INSTITUTO POTOSINO DE INVESTIGACIÓN CIENTÍFICA Y TECNOLÓGICA, A.C.

POSGRADO EN CIENCIAS EN BIOLOGÍA MOLECULAR

Análisis del transcriptoma de la interacción entre *Clavibacter michiganensis* subsp. *michiganensis* y *Solanum* spp.

Tesis que presenta

Leonardo Isaac Pereyra Bistraín

Para obtener el grado de

Doctor en Ciencias en Biología Molecular

Codirectores de la Tesis:

Dr. Ángel Gabriel Alpuche Solís

Dra. Dulce Alejandra Rougon Cardoso

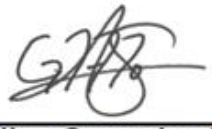
San Luis Potosí, S.L.P., diciembre de 2021

Constancia de aprobación de la tesis

La tesis "**Análisis del transcriptoma de la interacción entre *Clavibacter michiganensis* subsp. *michiganensis* y *Solanum* spp.**" presentada para obtener el Grado de Doctor en Ciencias en Biología Molecular fue elaborada por **Leonardo Isaac Pereyra Bistraín** y aprobada el **02 de diciembre de 2021** por los suscritos, designados por el Colegio de Profesores de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Dr. Ángel Gabriel Alpuche Solís
Codirector de la tesis

Dra. Dulce Alejandra Rougon Cardoso
Codirectora de la tesis



Dra. Carolina Granados Mendoza
Miembro del Comité Tutoral

Dr. Gerardo Rafael Argüello Astorga
Miembro del Comité Tutoral

Créditos Institucionales

Esta tesis fue elaborada en el Laboratorio de Biología Molecular de Plantas de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la dirección del Dr. Ángel Gabriel Alpuche Solís y la Dra. Dulce Alejandra Rougon Cardoso

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología (No. 330869) y del Instituto Potosino de Investigación Científica y Tecnológica, A. C. Se contó con los apoyos para investigación de los fondos: FMSLP-2013-C01-209337, CONACYT 316201, UNAM-DGAPA-PAPIME PE200618 y UNAM-PAPIIT IN220218.

IPICYT

Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Acta de Examen de Grado

El Secretario Académico del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 127 del Libro Segundo de Actas de Exámenes de Grado del Programa de Doctorado en Ciencias en Biología Molecular está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 13 días del mes de diciembre del año 2021, se reunió a las 12:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

Dr. Gerardo Rafael Argüello Astorga	**Presidente**	**IPICYT**
Dr. Ángel Gabriel Alpuche Solís	**Secretario**	**IPICYT**
Dra. Carolina Granados Mendoza	**Sinodal externo**	**UNAM**
Dra. Dulce Alejandra Rougon Cardoso	**Sinodal externo**	**UNAM**

a fin de efectuar el examen, que para obtener el Grado de:

DOCTOR EN CIENCIAS EN BIOLOGÍA MOLECULAR

sustentó el C.

Leonardo Isaac Pereyra Bistraín

sobre la Tesis intitulada:

Análisis del transcriptoma de la interacción entre Clavibacter michiganensis subsp. michiganensis y Solanum spp.

que se desarrolló bajo la dirección de

Dr. Ángel Gabriel Alpuche Solís
Dra. Dulce Alejandra Rougon Cardoso (UNAM)

El Jurado, después de deliberar, determinó

APROBARLO

Dándose por terminado el acto a las 15:00 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe el Secretario Académico del Instituto.

A petición del interesado y para los fines que al mismo convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 13 días del mes de diciembre de 2021.

Dr. Marcial Bonilla Marín
Secretario Académico

Mtra. Ivonne Lizette Cuevas Vélez
Jefa del Departamento del Posgrado

IPICYT
SECRETARÍA ACADÉMICA
INSTITUTO POTOSINO DE INVESTIGACIÓN CIENTÍFICA Y TECNOLÓGICA, A.C.

Dedicatorias

Con mucho cariño para mis padres, mi hermana y Jared por su cariño, apoyo y confianza.

A mis mejores amigos por siempre estar ahí compartiendo todas nuestras experiencias.

A Cintia por todo el apoyo, pero más compartir este viaje profesional y de vida.

Con mucho cariño para Said por ser una gran fuente de motivación. Espero que la vida te permita seguir y alcanzar todos sueños.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por otorgarme una beca para estudios de posgrado y de estancia de investigación.

Al Dr. Ángel Gabriel Alpuche Solís y la Dra. Alejandra Rougon Cardoso por su valioso apoyo y confianza para realizar este trabajo.

Al Dr. Cesaré Ovando Vázquez por el importante aporte y las mentorías durante este proyecto.

Al Dr. Gerardo Rafael Argüello Astorga y la Dra. Carolina Granados por sus siempre enriquecedores comentarios.

Al Dr. Sergio Casas Flores por apoyarnos para consegir semillas de tomate de *S. arcanum* usadas en este trabajo.

Al Dr. Tolga Bozkurt por abrirme las puertas de su laboratorio durante mi estancia de investigación.

A los técnicos: Q.F.B Rosalba Castillo y Biol. Salvador Ambriz por su apoyo en todo momento.

A mis colegas y compañeros del Laboratorio de Biología Molecular de Plantas.

Al Laboratorio Nacional de Biotecnología Agrícola, Médica y Ambiental (LANBAMA) por las facilidades otorgadas.

Al Centro Nacional de Supercómputo del IPICYT (CNS-IPICYT) por las facilidades otorgadas.

Contenido

List of tables

List of figures

Annexes

Resumen

Análisis del transcriptoma de la interacción entre *Clavibacter michiganensis* subsp. *michiganensis* y *Solanum* spp.

El tomate es uno de los cultivos más importantes para el consumo humano. Su producción es afectada por el actinomiceto *Clavibacter michiganensis* subsp. *michiganensis* (*Cmm*), considerado uno de los patógenos bacterianos más devastadores de este cultivo. Hasta ahora, el mecanismo molecular asociado a esta interacción planta-patógeno no ha sido comprendido completamente. En este trabajo, mediante RNA-Seq, se contrastaron los transcriptomas de la especie de tomate silvestre resistente *S. arcanum* LA2157 y la especie susceptible *S. lycopersicum* cv. Ailsa Craig durante las primeras 24 horas de interacción con el patógeno. Se realizaron tres enfoques de análisis: mapeo del genoma de referencia SL3.0 de *S. lycopersicum*, ensamblaje del transcriptoma semi *de novo* y ensamblaje del transcriptoma *de novo*. A pesar de que *S. lycopersicum* no es capaz de superar la infección por *Cmm*, varios grupos funcionales relacionados con defensa fueron enriquecidos. A diferencia de *S. lycopersicum*, *S. arcanum* mostró más genes inducidos relacionados con defensa bacteriana y fúngica en etapas tempranas. Los cambios transcripcionales globales parecen conservarse entre ambas especies. Nuestro estudio reveló genes interesantes que pueden contribuir al genotipo de resistencia. Los datos generados en este estudio contribuirán a una comprensión más profunda del mecanismo de defensa contra *Cmm,* proporcionando información para nuevos enfoques para obtener variedades resistentes. Además, se generaron vectores de silenciamiento basado en el begomovirus virus moteado del tomate (ToMoV), mediante la construcción de hemidímeros o dímeros parciales que pueden inocularse por bombardeo de partículas, y por procedimientos más rápidos y económicos como el método de infiltración de hojas. Los ensayos de silenciamiento de rango de hospedadores mediante infiltración, se realizaron usando el gen de la quelatasa de magnesio (*Chl*I), en el organismo modelo *Nicotiana benthamiana*, la especie comercial *S. lycopersicum*, así como especies de tomate silvestre relacionadas, y *Physalis ixocarpa*, comúnmente conocido como tomatillo. Las mediciones fenotípicas demostraron que mediante el uso de este vector de silenciamiento basado en ToMoV es posible alcanzar eficiencias de silenciamiento desde el 42% al 88%, a los 28 días posteriores a la inoculación, según la especie de planta, y no se observan efectos secundarios significativos en las plantas silenciadas. Este trabajo destaca un nuevo sistema VIGS que se puede adaptar para el estudio funcional de genes en diversas especies de solanáceas. Además, este sistema puede ser útil para un análisis más detallado de las interacciones tomate-*Cmm*.

Palabras clave: *Solanum*, RNA-Seq, *Cmm*, genes de defensa, análisis funcional, VIGS, Tomate, ToMoV, hemidímero.

Abstract

Molecular analysis of the interaction between *Clavibacter michiganensis* subsp. *michiganensis* and *Solanum* spp.

Tomato is one of the most important crops for human consumption. Its production is affected by the actynomicete *Clavibacter michiganensis* subsp*. michiganensis* (*Cmm*), considered one of the most devastating bacterial pathogens of this crop. Currently the molecular basis associated with this plant-pathogen interaction is not fully understood. In this work, the transcriptomes of the resistant wild tomato species *S. arcanum* LA2157 and the susceptible species *S. lycopersicum* cv. Ailsa Craig were contrasted using RNA-Seq. We considered the first 24 hours of interaction with *Cmm*. Three analysis approaches were performed: mapping to *S. lycopersicum* reference genome SL3.0, semi *de novo* transcriptome assembly and *de novo* transcriptome assembly. Despite being unable to overcome *Cmm* infection, *S. lycopersicum* showed enriched functional groups related to defense. Moreover, *S. arcanum* displayed more induced genes related to bacterial and fungal defense at early stages than *S. lycopersicum*. Global transcriptional changes seem to be conserved between both species. Our study revealed interesting genes that may contribute to the resistance genotype. The data generated in this study will contribute to a deeper understanding of the defense mechanism against *Cmm,* providing information for new approaches to obtain resistant varieties. Moreover, silencing vectors were obtained based on the begomovirus tomato mottle virus (ToMoV), by constructing partial dimers or hemidimers that can be inoculated by particle bombardment, and additionally we applied a less expensive and less time-consuming procedure such as the leaf infiltration method. Host-range silencing assays were performed by infiltration testing the magnesium chelatase gene (*Chl*I), required for chlorophyll biosynthesis in the model organism *Nicotiana benthamiana*, the commercial species *S. lycopersicum*, as well as related wild tomato species, and *Physalis ixocarpa* commonly known as “tomatillo”. Phenotypic measurements demonstrated that by using this ToMoV-based silencing vector is possible to reach silencing efficiencies from 42% up to 88%, at 28 days post-inoculation depending on the plant species, and no significant side effects are visible in the silenced plants. This work highlights a new VIGS system that can be adapted for functional genomics in diverse solanaceous species. In addition, this system can be useful for further analysis of Tomato-*Cmm* interactions.

KEY WORDS. *Solanum*, RNA-Seq, defense genes, functional analysis, VIGS, Tomate, ToMoV, hemidimer.

1. Introduction

Historically, Mexico represents the center of origin, domestication, or diversification of important crops for human consumption such as avocado, bean, chili, cotton, maize, tomato, and others (Challenger, 1998). Consequently, agricultural activity plays a fundamental role in the economy of Mexico. Among different vegetables farmed, tomato (*Solanum lycopersicum*) has relevance, since it is the second vegetable most consumed worldwide, just after potato (*Solanum tuberosum*) (FAO, 2018). Mexico belongs to the top ten tomato world producers; in 2018 the production was nearly four million tons (Figure 1). Mexico also occupies first place in exportations with more than two billion dollars in annual incomes see Figure 2 (FAO, 2018).

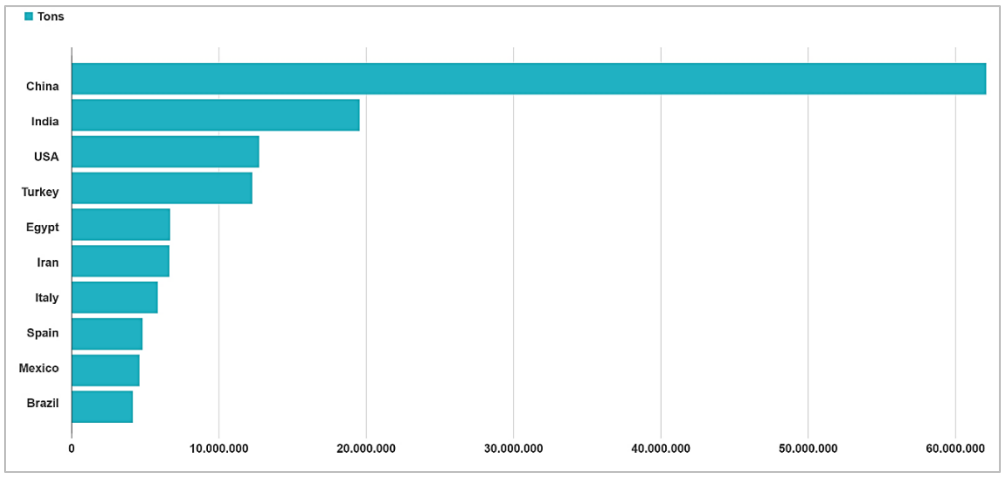


Figure 1. Top ten tomato producers. Source: FAO, 2018.

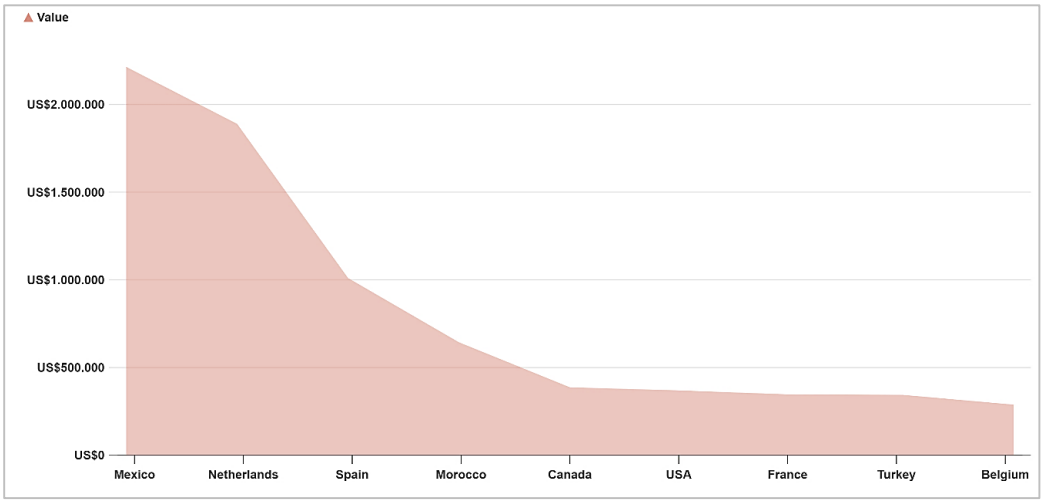


Figure 2. Top ten tomato exporters per annual income. Source: FAO, 2018.

Despite of tomato production statistics, there are important factors that can compromise tomato production yields, such as phytopathogenic diseases. Tomato crops are susceptible to approximately 200 diseases caused by bacteria, fungi, nematodes, and viruses (Lukyanenko, 1991). Some of them are considered the most important tomato diseases due to their destructiveness, yield, and economic losses, which can vary from 10% up to 90% of total production. Some examples include the Late blight caused by *Phytophthora infestans*, *Sclerotinia* rot caused by *Sclerotinia sclerotiorum*, *Fusarium* wilt caused by *Fusarium oxysporum,* and Bacterial wilt caused by *Ralstonia solanacearum,* among others (Kesavan and Chaudhary 1977; Jnra *et al.,* 2000; Foolad *et al.,* 2008; Lebeau *et al*., 2011; Singh *et al.,* 2017; Nandi *et al.,* 2018).

In this work, we are focused on the Bacterial canker of tomato caused by *Clavibacter michiganensis* subsp. *michiganensis* (*Cmm*). This disease was described more than 100 years ago in Michigan, United States (Smith, 1910) and now it is present in almost all tomato producing areas (CABI, 2021). *Cmm* is considered one of the most devastating pathogens of this crop (Mansfield *et al.*, 2012). If disease spreading is not properly controlled the tomato production can be dramatically affected in fields and protected crops producing severe economic losses. Countries such as Canada, have reported losses of up to 84% in producing fields; whereas in Michigan, US, there has been economic losses recorded for up to $300,000 dollars just in a year (Poysa, 1993; Hausbeck *et al.,* 2000; Sen e*t al.,* 2015). In Mexico, by 2006 about

200 hectares of protected crops were damaged, summing up to 40 million dollars in losses (Borboa Flores *et al.*, 2009).

Cmm is a pathogenic Gram-positive actinomycete (Stackebrandt *et al*., 1997), that can infect tomato plants trough wounds (Carlton *et al.*, 1998) and natural entries such as stomata, hydathodes and trichomes (Kontaxis, 1962; Carlton *et al*.,1998). The typical transmission sources are infected tomato seeds, infected young seedlings and residual plants in the soil (Strider, 1969; Gitaitis *et al*. 1991; Kleitman *et al*., 2008). Symptoms such as unilateral wilting of leaves, cankers on stems, and development of bird-eye spots in infected fruits are characteristic of bacterial canker of tomato (Gleason *et al*., 1993). *Cmm* colonizes ascendingly through the vascular system of tomato plants. At the beginning of the infection, *Cmm* multiplies and spreads laterally through xylem vessels, which causes the lateral wilting symptom. Later, *Cmm* degrades phloem tissues as the bacteria multiply (Pine *et al*., 1955; Wallis *et al*., 1977). Moreover, this pathogenic bacteria can survive in soil and plant debris for more than two years (Strider, 1967; Fatmi & Schaad, 2002), which represents an infection source for subsequent farming seasons; therefore crop rotation is recommended.

The molecular mechanism of tomato-*Cmm* interaction has not been fully understood. However, previous studies have uncovered key factors participating in such interaction. Regarding *Cmm*, three genomic features are particularly responsible for the pathogenesis. A 129 kb region in the chromosome with a low GC content near the origin of replication is recognized as the pathogenicity island (PAI), present in the *Cmm* strains that are pathogenic in tomato, and two plasmids, pCM1 (~27-59 kb) and pCM2 (~64-109 kb) (Meletzus *et al*., 1993; Gartemann *et al*., 2008; Thapa *et al*., 2017); the presence or absence of one or two of these plasmids compromises the pathogen virulence. Therefore, strains of *Cmm* can be defined as hypervirulent, hypovirulent, or nonvirulent (Alvarez *et al*., 2004). The PAI is divided into two subregions chp and tomA; the chp subregion contains genes that code serine-protease enzymes and other genes involved in virulence, and the tomA subregion contains genes participating in carbohydrate metabolism, as well as the *tomA* gene

that encodes a tomatinase responsible for the degradation of tomatin, which participates in tomato defense response (Kaup *et al*., 2005; Gartemann *et al*., 2008; Stork *et al*., 2008). Regarding pCM plasmids, two important virulence factors have been well characterized, the *celA* gene that encodes β-1,4 endoglucanase (Jahr *et al*., 2000) located on pCM1 and the *pat-1* gene on pCM2, which encodes a serine-protease enzyme (Dreier *et al*., 1997). Both are often used as markers to distinguish the *Clavibacter michiganensis* subsp. *michiganensis* subspecies from other *Clavibacter michiganensis* subspecies. Moreover, studies have revealed important traits involved in the *Cmm* pathogenicity, such as the transcriptional regulators Vatr1 and Vatr2, whose corresponding genes are located outside the PAI of the chromosome. These regulators play an important role since they are on the same molecular pathway responsible for *Cmm* virulence (Savidor *et al*., 2014).

Despite *S. lycopersicum* is unable to overcome the *Cmm* infection, it elicits a defense response, as well as the induction of several disease development factors associated with the pathogenesis process. Previous studies performed during the first eight days of tomato-*Cmm* interaction disclosed the induction of defense-related genes, pathogen-related proteins, production of free oxygen radicals, hormone synthesis, and the synthesis of ethylene, the last one characterized as an important factor in disease symptom development (Balaji *et al*., 2008; Savidor *et al*., 2011). However, several defense-related genes and proteins are also inhibited during this interaction, such as various class III peroxidases, the developmentally regulated plasma membrane polypeptide (DREPP), and a serine/threonine protein kinases required for signal transduction (Balaji *et al*., 2008; Savidor *et al*., 2011), which suggests that *Cmm* can down regulate the tomato defense response. Even when these studies have contributed to the partial understanding of the tomato-*Cmm* interaction, further efforts are necessary to discover new factors related to disease development. The study of such interactions remains challenging since there is not a well-established plant-*Cmm* model of study like in other diseases. Consequently, studies that require gene functional analysis becomes harder and time consuming since they are carried out mainly in commercial tomato; new strategies of study are required to overcome such limitations.

To date, there are not effective control methods against *Cmm* and most of them are preventive. Chemical treatments are principally based on cupper salts and antimicrobial compounds (e.g., streptomycin and mancozeb). Although these methods offers certain level of effectiveness, they can induce pathogen resistance, phytotoxicity, and phreatic surface pollution (Hausbeck *et al*., 2000; de León *et al.*, 2008; Nandi *et al.,* 2018). Another preventive method widely used is based on sanitary control, which is focused on hygienic good practices, such as, adequate decontamination of propagation material, green houses, workers, etc. (Gitaitis *et al.,* 1992; Gleason *et al.,* 1993). The use of beneficial microorganisms is another emerging control method. For example, treatments with *Bacillus subtilis*, *Trichoderma harzianum* and *Rhodosporidium diobovatum* can prevent the incidence of bacterial canker of tomato plants under greenhouse conditions (Utkhede and Koch, 2004). Furthermore, seed treatment with several *Bacillus* strains improves the quality of seeds (Kasselaki *et al.,* 2011). Likewise, some other bacteria and phages, have been studied due to their potential role for controlling bacterial canker of tomato (Haas and Defago, 2005; Wittmann *et al*., 2010; Paulin *et al*., 2017). The experimental employment of vegetal extracts such as cloves of *Allium sativum* and fruits of *Ficus carica* at different concentrations, resulted in effective disease control of up to 65% (*A. sativum*) and 38% (*F. carica*) of that of the standard copper treatment, indicating that these natural compounds could be useful for protecting tomato plants from *Cmm* and other tomato bacterial pathogens (Balestra *et al*., 2009). The proper use of these control methods either alone or combined, have a crucial role in preventing the bacterial canker of tomato. Nonetheless, once *Cmm* establishes in a tomato crop, can be very difficult to get a full recovery using these control methods. Knowing this, the development of new control strategies is essential.

Several desirable traits were selected during crop domestication such as bigger fruits, higher yields, evenness of ripening, flavor, among others. Nonetheless, this evolutionary process has conducted to a reduced genetic diversity affecting disease resistance principally to pathogens that are virulent on a limited range of hosts, often a single plant species or genus (Couch *et al*., 2005; Dwivedi *et al*., 2008; Bhupendra,

2013). It means that wild species closely related to domesticated crops still harbors such disease resistance traits (Jones *et al*., 1994; Yun *et al*., 2005; Liu *et al*., 2015). Some wild tomato species have been described as tolerant to *Cmm* infection such as *S. arcanum*, *S. chilense, S. habrochaites* and *S. pimpinenllifolium* (Sotirova *et al*., 1994; Van Heusden *et al*., 1999; Francis *et al.*, 2001; Kabelka *et al.*, 2002). Studies demonstrate that resistance is a multigenic characteristic and several quantitative trait loci (QTL), associated with defense response against this pathogenic bacterium (Thyr, 1976; Sandbrink *et al*., 1995; Van Heusden *et al*., 1999; Kabelka *et al.*, 2002). To date, no commercial resistant tomato varieties are available and only a few attempts carried out by traditional plant breeding have resulted in plants partially resistant to *Cmm* (Vulkova *et al*. 1993, 1976; Poysa *et al*., 1993). Emerging techniques in genetic engineering offer a broad range of applications for the study of Bacterial canker of tomato. Recent works have shown innovative approaches for crop improvement. For example, *de novo* domestication of wild tomato species using genome editing, empowers the future obtainment of new cultivars with characteristics of commercial crops, but preserving the natural resistance to biotic and abiotic stresses characteristic of wild species (Zsögön *et al*., 2018; Li *et al*., 2018).

In comparison to classic genomic' methods, the identification of genomic features of specific wild plant species has become relatively easier due to rapid development of next-generation sequencing (NGS). These technologies allow to generate massive information of genomic sequences, gene expression data and other varieties of massive data. Such information can be used to study plant diversity, evolutionary relationships, or to perform molecular breeding even without previous knowledge of genomic information (Unamba *et al*., 2015). NGS can also be applied to identify genetic sources of disease resistance and to provide new insights into plant-pathogen interactions (e.g., Frades *et al*., 2015). Once candidate genes or proteins are identified by NGS (e.g., RNA-Seq), the next step is their functional validation. For wild plant species this step remains challenging since functional analysis tools such as *Agrobacterium*-mediated plant transformation results difficult to apply in non-

model plants, which are mostly optimized for a few commercial plant species and model organisms.

Virus-induced gene silencing (VIGS), is another functional analysis system that has a key role in the study of plant-pathogen interactions, it can be adapted to a broad range of plant species including wild species (e.g., Esparza-Araiza *et al*., 2015; Krishnan *et al.,* 2015). VIGS is a less time-consuming method than stable transformation protocols. More important, it also allows the functional analysis of genes in plant species that are recalcitrant to stable transformation (e.g., Krishnan *et al*., 2015). Besides, since NGS is providing large amounts of data, VIGS enables the analysis of collections of candidate genes for hypothesis-driven research, in non-model plant species (Dommes *et al*., 2019).

The present research work addresses the Bacterial Canker of tomato by two main approaches divided into two chapters; the identification of potentially resistance-related genes associated to the Tomato-*Cmm* interaction, and the development of a functional analysis tool based on the VIGS technology.

Bibliography

Alvarez AM, Kaneshiro, WS, Vine BG. (2004) Diversity of *Clavibacter michiganensis* subsp. *michiganensis* populations in tomato seed: What is the significance. Acta Horticulturae 695, 205-214. doi: 10.17660/ActaHortic.2005.695.23

Balaji V, Mayrose M, Sherf O, Jacob-Hirsch J, Eichenlaub R, Iraki N, Manulis-Sasson S, Rechavi G, Barash I, Sessa G. (2008) Tomato transcriptional changes in response to *Clavibacter michiganensis* subsp. *michiganensis* reveal a role for ethylene in disease development. Plant Physiology 146, 1797–1809. https://doi.org/10.1104/pp.107.115188

Balestra GM, Heydari A, Ceccarelli D, Ovidi E, and Quattrucci, A. (2009) Antibacterial effect of *Allium sativum* and *Ficus carica* extracts on tomato bacterial pathogens. Crop Protection 28, 807-811. https://doi.org/10.1016/j.cropro.2009.06.004

Bhupendra C. (2013) Plant domestication and resistance to herbivory. International journal of Plant Genomics 572784. https://doi.org/10.1155/2013/572784

Borboa Flores J, Rueda Puente EO, Acedo Felix E, Ponce JF, Cruz M, Grimaldo Juárez O, Garcia Ortega AM. (2009) Deteccion de *Clavibacter michiganensis* subespecie *michiganensis* en el tomate del estado de Sonora, México. Revista Fitotecnia Mexicana 32, 319–326.

CABI. (2021) *Clavibacter michiganensis* subsp. *michiganensis*. In Invasive Species Compendium; CAB International: Wallingford, UK. Available online: www.cabi.org/isc

Carlton WM, Braun EJ, Gleason ML. (1998) Ingress of *Clavibacter michiganensis* subsp. *michiganensis* into tomato leaves through hydathodes. Phytopathology 88, 525–529. https://doi.org/10.1094/PHYTO.1998.88.6.525

Challenger A. (1998) Utilización y conservación de los ecosistemas terrestres de México. Pasado, presente y futuro. Conabio-Instituto de Biología, UNAM y Agrupación Sierra Madre, México, 847 p.

Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, van Kim P, Nottéghem JL, and Kohn, L.M. (2005) Origins of hostspecific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. Genetics 170, 613–630. doi: 10.1534/genetics.105.041780

De León L, Siverio F, López MM, Rodríguez A. (2008) Comparative efficiency of chemical compounds for in vitro and in vivo activity against *Clavibacter michiganensis* subsp. *michiganensis*, the causal agent of tomato bacterial canker. Crop Protection 27, 1277–1283.

Dommes AB, Gross T, Herbert DB, Kivivirta KI, Becker A. (2019) Virus-induced gene silencing: empowering genetics in non-model organisms. Journal of Experimental Botany, 70, 757–770. https://doi.org/10.1093/jxb/ery411

Dreier J, Meletzus D, Eichenlaub R. (1997) Characterization of the plasmid encoded virulence region pat-1 of phytopathogenic *C. michiganensis* subsp. *michiganensis*. Molecular Plant-Microbe Interactions 10, 195–206. https://doi.org/10.1094/MPMI.1997.10.2.195

Dwivedi SL, Upadhyaya HD, Stalker HT, Blair MW, Bertioli DJ, Nielen S, Ortiz R. (2008) Enhancing crop gene pools with beneficial traits using wild relatives. In Plant Breeding Reviews. doi: 10.1002/9780470380130.ch3

Esparza-Araiza MJ, Bañuelos-Hernández B, Argüello-Astorga GR, Lara-Ávila JP, Goodwin PH, Isordia-Jasso MI, Castillo-Collazo R, Rougon-Cardoso A, Alpuche-Solís ÁG. (2015) Evaluation of a SUMO E2 conjugating enzyme involved in resistance to *Clavibacter michiganensis* subsp. *michiganensis* in *Solanum peruvianum*, through a Tomato Mottle Virus VIGS assay. Frontiers in Plant Sciences 6, 1019. https://doi.org/10.3389/fpls.2015.01019

FAOSTAT. (2018) Food and Agriculture Organization of the United Nations. FAOSTAT Statistical Database. Available online: https://www.fao.org/faostat/en/#home

Fatmi M, Schaad NW. (2002) Survival of *Clavibacter michiganensis* ssp. *michiganensis* in infected tomato stems under natural field conditions in California, Ohio and Morocco. Plant Pathology 51, 149–154. https://doi.org/10.1046/j.1365-3059.2002.00675.x

Foolad MR, Merk HL, Ashrafi H. (2008) Genetics, genomics and breeding of late blight and early blight resistance in tomato. Critical Reviews in Plant Sciences 27, 75–107. doi: 10.1080/07352680802147353.

Frades I, Abreha KB, Proux-Wéra E, Lankinen Å, Andreasson E, Alexandersson E. (2015) A novel workflow correlating RNA-seq data to *Phythophthora infestans* resistance levels in wild Solanum species and potato clones. Frontiers in Plant Sciences *6*, 718. https://doi.org/10.3389/fpls.2015.00718

Francis DM, Kabelka E, Bell J, Franchino B, St Clair D. (2001) Resistance to bacterial canker in tomato (*Lycopersicon hirsutum* LA407) and its progeny derived

from crosses to *L. esculentum*. Plant Disease 85, 1171–1176. https://doi.org/10.1094/PDIS.2001.85.11.1171

Gartemann KH, Abt B, Bekel T, *et al.* (2008) The genome sequence of the tomato pathogenic Actinomycete *C. michiganensis* subsp. *michiganensis* NCPPB382 reveals a large island involved in pathogenicity. Journal of Bacteriology 190, 2138–2149. https://doi.org/10.1128/JB.01595-07

Gitaitis RD, Beaver RW, Voloudakis AE. (1991) Detection of *Clavibacter michiganensis* subsp. *michiganensis in symptomless tomato transplants*. Plant Disease 75, 834-838. https://doi.org/10.1094/PD-75-0834

Gitaitis RD, McCarter S, Jones J. (1992) Disease control in tomato trans-plants produced in Georgia and Florida. Plant Disease 76, 651–656

Gleason ML, Gitaitis RD, Ricker MD. (1993) Recent progress in understanding and controlling bacterial canker of tomato in eastern North America. Plant Disease 77, 1069–1076.

Haas D and Defago G. (2005) Biological control of soil-borne pathogens by fluo-rescent pseudomonads. Nature Reviews, Microbiology 3, 307-19. doi:10.1038/nrmicro1129

Hausbeck M K, Bell J, Medina-Mora C, Podolsky R, Fulbright DW. (2000) Effect of bactericides on population sizes and spread of *C. michiganensis* subsp. *michiganensis* on tomatoes in the greenhouse and on disease development and crop yield in the field. Phytopathology 90, 38-44. https://doi.org/10.1094/PHYTO.2000.90.1.38

Jahr H, Dreier J, Meletzus D, Bahro R, Eichenlaub R. (2000) The endo-beta-1,4-glucanase celA of *Clavibacter michiganensis* subsp. *michiganensis* is a pathogenicity determinant required for induction of bacterial wilt of tomato. Molecular Plant-Microbe Interactions 13, 703–714. https://doi.org/10.1094/MPMI.2000.13.7.703

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JD. (1994) Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266, 789–793. https://doi.org/10.1126/science.7973631

Jnra ML, Lopes CA, Silva WLC. (2000) Sclerotinia rot losses in processing tomatoes grown under centre pivot irrigation in central Brazil. Plant Pathology 49, 51–56. doi: 10.1046/j.1365-3059.2000.00394.x

Kabelka E, Franchino B, Francis DM. (2002) Two Loci from *Lycopersicon hirsutum* LA407 Confer Resistance to Strains of *Clavibacter michiganensis* subsp. *michiganensis*. Phytopathology 92, 504–510. https://doi.org/10.1094/PHYTO.2002.92.5.504.

Kaup O, Grafen I, Zellermann EM, Eichenlaub R, Gartemann KH. (2005) Identification of a tomatinase in the tomato-pathogenic actinomycete *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. Molecular Plant-Microbe Interactions 18, 1090–1098. https://doi.org/10.1094/MPMI-18-1090

Kasselaki AM, Goumas D, Tamm L, Fuchs J, Cooper J, Leifert C. (2011) Effect of alternative strategies for the disinfection of tomato seed infectedwith bacterial canker (*Clavibacter michiganensis* subsp. *michiganensis*). NJAS –Wageningen Journal of Life Sciences 58, 145–147. https://doi.org/10.1016/j.njas.2011.07.001.

Kesavan V, Chaudhary B. Screening for resistance to Fusarium wilt of tomato. (1997) SABRAO J 9, 51–65.

Kleitman F, Barash I, Burger A, Iraki N, Falah Y, Sessa G, Weinthal D, Chalupowicz L, Gartemann KH, Eichenlaub R. (2008) Characterization of a *Clavibacter michiganensis* subsp. *michiganensis* population in Israel. European Journal of Plant Pathology 121, 463-475. https://doi.org/10.1007/s10658-007-9264-z

Kontaxis DG. (1962) Leaf trichomes as avenues for infection by *Corynebacterium michiganense*. Phytopathology 52, 1306-1307.

Krishnan A, Mahadevan C, Mani T, Sakuntala M. (2015) Virus-induced gene silencing (VIGS) for elucidation of pathogen defense role of serine/threonine protein kinase in the non-model plant *Piper colubrinum* Link. Plant Cell Tissue and Organ Culture 122, 269–283. https://doi.org/10.1007/s11240-015-0764-9

Lebeau A, Daunay MC, Frary A, Palloix A, Wang JF, Dintinger J, Chiroleu F, Wicker E, Prior P. (2011) Bacterial wilt resistance in tomato, pepper, and eggplant: genetic resources respond to diverse strains in the *Ralstonia solanacearum* species complex. Phytopathology 101, 154–165. https://doi.org/10.1094/PHYTO-02-10-0048

Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C. (2018) Domestication of wild tomato is accelerated by genome editing. Nature Biotechnology. https://doi.org/10.1038/nbt.4273

Liu J, Zheng Z, Zhou X, Feng C, Zhuang Y. (2015) Improving the resistance of eggplant (*Solanum melongena*) to Verticillium wilt using wild species *Solanum linnaeanum*. Euphytica 201, 463–469. https://doi.org/10.1007/s10681-014-1234-x

Lukyanenko AN. (1991) Disease resistance in tomato. Genetic improvement of tomato, 99– 119. https://doi.org/10.1007/978-3-642-84275-7_9

Mansfield J, Genin S, Magori S, *et al*. (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. Molecular Plant Pathology 13, 614–629. https://doi.org/10.1111/j.1364-3703.2012.00804.x

Meletzus D, Bermphol A, Dreier J, Eichenlaub R. (1993) Evidence for plasmid-encoded virulence factors in the phytopathogenic bacterium *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. Journal of Bacteriology 175, 2131–2136. https://doi.org/10.1128/jb.175.7.2131-2136.1993

Nandi M, Macdonald J, Liu P, Weselowski B, Yuan ZC. (2018) *Clavibacter michiganensis* ssp. *michiganensis*: bacterial canker of tomato, molecular interactions and disease management. Molecular Plant Pathology 19, 2036-2050. https://doi.org/10.1111/mpp.12678

Paulin MM, Novinscak A, Lanteigne C, Gadkar VJ, Filion M. (2017) Interaction between 2,4-diacetylphloroglucinol- and hydrogen cyanide-producing *Pseudomonas brassicacearum* LBUM300 and *Clavibacter michiganensis* subsp. *michiganensis* in the tomato rhizosphere. Applied and Environmental Microbiology 83, e00073-17. https://doi.org/10.1128/AEM.00073-17

Poysa, V. (1993) Evaluation of tomato breeding lines resistant to bacterial canker Canadian Journal of Plant Pathology 15, 301-304. doi: 10.1080/07060669309501927

Pine TS, Grogan RG, Hewitt WB. (1955) Pathological anatomy of bacterial canker of young tomato plants. Phytopathology 45, 267-271.

Sandbrink JM, Vanooijen JW, Purimahua CC, Vrielink M, Verkerk R, Zabel P, Lindhout P. (1995) Localization of genes for bacterial canker resistance in *Lycopersicon peruvianum* using Rflps. Theoretical and Applied Genetics, 90, 444–450. https://doi.org/10.1007/BF00221988

Savidor A, Teper D, Gartemann KH, Eichenlaub R, Chalupowicz L, Manulis-Sasson S, Barash I, Tews H, Mayer K, Giannone RJ. (2011) The *Clavibacter michiganensis* subsp. *michiganensis*–tomato interactome reveals the perception of pathogen by the host and suggests mechanisms of infection. Journal of Proteome Research 11, 736–750. https://doi.org/10.1021/pr200646a

Savidor A, Chalupowicz L, Teper D, Gartemann KH, Eichenlaub R, Manulis-Sasson S, Barash I, Sessa G. (2014) *Clavibacter michiganensis* subsp. *michiganensis* Vatr1 and Vatr2 transcriptional regulators are required for virulence in tomato. Molecular plant-microbe interactions 27, 1035–1047. https://doi.org/10.1094/MPMI-02-14-0061-R

Sen Y, van der Wolf J, Visser RG, van Heusden S. (2015) Bacterial canker of tomato: current knowledge of detection, management, resistance, and interactions. Plant Disease 99, 4–13. https://doi.org/10.1094/PDIS-05-14-0499-FE

Singh VK, Singh AK, Kumar A. (2017) Disease management of tomato through PGPB: current trends and future perspective. 3 Biotech 7, 255. doi:10.1007/s13205-017-0896-1

Sotirova, V., Bogatsevska, N., and Stamova, L. (1994) Sources of resistance to bacterial diseases in tomato wild species. Acta Horticulturae 376, 353- 360. https://doi.org/10.17660/ActaHortic.1994.376.49

Smith, E.F. (1910) A new tomato disease of economic importance. Science, 31, 794–796.

Stackebrandt E, Rainey FA, WardRainey NL. (1997). Proposal for a new hierarchic classification system, Actinobacteria classis nov. International Journal of Systematic and Evolutionary Microbiology 47, 479–491. https://doi.org/10.1099/00207713-47-2-479

Strider DL. (1967) Survival studies with the tomato bacterial canker organism. Phytopathology 57, 1067–71.

Strider DL. (1969) Bacterial canker of tomato caused by *Corynebacterium michiganense*. A literature review and bibliography. North Carolina Agricultural Experiment Station 193, 1-110.

Stork I, Gartemann KH, Burger A, Eichenlaub R (2008) A family of serin proteases of *Clavibacter michiganensis* subsp. *michiganensis*: chpC plays a role in colonization of the host plant tomato. Molecular Plant Pathology 9, 599–608. https://doi.org/10.1111/j.1364-3703.2008.00484.x

Thapa SP, Pattathil S, Hahn MG, Jacques MA, Gilbertson RL, Coaker G. (2017) Genomic analysis of *Clavibacter michiganensis* reveals insight into virulence strategies and genetic diversity of a Gram-positive bacterial pathogen. Molecular Plant-Microbe Interactions 30, 786–802. https://doi.org/10.1094/MPMI-06-17-0146-R

Thyr BD. (1976) Inheritance of resistance to *Corynebacterium michiganense* in tomato. Phytopathology 66, 16-11 19.

Unamba CI, Nag A, Sharma RK. (2015) Next Generation Sequencing technologies: The doorway to the unexplored genomics of non-model plants. Frontiers in Plant Sciences 6, 1074. https://doi.org/10.3389/fpls.2015.01074

Utkhede R, Koch C. (2004) Biological treatments to control bacterial canker of greenhouse tomatoes. BioControl 49, 305–313. https://doi.org/10.1023/B:BICO.0000025373.69584.08

Van Heusden AW, Koornneef M, Voorrips RE, Brüggemann W, Pet G, Vrielink-Van Ginkel R, Chen X, Lindhout P. (1999) Three QTLs from *Lycopersicon peruvianum* confer a high level of resistance to *Clavibacter michiganensis* ssp. *michiganensis*. Theoretical and Applied Genetics 99, 1068–1074.

Vulkova ZV, Sotirova VG. (1993) Study of the three-genome hybrid *Lycopersicon esculentum* Mill. – *L. chilense* Dun. – *L. peruvianum* var 'humifusum' Mill. and its use as a source for resistance. Theoretical and Applied Genetics 87, 337-342. https://doi.org/10.1007/BF01184920

Wallis FM. (1977) Ultrastructural histopathology of tomato plants infected with *Corynebacterium michiganense*. Physiological Plant Pathology 11, 333- 342.

Wittmann J, Eichenlaub R, Dries EB. (2010) The endolysins of bacteriophages CMP1 and CN77 are specific for the lysis of *Clavibacter michiganensis* strains. Microbiology, 156, 2366–2373. https://doi.org/10.1099/mic.0.037291-0

Yun SJ, Gyenis L, Hayes PM, Matus I, Smith KP, SteVenson BJ, Muehlbauer GJ (2005) Quantitative trait loci for multiple disease resistance in wild barley. Crop Science 45, 2563–2572. https://doi.org/10.2135/cropsci2005.0236

Zsögön A, Čermák T, Naves ER, *et al*. (2018) De novo domestication of wild tomato using genome editing. Nature Biotechnology. https://doi.org/10.1038/nbt.4272

2. Objective

To identify possible sources of genetic resistance to bacterial canker of tomato, by the identification of candidate genes found in wild tomato species resistant to *Clavibacter michiganensis* subsp. *michiganensis.*

Specific goals

- To identify putative resistance genes through the analysis of the *S. arcanum* LA2157 and *S. lycopersicum* transcriptomes obtained during the interaction with *Clavibacter michiganensis* subsp. *michiganensis.*
- To improve a VIGS system for functional genomics analysis, based on the tomato mottle virus (ToMoV).

Chapter I

Comparative RNA-Seq analysis reveals potentially resistance-related genes in response to bacterial canker of tomato

Pereyra-Bistraín LI, Ovando-Vázquez C, Rougon-Cardoso A, Alpuche-Solís ÁG. (2021) Comparative RNA-Seq analysis reveals potentially resistance-related genes in response to bacterial canker of tomato. Genes 12, 1745. https://doi.org/10.3390/genes12111745

1. Background

Solanum arcanum LA2157 has been characterized with the highest degree of resistance against the bacterial canker of tomato (Van Heusden *et al.*, 1999; Sen *et al.*, 2013). This wild tomato species can be considered a natural source of resistance that could be used for the improvement of the commercial tomato species *S. lycopersicum* by incorporation of resistance-related genes. Hence, the importance to find response elements against this disease in the wild species.

Studies using techniques such as microarrays for differential expression analysis, have been useful to identify important genes that are probably associated to disease symptoms development, in particular, genes related to host-derived ethylene production. These genes play an important role in regulation of tomato's susceptible response during *Cmm* and *S. lycopersicum* interaction (Balaji *et al.*, 2008). Using cDNA-AFLP method, several putative tomato resistance genes have been identified and described, including genes involved in the ubiquitin mediated protein degradation pathway and oxidative burst. Such results were obtained by comparing gene expression levels during the interaction of *Cmm* with two resistant wild tomato species versus a susceptible tomato species *S. lycopersicum* (Lara-Ávila *et al.*, 2012).

RNA-Seq is a next generation sequencing technology that has been widely used in the last ten years. It allows to obtain almost complete gene expression profiles in a specific condition, overcoming the limitations of other transcriptomic analysis tools

such as microarrays. Furthermore, RNA-Seq downstream analyses permit the classification and quantification of the whole diversity of RNA sequences, either by using an annotated genomic reference, or by *de novo* assembling the obtained sequences. This *de novo* strategy also allows for the detection of new genes and alternative splicing events (Ameur *et al.*, 2010; Ozsolak and Milos, 2010; Slatko *et al.*, 2018).

In this study, we aimed to identify genes potentially related to an early defense response against *Cmm* using RNA-Seq. For that purpose, we contrasted the global transcriptional changes of the non-model resistant wild tomato species *S. arcanum* LA2157 and the susceptible species *S. lycopersicum* cv. Ailsa Craig during the first 24 hours of interaction with *Cmm*. This was performed using three different approaches of data analyses: mapping the RNA-Seq fragments to *S. lycopersicum* reference genome SL3.0 (MR), performing a semi *de novo* transcriptome assembly (STA) and a *de novo* transcriptome assembly (DA). Here we report a multi-bioinformatics approach to compare the transcriptomes of two tomato species during the *Cmm* infection. Data provided in this study contributes to the screening of defense response genes on Solanum species.

2. Methods

2.1. Pathogenic *Cmm* strain and plant material

S. lycopersicum cv. Ailsa Craig tomato seeds were obtained from University of Nottingham (Nottingham, UK). *S. arcanum* LA2157 seeds were acquired from Tomato Genetics Resource Center (TGRC, Davis, CA, USA). *Cmm* 1569 strain was isolated from a farming area in Chiapas, Mexico and donated by Dr. Lara-Ávila.

2.2. Growth conditions and plant infection

All experiments were carried out in controlled conditions, as described below. Before germination, all tomato seeds were sterilized in a 10% extran solution for 15 min. Subsequently, three washes with deionized water and one wash with 70% ethanol solution were performed, followed by a 10% bleach solution wash for 10 min and rinsing three times in deionized water. Tomato seeds were germinated in a commercial mix substrate (Sunshine Mix #3, Sun Grow Horticulture, Vancouver, BC,

CA) and a bioclimatic chamber at 25°C and 16/8 h light and dark period. Two weeks later, plants were transferred to greenhouse conditions with a controlled temperature max 28°C, min 20°C and humidity of 55-60%.

Forty days after germination, both plant species were infected with the pathogenic strain *Cmm* 1569. The bacterial cells were propagated in 802 medium cultures (2g/L^{-1} yeast extract, 1g/L^{-1} polypeptone and 0.92 g/L^{-1} magnesium sulfate). For inoculation, three biological replicates per each tomato species (*S. lycopersicum* cv. Ailsa Craig and *S. arcanum* LA2157) were inoculated in the first proximal internode above ground using an insulin syringe with a concentration of $5x10^7$ CFU (OD = 0.2). Plants were sampled (true leaves next to inoculation site) at 0, 8 and 24 h post inoculation. Biological samples were immediately frozen in liquid nitrogen and stored at -80°C. The plants were maintained in greenhouse conditions until the appearance of the characteristic disease symptoms.

2.3. cDNA library preparation and RNA sequencing

The 18 samples (nine for each plant species) were sent to the Beijing Genome Institute (BGI. Shenzhen, GD, China). TruSeq paired-end libraries were produced and sequenced by Illumina Hiseq10X (Illumina, Inc., San Diego, CA, USA). Multiplex sequencing was performed and the average targeted length obtained per read was 150 bp.

2.4. RNA-Seq data processing

Map to the *S. lycopersicum* genome SL3.0

This pipeline was designed to quantify gene expression using the *S. lycopersicum* available genome as reference, available at the time this research was conducted. Raw reads quality assessment was performed using the program FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). A summary of all FastQC reports was performed with MultiQC (Ewels *et al.*, 2016). The paired-end reads of each sample were mapped to *S. lycopersicum* genome version SL3.0 (https://solgenomics.net/organism/Solanum_lycopersicum/genome) with the splice junction mapping aware software HISAT version 2.1.0 (Kim *et al.*, 2015) with default parameters and with the non-deterministic option. Mappings were sorted with

Samtools version 1.3 (Li *et al.*, 2009). Mapping results summaries were analyzed with an R in-house script (https://github.com/cesaremov/Solanum_Arcanum-Lycopersicum; R Development Core Team, 2008). Fragment quantification assignment to exons per gene ID was performed with FeatureCounts version 1.6.2 (Liao *et al.*, 2014). Quantification results were analyzed and processed with an R in-house script to obtain counts and counts per million (cpm) tables. A principal component analysis was performed using cpm tables as input to *prcomp* function from R stats package (Le *et al*., 2008). A counts table was used as input to the differential expression analysis with the R package EdgeR (Robinson *et al.*, 2010). The counts table was filtered to keep only transcripts with cpm>=4 in at least 9 samples. Dispersion was estimated with the EdgeR function *estimateDisp* with default parameters. Subsequently, the data were fitted to the negative binomial distribution with the Edge R function *glmQLFit* using default parameters. We compared the data of *S. arcanum* from 0 to 8 hpi with the data of *S. lycopersicum* from the same time transition; and data of *S. arcanum* from 8 to 24 hpi with the data of *S. lycopersicum* from the same time transition. The EdgeR function *glmQLMFTest* was used to compare between conditions. Multiple testing correction was performed using the *p.adjust* function from R base.

To asses Gene Ontology (GO) term enrichment (Ashburner *et al.,* 2000), a logistic classification model defined by the Generalized Linear Model (GLM) was employed. The *glm* function was used from the R stats package. The formula provided to the GLM was *index ~ statistics*, where the index is a boolean index indicating if genes belong to GO term. The statistics were defined by the *score=sign(logFC)*-log10(p.value*) for each species*.* Weights in GLM were defined as the ratio of genes not in GO term to genes in GO term. The link function was set to *logit.* To asses GO term enrichment when species are competing, we used GLM as previously mentioned, but using both species scores in the same model. The formula for the model with species competing was: *index ~ statistic_S._arcanum + statistic_S._lycopersicum*. Each statistic is defined as before for each species with scores of each species in the same model.

Map to the semi *de novo* transcriptome

In this pipeline, the paired RNA-Seq reads not mapped to genome were assembled. All unmapped paired reads were gathered together into one single sequence file. These unmapped reads were assembled using Trinity version 2.6.6 with default parameters to obtain unmapped transcripts (Grabherr *et al.*, 2011). The spliced junction mapped fragments were used (see first pipeline) to assembly mapped transcripts with StringTie (Pertea M *et al.*, 2015). Unmapped and mapped assembled transcripts were combined to get a unique *semi de novo* transcriptomic reference. Coding regions were predicted using Transdecoder (https://github.com/TransDecoder/TransDecoder/wiki). This semi *de novo* assembly was functionally annotated by using the *autoTrinotate.pl* program provided by Trinotate tool (https://github.com/Trinotate/Trinotate.github.io/wiki) with the programs Blast and Hmmer (Finn *et al.*, 2011). The databases SwissProt (https://web.expasy.org/docs/swiss-prot_guideline.html) and Pfam (Punta *et al.*, 2012) were used for the functional annotation. Each transcript could have one or more of the following functional annotations: blastx, blastp, Pfam, eggnog or gene ontology. All RNA-Seq paired-end reads for each sample were mapped to this semi *de novo* transcriptomic reference using Kallisto 0.42.4 (Bray *et al.*, 2016). The mapping results were analyzed as in pipeline one. We used Kallisto pseudo-counts (TPM) as input to differential expression analysis with EdgeR R package. Subsequently, the same steps from the first pipeline to perform differential expression analysis.

Map to the *de novo* transcriptome

All forward and reverse RNA-Seq paired fastq files were gathered to get only a pair of forward and reverse fastq files using in-house programs. This pair was used as input to Trinity version 2.6.6 (Grabherr *et al.*, 2011) with default parameters and minimum transcript length of 500 nucleotides. The result of Trinity was a *de novo* transcriptome assembly, which was used as reference in this third pipeline. Functional annotation was performed in the same way as in the second pipeline. The paired-end reads were mapped to this *de novo* transcriptome using Kallisto version

0.42.4 with default parameters (Bray *et al.*, 2016). As in the previous two pipelines, in-house R programs were applied to analyze mapping results. As mentioned in pipeline two, we got Kallisto quantifications for *de novo* (third) pipeline with in-house R programs. The same steps from first pipeline were followed to perform differential expression analysis.

2.5. Experimental approach to find transcripts showing up regulation in *S. arcanum* located at the regions spanning QTLs associated with resistance

The regions spanning the QTLs in chromosomes 5 [44803914-52431431], 7 [1319683-1554620] and 9 [1753139-2,313,950] published on Sen (2014), were used as a database from the *S. lycopersicum* reference genome SL3.0 to do a blast search against the transcripts obtained from the *de novo* assembly of *S. arcanum* reads.

2.6. Gene expression validation with qRT-PCR analysis

Total RNA from infected plants was used for the synthesis of single-stranded cDNA using Superscript II reverse transcriptase and oligo dT (Invitrogen, Carlsbad, CA). The cDNA samples were amplified with gene-specific primers using Primer 3.0 (Untergasser *et al.*, 2007). Possible dimers and heterodimers were predicted with Beacon designer tool (Premier Biosoft International, Palo Alto, CA; http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1). Possible amplicon secondary structures were detected with UNAFold tool (IDT Company, Coralville, IA, USA; https://www.idtdna.com/site/account/login?returnurl=%2FUNAFold). Quantification was performed on an Applied Biosystems, 7500 fast real-time-PCR system version 2.0 (Applied Biosystems, Foster City, CA, USA). The biological replicates of each condition were pooled together. The qRT-PCR reaction was carried out in 96-well plates, the thermal cycling was 94°C, 5 min; 40 cycles of 94°C, 25 s; 60°C, 30 s. Experiments were carried out using fast SYBR green Master Mix (Applied Biosystems, Foster City, CA, USA). Expression levels were normalized to ACTIN1 housekeeping gene. Fold changes in RNA expression was determined using threshold cycles by ΔΔCT method (Livak and Schmittgen 2001). For each

experiment, the qRT-PCR was performed in triplicates, and the present values represent average values and standard deviation.

3. Results

3.1. Infection of tomato plants with *Cmm* 1569

To obtain the genetic material for the analysis of *S. lycopersicum* and *S. arcanum* LA2157 during the interaction with *Cmm* 1569, we collected leaves of infected plants at 0, 8 and 24 h post-inoculation (hpi). We followed the disease development symptoms in both tomato species. *S. lycopersicum* infected plants developed the characteristic disease symptoms at 25 days post-inoculation (dpi). The symptoms were unilateral wilting of leaves and stem canker extended far from the inoculation site; whereas mock-inoculated *S. lycopersicum* control plants had normal stems and leaves (Figure 3a-d). In the resistant species *S. arcanum* LA2157, the infected plants did not show these typical disease symptoms. Instead, the inoculated plants only showed a small canker wound located in the inoculation site of each plant that did not spread through the stem, therefore, the phenotype of the *S. arcanum* LA2157 infected plants was very similar to the control plants (Figure 3e-h). In summary, both tomato species displayed the expected phenotype after *Cmm* inoculation.

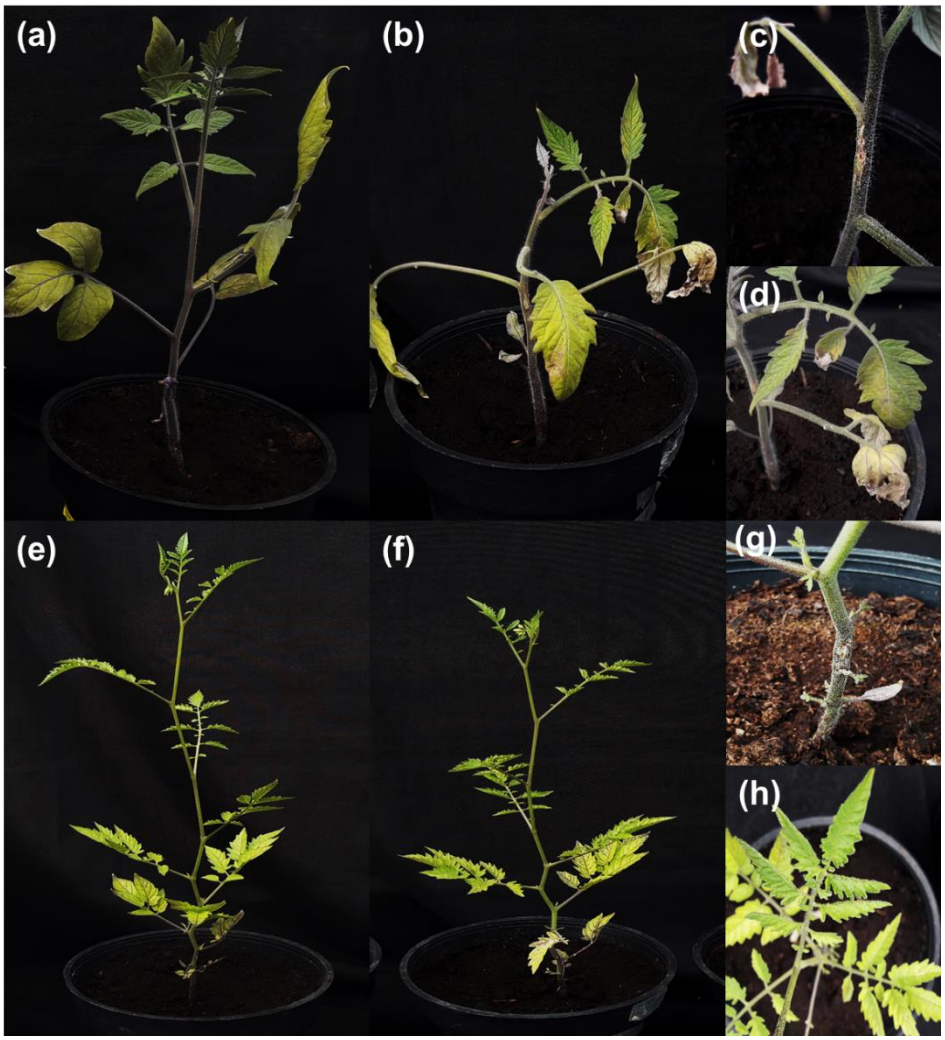


Figure 3. Development of disease symptoms in *S. lycopersicum* cv. Ailsa Craig and *S. arcanum* LA2157 infected with the *Cmm* 1569 strain. a) *S. lycopersicum* plant 25 days after mock-inoculation. b) *S. lycopersicum* 25 dpi, the plant showed the characteristic disease symptoms of bacterial canker. c,d) *S. lycopersicum* 25 dpi, leaves showing unilateral wilt and stem showing canker development. e) *S. arcanum* plant 25 days after mock-inoculation. f) *S. arcanum* 25 dpi. g,h) *S. arcanum* 25 dpi, leaves and stem without disease symptoms.

3.2. Mapping stats and transcript quantification

We sequenced 18 libraries resulting in fastq files ranging from 32-58 million reads of 150bp each. Considering the lack of a good quality reference genome for *S. arcanum* LA2157, we decided to analyze the sequencing data using three different strategies, which we will refer to as, MR, STA, and DA (see Materials and methods). Simultaneously, these strategies, allowed us to develop a comparative analysis

among the three strategies and to obtain a consensus result suitable for detecting genes potentially related to a defense response against *Cmm*.

Since *S. lycopersicum* and *S. arcanum* are phylogenetically relatively close species (Rodriguez, *et al*., 2009), the first strategy was to align the reads of both tomato species to the reference genome *S. lycopersicum* SL3.0 (Tomato Genome Consortium, 2012), (MR strategy; Figure 4). As a second strategy, the un-mapped reads resulting from the MR strategy were *de novo* assembled and concatenated to the coding sequences extracted from the reference genome, generating a new reference used for re-mapping all reads (STA strategy; Figure 4). Finally, the total reads of *S. lycopersicum* as well as those from *S. arcanum* were used to perform a *de novo* transcriptome assembly (DA strategy; Figure 4).

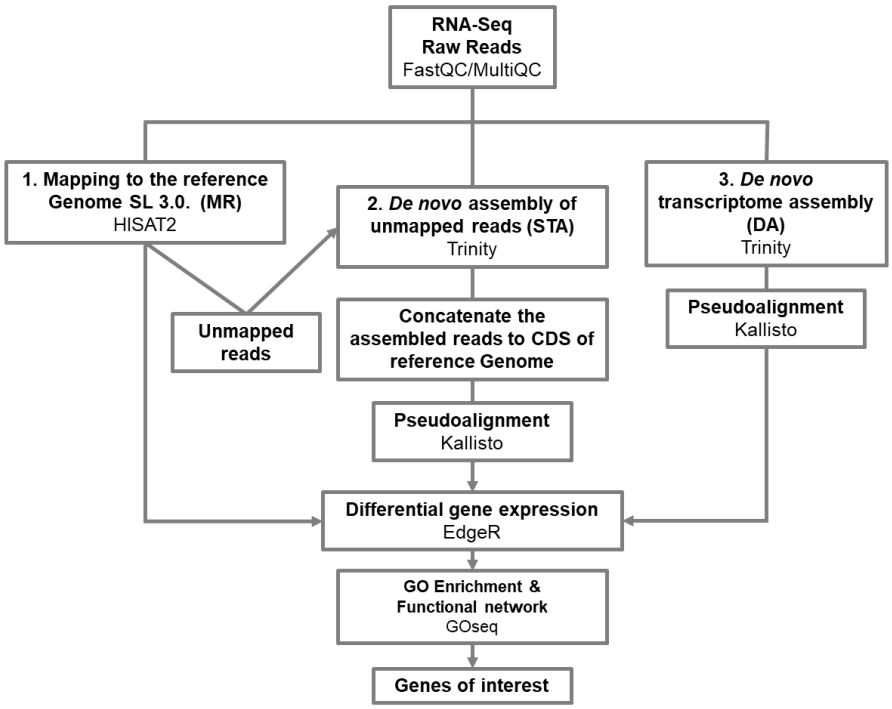


Figure 4. RNA-Seq data processing pipelines. The scheme summarizes the three strategies for data analysis: direct mapping to the reference genome of *Solanum lycopersicum* SL3.0, MR, (left), *de novo* assembly of unmapped reads, STA, (center), and *de novo* transcriptome assembly, DA, (right).

For the MR strategy, the percentage of mapped reads varied depending on the library, but in average 77.3% of *S. arcanum* LA2157 and 92.4% of *S. lycopersicum* reads mapped to the reference genome (mapped to exonic, intronic, intergenic and unassigned regions), resulting in a small percentage of unmapped reads to the reference (Figure 5a). In order to determine the sample similarity, we performed a principal component analysis (PCA) using as input the transcripts per million matrix obtained in each strategy (count matrix adjustment by Z-score function). The biological replicates grouped among them in each study condition (i.e., 0, 8 and 24 hpi; Figure 5b). In the STA strategy, the average percentage of mapped reads for *S. arcanum* species was 75.4% and 78.6% for *S. lycopersicum* (Figure 5c). The mapping percentage of the *S. arcanum* libraries slightly decreased with this strategy, and noticeably diminished for the *S. lycopersicum* ones. This last decrease could have occurred since we only considered the coding regions from the reference genome and the *de novo* assembled transcripts from unmapped reads for the STA strategy. The PCA obtained with this strategy presented the same distribution patterns for each group of biological replicates compared to the MR strategy (Figure 5d). With the DA pipeline, the average mapping percentage was 83.6% for *S. arcanum* LA2157 cDNA libraries and 86.2% for *S. lycopersicum* libraries (Figure 5e). The grouping patterns in the PCA with the DA strategy were very similar to the other two strategies (Figure 5f). Therefore, the application of either mapping approach did not affect the sample similarity of biological replicates. In parallel, we took advantage of the performed analyses to evaluate the computing time required for each strategy (we employed a cluster with 18 FAT nodes and 400 GB in RAM with 32 processors; Table 1). For the *de novo* transcriptome, the assembly took ~279 h compared to ~25 hours for the semi *de novo* transcriptome assembly and ~3 h for the mapping to the reference genome strategy (Table 1), pointing to considerable differences in the computing time required for each approach.

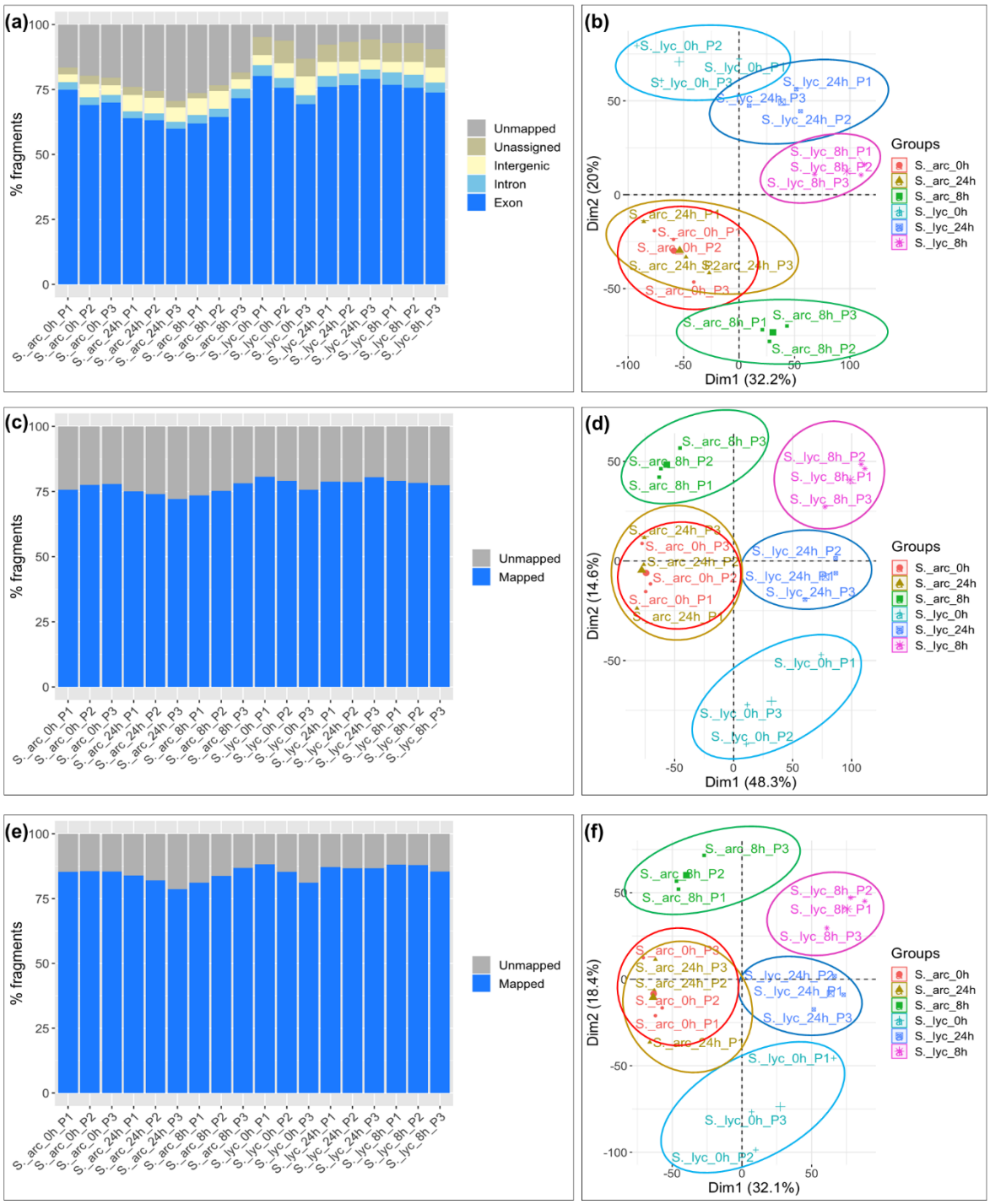


Figure 5. Mapping statistics of *S. arcanum* LA2157 and *S. lycopersicum* cDNA libraries and sample similarity determined by principal component analysis (PCA). a,b) Reads mapped to the *S. lycopersicum* SL3.0 reference genome (MR) and PCA. The average percentage of mapped reads was 77.3% for *S. arcanum* and 92.4% for *S. lycopersicum*. c,d) Mapping of sequence reads to the de novo assembled unmapped reads plus exonic regions of reference genome SL3.0 (STA) and PCA. The average percentage of mapped reads was 75.4% for *S. arcanum* and 78.6% for *S. lycopersicum*. e,f) Mapping of sequence reads to the de novo transcriptome assembly (DA) and PCA. The average

percentage of mapped reads was 83.3% for *S. arcanum* and 86.2% for *S. lycopersicum*. PCAs showed that biological replicates grouped correctly by study time (0, 8, and 24 hpi, accordingly) in all approaches. Count matrix adjusted by Z-score.

Table 1. Average computing time consumed per strategy.

Approach	Assembly	Annotation	Mapping & quantification (per library)	Total time
Mapping to the reference genome SL3.0	---	---	~ 3 h	~ 3 h
Semi *de novo* assembly	~ 7 h	~ 15 h	~ 3 h	~ 25 h
***De novo* assembly**	~ 240 h	~ 38 h	~ 1 h	~ 279 h

Note: This analysis was carried out in a cluster (Thubat Kaal 2.0), with 18 FAT nodes and 400 GB in RAM with 32 processors.

3.3. Differentially expressed genes in response to *Cmm* 1569

We determined the number of differentially expressed genes (DEGs) in each strategy of analysis. The statistical significance was evaluated by establishing a False Discovery Rate (FDR) limit of 0.01 for data mapped to the reference genome SL3.0 and an FDR value of 0.1 for data mapped to the semi *de novo* and the *de novo*-assembled transcriptomes. The 0.1 FDR value was established considering that the reads mapped to 77,515 and 49,340 transcripts for the STA and DA approaches, respectively (Table 2), even though there are about 35,000 genes annotated in the tomato genome (one transcript per gene). Therefore, the reads could be assigned to several transcript isoforms coming from the same gene, and a restrictive FDR cutoff could result in the assignment of less DEGs. Table 3 provides the number of DEGs detected in each analysis. By setting up these FDR thresholds, we obtained comparable numbers of differential transcripts in all three strategies.

Regarding the mapping of reads to the reference genome SL3.0 of the *S. lycopersicum* data, 1,421 induced and 1,483 repressed genes were detected in the 0 to 8 hpi transition, and 295 induced and 245 repressed in the 8 to 24 hpi transition.

With respect to the *S. arcanum* LA2157 data, 1,100 induced and 997 repressed genes were found in the 0 to 8 hpi transition; and 884 induced and 679 repressed genes in the 8 to 24 hpi transition. With reference to the reads mapping in the STA strategy, 1,602 induced and 2,249 repressed genes were found in the 0 to 8 hpi transition in the *S. lycopersicum* data; 500 induced and 277 repressed genes in the 8 to 24 hpi transition. In *S. arcanum*, 1,687 induced and 1,481 repressed genes were disclosed in the 0 to 8 hpi transition; 1,315 induced and 1,356 repressed genes were detected in the 8 to 24 hpi transition. Finally, with the DA pipeline, 1,409 induced and 1,968 repressed genes were detected in the 0 to 8 hpi transition in *S. lycopersicum*; and 419 and 247 up and down regulated genes, respectively, in the 8 to 24 hpi transition. The *S. arcanum* data exhibited 1,229 induced and 1,244 repressed genes in the 0 to 8 hpi transition; 1,152 induced and 908 repressed genes in the 8 to 24 hpi transition. The entire dispersion patterns of DEGs are shown in Figure 6. The average percentage of up-regulated genes from 0 to 8 hpi in the three approaches was similar in *S. lycopersicum* and *S. arcanum* (7.46% compared to 6.57%, respectively); meanwhile the average percentage of down-regulated genes in *S. lycopersicum* was higher compared to *S. arcanum* species at this same time (9.28% compared to 6.08%, respectively). In regard to the 8 to 24 hpi transition time, the average percentage of induced and repressed genes was higher in *S. arcanum* compared to *S. lycopersicum* (Table 3).

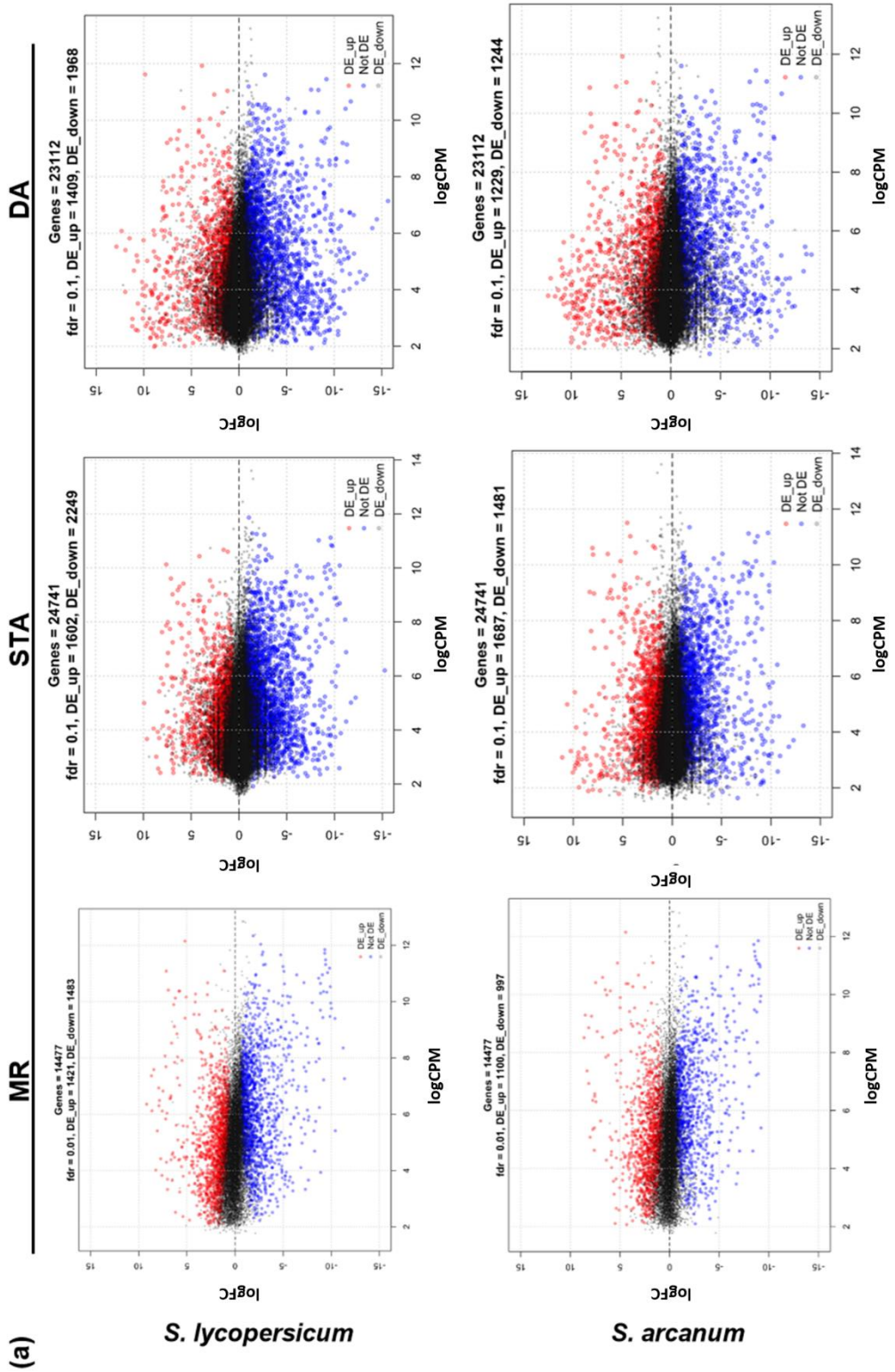
Table 2. Main features of generated transcriptomes.

Approach	Total transcripts	%GC content	N50	N90	Total bases	Predicted transcripts	Predicted transcripts per gene
Mapping to the reference genome SL3.0	35768	39.12	2226	801	54440110	0	1
Semi *de novo* assembly	77515	39.47	1157	581	166329048	123697	2.057214733
***De novo* assembly**	49340	38.62	2031	829	286104977	176548	3.578192136

Table 3. Differentially expressed genes identified in each strategy**.**

Plant species	0 – 8 hpi		8 – 24 hpi		Genes/Transcripts
Reference Genome SL3.0	**Up**	**Down**	**Up**	**Down**	**Total**
S. lycopersicum	1421	1483	295	245	14,477
S. arcanum LA2157	1100	997	884	679	FDR 0.01
Semi *de novo* assembly					
S. lycopersicum	1602	2249	500	277	24,741
S. arcanum LA2157	1687	1481	1315	1356	FDR: 0.1
***De novo* assembly**					
S. lycopersicum	1409	1968	419	247	23,112
S. arcanum LA2157	1229	1244	1152	908	FDR: 0.1
Average percentage of DEGs					
S. lycopersicum	7.46	9.28	1.95	1.29	
S. arcanum LA2157	6.57	6.08	5.46	4.69	
p-value	1.0	0.1	0.1	0.1	

Up: induced genes; Down: repressed genes


(a)
MR
STA
DA
S. lycopersicum
S. arcanum
Genes = 14477
fdr = 0.01, DE_up = 1421, DE_down = 1483
Genes = 14477
fdr = 0.01, DE_up = 1100, DE_down = 997
Genes = 24741
fdr = 0.1, DE_up = 1602, DE_down = 2249
Genes = 24741
fdr = 0.1, DE_up = 1687, DE_down = 1481
Genes = 23112
fdr = 0.1, DE_up = 1409, DE_down = 1968
Genes = 23112
fdr = 0.1, DE_up = 1229, DE_down = 1244
DE_up
Not DE
DE_down
logFC
logCPM

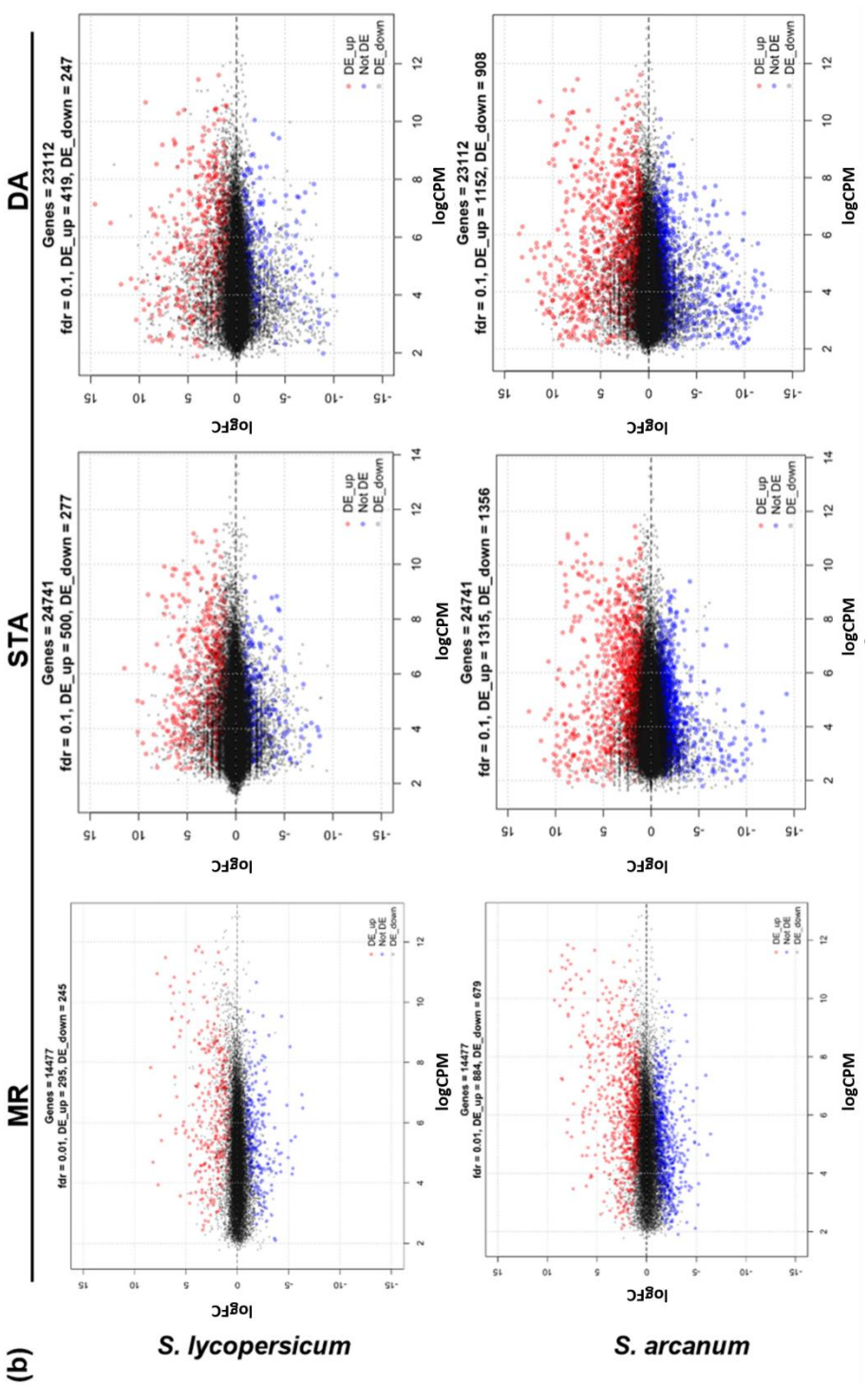


Figure 6. MA plots of DEGs per analysis approach. FDR 0.01 for MR analysis; FDR 0.1 for STA and DA analysis approaches. a) Differentially expressed genes dispersion from 0 to 8 hpi. b) Differentially expressed genes dispersion from 8 to 24 hpi. Red dots: repressed genes; blue dots: induced genes. x-axis = mean average abundance (Log_2CPM, counts per million); y-axis = log ratio (Log_2FC).

Differences in induced and repressed genes in *S. arcanum* and *S. lycopersicum* could be seen as a starting point to search for genes involved in a defense response to *Cmm*. We detected shared genes between *S. lycopersicum* and *S. arcanum* LA2157 considering the DEGs and transition times (Figure 7). With regard to induced genes in *S. arcanum* and repressed in *S. lycopersicum*, from 28 to 54 genes were detected from 0 to 8 hpi, indicating a similar number of genes in all strategies of analysis (Figure 7a, c, e). Whereas, from 8 to 24 hpi for shared induced genes in *S. arcanum* and repressed in *S. lycopersicum,* only one gene was detected by the DA strategy and four genes by the STA one (Figure 7b, d, f). Depending on the strategy, from 528 to 976 induced genes were detected in *S. arcanum* from 0 to 8 hpi (Figure 7a, c, e). At the same time, a similar number of induced genes was detected in *S. lycopersicum* only, around 900 genes in each analysis. Relative to the 8 to 24 hpi transition, between 689 and 1081 genes were induced in *S. arcanum* solely, and only from 100 to 267 in *S. lycopersicum* (Figure 7b, d, f). In regard to induced genes shared in both tomato species from 0 to 8 hpi, we detected from 476 to 657 genes per strategy (Figure 7a, c, e). As from 8 to 24 hpi, around 200 genes were detected in each approach (Figure 7b, d, f). A similar number of repressed genes shared in both species can be visualized in the Figure 7. Importantly, a higher number of repressed genes was detected solely in *S. lycopersicum* compared with *S. arcanum* from 0 to 8 hpi (Figure 7a, c, e), and this behavior was opposite from 8 to 24 hpi (Figure 7b, d, f), which suggests a downregulation provoked by *Cmm* in *S. lycopersicum*. Remarkably, the overall results presented the same tendencies regardless of the used strategy.

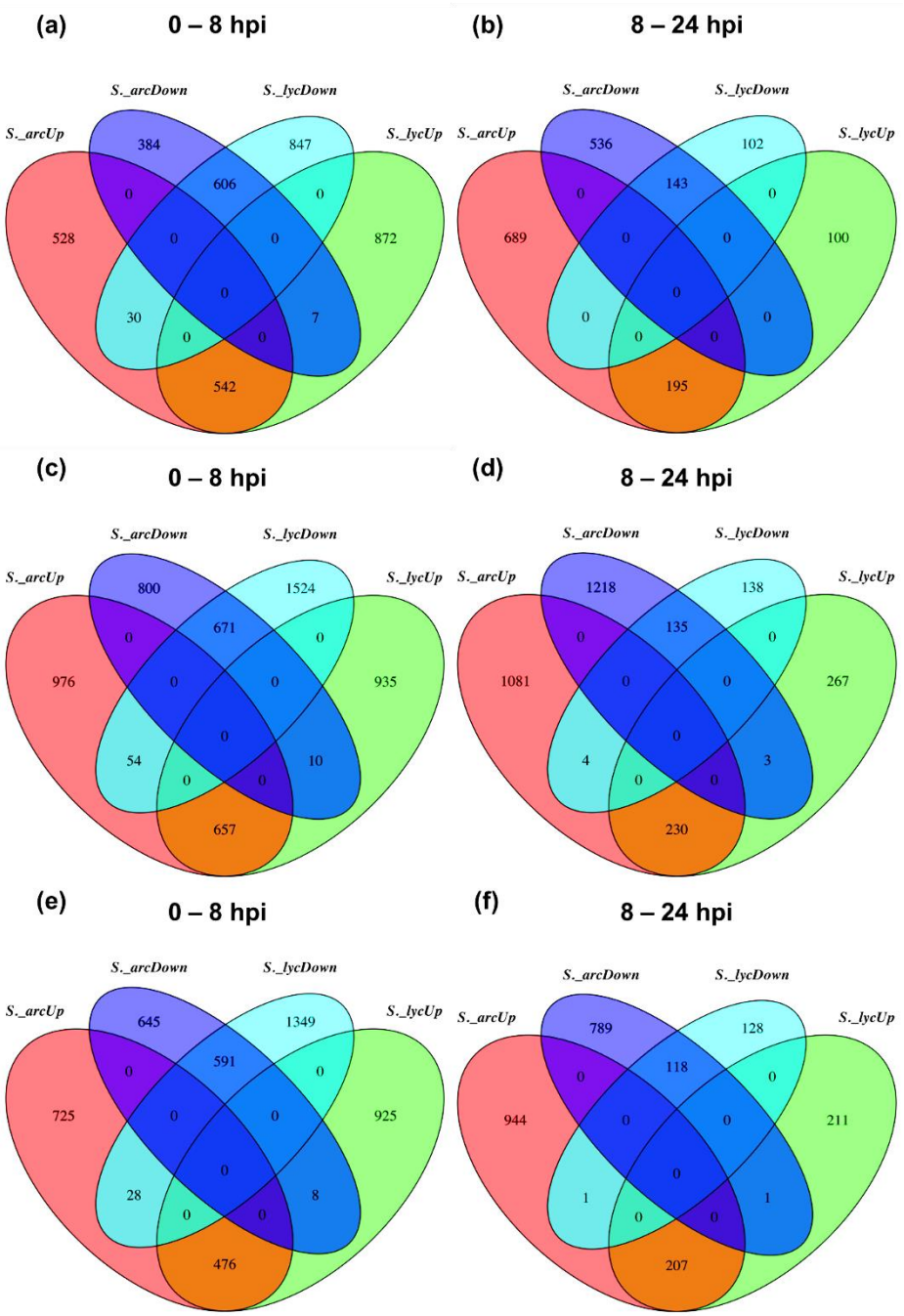


Figure 7. Venn diagrams showing DEGs contrasts between *S. lycopersicum* cv. Ailsa Craig and *S. arcanum* LA2157. a,b) Contrasts of DEGs from 0 to 8 hpi transition and 8 to 24 hpi transition. Approach: Mapping of reads to the reference genome SL3.0 (MR); FDR value 0.01. c,d) Contrasts of DEGs from 0 to 8 hpi transition and 8 to 24 hpi transition. Strategy: Mapping to the semi *de novo* assembled transcriptome (STA); FDR value 0.1. e,f) Contrasts of DEGs from 0 to 8 hpi transition and 8 to 24 hpi transition. Strategy: Mapping to the *de novo* assembled transcriptome (DA); FDR value 0.1.

3.4. Gene Ontology term enrichment analisys in *S. lycopersicum* and *S. arcanum* LA2157

In order to determine the global transcriptional context of both tomato species in response to *Cmm*, we performed a Gene Ontology (GO) enrichment analysis considering the Log Fold Change (LogFC) values and -log10 FDR values of DEGs. After extracting the enriched groups of each approach, we selected enriched GO terms that could be directly associated to an infection response caused by *Cmm*. In addition, enrichment scores of the selected functional groups of *S. lycopersicum* and *S. arcanum* were compared in order to detect which tomato species exhibited the higher enrichment.

Several functional groups were enriched in both *S. lycopersicum* and *S. arcanum* in the three strategies of analysis, including hydrolase activity (GO:0016787), proteasome complex (GO:0000502) and response to cytokinin (GO:0009735). However, they were enriched differently either by induced or repressed genes, depending on the tomato species and time of study (Figure 8). Similarly, we detected functional groups only enriched in one of the three strategies of analysis, such as nuclear proteasome complex (GO:0031595) and threonine type endopeptidase activity (GO:0004298), which were enriched only when we mapped the reads to the *de novo* assembled transcriptome (Figure 8c). Thus, we observed that some functional groups were not enriched using the MR and STA strategies, and could be only appreciated employing the DA approach.

Importantly, the more general functional group defense response (GO:0006952), was enriched in the three strategies of analysis. This functional group exhibited a higher enrichment by induced genes in *S. lycopersicum* from 0 to 8 hpi in all cases (Figure 8). However, the more specific functional groups defense response to bacteria (GO:0042742) as well as defense response to fungus (GO:0050832) showed a stronger enrichment by induced genes in *S. arcanum* from 0 to 8 hpi. This behavior was consistent in all three analysis approaches (Figure 8).

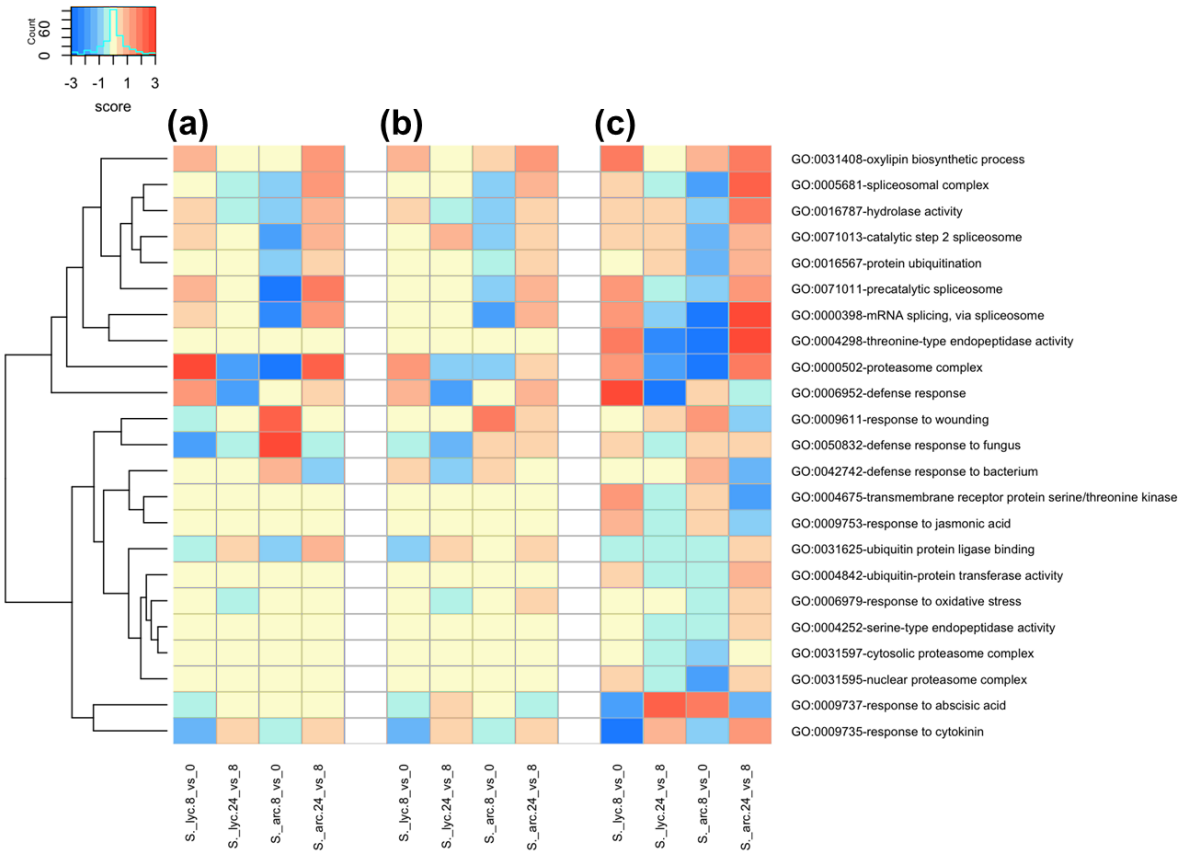


Figure 8. GO term enrichment analysis. Heatmaps representing the main enriched functional groups in each strategy of analysis, contrasting the enrichment level of each species. a) Mapping to the reference genome SL3.0 (MR). b) Mapping to the semi *de novo* assembled transcriptome (STA). c) Mapping to the novo assembled transcriptome (DA). The score values were obtained using the LogFC and the –log10 of FDR of DEGs.

3.5. DEGs presumably associated to the defense response against *Cmm* 1569

Using the results of all applied strategies, we aimed to obtain a list of genes potentially related to defense against *Cmm*. To do this, first we detected relevant genes in each approach considering the LogFC values, the comparisons (shared and unshared genes between the two tomato species), the GO term enrichment analysis, and the gene functional annotation. Then we compiled the genes in a single list (Table 4). Genes that appeared in more than one approach (same gene identifier) were arbitrarily reported just in one method to avoid redundancies.

We identified DEGs induced only in *S. arcanum* from 0 to 8 hpi, such as genes that codify a Polyphenol oxidase E (Solyc08g074620.3), an E3 ubiquitin protein-ligase (Solyc12g049330.2), a Putative late blight resistance protein (Solyc05g012900.3), and a Wall-associated receptor kinase-like (TRINITY_DN16553_c1_g3_i1). The last one also appeared from 8 to 24 hpi, but repressed (Table 4). Likewise, induced genes in *S. arcanum* and repressed in *S. lycopersicum* were detected from 0 to 8 hpi, such as a gene that codifies a Rust resistance Kinase Lr10 (Solyc02g081500.3; which is repressed in *S. arcanum* from 8 to 24 hpi), a Leucine-rich repeat receptor like serine/threonine protein kinase (Solyc11g017270.2), an unannotated gene with function related to Serine type endopeptidase inhibitory activity (TRINITY_DN15795_c0_g2_i2), an Arkynin repeat-containing protein (Solyc08g062360.3; later repressed in *S. arcanum* from 8 to 24 hpi), an L-type lectin-domain containing receptor kinase (TRINITY_DN22346_c2_g3_i1; later induced in *S. lycopersicum* from 8 to 24 hpi), a Mitogen-activated protein kinase 3 (TRINITY_DN15770_c0_g1_i7), and others (Table 4). Also, induced genes in both tomato species were detected from 0 to 8 hpi. However, some of them were strongly expressed in *S. arcanum* (Table 4), as is the case of a gene that codifies a Probable receptor-like serine/threonine protein kinase (Solyc01g109590.3; which is repressed in *S. arcanum* from 8 to 24 hpi), and a Class V chitinase (Solyc07g005090.3). We have also identified the top ten induced genes in *S. arcanum* with no transcripts assigned to *S. lycopersicum* detected with the DA approach (Table 5). Some of these induced genes may have an important role during the early defense response to *Cmm*, such as the genes that encode for a Bidirectional sugar transporter SWEET2a (TRINITY_DN13677_c0_g1_i8) or the Aquaporin PIP2-1 (TRINITY_DN15481_c0_g1_i3). Definitely, our study hosts many other genes of interest that can be identified by comparison of the two tomato species transcriptomes.

Table 4. Selected DEGs potentially related to resistant response against *Cmm*.

Gene/transcript ID	Description	*S. arc* LogFC (0-8h)	*S. lyc* LogFC (0-8h)	*S. arc* LogFC (8-24h)	*S. lyc* LogFC (8-24h)
Reference genome					
Solyc06g074030.1	CCR4-associated factor 1 homolog 9	3.14*	-3.85*	-2.16	-1.07
Solyc06g071280.3	Protein EDS1B	2.32*	-1.66*	-1.73*	-0.46
Solyc02g081500.3	Rust resistance kinase Lr10	2.62*	-1.87*	-1.80*	-0.23
Solyc02g092140.1	Protein PHLOEM PROTEIN 2-LIKE A10	5.37*	1.49*	-2.96*	-1.47
Solyc08g074620.3	Polyphenol oxidase E	5.65*	-0.38	0.07	3.03
Solyc05g051530.3	ABC transporter G family member 11	5.48*	-0.04	-4.90*	-2.55
Solyc11g017270.2	Leucine-rich repeat receptor-like serine/threonine-protein kinase	1.60*	-1.20*	-0.53	0.71
Solyc08g080830.3	Probable LRR receptor-like serine/threonine-protein kinase	1.05*	0.48	-0.69*	-0.23
Solyc12g009140.2	Proteasome subunit alpha type-6-B	-0.83*	-0.31	0.84*	0.22
Solyc01g109590.3	Probable receptor-like serine/threonine protein kinase	2.65*	1.58*	-2.16*	-1.09
Semi *de novo* Assembly					
TRINITY_DN15795_c0_g2_i2	Serine-type endopeptidase inhibitor activity	6.19*	-1.30*	0.97	0.04
Solyc08g062360.3	Ankyrin repeat-containing protein	2.14*	-2.67*	-2.70*	-0.88
Solyc06g066370.3	WRKY transcription factor WRKY24	1.33*	-1.38*	-0.96	0.51
Solyc12g049330.2	E3 ubiquitin-protein ligase SPL2	6.65*	-0.05	-1.15	-0.06
Solyc05g012900.3	Putative late blight resistance protein homolog R1B-23	5.36*	-1.62	-1.03	0.85
Solyc07g005090.3	Class V chitinase	2.84*	1.65*	0.83	0.07
Solyc06g083390.3	RPM1-interacting protein 4	1.29*	-1.55*	-0.74	-0.31
Solyc08g080830.3	Probable LRR receptor-like serine/threonine-protein kinase	0.98*	0.48	-0.52	-0.29
Solyc10g008010.3	Proteasome subunit alpha type-2	-0.67	0.10	1.02*	-0.09
Solyc12g007110.2	Proline-rich receptor-like protein kinase PERK1	1.36*	0.44	-0.81*	0.49
***De novo* Assembly**					
TRINITY_DN22346_c2_g3_i1	L-type lectin-domain containing receptor kinase	2.44*	-2.08*	0.91	2.82*
TRINITY_DN19315_c2_g1_i2	E3 ubiquitin-protein ligase RMA1H1	2.94*	1.84	-0.75	0.89
TRINITY_DN17399_c2_g1_i1	NDR1/HIN1-like protein 2	4.54	-3.19*	-0.57	-1.05
TRINITY_DN16733_c1_g1_i7	Annexin D4	2.72*	-5.55*	-1.32	1.74
TRINITY_DN15770_c0_g1_i7	Mitogen-activated protein kinase 3	2.12*	-1.74*	-0.54	-0.35
TRINITY_DN16314_c0_g3_i2	NAC domain-containing protein 2	2.30*	-2.04*	-0.56	1.16
TRINITY_DN22670_c0_g2_i2	Leucine-rich repeat receptor-like serine/threonine-protein kinase	1.67	-6.48*	-0.51	5.47
TRINITY_DN16553_c1_g3_i1	Wall-associated receptor kinase-like 20	3.20*	0.86	-1.85*	-0.05
TRINITY_DN20729_c2_g1_i1	Protein EDS1	2.09*	-1.58*	-1.83*	-0.44
TRINITY_DN14123_c3_g1_i2	Probable receptor-like protein kinase	1.75*	1.26*	-2.12*	-0.72

LogFC values of genes differentially expressed are highlighted with an asterisk*

Table 5. Top ten genes induced in *S. arcanum* with no transcripts assigned to *S. lycopersicum* after DA approach.

Gene/transcript ID	Description	*S. arc* LogFC
0-8 hpi		
TRINITY_DN14010_c2_g1_i5	Tetraspanin-3	4.48
TRINITY_DN24943_c4_g1_i9	Linoleate 13S-lipoxygenase 2-1	3.88
TRINITY_DN18887_c0_g4_i2	UDP-glycosyltransferase 85A8	3.80
TRINITY_DN16833_c0_g1_i15	Unknown	3.44
TRINITY_DN13677_c0_g1_i8	Bidirectional sugar transporter SWEET2a	3.16
TRINITY_DN23205_c0_g6_i3	Ethylene-responsive transcription factor 1	2.95
TRINITY_DN23169_c1_g1_i1	Gamma aminobutyrate transaminase 2	2.78
TRINITY_DN19385_c0_g1_i7	Heavy metal-associated isoprenylated plant protein 32	2.63
TRINITY_DN18065_c0_g1_i1	Allene oxide synthase 2 chloroplastic	2.48
TRINITY_DN20602_c0_g1_i4	Ethylene-responsive transcription factor 5	2.45
8-24 hpi		
TRINITY_DN21371_c0_g1_i15	Stromal 70 kDa heat shock-related protein	6.39
TRINITY_DN16003_c0_g1_i7	Glyceraldehyde-3-phosphate dehydrogenase	3.67
TRINITY_DN15481_c0_g1_i3	Aquaporin PIP2-1	3.26
TRINITY_DN17153_c0_g1_i7	Heat shock cognate 70 kDa protein 2	3.24
TRINITY_DN23165_c1_g1_i17	RuBisCO large subunit-binding protein subunit alpha	3.15
TRINITY_DN24736_c3_g1_i8	Protein LNK1	2.74
TRINITY_DN17225_c0_g1_i11	Protein WVD2-like 7	2.7
TRINITY_DN23140_c2_g1_i7	Ferric reduction oxidase 6	2.65
TRINITY_DN25011_c3_g1_i11	Heat shock 70 kDa protein 15	2.52
TRINITY_DN25230_c3_g1_i22	Serine/arginine-rich splicing factor SC35	2.45

Several QTLs have been previously related to the resistance against *Cmm* (van Heusden *et al*., 1999; Francis *et al*., 2001; Kabelka *et al*., 2002). We traced the regions spanning the QTLs in the *S. lycopersicum* genome and we found 70 transcripts on chromosome 5, 42 on chromosome 7 and 2,479 on chromosome 9. Interestingly, some of these transcripts were highly expressed in *S. arcanum* only at early stages of infection (Table 6). However, that is not the case in *S. lycopersicum*, where transcripts seemed to have little induction or even down regulation. Moreover, the annotation of these transcripts correlates with previously published proteins involved in disease immunity.

Table 6. Transcripts in *S. arcanum* located at the regions spanning QTLs associated with resistance.

Chromosome of QTL spanning region	Annotation	Transcript ID	*S._arc* LogFC		*S._lyc* LogFC	
			0-8hpi	8-24hpi	0-8hpi	8-24hpi
Ch5	UPF0481 protein At3g47200	TRINITY_DN21279_c1_g1_i2	**6.90**	-1.33	0.35	-1.33
Ch5	Hyoscyamine 6-dioxygenase	TRINITY_DN19945_c3_g1_i7	**4.37**	0.64	0.65	0.64
Ch7	Diacylglycerol kinase 7	TRINITY_DN20884_c0_g1_i10	**5.82**	0.26	0.45	-0.34
Ch7	Putative disease resistance RPP13-like protein 1	TRINITY_DN22407_c0_g2_i3	**1.69**	-1.12	0.21	-0.30
Ch9	TOM1-like protein 6	TRINITY_DN22806_c1_g1_i2	**10.52**	-0.07	-0.74	0.45

3.6. RNA-Seq validation using DEGs potentially related to resistance

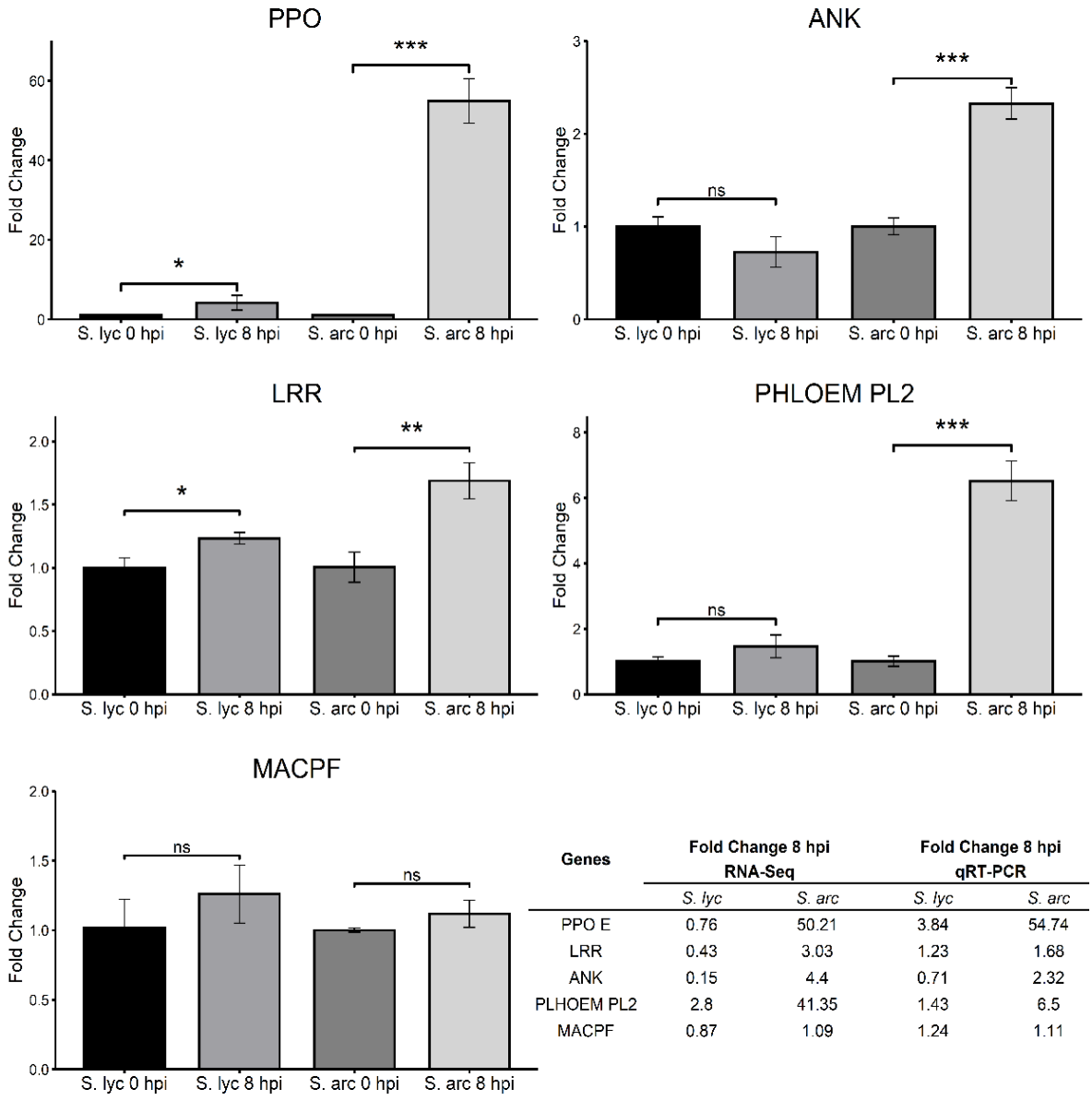
In order to validate the expression profiles of differentially expressed genes identified by the five bioinformatics analyses, we selected three genes potentially related to resistance against *Cmm*. We carried out a qRT-PCR analysis of samples obtained from independent infection assays in *S. lycopersicum* and *S. arcanum* plants (different infected plants from those used in the RNA-Seq experiment). We considered only the infection times of 0 hpi and 8 hpi, since our results suggested that the most notable expression changes occurred in this interval of time after infection. The selected genes were Polyphenol oxidase E (PPO E; Solyc08g074620.3), Leucine-rich repeat receptor-like serine-threonine-protein kinase (LRR; Solyc11g017270.2), the orthologous gene Ankyrin repeat-containing protein (ANK; Solyc08g062360.3.1), a Protein PHLOEM PROTEIN 2-LIKE A10 (PHLOEM PL2; Solyc02g092140.1), and an MACPF domain-containing protein CAD1 (MACPF; Solyc10g085710.2). The primers employed for qRT-PCR validation are shown in Table 7. With respect to the LRR gene, the expression levels measured by qRT-PCR were quite similar in both species, which varied slightly compared with the RNA-Seq obtained expression, since in *S. lycopersicum* the LRR expression was

reported as repressed by this latter method (Table 4). Even so, the expression level of LRR was higher in the wild tomato species (Figure 9). Results from the PPO E as well as the Protein PHLOEM PROTEIN 2-LIKE A10 (PHLOEM PL2; Solyc02g092140.1) gene indicated a clearer overexpression in *S. arcanum* LA2157 8 hpi compared with *S. lycopersicum* (Figure 9), which agreed with the RNA-Seq results (Figure 9; Table 4). In addition, the expression level detected for the gene that codifies for ANK also showed a higher accumulation in the *S. arcanum* species (Figure 9), not observed in *S. lycopersicum*, which also correlated with our RNA-Seq results (Figure 9; Table 4). According to our RNA-Seq results the MACPF domain-containing protein CAD1 (MACPF; Solyc10g085710.2) encoding gene was not differentially expressed neither in *S. lycopersicum* nor in *S. arcanum*, and the expression levels were alike to those shown in the qRT-PCR analysis (Figure 9). Overall, these results agreed with the transcriptome data analysis.

Table 7. List of primers employed for qRT-PCR validation.

Name	Primer Sequence (5´- 3')	Length (bp)	Amplicon size (bp)
PPO-E	F: TTGGTAAGGAAGTTGACACACC	22	187
	R: CAGATGTGAACCGGAGTATGAG	22	
LRR-At1g17230	F: GTTGCTGGTTCTGTCGGTTAT	21	183
	R: CCACTTGCTGTGCTTCTCAG	20	
Ankyrin	F: GGTAATGGACACAACATTCTGC	22	162
	R: TAGATCATTCACGTCCCTACGA	22	
PHLOEM PL2	F: GGAGAAATAGGAACTGGGTTGT	22	150
	R: CCACCATCTCAGCTAAAGAAAC	22	
MACPF	F: CTTACTCATCTCGGTCATGTCA	22	183
	R: CTTCTGTATTCAAGGCTACCAG	22	
ACT-TOM5	F: CCTCACCGAGAGAGGTTACATGT	23	61
	R: CATGTCGCGGACAATTTCC	19	

F: Forward primer; R: Reverse primer.



Genes	Fold Change 8 hpi RNA-Seq		Fold Change 8 hpi qRT-PCR	
	S. lyc	*S. arc*	*S. lyc*	*S. arc*
PPO E	0.76	50.21	3.84	54.74
LRR	0.43	3.03	1.23	1.68
ANK	0.15	4.4	0.71	2.32
PLHOEM PL2	2.8	41.35	1.43	6.5
MACPF	0.87	1.09	1.24	1.11

Figure 9. qRT-PCR analysis of selected DEGs. Genes analyzed: Polyphenol oxidase E (PPO E; Solyc08g074620.3), Leucine-rich repeat receptor-like serine-threonine-protein kinase (LRR; Solyc11g017270.2), Ankyrin repeat-containing protein (ANK; Solyc08g062360.3.1), Protein PHLOEM PROTEIN 2-LIKE A10 (PHLOEM PL2; Solyc02g092140.1) and MACPF domain-containing protein CAD1 (MACPF; Solyc10g085710.2). The expression levels were normalized to the actin expression (*ACTIN1*). Data represent fold change of gene expression 0 versus 8 hpi. The table compares the fold change values of RNA-Seq analysis versus fold change values from the qRT-PCR validation. The statistically significant differential gene expression between 0 hpi and 8 hpi was assayed by Student's t test; unpaired. ns: > 0.05. *: P ≤ 0.05. **: P ≤ 0.01. ***: P ≤ 0.001.

4. Discussion

We gained further insights about potentially resistance-related genes against *Cmm* as a result of the comparison of transcriptional changes in the resistant wild tomato species *S. arcanum* LA2157 and the susceptible species *S. lycopersicum* cv. Ailsa Craig. Additionally, mapping of RNA-Seq raw data of *S. arcanum* species to the *S. lycopersicum* reference genome is a reliable strategy in spite of the wild tomato species lacking a reference. Therefore, it is possible to use this MR approach for RNA-Seq analyses of tomato wild type species. Nevertheless, the combination of our three bioinformatics pipelines (MR, STA and DA) allowed us to obtain additional and complementary gene expression results.

4.1 From mapping to a reference genome to a *de novo* assembly transcriptome, a complementary approach

Transcriptomic studies in a not widely studied plant is always a challenge since there is not enough data available. That is the case of *S. arcanum* LA2157 wild tomato species where, at the time this research was conducted, no genome sequence was available for mapping. Usually, In an RNA-Seq analysis where there is no reference genome the first alternative is to perform a *de novo* transcriptome assembly employing the preferred transcriptome assembler tool, such as, Trinity, Oases, or SOAPdenovo-Trans (Grabherr *et al*., 2011; Schulz *et al*., 2012; Xie *et al*., 2014). However, the reference genome sequence of *S. lycopersicum,* which is relatively closed to *S. arcanum* (Rodriguez *et al.*, 2009) is available since 2012 (Sato *et al*., 2012). Considering this phylogenetic closeness, we decided to use the *S. lycopersicum* SL.3.0 reference genome for mapping the obtained reads of both tomato species (MR approach). Since this approach could deliver incomplete results, we decided to complement it and perform other two approaches, referred to as STA and DA (Figure 4).

A feature to highlight in an RNA-Seq analysis is the required computing time. As it has been reported, depending on the amount of reads, a *de novo* transcriptome assembly with Trinity requires long running time periods and high computing capacity (Grabherr *et al.*, 2011), which could be disadvantageous when the latter is

limited. According to our analysis, this approach consumed ~279 h compared with the MR approach which only consumed ~2 h of computing time (Table 1). As our data demonstrates, the analysis based on the semi *de novo* assembly approach is a good alternative for bioinformatics transcriptome analysis, since it required an intermediate computing time of ~25 h (Table 1). Additionally, it was able to produce a higher number of transcripts in comparison to the other two strategies (Table 2).

Regarding the MR approach, average mapping percentages of *S. arcanum* LA2157 libraries fluctuated from around 70 to 80%; about 15% less than the average mapping percentage obtained in *S. lycopersicum* (Figure 5a), which is a small difference; given they are different species. Furthermore, the amount of reads of both tomato species that correctly mapped to the reference genome varied from 70 to 90%, indicating good mapping proportions, which is in accordance with Sangiovanni *et al.*, (2019). Therefore, considering the quality of the reference genome SL3.0, our results indicate a proper mapping for *S. arcanum* LA2157. Compared to MR, the STA approach improved the average mapping in the wild tomato species (Figure 5c). A similar pattern is present when reads were mapped to the *de novo* assembled transcriptome (DA), with a higher mapping percentage of *S. arcanum* with respect to the STA (Figure 5e). As a consequence of the good performance obtained with the Trinity assembler (Honaas *et al.*, 2016; Hölzer and Marz, 2019), we were able to discover new specific features that are not shared between these two related tomato species or that were not available in the reference genome. These features can be found in the non-annotated genes that were only assigned to GO terms functions, which could be analyzed in future research. Overall, we suggest that technical barriers of MR could be compensated by comparing the results of the MR and STA approaches.

The PCA led us to determine that samples cluster by species and by study condition in all approaches (Figure 5). Additionally, the PCA indicates a cyclic pattern of the biological replicates, since it is possible to observe an overlap between the 0 hpi samples and 24 hpi samples of *S. arcanum*, not observed in *S. lycopersicum* (Figure 5). The above results suggest a faster global response of *S. arcanum* to the *Cmm*

infection in comparison to *S. lycopersicum*, which could have a delayed response as it has been previously proposed by Lara-Ávila *et al.*, (2012).

The total number of DEGs barely varied regardless of the used approach. The STA strategy showed more DEGs than the DA strategy (Table 3), and this can be explained because the STA transcriptome is larger (harbors information of annotated genes from the reference genome SL3.0 and additional transcripts obtained after the assembly of unmapped reads; Table 2). Interestingly, in all approaches we detected an approximately equal proportion of induced genes in both tomato species from 0 to 8 hpi (Table 3). Also, the proportion of repressed genes in *S. lycopersicum* compared to *S. arcanum* from 0 to 8 hpi was higher in all the three approaches (Table 3), which can be associated to the cyclic behavior observed in the PCA and the suggested delayed response (Figure 5). Moreover, after contrasting DEGs of *S. lycopersicum* and *S. arcanum*, gene groups are comparable in the three approaches; these groups are shared between species, whereas other genes are induced or repressed only in one species (Figure 7). This information is useful to select those that are up-regulated only in the *S. arcanum* species and perhaps play an important role in bacterial plant defense. Overall, these analyses indicate that the number of DEGs detected in each approach is in concordance and minor variations between one another are directly related to the number of genes/transcripts in the reference genome or assembled transcriptomes.

4.2. Global transcriptional profiling and evidence of resistance-related genes

According to GO term enrichment analysis, numerous functional groups revealed comparable patterns among MR, STA and DA approaches, and in *S. arcanum* and *S. lycopersicum*, suggesting that both tomato species share similarities in a global transcriptional context. Some of these enriched functions are directly related to plant defense (Figure 8). suggesting that both tomato species share important similarities in a global transcriptional context. For example, compared with the resistant species *S. arcanum*, the general defense response functional group exhibited a higher enrichment level in the susceptible species *S. lycopersicum* from 0 to 8 hpi. This result agrees with previous evidence which demonstrated that even when *S.*

lycopersicum is generating a defense response after sensing *Cmm*, it is not effective for counteracting the pathogen infection (Balaji *et al.*, 2008; Savidor *et al*., 2012).

Importantly, during the transition from 8 to 24 hpi the defense response group enrichment turned from induced genes to repressed genes in *S. lycopersicum*; while in *S. arcanum* this functional group was enriched mainly by induced genes at 24 hpi (Figure 8). These findings are in accordance with a recent comparative transcriptome analysis between two different *Solanum species* and *Cmm* interaction, where the defense response functional group was not found in the susceptible tomato species *S. lycopersicum* var. filinta, although this experiment was done at four and eight days after *Cmm* inoculation (dai). In contrast, this functional group was enriched 4 dai in the resistant tomato line *S. peruvianum* LA2157 (Basim *et al*., 2021). Moreover, the functional groups of defense response specific to bacteria, as well as defense response to fungi, were mainly enriched by induced genes from 0 to 8 hpi in *S. arcanum* (Figure 8), suggesting that *S. arcanum* is triggering a stronger and perhaps more specific early defense response in comparison to *S. lycopersicum* after the challenge.

In addition, Basim *et al*. (2021) reported a gene codifying for a TMV resistance protein N induced in *S. peruvianum* LA2157 at 4 dai. Interestingly it seems that this gene is also participating during the early defense response to *Cmm*, since we also found transcripts from TMV resistance protein N genes mainly induced in *S. arcanum* from 0 to 8 hpi. However, in our study we did not find the *Sn-2* gene that have been associated with a defensive role against *Cmm* (Balaji and Smart, 2012), probably, because these genes may have a relevant defensive role several days after the interaction of resistant wild tomato species with *Cmm*, as it has been recently showed in *S. peruvianum* LA2157 (Basim *et al*., 2021).

Although the results obtained by all strategies generally agreed, several functional groups showed enrichment in *S. arcanum* only in the STA and DA approaches. For example, in the MR approach, the oxylipin biosynthetic process was enriched only in *S. lycopersicum* from 0 to 8 hpi (Figure 8a). However, in the same transition time this functional group was enriched in *S. arcanum* together with *S. lycopersicum* when

we applied the STA and DA approaches. In contrast, response to wounding GO term, was enriched from 0 to 8 hpi in *S. arcanum* in all approaches, but it also appeared enriched in *S. lycopersicum* from 8 to 24 h in the DA approach only (Figure 8c). The enrichment of this group can be associated with the mechanical damage caused during *Cmm* inoculation. Several reasons may have contributed to the detection of enrichment of some functional groups only by STA and/or DA approaches. First, strategies employing assembly generated new transcripts and disclosed additional functional enrichment information. Second, since Trinity can predict transcript isoforms (Grabherr *et al.*, 2011), by using the DA approach high transcript abundance allowed the enrichment of particular functional groups that were not properly mapped to the reference genome SL3.0 in the MR approach. A third reason relies on the transcriptome functional annotation. Since the STA and DA annotations were independent, transcripts with a related GO term in the MR strategy might have different assigned GO terms in the STA and DA strategies. Consequently, the transcripts may correlate with different GO terms functionally related.

According to our GO terms analysis, several functional groups that have been directly or indirectly associated to defense response were enriched by induced and repressed genes. Notably, the enrichment patterns were consistent in all bioinformatics approaches and variations exhibited in each approach could be associated to our statistics test, number of genes/transcripts predicted per approach and isoforms generated in STA and DA approaches compared to MR. The oxylipin biosynthetic process GO term is defined as the chemical reactions and pathways resulting in the formation of oxylipins. Based on previous works that have shown the involvement of oxylipins in plant-pathogen interactions and wounding (Weber *et al*., 1997, 1999), high enrichment of this group in our analysis may be due to the infection with *Cmm,* as well as the wounding caused during inoculation. In *Arabidopsis thaliana* leaves, oxylipins tend to accumulate to high levels during hypersensitive response to *Pseudomonas syringae* pv tomato (*Pst*) expressing the avirulence gene *avRpm*1 (Vollenweider *et al.*, 2000). Correspondingly, silencing of an α-dioxygenase gene, which is involved in oxylipin biosynthesis, enhances the susceptibility to

several bacterial pathogens and suppresses the hypersensitive response in *Capsicum annum* (Hong *et al.,* 2017). The cytokinin GO term was similarly enriched in both tomato species, particularly with induced genes from 8 to 24 hpi (Figure 8). Cytokinins (Ck) are plant growth hormones that affect plant immunity to several pathogens, such as *Rhodococcus fascians*. This pathogen produces Ck recognized by *A. thaliana* AHK3 and AHK4 receptors leading to development of disease symptoms (Pertry *et al.*, 2009). Nonetheless, it has also been described that endogenous Ck promotes resistance in *Arabidopsis thaliana* in response to *Pst*, which does not secretes Ck into the host (Choi *et al.*, 2010). It is likely that *Cmm* does not secrete Ck into tomato plants, suggesting that endogenous Ck are having a protective effect in the wild tomato. Some other GO terms are enriched mainly in *S. arcanum*, such as mRNA splicing, protein ubiquitination, transmembrane receptor serine/threonine kinase, response to oxidative stress, ubiquitin ligase binding, etc., (Figure 8). These GO terms harbor traits that play several roles in plant defense response (Gomez-Gomez and Boller, 2000; Choi *et al.*, 2007; Lee *et al.*, 2011; Zhang *et al.*, 2014). Altogether, these results show some of the genes that may be involved in the defense response background of wild tomato species against *Cmm*.

Several genes are mainly induced in *S. arcanum* from 0 to 8 hpi in the MR approach, such as a gene that codifies to a PPO E. Polyphenol Oxidases are involved in the oxidation of phenols to quinones. Previous studies have demonstrated a defensive role in transgenic tomato plants overexpressing a PPO of potato (*Solanum tuberosum*). Such transgenic plants exhibited an increased resistance to the pathogen *Pst*, reducing the bacterial growth, as well as severity of disease symptoms (Li and Steffens, 2002). Moreover, down regulation of the PPO family in tomato, through the introduction of a chimeric antisense potato PPO cDNA, induced an increased susceptibility to *Pst*, suggesting that PPOs have a significant role in limiting disease development (Thipyapong *et al.*, 2004). In the STA approach we identified a non-annotated gene putatively related with serine-type endopeptidase inhibitory activity induced in *S. arcanum* from 0 to 8 hpi. Proteinase inhibitors have been characterized as enzymes involved in several biological processes including plant defense. Proteinase inhibitor StPI from a variety of *S. tuberosum* resistant to

Ralstonia solanacearum is strongly induced during the first 6 to 12 hours after the exposure with the pathogen and jasmonic acid treatment (Li *et al.*, 2007). Likewise, transgenic tobacco lines over-expressing the multi-domain proteinase inhibitor NA-PI from *Nicotiana alata* and a β-hordothionin from barley are resistant to the fungal pathogen *Botritys cinerea* and the bacterial pathogen *Pseudomonas solanacearum* (Charity *et al.*, 2015). Therefore, the fact that we observed induced genes, particularly in *S. arcanum*, associated to PPOs and putative proteinase inhibitors during the challenge with *Cmm* might indicate their relevance in this plant-pathogen interaction.

Our analysis showed DEGs encoding Leucine-rich repeat receptor-like kinases (LRR-RLKs) in *S. arcanum* and *S. lycopersicum* species from 0 to 8 hpi (Table 4). Several LRR-RLKs having a role in defense response have been previously identified. For example, the NILR1 receptor is involved in resistance to parasitic nematodes (Mendi *et al.,* 2017). In rice (*Oryza sativa*), the *Xa26* gene confers resistance against the bacterial pathogen *Xanthomonas oryzae* pv. *Oryzae* (*Xoo*) (Sun *et al.*, 2004). Similarly, the overexpression of the *ERECTA* gene in *A. thaliana* var Ler displays an increased disease resistance to *R. solanacearum*, as indicated by reduced wilt symptoms and impaired bacterial growth (Godiard *et al.*, 2003). In our study, the observed induction of this group of genes in both tomato species might indicate a role in the defense response against *Cmm*; and some specific members, like LRR receptor-like serine/threonine-protein kinase gene products (e.g., Solyc11g017270.2 and TRINITY_DN14123_c3_g1_i2) could be particularly involved.

This study showed the induction of genes encoding proteins related to Akyrin repeat-containing domain, mostly in *S. arcanum* from 0 to 8 hpi. In plants, the ankyrin proteins belong to a large family of proteins involved in diverse functions, including plant defense. In *A. thaliana* the ankyrin-repeat transmembrane protein BDA1 plays an important role in defense against bacteria. Loss of function mutations in the *BDA1* gene result in enhanced disease susceptibility to pathogen *Pst* DC3000, as well as in the non-pathogenic strain *Pst* DC30 *hrcC*; whereas, a gain of function allele of

bda1 constitutively activated cell dead and defense response (Yang *et al.*, 2012). Similarly, the XA21 binding protein 25 (XB25), a plant-specific ankyryn-repeat protein of rice (*Oryza sativa*), interacts with XA21, a protein that confers resistance to a broad spectrum of *Xoo*. In addition, the accumulation of XB25 is induced by *Xoo* infection. The repression of *Xb25* diminishes the accumulation of XA21 resulting in a compromised disease resistance, indicating that XB25 is required for maintaining XA21 mediated disease resistance (Jiang *et al.*, 2013). ANK (Solyc08g062360.3.1) was one of the transcripts differentially expressed from 0 to 8 hpi in *S. arcanum*, suggesting that this gene could also be participating during the early defense response to *Cmm*.

Another interesting gene detected was a PHLOEM PROTEIN 2-LIKE A10 (PHLOEM PL2; Solyc02g092140.1) gene. *Cmm* destroys xylem vessels and phloem tissues as the bacteria multiply (Gleason *et al*., 1993). Induction of phloem proteins has been associated with diverse external stresses including pathogenic attack (Beneteau *et al*., 2010; Zhang *et al*., 2011; Lee *et al*., 2014). In our experiment, PHLOEM PROTEIN 2-LIKE A10 was induced in *S. arcanum* from 0 to 8 hpi (Figure 9), suggesting a protective role against *Cmm* infection.

We found several genes that were induced only in *S. arcanum* (Table 5). For instance, an Aquaporin PIP2-1 upregulated at 8-24h, a class of small, membrane channel protein that can facilitate selective flux of various small molecules involved in numerous essential processes across membranes. Recently, evidence has pointed to aquaporins having a role in plant defense against pathogens (Li *et al*., 2020). After pathogen infection apoplastic H_2O_2 is rapidly translocated into the cytoplasm across the plasma membrane (PM). A recent work demonstrated that AtPIP1;4, which is one of the PIP family members in *Arabidopsis*, is involved in the transport of H_2O_2 across the PM (Tian *et al*., 2016). We also found a Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) which has been reported to have a role in immune responses to *Xanthomonas axonopodis* pv *manihotis* (Xam) in cassava (Zeng *et al*., 2018) and a Northern blot analysis showed the accumulation of cytosolic GAPDH gene transcripts in leaves and stems of inoculated potato plants with an

elicitor from the late blight fungal agent *Phytophthora infestan*s (Laxalt *et al*., 1996). In our analysis a *SWEET2* gene induced in *S. arcanum* from 0 to 8 hpi. SWEET proteins possess an important role during plant-pathogen interactions. For example, in roots of Arabidopsis, the loss-of-function *SWEET2* mutants are more susceptible to the infection by the oomycete *Pythium irregular* in comparison to *Arapidopsis* plants expressing *SWEET2*. Suggesting that SWEET2 activity contributes to resistance to *P irregulare* possibly by reducing the availability of glucose in the rhizosphere (Chen *et al*., 2015). Another interesting transcript is one annotated as LNK1. It has been suggested that biotic stress responses are regulated by the circadian clock in plants and that bacterial infection disrupts clock gene expression to attenuate immune responses. Recent research has demonstrated that infection with *Pseudomonas syringae* in wild-type (WT) plants downregulated the expression of several core clock genes, 1 h post-infection, including all members of the NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED (LNK) gene family, and this effect was attenuated in eds4 mutant which is highly susceptible mutant. Furthermore, lnk mutants were more susceptible than the WT to *P. syringae* infection (De Leone *et al*., 2020). Other protein found only in *S. arcanum* was a gamma aminobutyrate (GABA) transaminase 2. It has been described that recognition of bacterial and fungal pathogens has proven to trigger GABA accumulation in several plant species, including tomato (Wang *et al*., 2019). The accumulated GABA has been suggested to be transported to the mitochondria, where it undergoes deamination via GABA transaminase (GABAT) to be converted into succinic semialdehyde (SSA) (Shelp and Zarei, 2017).

For many years geneticists have attempted to clone the resistance genes present in the various QTLs identified for *C. michiganensis* resistance in *S. arcanum*. Although some progress has been made, the regions spanning the QTLs are large and there to our knowledge, a fine mapping has not been published yet. We have analyzed the presence of transcripts in the regions spanning some QTLs. We found 70 transcripts on chromosome 5, 42 on chromosome 7 and 2,479 on chromosome 9 (Table S10). Interestingly the most highly expressed transcript at 0-8h (TRINITY_DN21279_c1_g1_i2) that has homology to the region spanning the QTL

in chromosome 5 (Table 4), has the same annotation as (UPF0481 protein At3g47200) the one recently found on a high-density pepper linkage map, which is a possible candidate of a QTL also on chromosome 5, and it is related to *P. capsici* resistance (Li *et al*., 2021). Another interesting transcript is a putative Hyoscyamine 6-dioxygenase. These proteins generate free radical species, which is an early event associated with the Hypersensitive Response (HR). A putative hyoscyamine 6-dioxygenase was also found to be induced at the early stages of infection in a potato cultivar resistant to *Phytophthora infestans* (Birch *et al*., 1999).

On the region of chromosome 7 we found homology to a transcript annotated as a Diacylglycerol kinase (DGK) (Table 4). These are very important signaling enzymes that phosphorylate diacylglycerol (DAG) to produce phosphatidic acid (PA). There are two pathways by which PA can be produced, through phospholipase D (PLD) and the coupled phospholipase C (PLC)/DGK route (Escobar-Sepúlveda *et al*., 2017). The PLC/DGK pathway seems to be responsible for the rapid accumulation of PA during exposure to pathogen elicitors from bacteria, fungi or oomycetes (Arisz *et al*., 2009). Moreover, a critical reductive coenzyme (NADPH) is produced by DGK enzymes. An oxidative burst driven by NADPH oxidases takes place in most plant-pathogen interactions, especially in incompatible interactions. DGK enzymatic activity seems to be increased after elicitor treatment in tomato, tobacco, rice and Arabidopsis plants, as well as upon pathogen infection. It has been suggested that DGK is the principal producer of plant PA and that it is involved in the basal transcriptome regulation, the stimulation of callose accumulation in the apoplast, and the tolerance to the pathogen *Pseudomonas syringae* (Kue Foka *et al*., 2020). Moreover, the overexpression of the rice DGK gene (OsBIDK1) in tobacco, improved its tolerance to tobacco mosaic virus (TMV) and *Phytophthora parasitica* var. *nicotianae* infections (Zhang *et al*., 2008). On chromosome 7 we also found an RPP13-like protein 1. RPP13 is an Arapidopsis resistance gene that confers resistance to *Hyaloperonospora arabidopsidis* (Bittner-Eddy *et al*., 2000) and RPP13-like proteins have been found in many other plants associated with disease resistance. In the region of chromosome 9 we found a TOM1 like protein differentially regulated. It has been suggested that TOM1 like proteins have a role in

the passage of endocytosed ubiquitinated plasma membrane cargo and act as gatekeepers for the sorting of protein degradation to the vacuole (Roach *et al*., 2021). Silencing of a homologue of TOM1-L2 (TOL) in *Nicotiana benthamiana* has reduced bacterial growth, thus it has been proposed that TOL proteins negatively regulate the plant immune response (Conlan *et al*., 2018).

On the other hand, Coaker *et al*. (2004), analyzed the proteome of a susceptible tomato line and two isogenic tomato lines infected with *Cmm* and containing the QTLs Rcm 5.1 and Rcm 2 identified in *S. habrochaites* LA408, which are proteins related to *Cmm* resistance. This study revealed a role of oxidative stress in response to *Cmm*, since three distinct superoxide dismutase enzymes (SOD) were differentially regulated among genotypes analyzed. Interestingly, we also found transcripts from superoxide dismutase genes mainly induced in *S. arcanum* from 8 to 24 hpi y the three analysis approaches (Tables S3-S8); which supports the previous finding made by Coaker *et al*. (2004).

We validated our RNA-Seq results by qRT-PCR assays of three different genes related to the functions discussed above: a gene encoding a polyphenol oxidase E, an ankyrin repeat-containing protein and a LRR receptor-like serine/threonine-protein kinase (Table 4). The qRT-PCR results are in agreement with the gene expression profiles observed in the RNA-Seq (Figure 9), supporting the reproducibility of our results. The moderate variation in gene expression levels are due to the qRT-PCR sensitivity, together with the inherent biological variability resulting from two independent infection assays in two different tomato species (one for RNA-Seq, and one for qRT-PCR). Therefore, these results provide, not only statistical significance, but also biological significance to our analyses.

Most importantly, despite showing enriched GO terms related to defense, *S. lycopersicum* is unable to overcome *Cmm* infection. Our results suggest a delayed response in *S. lycopersicum* compared with the wild *S. arcanum*. Moreover, *S. arcanum* seems to have more induced genes at early stages after inoculation related to bacterial and fungal defense. Our mapping results also suggest high conservation levels between both species, which leads us to propose that such delayed response

might be associated with an upstream regulatory mechanism involved in the defense response, which probably diverged in *S. arcaum* and *S. lycopersicum* during domestication. However, we cannot rule out the possibility that those genes specific to *S. arcanum,* which we could only find in the *de novo* assembly, have an important contribution to defense against *Cmm*.

5. Concluding remarks

Our results demonstrated the feasibility of mapping to a closely related species but also the relevance of a *de novo* or a semi *de novo* approach. They also showed consistent results that revealed many interesting genes that contribute to the resistance phenotype. Additionally, a qRT-PCR assay confirmed differential gene expression on some of the candidates. Our functional analysis has unfolded a diversity of processes associated with the differential phenotype. The huge amount of data generated by this study will surely contribute to a deeper understanding of defenses against *Cmm*. However, further analysis of this defense signaling network, in order to find genetic variations and possibly divergent regulation, would be essential to unleash the resistance processes taking place in both species. Differentially expressed genes between resistant and susceptible varieties and their functional analysis contribute to the search for new options for generating resistant tomato varieties.

6. Bibliography

Ameur A, Wetterbom A, Feuk L, Gyllensten U. (2010) Global and unbiased detection of splice junctions from RNA-seq data. Genome Biology 11, R34. https://doi.org/10.1186/gb-2010-11-3-r34

Arisz SA, Testerink C, Munnik T. (2009) Plant PA signaling via diacylglycerol kinase. Biochimica et Biophysica Acta 1791, 869-875. doi:10.1016/j.bbalip.2009.04.006

Ashburner M, Ball CA, Blake JA, *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics 25, 25-29. https://doi.org/10.1038/75556

Balaji V, Mayrose M, Sherf O, Jacob-Hirsch J, Eichenlaub R, Iraki N, Manulis-Sasson S, Rechavi G, Barash I, Sessa G. (2008) Tomato transcriptional changes in response to *Clavibacter michiganensis* subsp. *michiganensis* reveal a role for ethylene in disease development. Plant Physiology 146, 1797–1809. https://doi.org/10.1104/pp.107.115188

Balaji V, Smart CD. (2012) Over-expression of snakin-2 and extensin-like protein genes restricts pathogen invasiveness and enhances tolerance to *Clavibacter michiganensis* subsp. *michiganensis* in transgenic tomato (*Solanum lycopersicum*). Transgenic Research 21, 23-37. doi:10.1007/s11248-011-9506-x

Basim H, Basim E, Tombuloglu H, Unver T. (2021) Comparative transcriptome analysis of resistant and cultivated tomato lines in response to *Clavibacter michiganensis* subsp. *michiganensis*. Genomics 113, 2455-2467. doi:10.1016/j.ygeno.2021.05.033

Beneteau J, Renard D, Marché L, Douville E, Lavenant L, Rahbé Y, Dupont D, Vilaine F, Dinant S. (2010) Binding properties of the N-acetylglucosamine and high-mannose N-glycan PP2-A1 phloem lectin in *Arabidopsis*. Plant Physiology 153, 1345-1361. doi:10.1104/pp.110.153882

Birch PR, Avrova AO, Duncan JM, Lyon GD, Toth RL. (1999) Isolation of potato genes that are induced during an early stage of the hypersensitive response to *Phytophthora infestans*. Molecular Plant-Microbe Interactions 12, 356-361. doi:10.1094/MPMI.1999.12.4.356

Bittner-Eddy PD, Crute IR, Holub EB, Beynon JL. (2000) RPP13 is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. The Plant journal: for Cell and Molecular Biology 21, 177-188, doi:10.1046/j.1365-313x.2000.00664.x

Bray NL, Pimentel H, Melsted P, Pachter L. (2016) Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 34, 525–527. https://doi.org/10.1038/nbt.3519

Charity JA, Hughes P, Anderson MA, Bittisnich DJ, Whitecross M, Higgins TJV. (2005) Pest and disease protection conferred by expression of barley beta-hordothionin and *Nicotiana alata* proteinase inhibitor genes in transgenic tobacco. Functional Plant Biology 32, 35–44.

Chen HY, Huh JH, Yu YC, Ho LH, Chen LQ, Tholl D, Frommer WB, Guo WJ. (2015) The Arabidopsis vacuolar sugar transporter SWEET2 limits carbon sequestration from roots and restricts *Pythium infection*. The Plant journal: for Cell and Molecular Biology 83, 1046-1058. doi:10.1111/tpj.12948

Choi HW, Kim YJ, Lee SC, Hong JK, Hwang BK. (2007) Hydrogen peroxide generation by the pepper extracellular peroxidase $CaPO_2$ activates local and systemic cell death and defense response to bacterial pathogens. Plant Physiology 145(3), 890–904. https://doi.org/10.1104/pp.107.103325

Choi J, Huh SU, Kojima M, Sakakibara H, Paek K, Hwang I. (2010) The cytokinin-activated transcription factor ARR2 promotes plant immunity via TGA3/NPR1-dependent salicylic acid signaling in Arabidopsis. Developmental Cell 19, 284–295. https://doi.org/10.1016/j.devcel.2010.07.011

Coaker GL, Willard B, Kinter M, Stockinger EJ, Francis DM. (2004) Proteomic analysis of resistance mediated by Rcm 2.0 and Rcm 5.1, two loci controlling resistance to bacterial canker of tomato. Molecular Plant Microbe Interactions 17, 1019-1028. doi:10.1094/MPMI.2004.17.9.1019

Conlan B, Stoll T, Gorman JJ, Saur I, Rathjen JP. (2018) Development of a rapid in planta BioID system as a probe for plasma membrane-associated immunity proteins. Frontiers in Plant Sciences 9, 1882. doi:10.3389/fpls.2018.01882

De Leone MJ, Hernando CE, Romanowski A, Careno DA, Soverna AF, Sun H, Bologna NG, Vázquez M, Schneeberger K, Yanovsky MJ. (2020) Bacterial infection disrupts clock gene expression to attenuate immune responses. Current Biology 30, 1740-1747. e1746. doi:10.1016/j.cub.2020.02.058

Escobar-Sepúlveda HF, Trejo-Téllez LI, Pérez-Rodríguez P, Hidalgo-Contreras JV, Gómez-Merino FC. (2017) Diacylglycerol kinases are widespread in higher plants

and display inducible gene expression in response to beneficial elements, metal, and metalloid ions. Frontiers in Plant Sciences 8, 129. doi:10.3389/fpls.2017.00129

Ewels P, Magnusson M, Lundin S, Käller M. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047-3048. https://doi.org/10.1093/bioinformatics/btw354

Finn RD, Clements J, Eddy SR. (2011) HMMER web server: interactive sequence similarity searching. Nucleic acids research 39, 29–37. https://doi.org/10.1093/nar/gkr367

Francis DM, Kabelka E, Bell J, Franchino B, St Clair D. (2001) Resistance to bacterial canker in tomato (*Lycopersicon hirsutum* LA407) and its progeny derived from crosses to *L. esculentum*. Plant Disease 85, 1171–1176. doi:10.1094/PDIS.2001.85.11.1171

Gleason ML, Gitaitis RD, Ricker MD. (1993) Recent progress in understanding and controlling bacterial canker of tomato in eastern North America. Plant Disease 77, 1069–1076. doi:10.1094/PD-77-1069

Godiard L, Sauviac L, Torii KU, Grenon O, Mangin B, Grimsley NH, Marco Y. (2003) ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. The Plant journal: for Cell and Molecular Biology 36, 353–365. https://doi.org/10.1046/j.1365-313x.2003.01877.x

Gómez-Gómez L, Boller T. (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Molecular Cell 5, 1003–1011. https://doi.org/10.1016/s1097-2765(00)80265-8

Grabherr MG, Haas BJ, Yassour M, *et al*. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29, 644-652. https://doi.org/10.1038/nbt.1883

Hausbeck M, Bell J, Medina-Mora C, Podolsky R, Fulbright D. (2000) Effect of bactericides on population sizes and spread of *Clavibacter michiganensis* subsp. *michiganensis* on tomatoes in the greenhouse and on disease development and crop yield in the field. Phytopathology 90, 38–44. https://doi.org/10.1094/PHYTO.2000.90.1.38

Hölzer M, Marz M. (2019) De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. GigaScience 8, giz039. https://doi.org/10.1093/gigascience/giz039

Honaas LA, Wafula E K, Wickett NJ, Der J P, Zhang Y, Edger PP, Altman NS, Pires JC, Leebens-Mack JH, dePamphilis CW. (2016) Selecting superior de novo transcriptome assemblies: Lessons learned by leveraging the best plant genome. PloS one 11, e0146062. https://doi.org/10.1371/journal.pone.0146062

Hong CE, Ha YI, Choi H, *et al*. (2017) Silencing of an α-dioxygenase gene, Ca-DOX, retards growth and suppresses basal disease resistance responses in *Capsicum annum*. Plant Molecular Biology 93, 497-509. https://doi.org/10.1007/s11103-016-0575-3

Jiang Y, Chen X, Ding X, Wang Y, Chen Q, Song WY. (2013) The XA21 binding protein XB25 is required for maintaining XA21-mediated disease resistance. The Plant journal: for cell and molecular biology 73, 814–823. https://doi.org/10.1111/tpj.12076

Kabelka E, Franchino B, Francis DM. (2002) Two Loci from *Lycopersicon hirsutum* LA407 Confer Resistance to Strains of *Clavibacter michiganensis* subsp. *michiganensis*. Phytopathology 92, 504–510. doi:10.1094/PHYTO.2002.92.5.504

Kim D, Langmead B, Salzberg SL. (2015) HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12, 357–360. https://doi.org/10.1038/nmeth.3317

Kue Foka, I.C.; Ketehouli, T.; Zhou, Y.; Li, X.W.;Wang, F.W.; Li, H. The emerging roles of diacylglycerol kinase (DGK) in plant stress tolerance, growth, and

development. Agronomy 2020, 10, 1375. https://doi.org/10.3390/agronomy10091375

Lara-Ávila JP, Isordia-Jasso MI, Castillo-Collazo R, Simpson J, Alpuche-Solís ÁG. (2012) Gene expression analysis during interaction of tomato and related wild species with *Clavibacter michiganensis* subsp. *michiganensis*. Plant Molecular Biology Reporter 30, 498–511. https://doi.org/10.1007/s11105-011-0348-8

Laxalt AM, Cassia RO, Sanllorenti PM, Madrid EA, Andreu AB, Daleo GR, Conde RD, Lamattina L. (1996) Accumulation of cytosolic glyceraldehyde-3-phosphate dehydrogenase RNA under biological stress conditions and elicitor treatments in potato. Plant Molecular Biology 30, 961-972. doi:10.1007/BF00020807

Le S, Josse J, Husson F. (2008) FactoMineR: an R package for multivariate analysis. Journal of Statistical Software 25, 1–18.

Lee JR, Boltz KA, Lee SY. (2014) Molecular chaperone function of Arabidopsis thaliana phloem protein 2-A1, encodes a protein similar to phloem lectin. Biochemical and Biophysical Research Communications 443, 18-21. doi:10.1016/j.bbrc.2013.11.034

Lee DH, Choi HW, Hwang BK. (2011) The pepper E3 ubiquitin ligase RING1 gene, CaRING1, is required for cell death and the salicylic acid-dependent defense response. Plant Physiology 156, 2011–2025. https://doi.org/10.1104/pp.111.177568

Li G, Chen T, Zhang Z, Li B, Tian S. (2020) Roles of aquaporins in plant-pathogen interaction. Plants 9, 1134. doi:10.3390/plants9091134

Li GC, Jin LP, Xie KY, Li Y, Qu DY. (2007) Cloning of proteinase inhibitor gene StPI in diploid potato and its expression analysis. Agricultural Science in China 6, 1315-1321.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. (2009) The

Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li L, Steffens JC. (2002) Overexpression of polyphenol oxidase in transgenic tomato plants results in enhanced bacterial disease resistance. Planta 215, 239–247. https://doi.org/10.1007/s00425-002-0750-4

Li YF, Zhang SC, Yang XM, Wang CP, Huang QZ, Huang RZ. (2021) Generation of a High-Density Genetic Map of Pepper (*Capsicum annuum L*.) by SLAF-seq and QTL Analysis of *Phytophthora capsici* Resistance. Horticulturae 7, 92. doi:10.3390/horticulturae7050092

Liao Y, Smyth GK, Shi W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656

Livak KJ, Schmittgen TD. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408. https://doi.org/10.1006/meth.2001.1262

Mendy B, Wang'ombe MW, Radakovic ZS, Holbein J, Ilyas M, Chopra D, Siddique S. (2017) Arabidopsis leucine-rich repeat receptor–like kinase NILR1 is required for induction of innate immunity to parasitic nematodes. PLoS Pathogens 13. https://doi.org/10.1371/journal.ppat.1006284

Ozsolak F, Milos PM. (2011). Nature reviews. Genetics 12, 87–98. https://doi.org/10.1038/nrg2934

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 33, 290–295. https://doi.org/10.1038/nbt.3122

Pertry I, Václavíková K, Depuydt S, *et al*. (2009) Identification of *Rhodococcus fascians* cytokinins and their modus operandi to reshape the plant. Proceedings of the National Academy of Sciences of the United States of America 106, 929–934. https://doi.org/10.1073/pnas.0811683106

Punta M, Coggill PC, Eberhardt RY*, et al*. (2012) The Pfam protein families database. Nucleic Acids Research 40, D290–D301. https://doi.org/10.1093/nar/gkr1065

Roach TG, Lång HK, Xiong W, Ryhänen SJ, Capelluto DG. (2021) Protein trafficking or cell signaling: A dilemma for the adaptor protein TOM1. Frontiers in Cell and Developmental Biology 9, 408. doi:10.3389/fcell.2021.643769

Robinson MD, Mccarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rodriguez F, Wu F, Ané C, Tanksley S, Spooner DM. (2009) Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history?. BMC Evolutionary Biology 9, 191. https://doi.org/10.1186/1471-2148-9-191

Sangiovanni M, Granata I, Thind AS, Guarracino MR. (2019) From trash to treasure: detecting unexpected contamination in unmapped NGS data. BMC Bioinformatics 20, 168. https://doi.org/10.1186/s12859-019-2684-x

Sato, S., Tabata, S., Hirakawa, *et al*. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485, 635–641. https://doi.org/10.1038/nature11119

Savidor A, Teper D, Gartemann KH, Eichenlaub R, Chalupowicz L, Manulis-Sasson S, Barash I, Tews H, Mayer K, Giannone RJ. (2011) The *Clavibacter michiganensis* subsp. *michiganensis*–tomato interactome reveals the perception of pathogen by the host and suggests mechanisms of infection. Journal of Proteome Research 11, 736–750. https://doi.org/10.1021/pr200646a

Schulz MH, Zerbino DR, Vingron M, Birney E. (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–1092. https://doi.org/10.1093/bioinformatics/bts094

Sen, Y. (2014) Bacterial canker resistance in tomato. Ph.D. Thesis, Wageningen University, Netherlands, ISBN:978-94-6257-076-4

Sen Y, Feng Z, Vandenbroucke H, van der Wolf J, Visser RGF, van Heusden AW. (2013) Screening for new sources of resistance to *Clavibacter michiganensis* subsp. *michiganensis* (*Cmm*) in tomato. Euphytica 190, 309–317. https://doi.org/10.1007/s10681-012-0802-1

Shelp BJ, Zarei A. (2017) Subcellular compartmentation of 4-aminobutyrate (GABA) metabolism in Arabidopsis: An update. Plant Signaling & Behavior 12, e1322244. doi:10.1080/15592324.2017.1322244

Slatko BE, Gardner AF, Ausubel FM. (2018) Overview of next-generation sequencing technologies. Current Protocols in Molecular Biology 122, e59. https://doi.org/10.1002/cpmb.59

Sun X, Cao Y, Yang Z, Xu C, Li X, Wang S, Zhang Q. (2004) Xa26, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. The Plant journal: for Cell and Molecular Biology 37, 517–527. https://doi.org/10.1046/j.1365-313x.2003.01976.x

Thipyapong P, Hunt MD, Steffens JC. (2004) Antisense downregulation of polyphenol oxidase results in enhanced disease susceptibility. Planta 220, 105–117. https://doi.org/10.1007/s00425-004-1330-6

Tian S, Wang X, Li P, Wang H, Ji H, Xie J, Qiu Q, Shen D, Dong H. (2016) Plant aquaporin AtPIP1; 4 links apoplastic H_2O_2 induction to disease immunity pathways. Plant Physiology 171, 1635-1650. doi:10.1104/pp.15.01237

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research 35, W71–W74. https://doi.org/10.1093/nar/gkm306

Van Heusden AW, Koornneef M, Voorrips RE, Brüggemann W, Pet G, Vrielink-Van Ginkel R, Chen X, Lindhout P. (1999) Three QTLs from *Lycopersicon peruvianum*

confer a high level of resistance to *Clavibacter michiganensis* ssp. *michiganensis*. Theoretical and Applied Genetics 99, 1068–1074.

Vollenweider S, Weber H, Stolz S, Chételat A, Farmer EE. (2000) Fatty acid ketodienes and fatty acid ketotrienes: Michael addition acceptors that accumulate in wounded and diseased Arabidopsis leaves. The Plant journal: for Cell and Molecular Biology 24, 467–476. https://doi.org/10.1046/j.1365-313x.2000.00897.x

Wang G, Kong J, Cui D, Zhao H, Niu Y, Xu M, Jiang G, Zhao Y, Wang W. (2019) Resistance against *Ralstonia solanacearum* in tomato depends on the methionine cycle and the γ-aminobutyric acid metabolic pathway. The Plant journal: for Cell and Molecular Biology 97, 1032-1047. doi:10.1111/tpj.14175

Weber H, Chételat A, Caldelari D, Farmer EE. (1999) Divinyl ether fatty acid synthesis in late blight-diseased potato leaves. The Plant Cell, 11, 485–494. https://doi.org/10.1105/tpc.11.3.485

Weber H, Vick BA, Farmer EE. (1997) Dinor-oxo-phytodienoic acid: a new hexadecanoid signal in the jasmonate family. Proceedings of the National Academy of Sciences of the United States of America 94, 10473-8. doi:10.1073/pnas.94.19.10473

Wei Z, Wang J, Yang S, Song Y. (2015) Identification and expression analysis of the LRR-RLK gene family in tomato (*Solanum lycopersicum*) Heinz 1706. Genome 58, 121–134. https://doi.org/10.1139/gen-2015-0035

Xie Y, Wu G, Tang J, *et al*. (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30, 1660–1666. https://doi.org/10.1093/bioinformatics/btu077

Yang Y, Zhang Y, Ding P, Johnson K, Li X, Zhang Y. (2012) The ankyrin-repeat transmembrane protein BDA1 functions downstream of the receptor-like protein SNC2 to regulate plant immunity. Plant Physiology 159, 1857–1865. https://doi.org/10.1104/pp.112.197152

Zeng H, Xie Y, Liu G, Lin D, He C, Shi H. (2018) Molecular identification of GAPDHs in cassava highlights the antagonism of MeGAPCs and MeATG8s in plant disease resistance against cassava bacterial blight. Plant Molecular Biology 97, 201-214. doi:10.1007/s11103-018-0733-x

Zhang C, Shi H, Chen L, *et al*. (2011) Harpin-induced expression and transgenic overexpression of the phloem protein gene AtPP2-A1 in Arabidopsis repress phloem feeding of the green peach aphid *Myzus persicae*. BMC Plant Biology 11:11. doi:10.1186/1471-2229-11-11

Zhang W, Chen J, Zhang H, Song F. (2008) Overexpression of a rice diacylglycerol kinase gene OsBIDK1 enhances disease resistance in transgenic tobacco. Molecules and Cells 26, 258-264.

Zhang Z, Liu Y, Ding P, Li Y, Kong Q, Zhang Y. (2014) Splicing of Receptor-Like Kinase-Encoding SNC4 and CERK1 is Regulated by Two Conserved Splicing Factors that Are Required for Plant Immunity. Molecular Plant 7, 1766–1775. https://doi.org/10.1093/mp/ssu103

Chapter II

Development of a new tomato mottle virus (ToMoV)-based gene silencing system, a versatile tool for VIGS in solanaceous species

1. Background

Plant viruses are the causal agents of important crop diseases, producing significant problems in terms of economic losses worldwide (Scholthof *et al.,* 2011; Rybicki 2015). Nonetheless, they also represent an important source of raw material for the development of wide range of molecular tools for plant research. Excellent reviews detail the applications of plant viruses in molecular biology and plant biotechnology (Dommes *et al*., 2019; Ibrahim *et al*., 2019, Abrahamian *et al.,* 2020). Remarkably, virus-induced gene silencing (VIGS) is a functional genomic tool that has been widely used in the last twenty years.

VIGS technology takes advantage of the defense mechanisms that plants have developed for counteracting plant viruses. Such systems are based in part, on the degradation of viral mRNA molecules by post-transcriptional gene silencing (PTGS) (Smith *et al*., 1994, Mueller *et al*., 1995; English *et al*., 1996; Sijen *et al*., 1996), mediated by small interference RNAs (siRNA), first described by Hamilton and Baulcombe (1999). PTGS occurs when viral double stranded RNAs (dsRNAs) are recognized and cleaved into 21-24 nt siRNAs by the RNase III-like ribonucleases Dicer-like (DCL) members family (Bernstein *et al.*, 2001; Bouché *et al.*, 2006; Garcia-Ruiz *et al.*, 2010). Such dsRNAs are mainly generated from RNA replication intermediates, highly structured RNAs or RNAs produced from bidirectional transcription overlaps (Chellappan *et al.*, 2004; Molnár *et al.*, 2005; Aregger *et al.,* 2012). The formed siRNAs can associate with proteins from the ARGONAUTE (AGO) family, establishing RNA-induced silencing complexes (RISC). Such complexes target and silence sequence-specific viral RNAs leading to PTGS (Morel *et al.,* 2002; Carbonell *et al.,* 2012; Garcia-Ruiz *et al.,* 2015). PTGS is not restricted to the infection site, but it can spread systematically throughout the plant (Palauqui

et al., 1997), which is mediated by an amplification process that generates *de novo* sequence specific siRNAs in tissues far from the infection site, helping to maintain the PTGS in the plant (Himber *et al.,* 2003). The insertion of partial sequences of plant endogenous genes into viral vectors results in the silencing of targeted plant genes, demonstrated by Kumagai *et al.,* (1995). This finding opened a new approach for functional genomics in plants by VIGS.

A wide range DNA and RNA virus have been manipulated in order to obtain useful VIGS systems (Dommes *et al*. 2019). The most employed VIGS vectors are based on the RNA tobacco rattle virus (TRV) (Ratcliff *et al*., 2001; Liu *et al*., 2002b). On the other hand, the most important VIGS DNA-based vectors come from members of the genus *Begomovirus* (*Geminiviridae*), the first silencing vector was based on the tomato golden mosaic virus (TGMV) (Kjemtrup *et al*., 1998).

Begomovirus genome consist of circular single-stranded DNA molecules that can be either monopartite or bipartite. In the case of bipartite viruses both DNA-A and DNA-B molecules are required for infectivity (Stanley 1983; Stanley and Gay 1983). DNA replication of geminiviruses occurs by rolling circle replication and is dependent of host DNA polymerases (Saunders *et al.*, 1991). Equally, RNA transcription is dependent of host RNA polymerases and occurs in a bidirectional manner from the common region (CR) of geminivirus genomes (Townsend *et al.*, 1985; Frischmuth *et al.,* 1991). Specifically for DNA viruses, such transcriptional mechanism could trigger PTGS due to formations of complementary dsRNAs. Also, is possible that viral transcripts can act either as templates for dsRNA synthesis by host RNA-dependent RNA polymerases (RdRP) or forming RNA secondary structures recognizable by DCLs (Townsend *et al.*, 1985; Dalmay *et al.,* 2000; Chellappan *et al.*, 2004; Vanitharani *et al.*, 2005).

DNA-based silencing vectors do not represent a biosafety risk, since viral encapsidation is defective due to removal of a portion of the coat protein (CP) gene, which is replaced with a fragment of the endogenous gene of interest (Kjemtrup *et al.*, 1998; Turnage *et al.*, 2002). The CP gene is located in the DNA-A component of bipartite viruses. In addition, an important consideration in DNA-based VIGS relies

on the inoculation method. Agroinoculation is one of the most common methods used to deliver VIGS vectors into plants (Ratcliff *et al.*, 2001; Liu *et al.*, 2002a, b; Lu *et al.*, 2003; Pandey *et al.*, 2009). For that purpose, viral genomes are cloned into binary vectors as partial dimers or hemidimers, such principle consists of the genome sequence bordered by two copies of the common region, which contains the virus replication origin, allowing the formation of replicative DNA monomers in the plant cells (Hayes *et al.*, 1988; Stenger *et al.*, 1991).

VIGS systems tend to display high silencing yields in *N. bethamiana*. Nonetheless, for non-model organisms VIGS induction can be hard to achieve, either because VIGS system is not suitable to silence a particular plant species or silencing yields are not efficient, in such cases additional modifications need to be performed (Hsieh *et al.,* 2013; Kim *et al.,* 2016; Singh *et al.*, 2018). However, in comparison to genome editing methods, VIGS does not require exhaustive stable transformation protocols, which are mostly optimized for a few plant species (Brooks *et al.*, 2014; Ma *et al.,* 2015). Moreover, knock-down of genes by VIGS allows the study of plant essential genes, otherwise difficult to analyze through their complete knock-out.

Even when there is a broad spectrum of VIGS vectors, there is still a need to develop and to improve VIGS tools for particular purposes. Esparza-Araiza *et al.*, (2015) established a VIGS system based on the begomovirus tomato mottle virus (ToMoV) (Abouzid *et al.*, 1992; Polston *et al.,* 1993). In this study, the resistant wild tomato species *Solanum peruvianum* (*S. arcanum* LA2172) was silenced in the *SCEI* gene, demonstrating its role in resistance to the pathogen *Clavibacter michiganensis* subsp. *michiganensis*. However, this vector includes a single DNA-A genomic component and can be inoculated only by particle bombardment and no host-range silencing analyses have been reported.

Here we report a new VIGS vector based on ToMoV, optimized for agroinoculation. Host-range silencing assays were performed by the silencing of magnesium chelatase gene (*Chl*I) producing a bleached phenotype. *Chl*I is required for chlorophyll biosynthesis. The results indicate that ToMoV DNA-A and DNA-B hemidimers can induce gene silencing in at least five solanaceous when are

inoculated by leaf agroinfiltration. This work highlights a new VIGS system that can be used for functional genomics.

2. Methods

2.1. Biological material and growth conditions

Seeds of *S. lycopersicum* cv. Ailsa Craig were obtained from The University of Nothingham, UK, and seeds of *S. arcanum* LA2157, *S. arcanum* LA2172 and *S. habrochaites* LA2128 were obtained from Tomato Genetics Resources Center, Davis, California (https://tgrc.ucdavis.edu). Seeds of *N. benthamiana* were donated by Dr. Gerardo Argüello. Seeds of *Physalis ixocarpa* were obtained from local sellers in San Luis Potosí, Mexico. All plant species were germinated and maintained under controlled conditions in a growth chamber at 25°C with a 16 h/ 8 h light/darkness photoperiod.

ToMoV A and B genome components were isolated from infected tomato leaves, then linearized through *Apa*I restriction site and cloned into pBluescript cloning vector (pBS II SK +/-, Stratagene, La Jolla, CA, USA), by R.F. Rivera-Bustamante, CINVESTAV, Mexico. The silencing vector ToMoV_C*hl*I, cloned into pBS II SK +/-, was developed by Esparza-Araiza *et al.,* (2015).

We employed *Escherichia coli* TOP10 competent cells for molecular cloning. Electro-competent cells of *A. tumefaciens* GV3101 strain were used to inoculate the silencing vector based on the ToMoV hemidimers improved for agroinoculation.

2.2. Construction of ToMoV hemidimers

All vectors were obtained by standard molecular cloning methods.

Construction of ToMoV A hemidimer

The DNA-A component of ToMoV cloned in pBS II SK +/- was digested with *Nco*I (*New England Biolabs Inc*.). The linearized product of 5.5 kb was end filled with the DNA polymerase I, large (Klenow) fragment 3'-5' exo- (*New England Biolabs Inc*.). This DNA product was simultaneously digested with *Apa*I and *Sma*I (*New England Biolabs Inc*.). Three sub-products were obtained: a 2.9 kb product (corresponding to the pBS II SK+/- vector), a 1.4 kb product (corresponding to a fragment of ToMoV-

A), and a 1.2 kb product (corresponding to the fragment with the common region, CR, of the DNA-A of ToMoV). These products were purified in column (PROMEGA) and ligated with T4 DNA ligase (PROMEGA). A 4.1 kb product was obtained corresponding to the DNA-A fragment with the CR inserted into pBS II SK +/-, named pBS-ToMoV-A_IR. Subsequently, the silencing vector ToMoV_C*hI* donated by Esparza-Araiza, was digested with *Sca*I (*New England Biolabs Inc*.), to linearize the cloning vector pBS II SK +/-. Later, the product was digested with *Apa*I for obtaining the complete monomer sequence of ToMoV_*ChI*. The linearized monomer of ToMoV_*ChI* was cloned into pBS-ToMoV-A_IR, previously digested with *Apa*I and dephosphorylated with Antarctic phosphatase (*New England Biolabs Inc.*). The resulting vector of 6.3 kb, was the ToMoV DNA-A hemidimer. However, since this vector had two pairs of *EcoR*I and *Hind*III repeated sites, the first pair located in the multiple cloning site (MSC) of the pBluescript vector, and the second in the deleted region of CP gene of ToMoV (essential for the insertion of DNA sequences of interest for gene silencing); we used as backbone another pBS II SK +/- cloning vector, where the *Sma*I-*Hind*III segment of the MCS was deleted, pBS II SK +/- ∆ *Sma*I-*Hind*III. In an independent reaction the ToMoV hemidimer was released with *Cla*I and *Kpn*I (*New England Biolabs Inc.*), and cloned into the pBS II SK +/- ∆ *Sma*I-*Hind*III, previously linearized with *Cla*I and *Kpn*I. The resulting vector of 6.3 kb, corresponded to the ToMoV hemidimer for gene silencing H-ToMoV-A_*ChI*.

In order to avoid particle bombardment and to use agroinoculation, a fast and less expensive approach, the ToMoV A hemidimer was subcloned into the binary vector pCAMBIA-1300 (Cambia, Australia). To achieve this, the H-ToMoV-A_*ChI* vector was digested with *Pvu*II (Thermo Scientific Inc.). A product of 2.5 kb was obtained (corresponding to a portion of pBS II SK +/- vector), and a second product of 3.7 kb, corresponding to the ToMoV hemidimer. The ToMoV hemidimer was purified from gel by column purification (Wizard® SV Gel and PCR Clean-Up System, Promega), and cloned into pCAMBIA-1300, which was previously digested with *Pvu*II and dephosphorylated with Antarctic phosphatase. The product of 12.4 kb was the silencing vector optimized for agroinfiltration pC_H-ToMoV_*ChI*.

Construction of ToMoV∆CP Hemidimer

To construct the ToMoV∆CP hemidimer improved for agroinoculation, the chelatase gene fragment (*Chl*I) was removed from the pC_H-ToMoV_*Chl*I silencing vector. To that end, the vector was double digested with *Eco*RI and *Hind*III. Then, the linearized product was filled with the DNA polymerase I, large (Klenow) fragment 3'-5' exo-. The resulting product was purified in column and ligated with T4 DNA ligase. The obtained vector of 12.2 kb was named pC_H-ToMoV∆CP.

Construction of ToMoV B Hemidimer

The DNA-B component of ToMoV cloned into pBS II SK +/- was digested with *Eco*RI and the linearized product was digested with *Apa*I. Three sub-products were generated, one of 1.6 kb corresponding to the fragment with the CR. This DNA product was purified from gel by column purification and cloned into pBS II SK +/- previously digested with *EcoR*I and *Apa*I. The vector of 4.5 kb was named pBS-ToMoV-B_IR. Afterwards, the DNA-B cloned in pBS II SK +/- was released with *Apa*I digestion. The resulting monomer was purified from gel by column purification, and cloned into pBS-ToMoV-B_IR previously linearized with *Apa*I. A vector of 7.1 kb was generated corresponding to the DNA-B ToMoV hemidimer named H-ToMoV_B.

We employed the binary vector pBI121 (Clontech, California, USA) as backbone for subcloning the ToMoV B hemidimer. Firstly, the H-ToMoV_B was linearized with *Sca*I and *Sac*I. A product of 1.7 kb was obtained (corresponding to a portion of pBS II SK +/- vector), and a second product of 5.3 kb, corresponding to the ToMoV B hemidimer. The 5.3 kb DNA fragment was purified from gel by column purification; cloned into pBI121 previously digested with *Sca*I and *Sac*I, and dephosphorylated with Antarctic phosphatase. The product of 17.5 kb was ToMoV B hemidimer optimized for agroinfiltration pBI_H-ToMoV_B.

2.3. Plant inoculation and detection of viral replication

Particle bombardment

The H-ToMoV-A_*Chl*I and H-ToMoV_B vectors were mixed in a 1:1 µg ratio. The DNA mixture was put together with a tungsten microprojectile suspension (tungsten

15 µg/mL, 2.5M $CaCl_2$, 0.1 M spermidine). Using the Biolistic® PDS-1000/He Particle Delivery System (Bio Rad, Hercules, CA, USA), two-week old *N. bethamiana* plants were bombarded at 600 psi, each with 2 µg of the DNA suspension. Inoculated plants were maintained in growth chamber conditions during the experiment.

Plant agroinfiltration

Agrobacterium tumefaciens GV3101 cells were transformed by electroporation with pC_H-ToMoV_*Chl*I, pC_H-ToMoV∆CP and pBI_H-ToMoV_B vectors, respectively. Colonies of *Agrobacterium* strains harboring each hemidimer were grown in LB medium, supplied with kanamicin (50 mg/mL) and gentamicin (50 mg/mL) for 18 h at 28°C. The *Agrobacterium* cells transformed with the mentioned constructions were harvested at 2,800 rpm and resuspended in agroinfiltration buffer (10 mM $MgCl_2$, 100 mM acetosyringone y 10 mM MES pH 5.7). Cell suspensions were adjusted to an OD_{600} of 0.5 each, and mixed in a 1:1 ratio (1mL of pC_H-ToMoV_*Chl*I or pC_H-ToMoV∆CP per 1 mL of pBI_H-ToMoV_B). The suspensions were incubated at room-temperature 2 h before agroinfiltration. For infiltration, two-week old plants of each species were infiltrated in the abaxial side of proximal leaves with a 1mL syringe without needle. Agroinfiltrated plants were maintained in growth chamber conditions during the experiment.

Viral replication assays

Total DNA was obtained from new leaves (leaves near to the apical meristem), four weeks after agroinoculation of ToMoV DNA-A and DNA-B hemidimers. As for *N. bethamiana* plants inoculated by particle bombardment, circular DNA molecules of ToMoV were detected by rolling circle amplification (RCA) with the TempliPhi DNA amplification Kit (Sigma-Aldrich, Inc.) and by the PCR amplification a 0.9 kb DNA fragment from the ToMoV DNA A component using the primers AL2-LDD forward 5'-GATCTAGATGNGYTCCCCTGTGCGTGAATCC-3' and DGRS reverse 5'-ATCTCGAGCCATCCRAACRTWCAGGGAGCT-3'. For plants inoculated by the agroinfiltration method, the 0.9 kb DNA fragment from the ToMoV DNA A component was detected.

2.4. Assessment of gene silencing

Phenotypic measurements were carried out based on the gene silencing phenotypic evaluation protocols used for VIGS (Senthil-Kumar *et al.*, 2007; Broderick and Jones 2014). Groups of ten plants per species were established to perform the phenotypic assessments and groups of three plants were set up as controls. Silencing frequency is the percent of plants with visible silencing; calculated from the groups of ten plants above mentioned. Silencing effectiveness, described as the percentage of leaves with visible silencing per plant, is determined from the plants that presented the bleaching phenotype in one assay. Silencing efficiency corresponds to percent of white colored area per leaf, analyzed using ImageJ software (Schneider *et al.,* 2012; https://imagej.nih.gov/ij). The estimation was carried out dividing the bleached area of the leaf by the total area of the leaf. Three silenced plants were randomly selected and one leaf per plant was analyzed, except for *S. arcanum* LA2172 and *P. ixocarpa* where only one plant was silenced; in this case, three leaves per plant were quantified. All phenotypic measurements were performed four weeks after agroinfiltration of ToMoV silencing vectors.

3. Results

3.1. Inoculation of ToMoV hemidimers by particle bombardment results in silencing of *N. benthamiana*.

Esparza-Araiza *et al.* (2015), reported the first silencing vector based on ToMoV, the ToMoV_*Chl*I, suitable for inoculation by particle bombardment. To improve the inoculation method, we developed a new silencing vector based on ToMoV_*Chl*I. The first step in the development of the silencing vectors was the construction of partial dimers. Since ToMoV is a geminivirus that comprises two DNA molecules (genome components A and B), both DNA-A and DNA-B were modified by molecular cloning (see Material and methods). Since the CR is located in the intergenic region (IR) of geminivirus genomes, an extra IR from the DNA-A component was added to the backbone of the silencing vector ToMoV_*Chl*I, resulting in the 6.28 kb hemidimer H-ToMoV-A_*Chl*I (Figure 10a, c). Similarly, an extra IR from the DNA-B component was added to the DNA-B backbone as a direct repeat, resulting in the 7.11 kb hemidimer H-ToMoV-B (Figures 10b, d).

Before testing an alternative inoculation method, we validated the new ToMoV hemidimers by testing the viral replication and induction of gene silencing. The ToMoV hemidimers were inoculated by particle bombardment into *N. bethamiana* plants. The new silencing vector H-ToMoV-A_*Chl*I has a deleted coat protein gene, with a 249 bp fragment from the *Chl*I gene from *S. arcanum* LA2172 inserted, that encodes for a magnesium chelatase (Figure 10a, c). From a three biological replicates assay, two plants presented a strong leaf bleaching spread over the whole plant at 50 days post-inoculation (dpi), and no disease symptoms related to ToMoV were visible (Figure 11d). The non-silenced plant did not show any change, with a phenotype comparable to plants in the control group at 50 dpi (Figure11a-d). We detected viral circular DNA molecules by rolling circle amplification (RCA) in the two silenced *N. bethamiana* plants at 50 dpi. The amplified DNA molecules were linearized by the unique restriction site *Apa*I in the sequence of ToMoV resulting in a ~2.3 kb DNA fragment (Figure 11e). Additionally, a 0.9 kb amplicon was detected by PCR from total genomic DNA obtained from both silenced plants, but it was not detected in the non-silenced replicate at 50 dpi (Figure 11f). These results indicate the capability of the ToMoV hemidimers to release themselves from the cloning vector and start the viral replication, resulting in the successfully induction of gene silencing.

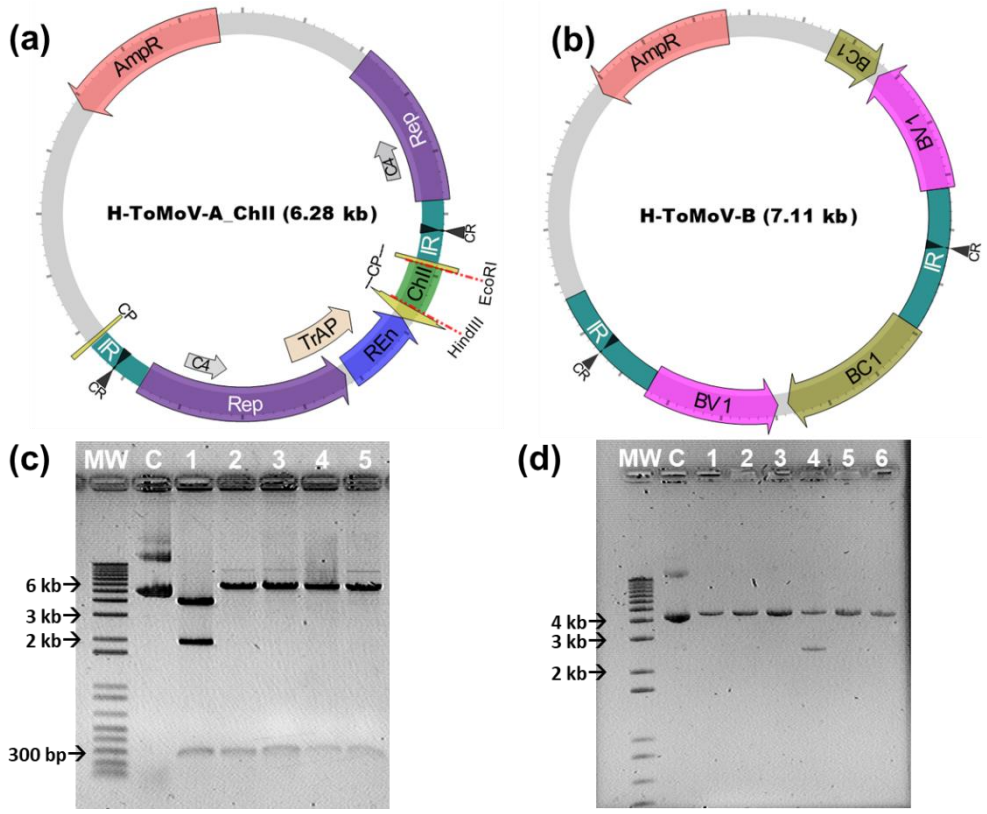


Figure 10. VIGS system based on ToMoV hemidimers. Organization of DNA-A and DNA-B hemidimers cloned into pBluescript plasmid. a) H-ToMoV-A_*Chl*I vector. The complete DNA-A sequence is flanked by two common regions, CRs (black arrows); a 249 bp DNA fragment from the *S. arcanum* LA2172 *Chl*I gene was inserted into the deleted region of *CP* gene. b) H-ToMoV-B vector. The complete DNA-B sequence is flanked by two CRs. Common regions are highlighted as "IR". Intergenic regions are shown in green color and highlighted as "IR". Genes *Rep*, *TrAP*, *REn*, *CP, BC1* and *BV1* encode the replication associated protein, transcription activator protein, replication enhancer, coat protein, and two movement proteins, respectively. The *Chl*I DNA fragment can be replaced with *EcoR*I and *Hind*III restriction enzymes. c) Digestion of the H-ToMoV-A_*Chl*I vector with *EcoR*I and *Hind*III, lanes 2-5 evidence the clones harboring the complete vector. Two DNA fragments resulted after digestion, a ~249 bp and 6 kb fragment, corresponding to the *Chl*I gene fragment and the rest of plasmid vector, respectively. d) Digestion of the H-ToMoV_B vector with *Apa*I, lane 4 evidences the clone harboring the complete vector. Two DNA fragments resulted after digestion; a 2.54 kb and 4.57 kb fragment, corresponding to DNA-B monomer and the rest of plasmid vector, respectively.

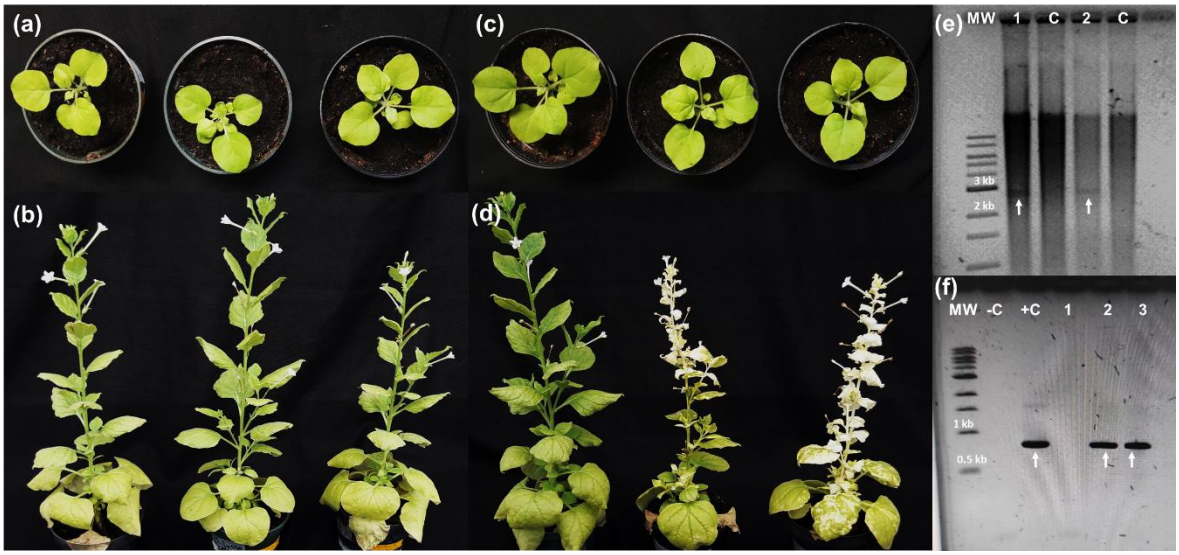


Figure 11. Validation of gene silencing of ToMoV hemidimers H-ToMoV-A_*Chl*I and H-ToMoV-B inoculated by particle bombardment in *N. benthamiana*. a) Mock-inoculated plants at 0 dpi. b) Mock-inoculated plants at 50 dpi. c) Plants inoculated with the ToMoV hemidimers at 0 dpi. d) Plants inoculated with the ToMoV hemidimers at 50 dpi; two of three plants showed bleaching due to silencing of *Chl*I gene. e) Detection of viral circular DNA molecules by TempliPhi amplification; lanes marked as 1 and 2 show the amplified DNA associated to silenced plants and subsequently digested with *Apa*I generating a ~2.3 kb DNA product. f) PCR amplification of a 0.9 kb DNA fragment of the DNA-A component of ToMoV; lanes marked as 2 and 3 show the amplified DNA associated to silenced plants. MW: DNA ladder 1 kb; C: Control, DNA of silenced plants not amplified with Phi29 DNA polymerase; -C: negative control of reaction; +C: positive control, H-ToMoV-A_*Chl*I vector; white arrows: amplified DNA of expected size.

3.2. Agroinoculation of ToMoV hemidimers induces gene silencing at least in five solanaceous

To perform the inoculation by agroifiltration, we subcloned the ToMoV hemidimers into binary vectors. The H-ToMoV-A_*Chl*I was subcloned into pCAMBIA-1300 resulting in the 12.4 kb pC_H-ToMoV_*Chl*I vector (Figure 12a, c). In addition, we obtained a 12.2 kb pC_H-ToMoV∆CP vector that includes the same DNA structure, but lacks the *Chl*I DNA fragment (see methods; Figure 12c). The H-ToMoV-B was subcloned into pBI121 resulting in the 17.5 kb pBI_H-ToMoV_B vector (Figure 12b, d).

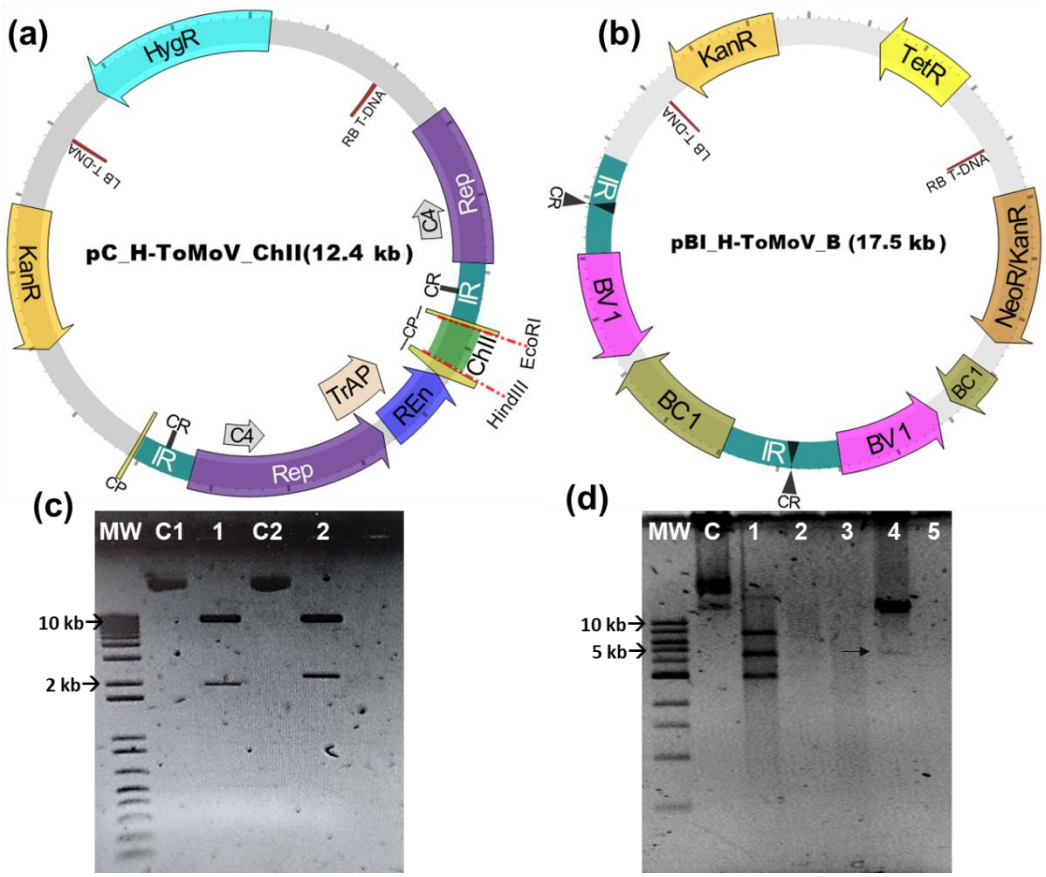


Figure 12. Organization of DNA-A and DNA-B hemidimers cloned into binary vectors. a) pC_H-ToMoV_*Chl*I vector. The DNA-A hemidimer flanked by two CRs and containing the 249 bp DNA fragment of the *Chl*I, was cloned into the backbone of the pCAMBIA-1300 binary vector. b) pBI_H-ToMoV_B vector. The DNA-B hemidimer is flanked by two CRs is cloned into the backbone of the pBI-121 binary vector. c) Digestion of the pC_H-ToMoV_ΔCP and pC_H-ToMoV_*Chl*I vectors with *Apa*I, lanes 1 and 2 evidences the clone harboring the complete vectors, respectively. Two DNA fragments resulted after digestion, a 1.95 kb fragment in the case of pC_H-ToMoV_ΔCP and a 2.18 kb fragment for pC_H-ToMoV_*Chl*I, as well as a 10.22 kb fragment, corresponding to DNA-A monomer and the rest of plasmid vector, respectively. d) Digestion of the pBI_H-ToMoV_B vector with *Sac*I and *Kpn*I, lane 4 evidences the clone harboring the complete vector. Two DNA fragments resulted after digestion, a 4.25 kb and 13.31 kb fragment, corresponding to DNA-B hemidimer and the rest of plasmid vector, respectively. MW: DNA ladder 1Kb and 1Kb plus; C: control sample corresponding to undigested vector.

ToMoV can replicate in several solanaceous genera including *Solanum, Physalis* and *Nicotiana* (Polston *et al.,* 1993). To investigate the host range in which gene silencing is inducible, groups of ten plants from five solanaceous were tested. We inoculated the ToMoV hemidimers by agroinfiltration using the *A. tumefaciens* strain

GV3101. Bleaching phenotype was visible at 7-10 dpi in *N. bethamiana*, *S. lycopersicum* cv. Ailsa Craig, the wild tomato species *S. habrochaites* LA2128 and *S. arcanum* LA2172, and *P. ixocarpa* (Figure 13). Bleaching level was variable among species and clearly spread throughout the plants at 28 dpi (Figure 14). In addition, replication of the silencing vector pC_H-ToMoV_*ChI*I was confirmed in the bleached plants at 28 dpi by PCR, and no replication was detected in non-silenced plants (Figure 15). Furthermore, the tomato *S. arcanum* LA2157 did not showed bleaching symptoms and no viral replication was detected (Figures 14g-h, 15d). No visible disease symptoms related to viral replication were visible in any species studied.

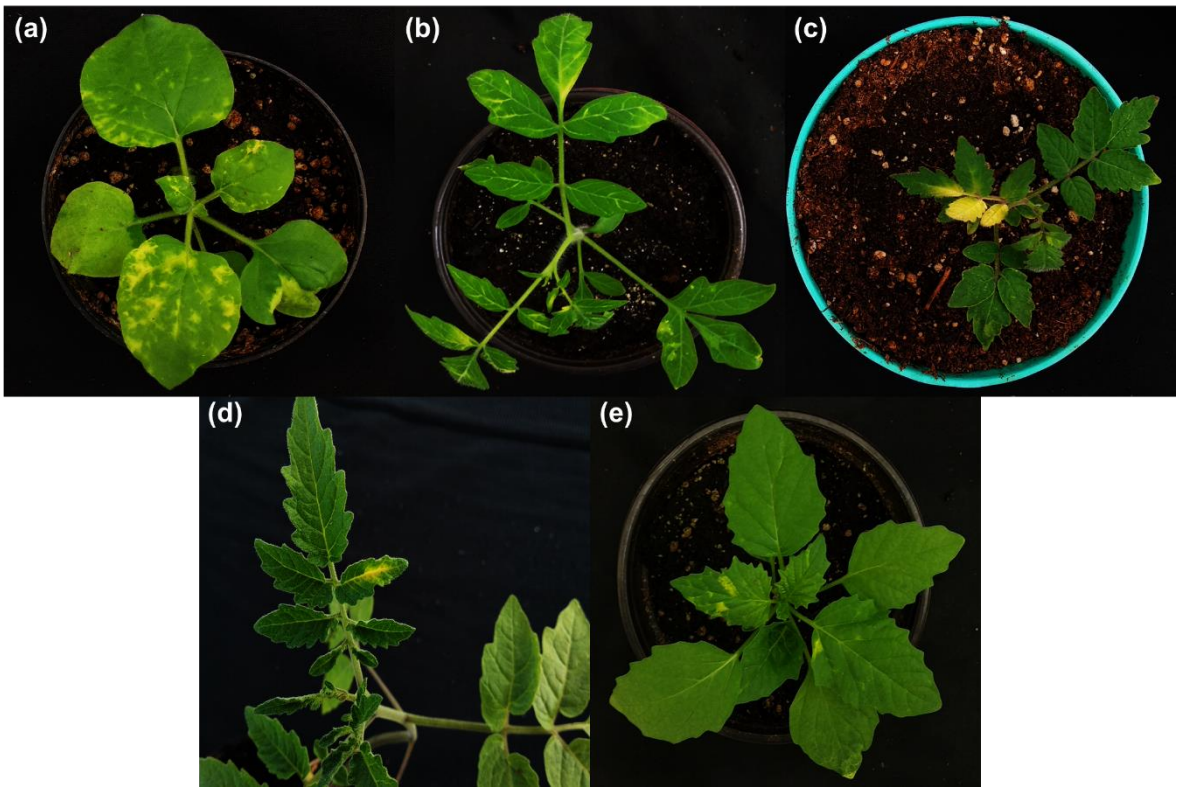


Figure 13. Emerging of bleaching phenotype in different plant species inoculated with the ToMoV hemidimer pC_H-ToMoV_*ChI*. a) Bleaching phenotype visible at 7 dpi in *N. benthamiana*. b) Bleaching phenotype visible at 7 dpi in *S. lycopersicum*. c) Bleaching phenotype visible at 7 dpi in *S. habrochaites* LA2128. d) Bleaching phenotype visible at 10 dpi in *S. arcanum* LA2172. e) Bleaching phenotype visible at 10 dpi in *P. ixocarpa*.

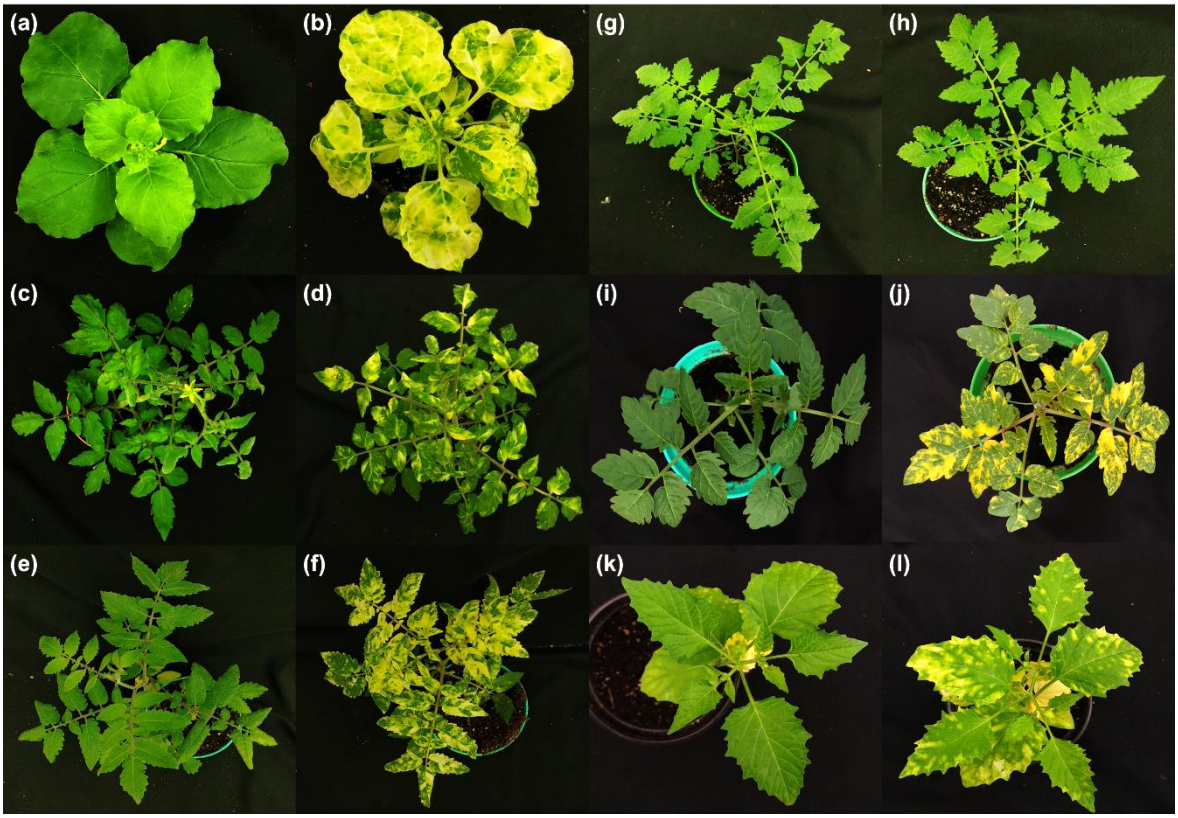


Figure 14. Host range silencing of ToMoV hemidimers inoculated by agroinfiltration. a,c,e,g,i,k) *Nicotiana benthamiana*, *S. lycopersicum*, *S. habrochaites* LA2128, *S. arcanum* LA2157, *S. arcanum* LA2172 and *P. ixocarpa,* respectively, at 28 dpi with empty vector pC_H-ToMoVΔCP. h) *S. arcanum* LA2157 at 28 dpi with silencing hemidimer pC_H-ToMoV_*ChI*I, with no visible bleaching effect. b,d,f,j,l) *N. benthamiana*, *S. lycopersicum*, *S. habrochaites* LA2128, *S. arcanum* LA2172 and *P. ixocarpa,* respectively, at 28 dpi with silencing hemidimer pC_H-ToMoV_*ChI*I, resulting in generalized bleaching visible all over the plant in all cases. The generalized bleaching level varies among the different plant species

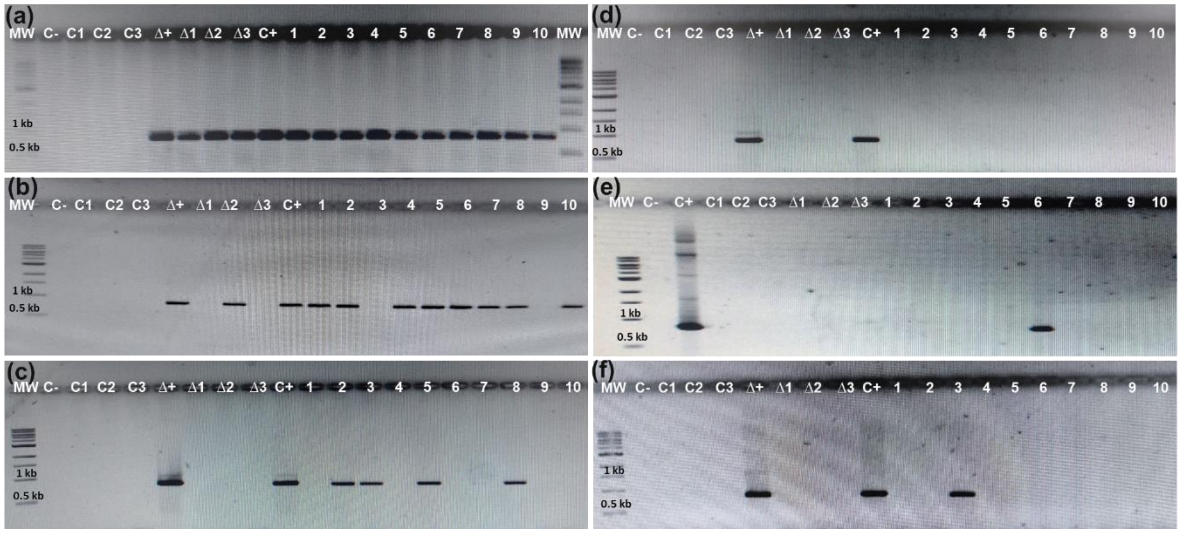


Figure 15. Viral replication assessment by PCR in host-range silencing assays. A 0.9 kb DNA fragment from the DNA-A component of ToMoV was observed in some samples. a) *N. benthamiana*. b) *S. lycopersicum*. c) *S. habrochaites* LA2128. d) *S. arcanum* LA2157. e) *S. arcanum* LA2172. f) *P. ixocarpa*. Lanes where the 0.9 kb DNA fragment was successfully amplified, corresponds to plants that showed bleaching phenotype; lanes where the 0.9 kb DNA fragment was not successfully amplified, corresponds to plants that did not show bleaching phenotype. MW: DNA ladder 1 kb; -C: negative control of reaction; +Δ: positive control, pC_H-ToMoV_*ChI*: C+: positive control, pC_H-ToMoV_ΔCP.

Since bleaching phenotype varied from one species to another, we carried out phenotypic assessments to determinate the silencing yields reached in each species (Senthil-Kumar *et al.*, 2007; Broderick and Jones 2014). To calculate the silencing frequency, the assays were set up in groups of ten plants per species. After performing two independent assays in *N. bentamiana*, each inoculated plant displayed a uniform bleaching phenotype, resulting in a 100% of silencing frequency at 28 dpi (Figure 16). A 65% and 35% was obtained in *S. lycopersicum* and *S. habrochaites* LA2128, respectively (Figure 16), and on the other hand a 10% of silencing frequency was reached for *S. arcanum* LA2172 and *P. ixocarpa* after two and one assays, respectively (Figure 16). In contrast, the silencing frequency in *S. arcanum* LA2157 was 0% after two independent assays since plants were not

silenced (Figures 14g-h, 15d). Thus, *S. arcanum* LA2157 was discarder for further analysis.

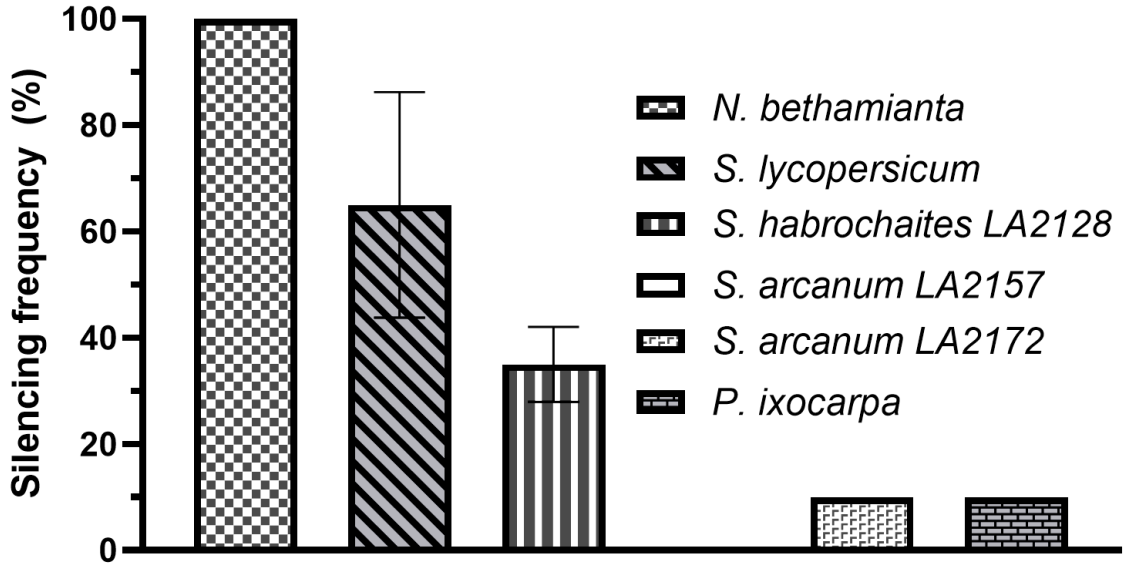


Figure 16. Silencing frequency of ToMoV hemidimers pC_H-ToMoV_*Chl*I and pBI_H-ToMoV_B inoculated by agroinfiltration in host-range silencing assays. The graph displays the percentage of silenced plants at 28 dpi. Silencing frequency resulted in 100% for *N bethamiana*, 65% for *S. lycopersicum,* 35% for *S. habrochaites* LA2128, 0% for *S. arcanum* LA2157 (non-silenced), 10% for *S. arcanum* LA2172 and 10% for *P. ixocarpa*. Except of *P. ixocarpa* (one assay), two independent assays were carried out for all plant species, with an n=10 per assay. Bars represent ±SD: Standard deviation.

Each solanaceous species exhibited a varied percentage of silenced leaves. To determine the silencing effectiveness, we analyzed the replicates that resulted in bleaching at 28 dpi from the first assay. In *N. bethamiana* 10 plants were analyzed and the 100% of leaves per plant presented bleaching. For eight *S. lycopersicum* replicates, the 69.97% of the leaves per plant exhibited bleaching. In *S. habrochaites* LA2128 three replicates were analyzed, resulting in 90.5% of silencing effectiveness. One plant was silenced from *S. arcanum* LA2172 and *P. ixocarpa* assays, showing an 82.05% and 57.14% of silencing effectiveness, respectively. These results

indicate a variation of the silencing effectiveness from 50 to 100% depending on the species (Figure 17).

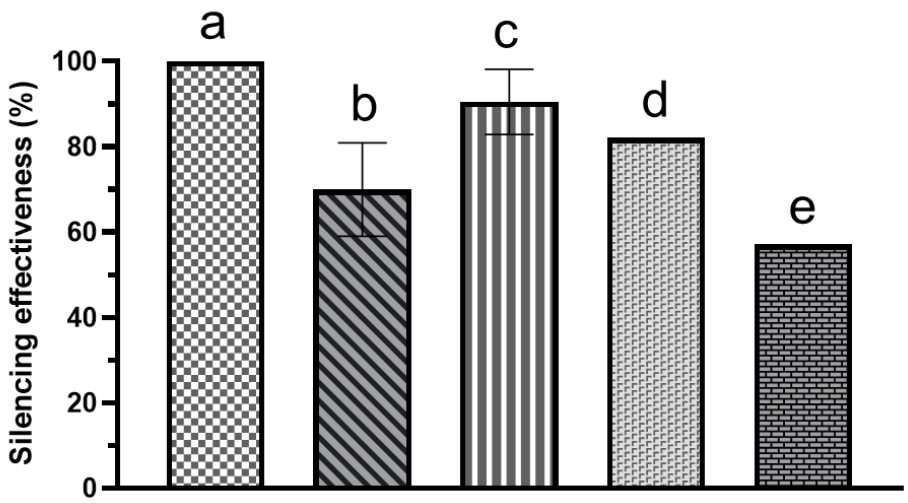


Figure 17. Silencing effectiveness of ToMoV hemidimers pC_H-ToMoV_*Chl*I and pBI_H-ToMoV_B inoculated by agroinfiltration in host-range silencing assays. The graph displays the percentage of silenced leaves per plant at 28 dpi. Silencing effectiveness resulted in: a) 100% for *N. bethamiana* (n=10); b) 69.97% for *S. lycopersicum* (n=8); c) 90.5% for *S. habrochaites* LA2128 (n=3); d) 82.05% for *S. arcanum* LA2172 (n=1); e) 57.14% for *P. ixocarpa* (n=1). Analyzed plants come from one of the independent assays employed to calculate silencing frequency. Bars represent ±SD: Standard deviation, calculated for those assays with an n ≥ 3.

Since the *Chl*I gene sequences for most of the plant species studied here are not completely known, it is difficult to determine the accurate transcript amounts in silenced plants. Therefore, we used chlorophyll content based on bleaching phenotypes as a parameter of silencing efficiency. To estimate the silencing efficiency of the pC_H-ToMoV_*Chl*I vector, we calculated the bleached area per leaf per plant species at 28 dpi (see materials and methods). The highest silencing efficiency was measured in *S. habrochaites* LA2128 and *N. bethamiana* with 88.3% and 85.5%, respectively (Figure 18). Followed by *S. lycopersicum* and *S. arcanum* LA2172, where the total bleached area varied from 62.8% to 52.9%, respectively. Leaves with less bleaching were calculated in *P. ixocarpa* with 42.5% in average

(Figure 18). Thus, the silencing efficiency varied from 40% up to 88% when the ToMoV hemidimers were inoculated by the agroinfiltration method using the *Agrobacterium* strain GV3101.

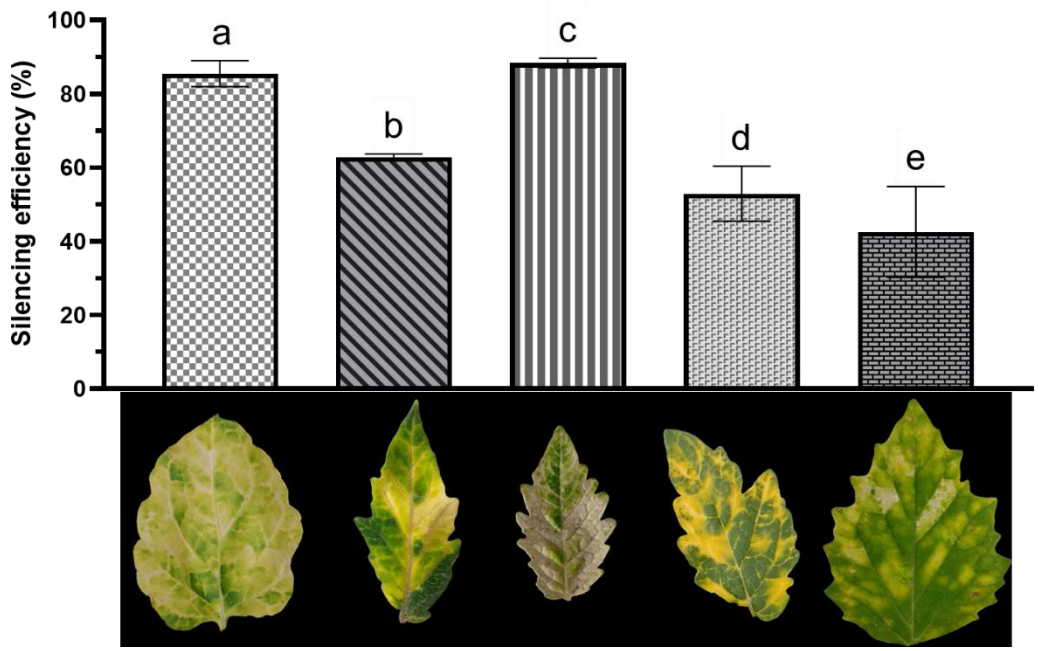


Figure 18. Silencing efficiency of ToMoV hemidimers pC_H-ToMoV_*Chl*I and pBI_H-ToMoV_B inoculated by agroinfiltration in host-range silencing assays. The graph displays the percentage of bleached area per leaf at 28 dpi. Silencing effectiveness resulted in: a) 85.5% for *N bethamiana*; b) 62.8% for *S. lycopersicum*; c) 88.3% for *S. habrochaites* LA2128; e) 52.9% for *S. arcanum* LA2172; e) 42.5% for *P. ixocarpa*. Three leaves were employed to calculate silencing efficiency. Bars represent ±SD: Standard deviation. Bleaching was quantified using the ImageJ software.

4. Discussion

In this work, we report a new optimized VIGS system based on the ToMoV_*Chl*I silencing vector previously designed by Esparza-Araiza *et al.* (2015). Our results reveal that using the agroinoculation method, the new ToMoV hemidimer induces gene silencing in the model organism *Nicotiana benthamiana*, the commercial species *Solanum lycopersicum* as well as related wild tomato species, and *Physalis ixocarpa* commonly known as “tomatillo” (Figure 14). Phenotypic measurements demonstrated that the employment of ToMoV hemidimer does not generates

significant side effects in silenced plants. This work highlights a new VIGS system that can be used for functional analysis in several plant species.

4.1. ToMoV-VIGS hemidimer an attractive tool for functional genomics in *N. bethamiana*

Particle bombardment is an excellent inoculation approach especially when host plants are recalcitrant to *Agrobacterium*-mediated inoculation or any other procedure, allowing the transformation of a wide variety of plant species (Luthra *et al.*, 1997). However, it is an expensive and time-consuming method to deliver VIGS systems into plants (Hagio, 1998; Rivera *et al.*, 2012), mainly for large-scale silencing assays. Besides, DNA degradation during the bombardment process has been reported (Krysiak *et al.*, 1999; Yoshimitsu *et al.*, 2009; Liu *et al.*, 2019), which could result in low silencing yields. The ToMoV_*Chl*I silencing vector (Esparza-Araiza *et al.*, 2015), is based on DNA-A and DNA-B components cloned as individual monomers into the pBluescript II SK (+/-) cloning vector. Such construction requires the use of restriction enzymes to release DNA-A and DNA-B components, in order to be suitable for particle bombardment inoculation. Therefore, we obtained an optimized ToMoV silencing vector based on the partial dimers approach, which involves adding an extra CR to each DNA component (Hayes *et al.*, 1988; Stenger *et al.*, 1991). The ToMoV hemidimers can be inoculated mechanically (Figure 11), but also by the less expensive and time-efficient leaf infiltration method (Figure 14).

We confirmed that ToMoV hemidimers were able to replicate and to induce gene silencing by following plant bleaching and viral replication in *N. benthamiana* plants inoculated by particle bombardment (Figure 11). This result shows that H-ToMoV-A_*Chl*I and H-ToMoV-B hemidimers induce gene silencing without the need to linearize the viral DNA before the inoculation, reaching about 66.6% of infectivity rate. Such a result agrees with other similar reports, in which dimeric infectious clones based on the begomovirus tomato yellow leaf curl virus (TYLCV), inoculated by particle bombardment can reach an average infectivity rate of 75% (Lapidot *et al.*, 2007). With the further improved construction of the ToMoV hemidimers pC_H-ToMoV_*Chl*I and pBI_H-ToMoV_B, which are cloned into binary vectors, we made

available agroinoculation for our VIGS system (Figure 12; see methods). Either the silencing vector H-ToMoV-A_*Chl*I or the pC_H-ToMoV_*Chl*I can be modified to induce gene silencing of plant endogenous genes. To do that, double digestion with the restriction enzymes *EcoR*I and *Hind*III allows the release of the *Chl*I DNA fragment cloned in both vectors (Figures 10, 12). The product obtained can be employed to insert the desired DNA fragment or to insert a multiple cloning site to make numerous VIGS constructions. To design VIGS constructs for gene silencing we recommend the VIGS tool available at https://vigs.solgenomics.net/ (Fernandez-Pozo *et al.*, 2015).

4.2. ToMoV-VIGS hemidimer acts in a wide range of solanaceous species

The co-inoculation of the ToMoV vectors pC_H-ToMoV_*Chl*I and pBI_H-ToMoV_B by the leaf infiltration using the *A. tumefaciens* strain GV3101 results in VIGS of the *Chl*I gene in five solanaceous (Figure 14). The *Chl*I fragment inserted into ToMoV silencing vector corresponds to a fragment obtained from *S. arcanum* LA2172 (Esparza-Araiza *et al.*, 2015), which shares about 54% to 100% identity to their corresponding heterologous *Chl*I gene fragments of the species studied here (Figure 19). Our results demonstrate that *N. bethamiana* plants inoculated by leaf infiltration render in a 100% of silencing frequency (Figure 16), a substantial improvement compared to the 66.6% when the plants are inoculated by particle bombardment (Figure 11). As expected, we detected viral replication in the control replicates inoculated with the pC_H-ToMoVΔCP by PCR and no disease symptom was identified, indicating that ToMoV hemidimer does not produce significant side effects (Figures 14a, 15a). Equally, 100% of leaves showed silencing, in which the bleached area per leave was 85.5% on average at 28 dpi (Figures 17a, 18a). Such silencing yields are slightly higher from those obtained when *N. bethamiana* plants are silenced in the phytoene desaturase gene (*PDS*) with the TRV-based silencing system of an RNA virus (Senthil-Kumar *et al.,* 2007), suggesting that our ToMoV VIGS system can be an excellent option for functional genomics in *N. benthamiana*.

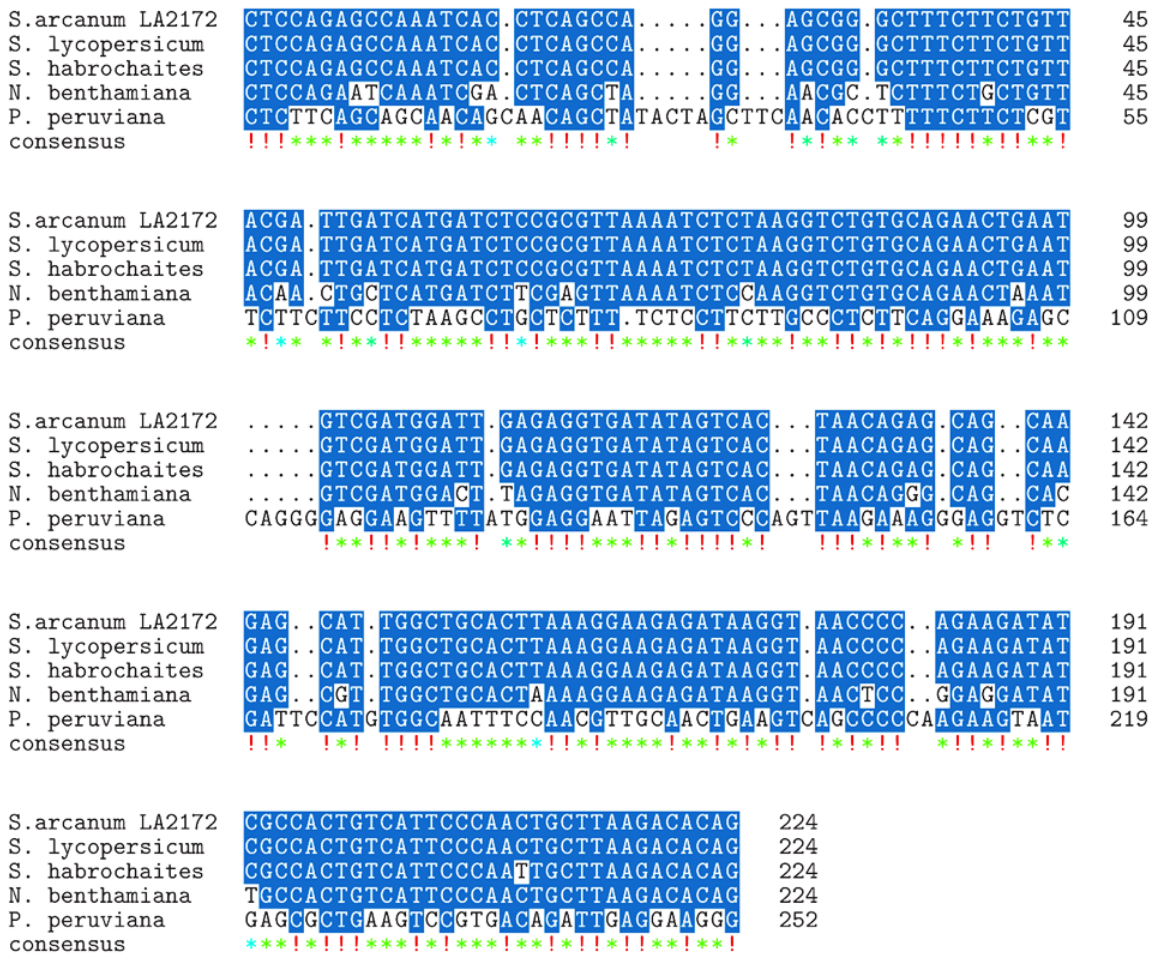


Figure 19. Percentage identities of *S. arcanum* LA2172 *Chl*I gene fragment and heterologous *Chl*I gene sequences of silenced solanaceous species. The *S. arcanum* LA2172 *Chl*I gene fragment shares a 100 and 99.55% percent identity with *S. lycopersicum* and *S. habrochaites,* respectively; an 88.39% percent identity with *N. bethamiana* and 54% with *Physalis peruviana*, the closest species related to *P. ixocarpa* with a *Chl*I gene sequence available in public data bases. Identity percentages were calculated with the multiple sequence alignment software T-Coffee, with default parameters.

Several VIGS systems have been employed to carry out functional genomics in *S. lycopersicum* (Liu *et al.,* 2002a,b; Cai *et al.*, 2007; Golenberg *et al.*, 2009), but only a few have been reported for tomato related wild species, mainly by the employment of TRV-based silencing vectors (Jablonska *et al.*, 2007; Slocombe *et al.*, 2008). Wild tomato species represent an important natural source of genetic resistance for abiotic and biotic stress (Rick and Chetelat 1995; Bolger *et al.*, 2014). Hence, the development of new functional analysis tools for their study is crucial. As our results indicate, viral replication was not detected in *S. arcanum* LA2157 (Figures 14g-h,

15d); therefore, the *Chl*I gene silencing was not induced. Since other report shows effective VIGS induction with TRV-based silencing system using *A. tumefaciens* strain GV3101 for inoculation (Jablonska *et al.,* 2007), we employed the same strain. Additionally, we applied particle bombardment for ToMoV inoculation in LA2157 (unpublished data) and no viral replication was detected. These results suggest that *S. arcanum* LA2157 is a resistant accession to ToMoV. On the other hand, we demonstrated that VIGS of *S. lycopersicum Chl*I gene reached up to 65% of threated plants when inoculated with the ToMoV hemidimers and an average of 35% and 10% of silencing frequency was measured in the wild species *S. habrochaites* LA2128 and *S. arcanum* LA2172, respectively (Figures 14 and 16). Once the *Chl*I gene silencing was established, the silencing effectiveness oscillated from 70% to 90% among these three tomato species at 28 dpi, being higher in both wild tomato species (Figure 17). Correspondingly, the silencing efficiency varied from 52% up to 88% in *S. habrochaites* at 28 dpi (Figure 18). In terms of silencing effectiveness and efficiency, silencing yields obtained with the new ToMoV hemidimer are comparable to other silencing systems applied in *S. lycopersicum* (Liu *et al*., 2002a, b; Senthil-Kumar and Mysore 2011). Altogether, our ToMoV-based silencing vector represents a valuable functional genomics tool for several *Solanum* species.

Physalis ixocarpa is an important crop particularly in Central America; however, only a limited number of methods for functional analysis are available for its study (e.g., Assad-García *et al.*, 1992). Here we show that our ToMoV hemidimer can induce gene silencing in *P. ixocarpa* with about 42.5% of silencing efficiency, but with a silencing frequency of up to 10% (Figures 16-18). Remarkably, we observed a clear bleached phenotype, even though *S. arcanum Chl*I sequence shares a 54% identity to *Phisalys peruviana,* which is the closest related species to *P. ixocarpa* with a *Chl*I sequence available in public data bases (Figure 19). Thus, we can hypothesize that the sequence identity with *P. ixocarpa Chl*I must be similar. This suggests that the addition of DNA fragments for gene silencing with an identity percent close to 100% must result in higher silencing efficiency. Even though the silencing frequency is not very high, the ToMoV silencing vectors represent a new tool that can be adapted for functional genomics in “tomatillo”.

Since silencing frequencies were relatively low in *S. habrochaites*, *S. arcanum* LA2172 and *P. ixocarpa,* certain adjustments need to be done. For example, according to Esparza-Araiza *et al.,* (2015), the silencing frequency obtained in *S. arcanum* LA2172 inoculated with the ToMoV_*Ch*I monomer by particle bombardment seems to be higher compared to our results; hence, the inoculation on the ToMoV hemidimer by particle bombardment may result in higher silencing frequency for *S. arcanum* LA2172, since the ToMoV hemidimer does not need of restriction enzymes to be released from the cloning vector. This assay needs to be tested. However, for some plant species, the employment of vacuum infiltration, *Agrobacterium* spray, and other *Agrobacterium* delivery methods results in higher number of silenced plants (Liu *et al.*, 2002a; Yan *et al.*, 2012). Moreover, by testing different *Agrobacterium* strains, better effects in the VIGS system can be obtained (e.g., Zeng *et al.*, 2019), as well as testing different temperature growth conditions (Brigneti *et al.*, 2004; Broderick and Jones 2014). Therefore, by adjusting such factors to a particular host plant, our ToMoV silencing vector can be adapted to improve silencing assays.

5. Concluding remarks

In conclusion, our results obtained with the host-range silencing analyses highlight valuable evidence about the potential applications of ToMoV-based silencing vectors for functional genomics. We demonstrated that the leaf infiltration of ToMoV hemidimer serves as a better inoculation method for some plant species than particle bombardment. According to the vast evidence originated from numerous VIGS vectors, it is necessary to carry out particular adjustments, such as control of temperature and *Agrobacterium* delivery methods, in order to reach higher silencing yields depending on the subject plant species. Considering that ToMoV has a broad spectrum of plant hosts (Polston *et al.,* 1993), surely this silencing vector based on hemidimeric clones, could be applied for large scale silencing assays in model plants such as *N. bethamiana*, but also in less studied species such as “tomatillo” and other plant species not included in this work.

6. Bibliography

Abouzid AM, Polston JE, Hiebert E. (1992) The nucleotide sequence of tomato mottle virus, a new geminivirus isolated from tomatoes in Florida. Journal of General Virology 73, 3225–3229. https://doi.org/10.1099/0022-1317-73-12-3225

Abrahamian P, Hammond RW, Hammond J. (2020) Plant virus-derived vectors: applications in agricultural and aedical biotechnology. Annual Review of Virology https://doi.org/10.1146/annurev-virology-010720-054958

Aregger M, Borah BK, Seguin J, Rajeswaran R, Gubaeva EG, Zvereva AS, Windels D, Vazquez F, Blevins T, Farinelli L, Pooggin MM. (2012) Primary and secondary siRNAs in geminivirus-induced gene silencing. PLoS Pathogens 8, e1002941. https://doi.org/10.1371/journal.ppat.1002941

Assad-García N, Ochoa-Alejo N, García-Hernández E, Herrera-Estrella L, Simpson J. (1992) *Agrobacterium*-mediated transformation of tomatillo (*Physalis ixocarpa*) and tissue specific and developmental expression of the CaMV 35S promoter in transgenic tomatillo plants. Plant Cell Reports 11, 558-562. https://doi.org/10.1007/BF00233092

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409, 363–366. https://doi.org/10.1038/35053110

Bolger A, Scossa F, Bolger M *et al*. (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nature Genetics 46, 1034–1038. https://doi.org/10.1038/ng.3046

Bouché N, Lauressergues D, Gasciolli V, Vaucheret H. (2006) An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. The EMBO Journal 25, 3347–3356. https://doi.org/10.1038/sj.emboj.7601217

Brigneti G, Martín-Hernández AM, Jin H, Chen J, Baulcombe DC, Baker B, Jones JD. (2004) Virus-induced gene silencing in *Solanum* species. The Plant Journal 39, 264–272. https://doi.org/10.1111/j.1365-313X.2004.02122.x

Broderick SR, Jones ML. (2014) An Optimized protocol to increase virus-induced gene silencing efficiency and minimize viral symptoms in petunia. Plant Molecular Biology Reports 32, 219–233. https://doi.org/10.1007/s11105-013-0647-3

Brooks C, Nekrasov V, Lippman ZB, Van Eck J. (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. Plant Physiology 166, 1292–1297. https://doi.org/10.1104/pp.114.247577

Cai XZ, Wang CC, Xu YP, Xu QF, Zheng Z, Zhou XP. (2007) Efficient gene silencing induction in tomato by a viral satellite DNA vector. Virus Research 125, 169–175. https://doi.org/10.1016/j.virusres.2006.12.016

Carbonell A, Fahlgren N, Garcia-Ruiz H, Gilbert KB, Montgomery TA, Nguyen T, Cuperus JT, Carrington JC. (2012) Functional analysis of three *Arabidopsis* ARGONAUTES using slicer-defective mutants. Plant Cell 24, 3613–3629. https://doi.org/10.1105/tpc.112.099945

Chellappan P, Vanitharani R, Pita J, Fauquet CM. (2004) Short interfering RNA accumulation correlates with host recovery in DNA virus-infected hosts, and gene silencing targets specific viral sequences. Journal of Virology 78, 7465–7477. https://doi.org/10.1128/JVI.78.14.7465-7477.2004

Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC. (2000) An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101, 543–553. https://doi.org/10.1016/s0092-8674(00)80864-8

Dommes AB, Gross T, Herbert DB, Kivivirta KI, Becker A. (2019) Virus-induced gene silencing: empowering genetics in non-model organisms. Journal of Experimental Botany 70, 757–770. https://doi.org/10.1093/jxb/ery411

Esparza-Araiza MJ, Bañuelos-Hernández B, Argüello-Astorga GR, Lara-Ávila JP, Goodwin PH, Isordia-Jasso MI, Castillo-Collazo R, Rougon-Cardoso A, Alpuche-Solís ÁG. (2015) Evaluation of a SUMO E2 conjugating enzyme involved in resistance to *Clavibacter michiganensis* subsp. *michiganensis* in *Solanum peruvianum*, through a Tomato Mottle Virus VIGS assay. Frontiers in Plant Sciences 6, 1019. https://doi.org/10.3389/fpls.2015.01019

English JJ, Mueller E, Baulcombe DC. (1996) Suppression of virus accumulation in transgenic plants exhibiting silencing of nuclear genes. Plant Cell 8, 179–188. https://doi.org/10.1105/tpc.8.2.179

Fernandez-Pozo N, Rosli HG, Martin GB, Mueller LA. (2015) The SGN VIGS tool: user-friendly software to design virus-induced gene silencing (VIGS) constructs for functional genomics. Molecular Plant 8, 486–488. https://doi.org/10.1016/j.molp.2014.11.024

Frischmuth S, Frischmuth T, Jeske H. (1991) Transcript mapping of abutilon mosaic virus, a geminivirus. Virology, 185, 596–604. https://doi.org/10.1016/0042-6822(91)90530-o

Garcia-Ruiz H, Takeda A, Chapman EJ, Sullivan CM, Fahlgren N, Brempelis KJ, Carrington JC. (2010) *Arabidopsis* RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during turnip mosaic virus infection. Plant Cell 22, 481–496. https://doi.org/10.1105/tpc.109.073056.

Garcia-Ruiz H, Carbonell A, Hoyer JS, Fahlgren N, Gilbert KB, Takeda A, Giampetruzzi A, Garcia Ruiz MT, McGinn MG, Lowery N, Martinez Baladejo MT, Carrington JC. (2015) Roles and programming of *Arabidopsis* ARGONAUTE proteins during turnip mosaic virus infection. PLoS Pathogens 11, e1004755. https://doi.org/10.1371/journal.ppat.1004755

Golenberg EM, Sather DN, Hancock LC, Buckley KJ, Villafranco NM, Bisaro DM. (2009) Development of a gene silencing DNA vector derived from a broad host range geminivirus. Plant Methods 5, 9. https://doi.org/10.1186/1746-4811-5-9

Hagio T. (1998) Optimizing the particle bombardment method for efficient genetic transformation. Japan Agricultural Research Quarterly 32, 239–247.

Hamilton AJ, Baulcombe DC. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286, 950–952. https://doi.org/10.1126/science.286.5441.950

Hayes R, Petty I, Coutts R, Buck K. (1988) Gene amplification and expression in plants by a replicating geminivirus vector. Nature 334, 179–182. https://doi.org/10.1038/334179a0

Himber C, Dunoyer P, Moissiard G, Ritzenthaler C, Voinnet O. (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. The EMBO Journal 22, 4523–4533. https://doi.org/10.1093/emboj/cdg431

Hsieh MH, Lu HC, Pan ZJ, Yeh HH, Wang SS, Chen WH, Chen HH. (2013) Optimizing virus-induced gene silencing efficiency with Cymbidium mosaic virus in *Phalaenopsis* flower. Plant Science 201-202, 25–41. https://doi.org/10.1016/j.plantsci.2012.11.003

Ibrahim A, Odon V, Kormelink R. (2019) Plant viruses in plant molecular pharming: toward the use of enveloped viruses. Frontiers in Plant Sciences 10, 803. https://doi.org/10.3389/fpls.2019.00803

Jablonska B, Ammiraju JS, Bhattarai KK, Mantelin S, Martinez de Ilarduya O, Roberts PA, Kaloshian I. (2007) The Mi-9 gene from *Solanum arcanum* conferring heat-stable resistance to root-knot nematodes is a homolog of Mi-1. Plant Physiology 143, 1044–1054. https://doi.org/10.1104/pp.106.089615

Kim KH, Lim S, Kang YJ, Yoon MY, Nam M, Jun TH, Seo MJ, Baek SB, Lee JH, Moon JK, Lee SH, Lee SH, Lim HS, Moon JS, Park CH. (2016) Optimization of a virus-induced gene silencing system with soybean yellow common mosaic virus for gene function studies in soybeans. The Plant Pathology Journal 32, 112–122. https://doi.org/10.5423/PPJ.OA.04.2015.0063

Kjemtrup S, Sampson KS, Peele CG, Nguyen LV, Conkling MA, Thompson WF, Robertson D. (1998) Gene silencing from plant DNA carried by a Geminivirus. The Plant Journal 14, 91–100. https://doi.org/10.1046/j.1365-313X.1998.00101.x

Krysiak C, Mazus B, Buchowicz J. (1999) Relaxation, linearization and fragmentation of supercoiled circular DNA by tungsten microprojectiles. Transgenic Research 8, 303–306. https://doi.org/10.1023/A:1008990712122

Kumagai MH, Donson J, della-Cioppa G, Harvey D, Hanley K, Grill LK. (1995) Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. Proceedings of the National Academy of Sciences of the United States of America 92, 1679-1683. https://doi.org/10.1073/pnas.92.5.1679

Lapidot M, Weil G, Cohen L, Segev L, Gaba V. (2007) Biolistic inoculation of plants with tomato yellow leaf curl virus DNA. Journal of Virological Methods 144, 143–148. https://doi.org/10.1016/j.jviromet.2007.04.011

Liu J, Nannas NJ, Fu FF, Shi J, Aspinwall B, Parrott WA, Dawe RK. (2019) Genome-scale sequence disruption following biolistic transformation in rice and maize. Plant Cell, 31, 368–383. https://doi.org/10.1105/tpc.18.00613

Liu Y, Schiff M, Dinesh-Kumar SP. (2002a) Virus-induced gene silencing in tomato. The Plant Journal 31, 777–786. https://doi.org/10.1046/j.1365-313x.2002.01394.x

Liu Y, Schiff M, Marathe R, Dinesh-Kumar SP. (2002b) Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. The Plant journal, 30, 415–429. https://doi.org/10.1046/j.1365-313x.2002.01297.x

Lu R, Martin-Hernandez A M, Peart JR, Malcuit I, Baulcombe DC. (2003) Virus-induced gene silencing in plants. Methods 30, 296–303. https://doi.org/10.1016/s1046-2023(03)00037-9

Luthra R, Varsha DRK, Srivastava AK, Kumar S. (1997) Microprojectile mediated plant transformation: a bibliographic search. Euphytica 95, 269–294. https://doi.org/10.1023/A:1002957229911

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, Wang B, Yang Z, Li H, Lin Y, Xie Y, Shen R, Chen S, Wang Z, Chen Y, Guo J, Chen L, Zhao X, Dong Z, Liu YG. (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Molecular Plant 8, 1274–1284. https://doi.org/10.1016/j.molp.2015.04.007

Molnár A, Csorba T, Lakatos L, Várallyay E, Lacomme C, Burgyán J. (2005) Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. Journal of Virology 79, 7812–7818. https://doi.org/10.1128/JVI.79.12.7812-7818.2005

Morel JB, Godon C, Mourrain P, Béclin C, Boutet S, Feuerbach F, Proux F, Vaucheret H. (2002) Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. Plant Cell, 14, 629–639. https://doi.org/10.1105/tpc.010358

Mueller E, Gilbert J, Davenport G, Brigneti G, Baulcombe DC. (1995) Homology-dependent resistance: transgenic virus resistance in plants related to homology-dependent gene silencing. The Plant Journal 7, 1001-1013. https://doi.org/10.1046/j.1365-313X.1995.07061001.x

Pandey P, Choudhury NR, Mukherjee SK. (2009) A geminiviral amplicon (VA) derived from tomato leaf curl virus (ToLCV) can replicate in a wide variety of plant species and also acts as a VIGS vector. Virology Journal 6, 152. https://doi.org/10.1186/1743-422X-6-152

Palauqui JC, Elmayan T, Pollien JM, Vaucheret H. (1997) Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. The EMBO Journal 16, 4738–4745. https://doi.org/10.1093/emboj/16.15.4738

Polston JE, Hiebert E, McGovern RJ, Stansly PA and Schuster DJ. (1993) Host range of tomato mottle virus, a new geminivirus infecting tomato in Florida. Plant Disease 77, 1181-1184. DOI: 10.1094/PD-77-1181.

Ratcliff F, Martin-Hernandez AM, Baulcombe DC. (2001) Technical Advance. tobacco rattle virus as a vector for analysis of gene function by silencing. The Plant Journal 25, 237-245. doi:10.1046/j.0960-7412.2000.00942.x

Rick CM, Chetelat RT. (1995) Utilization of related wild species for tomato improvement. Acta Horticulturae 412, 21–38. https://doi.org/10.17660/ActaHortic.1995.412.1

Rivera AL, Gómez-Lim M, Fernández F, Loske AM. (2012) Physical methods for genetic plant transformation. Physics of Life Reviews 9, 308-345. https://doi.org/10.1016/j.plrev.2012.06.002

Rybicki EP. (2015) A Top Ten list for economically important plant viruses. Archives of Virology 160, 17–20. https://doi.org/10.1007/s00705-014-2295-9

Saunders K, Lucy A, Stanley J. (1991) DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. Nucleic Acids Research 19, 2325–2330. https://doi.org/10.1093/nar/19.9.2325

Scholthof KBG, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, Hohn B, Saunders K, Candresse T, Ahlquist P, Hemenway C, Foster GD. (2011) Top 10 plant viruses in molecular plant pathology. Molecular Plant Pathology 12, 938–954. https://doi.org/10.1111/j.1364-3703.2011.00752.x

Schneider CA, Rasband WS, Eliceiri KW. (2012) "NIH Image to ImageJ: 25 years of image analysis". Nature Methods 9, 671-675. https://doi.org/10.1038/nmeth.2089

Senthil-Kumar M, Hema R, Anand A, Kang L, Udayakumar M, Mysore KS. (2007) A systematic study to determine the extent of gene silencing in *Nicotiana benthamiana* and other Solanaceae species when heterologous gene sequences are used for virus-induced gene silencing. New Phytologist 176, 782–791. https://doi.org/10.1111/j.1469-8137.2007.02225.x

Senthil-Kumar M, Mysore KS. (2011) Virus-induced gene silencing can persist for more than 2 years and also be transmitted to progeny seedlings in *Nicotiana benthamiana* and tomato. Plant Biotechnology Journal 9, 797–806. https://doi.org/10.1111/j.1467-7652.2011.00589.x

Sijen T, Wellink J, Hiriart JB, Van Kammen A. (1996) RNA-mediated virus resistance: Role of repeated transgenes and delineation of targeted regions. Plant Cell 8, 2277–2294. https://doi.org/10.1105/tpc.8.12.2277

Singh DK, Lee HK, Dweikat I, Mysore KS. (2018) An efficient and improved method for virus-induced gene silencing in sorghum. BMC plant biology 18, 123. https://doi.org/10.1186/s12870-018-1344-z

Slocombe SP, Schauvinhold I, McQuinn RP, Besser K, Welsby NA, Harper A, Aziz N, Li Y, Larson TR, Giovannoni J, Dixon RA, Broun P. (2008) Transcriptomic and reverse genetic analyses of branched-chain fatty acid and acyl sugar production in *Solanum pennellii* and *Nicotiana benthamiana*. Plant Physiology 48, 1830–1846. https://doi.org/10.1104/pp.108.129510

Smith HA, Swaney SL, Parks TD, Wernsman EA, Dougherty WG. (1994) Transgenic plant virus resistance mediated by untranslatable sense RNAs: expression, regulation, and fate of nonessential RNAs. Plant cell 6, 1441–1453. https://doi.org/10.1105/tpc.6.10.1441

Stenger DC, Revington GN, Stevenson MC, Bisaro DM. (1991) Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. Proceedings of the National Academy of Sciences of the United States of America 88, 8029–8033. https://doi.org/10.1073/pnas.88.18.8029

Stanley J. (1983) Infectivity of the cloned geminivirus genome requires sequences from both DNAs. Nature 305, 643- 645. https://doi.org/10.1038/305643a0

Stanley J, Gay MR. (1983) Nucleotide sequence of cassava latent virus DNA. Nature 301, 260-262. https://doi.org/10.1038/301260a0

Townsend R, Stanley J, Curson SJ, Short MN. (1985) Major polyadenylated transcripts of cassava latent virus and location of the gene encoding coat protein. The EMBO Journal 4, 33–37.

Turnage MA, Muangsan N, Peele CG, Robertson D. (2002) Geminivirus-based vectors for gene silencing in *Arabidopsis*. The Plant Journal 30, 107–114. https://doi.org/10.1046/j.1365-313x.2002.01261.x

Vanitharani R, Chellappan P, Fauquet CM. (2005) Geminiviruses and RNA silencing. Trends Plant Sciences 10, 144–151. https://doi.org/10.1016/j.tplants.2005.01.005

Yan HX, Fu DQ, Zhu BZ, Liu HP, Shen XY, Luo YB. (2012) Sprout vacuum-infiltration: a simple and efficient agroinoculation method for virus-induced gene silencing in diverse solanaceous species. Plant Cell Reports 31, 1713–1722. https://doi.org/10.1007/s00299-012-1285-1

Yoshimitsu Y, Tanaka K, Tagawa T, Nakamura Y, Matsuo T, Okamoto S. (2009) Improvement of DNA/metal particle adsorption in tungsten-based biolistic bombardment; alkaline pH is necessary for DNA adsorption and suppression of DNA degradation. Journal of Plant Biology 52, 524–532. https://doi.org/10.1007/s12374-009-9068-0

Zeng H, Xie Y, Liu G, Wei Y, Hu W, Shi H. (2019*) Agrobacterium*-mediated gene transient overexpression and tobacco rattle virus (TRV)-based gene silencing in cassava. International Journal of Molecular Sciences 20, 3976. https://doi.org/10.3390/ijms20163976

Annexes

Pereyra-Bistraín LI, Ovando-Vázquez C, Rougon-Cardoso A, Alpuche-Solís ÁG. (2021) Comparative RNA-Seq analysis reveals potentially resistance-related genes in response to bacterial canker of tomato. Genes 12, 1745. https://doi.org/10.3390/genes12111745