analyses by RT-PCR and electrophoresis separation of the viral RNA that encode crtE, crtB and crtI from the symptomatic systemic leaves infected with TEVΔNIB or TEVΔNIB-BIE. **(C)** Accumulation of phytoene and lycopene in the tobacco systemic leaves infected with TEVΔNIB or TEVΔNIB-BIE. Metabolites were extracted from infected tissues and analyzed by HPLC. The average abundance relative to total carotenoid contents in n=4 independent samples is represented. Error bars indicate SEM. **(D)** Confocal microscopy detection of chlorophylls and lycopene in the tobacco leaf cells infected with TEVΔNIB or TEVΔNIB-BIE. The overlay images of the autofluorescence emitted at 650–700 nm (chlorophylls, in red) or at 500–550 nm (lycopene, in green) after excitation at 488 nm are shown. Cell boundaries were drawn using the corresponding bright field images. Bar, 10 μm. **(E)** Light microscopy observation of red lycopene crystals (marked with arrows) in vascular cells from the leaves infected with TEVΔNIB-BIE. Bar, 10 μm.

NlaPro, was inserted into the TEV vector to generate viral clone TEVΔNIB-BIE (Figure 2A and Supplementary Figure 1). This recombinant virus was used to inoculate tobacco plants that constitutively express the viral NIB protein (Li and Carrington, 1995). As a control, we used an empty viral vector that expressed no recombinant protein (TEVΔNIB; Figure 2A and Supplementary Figure 1). At 10 days post-inoculation (dpi), systemic (noninoculated) leaves from the plants infected with TEVΔNIB-BIE exhibited spots and areas with a distinctly red coloration, suggestive of

lycopene accumulation (Figure 2A). RNA analysis via specifically designed primers and reverse transcription-polymerase chain reactions (RT-PCR) indicated that *crtE*, *crtB* and *crtI* were expressed in these tissues (Figure 2B). More importantly, HPLC analysis revealed the presence of phytoene (the product of *crtB*) and high lycopene levels in the leaf areas infected with TEV Δ NiB-BIE, but not in those infected with the empty TEV Δ NiB vector (Figure 2C). Lycopene is an intermediate of the carotenoid pathway that is readily converted into lutein, β -carotene and downstream xanthophylls in chloroplasts (Figure 1), and normally goes undetected given its very low abundance in tobacco leaves (Mann *et al.*, 2000; Wurbs *et al.*, 2007; Hasunuma *et al.*, 2008; Lu *et al.*, 2013). In leaves infected with TEV Δ NiB-BIE, however, lycopene comprised approximately 10% of the total carotenoid content in the symptomatic red tissues (Figure 2C). A microscopic examination of the red-colored areas of the TEV Δ NiB-BIE-infected leaves revealed the presence of extraplastidic needle-like-shaped structures that were not present in the TEV Δ NiB-infected leaf cells. It was deduced that these structures corresponded to lycopene crystals according to their autofluorescence properties (Figure 2D) (D'Andrea *et al.*, 2014), and their distinctive red color (Figure 2E). Some of these crystals were considerably large in size, and exceeded the diameter of chloroplasts by many times (Figure 2D and E), which reinforces the hypothesis of extraplastidial biosynthesis and localization. In agreement with this conclusion, green fluorescent protein (GFP)-tagged versions of the individual enzymes used in the synthetic pathway were confirmed to be excluded from chloroplasts in transient expression assays (Supplementary Figure 2).

A yellow visual marker to track the dynamics of virus infection in plants

To gain further insight into the virally expressed heterologous pathway, we constructed three recombinant TEV clones in which each *P. ananatis* biosynthetic enzyme was expressed alone (Figure 3A and Supplementary Figure 1). The NiB-expressing tobacco plants were inoculated with these recombinant viruses, and a few dpi plants started to show typical symptoms of TEV infection. The tissues infected with TEV Δ NiB-B exhibited a bright yellow coloration at 10 dpi (Figure 3A). No difference in coloration was noted between the tissues infected with TEV Δ NiB-E or TEV Δ NiB-I and the control tissues infected with TEV Δ NiB (Figure 3A). Stems of the plants infected with TEV Δ NiB-B also exhibited a distinctive yellow pigmentation that was not observed in the tissues infected with TEV Δ NiB (Figure 3B). This was unexpected

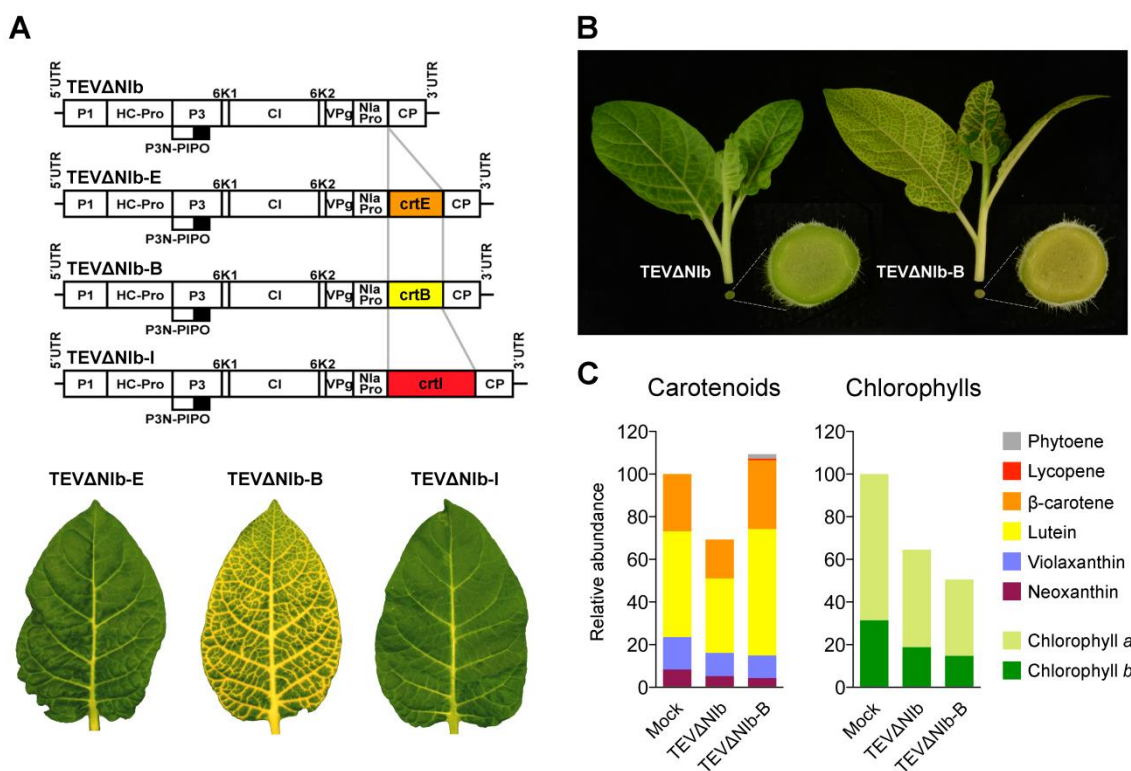


Figure 3. A yellow visual marker to track plant virus infection. **(A)** Schematic presentation of viral clones TEVΔNIB, TEVΔNIB-E, TEVΔNIB-B, and TEVΔNIB-I. Pictures in the lower part of the panel show systemic leaves of representative tobacco plants mock-inoculated or infected with TEVΔNIB, TEVΔNIB-E, TEVΔNIB-B, and TEVΔNIB-I. **(B)** Visual symptoms in systemic leaves and stems of tobacco plants infected with TEVΔNIB or TEVΔNIB-B. **(C)** Carotenoid and chlorophyll contents of tobacco leaves either infected with the indicated virus clones or treated with a mock solution. The average content in symptomatic systemic leaves from $n=4$ independent plants at 10 dpi is represented.

because phytoene, the direct product of *crtB* activity (Figure 1), is colorless. Interestingly, HPLC analysis of the carotenoid profile of the infected leaves showed that TEVΔNIB-B stimulated the accumulation of higher levels of chloroplastic (i.e. endogenous) carotenoids compared to empty vector control TEVΔNIB (Figure 3C). In particular, the orange carotenoid β-carotene and the yellow carotenoid lutein (Figure 1) were 70-75% more abundant in the symptomatic TEVΔNIB-B-infected yellow tissues (Figure 3C). These tissues also presented an approximate 50% concomitant reduction in chlorophylls (Figure 3C). A similar reduction in chlorophylls, but also in carotenoids, was observed when comparing the tissues infected with the empty TEVΔNIB vector to those from the mock-inoculated plants (Figure 3C). This indicates that chloroplast functions are negatively impacted by viral infection. Collectively, these findings suggested that the yellow pigmentation that derived from the virus-mediated expression of *crtB* was caused by a combination of factors, including viral infection, which caused

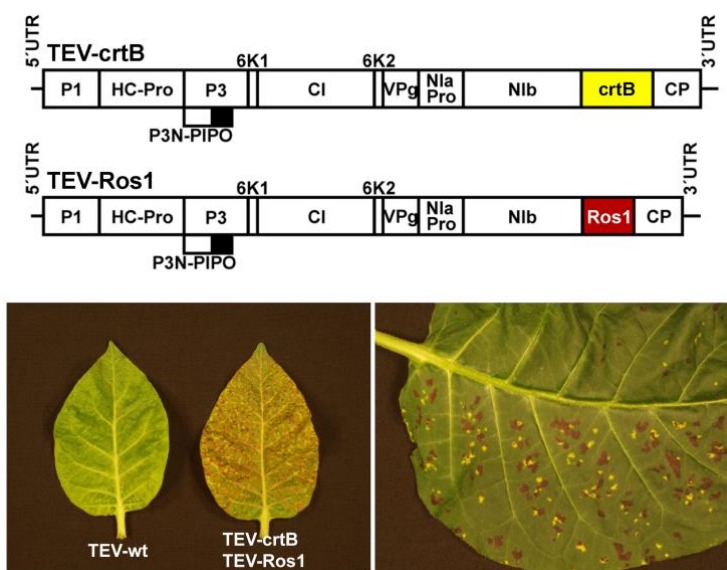


Figure 4. Simultaneous visual tracking of co-infecting plant viruses using crtB and Rosea1 markers. Schematic presentation of recombinant clones TEV-crtB and TEV-Ros1. The crtB and Rosea1 (Ros1) markers are represented with yellow and red boxes, respectively. Other details are as in Figure 2A. The lower left picture shows representative systemic leaves from the tobacco plants infected with wild-type TEV (TEV-wt) or co-infected with TEV-crtB and TEV-Ros1 at 15 dpi. The lower right picture is a magnification of a representative newly grown tobacco leaf co-infected with recombinant TEV viruses that harbor the crtB and Rosea1 markers at 25 dpi.

chlorosis due to a reduction in the levels of photosynthetic pigments, and crtB activity, which triggered increased carotenoid content.

Regardless of the molecular mechanism that caused the yellow pigmentation observed in the plants infected with TEV Δ NIB-B, we reasoned that *P. ananatis* crtB could be used as a reporter marker to visually track virus infection in plants. We previously described how the use of the transcription factor Rosea1 from *A. majus* (an activator of anthocyanin biosynthesis) allowed us to visually monitor infection dynamics in some plant-virus combinations based on the reddish color of the cells that accumulate anthocyanins (Bedoya *et al.*, 2010; Bedoya *et al.*, 2012; Majer *et al.*, 2013). To evaluate whether the simultaneous use of these two color-based visual markers could actually serve to monitor the infection of two different virus populations, we constructed a new TEV-crtB clone by inserting the sequence coding for crtB between the NIB and CP (coat protein) cistrons in the TEV genome (Figure 4 and Supplementary Figure 3), in exactly the same position as the Rosea1 gene in the TEV-Ros1 clone (Figure 4 and Supplementary Figure 3) (Bedoya *et al.*, 2012; Majer *et al.*, 2013). Note that, unlike the above-described TEV Δ NIB-B, the TEV-crtB construct contains the full-

length TEV genome, including Nib. As in TEV-Ros1, the crtB gene is also flanked by artificial sequences to promote its proteolytic processing from the viral polyprotein (Supplementary Figure 3). The third true leaves of wild-type tobacco plants were inoculated with both TEV-crtB and TEV-Ros1, with each recombinant virus in one half of the same leaf. After 4 days, plants started to show symptoms of infection and the symptomatic tissues became pigmented approximately two days later. Depending on which recombinant virus (TEV-crtB or TEV-Ros1) invaded the tissue, both inoculated and systemic symptomatic tissues showed alternative yellow or red coloration (Figure 4). Yellow and red spots nicely marked the particular tissue sections colonized by different viruses in systemic newly grown leaves at 25 dpi (Figure 4).

Then, we wondered whether *P. ananatis* crtB could be used as a more general marker in different plant-virus combinations. To answer this question, we constructed recombinant clones of two viruses that belong to different families: *Tobacco mosaic virus* (TMV; family *Virgaviridae*) and *Potato virus X* (PVX; family *Flexiviridae*), both of which harbored the crtB marker. The resulting clones, TMV-crtB and PVX-crtB (Figure 5A and Supplementary Figure 3), as well as TEV-crtB, were used to inoculate *Nicotiana benthamiana* plants. A few days later, plants developed symptoms of infection and the symptomatic tissues subsequently turned yellow (Figure 5A). The intensity and extension of yellow pigmentation differed according to virus, which indicate that each species has its particular infection dynamics in *N. benthamiana*. We also analyzed whether the crtB marker was useful for tracking virus infection in an important crop like tomato (*Solanum lycopersicum* L.) and in the model plant *Arabidopsis thaliana*. After inoculation with TEV-crtB, the symptomatic tissues of both the tomato (Figure 5B) and *A. thaliana* (Figure 5C) plants exhibited yellow coloration clearly discernible to the naked eye. Finally, we also analyzed the behavior of the new marker in cucurbitaceous plants, a family that includes many important crops. Cucurbits do not have the full anthocyanin pathway (Tadmor *et al.*, 2010), and our own previous observations have indicated that the Rosea1 marker does not function in this family. To test whether the crtB marker could be extended to track virus infection in cucurbits, we constructed two recombinant clones of *Zucchini yellow mosaic virus* (ZYMV; genus *Potyvirus*, family *Potyviridae*) that harbor the sequences, which encode either crtB (ZYMV-crtB) or Rosea1 (ZYMV-Ros1) (Figure 5D and Supplementary Figure 3). Zucchini (*Cucurbita pepo* L.) plantlets were inoculated with both recombinant viruses,

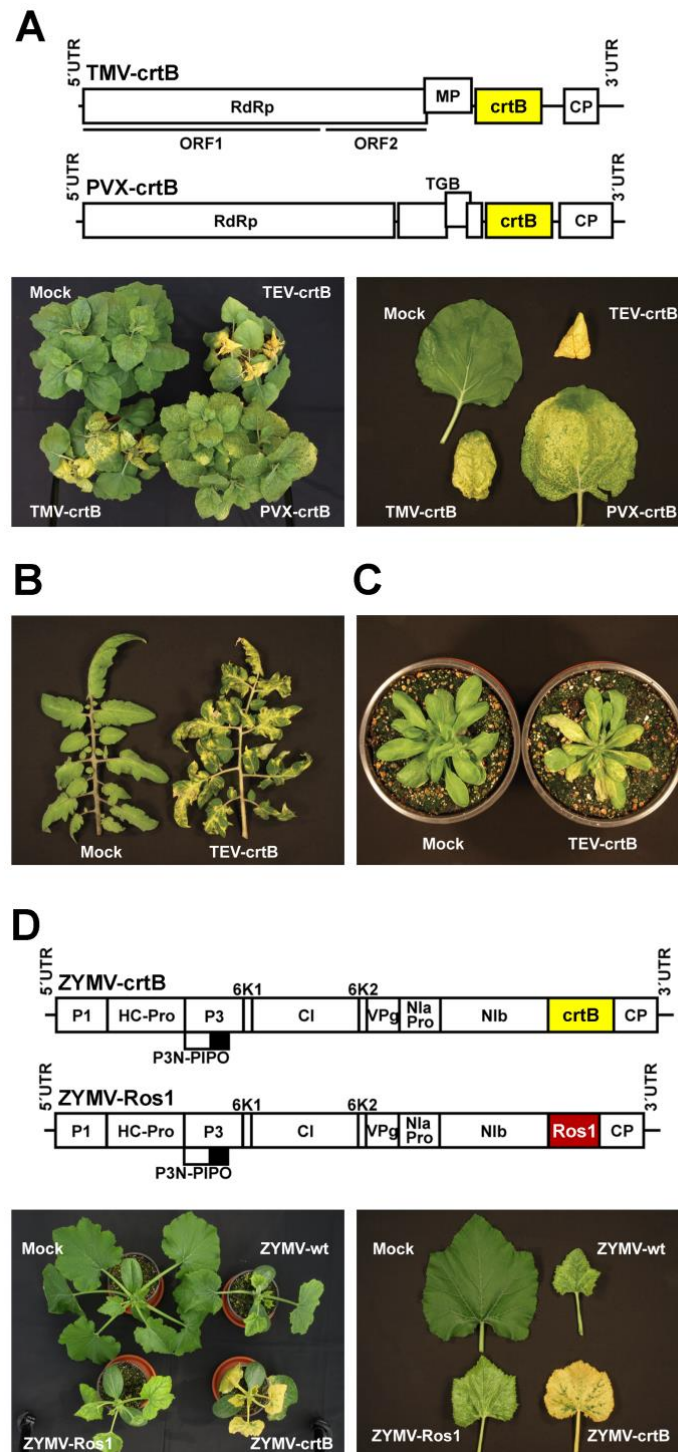


Figure 5. Performance of the *crtB* marker in different virus and plant species. **(A)** Schematic presentation of recombinant clones TMV-*crtB* and PVX-*crtB*. RNA-dependent RNA polymerase (RdRp), movement protein (MP), coat protein (CP) and triple gene block (TGB) proteins are represented by open boxes. 5' and 3' UTRs are represented by lines. *crtB* is denoted by a yellow box. The pictures in the lower part of the panel show *N. benthamiana* plants (left) and representative leaves from these plants (right) mock-inoculated or inoculated with TEV-*crtB*, TMV-*crtB* or PVX-*crtB* at 15 dpi. **(B)** Leaves from tomato plants mock-inoculated or infected with TEV-*crtB* at 15 dpi. **(C)** *A. thaliana* plants mock-inoculated or infected with TEV-*crtB* at 15 dpi. **(D)** Schematic presentation of the recombinant ZYMV clones tagged with *crtB* or Rosea1. Details are as in Figure 2A and 4A. Pictures in the lower part of the panel show representative zucchini plants (left) and leaves from these plants (right) mock-inoculated or infected with wild-type ZYMV (ZYMV-wt), ZYMV-Ros1 or ZYMV-*crtB* at 11 dpi.

and also with an untagged virus control. Some days later, all the inoculated plants developed symptoms of infection, but marker specific coloration was observed only for ZYMV-crtB (Figure 5D).

Discussion

Engineering plant metabolism to produce compounds of nutritional, industrial, or pharmaceutical relevance remains an important goal given the economic benefit of using sunlight to fuel sustainable production methods. However, this is currently a challenging process that is often hampered by the highly complex regulation of plant metabolism. Orthogonal systems, which are mechanistically and spatially separated from those hardwired by evolution and tightly regulated by the host, can potentially operate with improved performance. The feasibility and efficiency of our orthogonal approach was confirmed by the overproduction of lycopene and its accumulation as cytosolic crystals in tobacco leaf cells to levels that represented approximately 10% of total leaf carotenoid content (Figure 2). Most carotenoids, including lycopene, are highly hydrophobic and, hence, precipitate and form crystals when massively accumulated (Li *et al.*, 2001; Maass *et al.*, 2009; Cao *et al.*, 2012; Yuan *et al.*, 2015). In ripe tomato fruits and watermelons, for instance, lycopene naturally accumulates in the form of crystals (Yuan *et al.*, 2015). It was not previously known, or expected, that lycopene biosynthesis could be efficient enough in the cytosol of plant cells for crystals to eventually form. The transient expression and localized distribution of the used viral vector did not compromise plant survival, while allowing high carotenoid production in the cytosol. Nonetheless in our experiments, the red tissues that accumulated lycopene soon became necrotized, possibly because lycopene is not normally produced in the plant cell cytosol, and overaccumulation of such a lipophilic compound in crystals, or its association with cell membranes, may cause undesired effects on cell functions.

The fact that the TEV-mediated cytosolic expression of crtB turned infected tissue into a bright yellow color (Figure 3) was an unexpected observation. Virus-infected plants show strong metabolic alterations and typically display symptoms such as chlorosis and necrosis. In particular, chloroplasts and photosynthesis have been widely recognized as common targets by many plant viruses (Li *et al.*, 2015), and eventually result in low levels of photosynthetic pigments (chlorophylls and carotenoids), as we observed in the tissues infected with the empty vector (TEV Δ N1b)

(Figure 3). The increased production of chloroplast carotenoids observed in tissues infected with crtB-producing viruses, however, is more complex to explain and will require further experimental insights in order to unveil the exact underlying mechanism.

The tissue pigmentation caused by virus-mediated crtB expression aroused a second, and initially unanticipated, goal of the research herein; that is, its use as a visual marker to track infection dynamics of plant viruses. Tracking plant viruses with reporter genes has contributed significantly to our understanding of the infection process, and has helped to decipher the role of host factors and to screen resistance genes. Most current reporter marker systems used in plant virology rely on fluorescent proteins (Tilsner and Oparka, 2010), which require specialized instrumentation for detection purposes. A visual marker, unlike fluorescent proteins, facilitates high-throughput screenings and allows continuous monitoring of the infectious process (Bedoya *et al.*, 2012; Majer *et al.*, 2013). Previous reports showed that TMV-mediated expression of tomato PSY or pepper capsanthin-capsorubin synthase induced yellow pigmentation in infected tissues, although these plant enzymes were expressed including their corresponding transit peptides to target the chloroplasts. Authors recognized the utility of these visual markers in the analysis of viral insert stability, replication and movement (Kumagai *et al.*, 1995; Kumagai *et al.*, 1998). We previously demonstrated the effectiveness of plant metabolism-based reporter systems. Activating the biosynthesis of anthocyanins with *A. majus* transcription factor Rosea1 has well revealed the infected tissues with a reddish coloration in many plant-virus combinations (Bedoya *et al.*, 2012; Majer *et al.*, 2013). Pigment accumulation is cell-autonomous and occurs only in those tissues where the virus replicates. The Rosea1 visual marker is also quantitative and the amount of anthocyanins correlates with viral load. The addition of a second reporter gene that functions upon the utilization of a different metabolic pathway and displays an alternative output color clearly extends the experimental possibilities of the visual system, and allows the infection dynamics of two different viruses or of two different strains of the same virus in co-infected plants to be tracked (Figure 4). Our results indicated crtB to be a very versatile and robust marker that is capable of inducing yellow coloration when used with several viruses of different genus and families, including TEV, TMV, PVX, and ZYMV (Figure 5). Most interestingly, it works in a wide variety of host species, including model plants *A. thaliana* and tomato (Figure 5). Our experiments with cucurbitaceous plants showed that crtB may be a more general

marker than Rosea1 because, unlike the ubiquitous presence of the carotenoid pathway in plants, some taxonomic groups totally or partially lack the anthocyanin pathway (Tadmor *et al.*, 2010).

In summary, we herein engineered a heterologous complete pathway in an easy-to-manipulate viral vector and rewired the plant cell metabolism for efficient extraplastidic lycopene production, a valuable carotenoid present in minute quantities in leaves. Overall, the modification of plant metabolism by using engineered viruses offers a means of achieving compounds of nutritional or economic interest in a cost-effective and scalable manner. Further advances should combine the use of virus-derived vectors with stable genetic transformation, and even with the remodeling of the subcellular architecture in increasingly sophisticated engineering approaches. While our study joins others in demonstrating the enormous potential of viral vectors for engineering metabolic pathways in plants (Kumagai *et al.*, 1995; Kumagai *et al.*, 1998; Bedoya *et al.*, 2010; Sainsbury *et al.*, 2012; Kanagarajan *et al.*, 2012; Mozes-Koch *et al.*, 2012; Zhang *et al.*, 2013; Nielsen *et al.*, 2013; Geisler *et al.*, 2013; Sainsbury and Lomonosoff, 2014), we take a step forward by demonstrating that carotenoids can be efficiently produced and can accumulate outside plastids in plant cells, which represents a new way for future biotechnological approaches. We also show that *P. ananatis* crtB is an excellent visual marker for tracking infection dynamics of different viruses on several plant species, and that it can be combined with the Rosea1 marker in experiments with two different output colors.

Methods

Plasmid clones

Plasmid pGTEVa (Bedoya *et al.*, 2012) contains the cDNA of an infectious wild-type TEV (TEV-wt), with GenBank accession number DQ986288, but including two silent and neutral mutations, G273A and A1119G, flanked by Cauliflower mosaic virus (CaMV) 35S promoter and terminator in a binary vector that derives from pCLEAN-G181 (Thole *et al.*, 2007). This plasmid was used as a parental to construct plasmids with TEV clones TEV Δ N1b, TEV Δ N1b-BIE, TEV Δ N1b-E, TEV Δ N1b-B, TEV Δ N1b-I, by standard molecular biology techniques, including PCR amplification with high-fidelity Phusion DNA polymerase (Thermo Scientific); DNA assembly by

digestion with type-IIS restriction enzyme BsaI-HF (New England Biolabs), followed by ligation with T4 DNA ligase (Thermo Scientific) (Engler *et al.*, 2009); and Gibson DNA assembly (Gibson *et al.*, 2009) with the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). *P. ananatis* crtE, crtB and crtI were amplified from plasmid pACCRT-EIB (Cunningham *et al.*, 1993). The sequences of TEV Δ NIb and all the recombinant TEV Δ NIb-derived clones are seen in Supplementary Figure 1. A plasmid with recombinant clone TEV-crtB (Supplementary Figure 3) was also constructed from parental plasmid pGTEVa. Viral clone TEV-Ros1, which includes A. majus Roseal between the NIb and CP cistron (Supplementary Figure 3), was contained in the previously described plasmid pGTEV-Ros1(NIb/CP) (Bedoya *et al.*, 2012). Viral clone TMV-crtB (Supplementary Figure 3) derives from the TMV 30B expression vector, which includes the 3' end of *Tobacco mild green mosaic virus* (Shivprasad *et al.*, 1999). It was constructed in a binary plasmid that derived from pCLEAN-G181, including the CaMV 35S promoter and terminator and a ribozyme to produce the 3' end of the viral RNA that derived from that of the minus strand of the human Hepatitis delta virus (Schurer *et al.*, 2002). Viral clone PVX-crtB (Supplementary Figure 3) was constructed by starting from pgR107 (Lu *et al.*, 2003), but recombinant crtB was expressed from a heterologous promoter that derived from Bamboo mosaic virus and contained the deletion of the 29 initial codons of PVX CP (Dickmeis *et al.*, 2014). Plasmid pGZYMV, which contained a ZYMV (Supplementary Figure 3) infectious clone (GenBank accession number pending), was also a derivative of pCLEAN-G181, including the CaMV 35S promoter and terminator. This plasmid was used to construct recombinant clones ZYMV-crtB and ZYMV-Ros1 (Supplementary Figure 3). For the transient expression in *N. benthamiana* leaves, full-length crtE, crtB and crtI sequences were amplified from pACCRT-EIB and cloned into plasmid pDONR207 by the Gateway technology (Invitrogen). Sequences were then subcloned into plasmid pGWB405 to obtain fusions with a carboxy-terminal GFP tag (Nakagawa *et al.*, 2007). The sequences of all the plasmids were confirmed by standard DNA sequencing techniques.

Plant inoculation

The plasmids that contained the different viral clones were electroporated into *Agrobacterium tumefaciens* C58C1, which carried helper plasmid pCLEAN-S4837. The cultures of transformed *A. tumefaciens* were used to agroinoculate *N. benthamiana*

plants (Nakagawa *et al.*, 2007) that were cultivated in a growth chamber at 25°C with a 12-h photoperiod. As soon as symptoms were detected in systemic tissue, aliquots of symptomatic tissue were collected and stored frozen at -80°C. The crude extracts from these aliquots were used to mechanically inoculate 4-week-old *N. tabacum* cv. Xanthi nc, 5-week-old *N. benthamiana*, 4-week-old *A. thaliana* ecotype Landsberg erecta, 3-week-old tomato cv. Marglobe and 12-day-old zucchini cv. MU-CU16 plants, as previously described (Bedoya and Daròs, 2010). The transient expression in the *N. benthamiana* leaves was performed as described elsewhere (Sparkes *et al.*, 2006).

Analysis of metabolites in infected tissues

Carotenoids and chlorophylls were extracted from 4 mg of lyophilized tobacco leaf tissue using 1 ml cold extraction solvent, and were analyzed by HPLC in an Agilent 1200 series HPLC system (Agilent Technologies), as previously described (Llorente *et al.*, 2016).

RNA analysis

Leaf samples of the symptomatic tissue infected with TEVΔNIb or TEVΔNIb-BIE were harvested and total RNA was extracted with the Maxwell 16 LEV simplyRNA Tissue Kit (Promega) according to the manufacturer's instructions. Purified RNAs were quantified by spectroscopy in a NanoDrop apparatus (Thermo Scientific) and RNA integrity was evaluated by agarose gel electrophoresis. The First Strand cDNA Synthesis Kit (Roche) was used to generate cDNA according to the manufacturer's instructions, with random primers and 1 µg of total RNA. The PCR reactions for crtE, crtB and crtI detection were performed with the Go Taq Green Master Mix (Promega) and primers Pa.crtE-F (5'-AACTGCTGGACGATTTGACC-3'); Pa.crtE-R (5'-CTCGGGCCTAACAGATTGAC-3'); Pa.crtB-F (5'-CTACGGCGA-AGCAGGTTTAC-3'); Pa.crtB-R (5'-GCAGCAGCGTTAATTTTTTCG-3'); Pa.crtI-F (5'-ATTACATGCCTGGCTTACGG-3'); and Pa.crtI-R (5'-GCTCCACAGAAAAGG-CTGAG-3'). PCR products were analyzed by electrophoresis on 3% agarose gels and stained with ethidium bromide.

Microscopic analyses

Lycopene and chlorophyll detection in the *N. tabacum* tissues infected with

TEVΔNIb-BIE was performed as previously described (D'Andrea *et al.*, 2014). Light microscopy was done under a Zeiss Axiophot microscope. GFP and chlorophyll fluorescence was detected in the agroinfiltrated *N. benthamiana* leaves under a Leica TCS SP5 confocal laser scanning microscope.

Acknowledgments

We thank Verónica Aragonés and M. Rosa Rodríguez-Goberna for excellent technical assistance. This research was supported by Spanish Ministerio de Economía y Competitividad (MINECO) grants BIO2014-54269-R to J.-A.D., and BIO2014-59092-P and BIO2015-71703-REDT to M.R.-C. Financial support from the Generalitat Valenciana (PROMETEOII/2014/021), the Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (Ibercarot 112RT0445), and the Generalitat de Catalunya (2014SGR-1434) is also acknowledged. E.M. is the recipient of a pre-doctoral fellowship (AP2012-3751) from the Spanish Ministerio de Educación, Cultura y Deporte. B.L. is supported by a postdoctoral fellowship (FPDI-2013-018882) from MINECO.

References

- Bedoya L, Martínez F, Rubio L, Daròs JA. 2010. Simultaneous equimolar expression of multiple proteins in plants from a disarmed potyvirus vector. *J Biotechnol.* 150(2): 268-275.
- Bedoya LC, Daròs JA. 2010. Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* 149(2): 234-240.
- Bedoya LC, Martínez F, Orzáez D, Daròs JA. 2012. Visual tracking of plant virus infection and movement using a reporter MYB transcription factor that activates anthocyanin biosynthesis. *Plant Physiol.* 158(3): 1130-1138.
- Cao H, Zhang J, Xu J, Ye J, Yun Z, Xu Q, Deng X. 2012. Comprehending crystalline beta-carotene accumulation by comparing engineered cell models and the natural carotenoid-rich system of citrus. *J Exp Bot.* 63(12): 4403-4417.

- Carrington JC, Haldeman R, Dolja VV, Restrepo-Hartwig MA. 1993. Internal cleavage and *trans*-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus in vivo. *J Virol.* 67(12): 6995-7000.
- Cazzonelli CI, Pogson BJ. 2010. Source to sink: regulation of carotenoid biosynthesis in plants. *Trends Plant Sci.* 15(5): 266-274.
- Cunningham FX, Jr., Chamovitz D, Misawa N, Gantt E, Hirschberg J. 1993. Cloning and functional expression in *Escherichia coli* of a cyanobacterial gene for lycopene cyclase, the enzyme that catalyzes the biosynthesis of beta-carotene. *FEBS Lett.* 328(1-2): 130-138.
- D'Andrea L, Amenós M, Rodríguez-Concepción M. 2014. Confocal laser scanning microscopy detection of chlorophylls and carotenoids in chloroplasts and chromoplasts of tomato fruit. *Methods Mol Biol.* 1153: 227-232.
- Dickmeis C, Fischer R, Commandeur U. 2014. *Potato virus X*-based expression vectors are stabilized for long-term production of proteins and larger inserts. *Biotechnol J.* 9(11): 1369-1379.
- Engler C, Gruetzner R, Kandzia R, Marillonnet S. 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One.* 4(5): e5553.
- Fraser PD, Bramley PM. 2004. The biosynthesis and nutritional uses of carotenoids. *Prog Lipid Res.* 43(3): 228-265.
- Geisler K, Hughes RK, Sainsbury F, Lomonosoff GP, Rejzek M, Fairhurst S, Olsen CE, Motawia MS, Melton RE, Hemmings AM, Bak S, Osbourn A. 2013. Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc Natl Acad Sci USA.* 110(35): E3360-3367.
- Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods.* 6(5): 343-345.

Giuliano G. 2014. Plant carotenoids: genomics meets multi-gene engineering. *Curr Opin Plant Biol.* 19: 111-117.

Hasunuma T, Miyazawa S, Yoshimura S, Shinzaki Y, Tomizawa K, Shindo K, Choi SK, Misawa N, Miyake C. 2008. Biosynthesis of astaxanthin in tobacco leaves by transplastomic engineering. *Plant J.* 55(5): 857-868.

Kanagarajan S, Muthusamy S, Gliszczynska A, Lundgren A, Brodelius PE. 2012. Functional expression and characterization of sesquiterpene synthases from *Artemisia annua* L. using transient expression system in *Nicotiana benthamiana*. *Plant Cell Rep.* 31(7): 1309-1319.

Kelloniemi J, Mäkinen K, Valkonen JP. 2008. Three heterologous proteins simultaneously expressed from a chimeric potyvirus: infectivity, stability and the correlation of genome and virion lengths. *Virus Res.* 135(2): 282-291.

Kumagai MH, Donson J, della-Cioppa G, Harvey D, Hanley K, Grill LK. 1995. Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proc Natl Acad Sci USA.* 92(5): 1679-1683.

Kumagai MH, Keller Y, Bouvier F, Clary D, Camara B. 1998. Functional integration of non-native carotenoids into chloroplasts by viral-derived expression of capsanthin-capsorubin synthase in *Nicotiana benthamiana*. *Plant J.* 14(3): 305-315.

Li L, Paolillo DJ, Parthasarathy MV, DiMuzio EM, Garvin DF. 2001. A novel gene mutation that confers abnormal patterns of beta-carotene accumulation in cauliflower (*Brassica oleracea* var. *botrytis*). *Plant J.* 26(1): 59-67.

Li XH, Carrington JC. 1995. Complementation of tobacco etch potyvirus mutants by active RNA polymerase expressed in transgenic cells. *Proc Natl Acad Sci USA.* 92(2): 457-461.

Li Y, Cui H, Cui X, Wang A. 2015. The altered photosynthetic machinery during compatible virus infection. *Curr Opin Virol.* 17: 19-24.

Llorente B, D'Andrea L, Ruiz-Sola MA, Botterweg E, Pulido P, Andilla J, Loza-Alvarez P, Rodriguez-Concepcion M. 2016. Tomato fruit carotenoid biosynthesis is

adjusted to actual ripening progression by a light-dependent mechanism. *Plant J.* 85(1): 107-119.

Lu R, Malcuit I, Moffett P, Ruiz MT, Peart J, Wu AJ, Rathjen JP, Bendahmane A, Day L, Baulcombe DC. 2003. High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. *EMBO J.* 22(21): 5690-5699.

Lu Y, Rijzaani H, Karcher D, Ruf S, Bock R. 2013. Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc Natl Acad Sci USA.* 110(8): E623-632.

Maass D, Arango J, Wust F, Beyer P, Welsch R. 2009. Carotenoid crystal formation in *Arabidopsis* and carrot roots caused by increased phytoene synthase protein levels. *PLoS One.* 4(7): e6373.

Majer E, Daròs JA, Zwart MP. 2013. Stability and fitness impact of the visually discernible Roseal marker in the *Tobacco etch virus* genome. *Viruses.* 5(9): 2153-2168.

Mann V, Harker M, Pecker I, Hirschberg J. 2000. Metabolic engineering of astaxanthin production in tobacco flowers. *Nat Biotechnol.* 18(8): 888-892.

Meléndez-Martínez AJ, Mapelli-Brahm P, Benítez-González A, Stinco CM. 2015. A comprehensive review on the colorless carotenoids phytoene and phytofluene. *Arch Biochem Biophys.* 572(188-200).

Misawa N, Nakagawa M, Kobayashi K, Yamano S, Izawa Y, Nakamura K, Harashima K. 1990. Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. *J Bacteriol.* 172(12): 6704-6712.

Mozes-Koch R, Gover O, Tanne E, Peretz Y, Maori E, Chernin L, Sela I. 2012. Expression of an entire bacterial operon in plants. *Plant Physiol.* 158(4): 1883-1892.

Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, Watanabe Y, Nakamura K, Kimura T, Ishiguro S. 2007. Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci Biotechnol Biochem.* 71(8): 2095-2100.

- Nielsen AZ, Ziersen B, Jensen K, Lassen LM, Olsen CE, Moller BL, Jensen PE. 2013. Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth Biol.* 2(6): 308-315.
- Nisar N, Li L, Lu S, Khin NC, Pogson BJ. 2015. Carotenoid metabolism in plants. *Mol Plant.* 8(1): 68-82.
- O'Connor SE. 2015. Engineering of secondary metabolism. *Annu Rev Genet.* 49: 71-94.
- Rodríguez-Concepción M, Boronat A. 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol.* 130(3): 1079-1089.
- Ruiz-Sola MA, Rodríguez-Concepción M. 2012. Carotenoid biosynthesis in Arabidopsis: a colorful pathway. *Arabidopsis Book.* 10: e0158.
- Sainsbury F, Saxena P, Geisler K, Osbourn A, Lomonossoff GP. 2012. Using a virus-derived system to manipulate plant natural product biosynthetic pathways. *Methods Enzymol.* 517: 185-202.
- Sainsbury F, Lomonossoff GP. 2014. Transient expressions of synthetic biology in plants. *Curr Opin Plant Biol.* 19: 1-7.
- Schürer H, Lang K, Schuster J, Mörl M. 2002. A universal method to produce *in vitro* transcripts with homogeneous 3' ends. *Nucleic Acids Res.* 30(12): e56.
- Shivprasad S, Pogue GP, Lewandowski DJ, Hidalgo J, Donson J, Grill LK, Dawson WO. 1999. Heterologous sequences greatly affect foreign gene expression in tobacco mosaic virus-based vectors. *Virology.* 255(2): 312-323.
- Sparkes IA, Runions J, Kearns A, Hawes C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc.* 1(4): 2019-2025.
- Tadmor Y, Burger J, Yaakov I, Feder A, Libhaber SE, Portnoy V, Meir A, Tzuri G, Saar U, Rogachev I, Aharoni A, Abeliovich H, Schaffer AA, Lewinsohn E, Katzir N. 2010. Genetics of flavonoid, carotenoid, and chlorophyll pigments in melon fruit rinds. *J Agric Food Chem.* 58(19): 10722-10728.

Thole V, Worland B, Snape JW, Vain P. 2007. The pCLEAN dual binary vector system for *Agrobacterium*-mediated plant transformation. *Plant Physiol.* 145(4): 1211-1219.

Tilsner J, Oparka KJ. 2010. Tracking the green invaders: advances in imaging virus infection in plants. *Biochem J.* 430(1): 21-37.

Wurbs D, Ruf S, Bock R. 2007. Contained metabolic engineering in tomatoes by expression of carotenoid biosynthesis genes from the plastid genome. *Plant J.* 49(2): 276-288.

Yuan H, Zhang J, Nageswaran D, Li L. 2015. Carotenoid metabolism and regulation in horticultural crops. *Hortic Res.* 2: 15036.

Zhang H, Wang L, Hunter D, Voogd C, Joyce N, Davies K. 2013. A *Narcissus mosaic* viral vector system for protein expression and flavonoid production. *Plant Methods.* 9(1): 28.

TTTGGTCCCTTCATCATTACAAAACAGCACTGTTTGTAGAAAGATAATGGAACACTGTTGGTCCAATCACTACATGGTGTATTCAAGGTCAAGAACACCAC
 GACTTTCGAACAACACCTCATTGATGGGAGGACATGATAATTTCCGCATGCTAAGGATTTCCACCATTTCCTCAAAAAGTGAAATTTTAGAGAGCCAC
 AAAGGGAAGAGCGCATATGTCTTGTGACAACCAACTTCCAAACTAGAGCATGTCTAGCATGGTGTGACAGACTAGTTGCACATTCCTTTCATCTGATGGC
 ATATTTCTGGAGAGTATGGATTTCAAACCAAGGATGGGAGTGTGGCAGTCCATATAGTATCAACTAGAGATGGGTTCATTTGGTATACAGCTCAGCATCGAA
 TTTCCACCAACACAAAACAAATTTTTCACAAGCGTGCAGAAACACTTCATGGAAATTTGTTGACAATAACAGGAGCGCAGCAGTGGGTAGTGGTGGCGATTA
 ATGCTGACTCAGTATTGTTGGGGGGCCATAAAGTTTTTCATGAGCAAACCTGAGAGCCCTTTTCAGCCAGTTAAGGAAGCGACTCAACTCATGATGTAATG
 GTGACTCGCAAGTGGCAGCTGTGGGTGTGGTGTGACGCTGGTAAAGAAAGATCAAAGGATGATAAAGTCGCTGAGCAGGCTTCAAAGGATAGGGA
 TGTAAATGCTGGAACCTCAGGAACATTTCTCAGTTCACGAAATAAATGCTATGGCCACAAAACCTCAATATCCAAGGATGAGGGGAGAGGTGGTTGTAAC
 TGAATCACCCTTTAGGATACAAGCCACAGCAATTTGATTTGTCAAATGCTCGAGCCACACATGAGCAGTTTCCCGGTGGCATCAGGCACTGATGACAGCC
 TATGGAGTGAATGAAGAGCAAAATGAAATATTTGCTAAATGGATTTATGGTGTGGTGCATAGAAAATGGGACTTCCCAAAATTTGAACGGAACTTGGGTAT
 GATGGATGGTGGAGCAAGTTTCATACCCGCTGAAACCAATGGTTGAAAACGCGCAGCCAACTGAGGCAAAATTTAGACACACTTCAGTGACCTGGCTG
 AAGCGTATATTGAGATGAGGAATAGGAGCGACCATACATGCTAGGTATGGTCTACAGAGAAACATTACAGACATGAGTTTGTACGCTATCGCTTCGCAC
 TTCTATGAGCTAACTTCAAACACCTGTTAGAGCGAGGAGGCGCATATGCAAATGAAAGCTGCTGCAGTACGAAACAGTGGAACTAGTTATTTGGTCT
 TGATGCGCAACTGGGATCTCGAGGAGAACACTGAAACGGCACACAGGCGCATGTGAAACCGTAAACATGCACACACTATTAGGGTCCGGCACTGATAGT
 TTTCCGGTGTCTTGGCTTCCGCTTTTAAGCTTATTGTAATATATATGAAATAGCTATTCACAGTGGGACTTGGTCTTGTGTGAAATGATCTTATATGTT
 TTAATATGCTTATTAGTCTCATTACTTAGGCGAACGACAAAAGTGAGGTACCCTCGGTCTAATTTCTCTATGTAGTGGGAAAAAATTTTTTTTTTT
 AAAAAAAAAAAAAAAAAAAAAA

>TEVΔIb-BIE (insert between positions 6981-6982 of TEVΔIb)

GGGAGAAGATGAATAATCCGTCGTTACTCAATCATGCGGTGCAAAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGATGCAAA
 AACCCGGCGCAGCGTACTGATGCTCTACGCCGTTGGTCCGCCATTTGTGACGATGTTATTTGACGATCAGACGCTGGGCTTTCAGGCCCGGACGCTGCCTTAC
 AACCGCCGCAACACGCTCTGATGCAACTTGAGATGAAAACCGCCAGGCGCTATGCGAGATCGCAGATGCAGAACCGCGCTTTGCGGGTTTTTCAGGAAGTG
 GCTATGGCTCATGATACTGCCCCGCTTACGCGTTTTGATCATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGATGATACGCT
 CGCTATTGCTATCAGTTGCGAGCGTTGTCCGGTTGATGATGGCCAAATCATGGGCGTGGGGATAACGCCACGCTGGACCGCGCTGTACCTTGGCC
 TGGCATTTCACTGACCAATATTTGCTCGCGATATTGTGGACGATGGCATGCGGGCCGCTGTTATCTGCCGGCAAGCTGGCTGGAGCATGAAGTCTGAAAC
 AAAGAAATTTAGCGGCTGAAAACCGCTCAGGCGCTGAGCCGCTGTTGCCCGCTGTTTGGTGGAGGAAGCAGAACCTTACTATTTGTCTGCCACCGCGG
 CTTGGCAGGTTGCCCCGTGCTTCCGCTGGCAATCGTTACGGCAAGCAGGTTTACCGGAAAATAGGTTGTCAAAAGTTGAACAGGCCGCTCAGCAAGCCT
 GGGATCAGCGGCAGTCAACGACCAGCCGAAAATTAACGCTGCTGCTGGCCGCTTGGTTCAGGCCCTTACTTCCCGGATGCGGGCTCATCTCCCGCG
 CTTGCCCATCTCTGGCAGCGCCGCTACGACTGAAAACCTTGTACTGCTAAGCGGCACTATGAAACCAACTACGGTAATTTGGTGCAGGCTTCGGTGGCT
 GGCACGTGGCAATTCGTCACAAGCTGCGGGGATCCCGCTTACTGCTGAAACAACGTTGATAAAACCGCGGCTGGGCTTATGCTACGAGATCAGGGT
 TTACCTTTGATGCAAGCGCCGAGGTTATCACCAGTCCAGTGGCATTGAAGAACTGTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACCTGCTG
 CCGGTTACCGCGTTTACCCTGTGTTGGGATCGAGGAAAGTCTTTAATACGATAACGATCAAACCGGCTCGAAGCGCAGATTACAGCAGTTAATACC
 CGCGATGTCGAAGGTTATCGTCACTTTCTGGACTATTCACGCGCGGTGTTTAAAGAAAGGCTATCTAAAGCTCGGTACTGTCCTTTTTTATCTGCTCAGAG
 ACATGCTTCGCGCCGCACTCAACTGGCGAACTGCAGGCATGGAGAAGCGTTTACAGTAAAGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGCG
 TTTTCTTCCACTCGCTGTTGGTGGCGGCAATCCCTTCGCCACTCATCCATTTATACGTTGATACCGCGCTGGAGCGTGAAGTGGGCGCTGGTGT
 CGGTGGCGGCACCGCGCATTTAGTTTCCGGGATGATAAAGCTGTTTCCAGGATTCACGCGAGCTGATGACGAAATTTTAAATCATGATGGCCATGGAACAC
 CAGGAAAACAAGATTGAAGCGTGCATTTAGAGGACGCTCGCAGTTTCTGACGCAAGCGCTCGCGTCAAATGCAGATGTTGGTTCATACCTATCGCGACTG
 TTAAGCCAGCACCTGCCCCGTTAAGCAGTCAACAACCTGCAAGCTAAGCGCATGATGAACTCTGTTTGTGCTCATTTTGTGTTGAATCACCATTC
 TGATCAGCTCGCCATCACAGGTTTGTTCGCGCCGCTTACCAGCGAGCTGATGACGAAATTTTAAATCATGATGGCCATGCGCAGGACTTCTCACTTT
 ATCTGCAGCGCCCTGTGTCAGGATTCGTCACGCGCTGAAGTTGCGGCAGTTACTATGTTGGCGCGGTCGCCATTTAGCACCGCGAACCTC
 GACTGGACGTTGAGGGGCAAACTACCGCACCGTATTTTGGGTACTTTGACGAGCATACATGCTGGCTACGGAGTCAAGCTGGTCAAGCCAGGAT
 GTTTACGCGGTTTGGATTTTTCGCGACCACTTAATGCCTATCATCGCTCAGCTTTCAGGATCTGGTGGAGCGGTTTCTTACCAGAGCGGATTTCCGCGCATA
 ACCCGGATAAAACCATTAATACTCTACCTGCTGGCGCAGGACCGATCCCGCGCAGGCACTTCTGGCGTCTCGGCTCGGCAAAAGCGACAGCAGGT
 TTGATGCTGGAGGATCTGATACTACAGAGAACCCTACTTTCAATCAGGTACAATGACGGTCTGCGCAAAAAACAGTTTCACTCTCACTCGCGATGCTG
 GGAGCAGTTACTCGCTGATATTGATCGACCGCTTGATCAGTTATTTGCCCTGGAGGAGAAACGGGATGTTGGTGGTCCCGGATCGTGAAGTTGCGCTG
 CACCGGAAAACGATTTTCGCCCATGTTGCTGTTGCTGACCGCCGCGATCTGGGTTGGCTGTGAGCATTGACGGATTTACTGGATTTGGCTGTGGCGG
 GAAATGGTCCACCGCGCTTCGCTGATCCTTGACGATATGCCCTGCATGGAGCATGCGAAGCTGCGGGCGGACCGCTTACCATTCACTTCTCATTCCGGAGA
 GCAATGGCCAACTGCGCGCGTTCGCTGCTGAGTAAAGCCTTTGGCCTTAAATGGCCATGCGCAGATGCGCTCACGCGCTGGCAAAAATTCGCGCGGCTT
 CTGAATGTCAAACGCCATCGGCATGCAAGGATTTGGTTACGGGTCAGTTCAAGGATCTGTTGAAAGGGATAAGCCGCGCAGCGTGAAGCTATTTGATG
 ACGAATCACTTTAAAACCGACCGCTGTTTGTGCTCCATGCAAGTGGCCGATGTTGTTGCGAATGCTCCAGCGAAGCGCGTGAATGCTGCACTGCTT
 TTTCACTTGACTTTGGTCAGGCAATTTCAACTGCTGGACGATTTGACCGATGACCGCACACCGGTAAGGATAGCAATCAGGACCGCGGTTAAATTCAGCCG
 TGGTCAATCTGTTAGGCCCGAGGCGGTTGAAGAACCTCTGAGACAACATCTTACGTTGGCAGTGAGCATCTCTCTCGCGCTGCCAACCGGCGACCG
 ACTCAACATTTTATTCAGGCTGGTTTGGCAAAAAAATCGCTGCGGTCAGTACGACTGAGAATCTTTATTTTTCAG

>TEVΔIb-E (insert between positions 6981-6982 of TEVΔIb)

GGGAGAAGATGACGGTCTGCGCAAAAAACAGTTTCACTCTACTCGCATGCTGCGGAGCAGTTACTGGCTGATATTGATCAGCGCCTTATCAGTTATTT
 GCCGTGGAGGGAAGCCGGATGTTGGTGGTCCCGCATCGCTGAGGTCGCGTGGCAGCCGCAAAACAGTATTCGCCCGATGTTGCTGTGCTGACCGCC
 CGCATCTGGGTTGCGCTGTAGCAGCTGACGGATTTGATTTGGCTGTGGCTGGAAAATGGTCCACCGCGCTTCGCTGATCTTACGATATCTTCCGCTG
 ATGGACGATGCGAAGCTGCGCGCGGACGCCCTACCATTCACTTCAATACGGAGAGCATGTGGCAATACTGGCGCGGTTGCTTGTGAGTAAAGCCTT
 TGCGCTAATTTGCCAGTCAGATGGCTCAGCGCTGGCAAAAATCGGCGGCTTTCTGAACTTCAAACGCCATCGCATGCAAGGATTTGGTTCAGGCTG
 AGTTCAAAGTATGCTTGAAGGGATAAGCCCGCGCAGCTGAAGCTTATTGATGACGAACTCACTTAAAACCGCAGCAGCTGTTTGTGCTCCATCGAG
 ATGGCTCGATTGTTGCGAATGCTCCAGGAGCGCGTATGCTGCTGATGCTTTTCACTGATCTTGGTCAAGCATTCAACTGCTGGACGATTTGAC
 CGATGGCATGACCGACCGGTAAGGATAGCAATCAGGACCGCGGTAATCGACGCTGGTCAATCTGTTAGGCCCGAGGCGGTTGAAGAAGCTCTGAGAC
 AACATCTTACGTTGGCAGTGAGCATCTCTCTCGGCGCTGCCAACCGGCGACGCCACTCAACATTTTATTCAGGCTGGTTGACAAAAAATCGCTGCC
 GTCAGTACGACTGAGAATCTTTATTTTTCAG

>TEVΔIb-B (insert between positions 6981-6982 of TEVΔIb)

GGGAGAAGATGAATAATCCGTCGTTACTCAATCATGCGGTGCAAAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGATGCAAA
 AACCCGGCGCAGCGTACTGATGCTCTACGCCGTTGGTCCGCCATTTGTGACGATGTTATTTGACGATCAGACGCTGGGCTTTCAGGCCCGGACGCTGCCTTAC
 AACCGCCGCAACACGCTCTGATGCAACTTGAGATGAAAACCGCCAGGCGCTATGCGAGATCGCAGATGCAGAACCGCGCTTTGCGGGTTTTTCAGGAAGTG
 GCTATGGCTCATGATACTGCCCCGCTTACGCGTTTTGATCATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGATGATACGCT
 CGCTATTGCTATCAGTTGCGAGCGTTGTCCGGTTGATGATGGCCAAATCATGGGCGTGGGGATAACGCCACGCTGGACCGCGCTGTGACCTTGGCC
 TGGCATTTCACTGACCAATATTTGCTCGCGATATTGTGGACGATGGCATGCGGGCCGCTGTTATCTGCCGGCAAGCTGGCTGGAGCATGAAGTCTGAAAC
 AAAGAAATTTAGCGGCTGAAAACCGCTCAGGCGCTGAGCCGCTGTTGCCCGCTGTTTGGTGGAGGAAGCAGAACCTTACTATTTGTCTGCCACCGCGG
 CTTGGCAGGTTGCCCCGTGCTTCCGCTGGCAATCGTTACGGCAAGCAGGTTTACCGGAAAATAGGTTGTCAAAAGTTGAACAGGCCGCTCAGCAAGCCT
 GGGATCAGCGGCAGTCAACGACCAGCCGAAAATTAACGCTGCTGCTGGCCGCTTGGTTCAGGCCCTTACTTCCCGGATGCGGGCTCATCTCCCGCG
 CTTGCCCATCTCTGGCAGCGCCGCTACGACTGAGAATCTTTATTTTTCAG

>TEVΔIb-I (insert between positions 6981-6982 of TEVΔIb)

GGGAGAAGATGAAACCACTACGGTAATTTGGTGCAGGCTTCGGTGGCCTGGCAATTCGCTACAAAGCTGCGGGGATCCCGCTTACTGCTTGA
 ACAACCTGATAAACCAGCGCTGGGCTTATGCTACGAGGATCAGGGTTTACCTTTGATGACGCGCCGAGCTTATCACCGATCCCGAGTGCATTTGA
 AACTGTTGACACTGGCAGGAAAACAGTTAAAAGAGTATGCTGCAACTGTCGCGCTTACCGGCTTTACCGCGTGTGTTGGAGTCAAGGAAGGCTTTAAT
 TACGATAACGATCAAACCGCGCTGAAAGCGCAGTTCAGCAGTTAATCCCGCGAGTGTGCAAGGTTATCGTCACTTCTGGACTATTCAGCGGGGTT
 TAAAGAAGGCTAICTAAAAGCTCGGTACTGTCCTTTTTATCGTTACAGACATGCTTTCGCGCCGCACTCAACTGGCGAACTCAGCAAGTGGGAAGCG
 TTTACAGTAAAGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGCGTTTTCTTCCACTCGCTGTTGGTGGCGGCAATCGCTTCCGCACTCATCTC

```

ATTATACGTTGATACACGGCGTGGAGCGTGAGTGGGGCGTCTGGTTCCGCGTGGCGGCACCGGCGCATTAGTTCAGGGGATGATAAAGCTGTTTCAGGA
TCTGGGTGGCGAAGTCGTGTAAACGCCAGAGTCAGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGGTCGCAGGTTCTGA
CGCAAGCCGTCGGCTCAAATGCAGATGTGGTTCATACCTATCGCGACCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAACTGCAGACTAAG
CGCATGAGTAACTCTCTGTTTGTGCTCTATTTGGTTTGAATCACCATCATGATCAGCTCGCGCATCACACGGTTGTTTCGGCCCGGTTACCGCGAGCT
GATTGACGAAATTTTAAATCATGATGGCCTCGCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTACGGATTCTGCACTGGCGCCTGAAGGTTGCC
GCAGTTACTATGTGTTGGCGCCGGTGGCCGATTAGGCACCGCAACCTCGACTGGACGGTTGAGGGGCCAAAACACTACGCGACCGGATTTTTGCGTACCTT
GAGCAGCATTACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTTGATTTTCGCGACCAGCTTAATGCCATCATGGCTCAGC
CTTTTCTGTGGAGCCCGTTCTTACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCTGGTCGGCGCAGGCACGCATC
CCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGCGACAGCAGGTTTGATGCTGGAGGATCTGATAACGACTGAGAATCTTTATTTTCA

```

Figure S2. Subcellular localization of the indicated GFP-tagged proteins in agroinfiltrated *N. benthamiana* leaves visualized by confocal microscopy. A mock-treated control and an empty vector control (with only the GFP protein) are also shown. Chlorophyll autofluorescence (in red); GFP fluorescence (in green) and merged images of both are shown. Bar, 10 μ m.

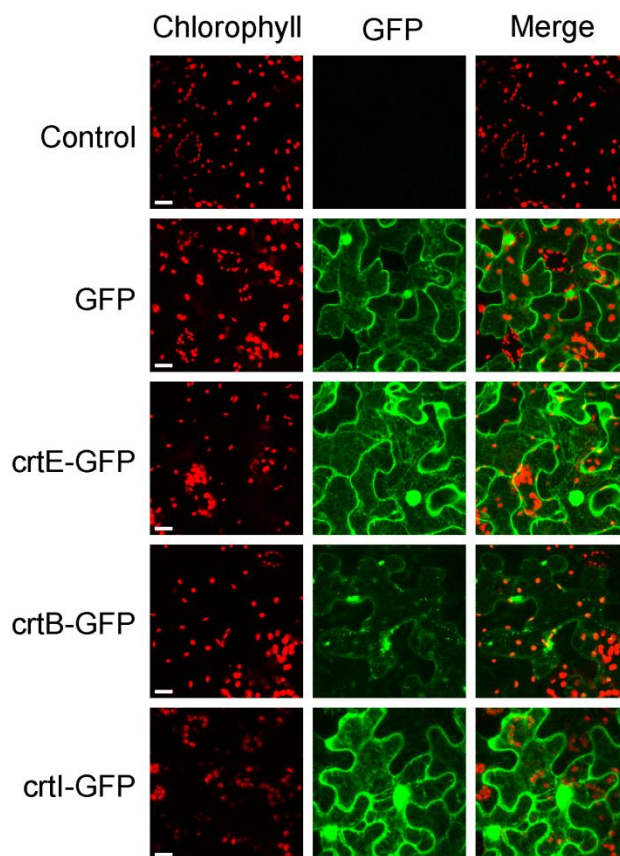


Figure S3. Full sequences of virus clones TEV-wt, TEV-crtB, TEV-Ros1, TMV-crtB, PVX-crtB, ZYMV-wt, ZYMV-crtB and ZYMV-Ros1. Some details of each particular virus clone are indicated below each sequence.

>TEV-wt

AAAATAACAAATCTCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGCAGCAATTTAAATCATTTCTTTAAAGCA
 AAAGCAATTTTCTGAAATTTTACCATTTCAGAACGATAGC**CA**TGGCAGCTCATCTTTGGCACAGTCAACGCTAACATCTCTGAAGGAAGTGTTCGGTGGAG
 CTGATATGGCTTTCGGTTACCAGCGCACATATGGCTGGAGCGAATGGAGCATTTTGAAGAAGGAGAGAA**AA**CCCTCTCGTGAATCATGCACAAACAGTG
 ATCTTCGGAGAAAGACTACATTACCAGGCGAGACTTGGCTTACACACCCTCCATTTAGAGTTCGATGCTGAAATGGAGCGGATGTTATTCTTGGTTCGTCG
 CGCGCTCACCCATGGCAAGAGACGCAAAAGTTTCTGTGAATAACAGAGGAACAGGAGAGAAAGTGGCCAAAACGTACGTGGGCGGTGATTCATTGTTG
 AGAAGATTGTAGTGGCCACACCGAGAGAAAGTTTATACACAGCAGCAGTGGAGAACATTTGCAATGAAGCTTACCCTCAACTTGTGCATAATAGTATG
 CCAAAGCGTAAGAGCAGAAAACCTTTCGCCGCCACTTCACTAAAGTAACTGTATGCCCAAACCTGGAGCATAGTCGGCAAAACCCATATGCAAGTGGGA
 GATCATTAGCAAGAGCGTCCGAGCGAGGGTCAAGAGATTGAGGGCTCGGTGCAATTTGTCGCAAGTGTGGTGCACATGTATGGCGAGAGGAAAGGG
 TGGACTTACGTATTGCAACTGGCAGCAAGAGACACTTCTAGACCTTGCTAAAAGATTAAAGAAATGAGAGAGTGGATCAATCGAAGCTCACTTTTGGTTCA
 AGTGGCCATGTTTTGAGCGAAGGCTCGTACGGACCTGCGCATTTGGTATCGACATTTGGTATGTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG
 TCGTGGGAAGGTAACTGTTTCTGCTCACTCAATGACACATTAT**TA**CGCACAAATCAATCTCTGAGGCATCTTTCATACCATACTCTTAAGAAATCTTTGG
 AGTTGAG**CG**CAGATGGAATCTCCCATGAGTGTACAAGAGGAGTATCAAGTTGAGCGGTGGCGTGGAGTGGCTGCAATCTGCACACAAGCACTTTCACCCTGT
 GGTAAAGATCACATGCAAAACGTTGCATGGTTGAACACCTGACATTTGTTGAGGGTGAAGTTCGGGAGACAGTGTCAACCAACCAAGGTAAGCTCCAGCAATGT
 GAAAGAACAGTATCCAGATTTCCCAATGGCCGAGAACTACTCACAGGTTTTGCAACAGAAATCACTAGTAATAACAAATTTGACAGCCTGCGTGGAGCG
 TCAAAACAATCATTGGTGAACCGCAAAACAGCTCCATTCACACACGCTACTGGCTGTACAGCAAAATCTGTTTAAAGGCAATAAACTAACAGGGGCGCATCTC
 GAAGAGGCAAGCACACATATGCTTGAATAGCAAGGTTCTTGAACATGCAAGGTTCTTGAACAAATATGCGCATTTGGCCAACTTGGTCTTTCAGAAATAAAATCTC
 ATCGAAGGCCCATGTGAATAACGCACATATGTTGATAATCAACTGATCAGATGGGAATTTTATTGGGGACTAAGGGGTGCACGCAAGAGGTTTCT
 TTAAGGATTTTCACTGAGATTGCCCAAAATGAAGGATACGATAAGTATGTTATCAGGAACATATCAGGGGTAGCAGAAAGCTAGCAATTTGGCAATTTG
 ATAATGTTCAACTGACTTCCAGAGCTCAGCAACAAATCAAGGCGAAACTATGAGCGTAAAGAAATTTGGAAATCAGTATTCGGAAATCGGAAATCGGATA
 TTACGTGATCCCATGTTGTTGTTACTCTTGAAGATTGTAAGGCTCAATATTGGATCTAAAGCATCAACAGAGAGACATCTGGTCAATTTGGCAACTCTG
 CGGATTCAAAGTACCTAGACCTCCAGTTCTCAATGAAGAGAAATGTATAGCTAATGAAGTTTATGCTACATGAACATTTCTTTGCTCTACTAGTG
 AATGTCAAGGAAAGAGGATCCAAAGCTTCCCAAGTTTATAAGGGACACAATTTGTTCCAAAGCTTGGAGCGTGGCCAAACCAAGTCAAGTATGCAAGTGC
 ATGCTACTTACTTCCATCTTTACCAGATGTCCTGAGTGTCAAATACCAGAAATTTGGTTGATCATGACACAAAACCAATGCATGTTTGGATTTCGT
 ATGGGTCTAGAAGCAGGATCACCATGTTGAAATGAACACAACATCCAGCTAATGAATTCGTTCAATTCAGGTTTGGAAATCCGAAATGAAACTTTAC
 AATGTTGG**AG**GGATGAACCGAGATATGGTCAACAAGGTCGAATTTGAGATGTTGATCAAGTCCATATACAAACCACTCTCATGAAGCAATTTCTATGGAG
 GGAGCCATACATAATGTTCTCGCAATAGTCTCCCTTCAATTTAATGGCCATGTACAACCTTGGAACTTTTGGAGCGGGTACAAATTTGGATTGGCCAA
 ATACAATGAGGTTAGTCACTCGCTGCCATCTTGTGACGCTTGGCCGAAAAGTTAACTTTGGCAGACTTGTTCGTCAGCAGCGTAATTTGATTATGAG
 TATGGCCAGTAAATTTTGGCAATCTGATTGACGGTGTCAAGGTTAAACATTCGCTATCCCTAGCAATGGAAATTTGATTATTTAGTATTTGCAACCCAA
 GATGGACATGGCTGAGGGAAAGTGGCTATGCTGTGACCTCTGAAAAGGTGCATGAATTTGGAAAAAACTATGTAAGGATTTGAAGGATGCAATGG
 ACGAATTAACCTGGTTGGAAAAATCTCCGCAATCAGGCTTCAAGAAAGCTCTTGAATTTGGCGAAAGCTTTAATAGCTGAAACACCCCTAGATTGCG
 GCGGACATAGACTTGTCTGTGAAATCGCTTTTCAAGTTCCACTTGGAACTCTGAAGGAAACCATCTCAAGAGGTAAGCTTAAATGGTGGTGCAGAAAGGTT
 AAGAGTAGCGAAGAAATGCCATGACAAAAGGGGTTTTTCTCAAAATCTACAGCATGCTTCTGACGCTTACAAGTTTATCACAGCTCGAGTGTCTTTCT
 TGTGTTGACTTCTTATTTCAAATGACTGCTATGATAAGGCGACACCGAGGCGAAGGTTGCTGCACAGTTGCAGAAAGAGAGATGGGACAAATAT
 ATCAATAGAATTTCCAGTATTTAAGCTTGAATTCCTATTTGGTATTCGCTCTACAGCGGAGGAAAGACTCCAATCAGAACACCCCGAGGCTTTCGAGTA
 CTACAAGTTTGTGATGGAAAGGAGACCTCGTTGAACA**AG**CAAAAACACCCGAGATAGCATACTTTGAAAAGATTATAGCTTTCATCACACTGTATTTAA
 TGGCTTTTGACGCTGAGCGGAGTGTGGAGTGTTCGAAGTACTCAATTAAGTTCAAGAGAAATCTGAGCTCAACCGAGAGGAGATGATCTACACGA**AG**AT
 TTGGATGATTACGTTACAACCTTTGATGACAATATGACAATCAACCTCGAGTTGAATATGGATGAACCTCACACAAGCAGGCTTCTCGGATCACTTTTAA
 GCAATGGTGGAAACAACTAACGCGAGGCAACCTGAAAGCCATGATGAAGCTGAGGGGCACTTCAATGGAGTTTACAGAGATATTCGCGCATCGGTTG
 CCAGCAGATATCACACTCACCCGCAAGAGATTTTCTTGTGAGAGTGTGCTTGGATCTGGAAATCCACAGGACTTCCATACCATTTCAATAAGAGAGGG
 AGAGTGTAAATGCTTGGCTTACCAGACCTCACAGATAACGTGCACAAGCAACTGAGAAGTGAACATTTAACTGCTTCCCACTTTGAGGATGAGAGG
 GAAGTCAACTTTGGGTCATCACCGATTACAGTCACTGACTAGTGGATTTCGCTTACACCAATTTGACGAAACATAGCTGAGGTAAGAAACATGACATTTG
 TCATAATTTGATGAATGCTGTTGAATGATGCTTCTGCTATAGCTTTAGAACTTCTGTTTGAACATGAATTTGAAGGAAAGTGTTCGAAAGTTCAGCC
 ACACCACAGGTAGAGAAGTTGAATTCACAACCTCAGTTTCCCGTGAACCTCAAGATAGAAGAGGCTCTAGCTTTCAGGAATTTGTAAGTTTACAAGGGAC
 AGGTGCCAACCGCGATGATGATTAGTTGTGGCGACAACATACTAGTATATGTTGCTAGCTACAATGATGTTGATAGTCTTGGCAAGCTCCTTGTGCGAAAAG
 GATTACAAAGTTGCGAAGATTGATGGAAAGAACAATGAAGAGTGGAGGACTGAAATAAATCACTGAAGGTACTTCAAGTGAAGGAAATTTCAACCCACTGCTG
 AATATTTATGAAATGGTGAACATTTGACATTTGATGATGTTGAGTTTGGGACTAAGTGTACCAGTTTGGATGTGGACAATAGAGCGGTGCAGTA
 CAACAACCTGTTGGTGAATTTGGGAGCGCATCCAAAGACTCGGTAGATTGGGCGACACAAGGAGGAGTAGCATTCCGAATTTGGCCAAACAAATAAAA
 CACTGGTTGAAATTTCCAGAAATTTGCCACTGAAGCTGCTTTCTATGCTTCAATTTGCAATTTGCCAGTGACACACAGAGTGTTCACCCACTGCTG
 GAAATGCCACATTTATCAAGCTAGAATTTGGCACAGTTGAGCTATCATATTTTACACAATTAATTTTGTGCGATTTGATGGTAGTATGCATCCAGT
 CATACTGACAAGCTGAAGCGCTTAAAGCTTACACATTTGTGAGACATTTCTCATAAAGTTGGCGATCCAAATAAAGGCTTATCTCTTGGCTTACGAGTG
 GAGAGTAAAGGACTTGGTTTATAGCAGAGGATGCTGGCATAGAATCCAAATCTGCTGTGCAAAAGAAATTTCCAGACTTCTGATGCAAGAAATTTGGCC
 ATTTGATGCGCCATAAAGGTGACTCGGGTATTGGGAGGCTCACTAGCCTACAGGCAAGGTTGTTTATCTCTGCAACCGGATGTGCACTCAATTCG
 GAGGACTTACGATGATCAATGACTCATAGCAGATGAACAAATGAAGCAGAGTCAATTTGAAGCGCAACTGGGAGAGGATTTCTCTCACAATTTACT
 CAATAACAAGCATTTTGAACCGCTGAAAGCAAAATTTGCTACAAAAGCTGACAAAAGAAATATTTGCAAGTCTTCCAGCAGGCAAAAGATCAATGCTAG
 TTTTGAACCTAGCAAAGGATCAAGATGTCAGGGTATCATCCAGACTTCAATCACTGGAAATATCTATCTCCA**AT**TCAGATAGCAAGTGGTCAAGCA
 TCTGAAGCTTAAAGTCACTGGAATAAAAGCCAAATCACTAGGACATCATATAGCTTTGCTGTGTTAATTTGGTGGTGGATGGATGCTTGAACCTACT
 TCAAGGACAAGTTCAATGAACAGTCTATTTCCA**AG**GGAAGAAGAATCAGAAGCACAAGCTTAAAGTAGAGAGAGGCGGTGGGGCTAGAGGCAATATGAG
 GTTGCAGCGGAGCCAGAGGCGCTAGAACATTTCTTTGGAAGCGCATATAATAACAAGGAAAGCGCAAGGGCACCCAGAGGAAATGGGTGCAAAAGTCTCG
 GAAATTCATAAACATGATGGTGGTTCCTCAACTGATTTTTCATACATTTAGGTTTGTGGATCCATTGACAGGTCACTATTTGATGAGTCCACAACCGCAC
 CTATTTGATTTAGTCAGCATGAGTTGGAAAGGTTAGAACACCGCATGTTAATTTGACGATGAGATAGAGCTCAAAGTCTTAGCACCCACACCAACTCCAT
 GCTTATTTGGTGAATGATGGCAGGAAAGTCTTAAAGTGTGATTTAACACCACACTCGTGTGCTAGTGGAGTGAAGAAATCAACAGCAATATGGGATTT
 TCTGAAAGGGAGAAATGATGGTCAAACCGCATGGCAGTGGCAGTGTGATGATCAATTTGCCACAAAGAGTGGAGACTTGCAGTTTGA**AG**GAGAAA
 GCTTTGTTAAAGGACCCAGTGAATACAACCCGATATCGAGCACCATTGTTCTACTTGAACAAATGAATCTGATGGGCACACAACATCTGTTGATGGTATTTGGA
 TTTGGTCCCTTTCATATACAACAAGCACTTGTTTAGAAGAAATATGGAACACTGTTGGTCCAATCACTACATGGTGTATCAAGTCAAGAACACCC
 GACTTTGCAACAACCTCATTGATGGGAGGACATGATAATTTATTCGCATGCTCAAGGATTTCCACCATTTCTCAAAAGCTGAAATTTAGAGAGCCAC
 AAAGGAAAGCCATGTTGTTGTGACAACCACTTCCAACTAAAGAGATGCTAGCATGTTGAGCAGTGTGAGCAGTTCCTTCACTTCAATGATGGC
 ATATCTGGAAAGCATTTGATTTCAACCAAGGATGGGAGTGTGGCAGTCCATTAGTATCAACTAGAGATGGGTTCAATGTTGGTATACACTCAGCATCGAA
 TTTCCACAACACAACAAATTTTTCACAAGCGTGGCAAAACCTTCAATGAAATTTGTGCAAAATCAGGAGGCGCAGCAGTGGGTTAGTGGTGGCGATTAA
 ATGCTGACTCAGTATTTGGGGGGGGCCATAAAAGTTTTCATGAGCAAACTGAAGAGCTTTTCCAGCAGTTAAAGGAAAGCAGCACTCAACTCATGAGTGAATTG
 GTGACTCGCA**AG**GGAGAAGGAAATGGGTCGTGGAAGCACCTGTCAGGGAACCTGAGGCGAGTGGCTGAGTTCGCCAGTCACTGATGCAAAAGCATGT
 GGTAAAGGAAAGTGTCCCTCTTTGAGCTCTACTTGCAGTTGAATCCAGAAAGGAGGACATATTTAAACCGATGATGGGAGCATATAAGCAGGCTGAG
 TTAATAGAGAAAGTTCCTCAAGGACATTTAAATATGCTAGTGAATTTGAGATTTGGAAATGGGATTTGACTTGTGGGATTTGACTTGTGAAATGAAGTACTC
 ATCACAAGCTCAAGCGGTTAGGATTTCCCACTGTAACATACACTGACCCAGGAAATTTTATGTCATTTGAAATGAAGGAGCACTATGGGAGCACT
 ATACAAGGCAAGAAAGAAAGCTCTCAGCGACTCACACTAGATGAGCAGGAGGCAATGCTCAAAGCAAGTTGCCCTGCGACTGATACGGGAAGCTGG
 GAATTTGAAATGGCTATTTGAAAGCAGAGTTGCGTCCAATTTGAGAAGTTGAAAACAAACAAACCGCAACTTTACAGCAGCAGCAATAGACACTCTTCTT
 GCTGGTAAAGTTTGCGTGGATGATTTCAACAATCAATTTTATGATCTCAACATAAAGGACCACTGGACAGTTGGTATCACTGAATTTTATAGGGGTTGAA
 TGAATTTGATGGAGCTTTACCAAGTGGTGGTGTATTGTGACCACTGATGGTTCGCAATTCGACAGTTCTTGAATTCCTCATTTAGCTGATTGA
 AAGTGGCACTTGCCTTCATGGAGGAATGGGATATTTGGTGAACAACTGCTGGAAATTTGTACTGAGATAGTGTATACCAATCTCCACCGGATTTGGT

```

ACTATCATTAAGAAGCATAAAGGCAACAATAGCGGGCAACCTTCAACAGTGGTGGACAACACACTCATGGTCATTATTGCAATGTTATACACATGTGAGAA
GTGTGGAATCAACAAGGAAGAGATTGTGTATTACGTCAATGGCCGATGACCTATTGATTGCCATTCAACCCAGATAAAGCTGAGAGGTTGAGTGGATTCAAAG
AATCTTTCCGGAGAGTTGGCCCTGAAATATGAATTTGACTGCACCACCAGGACAAGACACAGTTTGGTTCATGTCACACAGGGCTTTGGAGAGGGATGGC
ATGTATATACCAAAGCTAGAAGAAGAAAGGATTGTTCTATTTTGGAAATGGGACAGATCCAAAGAGCCGTACATAGGCTTGAAGCCATCTGTGCATCAAT
GATCGAAGCATGGGGTTATGACAAGCTGGTTGAAGAAATCCGCAATTTCTATGATGGGTTTTGGAACAAGCCGCTTATCACAGCTTGCAGAAAGGAA
AGGCGCCATATCTGGCTGAGACTGCGCTTAAAGTTTTTGTACACATCTCAGCACGGAACAACCTCTGAGATAGAAGATTTTAAAGTGTGTATGATTAC
GATATTTCCAACGACTGAAATCTTTATTTTCAAGTGGCACTGTGGTGTCTGGTGTGACGCTGGTAAAGAAGAATCAAAAGGATGATAAAGTCGCTGA
GCAGGCTTCAAAGGATAGGATGTTAATGCTGGAACTTCAGAAACATTTCTCAGTCCACGAATAAATGCTATGGCCACAAAACCTCAATATCCAAGGATGA
GGGAGAGGTTGGTGTAACTTGAATCACCTTTTAGGATACAAGCCACAGCAATGATTGTCAAATGCTCGAGCCACACATGAGCAGTTTCCCGCTGG
CATCAGGCAGTGTACAGCCTATGGAGTGAATGAAGAGCAAAATGAAAATATTGCTAAATGGATTATGGTGGTGGTATAGAAAATGGGACTTCCCCAAA
TTTGAACGGAACTTGGGTTATGATGGATGGTGGAGGACAAGTTTCATACCCCGTGAACCAATGGTTGAAAACCGCGAGCCAACTGAGGCAAAATATGA
CACACTTCAGTGACCTGGCTGAAGCGTATATTGAGATGAGGAATAGGGAGCGCACCATACATGCCTAGGTATGGTCTACAGAGAAACATTACAGACATGAGT
TTGTACGCTATGCGTTGCACTTCTATGAGCTAACTTCAAAAACCGCTGTAGAGCGAGGGAGGGCCATATGCAAAATGAAAGCTGCTGCAGTACGAAAACAG
TGGAACTAGGTTATTTGGTCTTGATGGCAACGTTGGTACTGCAGAGGAAGACACTGAACGGCACACAGCGCACAGCTGTGAACCCGTAACATGCACACACTAT
TAGGGTCCGCCAAGTATAGTCTTCTGCGTCTTTGCTTTCCGCTTTTAAAGCTTATTGTAATATATATGATAGCTATTCACAGTGGGACTTGGTCTTGTG
TTGAATGGTATCTTATATGTTTTAATATGCTTATTAGTCTCATATTAGGCGAACGACAAAAGTGAGGTACCTCGGTCTAATTTCTCCTATGTAGTGCGA
GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

TEV sequence corresponds to Genbank accession number DQ986288 including two silent mutations (G273A and A1119G, in red). Limits between TEV cistrons are marked on blue background.

>TEV-crtB (insert between positions 8517-8518 of TEV-wt)

```

TCAGGTACAATGAATAATCCGTCGTTACTCAATCATCGGTCGAAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTGATGCAAA
AACCCGGCCGAGCGTACTGATGCTCTACGCTTGGTGGCCCATTTGTGACGATGTTATTGACGATCAGACGCTGGGCTTTACAGCCCGGCGCCTGCCCTTAC
AAACGCCCAACACGCTCTGATGCAACTTGAGATGAAAACGCGCCAGGCTTATGCAGGATCGCAGATGCACGAAACCGGCTTTGCGGCTTTTCAGGAAGTG
GCTATGGCTCATGATATCGCCCCGGCTTACCGCTTTGATCATCTGGAAGGCTTCGCCATGGATGTCGCGAAGCGCAATACAGCCAACTGGATGATACGCT
GGCCTATTGCTATCACGTTTCAGGCTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAAATCATGGGCGTGCGGGATAACGCCACGCTGGACCCGCGCTGTGACCTTGGCC
TGGCATTTCAGTTGACCAATATTGCTCGCGATATTGTTGGACGATGCGCATGCGGGCCGCTTATCTGCCGGCAAGCTGGCTGGAGCATGAAGTCTGAAC
AAAGAGAATATTGCGCCACCTGAAAACCGCTCAGGCGCTGAGCCGATCGCCCGTCTGTTGGTGCAGGAAGCAGAACCCTACTATTTGCTGCCACAGCCGG
CTTGGCAGGGTTGCCCTGGCTTCCGCTGGGCAATCGCTACGGCGAAGCAGGTTTTACCGGAAAATAGGTGTCAAAGTTGAACAGCCGCGCTCAGCAAGCTT
GGGATCAGCGGCGAGTCAACGACCAGCCGAAAATTAACGCTGCTGCTGGCCGCTCTGGTCAGGCGCTTACTTCCCGGATGCGGGGCTCATCTCCCGCC
CTTGGCCTACTCTGGCAGCGCCCGCTCACTACAGAGAACCTTACTTTCA

```

cDNA corresponding to *P. ananatis crtB* is on yellow background. Sequences corresponding to artificial TEV NlaPro cleavage sites are in blue.

>TEV-Ros1 (insert between positions 8517-8518 of TEV-wt)

```

TCAGGTACAATGAAAAGAATTTCGTTGGAGTGAGAAAAGGTAAGTGGACCAAAGAAGAAGCACTCTCTTGGGCAATGTATAGAAGATATGGTGAAGG
GAAAATGGCATCAAGTTCCACACAGAGCAGGTTGAACCGGTGTAGGAAGAGTTGCAGGCTGAGGTGGTTGAATATCTPGAGGCCAAATATCAAAGAGGTC
GGTTTTCCAGAGATGAAGTGGACCTAATTTGTAGGCTTTCATAAGCTGTGGGTAACAAATGGTCGCTGATTGCTGGTGAATTCCTGGAAGGACAGCTAAT
GACGTGAAGAACTTTTGAATACTCATGTGGGAAGAAATTTAGCGAGGATGGAGAAGCATGCCGGAATAATGTTATGAAACAAAAACCATTAAGCTGAC
TAATATCGTAAGACCCGAGCTCGGACCTTCCCGGATTCACCGGATTCGCACTTCTGGCGAGAAAGTTCGGAATAACCCGATGAAATTTCAAATGTCGCGTTAAACA
CTGATGAGATTTCCAGATTTGTAAGCAACCGCAATTTTACAATGATGTTGCGTCGCCACAAGATGAAGTTGAAGACTGCATTCAAGTGGTGGAGTAAGTTG
CTAGAACAACCGGAGGATGGGAATTAGGAACCTATTCGAGGAGGCCAACAAATGGAAATCACTACAGAGAACCTTACTTTCA

```

cDNA corresponding to *A.majus Ros1* is on red background. Sequences corresponding to artificial TEV NlaPro cleavage sites are in blue.

>TMV-crtB

```

GTATTTTACAACAATTACCAACAACAACAACAACAGACAACATTACAATTAATTTACAATTACAATGGCATACACACAGACAGTACCACATCAGCT
TTGCTGGACTGTCCGAGGAAAACACTCCTTGGTCAATGATCAGAAAAGCCGCTCTTTACGACACAGCGGTTGAAGAGTTTAAACGCTCGTGACCCGAG
GCCAAGGTTAACTTTTCAAAGTAAATAAGCGAGGACAGACGCTTATGCTACCCGGGCGTATCCAGAAATCCAAATTAATTTTAAACAGCCAAAATG
CCGTGCATTCGCTTGCAGGTTGGATTGCGATCTTTAGAACTGGAATATCTGATGATGCAAAATCCCTACGGATCATGACTTATGACATAGCGGGAAATTTT
GCATCGCATCTGTTCAAGGACGAGCATATGTACACTGCTGCATGCCAACCTGGACGTTTCGAGACATCATGGCCACGAAGGCCAGAAAGACAGTATTGA
ACTATACCTTTCTAGGCTAGAGAGAGGGGGGAAAACAGTCCCCAACTTCCAAAAGGAAGCATTTGACAGATACGCAAGAAATTCCTGAAGACGCTGTCTGTC
ACAATACTTTCCAGACATGGCAACATCAGCGATGCAGCAATCAGGAGAGTGTATGCCATTGCGCTACACAGCATATATGACATACCAGCGGATGAGTTC
GGGGGGCACTCTTGGAGAAAATGTCCATACGTGCTATGCGGCTTTCCACTTCCGAGAACCTGCTTCTGAAGATTCATCGCTCAATTTGGACGAAAT
CAACCGGTGTTTTTCCGCGGATGGAGACAAGTTGACCTTTCTTTTGCATCAGAGAGTACTCTTAATTAAGTGTATGTTTAAATTTCTTAAGTATG
TGTGCAAACTTACTTCCCGGCTCTAATAGAGAGGTTTACATGAAGGAGTTTTTGTACCCAGAGTTAATACCTGGTTTTGTAAGTTTTCTAGAAATAGAT
ACTTTCTTTTGTACAAGGTTGGCCATAAAGTGTAGATAGTGAGCAGTTTTTACTGCAATGGAAGACCGATGGCATTACAAAAGACTCTTGCAT
GTGCAACAGCGAGAGAATCCTCCTTGAGGATTCATCATCAGTCAATTAAGTGGTTTTCCAAAATGAGGGATATGGTCACTGACCATTTTCGACATTTCTT
TGGAGACTAGTAAGAGGACGCGCAAGGAAAGTCTTAGTGTCCAAGGATTTGCTGTTTACAGTGTCTAACACATTCGAACATACCAGCGGAAAGCTTTCACA
TACGCAATGTTTTGCTTTCGTCGAATCGATTTCGATCGAGGTAATCAATTAACGTTGACAGCGAGGTCGAAGTGGATGGGACAAATCTTTGTTACA
ATCCTTGTGCATGACGTTTTTACGCTGCACTAAGCTTGGCTTCTAAAGATGACTTACTGATTAGCAAGTTTAACTCTCGTTCGTTGAAAACCGTGTGCCAGC
ATGTGTGGGATGAGATTTCCCTGGGCTTTGGGAACGCAATTTCCCTCCGTGAAAAGAGAGGCTTTCGAACAGGAACTTATCAGAGTGGCAGCGCACGATTA
GAGATCAGGTGCTGATCTATATGTTGACCTTCCAGCAGAGATTAGTACTGACTAGTACAAGGCTTCTGTGACATGCTTGGCTTGAATTAGAAGAAGAT
GGAAAGAACGGAAGTGTACAATGCACTTTCAGAATATTCGGTGTAAAGGAGTCTGACAAAATTCGATGTTGATGTTTTTCCAGATGTGCAATCTT
TGGAAATGACCAATGACGGCAGCGAAGGTTATAGTCCGGTCAATGAGCAATGAGAGCGGCTGACTCTCACATTTGAAACGACTTACGAGGCAATGTT
GCCTAGCTTTACAGGATCAAGAGAAGGCTTCAGAAGGTGCATTTGGTACTTACGTAAGAGAAGTTGAAGAACCGCTCAAAAGTTTGGAGTCTGGATGCTAGGAAGT
AGAGTTACAATTAGCTGGTCTTGTGGAGATCATCCGGAATCGTCTTATCTAAGAACGAGGAGATAGAGTCTTTAGAGCAGTTTCATATGGCCAGCGCAG
ATTCGTTAATTCGTAAGCAGATGAGCTCGATTTGTTACACGGGTCGATTAAGTTTCAGCAAAATGAAAACCTTTATCGATAGCCTGGTAGCATCACTAT
GCTGCGGTGCGAATCTCGTCAAGATCCTCAAAGATACAGCTGCTATTGACTTGAACCCGCTCAAAAGTTTGGAGTCTGGATGCTAGGAAGTGTG
GTTAATCAAACCAAGGATGATGATGATGGGGTGTGTTGTTGAAACCCAGCGAGGAGTATCATGTTGGGCTTTTGGAAATGATGAGCAGGGTGTGG
TGACATGCGATGATGGAGAAAGATGACTGTTAGCTCTGAGTCTGTTGTTTATTCGACATGGCGAACTCAGAACTCTGCAGCAGCTGCTTCAAAACGGA
GAACCGCATGTCAGTAGCGCAAGGTTGTTCTGTGGACGGAGTTCCGGCTGTGGAAAACCAAAGAAATTTCTTCCAGGTTAATTTGATGAAGATCT
AATTTTAGTACCTGGGAAGCAAGCCGCGAAATGATCAGAAGACGTGCGAATTCCTCAGGATTTATTGTTGCCAGGAGGACAACGTTAAAACCGTTGATT

```

CTTTCATGATGAATTTTGGGAAAAGCACACCGTGTCAAGGTTTATTCATTGATGAAGGGTTGATGTTGCATACTGGTTGTGTTAATTTCTTGTG
 GCATGTCAATTTGGCAAAATGGCATATGTTTACGGAGACACACAGCAGATTCCATACATCAATAGAGTTTCAGGATTTCCCGTACCCCGCCATTTTGCCAA
 ATTGGAAGTTGACGAGGTGGAGACACGCAGAACTACTCTCCGTTGTCAGCCGATGTCACACATTATCTGAACAGGAGATATGAGGGCTTTGTCATGAGCA
 CTTCTCCGGTTAAAAGTCTGTTTCCGAGGAGATGGTCGGCGGAGCCGGTGTATCAATCCGATCTCAAAACCCCTGCGATGGCAAGATCTTGACTTTTACC
 CAATCCGATAAAGAGTCTGCTTTTCAAGAGGTTTTCAGATGTTCCACATGTGCATGGAAGTGAAGGCAAGGCGAGACATCTCTGATGTTTCACTTTAGT
 AACCCCTACACCGGCTCCATCATTTGCAGGAGACAGCCACATGTTTTGGTCGATTTGCAAGGCACACCTGTTCCGTCGAAGTACTACACTGTTGTTATG
 ATCCTTTAGTTAGTATCATTAGAGATCTAGAGAACTTAGCTCGTACTGTTAGATATGTATAAGGTCGATGCAGGAACACAATAGCAATACAGATTTGAC
 TCGGTGTTCAAAGGTTCCAAATCTTTTTGTTGTCAGCGCAAGACTGGTGAATTTCTGATATGCAGTTTACTATGATAAGTGTCTCCAGGCAACAGCAC
 CATGATGAATAATTTGATGCTGTTACCATGAGGTTGACTGACATTTCAATGATGTCAAAGATTGCATATGGATATGCTAAGTCTGTTGCTCGCCCTA
 AGGATCAAAATCAAAACCAATAACTTATGTTACGAAACCGCCGAGAAATGCCACGCCAGACTGGACTATTGGAAAAATTTAGTGGCGATGATTAAGAAAC
 TTTAACGCAACCCGAGTTGCTCGGCATCATTGATATTGAAAAATACTGCATCTTTGGTTGTAGATAAGTTTTTGTAGATGTTATTTGCTTAAAGAAAAAGAAA
 ACCAAATAAAAAATGTTTCTTTGTTCACTAGAGAGTCTCTCAATAGATGGTTAGAAAAGCAGGAACAGGTAACAATAGGCCAGCTCCGAGATTTGATTTTG
 TGGATTTGCCAGAGTTGATCAGTACAGACACATGATTAAGCACAACCAACAAAAAGTTGGACACTTCAATCCAAACCGGAGTACCCGGCTTTGCGAGCG
 ATTTGTGTACCATTCAAAAAAGATCAATGCAATATTCCGGCCGTTGTTTAGTGTGACTTACAGGCAATACTGGACAGTTTCTGAGCAGATTTTGTGTT
 TTTCAAGAAAGACACAGCGCAGATTGAGGATTTCTCGGAGATCTCGACAGTCTGTTCCGATGGATGCTTGGAGCTGGATATATCAAAATACGACA
 AATCTCAGATGAATTTCCACTGTGCAGTAGAATACGAGATCTCGGCAAGATTTGGGTTTCCGAACTCTTTGGGAGAAGTTTGGAAACAGGGCATAGAAA
 ACCACCTCAAGATTATACCGCAGGTATAAAAACTTGCATCTGGTATCAAGAAAGAGCGGGGACCTCACGACGTTTCAATGAAACACTGTCACTGATTC
 TGCATGTTGGCCTCGATGCTCCGATGGAGAAATAATCAAAAGACCTTTTCCGGTGCAGATAGTCTGCTGACTTTCCAAAGGGTTGTGAGTTTCCGG
 ATGTGCAACACTCCGGAACTTATGTGGAATTTTGAAGCAAACTGTTTAAAAAACAGTATGGATACTTTTGGGAAAGATATGTAATACATCACGACAGA
 GGATGCAATTTGTGATTCAGATCCCTTAAAGTTGATCTCGAACTGTTGTTTAAAGAGTGTATGTGTTTCCAAAGTTGATAAAAATAAGTTGTTCAATGAAATGATGC
 TGATGTTGCTGTTTCTGTGAAACAAATGTGCGTATACACACAGTTGGACGACGCTGTATGGGAGGTTCAAGACCGCCCTCCAGGTTCTGTTGTTTATA
 AAAGTCTGTTGAAGTATTTGCTGTATAAAGTCTTTTTTAGAAGTTTGTATATAGATGGCTCTAGTGTGTTAAAGGAAAAGTGAATATCAATGAGTTTATCGA
 CCTGCAAAAAATGGAGAAGATCTTACCCTCGATGTTTACCCTGTAAGAGAGTGTATGTGTTTCCAAAGTTGATAAAAATAAGTTGTTCAATGAAATGATGC
 TGTGAGGGTGAACCTTCTTAAAGGAGTTAAGCTTATGATAGTGGATAGCTGTTTGGTGGTTCAGGGCAGTGGAACTTGCCTGACAAT
 TGCAGAGGAGGTGTGAGCTGTGCTGGTGGCAAAAAGGATGGAAAGAGCCGACGAGGCCATCTCGGATCTTACTACACAGCAGCTGCAAAAGAAAAGATT
 TCAGTTCAAGGTCGTTCCCAATTTATGCTATAACCCACCCAGGACCGGATGAAAACACTGTCGCAAGTTTATGTTTAAATATGAAATGATGCACGCGG
 GTTTCTGTCGCTTTCTCTGGAGTTTGTGTCGTTGTGTTTATAAATAATATAAAATAGGTTTGGAGAGAGAAGATACAAACCTGAGAGACGGA
 GGGCCATGGAACTTACAGAAGAAGTCGTTGATGAGTTTCAATGGAAGATGTCCTATGTGCATCAGGCTTGCAAAGTTTCGATCTCGAACCGGAAAAAGAG
 TGATGTCGCAAAAGGAAAAATAGTAGTGTGATCGGTGAGTCCGCAACAAGAACTATAGAAATGTTAAGGATTTTGGAGGAATGAGTTTAAAAAGAAATA
 ATTTAATCGATGATGATTCGGAGGCTACTGTCGCCGAATCGGATTCGTTTAAATAGACTTACAGATATCACTACTCCATCTCAGTTGCTGTCTTGTTCAT
 TAATAGAAATAATCCGCTCTACTCAATCATCGGCTCGAAACGATGGCAGTTGGCTCGAAAGTTTGGCAGCGCTCAAAGTTTATGATGCAAAAACCC
 GCGCAGCGTACTGATGCTCTACGCTCTACGCTTGTGAGTGTATTGACGATGTTTACGACTCAGACGCTGGGCTTTCCAGCCCGGACGCTCCCTTACAAACG
 CCCGAAACAGCTGATGACAACTTGGAGATGAAAACCGCCAGGCTATGCAAGTATCCAGATGCACCAACCGCGGTTTGGCGCTTTTCAAGAAATGGCTAT
 GGCTCATGATATCGCCCGGCTTACGCTTTGATCATCTGGAAGGCTTCCGCTAGGATGTACCGAAGCGCAATACAGCCAACTGGATGATACGCTCGCT
 ATTGCTATCAGTTTCAAGGCTTGTGCGCTGATGATGGCGCAAAATCATGGGCTGCGGGATAACGCCACGCTGGACCGCCGCTGCTGCTTGGCTGGCA
 TTTCAATGCAAAATATGCTCGCGATATTGTTGGACGATGCGCATGCGGGCCGCTGTTATCTGCGCGCAAGCTGGCTGGAGCATGAGGCTGTGAACAAGA
 GAATATGCGGCCACTGAAACCGTCAAGGCTGAGCCGCTATCGCCGCTGTTTGGTGCAGGAAGCAGAACTTACTATTGCTCGCAACAGCCCGCTGG
 CAGGTTTGGCCCTGCTTTCGCTTCCGCTACCGCGCAAGCAGGTTTACCGGAAAAATAGGTTGTCAAAAGTTGAACAGGCGCGTCAAGAACCTGGGAT
 CAGCGCAGTCAACAGCACCGCCGAAAAATTAACGCTGCTGCTGGCCGCTCTGTTGTCAGGCGCTTACTTCCCGGATGCGGGCTCATCTCCCGCCCTG
 GCATCTCTGGCAGCGCCGCTTAAAGGGTAGTCAAGTGCATATAAATAACGATTTGTTGTCGCTAATCACAGTGGGTAGCATAAGCAGATGATGTT
 TTTCCCTCCACTTAAATCGAAGGGTTGTGCTTGGATCGCGGGGTTCAAATGTATATGTTTACATATACATCCGACGACGTAATAAAGCGAGGGGTTCCG
 GTCGAGGTCGGCTGTGAACTCGAAAAGGTTCCGGAAAAACAAAAAGAGAGTGGTAGGTAATAGTGTATAAATAAGAAAATAATAATAGTGGTAAGAAA
 GGTTTGAAATTTGAGGAATTTGAGGAATGTGAAGTGTAGCAGTATGACGATCTCGCTATCGAGTACGTTTTAATCAATAGCTTATCACTCACTCACTCCG
 AGCCAATTTGTTTACTTAAGTCCGCTTATGCAGATCCTGTGCAGTGTACATCTGTGTACAAATGCATTGGGTAACAGTTTCAAACGCAACAGCTAG
 GACAAACGTCACACAGCAATTTCCGGATGCTGGAAACCTGTGCTAGTATGACAGTGTGATTTCCGTCATCGGATTTCTATGTGATAGATATAATTCGA
 CGTGTGATCGCTGATCAGCGCTTATTAAGTCTCGATACAGTATGAAATAGAGTTGATAATCAACCCCGCAGCAATTAATGACAACTCACTCACTCCG
 AACCGACTCAGAGGTAGACGATGCGACTGTAGCTATAAGGCTTCAATCAATAATTTGGCTAATGAAGTGGTTCGTTGAACTGGCATGTTCAATCAAGC
 AAGCTTTGAGACTGCTAAGTGGACTTGTCTGGACCACAACCTCCGCTACTTAGCTATTGTTGTGAGATTTCTTAAATTAAGCTCACTGAGACTTAAATTC
 AGGGTGGCTGATGACAAAATCAGCAGTGGTTGTTCTGCTCACTTAAATAGACTATGTCATATCTGGATCCAAACAGTAAACCAATGATGTTGTTACTGT
 TGGATGCGGTAACCAACGAAAAAGTCCGTGAAGACTTAAATTCAGGTTGGCTGATACCAAAATCAGCAGTGGTTGTTCTGCCACTTAAAAATAACGAT
 TGTCAATCTGGATCCAAACAGTTAAACCAATGTTGATGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
 GCCCA

TMV sequence corresponds to vector TMV 30B, which includes the 3' end of *Tobacco mild green mosaic virus* (see reference 41). cDNA corresponding to *P. ananatis crtB* is on yellow background.

>PVX-crtB

GAAAACTAAAACATACACCACCAACAACCAAAACCCACCAGCCCAATTTGTTACACCCCGCTTAAAAAGAAAGTTTAAACAAATGGCCAAAGGTGCGCGA
 GGTTTACCAATCTTTTACAGACTCCACCACAAAACCTCTCATCCAAAGTAGAGGCTTATAGAAAACATTCGCCCCATCATGGAGGCTCAAGAACTAGCTTAACC
 CTTACGCTCAAACGGTTGAAGCGGCTAATGATCTAGAGGGGTTCCGCATAGCCCAATCCCTATAGCATTGAATGTCATACACATGCAGCCGCTAAGACC
 ATAGAGAAATAAATCTTAGAGGTTGCTTGGTTCCATCTACCAAGAACCTGTGTACATTTATGTTTCTTAAACCCAGAAAGCTAACTTACATGAGAAGAAA
 CCGCGGATCAAGGACATTTTCCAAAATGTTGCCATTGAACCAAGAGACGTTAGCCAGGTACCCCAAGGAAACAATAATTTGACAACTCAGAGACTCACA
 CGGAAACAGCATACATTTAGTACACTCTGCATCTTGGATCCGAGCTACATAGTGGAGACATCCAAAACCTGCCAAAATTTGCAAAACATTGATGCGACC
 TTAGTTTCCCGGTTGAGGAGCCTTTAAAATGGAAAAGCCTCCACCGAATATACAGCCCTCAAATACTTCCGAGATGTTTCCAGTATATACAGGCA
 CCATGGTGGCGGGGATACCATGAAATTCGCTCATCTACAATGGCTCAAAGTGGGAAAGATCAAGTGGAGGGACCCCAAGGATAGCTTTCTCGGACATC
 TCAATTACAGACTGAGCAGGTTGAGATGCACACAGTACAGTACAGTTGCAGGAATCGTTCCGCGCAACCACTTGTACTGCATCAGGAGAGGAGACTTG
 CTCACACCGGAGGTGCGCACTTTCCGGCAACCTGACAGGTACGTTGTTCCACCACAGATCTTCTCCCAAAAGTTTCACAACTGCAAGAAAGCGGATTTCTCA
 GAAAACATGATGACAGCTCTTCTGATGTTAGGACAGTCAAGGTCGCAAAAATTTGACATTTTGGCAAGTCAAGAAATTAATAATCACTGATGATG
 TGGCAAAATACTGCTGTGGAAGTGGTTTACTTAGTAAAGTACATGGAGTTTCTTCCGATTTCAAGACTACCACCTGCTTCTCAGACACTTTCTGGT
 GGCTTGTCAAAAAGACCTTGCACCAGGTTGAGGGCTTGGATACAAAGAAAAAGATGCAGCTGTTTGGTCTTGGAGACTACGCGAAGTTAGTCAAAAGCAGT
 TGATTTCCACCCCGGTTGATTTTCTTCAAAGTGGAACTTGGACTTCCAGTCTCACCCCTTGCAAGCGTGGAAAGCTTCCGACAGGAGGAAAGCTTCGG
 ATGTAGAGGAAATGGAAGTTTGTCTCAGATGGGACCTGCTTGTGTTTCCACAAGAAATGCCAGCTTATGCGGTAACCGCAGAGGAAGATTTAGCTGCA
 ATCAGGAAAACCGCCGAGATGGATGTCGGTCAAGAAGTTAAAGAGCTGCAAGGACAGAAATCAATACTCAAACCTGCAGAACTTTCTCAACAAGCT
 CCACAGGAAACACAGTAGGAGGTGAAACACCCAGGCGCAAGAAAGCTAAACGCTTAGCTGAAATCCAGGAGTCAATGAGAGCTCAAGGATGAGGCAATGATCA
 CAAATGAAATAGCGGACGATGGGGCAATACCCAGCAACCGCAACTCCTGGCAGAAATGATGCCAGACAAGAACTCACACTCCCAACCACTAAACCT
 GTCCCTGCAAGGTGGGAAGATGCTTCAATCAAGATTTAGTGTGGAAGAGGAGCAGGTTAAACTCCTTGGAAAAGAAACCGTTGAAACAGCGCAGCAACA
 AGTCAATCGAAGGACTTCCCTTGAACCACTGGATTTCCCTCAATTAATGCTGTGGATTTCAAGGCGCTGGAAAATTCAGAGGATAGGAGTGAAGGATGATCA
 TGCCATCACAGAAATGGTCTCCGGGCTGAAAAAGAGGACTTCCCTGAAGGAACTCCAAAAGAGTTGGCACGAGAAATGTTGCTGATGACAGAACTGATCA
 GCCACACTCCCTTTGGACCTGCTTAGAGCCAGACTACGGCAGTGTATGAAAGAACAAAGAAATGGTGCATCACAAAGACACAGGCAACAGGTTGGGG
 CGAAATCTGACAGGAAAGATAGAAAGCTTAACTGAGAGAAAGTTCGCACTGTTGTCATGTCATGAGGCTGAGGAGTGGGATGAGGATGAGGATGAGGATG
 AGGCAATGAGAGAAATTTGCAAGGGCTCGGACATCACTGTAGTCTGCCAGCAATGAACCTGCGGCTAGATTTGGAGTAAAGAAAGTGCCTAACACTGAGCC
 TATATGTTCAAGACCTTCAAAAAGGGCTTAATTTGGGGAACAGGCAGCATAGTCACTTTGACGATTACTCAAACCTTCCCTCCGCTTACATAGAAAGCTT
 AGTCTGTTTCTACTTAAATCAAGCTAACTTCTAACAGGAGATAGCAGCAAAAGCGTCTACCATGAACTGCTGACCCGCTCCATCAGGCAATTTGG

GACCAGCAACAGACTACTTCTCAAATACTGCCGACTACTCTCAATGCCACACACCCGCAACAAGAAAGATCTTGGCAACATGCTTGGTGCTACAGTGAG
 AGAACCGGATACCCGAAATCAGCATGAGCGCCGAGTTCTTAGAAGAAATCCCAACTTTGGTACCCTCGGATGAGAAGAGAAAGCTGTACATGGCCACCGG
 GAGGAATGACACGTTTACATACGCTGGATGCCAGGGGCTAACTAAGCCGAGGTACAAAATAGTGTGGACCACAACCCCAAGTGTGTAGCCGGAATGTGA
 TGTACACGGCCTTCTTAGAGCCACCGATAGGATTCCTCTGTAACACAAAGTCAAAATAGTGTGGACCACAACCCCAAGTGTGTAGCCGGAATGTGA
 AAGACTTTCCTTACGTGGTGGAGAACCAAGCACTCAGGAGATCAGCGGCGAGGCAAGTTCAGAGCCCTGAGCCCCAGACACCATGTGTGT
 CGAGAATGAGGAGTCCGTGCTAGAAGAGTACAAGAGGAAGTCTTGGAAAAAGTTTACAGAGAGATCCACTCTGAATCCCATGGTCATTTAAACTGTGTCC
 AAATGAAGACAAACCATTGATGTTTTCCGATCAACAAAGCAAAAGATGAGACCCCTCTCTGGGGACTATAGATGCGCGGCTCAAGACCAGCAATCAA
 GAAACAACCTCCGGAATTCCTGAGCAAGAAGGACATTTGGGGACTTCTGTTTTTAAACTACCAAAAGAGCTATGGGTTTACCCAAAGAGCGGTATTCCTTT
 TTCCAAAGAGGCTGGGAAGCTTGTGCCACGAAGTACAAGCAAGTACCTCAGCAAGTCAAAGTCAACTTGATCAATGGGACTGTGAGACAGAGCCAG
 ACTTCGATGAAAATAGATTATGGTATTCCTCAAGTGCAGTGGGTACAAAGGTGGAAAAACTAGGTCTACCCAAAGATTAGCCAGGTCAAAACCATAGCA
 GCCTTTTACCAGCAGACTGTGATGCTTTTTGGAACTATGGCTAGGTACATGCGATGGTTTACAGACAGGCTTTCAGCCAAAAGAGTCTTCATAAACTGTGA
 GACCACGCCAGATGACATGTCTGCATGGCCCTTGAACAACCTGGAATTTACAGACAGCTAGCTTGGCTAATGACTACACAGCTTTCAGCCAGTCTCAGGATG
 GAGCCATGTTGCAATTTGAGGTGCTCAAAGCCAAACACCCTGCATTCAGAGCAAGAACTATTAGGATACATAGATATTAGACTAATGCACAGATTTTC
 CTAGCCACGTTATCAATTTAGCCTGACTGGTGAAGTCCCACTTTTGTGCAAACTGAGTGCAACATAGCTTACACCATAAGAAATTTGACATTCCTCC
 AGCCGGAACCTGCTCAAGTTTATGCAAGGAGACTCCGCACTGGACTGTGTCCAGAAGTGAAGCATAGTTTCCACAGGCTTGGAGCAAAATTACTCTCTAA
 AGTCAAAGCCTGTAATCAGCCAGCAAAAGAGGGCAGTTGGCTGAGTTTTGTGGTTGGCTGATCACACAAAAGGGGTGATGAAAGACCAATTAAGCTC
 CATGTTAGCTTAAATTTAGCTGAGCTAAAGCTAAAGGCTTGCATTTCCAGACCAACTCCAGGATGATGAAATTTGATCTGAGTTAGCCATGATGACCAAGGACTCTCT
 GCATGACTTGTTCGATGAGAAGAGTGTCAAGCACAACACTCACTTGCAGAACACTAATCAAGTCAAGGAGAGGCACTGTCTCACTTCCCGCTCAGAA
 ACTTTCTTTAACCGTTAAGTTACTCTTAGAGATTTGAATAAGATGGATATTCATCAGTAGTTTGAAGTTTAGGTTATTCTAGGACTTCCAATCTTTA
 GATTGAGCACTTTGTAGTACATGCGATAGCCGGAGCCGTAAGTCCACAGCCCTAAGGAAGTTGATCCTCAGACACCCAACTTACCGTGCATACACT
 CGTGTCTCCCTGACAAGTGAATATCAGAACTAGAGGCATACAGAGCCAGGACTATTCCTGAGGGCACTTCCGCAATCTCGATGAGTATACTTTGAGCA
 ACACCAAGAACTCTAACCCAGGCACTTTTTGCTGACCCCTTATCAGGCACCGGAGTTTAGCCTAGAGCCCACTTCTACTTGGAAACATCATTTCGAGTT
 CCAGGAAAGTGGCAGATTTGATAGCTTGGCTTGCATTTCCAGACCAACTCCGGAAAGAGGGCACTTAGAGATCACTGGCATTTCAGAGGCTTCAAAGGCTC
 CTTACTCGAAAGGTGATAGCCATTGATGAGGAGTCTGAGACAACACTGTCCAGGCATGGTGTGAGTTTGTAAAGCCTGCCAAGTACCCGAGCTTGAAT
 TCAAAGTAGTCACTATTGTGCTGCCGCAACATAGAGGAAATGGCCAGTCCACAGCTTCTACAACGCTATCACCAGGTCAAAGGGATTGACATATGTC
 CGCGAGGGCCATAGGCTGACCTCCGGTCAATTTCTGAAAAGTTACATAGTATTAGGTTCTATCATTTGCTTTAGTTTCAATTTAGCTTCAAGGCTTCA
 GAAATAGCTTACCACCGTCCGGTGAACAACATTACAGCTTGCACACCGGAGGACTTACAGAGACGGCACCAAAGCAATCTTGTACAACCTCCCAAAATCTA
 GGGTCCAGAGTGAAGTACACAACGAAAGAACGCAGCATTTGCTCCGCTTTGCTACTGACTTTGCTGATCTATGGAAGTAAATACATATCTCAACGCA
 TCATACTTTGGCTTTGGTAAACAATCATAGCAGTCAATAGCAGTTCCTTAGTAGGACTGAACCTTGTGTCATCAAGTTTACGGGAAATCAATCAGAGT
 TTGGCTTGAACCTAGATGAGAAACCAATAGGGCCATTGCCGATCTCAAGCCTACTCCGTTGAACGGTTAAGTTTCCATTGATCTCGAAAGAGGTCAG
 CACCAGTACGTAAGATACTCCGCTGCTTACTCAATCATGCGGTGAAACAGTGGCAGTGGCTCGAAAAGTTTGGCACAGCCTCAAAGTTATTGATGCA
 AAAACCGGGCGCAGCTAGCTGAGCTTACGCTTCCGCTTGGTCCGCAATTTGTGACGATGTTATTGACGATCAGACGCTGGGCTTTCAGGCCAGGCACTGCTT
 ACAAAACCGCCGAAACAGCTGATGCAACTTGAATGAAACCGCCAGGCTTACAGGATCGCAGATGCACGAACCGCGCTTTCGCGCTTTTCAGGAAG
 TGGCTATGGCTATGATATCGCCCGGCTTACCGCTTTGATCATCTGGAAGGCTTCGCCATGGATGTACCGCAAGCGCAATACAGCCAACTGGATGATACG
 CTGCGCTATTGCTATCAGCTGCAGGCTTGTGCGCTTGTATGATGGCGCAAACTATGGGCGTGCAGGATAACGCCACGCTGGACCCTGCTGACTTTGG
 GCTGGCATTTCAGTTGACCAATTTGCTCGCATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCGCAAGCTGGCTGGAGCATGAAGGTTCTGA
 ACAAAGAGAATTATGCGCCACTGAAACCGTCAAGCCGCTGAGCCGCTGCTGCCGCTGCTTGGTGCAGGAAGCAGAACTTACTATTGCTGCCACAGCC
 GCACTGGCAGGTTTCCCGCTGCTTCCGCTGGCAATCGTACGGCGAAGCAGGTTTACCAGAAAATAGGTTCAAAGTTGAACAGCCGCTCAGCAAGC
 CTGGGATCAGCGGCACTCAACAGCACCAGCCGCAAAAATTAACCGTGTGCTGGCCGCTTGGTCAAGGCTTACTTCCCGGATGCGGGCTCATCTCCCTC
 CCCTTCGCACTCTGTCGCAAGCCGCTCAGGCTTACGCTTTCCTTTTACTCGAAAAGATGTCAGGCTGTTACCATCCCGGATGGGATTTCTT
 TAGTACAGCCCGTGCATAGTAGCCAGCAATGCTGTGCAACAATAAGGACTCAGCAAGATTGAGGCTATTTGGAAGGACATGAAGGTGCCACAGACA
 CTATGGCACAGGCTGCTGGGACTTAGTCAAGACTGTGCTGATGATAGGATCATCCGCTCAAAAGAAATGATAGATACAGGCTCCCTTTCCAAGCCGATCA
 AGCAGAGCTAGACTGGCAGAGCAATTAAGAGGTGTGCACACTTGGCAATTTGATGAAATGATGCTCCAGTGGATGGAATTTAACTAACACA
 CAGTCCACCTGCTAACTGGCAAGCACAAGGTTTCAAGCCTGAGCACAATACTGCTGCAATTCGACTTCTTCAATGGAGTCAACACCCAGCTGCCATCATGC
 CAAAGAGGGGCTCATCCGCCACCCTGTAAGCTGAAATGAAATGCTGCCCAAACCTGCTGCTTTGTGAAGATTACAAGGCGAGGGCACAATTCACACGAC
 TTTGCCAGCCTAGTCAAGCTGTCTCAGCTGATCAGTGAACAACAGCCTGAGGCTGTTGTCAGTCTTACCACCCATACTAGCTCTCACTATAA
 CCGACGCTTACCAGTTTCATAGTATTTCTGGTTTGTATGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
 AA

PVX sequence corresponds to M95516.1, but contains a heterologous promoter that derives from *Bamboo mosaic virus* and the deletion of the 29 initial codons of PVX coat protein (see reference 44). cDNA corresponding to *P. ananatis crtB* is on yellow background.

>ZYMV-wt

AAATTAACAACAAATCACAAAGACTACAGAAATCAACGAACAACAACGAATTTTAAACCGTGTAAACAAACAGCAATTTATAATTGCGACAGCATCAAG
 AATTTCTGCAATCATTTTGTATTATTAGACACAACAATGGCCCTCAGTATGATTTGGTTCAATCTCCGTACCCATCGCACAGCCTGCGCAGTGTGCAAAACA
 CCCAAGCGAGCAACCGGGTTAACGTAGTGGCACCTGGCCACATGGCAACATGCCACCATCATTTGAAAACGCACACATACTACATGCATGAGCTAAGAAG
 TTGATAAATTCAGATAAAAAGCAATGAAATTTTGAACAAATTTCTTTAACAATGATGAGATGAAATTCGCGCTCACTAAGATGAAATGAGTAAGGTGAAAAA
 GGGTCCGAAATGGGAGGTAGTTCTTCGCAAAACCAGCAAGCAGCGGGTTTTCGTCCGGATCGACAGGATGAGGCAGCACGAAGGAAGAAGCTTTTTCC
 TTGCAGGAAATATGATGGTTCGATCACAATCTAGTGTAGTGTCTTCCATCTGAAATGACTCGCGATGTGATGCGAGTTTGGATCACCATTTTACAAG
 CGCACATATAAAGAGGACAGAAAAGAAAGTGGCCGAAAAGCAAACTCGCCAGGCACCCTTAAACAGCTTGTGCACACGTTCTTAAAATTTGCCGCAATAG
 AAACATCCCTGTTGAGATAATTGGCAACAAGAAAGCAAGACATACACTTACCTTCAAGAGGTTTAGGGGATGTTTTGTTGGAAGGTTGCAGTTGCACATG
 AGGAAGGACGAATGCAACACTGAGATATACAGAGCAGTTTGAATGGATTTACAAGCTATTTGTCGGGTACTTATACAGAGCAATTCGTGAGGAA
 GACATTAACACAGGTTGATGTTGGTGGTGTTAGGCACTGATCATACAACTGACCAAGAGATATTCAAGATTGCCACATCTGGTAATTCAGGTTAGAGATGA
 TGACGGGATTTGTAACCGCTGGAACCGGTTTATTTACAGCGAAAGTGAACCACTTTCGTCGCAACCGGAAAGTTCAGTTCTTCCAAGGATGGCCAGCA
 GTTTCGACAAGTTTAGACCCGCTCAGATCATGTGTGCAAAAGTTGATCACACAATGAGGAGTGTGGTGTAGTGTAGCAGCAATCTTTGTCAGGCTTTGTTTC
 CCAGTAGTGAAATTTCTCGTGCACAAACATGCAGAGAAAAGCTTAGTAGTGTAGCTTCGAGGAAATCAAAGATTCTTTGAAATCAAACTTTATTTACCATAA
 GGACGAATGGGATAGTTTCAAAGAGGCTCTCATACGATAAATTTCAAATTAATTAAGTAGCAACACAGGTAATCAAGATCTCAAGCTCTCACTCT
 AAGTAAAGACTTCTTTGAGGAAGTTATACCAAGTGAAGGATATACAAATACGATGATGCGAAACTTCCAAAAGTTCATGTAAGTTTCCGCAAGGCTCA
 TTGACTGTTGGCTTTGAAACAGCTTCTAGAAATGACTCAGTGGTTTAAAGAACCATGCACCTGACTGGTGAAGAAGCATTGAAGATGTTGAGAAACAAGCG
 CTCATGCAAGGCTATGATAAATCCTAACCTTTTATGTGATAACAGTGTGCAAAAATGAAAACCTTTGCTGGGGAGAAAGAGGATATCATTTCAAGGAT
 TATTTCAAAGACTTCTTTGAGGAAGTTATACCAAGTGAAGGATATACAAATACGATGATGCGAAACTTCCAAAAGTTCATGTAAGTTTCCGCAAGGCTT
 GTGAATGTTAATGAGAACGAGGCAAGGATTTACCACAAATGATTCGTGATGTTTTGATCCCTATGCTTGGCAATGGCCCTTCTGTATGGATGTCGCGAC
 TCAGACATACATTTTAGGTGATTTCCATCTGAAACCGCATGGCTGAAATACCTAGGATTTTGTGACCATGCTAGCAGCAACCATGCATGCTGATTTGAT
 CATATGGATTAATTAAGCTTTGGTTATCACGTGCTCAAGCGCGAAAGTGTCAATATTAAATCAATTTGCTCAAAATGCTCAGAGCGAGATGAAGCAT
 TACAGAGTCCGGCAAGCCCAACACAGCGCATAAAATTTAGGAGCAATTTGATTAAGGAATCTTCAAACCAAACTTATGATGCAGCTTTCACAGATGA
 CCCATACATATTTAGCTGGCATGATCTCACCACCATCTTCTGTGCATGATATGAGGATGCGTCATTTGAGCGAGGATTTGAAATATGGATTAAGAGAG
 ATCATGAAATTTGAAAGGATTTTCGTCATATTGGAACAGCTCACAGGAAAGTGTCTGGCTGAAGTTCTTGTGATCAGCTGATTTGATTAAGTGAAGCT
 TCACCACATTTACTTGAATCATGAGGGTTGTCAAGATAATCAAAGGCGTATGTACCTGCGCTGGATCTGTTAACGATACAAGTAGAGCGTGTGTTTTT

AAATAAAGAACTTAAACCAATGGCTATCCAGATTTGCAGCAAACGTTGTTTATGATATGAGAGAAAAATGTATGCAAAGCAGTTGCACAGTTTCATGGCAAG
AGCTAAGCTTGCAGAAAAATCTTGTGTAACCGTGCAGTTGAAGCAATTCGATTTTACGGAAGAAATTTAATCCAGCGAGCAAAAGAGGAAAGCGC
ACATCTTCGCTACAATTTGTTACAGAGTGCCTTATACAGCACCCGAGTACATGCGAAGAGCATTTCGCGATGCAGGCGTGCAGCAAGTAAATGAGGCTCTCGT
TGAACTTGTAAAGTCTTTTTCTCTGTTGTTTCAAGATTTTTCGCGGGTTCACAGGACATATATACCTTGTGAACGTGTGTTTGGTATCTCTCTCTG
TGTACAAATGTCTAACTGTGTCGCAACATGATAGCGGGCAGCAAGGAAAGAGAGAGCGATGGCAAAATAAGCTGATGAAATGAAAGACCTGTA
ATGCACATGTACCACATTTTCAGTAAGAAGCAGGATGAAGCGCCCATATATAACGACTTTCTTGAACATGTTTCGCAATGTAAGACCAGATCTTGAGAAAC
CCTCTGTACATGGCTGGTGCAGAGTGTGTGCAACACAAGCAAGTACAGCGTTTCAGATTTCAGTTTCGAGAAAGATTATAGCTGTATTTGGCCTGCTCCTA
TGTGCTTTGACCGCGAAGAAAGTGGAGGCAATTTTCAAGATTTTGAAGAAAGTTCGCAAAAGCTTTCGCAAAAGTTCGCAAAAGTTCGCAAAAGTTCG
GAAGATTTGAGAGCTTGAAGACGATAAGAGACTCAAAATGACTTTGATATTAACAGCAATGAGGCTCAATCGTCGCAACGTTTGTGTTTCATTTTGA
TGATTTGGTGGAAATCGGCAGCTACAGCAAAATCGCACAGTTCCACATATATAGGACCACAGGTTAAATTCCTCGAAATTCACCAGAAGCACTGCAGCTTTTGG
CTAATGAAATAGCATCACTAAGTGAAGGAGAAATTTTAGTTAGAGGAGCAGTGGTTCGCAAAATCAACGAGCTTGCTGCGCATCTTTCCAAAGAGGTT
AAAGTATTACTACTCGAACCCACAGCCCTTTGGCGGAGAATGTAGTACAGATTGGCAGGCGATCTTTTTTCCAAAACGTCACACTCAGAAATGAGAGG
GCTAAATTTGCTTTGGTTCAAGTAACATTACAGTGTGATGACGAGTGGATTTGCTTTTCTACTACTATGTTAATAATCCACACCAATTAATGAAATTTGACTTTA
TTATCATAGACGAATGCCATGTTACGACAGTGCATATAGCTTTCAATTCGCGCATTAAGGAGTAAATTTTGGTGGCAATTTGCTGGCAATTTGATGAAATTTGA
ACGCCCGCAGGGAGGAGTGTGATTTGATACGCAATTCGCGGTGAAAGTCAAACCGGAGGACCACTTTTCAATTCGATTCGTTTGGCGCAGAGAAGAC
TGTTTCAAACGCTGATGTTTCAGCATGGCAATAACATACTTGTGTATGTTGCAAGTTATAACGAAAGTAGACATGCTTTCCAAATTTACTCACTGAGCGAC
AATTTTCAGTGCAGAAAGTGTGAGTGCAGTTCGCAACGAAATGCAACTTTGGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTG
GAAATGTCACAGTAAACAGATGAGATGTGCTTTGAATTTGAGTTAATTCCTTTTACCCTCAATTAATTCGCAATGATGGTAGCATGCACCACT
GATACAGAAAGAACTGAAACAATTTAAACTCAGGACTCAGAAATGGTGTCAACAAGTTGCATTACCTCACCAAATTTGTGAGTCAATGGTGGATCAAA
GTAGATGAGACGATTTGGAGTGCAGCTTCAATGCCATGAGAGCACAGCAGCTTATCATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCA
AAGTGCATACAGAAACAAGATGATGCGGTTTTGGTAAGCTTCAAGTGTGTTGCTGACTAAGTTAGTTATACACTCAGCACTGATCCAGCAGCAT
ACCCAGAATTTGCAATCATCGACCACCTGCTTGGCGAGAAATGATGAAGCGGAATCACTTCGACACGATCAGCTCAGCTGTAAACGGGCTATTCATTTT
CCCTCGCTGGAATTTGCTGATTTCTTTAGGAAAAGGATATATGCGCGATTAACACAGCGGATAACATTTGCAATTTCCAAACGCAAGTGCAGTGGCTGCGA
TTCAATAGTAAAAATGTGAACATCAACAACCTGTCGATCGGAAGAAATTTGGAGTTAATAAGTCGGTGGTGTGCAAGTAAAGCAAGAGGTGCAAGCACT
CCTAGGACTTCGCGGTAAATGGGATGGCGGAAATTTGCGAATGATGTGATATTTGGCGATCATGACACTCTTAGGAGTGGATGGTTTATGTTGGAAATCT
TCAGAAAAGAGTCAATGAACCGCTGCGCGTTGAAGCAAGAAACCGCGATCTCAAAGTTGAAATTCAGGATGCGTACAGATAGGAAAGTGCAGCGTGGAG
ATTTTGGCGATGATGACACAAATTTGGCGCACTTTTGGCGAAGCTTACAGGAGAGAGAAAGTCAAAGGAAACAACAGCAAAAAGAAATGGGACGGAA
AACTCGCAATTTTGTGATTTTATGTTGTTGAGCCTGAGAATACAGCTTTATAGATTTGTGGACCTCTCAGTGGCCATACATTTGGACGAAAGCACC
ATACAGCATTTGCTAGTGCAGGAGGAGTTTGAAGTATTTAGAGAAATTTAGAGAAATGTTTGAAGAAATGTTTAACTCGAGGCAAGTAAATTAACAAAGCTG
ATTCAGCATATTTTATGGCAAGGGCACCAGAAAGCACTCAAAGTTGATTTGACFCTCATGTACCATTGCTTCTGTGAGAAACCAACATGCCATTGG
GGGTATCCAGAGAGAAAATGAGTTGAGACAACCGCACACCAAGTTAAGTTTCTTTAAAGACGTGCCAGAGAAAACAACATGTCGATTTGGAAGCA
GCAAAATTCGCAAAAGAGTGCAGCATTTCAATGGCAATTTCAACAACCTGCTGTCAATTAACGAAACGATTTGATGGCTCAAAGTATGATATGATTTGGT
ATTTGGCTACGGGGCGATAATCATTACTAATGGGCACTCTTCAGAAAACAATGGTACACTTCTAGTCAGGCTTGGCATGGTGAATTCAGCTGTTAAAAA
TACCACACGCTCAAAGTCCATTTATAGAAAGGAGGATTTGTTTGTAGTGCATGCAAAAGGACTTTCCACCGTTCAAAGCAACGCTTTCTTTAGAG
CGCAAAAACGGGAAACGAGCATGCTTGGTGGAAACAATTTTCAAGAAAGAGTCTCCGCTCCACTGTTTTCAGAACTTCAATGACAATTAACGAAAGGAA
ACTGGCTCATATTTGATTTTCAAGCAATTAAGGAGTTCACCAATTAAGGGGACTGCGGATTTACCAATGGTTTCAACAACGGATGGTAAAGATAATTTGAGCTTCATGGAT
GGCTTCCACAGTCTCATTAAGAAATTTTGTCCCATTCATGATGATTTTATAGCCAGCACTTTGAGCAAGCTTTGATGATTCACATGGACTCAGCAT
GGCTATGGCAACCTAGCAAAATCGCTGGGGAAACGCTCAACTTAGTTGATGAACAACAGGGCCGTAATTTTCGATTTTCAAACTAGTCAAGGATTTGTTT
ACTTCTGGTTGTAACACACAAGCAAGCGGAAAGTGGTCTACGAAAGCTGGAAGGAAACCTTCGAGCTGTTGGAATCGCCGATTCAGGCTTAGTCAC
CAAACATGTTGTCAGGGCAAGTGTCTTTCTCGAAGAATTTTGAAGAAATTTTGAACAACACAGCAAGAAAGGAGGCTTATTTGACACCTTAATGGGAGTTTACAGC
CGAGCAAGTTGAACAAGAGGCTTCAAAGAGGATTTCTTTAAATACAACAACCCGTCCTGTTAATCAATTTGGATCATGATAAATTTTGAAGCAGTG
GATGGGTTATACGATGATGTTGACTTTGAGTTCAATGAATGCCGATTCATTACAGATCCCGAGGAAATTTCAACTCTCTGAAATGAAGAGCAGCAAT
TGGAGCCCAATATAGAGGAAAGAAAGAAATATTTTGAAGGGCTGATGATTTCTGCAAGCAAGCAAGAAAGGAGGCTTATTTCAAAGTGTGTAAGT
GCTATAAAGTTTGTGGAATGGATCTTTAAAGGCTGAGCTCAGGCCGCTTGAAGAAAGTCAAGGCTAACAAAACAGCACTTTTACAGCAGCGCCAAATTTGAT
ACATTTGCTGGAGCTAAAGTTTTCGCTGGATGATTTCAATAATGAATTTTACAGAAAATCTCAAGTGTCCATGGACGGTGGCATGACAAAATTTTATGG
TGGTTGGGATAAATGATGAGATCGTTACCTGATGGTTGGTTATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
CAGTCTTATAATCAGTCTTTTTACATGGAAGATTGGTGGTTGGCCAAGAGATGCTCGAAAATCTCTATGCTGAGATTGTGTACTCCAATTTCTTGCT
CCGATGGAACAATTTTCAAGAAATTTAGAGGTAACAACAGTGGGCAACCTCAACAGTGGTGGATAACACACTAATGGTTGATGATGATGATGATGATG
GTGATGAAGTTTGGGTTGGAATTTGCGAGGAAATTTGAGAAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAAT
GCTTACTTGATAACATGTCATCTCTTTTCCGAACCTGGACTGAATTACGATTTTTCGGAACGCACGCAAAAAGAGAAGATCTTTGGTTCATGCTCCAC
CAGGCAATGTTAATGATGAAATGTACATCCAAAATCTGAGAAGAGAGAAATTTGTTCAATTTAGAAATGGATAGAAATGAAAGAAATGCAACGAAAC
AGAGGCTATTTGGCTGCAATGATTTGAAGCATGGGGACACACAGGATGTTTCAAGGAAATCAGAAAAGTTTTACCTGTTGTTGTTGTTGTTGTTGTTG
GAGAATTTGGCTGCCCTCGAAAAGCTCCATACATAGCTGAGCAGCTCTTCGTAAGTTATACACTGACAAGGGAGGCGGATACGAGTGAACGGCAGCTAT
CTACGAGCCCTCCATCAAGATACTTCTTTGAACAAGGAGACACTGTAATGCTCCAATCAGGCACTCAGCAACTGTGGCAGAGCCTGGGGCTACAAAAGAA
AGCAAAAGAAAGATGACAAAAGGAAACAAAGGATGTTTCAAGGAAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTG
GTTCTCATGGGAAGATCGTCCCGCTCTTCAAAGATCACAAGAAAATGTCCTGCCACCGCTGAAAGGAAATGTGATCTCGATATCGATCATTGTGCTG
GAATATAAACCGGATCAAATCGAGTTGTAACAACACAGCAGCGCTCTCATCAGCAATTCGCTTCTGTTTAAACCAAGTTAAACAGAAATATGATCTGAATGA
GCAACAGATGGGAGTTGTAATGAATGGTTTTCATGGTTTGGTGTATGAAATGGCACTCACCAGCATTAATGGAGTGTGGGTTGATGATGAGGAAAGT
AGCAAGTTGAATATCTTTGAAACCAATAGTTGAAAATGCAAGGCCAACGCTGGCGCAAAATATGATCATTTTTTCAGATGCGAGCGGACATACATAGAG
ATGAGAAATGCAAGGACCATACATGCCAGGATGATGGTTGCTTCAAACCTACGGGATAGGAGTTTGGCAGCATATGCTTTGATTTCTATGAAGTCAA
TTCTAAAACCTCTGAAAGAGCCGCTGAAGCTGTTGCGCAAAATGAAAGCAGCAGCTCTTAGCAATGTTTCTTCAAAGTTGTTTGGCTTGTGAAATGTTG
CCACTACTAGCGAAGACACTGAAACGGCACACTGCACGTGATGTTAATGAAACATGCAACCTTACTAGTGTGAAATGCAATGCAATGCAATGCAATGCAAT
CTACCTAGGTTATGTTTTCGCTGCCAGCTAATCTAATATTTACCCTTTATTTGATATCTTTAAATTTCTAGAGTGGGCTTCCACCCTTAAAGCGTAA
AGTTTATGTTAGTTGTCAGGAGTGCCTGAGTCTTGTGAGGCTTGTAGTGGAGCCTCTCACGAAATAGCTCGAGATTAGACTCCGTTTCCAGGCTTAA
AA

This ZYMV sequence is pending to deposit in GenBank. Limits between ZYMV
cistrons are marked on blue background.

>ZYMV-crtB (insert between positions 8541-8542 of ZYMV-wt)
TCTGGCACAATGAATAATCCGTCGTTACTCAATCATGCGGTGCAAAACGATGGCAGTTGGCTCGAAAAGTTTTGGCAGAGCCTCAAAGTTATTTGATGCAAA
AACCCGGCGAGCGGACTGATGCTTACGCTGTTGGCGCCATTTGTGACGATTTATGACGATCAGACGCTGGGCTTTCCAGGCCCGCAGCTTCCCTTAC
AAACCCCGCAACACTCTGATGCACTTGAGTGAAGAACGCGCCAGCCTATGCAAGATGCAAGATGCAAGCAAGCGGCTTTGCGGCTTTTTCAGGAAGTG
GCTATGGCTCATGATATCGCCCCGGCTTACGCGTTTGTATCATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGATGATACGCT
CGCCTATTTGCTATCAGTTGCAGCGCTTGTGCGCTTGTATGATGGCGCAAAATCACTGGGCGCTGCGGGATAACGCCACGCTGGACCGCGCTGTGACCTTGGCG
TGCCATTTTCAGTTGACAAATATTTGCTCGCGATTTGTGGACGATGGCCTACGCGGCGCTGTATCTGCGGGCAAGCTGGGATGGAGCATGAAAGTTGCAAC
AAAGAGAATTTGCGGCACTGAAAACCGCTCAGGCGCTGAGCCGATGTCGCCCTGCTTGGTGCAGGAAGCAGAACTTACTATTTGTCTGCCACAGCCGG
CCTGGCAGGTTGCCCTGCGTTCGCCCTGGGCAATCGCTACGCGCAAGCAGGTTTACCAGAAAATAGGTTCAAAAGTTGAACAGGCGGCTCAGCAAGCCT
GGATCAGCGGAGTCAACGACCACGCGCCGAAAATTAACGCTGCTGCGCGCCCTGGTTCAGGCGCTTACTTCCCGGATGCGGGCTCATCTCCCGG
CCTGCGCATCTGCGCAGCGCCCGCTCAGGAGATACTGTGATGCTTCA

cDNA corresponding to *P. ananatis crtB* is on yellow background. Sequences corresponding to artificial ZYMV NIaPro cleavage sites are in blue.

>TEV-Ros1 (insert between positions 8541-8542 of ZYMV-wt)

```

CTGGCACATGGAAAAGAATTGTCGTGGAGTGAGAAAAGTACTTGGACCAAAGAAGAAGACACTCTCTTGAGGCAATGTATAGAAAGATATGGTGAAGG
GAAATGGCATCAAGTTCCACACAGAGCAGGGTTGAACCGGTGTAGGAAGAGTTGCAGGCTGAGGTGGTGAATTATCTGAGGCCAAATATCAAAAAGAGGTC
GGTTTCGAGAGATGAAGTGGACCTAATTGTGAGGCTTCATAAGCTGTGGTTAACAAATGGTCGCTGATTGCTGGTAGAATTCCTGGAAGGACAGCTAAT
GACGTGAAGAACTTTTGGAACTCATGTGGGAAGAATTTAGGCGAGGATGGAGAACGATGCCGGAAAAATGTTATGAACACAAAAACCATTAAGCTGAC
TAATATCGTAAGACCCCGAGCTCGGACCTTCACCGGATTCACAGTTACTTGGCCGAGAGAAGTCGGAAAAACCGATGAATTTCAAAATGTCGGTTAACAA
CTGATGAGATCCAGATTGTGAGAAGCAAACGCAATTTACAATGATGTTGCGTCGCCACAAGATGAAGTTGAAGACTGCATTCAGTGGTGGAGTAAGTTG
CTAGAAACAACGGAGGATGGGGAATTAGGAAACCTATTCGAGGAGGCCCAACAAATTTGGAAAATCAGGGAGATACTGTGATGCTTCA

```

cDNA corresponding to *A. majus Ros1* is on red background. Sequences corresponding to artificial ZYMV NIaPro cleavage sites are in blue.

General Discussion

Photosynthetic organisms like plants are the backbone of life on Earth, able to transform solar energy into carbohydrates and oxygen. They are essential for human life since they serve as foodstuffs, feedstock, building material and fuel. Furthermore, plants are important sources of medicines and health promoting molecules, since they produce thousands of biologically active compounds in which the majority of the human population still relies as primary health care (Gurib-Fakim, 2006). Therefore, it is not surprising that humans tried to enhance plant productivity from ancient times. In the past thirty years, biotechnological tools enabled scientists make important agronomic trait improvements (Cressey, 2013; Chen and Lin, 2013). In addition, plants have been developed as biofuels as well as alternative platforms for pharmaceutical protein production (Ma *et al.*, 2003; Xu *et al.*, 2012). Many attempts have been also made to re-engineer metabolic pathways to induce the production of health promoting compounds in natural host plants (Farré *et al.*, 2014). Plant viruses have contributed to these objectives as transient expression systems enabling the fast and high level expression of several recombinant proteins (Sainsbury and Lomonosoff, 2014).

Plant viruses are obligate intracellular parasites that debilitate the host and often cause enormous economic losses in crop yields worldwide. Some of them –mainly those infecting staple crops in developing countries– have such devastating effects on plant production that occasionally could originate serious problems in food supply (Rybicki, 2015). The economic damage caused by plant viruses stimulated the research on their molecular biology, which also contributed to the understanding of eukaryotic biology and plant biology. Furthermore, the capacity of plant viruses to synthesize high amounts of viral proteins prompted plant virologists to take advantage of this fascinating attribute. The new technologies of genome manipulation enabled the construction of infectious clones of plant viruses that both served as experimental purposes and heterologous protein production systems in plants. Several plant viruses were successfully used to construct vectors for foreign protein expression, making in effect an ‘ally from the enemy’ (Pogue *et al.*, 2002; Gleba *et al.*, 2014; Peyret and Lomonosoff, 2015). These expression systems benefitted from plant virus gene expression strategies to build new vectors enabling protein production in microgram quantities per kilogram of fresh tissue. Nonetheless, the use of plant virus-based expression vectors requires a fine design and a deep understanding of virus biology.

Their relatively small genome helps gene manipulation, but the complexity of genome organization hinders the simple insertion of foreign cDNAs for protein expression.

Bedoya *et al.* (2010) developed a potyvirus-based gene expression vector which takes advantage of the genome expression strategy of the viral genus and expresses several heterologous proteins in equimolar amounts in the same subcellular localization. Potyviruses produce a large polyprotein, which is processed co- and post-translationally by three viral encoded proteases resulting in ten mature gene products in equimolar amounts (Urcuqui-Inchima *et al.*, 2001; Ivanov *et al.*, 2014). An additional gene product, P3N-PIPO, is expressed as the result of a frame-shift mechanism (Chung *et al.*, 2008; Mingot *et al.*, 2016). Despite of their apparent simplicity of this genome expression strategy, the processing of potyviral polyprotein seems to be finely regulated resulting in the production of viral proteins in the right amount, time and space (Merits *et al.*, 2002; Ivanov *et al.*, 2014). The regulation is mainly based on the recognition efficiency of a viral encoded protease, NIaPro, responsible for the proteolytic processing of the main genome block of potyviruses from P3 to the end of the polyprotein (Carrington *et al.*, 1993; Parks *et al.*, 1995; Kim *et al.*, 1995). Several studies demonstrated that there are some intergenic positions that enable insertion of foreign genes without interrupting viral functions resulting in viable chimeric viruses. These sites are located between the P1/HC-Pro and NIb/CP junctions and has been frequently used by many scientists to insert the cDNA of heterologous proteins (Dolja *et al.*, 1997; Fernández-Fernández *et al.*, 2001; Beauchemin *et al.*, 2005; Dietrich and Maiss, 2003). In addition, an insertion site near to the amino terminus of *Potato virus A* polyprotein was also identified (Rajamäki *et al.*, 2005).

For a deeper understanding of regulatory elements on potyvirus genome, and for the identification of possible new insertion sites in the genome, an experimental setup was designed in which TEV genome was reorganized by relocating the NIb cistron to all possible intergenic positions of the genome –excluding P3N-PIPO– and the viability of chimeric viruses was tested. NIb, the RNA-dependent RNA polymerase, was chosen because it is known that this gene can be supplemented in *trans* either from a transgene or from a compatible viral vector for efficient virus replication (Li and Carrington, 1995; Bedoya *et al.*, 2010). Chimeric virus constructs were assayed both in wild-type and transgenic plants in which NIb was supplemented in *trans* from a transgene. Recombinant viruses with NIb relocated to the most amino-terminal end of

the polyprotein and between cistrons P1 and HC-Pro were the only viable virus constructs both in wild-type and transgenic plants. Chimeric virus harboring the relocated NIb between HC-Pro and P3 were rescued partially in transgenic plants. All other recombinant viruses showed to be non-viable. Our hypothesis is that in these last cases, the insertion of NIb induced lethal effects in polyprotein processing or disrupted partially processed gene products or regulatory RNA elements. Our results demonstrated that potyviral genome organization is much more complex than just a set of individual cistrons. These results served as the basis for further genome organization studies (Willemsen *et al.*, 2016). We found that TEV genome only tolerates insertions in certain sites like P1/HC-Pro and NIb/CP junctions, as previously described, and at the here discovered amino-terminal end of the polyprotein. This newly described insertion position opened the possibility to explore new questions of recombinant protein expression using the TEV-based vector.

Since plants are natural source of various specialized secondary metabolites which have a wide range of biological activities and are used for pharmaceutical, cosmetic and agrochemical purposes, huge efforts have been made to understand and manipulate metabolic pathways in plants, to boost the production of desired compounds with special attention on those which are conferring nutritional benefits for humans. The capacity of plant viruses to produce high amounts of proteins avoiding the time-consuming generation of transgenic plants has been exploited in this field. Plant viruses were used to elucidate endogenous biosynthetic pathways (Robertson, 2004) and re-engineer plant metabolism by expressing active enzymes and transcription factors (Kumagai *et al.*, 1995; Kumagai *et al.*, 1998; Bedoya *et al.*, 2010; Mozes-Koch *et al.*, 2012; Zhang *et al.*, 2013; Nielsen *et al.*, 2013). Between them, the potyvirus-based expression vector developed by Bedoya *et al.* (2010) was used to induce a large accumulation of anthocyanins, plant pigments with antioxidant properties, in tobacco leaves through the expression of two interacting transcription factors (Delila and Rosea1) from *A. majus*.

However metabolic pathway manipulation is not as easy as the expression of a sole enzyme even in the natural host and metabolic engineering of plants faces many challenges because of the tight regulation of the host genes (Farré *et al.*, 2014). One of them is the identification of the most suitable subcellular compartments of metabolic reactions and targeting the selected enzymes to these locations (Heinig *et al.*, 2013).

Therefore, our next goal was to study the important question of subcellular targeting of expressed proteins from the TEV-based viral vector. We explored if heterologous protein delivery is possible to different subcellular compartments and which insertion site is the most suitable for this purpose. Recombinant TEV clones harboring GFP at the amino-terminal site or at an internal site (between NIa-Pro and CP, replacing the NIb cistron) of the polyprotein with sequence motifs fused to the amino-terminal of GFP for chloroplasts, nuclei and mitochondria delivery were constructed. Our results revealed that the insertion position in the viral genome is crucial for protein delivery. Proteins expressed from the amino-terminal insertion site were successfully delivered, but only the GFP variant targeted to the nucleus was able to reach the corresponding organelle from the internal insertion position. These results may reflect the difference between the translocation process of proteins targeted to chloroplasts or mitochondria and proteins targeted to the nuclei. In chloroplasts and mitochondria targeted proteins throughout synthesis, the amino-terminal sequence are recognized by specific factors of the translocation complex, which keeps preproteins in an unfolded state to facilitate the translocation into the corresponding organelle. Nuclear transport of proteins differs essentially, since it occurs through the nuclear pore complex, which allows the transit of completely folded proteins. In the potyvirus vector context, the signal sequence of recombinant proteins inserted into the internal position may not be efficiently recognized by chloroplasts and mitochondria translocons because the tertiary structure of the polyprotein could mask the targeting information and therefore the amino-terminal signal peptide may not be efficiently loaded by the translocation machinery. In case of nucleus targeting, the translocation is successfully completed from both insertion positions, since nucleus targeting sequences are recognized independently from the folded state of the protein and the location of these signals. In summary, our results indicate that for delivering recombinant proteins to the chloroplasts and mitochondria from the TEV-based vector, proteins have to be inserted in the amino terminus of the viral polyprotein but for nuclear delivery, both insertion positions are suitable. Based on these findings, we were able to target different fluorescent proteins into chloroplasts and the nuclei from the same viral vector. We also studied the effect of the insertion site and the signal sequence on the accumulation of targeted heterologous proteins. Our results showed that the type of signal peptide and the insertion position both affect recombinant protein accumulation. In general we observed that the fusion of signal peptides can negatively alter protein accumulation, but successful subcellular

targeting of heterologous proteins from the viral vector can result in higher accumulation, which corroborate previous studies (Fischer *et al.*, 2004). Our results also showed that the position of extra genes in the potyvirus genome affects the stability of recombinant viruses and amino-terminal insertion position seems less stable than the internal site. Furthermore, we observed milder symptoms of virus constructs with amino-terminal expression of heterologous proteins which could be beneficial from a biotechnological point of view.

Based on the promising previous results, inducing anthocyanin accumulation in infected tissues with the viral vector (Bedoya *et al.*, 2010), we wanted to go further and demonstrate that the potyvirus-based vector is able not just to express transcription factors which activate several genes, but also an entire heterologous biosynthetic pathway in plant cells. For this aim we cloned the cDNAs coding for three enzymes from bacterial origin into the potyvirus-based vector which are responsible for the conversion of lycopene, a type of carotenoid, in a three-step biosynthetic pathway. In plants, carotenoids, a group of isoprenoids, are biosynthesized in plastids through the MEP pathway from the universal isoprenoid precursors IPP and DMAPP, but these molecules are also synthesized in the cytosol through the MVA pathway where they are mainly used for the biosynthesis of sterols (Rodríguez-Concepción and Boronat, 2002). Much effort have been made on engineer carotenoid metabolic pathway in plants and maximize the accumulation of β -carotene and lycopene, because of their important role in human health and diet (Giuliano, 2014). Nevertheless, engineering high level carotenoid content has proven to be extremely challenging because of the existing multilevel regulator constrains of the pathway and the time-consuming and resource-intensive creation of stable transgenic lines (Cazzonelli and Pogson, 2010; Ruiz-Sola and Rodríguez-Concepción, 2012; Nisar *et al.*, 2015). By expressing the three enzyme pathway in the cytosol with the viral vector, we expected to redirect carbon flow and produce lycopene outside the plastids where this compound is not synthesized naturally. With this experimental setup, we also expected to gain interesting insights into carotenoid biosynthetic pathway regulation. Carotenoids are present in green leaves, fruits and flowers, but they are quantitatively and qualitatively different in these tissue types. In green tissues, yellow colored xanthophylls associated to photosynthetic functions are predominant, but their color is masked by green chlorophylls. Carotenoids in non-photosynthetic tissues are visual and act as colorants calling the attention of

pollinating and seed-dispersing animals (Glover, 2011). Lycopene accumulates mostly in fruits, but it is also synthesized in green tissues where it is rapidly converted to other types of carotenoids, mainly xanthophylls with photosynthetic functions.

The recombinant TEV clone expressing bacterial enzymes was constructed by inserting the cDNA of these genes in place of the N1b cistron, and flanking them with artificial processing sites of the N1aPro. Ten dpi symptomatic leaves exhibited reddish coloration suggesting lycopene accumulation which was confirmed by HPLC. The analysis revealed the presence of colorless phytoene and high levels of orange lycopene which comprised approximately 10% of the total carotenoid content in leaves, never described before. The light microscopic examination of the infected tissue revealed that the accumulated lycopene was organized in extraplasmidic structures. This was not surprising since these compounds are highly hydrophobic and tend to form crystals when accumulated in high amounts (Li *et al.*, 2001; Maass *et al.*, 2009; Cao *et al.*, 2012; Yuan *et al.*, 2015), and the increase in the carotenoid content usually induces the proliferation of carotenoid sequestering structures (Rabbani *et al.*, 1998). According to our hypothesis, the newly engineered pathway expressed with the viral vector resulted in the production of lycopene outside the plastids where this compound can be accumulated since no carotenoid biosynthetic enzymes are present to use it as substrate. Nevertheless, necrosis of lycopene expressing tissue was observed which probably indicates that lycopene accumulation outside plastids is not stable. Further studies are needed to engineer storage of accumulated lycopene, which could be accomplished by inducing carotenoid-sequestering organelle production, formerly used to store these compounds in cauliflower (Li *et al.*, 2001; Paolillo *et al.*, 2004).

To gain further insight into the virally expressed heterologous pathway, we expressed the three carotenoid biosynthetic enzymes separately with the TEV-based vector. Interestingly, the tissue infected with the recombinant virus expressing the bacterial phytoene synthase, crtB (TEV Δ N1b-B) exhibited a bright yellow coloration, which was unexpected since phytoene, the direct product of crtB, is colorless. The phytoene synthase was identified as the most important regulatory step in the pathway and crtB was used to re-engineer the pathway in many important crops (Shewmaker *et al.*, 1999; Fraser *et al.*, 2002; Ducreux *et al.*, 2005; Fujisawa *et al.*, 2008). HPLC analysis of infected leaves revealed that virally expressed crtB stimulated the accumulation of chloroplastic carotenoids in particular β -carotene and lutein which was

accompanied with the concomitant reduction of chlorophylls, as the result of viral infection.

The yellow coloration of the infected tissue incited us to study crtB, as a reporter system to visually track viral infection. Previously tested red virus marker, Rosea1, resulted to be an excellent plant virus visualizing system especially for experiments involving massive inoculation of plants. Rosea1 enabled the easy detection and monitoring of viral infection, without the need of any specific devices in a non-destructive manner (Majer *et al.*, 2013; Bedoya *et al.*, 2012). Additionally, Rosea1 permitted the quantification of viral load in infected tissue with an easy and cheap anthocyanin extraction method and a colorimetric measurement of pigment content (Bedoya *et al.*, 2012). CrtB was able to report infection of several plant viruses, like TEV, TMV, PVX in *N. benthamiana*, in *A. thaliana* and tomato. Furthermore crtB resulted to be a more general marker than Rosea1, since it was functional in plant species where Rosea1 failed to detect virus presence, like zucchini. CrtB operates in a different biosynthetic pathway, therefore complements Rosea1, enabling the plant virus detection in species where the anthocyanin biosynthetic pathway is deficient or inexistent. In addition, the co-infection of TEV tagged with crtB and Rosea1, demonstrated the capability of these visual markers to simultaneously monitor two different plant virus populations.

Definitively, this work reflects the versatility of plant viruses as biotechnological tools. Here we explored foreign protein insertion sites in potyvirus genome, new possibilities to delivery heterologous proteins to specific subcellular compartments from the viral backbone, and vector capacity to manipulate plant metabolism resulting in the accumulation of secondary metabolites with dietary and medical importance.

Conclusions

1. The relocation of the NIb cistron to all intercistronic positions of the TEV genome demonstrated the complexity of potyviral genome organization suggesting the existence of regulatory elements unknown yet. Our results showed that even though NIb can be provided in *trans*, it can only be relocated to the amino terminus of the polyprotein or in between P1 and HC-Pro without affecting virus viability. The insertion of NIb into any other intercistronic positions results in nonviable viruses most probably because of introduction of lethal effects in polyprotein processing, disruption of partially processed gene products with distinctive roles during infection, or disruption of regulatory RNA elements in the genome.
2. Our results also demonstrated the existence of a novel insertion position at the outermost amino-terminal end of the polyprotein where sequences coding heterologous proteins can be inserted.
3. Targeting heterologous proteins to chloroplasts, nuclei and mitochondria can only be reached from the amino-terminal insertion position of the potyvirus-based vector but for nucleus targeting, both the amino-terminal and internal insertion positions are suitable. This difference most probably reflects the necessary coordination between protein synthesis and translocation to the chloroplast or mitochondria. Based on these observations two different subcellular localizations –chloroplasts and nuclei– could be reached from a single viral backbone.
4. Accumulation of targeted recombinant proteins depends on the signal peptide and the insertion position in the viral polyprotein. GFP variants expressed from the amino-terminal positions were accumulated in higher amounts, in contrast to all tagged GFP variants expressed from the internal site.
5. The position of the extra genes in the potyvirus genome affects the stability of recombinant viruses and, according to our results, cDNAs inserted into the internal position are more stable than those inserted at the amino-terminal site.
6. The reconstruction of an entire heterologous multistep biosynthetic pathway was achieved with the potyvirus-based vector. The heterologous pathway composed of three bacterial enzymes successfully rewired metabolic circuits to synthesize carotenoids from cytosolic isoprenoid precursor. We also demonstrated that the bottleneck of this pathway is the phytoene synthase, crtB, which is enough to induce carotenoid accumulation in tobacco leaves to visual levels.

7. We exploited the possibilities of crtB as visual marker to track virus infection in plants. CrtB proved to be a more general marker than the previously described Roseal for visual tracking and succeeded in reporting viral presence in various plants of both scientific and agronomic importance.

Bibliography

- Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel WJ, Verstappen FW, Verhoeven HA, Jongma MA, Schwab W, Bouwmeester HJ. 2003. Terpenoid metabolism in wild-type and transgenic *Arabidopsis* plants. *Plant Cell*. 15(12): 2866-2884.
- Ahlquist P, Janda M. 1984. cDNA cloning and in vitro transcription of the complete brome mosaic virus genome. *Mol Cell Biol*. 4(12): 2876-2882.
- Ahlquist P, French R, Janda M, Loesch-Fries LS. 1984. Multicomponent RNA plant virus infection derived from cloned viral cDNA. *Proc Natl Acad Sci USA*. 81(22): 7066-7070.
- Aluru MR, Rodermeel SR, Reddy MB. 2011. Genetic modification of low phytic acid 1-1 maize to enhance iron content and bioavailability. *J Agric Food Chem*. 59(24): 12954-12962.
- Arazi T, Slutsky SG, Shibolet Y, Wang Y, Rubinstein M, Barak S, Yang J, Gal-On A. 2001. Engineering zucchini yellow mosaic potyvirus as a non-pathogenic vector for expression of heterologous proteins in cucurbits. *J Biotechnol*. 87(1): 67-82.
- Avesani L, Marconi G, Morandini F, Albertini E, Bruschetta M, Bortesi L, Pezzotti M, Porceddu A. 2007. Stability of *Potato virus X* expression vectors is related to insert size: implications for replication models and risk assessment. *Transgenic Res*. 16(5): 587-597.
- Bachmann MF, Jennings GT. 2010. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat Rev Immunol*. 10(11): 787-796.
- Bai L, Kim EH, DellaPenna D, Brutnell TP. 2009. Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. *Plant J*. 59(4): 588-599.
- Balsera M, Soll J, Bolter B. 2009. Protein import machineries in endosymbiotic organelles. *Cell Mol Life Sci*. 66(11-12): 1903-1923.

- Baulcombe DC, Chapman S, Santa Cruz S. 1995. Jellyfish green fluorescent protein as a reporter for virus infections. *Plant J.* 7(6): 1045-1053.
- Beauchemin C, Bougie V, Laliberté JF. 2005. Simultaneous production of two foreign proteins from a polyvirus-based vector. *Virus Res.* 112(1-2): 1-8.
- Bedoya L, Martínez F, Rubio L, Daròs JA. 2010. Simultaneous equimolar expression of multiple proteins in plants from a disarmed potyvirus vector. *J Biotechnol.* 150(2): 268-275.
- Bedoya LC, Daròs JA. 2010. Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* 149(2): 234-240.
- Bedoya LC, Martínez F, Orzáez D, Daròs JA. 2012. Visual tracking of plant virus infection and movement using a reporter MYB transcription factor that activates anthocyanin biosynthesis. *Plant Physiol.* 158(3): 1130-1138.
- Belshaw R, Pybus OG, Rambaut A. 2007. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* 17(10): 1496-1504.
- Beyer P. 2010. Golden Rice and 'Golden' crops for human nutrition. *N Biotechnol.* 27(5): 478-481.
- Busch M, Seuter A, Hain R. 2002. Functional analysis of the early steps of carotenoid biosynthesis in tobacco. *Plant Physiol.* 128(2): 439-453.
- Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J, Martin C. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat Biotechnol.* 26(11): 1301-1308.
- Cañizares MC, Liu L, Perrin Y, Tsakiris E, Lomonossoff GP. 2006. A bipartite system for the constitutive and inducible expression of high levels of foreign proteins in plants. *Plant Biotechnol J.* 4(2): 183-193.

- Cao H, Zhang J, Xu J, Ye J, Yun Z, Xu Q, Xu J, Deng X. 2012. Comprehending crystalline beta-carotene accumulation by comparing engineered cell models and the natural carotenoid-rich system of citrus. *J Exp Bot.* 63(12): 4403-4417.
- Caplan JL, Mamillapalli P, Burch-Smith TM, Czymmek K, Dinesh-Kumar SP. 2008. Chloroplastic protein NRIP1 mediates innate immune receptor recognition of a viral effector. *Cell.* 132(3): 449-462.
- Carrington JC, Cary SM, Parks TD, Dougherty WG. 1989a. A second proteinase encoded by a plant potyvirus genome. *EMBO J.* 8(2): 365-370.
- Carrington JC, Freed DD, Sanders TC. 1989b. Autocatalytic processing of the potyvirus helper component proteinase in *Escherichia coli* and in vitro. *J Virol.* 63(10): 4459-4463.
- Carrington JC, Freed DD, Oh CS. 1990. Expression of potyviral polyproteins in transgenic plants reveals three proteolytic activities required for complete processing. *EMBO J.* 9(5): 1347-1353.
- Carrington JC, Haldeman R, Dolja VV, Restrepo-Hartwig MA. 1993. Internal cleavage and *trans*-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus in vivo. *J Virol.* 67(12): 6995-7000.
- Cazzonelli CI, Pogson BJ. 2010. Source to sink: regulation of carotenoid biosynthesis in plants. *Trends Plant Sci.* 15(5): 266-274.
- Chapman S, Kavanagh T, Baulcombe D. 1992. Potato virus X as a vector for gene expression in plants. *Plant J.* 2(4): 549-557.
- Chen CC, Chen TC, Raja JA, Chang CA, Chen LW, Lin SS, Yeh SD. 2007. Effectiveness and stability of heterologous proteins expressed in plants by *Turnip mosaic virus* vector at five different insertion sites. *Virus Res.* 130(1-2): 210-227.
- Chen H, Lin Y. 2013. Promise and issues of genetically modified crops. *Curr Opin Plant Biol.* 16(2): 255-260.

Chen Q, He J, Phoolcharoen W, Mason HS. 2011. Geminiviral vectors based on bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in plants. *Hum Vaccin*. 7(3): 331-338.

Chilton MD, Drummond MH, Merio DJ, Sciaky D, Montoya AL, Gordon MP, Nester EW. 1977. Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell*. 11(2): 263-271.

Chirico N, Vianelli A, Belshaw R. 2010. Why genes overlap in viruses. *Proc Biol Sci*. 277(1701): 3809-3817.

Choi IR, Stenger DC, Morris TJ, French R. 2000. A plant virus vector for systemic expression of foreign genes in cereals. *Plant J*. 23(4): 547-555.

Chung BYW, Miller WA, Atkins JF, Firth AE. 2008. An overlapping essential gene in the Potyviridae. *Proc Natl Acad Sci USA*. 105(15): 5897-5902.

Ciccone MM, Cortese F, Gesualdo M, Carbonara S, Zito A, Ricci G, De Pascalis F, Scicchitano P, Riccioni G. 2013. Dietary intake of carotenoids and their antioxidant and anti-inflammatory effects in cardiovascular care. *Mediators Inflamm*. 2013(782137).

Cressey D. 2013. Transgenics: A new breed. *Nature*. 497(7447): 27-29.

Cunningham FX, Jr., Chamovitz D, Misawa N, Gantt E, Hirschberg J. 1993. Cloning and functional expression in *Escherichia coli* of a cyanobacterial gene for lycopene cyclase, the enzyme that catalyzes the biosynthesis of beta-carotene. *FEBS Lett*. 328(1-2): 130-138.

D'Ambrosio C, Giorgio G, Marino I, Merendino A, Petrozza A, Salfi L, Stigliani A, Cellini F. 2003. Virtually complete conversion of lycopene into beta-carotene in fruits of tomato plants transformed with the tomato lycopene beta-cyclase (tlyc-b) cDNA. *Plant Sci*. 166: 207-214.

- D'Andrea L, Amenos M, Rodriguez-Concepcion M. 2014. Confocal laser scanning microscopy detection of chlorophylls and carotenoids in chloroplasts and chromoplasts of tomato fruit. *Methods Mol Biol.* 1153: 227-232.
- Daròs JA, Schaad MC, Carrington JC. 1999. Functional analysis of the interaction between VPg-proteinase (NIa) and RNA polymerase (NIb) of tobacco etch potyvirus, using conditional and suppressor mutants. *J Virol.* 73(10): 8732-8740.
- Day RN, Davidson MW. 2009. The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev.* 38(10): 2887-2921.
- Dayan FE, Owens DK, Tranel PJ, Preston C, Duke SO. 2014. Evolution of resistance to phytoene desaturase and protoporphyrinogen oxidase inhibitors--state of knowledge. *Pest Manag Sci.* 70(9): 1358-1366.
- DellaPenna D, Pogson BJ. 2006. Vitamin synthesis in plants: tocopherols and carotenoids. *Annu Rev Plant Biol.* 57: 711-738.
- Di Fiore S, Li Q, Leech MJ, Schuster F, Emans N, Fischer R, Schillberg S. 2002. Targeting tryptophan decarboxylase to selected subcellular compartments of tobacco plants affects enzyme stability and in vivo function and leads to a lesion-mimic phenotype. *Plant Physiol.* 129(3): 1160-1169.
- Dickmeis C, Fischer R, Commandeur U. 2014. *Potato virus X*-based expression vectors are stabilized for long-term production of proteins and larger inserts. *Biotechnol J.* 9(11): 1369-1379.
- Dietrich C, Maiss E. 2003. Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected *Nicotiana benthamiana* plants. *J Gen Virol.* 84: 2871-2876.
- Diretto G, Tavazza R, Welsch R, Pizzichini D, Mourgues F, Papacchioli V, Beyer P, Giuliano G. 2006. Metabolic engineering of potato tuber carotenoids through tuber-specific silencing of lycopene epsilon cyclase. *BMC Plant Biol.* 6: 13.

- Diretto G, Al-Babili S, Tavazza R, Papacchioli V, Beyer P, Giuliano G. 2007. Metabolic engineering of potato carotenoid content through tuber-specific overexpression of a bacterial mini-pathway. *PLoS One*. 2(4): e350.
- Dixon RA, Liu C, Jun JH. 2013. Metabolic engineering of anthocyanins and condensed tannins in plants. *Curr Opin Biotechnol*. 24(2): 329-335.
- Dolja VV, McBride HJ, Carrington JC. 1992. Tagging of plant potyvirus replication and movement by insertion of β -glucuronidase into the viral polyprotein. *Proc Natl Acad Sci USA*. 89(21): 10208-10212.
- Dolja VV, Hong J, Keller KE, Martin RR, Peremyslov VV. 1997. Suppression of Potyvirus Infection by Coexpressed Closterovirus Protein. *Virology*. 234(2): 243-252.
- Domonkos I, Kis M, Gombos Z, Ughy B. 2013. Carotenoids, versatile components of oxygenic photosynthesis. *Prog Lipid Res*. 52(4): 539-561.
- Ducreux LJ, Morris WL, Hedley PE, Shepherd T, Davies HV, Millam S, Taylor MA. 2005. Metabolic engineering of high carotenoid potato tubers containing enhanced levels of beta-carotene and lutein. *J Exp Bot*. 56(409): 81-89.
- Engler C, Gruetzner R, Kandzia R, Marillonnet S. 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One*. 4(5): e5553.
- Etminan M, Takkouche B, Caamano-Isorna F. 2004. The role of tomato products and lycopene in the prevention of prostate cancer: a meta-analysis of observational studies. *Cancer Epidemiol Biomarkers Prev*. 13(3): 340-345.
- Evans DJ. 2008. The bionanoscience of plant viruses: templates and synthons for new materials. *J Mater Chem*. 18: 3746-3754.
- Farhi M, Marhevka E, Ben-Ari J, Algamas-Dimantov A, Liang Z, Zeevi V, Edelbaum O, Spitzer-Rimon B, Abeliovich H, Schwartz B, Tzfira T, Vainstein A. 2011.

Generation of the potent anti-malarial drug artemisinin in tobacco. *Nat Biotechnol.* 29(12): 1072-1074.

Farré G, Blancquaert D, Capell T, Van Der Straeten D, Christou P, Zhu C. 2014. Engineering complex metabolic pathways in plants. *Annu Rev Plant Biol.* 65: 187-223.

Faye L, Gomord V. 2010. Success stories in molecular farming-a brief overview. *Plant Biotechnol J.* 8(5): 525-528.

Fernández-Fernández MR, Mouriño M, Rivera J, Rodríguez F, Plana-Durán J, García JA. 2001. Protection of rabbits against rabbit hemorrhagic disease virus by immunization with the VP60 protein expressed in plants with a potyvirus-based vector. *Virology.* 280(2): 283-291.

Fischer R, Stoger E, Schillberg S, Christou P, Twyman RM. 2004. Plant-based production of biopharmaceuticals. *Curr Opin Plant Biol.* 7(2): 152-158.

Fraser PD, Romer S, Shipton CA, Mills PB, Kiano JW, Misawa N, Drake RG, Schuch W, Bramley PM. 2002. Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner. *Proc Natl Acad Sci USA.* 99(2): 1092-1097.

Fraser PD, Bramley PM. 2004. The biosynthesis and nutritional uses of carotenoids. *Prog Lipid Res.* 43(3): 228-265.

Fraser PD, Enfissi EM, Bramley PM. 2009. Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. *Arch Biochem Biophys.* 483(2): 196-204.

Fray R, Wallace A, Fraser P, Valero D, Hedden P, Bramley P, Grierson D. 1995. Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. *Plant J.* 8(5): 693-701.

Fray RG, Grierson D. 1993. Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant Mol Biol.* 22(4): 589-602.

Freitas N, Cunha C. 2009. Mechanisms and signals for the nuclear import of proteins. *Curr Genomics.* 10(8): 550-557.

Fujisawa M, Watanabe M, Choi SK, Teramoto M, Ohyama K, Misawa N. 2008. Enrichment of carotenoids in flaxseed (*Linum usitatissimum*) by metabolic engineering with introduction of bacterial phytoene synthase gene *crtB*. *J Biosci Bioeng.* 105(6): 636-641.

Fujisawa M, Takita E, Harada H, Sakurai N, Suzuki H, Ohyama K, Shibata D, Misawa N. 2009. Pathway engineering of *Brassica napus* seeds using multiple key enzyme genes involved in ketocarotenoid formation. *J Exp Bot.* 60(4): 1319-1332.

Geisler K, Hughes RK, Sainsbury F, Lomonossoff GP, Rejzek M, Fairhurst S, Olsen CE, Motawia MS, Melton RE, Hemmings AM, Bak S, Osbourn A. 2013. Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc Natl Acad Sci USA.* 110(35): E3360-3367.

German-Retana S, Candresse T, Alias E, Delbos RP, Le Gall O. 2000. Effects of green fluorescent protein or beta-glucuronidase tagging on the accumulation and pathogenicity of a resistance-breaking *Lettuce mosaic virus* isolate in susceptible and resistant lettuce cultivars. *Mol Plant Microbe Interact.* 13(3): 316-324.

Gibbs A, Ohshima K. 2010. Potyviruses and the digital revolution. *Annu Rev Phytopathol.* 48: 205-223.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods.* 6(5): 343-345.

- Gil-Humanes J, Piston F, Tollefsen S, Sollid LM, Barro F. 2010. Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. *Proc Natl Acad Sci USA*. 107(39): 17023-17028.
- Gils M, Kandzia R, Marillonnet S, Klimyuk V, Gleba Y. 2005. High-yield production of authentic human growth hormone using a plant virus-based expression system. *Plant Biotechnol J*. 3(6): 613-620.
- Giorio G, Stigliani AL, D'Ambrosio C. 2008. Phytoene synthase genes in tomato (*Solanum lycopersicum* L.) - new data on the structures, the deduced amino acid sequences and the expression patterns. *FEBS J*. 275(3): 527-535.
- Giovannucci E. 2002. A review of epidemiologic studies of tomatoes, lycopene, and prostate cancer. *Exp Biol Med (Maywood)*. 227(10): 852-859.
- Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, Gleba Y. 2006. Rapid high-yield expression of full-size IgG antibodies in plants coinfecting with noncompeting viral vectors. *Proc Natl Acad Sci USA*. 103(40): 14701-14706.
- Giuliano G, Tavazza R, Diretto G, Beyer P, Taylor MA. 2008. Metabolic engineering of carotenoid biosynthesis in plants. *Trends Biotechnol*. 26(3): 139-145.
- Giuliano G. 2014. Plant carotenoids: genomics meets multi-gene engineering. *Curr Opin Plant Biol*. 19: 111-117.
- Gleba Y, Marillonnet S, Klimyuk V. 2004. Engineering viral expression vectors for plants: the 'full virus' and the 'deconstructed virus' strategies. *Curr Opin Plant Biol*. 7(2): 182-188.
- Gleba Y, Klimyuk V, Marillonnet S. 2005. Magniffection—a new platform for expressing recombinant vaccines in plants. *Vaccine*. 18(23): 2042–2048.
- Gleba Y, Klimyuk V, Marillonnet S. 2007. Viral vectors for the expression of proteins in plants. *Curr Opin Biotechnol*. 18(2): 134-141.

Gleba YY, Tuse D, Giritch A. 2014. Plant viral vectors for delivery by *Agrobacterium*. *Curr Top Microbiol Immunol.* 375(155-192).

Glover BJ. 2011. Pollinator attraction: the importance of looking good and smelling nice. *Curr Biol.* 21(9): R307-309.

Gray MW. 1992. The endosymbiont hypothesis revisited. *Int Rev Cytol.* 141: 233-357.

Grimsley N, Hohn B, Hohn T, Walden R. 1986. "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. *Proc Natl Acad Sci USA.* 83(10): 3282-3286.

Guo HS, López-Moya JJ, García JA. 1998. Susceptibility to recombination rearrangements of a chimeric plum pox potyvirus genome after insertion of a foreign gene. *Virus Res.* 57(2): 183-195.

Gurib-Fakim A. 2006. Medicinal plants: traditions of yesterday and drugs of tomorrow. *Mol Aspects Med.* 27(1): 1-93.

Haldeman-Cahill R, Daròs JA, Carrington JC. 1998. Secondary structures in the capsid protein coding sequence and 3' nontranslated region involved in amplification of the tobacco etch virus genome. *J Virol.* 72(5): 4072-4079.

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES. 2008. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science.* 319(5861): 330-333.

Hasunuma T, Miyazawa S, Yoshimura S, Shinzaki Y, Tomizawa K, Shindo K, Choi SK, Misawa N, Miyake C. 2008. Biosynthesis of astaxanthin in tobacco leaves by transplastomic engineering. *Plant J.* 55(5): 857-868.

Heinig U, Gutensohn M, Dudareva N, Aharoni A. 2013. The challenges of cellular compartmentalization in plant metabolic engineering. *Curr Opin Biotechnol.* 24(2): 239-246.

Herrera-Estrella L, Block MD, Messens E, Hernalsteens JP, Montagu MV, Schell J. 1983. Chimeric genes as dominant selectable markers in plant cells. *EMBO J.* 2(6): 987-995.

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V. 2011. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J Exp Bot.* 62(8): 2465-2483.

Hirschberg J. 2001. Carotenoid biosynthesis in flowering plants. *Curr Opin Plant Biol.* 4(3): 210-218.

Huang Z, Santi L, LePore K, Kilbourne J, Arntzen CJ, Mason HS. 2006. Rapid, high-level production of hepatitis B core antigen in plant leaf and its immunogenicity in mice. *Vaccine.* 24(14): 2506-2513.

Huang Z, Chen Q, Hjelm B, Arntzen C, Mason H. 2009. A DNA replicon system for rapid high-level production of virus-like particles in plants. *Biotechnol Bioeng.* 103(4): 706-714.

Huang Z, Phoolcharoen W, Lai H, Piensook K, Cardineau G, Zeitlin L, Whaley KJ, Arntzen CJ, Mason HS, Chen Q. 2010. High-level rapid production of full-size monoclonal antibodies in plants by a single-vector DNA replicon system. *Biotechnol Bioeng.* 106(1): 9-17.

Hull R. 1978. The possible use of plant viral DNAs in genetic manipulation in plants. *Trends Biochem Sci.* 3(4): 254-256.

Hurt EC, Müller U, Schatz G. 1985. The first twelve amino acids of a yeast mitochondrial outer membrane protein can direct a nuclear-coded cytochrome oxidase subunit to the mitochondrial inner membrane. *EMBO J.* 4(13A): 3509-3518.

Ivanov KI, Eskelin K, Lohmus A, Makinen K. 2014. Molecular and cellular mechanisms underlying potyvirus infection. *J Gen Virol.* 95(Pt 7): 1415-1429.

Jarvis P, López-Juez E. 2013. Biogenesis and homeostasis of chloroplasts and other plastids. *Nat Rev Mol Cell Biol.* 14(12): 787-802.

Jefferson RA, Kavanagh TA, Bevan MW. 1987. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6(13): 3901-3907.

Juárez P, Presa S, Espí J, Pineda B, Antón MT, Moreno V, Buesa J, Granell A, Orzaez D. 2012. Neutralizing antibodies against rotavirus produced in transgenically labelled purple tomatoes. *Plant Biotechnol J.* 10(3): 341-352.

Kaiser CR, Flenniken ML, Gillitzer E, Harmsen AL, Harmsen AG, Jutila MA, Douglas T, Young MJ. 2007. Biodistribution studies of protein cage nanoparticles demonstrate broad tissue distribution and rapid clearance in vivo. *Int J Nanomedicine.* 2(4): 715-733.

Kalderon D, Roberts BL, Richardson WD, Smith AE. 1984. A short amino acid sequence able to specify nuclear location. *Cell.* 39(3 Pt. 2): 499-509.

Kanagarajan S, Muthusamy S, Gliszczynska A, Lundgren A, Brodelius PE. 2012. Functional expression and characterization of sesquiterpene synthases from *Artemisia annua* L. using transient expression system in *Nicotiana benthamiana*. *Plant Cell Rep.* 31(7): 1309-1319.

Kelloniemi J, Mäkinen K, Valkonen JP. 2008. Three heterologous proteins simultaneously expressed from a chimeric potyvirus: infectivity, stability and the correlation of genome and virion lengths. *Virus Res.* 135(2): 282-291.

Kim DH, Park YS, Kim SS, Lew J, Nam HG, Choi KY. 1995. Expression, purification, and identification of a novel self-cleavage site of the N1a C-terminal 27-kDa protease of turnip mosaic potyvirus C5. *Virology.* 213(2): 517-525.

Köhler RH, Zipfel WR, Webb WW, Hanson MR. 1997. The green fluorescent protein as a marker to visualize plant mitochondria in vivo. *Plant J.* 11(2): 613-621.

- Komarova TV, Skulachev MV, Zvereva AS, Schwartz AM, Dorokhov YL, Atabekov JG. 2006. New viral vector for efficient production of target proteins in plants. *Biochemistry (Mosc)*. 71(8): 846-850.
- Kumagai MH, Donson J, della-Cioppa G, Harvey D, Hanley K, Grill LK. 1995. Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proc Natl Acad Sci USA*. 92(5): 1679-1683.
- Kumagai MH, Keller Y, Bouvier F, Clary D, Camara B. 1998. Functional integration of non-native carotenoids into chloroplasts by viral-derived expression of capsanthin-capsorubin synthase in *Nicotiana benthamiana*. *Plant J*. 14(3): 305-315.
- Landrum JT, Bone RA. 2001. Lutein, zeaxanthin, and the macular pigment. *Arch Biochem Biophys*. 385(1): 28-40.
- Lassner MW, Jones A, Daubert S, Comai L. 1991. Targeting of T7 RNA polymerase to tobacco nuclei mediated by an SV40 nuclear location signal. *Plant Mol Biol*. 17(2): 229-234.
- Lau W, Fischbach MA, Osbourn A, Sattely ES. 2014. Key applications of plant metabolic engineering. *PLoS Biol*. 12(6): e1001879.
- Li C, Jiang Y, Guo W, Liu Z. 2013. Production of a chimeric allergen derived from the major allergen group 1 of house dust mite species in *Nicotiana benthamiana*. *Hum Immunol*. 74(5): 531-537.
- Li F, Vallabhaneni R, Wurtzel ET. 2008. PSY3, a new member of the phytoene synthase gene family conserved in the Poaceae and regulator of abiotic stress-induced root carotenogenesis. *Plant Physiol*. 146(3): 1333-1345.
- Li L, Paolillo DJ, Parthasarathy MV, Dimuzio EM, Garvin DF. 2001. A novel gene mutation that confers abnormal patterns of beta-carotene accumulation in cauliflower (*Brassica oleracea* var. *botrytis*). *Plant J*. 26(1): 59-67.

- Li L, Yuan H. 2013. Chromoplast biogenesis and carotenoid accumulation. *Arch Biochem Biophys.* 539(2): 102-109.
- Li XH, Carrington JC. 1995. Complementation of tobacco etch potyvirus mutants by active RNA polymerase expressed in transgenic cells. *Proc Natl Acad Sci USA.* 92(2): 457-461.
- Li Y, Cui H, Cui X, Wang A. 2015. The altered photosynthetic machinery during compatible virus infection. *Curr Opin Virol.* 17(19-24).
- Lico C, Benvenuto E, Baschieri S. 2015. The two-faced Potato virus X: from plant pathogen to smart nanoparticle. *Front Plant Sci.* 6: 1009.
- Lindbo JA. 2007. TRBO: a high-efficiency tobacco mosaic virus RNA-based overexpression vector. *Plant Physiol.* 145(4): 1232-1240.
- Liu L, van Tonder T, Pietersen G, Davies JW, Stanley J. 1997. Molecular characterization of a subgroup I geminivirus from a legume in South Africa. *J Gen Virol.* 78 (Pt 8): 2113-2117.
- Liu L, Lomonosoff G. 2002. Agroinfection as a rapid method for propagating Cowpea mosaic virus-based constructs. *J Virol Methods.* 105(2): 343-348.
- Liu L, Grainger J, Canizares MC, Angell SM, Lomonosoff GP. 2004. Cowpea mosaic virus RNA-1 acts as an amplicon whose effects can be counteracted by a RNA-2-encoded suppressor of silencing. *Virology.* 323(1): 37-48.
- Liu Y, Schiff M, Dinesh-Kumar SP. 2002. Virus-induced gene silencing in tomato. *Plant J.* 31(6): 777-786.
- Lizotte PH, Wen AM, Sheen MR, Fields J, Rojanasopondist P, Steinmetz NF, Fiering S. 2016. In situ vaccination with cowpea mosaic virus nanoparticles suppresses metastatic cancer. *Nat Nanotechnol.* 11(3): 295-303.

- Llorente B, D'Andrea L, Ruiz-Sola MA, Botterweg E, Pulido P, Andilla J, Loza-Alvarez P, Rodriguez-Concepcion M. 2016. Tomato fruit carotenoid biosynthesis is adjusted to actual ripening progression by a light-dependent mechanism. *Plant J.* 85(1): 107-119.
- Lopez-Juez E, Pyke KA. 2005. Plastids unleashed: their development and their integration in plant development. *Int J Dev Biol.* 49(5-6): 557-577.
- Lu R, Malcuit I, Moffett P, Ruiz MT, Peart J, Wu AJ, Rathjen JP, Bendahmane A, Day L, Baulcombe DC. 2003. High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. *EMBO J.* 22(21): 5690-5699.
- Lu Y, Rijzaani H, Karcher D, Ruf S, Bock R. 2013. Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc Natl Acad Sci USA.* 110(8): E623-632.
- Ma JK, Drake PM, Christou P. 2003. The production of recombinant pharmaceutical proteins in plants. *Nat Rev Genet.* 4(10): 794-805.
- Maass D, Arango J, Wust F, Beyer P, Welsch R. 2009. Carotenoid crystal formation in Arabidopsis and carrot roots caused by increased phytoene synthase protein levels. *PLoS One.* 4(7): e6373.
- Mackenzie S, McIntosh L. 1999. Higher plant mitochondria. *Plant Cell.* 11(4): 571-586.
- Majer E, Daròs JA, Zwart MP. 2013. Stability and fitness impact of the visually discernible Roseal marker in the Tobacco etch virus genome. *Viruses.* 5(9): 2153-2168.
- Majer E, Salvador Z, Willemsen A, Elena SF, Daròs J-A. 2014. Relocation of the NIB gene in the tobacco etch potyvirus genome. *J Virol.* 88(8): 4586-4590.
- Mann V, Harker M, Pecker I, Hirschberg J. 2000. Metabolic engineering of astaxanthin production in tobacco flowers. *Nat Biotechnol.* 18(8): 888-892.

- Mardanova ES, Kotyarov RY, Ravin NV. 2009. The optimization of viral vector translation improves the production of recombinant proteins in plants. *Mol Biol (Mosk)*. 43(3): 524-527.
- Marillonnet S, Giritch A, Gils M, Kandzia R, Klimyuk V, Gleba Y. 2004. In planta engineering of viral RNA replicons: efficient assembly by recombination of DNA modules delivered by *Agrobacterium*. *Proc Natl Acad Sci USA*. 101(18): 6852-6857.
- Marillonnet S, Thoeringer C, Kandzia R, Klimyuk V, Gleba Y. 2005. Systemic *Agrobacterium tumefaciens*-mediated transfection of viral replicons for efficient transient expression in plants. *Nat Biotechnol*. 23(6): 718-723.
- Martínez F, Daròs JA. 2014. *Tobacco etch* virus protein P1 traffics to the nucleolus and associates with the host 60S ribosomal subunits during infection. *J Virol*. 88(18): 10725-10737.
- Mason HS, Warzecha H, Mor T, Arntzen CJ. 2002. Edible plant vaccines: applications for prophylactic and therapeutic molecular medicine. *Trends Mol Med*. 8(7): 324-329.
- Masuta C, Yamana T, Tacahashi Y, Uyeda I, Sato M, Ueda S, Matsumura T. 2000. Development of clover yellow vein virus as an efficient, stable gene-expression system for legume species. *Plant J*. 23(4): 539-546.
- Matz MV, Lukyanov KA, Lukyanov SA. 2002. Family of the green fluorescent protein: journey to the end of the rainbow. *Bioessays*. 24(10): 953-959.
- May T, Soll J. 2000. 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. *Plant Cell*. 12(1): 53-64.
- Mazur BJ, Chui CF. 1985. Sequence of a genomic DNA clone for the small subunit of ribulose bis-phosphate carboxylase-oxygenase from tobacco. *Nucleic Acids Res*. 13(7): 2373-2386.

- Meléndez-Martínez AJ, Mapelli-Brahm P, Benítez-González A, Stinco CM. 2015. A comprehensive review on the colorless carotenoids phytoene and phytofluene. *Arch Biochem Biophys.* 572: 188-200.
- Melnik S, Stoger E. 2013. Green factories for biopharmaceuticals. *Curr Med Chem.* 20(8): 1038-1046.
- Merits A, Rajamaki ML, Lindholm P, Runeberg-Roos P, Kekarainen T, Puustinen P, Makelainen K, Valkonen JP, Saarma M. 2002. Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. *J Gen Virol.* 83(Pt 5): 1211-1221.
- Mingot A, Valli A, Rodamilans B, San León D, Baulcombe DC, García JA, López-Moya JJ. 2016. The P1N-PISPO *trans*-frame gene of sweet potato feathery mottle potyvirus is produced during virus infection and functions as an RNA silencing suppressor. *J Virol.* 90(7): 3543-3557.
- Misawa N, Nakagawa M, Kobayashi K, Yamano S, Izawa Y, Nakamura K, Harashima K. 1990. Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. *J Bacteriol.* 172(12): 6704-6712.
- Moise AR, Al-Babili S, Wurtzel ET. 2014. Mechanistic aspects of carotenoid biosynthesis. *Chem Rev.* 114(1): 164-193.
- Mor TS, Moon YS, Palmer KE, Mason HS. 2003. Geminivirus vectors for high-level expression of foreign proteins in plant cells. *Biotechnol Bioeng.* 81(4): 430-437.
- Mozes-Koch R, Gover O, Tanne E, Peretz Y, Maori E, Chernin L, Sela I. 2012. Expression of an entire bacterial operon in plants. *Plant Physiol.* 158(4): 1883-1892.
- Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, Watanabe Y, Nakamura K, Kimura T, Ishiguro S. 2007. Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci Biotechnol Biochem.* 71(8): 2095-2100.

- Naqvi S, Zhu C, Farre G, Ramessar K, Bassie L, Breitenbach J, Perez Conesa D, Ros G, Sandmann G, Capell T, Christou P. 2009. Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proc Natl Acad Sci USA*. 106(19): 7762-7767.
- Nelson RS, Citovsky V. 2005. Plant viruses. Invaders of cells and pirates of cellular pathways. *Plant Physiol*. 138(4): 1809-1814.
- Neupert W, Herrmann JM. 2007. Translocation of proteins into mitochondria. *Annu Rev Biochem*. 76: 723-749.
- Nguyen M, Haenni AL. 2003. Expression strategies of ambisense viruses. *Virus Res*. 93(2): 141-150.
- Nielsen AZ, Ziersen B, Jensen K, Lassen LM, Olsen CE, Moller BL, Jensen PE. 2013. Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth Biol*. 2(6): 308-315.
- Nisar N, Li L, Lu S, Khin NC, Pogson BJ. 2015. Carotenoid metabolism in plants. *Mol Plant*. 8(1): 68-82.
- Nogueira M, Mora L, Enfissi EM, Bramley PM, Fraser PD. 2013. Subchromoplast sequestration of carotenoids affects regulatory mechanisms in tomato lines expressing different carotenoid gene combinations. *Plant Cell*. 25(11): 4560-4579.
- Noris E, Poli A, Cojoca R, Ritta M, Cavallo F, Vaglio S, Matic S, Landolfo S. 2011. A human papillomavirus 8 E7 protein produced in plants is able to trigger the mouse immune system and delay the development of skin lesions. *Arch Virol*. 156(4): 587-595.
- O'Connor SE. 2015. Engineering of Secondary Metabolism. *Annu Rev Genet*. 49: 71-94.
- Ohara K, Ujihara T, Endo T, Sato F, Yazaki K. 2003. Limonene production in tobacco with *Perilla* limonene synthase cDNA. *J Exp Bot*. 54(393): 2635-2642.

- Oparka K, Roberts G, Santa Cruz S, Boevink P, Prior D, Smallcombe A. 1997. Using GFP to study virus invasion and spread in plant tissues. *Nature*. 388: 401-402.
- Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science*. 273(5280): 1392-1395.
- Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R. 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat Biotechnol*. 23(4): 482-487.
- Paolillo DJ, Jr., Garvin DF, Parthasarathy MV. 2004. The chromoplasts of Or mutants of cauliflower (*Brassica oleracea* L. var. botrytis). *Protoplasma*. 224(3-4): 245-253.
- Park H, Kreunen SS, Cuttriss AJ, DellaPenna D, Pogson BJ. 2002. Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. *Plant Cell*. 14(2): 321-332.
- Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG. 1995. Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology*. 210(1): 194-201.
- Peyret H, Lomonosoff GP. 2013. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol Biol*. 83(1-2): 51-58.
- Peyret H, Lomonosoff GP. 2015. When plant virology met *Agrobacterium*: the rise of the deconstructed clones. *Plant Biotechnol J*. 13(8): 1121-1135.
- Pogson BJ, Rissler HM. 2000. Genetic manipulation of carotenoid biosynthesis and photoprotection. *Philos Trans R Soc Lond B Biol Sci*. 355(1402): 1395-1403.
- Pogue GP, Lindbo JA, Garger SJ, Fitzmaurice WP. 2002. Making an ally from an enemy: plant virology and the new agriculture. *Annu Rev Phytopathol*. 40: 45-74.

- Pokorski JK, Steinmetz NF. 2011. The art of engineering viral nanoparticles. *Mol Pharm.* 8(1): 29-43.
- Potrykus I. 2003. Nutritionally enhanced rice to combat malnutrition disorders of the poor. *Nutr Rev.* 61(6 Pt 2): S101-104.
- Qbadou S, Becker T, Mirus O, Tews I, Soll J, Schleiff E. 2006. The molecular chaperone Hsp90 delivers precursor proteins to the chloroplast import receptor Toc64. *EMBO J.* 25(9): 1836-1847.
- Rabbani S, Beyer P, Lintig J, Hugueney P, Kleinig H. 1998. Induced beta-carotene synthesis driven by triacylglycerol deposition in the unicellular alga *dunaliella bardawil*. *Plant Physiol.* 116(4): 1239-1248.
- Rajamäki ML, Kelloniemi J, Alminaita A, Kekarainen T, Rabenstein F, Valkonen JP. 2005. A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. *Virology.* 342(1): 88-101.
- Ratcliff F, Martin-Hernandez AM, Baulcombe DC. 2001. Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing. *Plant J.* 25(2): 237-245.
- Restrepo MA, Freed DD, Carrington JC. 1990. Nuclear transport of plant potyviral proteins. *Plant Cell.* 2(10): 987-998.
- Riechmann JL, Laín S, García JA. 1992. Highlights and prospects of potyvirus molecular biology. *J Gen Virol.* 73: 1-16.
- Riedel D, Lesemann DE, Maiss E. 1998. Ultrastructural localization of nonstructural and coat proteins of 19 potyviruses using antisera to bacterially expressed proteins of plum pox potyvirus. *Arch Virol.* 143(11): 2133-2158.
- Robertson D. 2004. VIGS vectors for gene silencing: many targets, many tools. *Annu Rev Plant Biol.* 55: 495-519.

- Rodriguez-Concepcion M, Boronat A. 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol.* 130(3): 1079-1089.
- Römer S, Fraser PD, Kiano JW, Shipton CA, Misawa N, Schuch W, Bramley PM. 2000. Elevation of the provitamin A content of transgenic tomato plants. *Nat Biotechnol.* 18(6): 666-669.
- Roossinck MJ. 2011. The big unknown: plant virus biodiversity. *Curr Opin Virol.* 1(1): 63-67.
- Rosati C, Aquilani R, Dharmapuri S, Pallara P, Marusic C, Tavazza R, Bouvier F, Camara B, Giuliano G. 2000. Metabolic engineering of beta-carotene and lycopene content in tomato fruit. *Plant J.* 24(3): 413-419.
- Roy G, Weisburg S, Foy K, Rabindran S, Mett V, Yusibov V. 2011. Co-expression of multiple target proteins in plants from a tobacco mosaic virus vector using a combination of homologous and heterologous subgenomic promoters. *Arch Virol.* 156(11): 2057-2061.
- Ruiz-Sola MA, Rodríguez-Concepción M. 2012. Carotenoid biosynthesis in Arabidopsis: a colorful pathway. *Arabidopsis Book.* 10: e0158.
- Ruiz MT, Voinnet O, Baulcombe DC. 1998. Initiation and maintenance of virus-induced gene silencing. *Plant Cell.* 10(6): 937-946.
- Rybicki EP. 2015. A Top Ten list for economically important plant viruses. *Arch Virol.* 160(1): 17-20.
- Sainsbury F, Lavoie PO, D'Aoust MA, Vézina LP, Lomonossoff GP. 2008. Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. *Plant Biotechnol J.* 6(1): 82-92.

- Sainsbury F, Lomonossoff GP. 2008. Extremely high-level and rapid transient protein production in plants without the use of viral replication. *Plant Physiol.* 148(3): 1212-1218.
- Sainsbury F, Thuenemann EC, Lomonossoff GP. 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J.* 7(7): 682-693.
- Sainsbury F, Cañizares MC, Lomonossoff GP. 2010. *Cowpea mosaic virus*: the plant virus-based biotechnology workhorse. *Annu Rev Phytopathol.* 48: 437-455.
- Sainsbury F, Saxena P, Geisler K, Osbourn A, Lomonossoff GP. 2012. Using a virus-derived system to manipulate plant natural product biosynthetic pathways. *Methods Enzymol.* 517: 185-202.
- Sainsbury F, Lomonossoff GP. 2014. Transient expressions of synthetic biology in plants. *Curr Opin Plant Biol.* Jun(19): 1-7.
- Salazar-González JA, Bañuelos-Hernández B, Rosales-Mendoza S. 2015. Current status of viral expression systems in plants and perspectives for oral vaccines development. *Plant Mol Biol.* 87(3): 203-217.
- Sánchez F, Sáez M, Lunello P, Ponz F. 2013. Plant viral elongated nanoparticles modified for log-increases of foreign peptide immunogenicity and specific antibody detection. *J Biotechnol.* 168(4): 409-415.
- Sandmann G, Romer S, Fraser PD. 2006. Understanding carotenoid metabolism as a necessity for genetic engineering of crop plants. *Metab Eng.* 8(4): 291-302.
- Santi L, Giritch A, Roy CJ, Marillonnet S, Klimyuk V, Gleba Y, Webb R, Arntzen CJ, Mason HS. 2006. Protection conferred by recombinant *Yersinia pestis* antigens produced by a rapid and highly scalable plant expression system. *Proc Natl Acad Sci USA.* 103(4): 861-866.

- Schaad MC, Haldeman-Cahill R, Cronin S, Carrington JC. 1996. Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *J Virol.* 70(10): 7039-7048.
- Schurer H, Lang K, Schuster J, Morl M. 2002. A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic Acids Res.* 30(12): e56.
- Schwinn K, Venail J, Shang Y, Mackay S, Alm V, Butelli E, Oyama R, Bailey P, Davies K, Martin C. 2006. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell.* 18(4): 831-851.
- Sharoni Y, Agbaria R, Amir H, Ben-Dor A, Bobilev I, Doubi N, Giat Y, Hirsh K, Izumchenko G, Khanin M, Kirilov E, Krimer R, Nahum A, Steiner M, Walfisch Y, Walfisch S, Zango G, Danilenko M, Levy J. 2003. Modulation of transcriptional activity by antioxidant carotenoids. *Mol Aspects Med.* 24(6): 371-384.
- Shewmaker CK, Sheehy JA, Daley M, Colburn S, Ke DY. 1999. Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. *Plant J.* 20(4): 401-412X.
- Shivprasad S, Pogue GP, Lewandowski DJ, Hidalgo J, Donson J, Grill LK, Dawson WO. 1999. Heterologous sequences greatly affect foreign gene expression in tobacco mosaic virus-based vectors. *Virology.* 255(2): 312-323.
- Soll J, Schleiff E. 2004. Protein import into chloroplasts. *Nat Rev Mol Cell Biol.* 5(3): 198-208.
- Sparkes IA, Runions J, Kearns A, Hawes C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc.* 1(4): 2019-2025.
- Stahl W, Sies H. 2003. Antioxidant activity of carotenoids. *Mol Aspects Med.* 24(6): 345-351.

- Steinmetz NF, Evans DJ. 2007. Utilisation of plant viruses in bionanotechnology. *Org Biomol Chem.* 5(18): 2891-2902.
- Steinmetz NF, Lin T, Lomonosoff GP, Johnson JE. 2009. Structure-based engineering of an icosahedral virus for nanomedicine and nanotechnology. *Curr Top Microbiol Immunol.* 327: 23-58.
- Szeto WW, Hamer DH, Carlson PS, Thomas CA, Jr. 1977. Cloning of cauliflower mosaic virus (CLMV) DNA in *Escherichia coli*. *Science.* 196(4286): 210-212.
- Tadmor Y, Burger J, Yaakov I, Feder A, Libhaber SE, Portnoy V, Meir A, Tzuri G, Sa'ar U, Rogachev I, Aharoni A, Abeliovich H, Schaffer AA, Lewinsohn E, Katzir N. 2010. Genetics of flavonoid, carotenoid, and chlorophyll pigments in melon fruit rinds. *J Agric Food Chem.* 58(19): 10722-10728.
- Tamura K, Hara-Nishimura I. 2014. Functional insights of nucleocytoplasmic transport in plants. *Front Plant Sci.* 5: 118.
- Tanaka Y, Sasaki N, Ohmiya A. 2008. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54(4): 733-749.
- Thole V, Worland B, Snape JW, Vain P. 2007. The pCLEAN dual binary vector system for *Agrobacterium*-mediated plant transformation. *Plant Physiol.* 145(4): 1211-1219.
- Tilsner J, Oparka KJ. 2010. Tracking the green invaders: advances in imaging virus infection in plants. *Biochem J.* 430(1): 21-37.
- Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 5(2): 123-135.
- Turpen TH, Reinf SJ, Charoenvit Y, Hoffman SL, Fallarme V, Grill LK. 1995. Malarial epitopes expressed on the surface of recombinant tobacco mosaic virus. *Biotechnology (N Y).* 13(1): 53-57.

- Urcuqui-Inchima S, Haenni AL, Bernardi F. 2001. Potyvirus proteins: a wealth of functions. *Virus Res.* 74(1-2): 157-175.
- Verchot J, Koonin EV, Carrington JC. 1991. The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology.* 185(2): 527-535.
- Vitale A, Denecke J. 1999. The endoplasmic reticulum-gateway of the secretory pathway. *Plant Cell.* 11(4): 615-628.
- Waegemann K, Soll J. 1996. Phosphorylation of the transit sequence of chloroplast precursor proteins. *J Biol Chem.* 271(11): 6545-6554.
- Wang C, Zeng J, Li Y, Hu W, Chen L, Miao Y, Deng P, Yuan C, Ma C, Chen X, Zang M, Wang Q, Li K, Chang J, Wang Y, Yang G, He G. 2014. Enrichment of provitamin A content in wheat (*Triticum aestivum* L.) by introduction of the bacterial carotenoid biosynthetic genes *CrtB* and *CrtI*. *J Exp Bot.* 65(9): 2545-2556.
- Webster DE, Wang L, Mulcair M, Ma C, Santi L, Mason HS, Wesselingh SL, Coppel RL. 2009. Production and characterization of an orally immunogenic *Plasmodium* antigen in plants using a virus-based expression system. *Plant Biotechnol J.* 7(9): 846-855.
- Wei T, Zhang C, Hong J, Xiong R, Kasschau KD, Zhou X, Carrington JC, Wang A. 2010a. Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *PLoS Pathog.* 6(6): e1000962.
- Wei T, Huang TS, McNeil J, Laliberté JF, Hong J, Nelson RS, Wang A. 2010b. Sequential recruitment of the endoplasmic reticulum and chloroplasts for plant potyvirus replication. *J Virol.* 84(2): 799-809.
- Welsch R, Wust F, Bar C, Al-Babili S, Beyer P. 2008. A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes. *Plant Physiol.* 147(1): 367-380.

Werner S, Marillonnet S, Hause G, Klimyuk V, Gleba Y. 2006. Immunoabsorbent nanoparticles based on a tobamovirus displaying protein A. *Proc Natl Acad Sci USA*. 103(47): 17678-17683.

Whitham SA, Yamamoto ML, Carrington JC. 1999. Selectable viruses and altered susceptibility mutants in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*. 96(2): 772-777.

Wigdorovitz A, Perez Filgueira DM, Robertson N, Carrillo C, Sadir AM, Morris TJ, Borca MV. 1999. Protection of mice against challenge with foot and mouth disease virus (FMDV) by immunization with foliar extracts from plants infected with recombinant tobacco mosaic virus expressing the FMDV structural protein VP1. *Virology*. 264(1): 85-91.

Willemsen A, Zwart MP, Tromas N, Majer E, Daròs JA, Elena SF. 2016. Multiple barriers to the evolution of alternative gene orders in a positive-strand RNA virus. *Genetics*. 202(4): 1503-1521.

Wu S, Schalk M, Clark A, Miles RB, Coates R, Chappell J. 2006. Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nat Biotechnol*. 24(11): 1441-1447.

Wu S, Chappell J. 2008. Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. *Curr Opin Biotechnol*. 19(2): 145-152.

Wurbs D, Ruf S, Bock R. 2007. Contained metabolic engineering in tomatoes by expression of carotenoid biosynthesis genes from the plastid genome. *Plant J*. 49(2): 276-288.

Xu J, Dolan MC, Medrano G, Cramer CL, Weathers PJ. 2012. Green factory: plants as bioproduction platforms for recombinant proteins. *Biotechnol Adv*. 30(5): 1171-1184.

Ye X, Al-Babili S, Kloti A, Zhang J, Lucca P, Beyer P, Potrykus I. 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*. 287(5451): 303-305.

Yildiz I, Shukla S, Steinmetz NF. 2011. Applications of viral nanoparticles in medicine. *Curr Opin Biotechnol.* 22(6): 901-908.

Yuan H, Zhang J, Nageswaran D, Li L. 2015. Carotenoid metabolism and regulation in horticultural crops. *Hortic Res.* 2: 15036.

Zambryski P, Joos H, Genetello C, Leemans J, Montagu MV, Schell J. 1983. Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *EMBO J.* 2(12): 2143-2150.

Zhang H, Wang L, Hunter D, Voogd C, Joyce N, Davies K. 2013. A *Narcissus mosaic* viral vector system for protein expression and flavonoid production. *Plant Methods.* 9(1): 28.

Zhang Y, Butelli E, Martin C. 2014. Engineering anthocyanin biosynthesis in plants. *Curr Opin Plant Biol.* 19(81-90).

Zhu C, Naqvi S, Breitenbach J, Sandmann G, Christou P, Capell T. 2008. Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc Natl Acad Sci USA.* 105(47): 18232-18237.

Zwart MP, Willemsen A, Daros JA, Elena SF. 2014. Experimental evolution of pseudogenization and gene loss in a plant RNA virus. *Mol Biol Evol.* 31(1): 121-134.

Acknowledgements

I would like to express my deep gratitude to Dr. José Antonio Daròs, my Director, for giving me the opportunity to work in his laboratory and collaborate in many different projects. He helped me any time when I needed and I'm very thankful for his patient guidance and useful critiques.

I am particularly grateful to Dr. Carmelo López del Rincón, my Tutor, for his support and kindness with which he welcomed me every time when I went to his laboratory.

I'm greatly indebted to Dr. Belén Picó Sirvent, who helped me out so many times with an admirable predisposition to go through all of the administration steps.

This work would not have been possible without the support of my colleagues who contributed to this doctoral thesis. I would like to thank Professor Santiago F. Elena, Dr. Mark P. Zwart, Dr. Anouk Willemsen and Zaira Salvador for their support in the genome re-organization project, to Dr. José Antonio Navarro for his valuable advices in the protein targeting study, to Dr. Manuel Rodríguez-Concepción and Dr. Briardo Llorente for their creative ideas and support in the metabolic pathway rewiring study.

I would like to offer my special thanks to Dr. Guillermo Rodrigo for including me in many of his projects and giving me useful advices.

I am very grateful for my labmates, who accompanied me during these five years not just as colleagues but also as friends; special thanks to Verónica Aragonés, Fernando Martínez, Teresa Cordero, Arantxa Rosado and Alberto Carbonell. I can't forget to mention Miryam Pérez, who was my upholder during the 'El Camino' and my first half marathon.

I am very grateful for the help supported by the whole staff of the IBMCP; thanks to Marisol Gascón and Eugenio Grau for their excellent work and to Toni Villar for taking care of my plants in the greenhouse.

I want to thank to all of my friends in the institute with whom I shared some unforgettable moments: Pilar Rojas, Camilo Alvarez, Liliam Martínez, Gerardo Sánchez, Carito Gallego, Rim Hamza, Pepe Moya, Pablo Torres, Mariana Krüger, Priscilla De Barros Rossetto, Victor Zuñiga and Humberto Herrera. My grateful thanks

to Lesia Rodríguez, who was my good friend in this last year, carefully listened to me and gave me some excellent advices.

I would like to give many thanks to Horacio M. Pace Bedetti. His moral support, original ideas and constructive criticism makes me feel constantly be loved and helps me to become a better person.

Finally, I would like to dedicate this work to my family: Köszönöm a támogatást szerető szüleimnek, akik emberségükkel és kemény munkájukkal egész életemben például szolgáltak és a jövőben is annak szolgálnak majd.

Plant viruses are obligate intracellular parasites which are able to force host cells to synthesize high amounts of viral proteins. This fascinating capacity was exploited with recombinant virus vectors which express foreign proteins in plant biofactories. Some attempts were also made to use viral vectors in endogenous metabolic pathway modification of valuable natural products. The main limitation of many plant virus-based systems is the difficulty to co-express various heterologous proteins in the same cell with proper subcellular localization, which is a crucial question in metabolic engineering, since these pathways are composed of many enzymes located in different subcellular organelles.

This work provides a solution to overcome this limitation by using a potyvirus-based vector system which was manipulated to express several foreign proteins in equimolar amounts, target them to different subcellular compartments and express a whole heterologous biosynthetic pathway to induce valuable secondary metabolite production in infected tobacco leaves.

