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

POSGRADO EN CIENCIAS EN BIOLOGIA MOLECULAR

Predictive and *in-silico* functional analysis of small RNAs in the *Trichoderma atroviride*-*Arabidopsis thaliana* interaction

Tesis que presenta

Daniel Rafael Saldaña Torres

Para obtener el grado de

Maestro en Ciencias en Biología Molecular

Codirectores de la Tesis:

Dr. J. Sergio Casas Flores

Dr. Cesaré Moisés Ovando Vázquez

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

Constancia de aprobación de la tesis

La tesis "**Predictive and *in-silico* functional analysis of small RNAs in the *Trichoderma-Arabidopsis* interaction**" presentada para obtener el Grado de Maestro en Ciencias en Biología Molecular fue elaborada por **Daniel Rafael Saldaña Torres** y aprobada el **22 de Junio 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. J. Sergio Casas Flores
Codirector de la tesis

_____

Dr. Cesaré Moisés Ovando Vázquez
Codirector de la tesis

_____

Dr. Gerardo R. Argüello Astorga
Miembro del Comité Tutoral

Créditos Institucionales

Esta tesis fue elaborada en el Laboratorio 8 de Genómica Funcional y Comparativa de la División de Biología Molecular, y el Laboratorio de Inteligencia Artificial y Bioinformática del Centro Nacional de Supercómputo, ambos del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la codirección de los doctores J. Sergio Casas Flores y Cesaré Moisés Ovando Vázquez.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología con número de registro: (1007772) y del Instituto Potosino de Investigación Científica y Tecnológica, A. C. Este trabajo fue apoyado en sus inicios por los proyectos UC-MEXUS CN-14-6 2014 y FC-2016-1538 otorgado a J. Sergio Casas-Flores.

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 220 del Libro Segundo de Actas de Exámenes de Grado del Programa de Maestría en Ciencias en Biología Molecular está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 12 días del mes de julio 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. Cesaré Moises Ovando Vázquez	**Secretario**	**IPICYT**
Dr. J. Sergio Casas Flores	**Sinodal**	**IPICYT**

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

MAESTRO EN CIENCIAS EN BIOLOGÍA MOLECULAR

sustentó el C.

Daniel Rafael Saldaña Torres

sobre la Tesis intitulada:

Predictive and in-silico functional analysis of small RNAs in the Trichoderma atroviride-Arabidopsis thaliana interaction

que se desarrolló bajo la dirección de

Dr. Cesaré Moises Ovando Vázquez
Dr. J. Sergio Casas Flores

El Jurado, después de deliberar, determinó

APROBARLO

Dándose por terminado el acto a las 14:14 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 12 días del mes de julio 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

A mi familia, que siempre me ha apoyado en todo lo que me he dedicado a realizar. Gracias

Agradecimientos

Al Dr. Cesaré Moisés Ovando Vázquez, por ser mi mentor en al área de la bioinformática y la estadística, a su vez de ser una gran fuente de inspiración para mi futuro.

Al Dr. J Sergio Casas Flores, por siempre ayudarme a no olvidar el énfasis biológico de este proyecto y atender siempre mis dudas.

Al equipo de trabajo del Laboratorio de Inteligencia Artificial y Bioinformática, por siempre hacer mis días de trabajo momentos amenos.

Al Dr. Gerardo Rafael Argüello Astorga por darse el tiempo y la paciencia de leer mi proyecto y revisarlo.

Al Laboratorio 8 de Genómica Funcional y Comparativa por los datos biológicos con los cuales se generó este proyecto y su siempre atenta colaboración a la mejora de esta tesis.

A mis codirectores, por tomarse el tiempo y paciencia para poder perfeccionar mis habilidades de escritura en la realización de esta tesis.

A mis compañeros: Leonardo Daniel Villalana Alvarez, Antonio de Jesús Barrera Juárez, Ares Orlando Cuellar Santoyo y Raquel Angélica Castorena Vazuez, por su valiosa amistad y compañía.

A Genessis Montalvo Guevara, por su cariño, apoyo y cuidado a mi salud en todo momento durante todo este proyecto.

Al Centro Nacional de Supercómputo del IPICYT, por brindarme acceso a la supercomputadora para el desarrollo de análisis de datos masivos empleados en este proyecto mediante el proyecto TKII-R2018-COV1.

Contenido

Lista de tablas

Lista de figuras

Anexos

Abreviaturas

AGO	Argonaute protein
AGO RIP-Seq	AGO-RNA immunoprecipitation followed by high-throughput Sequencing
ANG	Angiogenin
***At*-ck-tRF**	*A. thaliana* cross-kingdom tRF
Aux/IAA	Auxin/Indole-3-Acetic Acid
Avr	Avirulence factor
Bc-sRNA	*Botrytis cinerea* sRNA
Bj-tRF	*Bradyrhizobium japonicum* tRF
BP	Biological Process
cDNA	complementary DNA
ck-RNAi	cross-kingdom RNAi
ck-sRNA	cross-kingdom sRNA
CPM	Counts Per Million
DCL	Dicer-Like ribonucleases
DCTN1	Dynactin complex
DE	Differentially Expressed
DNA	Deoxyribonucleic Acid
dsRNA	double-stranded RNA
Dusp1	Dual specificity protein phosphatase 1
ETI	Effector-Triggered Immunity
ETS	Effector-Triggered Susceptibility
GO	Gene Ontology
HAM4	Hairy Meristem 4
hcc	hours of coculture
hpi	hours post-inoculation
HPPD	4-hydroxy-phenyl pyruvate dioxygenase
hpRNA	hairpin structured RNA
HR	Hypersensitive Response
Il33r	interleukin 1 receptor like 1
lncRNAs	long non-coding RNAs
LRX5	Leucine-Rich repeat extension-like 5
MAMP	Microbial Associated Molecular Pattern
MAPKKK4	Mitogen-Activated Protein Kinase Kinase Kinase 4

MiSSP	Mycorrhiza-induced Small-Secreted Protein
MPK	Mitogen-activated Protein Kinase
MTI	MAMP Triggered Immunity
MVB	Multi-Vesicular Bodies
NB-LRR	Nucleotide-Binding and Leucine-Rich Repeat protein
ncRNA	non-coding RNA
nt	Nucleotides
PAMP	Pathogen Associated Molecular Pattern
PKA-R	cAMP-dependent protein kinase
PRR	Pattern Recognition Receptor
PTGS	Post-Transcriptional Gene Silencing
PTI	PAMP Triggered Immunity
qRT-PCR	quantitative Reverse Transcription Polymerase Chain Reaction
RdDM	RNA-directed DNA Methylation
RHD3	Root Hair Directive 3
RISC	RNA-Induced Silencing Complex
RNA	Ribonucleic Acid
RNAi	RNA interference
RNS	RNAses T2
rRF	rRNA derived RNA fragment
rRNA	ribosomal RNA
SAC1	Suppressor of Actin
snoRNAs	small nucleolar RNAs
snRNAs	small nuclear RNAs
sRNAs	small RNAs
STTM	Short Tandem Target Mimic
***Ta*-ck-tRF**	*T. atroviride* cross-kingdom tRF
TGS	Transcriptional Gene Silencing
tRF	tRNA derived RNA fragment
tRH	tRNA-Halve

tRNA	transfer RNA
Vps51	Vacuolar protein sorting 51

Resumen

Análisis *in-silico* predictivo y funcional de RNAs pequeños en la interacción *Trichoderma-Arabidopsis*.

Las plantas y sus microorganismos han desarrollado mecanismos moleculares, producto de su coevolución, los cuales son fundamentales para poder establecer una relación simbiótica entre ellos. Ejemplos de estos mecanismos es el uso de RNAs pequeños (sRNAs) para modular la expresión génica entre reinos, suceso denominado ck-RNAi (por sus palabras en inglés cross-kingdom RNAi). La biogénesis canónica de un sRNA involucra el reconocimiento de una molécula precursora de RNA la cual madura a un RNA < 200 nucleótidos (nt) de longitud. Esto mediante la participación de proteínas como las ribonucleasas tipo III, denominadas en plantas como proteínas similares a Dicer (DCLs, por sus siglas en inglés). Recientemente se ha reportado que los sRNAs pueden ser generados a partir de moléculas de tRNA o rRNA, pero poco se conoce de sus mecanismos moleculares. De la misma manera se tiene poco conocimiento acerca de ck-RNAi en interacciones simbióticas de origen mutualista. Se estudiaron 8 genotecas de secuenciación masiva de sRNAs en la interacción *Trichoderma-Arabidopsis* durante 24, 48, 72 y 96 horas de cocultivo (hcc) y 18 genotecas de secuenciacion masiva de mRNAs a 48, 72 y 96 hcc. Se implementaron análisis bioinformáticos para detectar sRNAs que posiblemente participen en un evento de ck-RNAi; determinar su organismo productor, moléculas precursoras y sus posibles funciones. A su vez también se infirió una red de co-expresión sRNAs-mRNAs para identificar las diversas asociaciones (co-expresión y regulación) entre estos. Durante los tiempos de interacción, *Arabidopsis* y *Trichoderma* produjeron diferentes cantidades de sRNAs. Aquellos sRNAs que presentaron mayores cambios a lo largo de los tiempos estudiados fueron anotados como tRFs. Los blancos predichos de estos tRFs están involucrados en la defensa de la planta, en su actividad metabólica y de traduccion. Se observó una red de regulación de estos tRFs con diversos blancos al compararlos con datos de expresión génica, siendo principalmente genes asociados involucrados en la defensa de la planta. Los resultados sugieren una posible comunicación ck-RNAi principalmente compuesta de fragmentos de RNA derivados de tRNAs (tRFs), siendo producidos predominantemente por *Arabidopsis* durante 48 y 72 hcc y por *Trichoderma* durante las 96 hcc, afectando principalmente procesos de defensa en la planta.

PALABRAS CLAVE. Simbiosis, Comunicación entre reinos, secuenciación de RNAs pequeños, Bioinformática, tRF, RNA de interferencia.

Abstract

Predictive and *in-silico* functional analysis of small RNAs in the *Trichoderma-Arabidopsis* interaction.

Plants and microorganisms have co-evolved molecular mechanisms to establish symbiotic relationships. One of these mechanisms is the use of small RNAs (sRNAs) to modulate gene expression to each other, which is called cross-kingdom RNAi (RNA interference). The canonical sRNA biogenesis includes recognition of a precursor RNA and the activity of Ribonucleases type III, called in plants Dicer-Like proteins (DCLs) to produce RNAs < 200 nucleotides (nt) in length. The biogenesis of sRNAs derived from tRNA and rRNA molecules has been recently reported, although little is known about their biogenesis and function in gene regulation. Currently, information related to cross-kingdom RNAi in mutualistic symbiotic interactions is scarce. In this study, 8 sRNA-seq libraries and 18 mRNA-seq libraries from the interaction between *Trichoderma atroviride*, a mutualistic fungus, and the model plant *Arabidopsis thaliana* during 24-, 48-, 72-, and 96 hours of coculture (hcc) were analyzed. Bioinformatic analyses were carried out to dissect the sRNAs that putatively participate in a cross-kingdom interaction (ck-sRNAs), as well as to determine their producing organism, precursor molecules, and their potential roles. Additionally, an sRNA-gene co-expression network was inferred to identify diverse associations (co-expression and regulation) between them.

Arabidopsis and *Trichoderma* produce different amounts of sRNAs through their time of interaction. tRNA-derived RNA fragments (tRFs) are the main sRNA found in both organisms, changing their abundances through time. Targets for these sRNAs were associated with plant defense, metabolic activity, and translational processes in the plant. The sRNA-gene co-expression network has, in their majority, targets involved in plant defense response. Results suggest a possible ck-RNAi communication derived from tRFs at different times of the interaction (24- to 96 hcc), being mainly produced by *Arabidopsis* at 48- and 72 hcc and by *Trichoderma* at 96 hcc. These regulations mainly affected plant defense biological processes; thus plant defense mechanisms are the main regulatory objective during *Trichoderma-Arabidopsis* interaction in 24-, 48-,72-, and 96 hcc.

KEYWORDS. Symbiosis, Cross-Kingdom, sRNA-seq, Bioinformatics, tRF, RNAi.

Introduction

Plants depend on each cell's innate immune response to confront pests and pathogens. The plant immune response is triggered by the detection of microorganisms through the cell membrane pattern recognition receptors (PRRs), which interact with microbial-/ pathogen-associated molecular patterns (MAMPs and PAMPs, respectively), such as bacterial flagellin or fungal chitin, to stimulate a MAMP- / PAMP-triggered immunity (MTI and PTI, respectively). In consequence, microorganisms release effectors molecules (virulence factors or toxins) altering the cell functions, facilitating the infection process by interfering MTI/PTI, generating an effector-triggered susceptibility (ETS). Effectors can trigger the defense responses as well. Effectors can be grouped into two classes: Apoplastic effectors, which are secreted into the plant apoplast to interact with extracellular targets and receptors, or Cytoplasmic effectors, that are translocated inside the plant cell (Selin et al., 2016). However, plants counterattack by detecting such effectors through resistance (R) proteins, such as nucleotide-binding and leucine-rich repeat proteins (NB-LRRs), stimulating a second immune layer, called effector-triggered immunity (ETI), leading to the hypersensitive response (HR), a kind of programmed cell death, to avoid the growth of biotrophic and hemibiotrophic pathogens (Jones & Dangl, 2006).

Plants can develop a systemic protection against future attacks when interacting with microorganisms, a process called induced resistance. Plant-pathogen interaction can result in the triggering of the systemic acquired resistance (SAR), wherein accumulation of pathogenesis-related (PR) proteins and salicylic acid (SA) takes place. Instead, the induced systemic resistance (ISR), is stimulated by mutualistic microbes and is characterized by the accumulation of Jasmonic Acid (JA) and Ethylene (ET) (Pieterse et al., 2014).

Early studies about the interaction between the mutualistic fungus *Glomus* and alfalfa (*Medicago sativa*) revealed that expression of defense- and stress-related genes are enhanced during the early stages of the interaction and then declines as symbiosis developed (Kapulnik et al., 1996). Later, a genomic study and genome-scale expression profiling in *Laccaria bicolor*, other mutualistic fungus, found proteins that shared significant similarity with pathogen effectors used during the infections of pathogenic basidiomycetes (Martin et al., 2008). These data suggested that plants initially perceive mutualistic fungi as potential invaders and perform defense responses, which were then countered by effector-like molecules by mutualistic fungal symbionts (Zamioudis & Pieterse, 2012).

Further studies about the identification of these effector-like molecules and how they could be regulating different aspects of plant defense during early times of the mutualistic interactions have been performed. For instance, the mycorrhiza-induced small-secreted protein 7 (MiSSP7) is released by *L. bicolor* entering the plant cell via endocytosis like pathogenic effectors. Then, MiSSP 7 is transported to the plant nucleus where it positively regulates auxin-responsive gene families like the auxin/indole-3-acetic acid (Aux/IAA); cell wall remodeling genes (e.g., beta-glucosidase, pectinase, and extensin) and reactive oxygen species production-related genes (GRIM REAPER). (Plett et al., 2011).

Non-coding RNAs (ncRNAs) are RNA molecules that are not translated to proteins, and they can be divided, according to their regulatory roles, into two categories: 1) the housekeeping ncRNAs, which are ubiquitously expressed in cells, participating generic cellular functions and, 2) the regulatory ncRNAs, which regulate gene expression at transcriptional, and post-transcriptional levels.

Housekeeping ncRNAs range in size from 50 nucleotides (nt) to 500 nt and includes ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs). They are essential in protein synthesis, RNA splicing, and RNA modifications, for rRNAs and tRNAs, snRNAs, and snoRNAs respectively (Zhang et al., 2019).

On the other hand, regulatory ncRNAs could be further classified as long non-coding RNAs (lncRNAs) with sizes greater than 200 nt and small non-coding RNAs (sRNAs), smaller than 200 nt (Zhang et al. 2019).

LncRNAs also lack the protein-coding ability and concerning their regulatory effects, these RNAs can be classified as cis-lncRNAs (cis-acting lncRNAs) or trans-lncRNAs (trans-acting lncRNAs) that regulate the expression of close and distant genes, respectively (Zhang et al. 2019).

Small RNAs regulate gene expression by sequence complementarity at transcriptional (TGS) or post-transcriptional levels (PTGS) in Eukarya (Zhang, 2009). The main classes of sRNAs are microRNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs). The biogenesis process of an sRNA involves the maturation of a double-stranded RNA (dsRNA) or hairpin structured RNA (hpRNA) precursor through the action of Dicer or Dicer-like (DCL) type III ribonucleases to produce a mature double-stranded RNA (Henderson et al., 2006). Subsequently, one of the two strands of mature RNA (-3p or -5p) is loaded into Argonaute proteins (AGOs) and used to guide the RNA-induced silencing complex (RISC) by sequence complementarity to a specific mRNA target in case of PTGS (Fire et al., 1998). In the case of TGS, the RISC complex is transported to the nucleus where it interacts with the RNA transcribed from the target DNA and recruits other factors like Domains Rearranged Methylase2 (DRM2), resulting in an RNA-directed DNA Methylation (RdDM) (Guo et al., 2016).

It has been described that housekeeping ncRNAs like tRNAs and rRNAs can act as precursor molecules of sRNAs under certain conditions (Gebetsberger et al., 2017; Locati et al., 2018) producing tRNA derived fragments (tRFs) or rRNAs derived RNA fragments (rRFs), which can be 14 ~ 40 nt long in both cases (Lambert et al., 2019; Loher et al., 2017; Xie et al., 2020).

In the case of tRFs, the strict use of dicer enzymes is not a fundamental step, instead, other animal RNase called angiogenin (ANG) can produce the sRNAs under certain conditions of stress (Su et al., 2019).Their products are tRNA-halves (tRHs), 30 ~ 40 nt tRFs molecules, and results from the cleaving within the anticodon loops of mature tRNAs (Thompson & Parker, 2009). In plants, the

participation of the RNAse T2 (RNS) has been identified as a major player in tRFs biogenesis instead of Dicer-likes (Chen et al., 2013; Ivanov et al., 2011).

Some sRNAs can be transferred from different kingdoms. This phenomenon has been called cross-kingdom RNAi (ck-RNAi), and the sRNAs involved, cross-kingdom sRNAs (ck-sRNAs) (Hou et al., 2019).

Examples of ck-RNAi have been mainly observed in parasitic interactions. One common example is the resistance to the malaria disease of humans with sickle cell disease and *Plasmodium falciparum*. It has been observed that erythrocytes from sickle cell individuals have an expression enrichment of miR451 and let-7i miRNAs. These miRNAs translocate into the parasitic cells and bind to *P. falciparum* mRNAs, like cAMP-dependent protein kinase (PKA-R), leading to specific translation inhibitions reducing their parasitemia (Lamonte et al., 2012).

During pathogen-plant interactions, ck-sRNAs can function as effectors molecules through sequence complementarity with its target mRNA attenuating the immune response. An example of this ck-RNAi is the cross-kingdom communication of the pathogenic fungus *Botrytis cinerea* when infecting *A. thaliana* and tomato plants (Weiberg et al., 2013). In that work, *B. cinerea* mature sRNAs (Bc-sRNAs) are exported into the plant cells where they bind to and kidnap the endogenous plant's AGO1 protein, and silence genes, like mitogen-activated protein kinases (MPKs) MPK1 and MPK2 and Mitogen-activated protein kinase kinase kinase 4 (MAPKKK4), that participate positively during plant defense response.

In the plant side, it has been observed that *A. thaliana* cells infected with *B. cinerea* secrete ck-sRNAs to silence vacuolar protein sorting 51 (*Vps51*), dynactin complex (*DCTN1*), and suppressor of actin (*SAC1*) encoding genes, all of them related to vesicle-trafficking, a main factor in virulence during infection of plants by *B. cinerea* (Cai et al., 2018).

Cross-kingdom sRNAs have been suggested to regulate more pathogen-host interactions, such as *Plasmopara viticola*-*Vitis vinifera* (Brilli et al., 2018), Cotton plants infected with *Fusarium oxysporum* or *Verticillium dahliae* (Shapulatov et al., 2016; T. Zhang et al., 2016), *A. thaliana* challenged against *Phytophthora capsici* (Hou et al., 2019) and *A. thaliana* versus *Cuscuta campestris* (Shahid et al., 2018).

Some studies involving the participation of ck-sRNAs in the mutualistic symbiotic association between microorganisms and plants have been reported. For instance, the interaction of the rhizobacterium *Bradyrhizobium japonicum* with soybean (Glycine max), where sRNAs derived from tRNAs (Bj-tRFs) of *B. japonicum* are involved as effector molecules by silencing the genes: root hair directive 3 (GmRHD3a/GmRHD3b), hairy meristem 4(GmHAM4a/GmHAM4b) and leucine-rich repeat extension-like 5(GmLRX5RHD3), that negatively regulate the generation of root hairs, thus promoting the development of nodules in the roots (Ren et al., 2019).

Recently, the participation of small RNAs in fungal-plant mutualistic interaction has been suggested, however information about their mechanistic participation

is scarce (Ramírez-Valdespino et al., 2019). For instance, *Populus* spp. plants present a different sRNA production profile when co-inoculated with ecto- and endo-mycorrhizal fungi like *Rhizophagus irregularis* or *L. bicolor*. Some of the plant sRNAs have predicted putative targets in fungal genes (Mewalal et al., 2019). Thus, these data suggest that potentially a ck-RNAi communication could be carried out during early times of mutualistic *fungi*-plant interactions, however, more investigation is needed.

Trichoderma spp. are filamentous fungi that are common inhabitants of soil where they colonize the plant roots establishing a mutualistic relationship. It has been observed that *Trichoderma* spp. can improve plant growth and their resistance against abiotic stresses (Kleifeld & Chet, 1992; Salas-Marina et al., 2011; Guler et al., 2016; González-Pérez et al., 2018). The usefulness of *Trichoderma* spp. as biocontrol agents has been documented, due to its ability to produce antibiotics (Oh et al., 2002) and feed on other fungi (mycoparasitism) (Rocha-Ramírez et al., 2002).

The genome of *Trichoderma atroviride* bears two dicer (dcr1 and dcr2), three Argonaute (ago1-ago3), and three RNA-dependent RNA polymerases (rdr1-rdr3) homologs genes. Although mutation of these genes altered reproductive development and vegetative growth (Carreras-Villaseñor et al., 2013), there is no direct evidence for modulation of plant immunity by *Trichoderma* sRNAs.

Early stages of plant root colonization at 24 to 96 hours post-inoculation (hpi) by *Trichoderma*, have been associated with chromatin modifications, modulating beneficial effects like plant growth, and defense against pathogens (Estrada-Rivera et al., 2019). In our Lab group, sequencing of sRNAs during *Arabidopsis–T. atroviride* interaction at different times revealed the production of 37 fungal sRNAs that potentially regulate *A. thaliana* host genes (Dautt-Castro, unpublished; Ramírez-Valdespino et al., 2019).

In this present study, we aim to identify the composition and possible participation of ck-sRNAs in a *Trichoderma-Arabidopsis* mutualistic interaction model after 24-, 48-, 72-, and 96 hcc by means of different bioinformatic studies like sRNA annotation, differential expression analysis (DE), target prediction, and GO enrichment analyses, using sRNA-seq and mRNA-seq data. Finally, a co-expression network was inferred to allow us a better understanding of the potential sRNA-target associations.

Results

The composition of small RNAs producing *loci from A. thaliana and T. atroviride changed during different interaction times*

Although previous studies involved ck-sRNAs participation in different plant mutualistic interactions (Mewalal et al., 2019; Ren et al., 2019; Silvestri et al., 2019), little is known about *Trichoderma*'s sRNAs and their role to establish a mutualistic interaction with plants. Thus, it is of our interest to know a possible bidirectional regulation by ck-sRNAs.

Eight sRNA-seq libraries from *A. thaliana*-*T. atroviride* at 24-, 48-, 72- and 96 hours of coculture (hcc), which are part of our Lab collection were used through this work.

To determine the *A. thaliana* and *T. atroviride* production of 18 to 40 nt sRNAs during their interaction, an *A. thaliana*-*T. atroviride* concatenated reference genome was generated (see methods section for more details). We calculated the percentage of reads that mapped to the reference genome for each of the 8 analyzed sRNA-seq libraries (24-, 48-, 72-, and 96 hcc, Table 1).

Table 1. Total number of trimmed sRNA-Seq reads mapped to *A. thaliana* and *T. atroviride* genomes.

Sample	Total reads	Mapped reads	*A. thaliana* mapped reads	*T. atroviride* mapped reads
24h_R1	12,620,955	12,102,228 (96%)	11,184,944 (89%)	634,970 (5%)
24h_R2	14,062,629	12,974,002 (92%)	6,620,906 (47%)	6,122,744 (43%)
48h_R1	59,106,424	55,454,102 (93%)	15,271,481 (26%)	38,951,180 (66%)
48h_R2	19,368,598	18,971,041 (98%)	17,025,660 (88%)	1,613,636 (8%)
72h_R1	10,740,056	10,508,052 (98%)	9,938,672 (92%)	293,400 (3%)
72h_R2	14,483,650	14,081,090 (97%)	12,836,612 (89%)	1,024,901 (7%)
96h_R1	15,096,804	14,049,699 (93%)	5,858,711 (39%)	7,968,291 (53%)
96h_R2	10,338,594	9,991,199 (97%)	2,070,074 (20%)	7,709,867 (75%)

% Corresponds to the percentage of mapped reads recovered from their total sample reads.

From every interaction time analyzed, about 90% of the reads mapped to the concatenated reference genome (Fig. 1A). The mapped reads were mainly composed of *A. thaliana* multimapping reads (At_Multi_map, reads that mapped

multiple times only over *A. thaliana* genome) at 24- and 72 hcc, with 64% and 84% of total mapped reads, respectively. On the other hand, there were more reads of *Trichoderma* at 96 hcc (65% of total mapped reads), composed in a 50% of the total mapped reads, by unique mapping reads (Ta_Unique_map, reads that mapped a single time only over *T. atroviride* genome) (Fig. 1B).

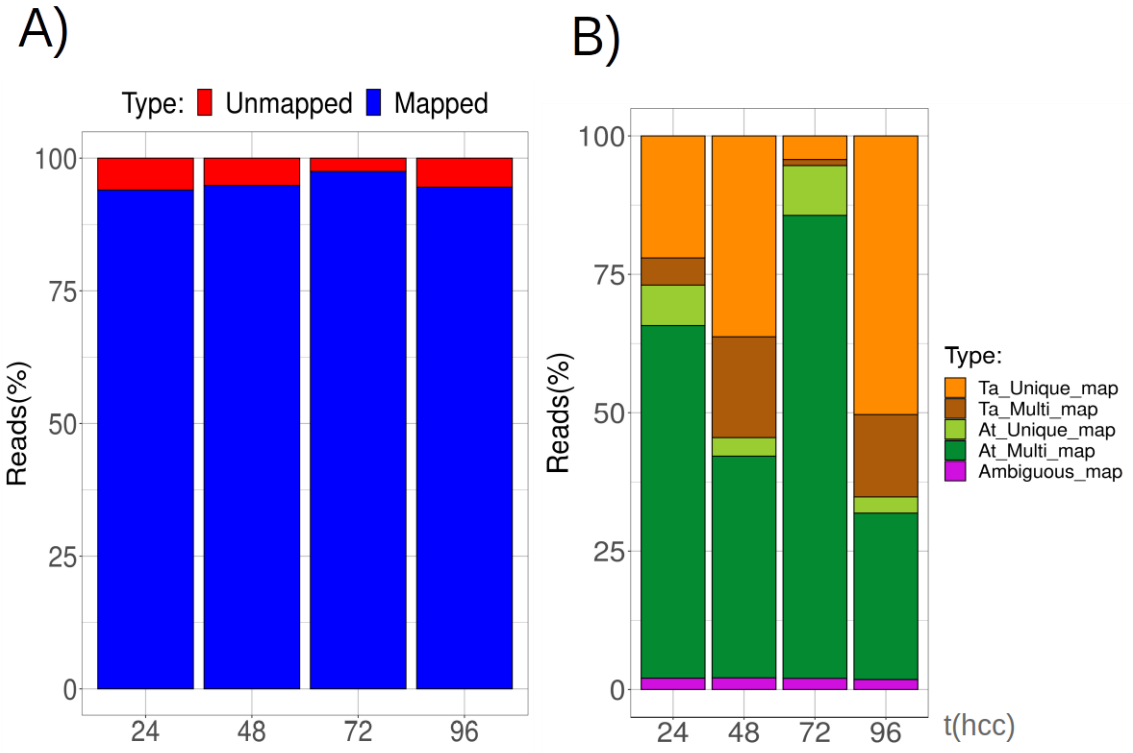


Fig. 1. *Arabidopsis thaliana* and *Trichoderma atroviride* varied in their sRNA compositions through the time of interaction. A) The mean mapped (blue)/unmapped (red) genome reads percentage. (B) The mean mapping types in percentage. Ta_Unique_map (orange) = *T. atroviride* unique mapping reads; Ta_Multi_map (brown) = *T. atroviride*-specific multimapping reads; At_Unique_map (green) = *A. thaliana* unique mapping reads; At_Multi_map (dark green) = *A. thaliana*-specific multimapping reads; Ambiguous_map (purple) = ambiguous reads.

These results suggest that the origin of the analyzed sRNAs is not equally distributed between the two organisms through time, being *Arabidopsis* the main producer at 24 and 72 hcc, and *Trichoderma* at 96 hcc.

Arabidopsis thaliana bears most of the variety of sRNA producing loci

tRFs have been found during the interaction of *B. japonicum*, a mutualistic symbiotic bacterium with *G. max*, whose main effect is in promoting the development of nodules in soybean roots (Ren et al., 2019). So, we wanted to know if there was a contribution of sRNAs derived from tRNAs (tRFs) and rRNAs (rRFs) molecules in our *Trichoderma-Arabidopsis* interaction model.

To associate the mapped reads to the different classes of sRNAs *loci* (miR, rRNA, tRNA, DCR, Other), we performed annotation of sRNAs producing *loci* throughout the concatenated reference genome using ncRNAs annotation tools and databases (see methods).

The 28, 764 total *loci* were distributed in similar proportions between *T. atroviride* (52%) and *A. thaliana* (48%) (Fig. 2A). Considering total *loci* in both organisms, 94% corresponded to possible Dicer cleavage (DCR *loci*). This percentage represented a 92% composition of the total *A. thaliana* annotated *loci* and a 96% for *T. atroviride*.

Considering remaining *loci* types (6%) a non-equally distribution of producing *loci* were found between both organisms, being primarily originated from *A. thaliana* genome (4%) and in less proportion from *T. atroviride* (2%) (Fig. 2 B). From these 6% annotated *loci* we found miRNAs (0.4%), tRNAs (potential tRFs, 2.7%), rRNAs (potential rRFs, 0.9%), and other kinds of ncRNAs *loci* (2%).

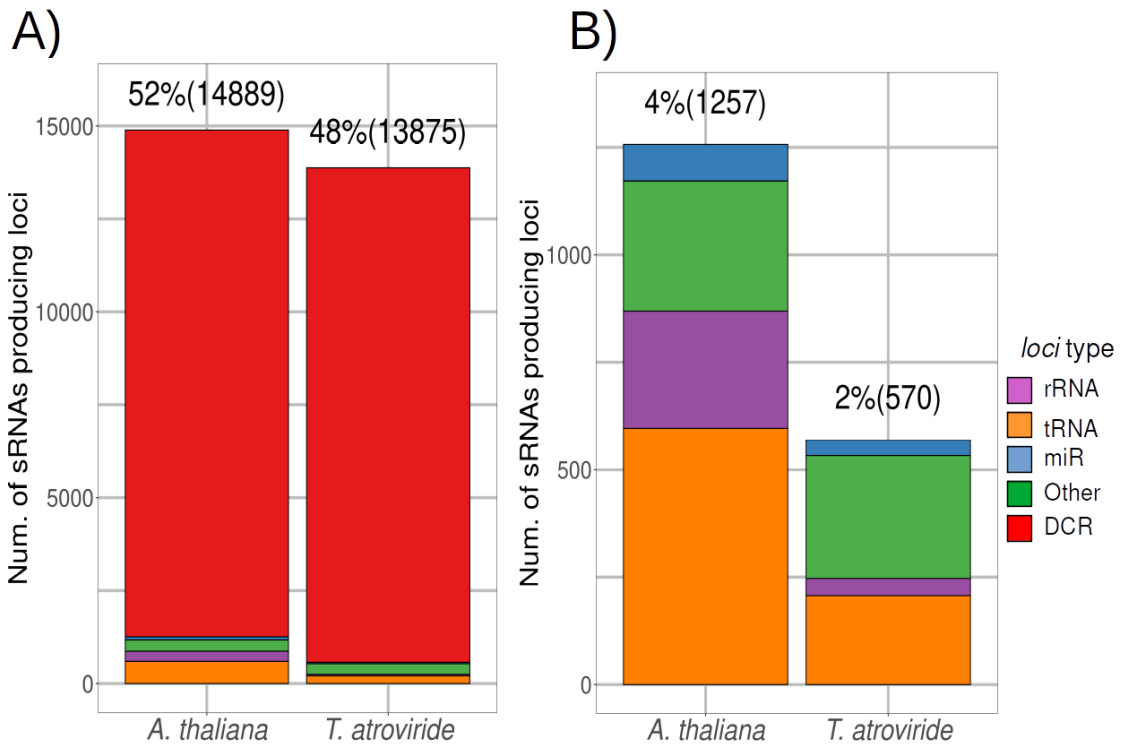


Fig. 2. DCR sRNAs were the main produced *loci* in *A. thaliana* and *T. atroviride*. (A) Percentage of *loci* type composition in the two organisms for all sRNAs annotated *loci*. (B) Percentage of *loci* type composition and their producing organism, for non-DCR type *loci*.

Excluding the participation of DCR-type producing *loci*, *Arabidopsis* contributed with most of the variety of the annotated *loci* during its interaction with *Trichoderma*.

The tRNAs were the main producing loci that statistically change in A. thaliana.

As previous work on these eight sRNA-seq libraries, showed *Trichoderma* sRNAs being accumulated at interaction with *Arabidopsis* (Ramírez-Valdespino et al., 2019), we wanted to know if there was a temporal association with changes in sRNAs abundances during early stages of interaction.

To determine which annotated sRNAs producing *loci* could be modulating their sRNA abundances through interaction times, a differential expression analysis of such sRNAs was performed.

Counts associated with the 28,764 producing *loci* were transformed to counts per million (CPM) values. Those *loci* with CPM < 4 and found in at least two libraries were filtered out. The remaining 2,954 (10.3%) *loci* were used for differential expression analysis through an ANOVA test.

From these 2,954 *loci*, 230 (7.8%) sRNAs producing *loci* were determined with dissimilar abundances (Fig. 3A). From these *loci*, we observed that 141 were potentially generated in the *T. atroviride* genome. The higher quantity of *loci* type were DCR (107, 46.4%), followed by tRNA (23, 10%), rRNA (9, 3.9%), and other kinds of ncRNAs *loci* (2, 0.9%) (Fig. 3B). The remaining 89 *loci* corresponded to *Arabidopsis* and were mainly composed by tRNA (57, 24.8%) and DCR *loci* (17, 7.4%). In a minor quantity rRNA (11, 4.8%), miRNA (2, 0.9%) and other kinds of ncRNAs (2, 0.9%) conformed the total of *Arabidopsis loci* (Fig. 3B).

These 230 *loci* augmented their tRNAs composition in comparison to the total annoated *loci* type composition, mostly in *Arabidopsis*.

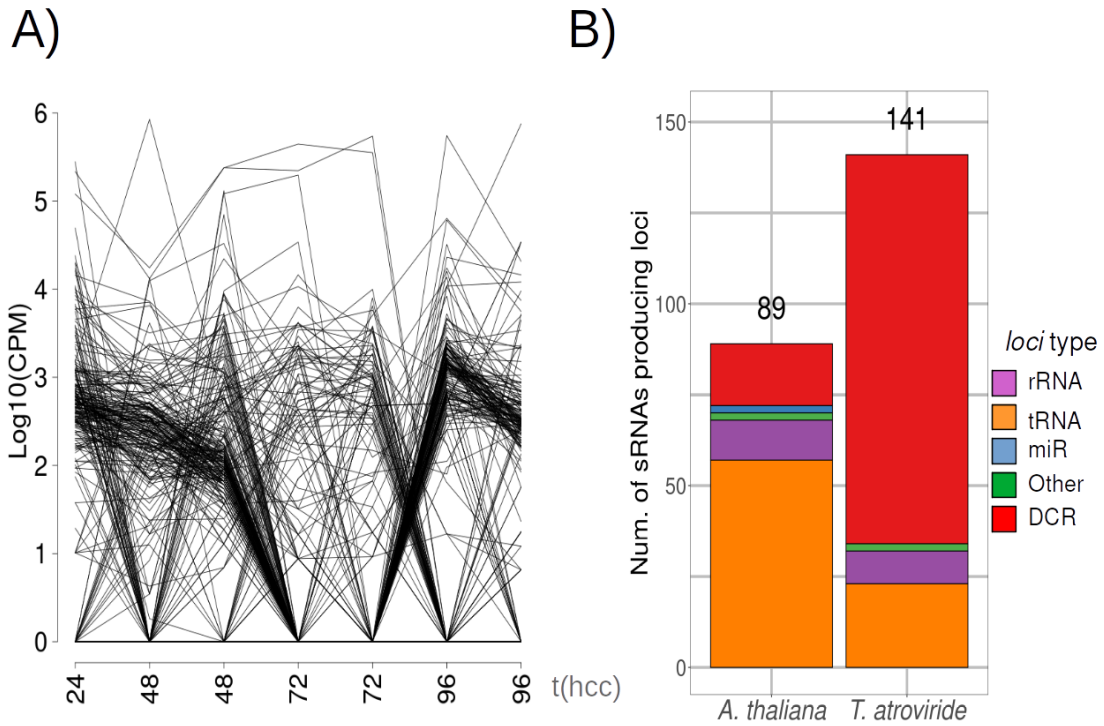


Fig. 3. *Trichoderma* DCR was the main producing *loci* with statistical differences through time. (A) Log_{10} (CPM) abundances of annotated *loci* with differences in their abundances through time (FDR ≤ 0.1). Each line corresponds to one of the 230 *loci* with dissimilar abundances. (B) Percentage and type composition of *loci* belonging to *A. thaliana* and *T. atroviride*, total organism-associated *loci* are depicted above each bar.

In summary, 230 *loci* showed statistical dissimilar abundances through time, being mainly originated from *Trichoderma*. DCR was the principal sRNA-

producing *loci*-type that changed their abundance in *Trichoderma*. On the other hand, *Arabidopsis* sRNAs principally were composed of tRNA-producing *loci*.

A. thaliana and T. atroviride sRNAs were mainly abundant at 48-72 and 96 hcc, respectively.

To identify producing *loci* with similar abundance profile patterns through time, the 230 *loci* with statistical changes were clustered into 4 groups using the k-means algorithm (k = 4) (see methods section for details) (Hartigan & Wong, 1979) (Supplementary Fig. 1).

Accordingly, to their overall profile patterns through time, these 4 clusters can be grouped into two more general groups. The first group correspond to clusters 1 and 2 that progressively dropped their abundances at 96 hcc. In the contrary, the second group conformed by clusters 3 and 4, showed an increase in their abundances at 96 hcc (Fig. 4A).
These first two groups (clusters 1 and 2) originated mainly from the *A. thaliana* genome (Fig. 4B) and were predominantly composed of tRNA *loci* (tRFs). In contrast, clusters 3 and 4 originated mainly from the *T. atroviride* genome (Fig. 4B) and are mainly composed of *loci* annotated as DCRs or rRNAs (rRFs), respectively.

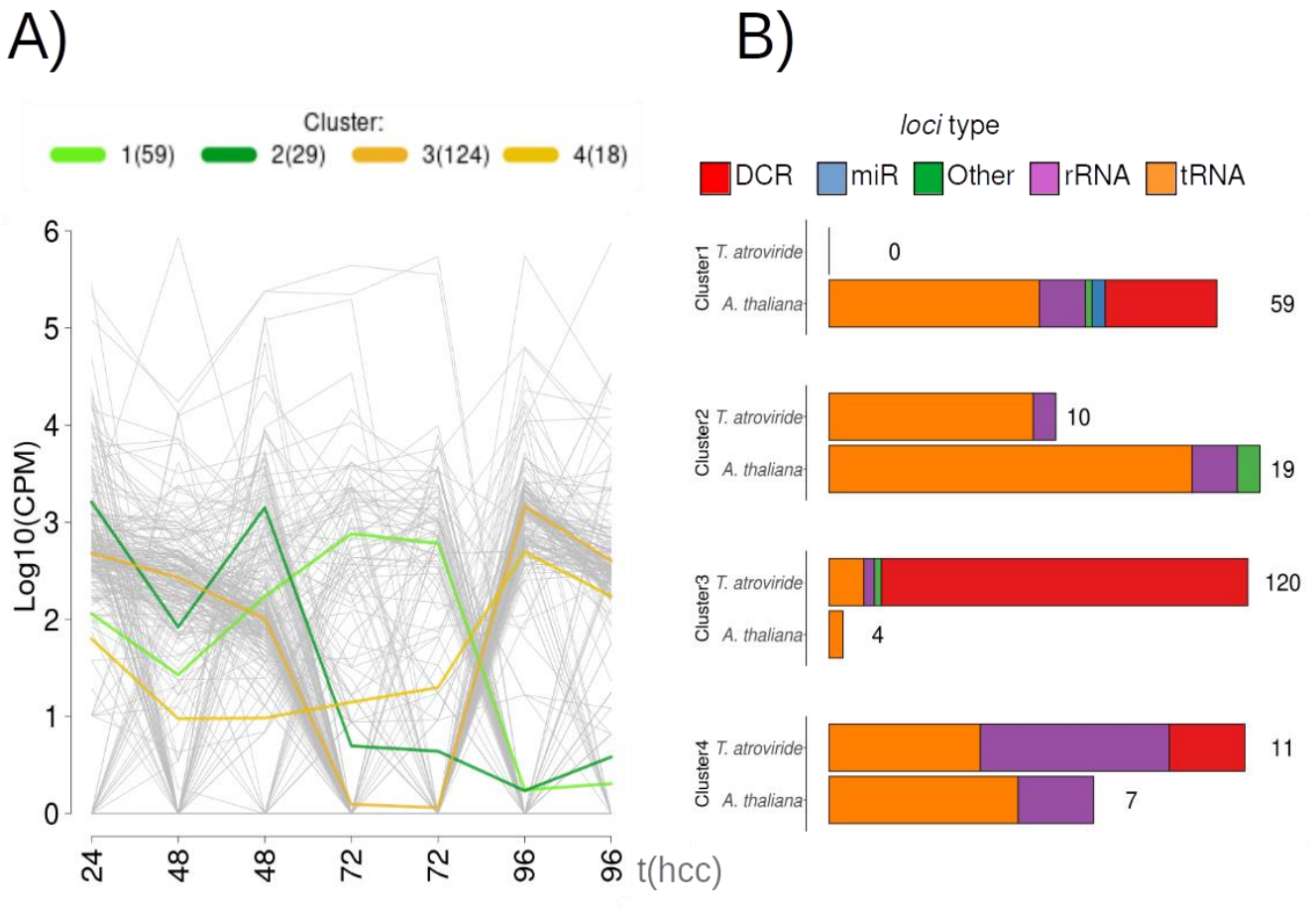


Fig. 4. *Arabidopsis* sRNAs were mainly abundant at 48-72 hcc and at 96 hcc by *Trichoderma* sRNAs. (A) Log_{10} (CPM) abundances through time of the cluster *loci* and its representative cluster pattern, each cluster representative is colored in a different grade of cyan (total cluster corresponding *loci* are noted between parenthesis on legend). (B) The amount and type composition of cluster *loci* that mapped to *A. thaliana* or *T. atroviride* genomes, total organism-associated *loci* are depicted above each bar.

In summary, *Arabidopsis*-derived sRNAs contributes to the major sRNA abundances during the intermediate times of the interactions analyzed (48- and 72 hcc, cluster 1 and 2). Cluster 1 and 2 were mainly composed of *A. thaliana* sRNAs (100% and 65%, respectively). In these clusters the most represented sRNAs were tRFs (54% and 86%, respectively).

On the other hand, we observed that sRNAs were mainly derived from *Trichoderma* producing *loci* at 96 hcc (cluster 3 and cluster 4, 97% and 61% respectively). From these clusters, 85% of sRNAs derived from DCR *loci* for cluster 3 and 50% to tRNA for cluster 4.

Representative sRNAs derived from tRNAs loci

To in deep analyze the characteristics of the 230 sRNAs during early root colonization by *Trichoderma*, we selected for every cluster the top 4 representative sRNAs based on their $-\log_{10}$(FDR) values.

A total of 28 representative sRNAs were obtained, being the lowest $-\log_{10}$(FDR) a value of 1.4 (FDR = 0.04) (Fig. 5). From these 28 sRNAs, the major part corresponded to tRFs (22, 78.57%) and rRFs (5, 17.86%). These sRNAs have $-\log_{10}$(FDR) values ranging from 1.4 to 5.2 (FDR = 0.04 and 6.31×10^{-6} respectively). In contrast, only one DCR (3.57%) was observed.

Due that most of the DCRs, with significant changes in their abundance, presented lower $-\log_{10}$(FDR) values (higher FDRs) (Fig. 5) and being the highest one a $-\log_{10}$(FDR) of 2.6 (FDR=2.51×10-3), DCR derived sRNAs were not selected. Additionally, we considered an sRNA annotated as a miRNA (ncRNA_5780) that has changes in their abundance through time.

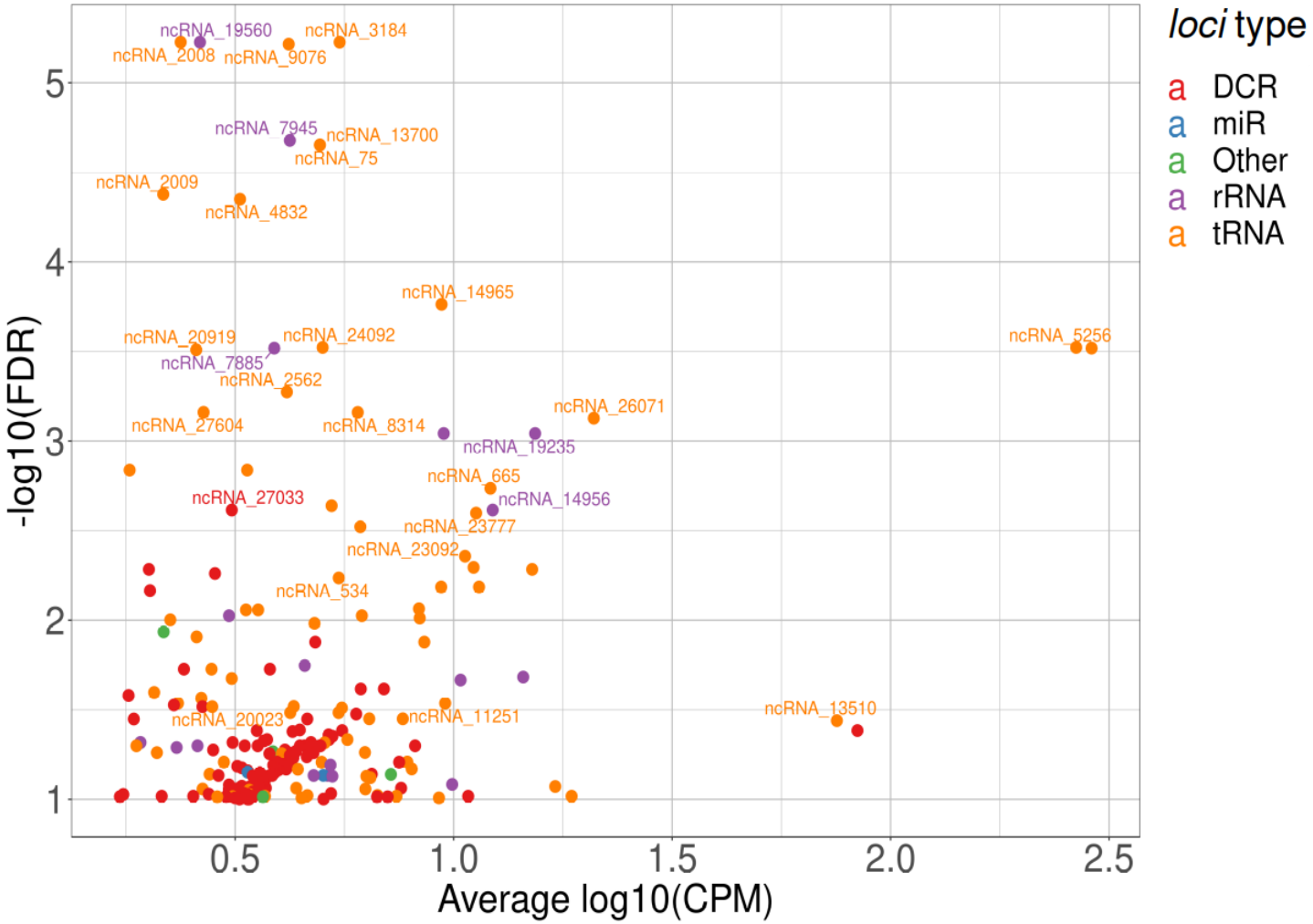


Fig. 5. Representative sRNAs clusters were mainly composed of tRFs in both organisms. Average abundance/statistical significance distribution of the 230 cluster sRNAs *loci*. Labeled points refer to the top selected cluster representative sRNAs *loci*.

To conclude, tRFs are the main *loci* type of sRNAs associated with changes in abundance through early root colonization by *Trichoderma*, representing them as potential players in a ck-RNAi communication.

tRFs would be implicated in regulating plant defense-related genes during interaction with Trichoderma

Prediction of putative targets and Gene Ontology (GO) enrichment analysis were performed with the objective to associate the 28 sRNAs with possible biological functions.

Target predictions of sRNAs were performed using psRNATarget (Dai et al., 2018). Due that a previous work reported that tRHs-mRNA binding could be performed by different parts of the sRNA, like the center or the 3' (Jehn et al., 2020) and 79% of our sRNAs have sizes ranging from 30 to 41nt in length, we opted to split out these sRNAs into k-mers of 22 nt, covering the 5', center, and 3' regions of the original sRNA sequence predicted, this to infer putative interactions for each region. The scores for each region were added up by summing the expectation, type of inhibition, and site accessibility values, obtaining a representative prediction score for each target. Then, target predictions (representative score) for every sRNA were used to perform GO enrichment analysis (see Supplementary Fig. 3 and methods).

Additionally, we used floral development *Arabidopsis* microRNAs: ath-miR172a (MIMAT0000203) and ath-miR156a-5p (MIMAT0000166), as controls for sRNA prediction and GO enrichment results. Expecting that such microRNAs will target genes related to floral development.

Control miRNA ath-miR156a-5p showed top target enriched GO biological processes (BP) related to anther development (GO:0048653, $p = 0.03$) and regulation of timing of transition from vegetative to reproductive stages (GO:0048510, $p = 0.06$). In the case of ath-miR172a, the top enriched processes were meristem maintenance (GO:0010073, $p = 0.14$) and specification of floral organ identity (GO:0010093, $p = 0.14$). In both cases predicted targets were mainly related to floral development biological processes.

From the 28 representative sRNAs, 24 have predicted targets, and 4 do not have any predicted target. The top 4 enriched target biological processes per representative sRNA were obtained. A total of 77 different terms were obtained, all of them derived from *A. thaliana* targets.

tRFs showed most frequently BP enrichments related to the structure of cell cytoskeleton (GO:0030036 and GO:0051016), plant immune response (GO:0042742 and GO:0002238), and metabolism (GO:0019516 and GO:0009854). Rather, rRFs have sRNA frequencies associated with DNA maintenance (GO:0006302) and segregation (GO:0007062) (Table 1).

Table 2. Top ten most frequently Biological Process terms (Gene Ontology) found in the enrichment analysis (out of 77).

Term	*loci*	Organism	sRNA frequency	Enrichment p-values	GO ID
actin cytoskeleton organization	tRNA	*A.thaliana*	5	0.44 - 1	GO:0030036
barbed-end actin filament capping	tRNA	*A.thaliana*	4	0.44 - 1	GO:0051016
defense response to bacterium, incompatible interaction	tRNA	*A.thaliana*	3	0.33	GO:0042742
dephosphorylation	tRNA	*A.thaliana*	3	0.42	GO:0016311
glucosinolate biosynthetic process	tRNA	*A.thaliana*	3	0.28	GO:0019761
response to molecule of fungal origin	tRNA	*A.thaliana*	3	0.42	GO:0002238
double-strand break repair	rRNA	*A.thaliana*	2	0.32	GO:0006302
sister chromatid cohesion	rRNA	*A.thaliana*	2	0.32	GO:0007062
lactate oxidation	tRNA	*A.thaliana*	2	0.65	GO:0019516
oxidative photosynthetic carbon pathway	tRNA	*A.thaliana*	2	0.65	GO:0009854

From these results, we can suggest that in general, tRFs could regulate biological processes related to plant defense, cell structure, and metabolism. On the other hand, rRFs could be orchestrating genomic modulation, by affecting genomic cell maintenance and replication.

Some predicted tRFs have cross-kingdom targeting in both organisms

Gene expression analysis over 18 mRNAseq libraries of *Trichoderma-Arabidopsis* interactions at 48-, 72- and 96 hcc was performed with the aim to select the potential target candidates that could be regulated by these 24 sRNAs. *Arabidopsis* seedlings growing alone at the same times were included as controls (9 mRNA-seq libraries).

The top 2 targets with the highest prediction score for each sRNA were selected. Then, we calculated the correlation (Spearman correlation method) between the abundances of the selected sRNA and the expressions of their top targets through interaction times, 48-, 72- and 96 hcc (see Supplementary Fig. 2B and methods).

In the case of the flowering microRNAs, AT1G27360 and TRIATDRAFT_147585 were predicted as putative targets for ath-miR156a-5p, being related to flowering (GO:0048510) and a putative metal ion transporter respectively; AT5G60120 and AT4G36920 were putative targets for ath-miR172a, being both implicated in flowering processes (GO:0048409). Although changes in expression were observed in the best scored putative targets, abundance of such miRNAs was not detected (Fig. 6), which may indicate that they are not induced during the interaction.

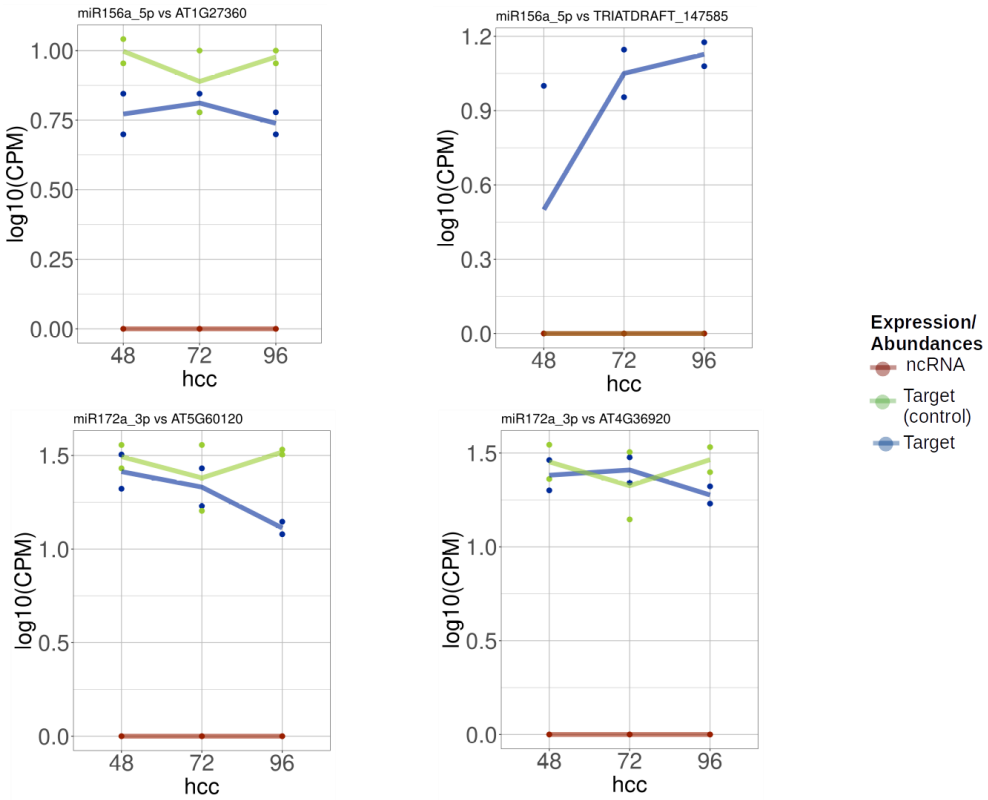


Fig. 6. Control flowering miRNAs, miR156a_5p and miR172a_3p, were not induced during *Trichoderma-Arabidopsis* interaction. Changes in $\log_{10}$ (CPM) abundances through 48, 72, and 96 hcc of control miRNAs (maroon) and their targets expression during *Trichoderma-Arabidopsis* interaction (blue) and in not inoculated *Arabidopsis* (green).

From the target prediction analysis of the 24 sRNAs, 8 predicted targets resulted with a representative psRNATarget score > 2 and were negatively correlated with their corresponding sRNAs, thus being considered the targets with the best potential (Fig. 7).

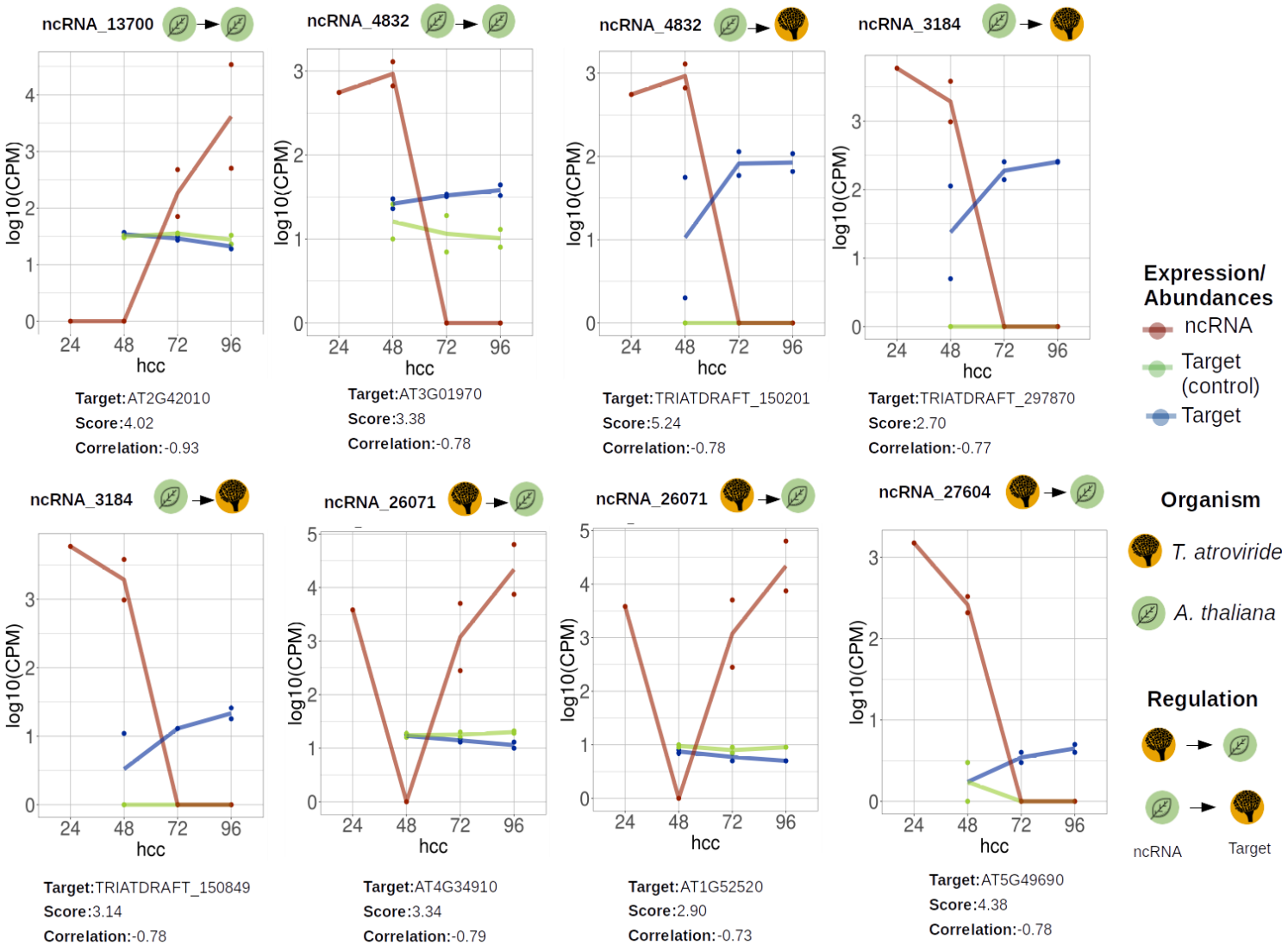

ncRNA_13700
log10(CPM)
hcc
Target:AT2G42010
Score:4.02
Correlation:-0.93
ncRNA_4832
Target:AT3G01970
Score:3.38
Correlation:-0.78
ncRNA_4832
Target:TRIATDRAFT_150201
Score:5.24
Correlation:-0.78
ncRNA_3184
Target:TRIATDRAFT_297870
Score:2.70
Correlation:-0.77
Expression/ Abundances
ncRNA
Target (control)
Target
ncRNA_3184
Target:TRIATDRAFT_150849
Score:3.14
Correlation:-0.78
ncRNA_26071
Target:AT4G34910
Score:3.34
Correlation:-0.79
ncRNA_26071
Target:AT1G52520
Score:2.90
Correlation:-0.73
ncRNA_27604
Target:AT5G49690
Score:4.38
Correlation:-0.78
Organism
T. atroviride
A. thaliana
Regulation
ncRNA
Target

Fig. 7. Top eight target candidates would be regulated by tRFs, which suggest endogenous and cross-kingdom regulations. Changes in $\log_{10}$ (CPM) abundances through 48-, 72-, and 96 hcc of sRNAs (maroon) and their targets (green in control and blue in interaction condition). Above each graph, the sRNA IDs, *loci* types, and regulations are depicted as shown in the legend. Target IDs, prediction score, and sRNA-Target correlation are specified below each graph. The fungus (yellow circles) or plant (green circles) cartoons into the yellow or green circles followed by an arrow indicates a potential endogenous or cross kingdom-regulation.

All these 8 target candidates were predicted as tRFs targets. The main sRNA regulation observed was a cross-kingdom regulation (6 of 8 candidates) being equally distributed by *Arabidopsis*-*Trichoderma* and *Trichoderma-Arabidopsis* interactions. The remaining 2 targets were predicted as endogenous regulation in *Arabidopsis*.

Only 3 of the 8 candidates have a BP GO annotation. From the resting 5, BP was deduced by analyzing GO annotations from their interacting proteins reported in the STRING database (Table 2).

Table 3. Top eight sRNA candidates, showing their corresponding gene targets and functional information.

sRNA	Target	score	Correlation, Spearman	Correlation p-value	GO ID	Function
ncRNA_4832	TRIATDRAFT_150201	5.23	-0.77	0.068		DNA binding protein
ncRNA_27604	AT5G49690	4.37	-0.78	0.065		Metabolism
ncRNA_13700	AT2G42010	4.02	-0.93	0.007	GO:0009816	Defense response to bacterium, incompatible interaction
ncRNA_4832	AT3G01970	3.37	-0.77	0.068	GO:0006817	Phosphate ion transport
ncRNA_26071	AT4G34910	3.34	-0.79	0.059		Ribosome biogenesis
ncRNA_3184	TRIATDRAFT_150849	3.14	-0.78	0.062		Ribosomal protein
ncRNA_26071	AT1G52520	2.90	-0.73	0.095		Light control of development
ncRNA_3184	TRIATDRAFT_297870	2.70	-0.77	0.068		GTP Binding protein

T. atroviride TRIATDRAFT_150201, TRIATDRAFT_150849 and TRIATDRAFT_297870 were predicted as putative targets of the *Arabidopsis* sRNAs ncRNA_4832 and ncRNA_3184, respectively. These *Trichoderma* targets have similar expression patterns between them (having an increment in their expression from 72- and 96 hcc) and similar negative correlations with their corresponding sRNAs. None of these targets have their own biological annotations, neither inferred by their interacting proteins. Functional domains and descriptions associate TRIATDRAFT_150201 to DNA binding, whereas TRIATDRAFT_150849 to ribosomal protein and TRIATDRAFT_297870 to GTP binding protein.

A. thaliana candidate targets that were related to cross-kingdom sRNA targeting, have biological functions related to metabolism (AT5G49690), translation (AT4G34910), and flowering (AT1G52520). Having the last two, similar expression behaviors.

A. thaliana candidate targets with endogenous sRNA targeting, were implicated in plant immune response (AT2G42010) and phosphate transportation (AT3G01970). These targets have opposite expression behavior through the interaction time with *Trichoderma*.

The biological functions for these target candidates (Table 3) and their changes in expression through interaction times (Fig .7) suggest that *T. atroviride* cross-kingdom tRFs (Ta-ck-tRFs) would be regulating metabolism, translation, and development processes in *Arabidopsis*. On the other hand, *A. thaliana* cross-kingdom tRFs (At-ck-tRFs) would be regulating targets with putative functions in transcriptional and translational mechanisms in *Trichoderma*, and endogenously modulating plant defense mechanisms and metabolism, respectively.

Arabidopsis-Trichoderma sRNA-gene co-expression reveals complex RNA modulation communities mainly affecting plant defense mechanisms

An sRNA-gene co-expression network was inferred to assess possible sRNA-target associations.

Cluster representatives of differentially expressed (DE) genes with changes through interaction times were obtained by differential expression analysis and subsequent k-means clustering.

From a total of 23,114 DE genes, 221 have differential expressions with FDR values ≤ 0.1. From these 221 DE genes, 160 belong to *A. thaliana* genes, whereas the remaining 61 pertain to *T. atroviride*. From these 221 DE genes, 78 (35 %) were downregulated and 143 (65%) upregulated at 72 hcc, whereas 80 (36 %) were downregulated and 141 (64 %) upregulated at 96 hcc.

These 221 DE genes, as well as the top 8 target candidates, were clustered into 8 and 6 clusters, for *A. thaliana* and *T. atroviride* respectively (Supplementary Fig. 3 and Supplementary Fig. 4).

For each cluster, the top of 3 representative genes were selected based on their $-\log_{10}$(FDR) values. A final set of 50 genes (42 cluster representatives and 8 target candidates) were obtained. This final set was used to infer the sRNA-gene co-expression network. The network was based on the correlation, obtained from the Spearman correlation method, of the 50 cluster representative gene expression and the 24 representative sRNAs abundances. In addition, mRNAs target scores were used to indicate a possible regulation by each of the 24 sRNAs. The resulting network consisted of 2,701 correlations (Supplementary Fig. 5).

To consider sRNA-gene associations with relevant correlations among the network, correlation with p-values > 0.1 were filtered out. Two resulting networks of 904 and 1 correlations were generated (Fig. 8A). These networks kept all the 74 representative RNAs. In this network, there was a higher percentage of positive correlations (483, 53%) than negative ones (423, 47%) (Pearson's chi-squared test, $p = 0.003$).

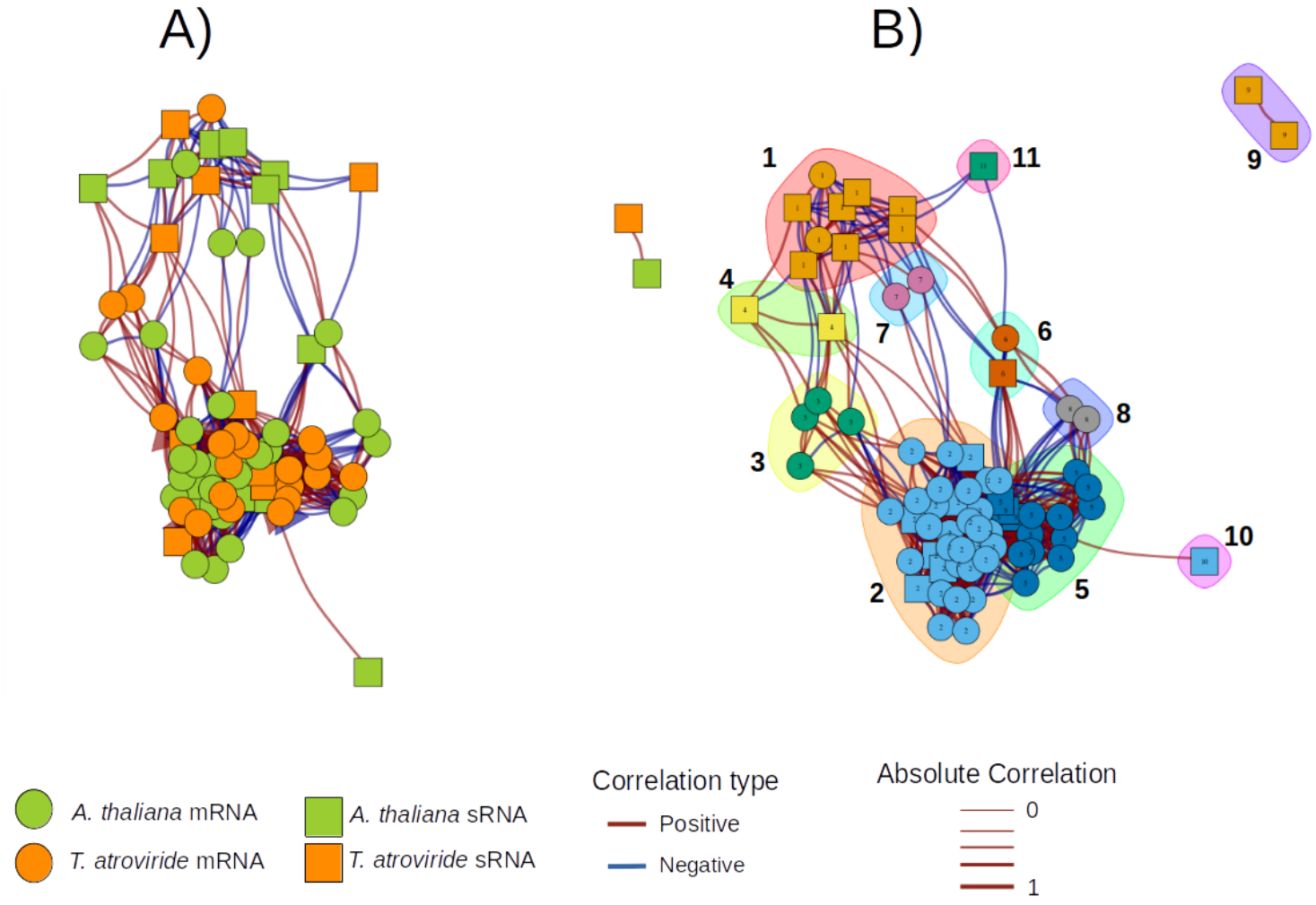


Fig. 8. *Trichoderma-Arabidopsis* interaction sRNAs would regulate different biological functions, mainly related to defense response in plants. The sRNA-target co-expression networks (A) and their correlation-based communities (B).

To find densely connected associations (subnetworks also called communities) inside the networks we use the random walk clustering algorithm (Pons & Latapy, 2005). This methodology finds 11 communities with different feature associations (Fig. 8B). Communities 1, 2, 5, and 6 were considered for subnetwork properties analysis, due to their density of associations and their predicted targeting.

Community number 1 was composed of 20 associations. Inside community number 1 an *Arabidopsis* gene (AT2G14610) with function in defense response in the plant (GO:0009814) was negatively correlated with the *Trichoderma*, ncRNA_19235, and ncRNA_14956, both being rRFs. In the same community, *Trichoderma*'s gene, TRIATDRAFT_80087 (EHK44781) a hypothetical protein, was negatively correlated with 5 *Arabidopsis* sRNAs (ncRNA_5780, ncRNA_5256, ncRNA_2009, ncRNA_2008, ncRNA_7945), of which 1 was predicted as miRNA, 1 as rRF, and 3 as tRFs. The product for this gene is an hypothetical protein.

Community number 2 has a total of 435 associations. *Arabidopsis* targets from this community were involved in plant defense, response to stress, metabolism or were hypothetical proteins. In contrast, *Trichoderma* genes were mainly related to metabolism. In this community, two previously selected target candidates, AT2G42010 and AT4G34910, with previously annotated sRNAs (ncRNA_13700 and ncRNA_26071 respectively) were found with negative correlations with other 3 sRNAs (ncRNA_7885, ncRNA_19560, ncRNA_75).

Interestingly *Trichoderma*'s target TRIATDRAFT_247300 (EHK43110), a chitinase, was positively correlated with the abundance of the *Trichoderma* tRF ncRNA_26071 and presented a targeting prediction score of 2.24.

Community 5 was composed of 102 associations. *Trichoderma* targets were dominant in this community, composing 8 of the 11 genes, we found association to metabolism functions. Although, putative and hypothetical proteins were observed as well. The resting 3 *Arabidopsis* targets have functions in plant metabolism (GO:0033559), DNA binding (GO:0006355) or was an hypothetical protein. Inside this community, *Trichoderma* candidate targets TRIATDRAFT_150201 (EHK43765) and TRIATDRAFT_297870 (EHK49216) were found. Additionally, AT1G06080 the gene associated with metabolism, has a positive correlation and a targeting prediction score of 1.95 with a *Trichoderma* tRF (ncRNA_27604).

Community 6 was composed of an individual sRNA-target association between the *Arabidopsis*-derived tRF ncRNA_534 and the *Arabidopsis* gene AT2G26020. GO annotation for this gene was related to the plant defense response to other organisms (GO:0050832).

These results suggest complex cross-kingdom sRNA-gene co-expression associations, particularly regulating metabolic and defense mechanisms in the plant. On the other hand, 2 predicted targets for *Trichoderma* tRFs were positively correlated in their expression, one for an *Arabidopsis* target (AT1G06080) and the other for *Trichoderma* (TRIATDRAFT_247300). These results would suggest

possible positive sRNA-gene regulation occurring in a cross-kingdom way as endogenously by some *Trichoderma*'s tRFs.

Discussion

Plant ck-RNAi has been mainly studied on host-pathogen interactions, like *A. thaliana* infection by the pathogen *B. cinerea*, where the fungus- and plant-derived sRNAs exert an specific gene expression regulation each other (Cai et al., 2018; Weiberg et al., 2013) with the objective of infecting or avert the infection. On the other hand, little is known about ck-RNAi communication in a mutualistic interaction, where the main objective in this symbiosis is to achieve a mutually beneficial relationship. Here, we applied high-throughput RNA sequencing to profile a ck-RNAi communication between *A. thaliana* roots and the mutualistic fungus *T. atroviride* after 24-, 48-, 72-, and 96 hcc.

By means of read mapping and sRNA annotation on an *Arabidopsis-Trichoderma* concatenated genome, we identified different compositions of fungus- and plant-derived sRNAs during the time points analyzed, being times of 24- and 72 hcc mainly composed of *A. thaliana* sRNAs and in 96 hcc for *T. atroviride* sRNAs. The annotated sRNAs *loci* were mainly annotated as tRNAs, where DCRs derived sRNAs were taken out.

It is known that *A. thaliana* as well as fungi can produce tRNA fragments (tRFs) in a regulated manner (Jöchl et al., 2008; Lalande et al., 2020). Moreover, tRFs have been implicated as ck-sRNAs during *Homo sapiens* infection by *Trypanosoma cruzi* (Garcia-Silva et al., 2014), and *B. japonicum-G. max* mutualistic interaction (Ren et al., 2019), thus promoting susceptibility or altering developmental functions, respectively.

Our analysis of differential expression and clustering of the analyzed sRNAs revealed that sRNAs that changed their abundances through time can be clustered in 4 different profile patterns, where 48- or 72 hcc belong to *A. thaliana* sRNAs, whereas 96 hcc pertained to *T. atroviride*.

The main type of sRNAs associated with dissimilar abundances through time of interaction were tRFs with an 82% of the total of tRFs of 30 to 37 nt in size (thus called tRHs). Although little is known about the molecular biogenesis mechanisms of a tRH molecule or their functionality, they have been reported being produced under oxidative stress conditions in yeast (*Saccharomyces cerevisiae*), mammalian cells (HeLa cells), and plants tissues (*A. thaliana*), suggesting that tRH production is a conserved aspect of the response to oxidative stress in eukaryotes (Thompson et al., 2008; Xie et al., 2020). Moreover, tRHs have been found as a principal source of sRNAs in *T. cruzi* ck-RNAi interaction with human cells (Garcia-Silva et al., 2014) and even in transgenerational epigenetic inheritance in mice (Chen et al., 2016). Whereby, the idea that ck-RNAi communication between *T. atroviride* and *A. thaliana* would, mainly, be performed by tRFs, specifically tRH molecules, is plausible.

Our GO enrichment results show that tRFs could be regulating targets associated with plant defense responses as well as cell structure and metabolic changes in the plant cell. Moreover, *Arabidopsis* target candidates with the best target prediction scores were related to defense and metabolic functions (AT2G42010 and AT5G49690 respectively).

AT2G42010 codes for a phospholipase Dβ1 (PLDβ1), an enzyme that hydrolyzes membrane phospholipids to generate phosphatidic acid (PA). Deficient *Arabidopsis* mutants in PLDβ1 show susceptibility to the fungal pathogen *B. cinerea* (Zhao et al., 2013). The AT2G42010 expression pattern reveals that this target would be negatively regulated at 72- and 96 hcc by the endogenous sRNA ncRNA_13700, suggesting a putative role in regulating the plant immune response to allow mutualistic symbiosis establishment.

On the other hand, AT5G49690 codes for a UDP-glycosyltransferase (UGT91C1), an enzyme which predicted to perform glycosylations in soya saponins (triterpenoids that form protective surfactants in plant cells). A previous study links the activity of UGT91C1 to pesticide detoxification in *Arabidopsis*, by glycosylating sulcotrione, a triketone herbicide. As sulcotrione mode of action is the inhibition of the enzyme 4-hydroxyphenylpyruvate dioxygenase (HPPD) activity, causing feedback suppression in tyrosine metabolism, glycosylation of sulcotrione resulted in the enhancing of the plant tyrosine metabolic pathway genes (Huang et al., 2021). There is evidence that *Arabidopsis* challenging against *Pseudomonas syringae* pv. tomato strain DC3000 (DC3000) suffers metabolite reconfigurations after 12 hcc that drastically enhance the production of tyrosine (Ward et al., 2010). These findings, together with the observed expression patterns here, suggests that UGT91C1 would be implicated in the defense response by glycosylating un yet described targets and being negatively regulated by *T. atroviride* during 24- and 48 hcc.

Taking into account that the possible regulation of these two putative targets occurs in separated time lapses, UGT91C1 at 24- to 48 hcc and PLDβ1 at 72- to 96 hcc, and in different contexts (UGT91C1 is targeted by *Trichoderma*'s sRNA and PLDβ1 by *Arabidopsis* sRNA) would suggest that *Arabidopsis* immune modulation is an essential and complex biological process that needs to be regulated in different times by the two organisms through tRFs to establish a mutualistic interaction.

In our results, *T. atroviride* candidate targets were modulated similarly during the interaction times analyzed. Although *Trichoderma*'s candidate targets could not be associated with a biological process, non-redundant BLASTp analyzes on TRIATDRAFT_150201 and TRIATDRAFT_150849 resulted in matches with metallothionein expression activators and ribosomal subunit S7 proteins, respectively. Metallothioneins (MTs) are metal-binding proteins involved in diverse processes, including metal homeostasis and detoxification (Takahashi, 2015). These findings suggest that *Arabidopsis* tRFs would be regulating protein translation and survival capabilities in *Trichoderma* during 24- to 48 hcc.

Our network analysis clustered 24 dissimilar abundant sRNAs and 50 DE genes during *Trichoderma-Arabidopsis* interaction into 11 communities, from which 4 presented associations that would lead to an sRNA-target regulation. Three out of four communities were associated with the plant's defense response (as is the case of communities 1, 2, and 6) implying that immunity in plants is regulated in different ways by the tRFs.

Another interesting aspect of the network analysis is that it allowed us to find sRNA target interactions with positive correlations. These occurring for ncRNA_26071 and ncRNA_27604 (*Trichoderma'*s tRHs) with TRIATDRAFT_247300 and AT1G06080, respectively. This suggests a positive regulation in gene expression by the action of tRHs. Recently, tRHs have been reported modulating positively mRNA targets on primate hippocampal tissues, being defined as ”shelter target” due to the possibility that tRH interaction with their targets could be stabilizing the mRNAs (Jehn et al., 2020). As tRFs can interact with other effector proteins or act protein-independent to regulate genes (Deng et al., 2015; Luo et al., 2018), tRFs would conduct different sRNA regulation processes that are not currently known, which would lead to sheltering of their targets in *Trichoderma-Arabidopsis* interaction.

Although our analyses suggest that a ck-RNAi seems to be occurring during *Trichoderma-Arabidopsis* interaction, reverse transcription-polymerase chain reaction (qRT-PCR) are needed to validate the changes in abundance and expression of sRNAs and their targets, respectively. In a same way, validation of sRNA biogenesis mechanisms would be performed by qRT-PCR over DCL, AGO and RNAseT2 mutants in both organisms.

Further experimental analyses, like AGO-RNA immunoprecipitation followed by high-throughput sequencing (AGO RIP-Seq), short tandem target mimics (STTM) and qRT-PCR over sRNA mutants, are needed to validate these sRNA-target associations, as well as to deduce functional regulations through phenotype observations.

Deeper bioinformatic analyses involving gene annotation, to annotate genes with no annotation or unknown functions, and sRNA-mRNA-AGO molecular docking are needed to better describe the interaction mechanisms of selected tRHs with their corresponding targets and the possible modulations in both organisms.

Conclusions

Both *Arabidopsis* and *Trichoderma* produce sRNAs at distinct times of interaction through 24- to 96 hcc. From these sRNAs, tRFs (more specifically tRHs), are the main type of abundant sRNA during *Arabidopsis*-*Trichoderma* interaction, having the most notorious changes through 48- to 72 hcc in *Arabidopsis* and at 96 hcc in *Trichoderma* (Fig. 9A).

Cross-kingdom communication by tRFs would be performed by both organisms modulating as principal aspect defense response of the plant. Although *Trichoderma* ck-RNAi regulation processes by these tRFs are unknown, functional analysis results point out to affect the survival and protein translation processes in *Trichoderma* from 24- to 48 hcc (Fig. 9B).

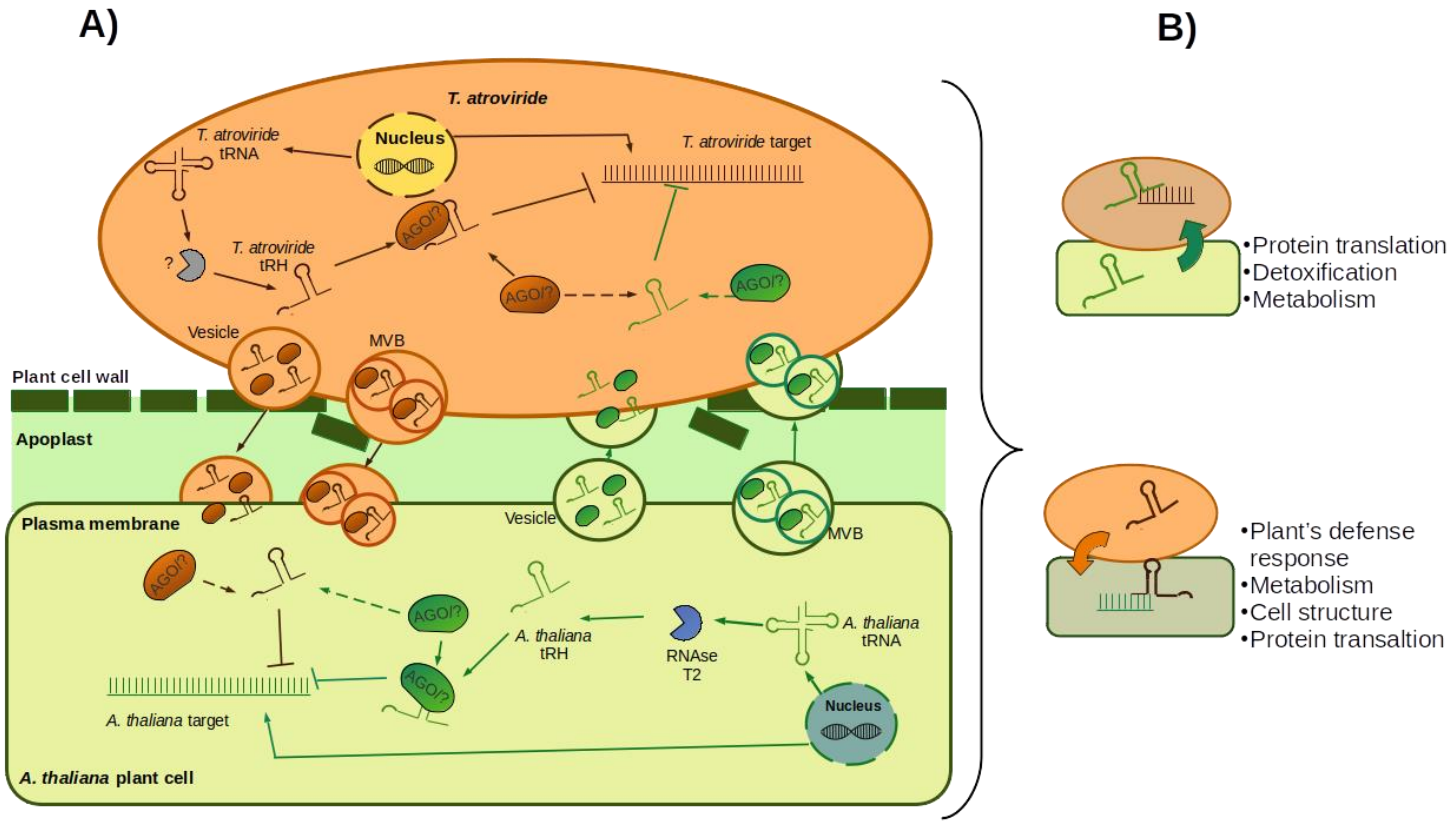


Fig. 9. Plant defense and fungus survival functions would be potentially affected by cross-kingdom communication, guided by tRHs, through the *Trichoderma-Arabidopsis* 24-, 48-, 72- and 96 hcc. (A) An schema of *T. atroviride-A. thaliana* tRH mediated ck-RNAi. Both organisms produce, through the cleavage of endogenous tRNAs, tRH products. This processing could be performed by RNAse T2 (RNS) in *A. thaliana* and by unknown RNAse in *T. atroviride*. Then, tRHs would be transported through ck-RNAi transport mechanisms, like vesicles or multivesicular bodies (MVB), to another organism. The tRHs associate with cargo proteins like endogenous argonautes proteins (AGO) that would be transported with them, other organism argonauts, or other proteins with tRH cargo capacity. Assembled tRHs perform target regulation by binding to the mRNA. (B) Biological functions that were observed being affected by tRH target prediction or by co-expression negative correlations in *T. atrovirde* (Top) and *A. thaliana* (Bottom).

Methods

RNA sequencing

A. thaliana roots inoculation with *T. atroviride* was performed and then collected from the interaction zone at 24, 48, 72, and 96 hcc. Total RNA was extracted and sRNA-seq libraries were generated according to the sRNA cloning protocol provided by Solexa.

RNA-seq data processing

Sequence quality checks were done using the MultiQC software (Ewels et al., 2016) using the function multiqc. Then removal of adapters (trimming) and size selection of read sequences with 18 to 50 nt was performed using reaper and Tally (Davis et al., 2013). The functions reaper and tally were used with the modified parameters "3p-global 12/2/1 -3p-prefix 8/2/1 -3p-head-to-tail 1 -nnn-check 3/5 -dust-suffix 20 -polya 5 -qqq-check 35/10 -mr-tabu 14/2/1" and "-l 18 -u 50 -tri 50", respectively.

Generation of a concatenated reference genome and transcriptome for Trichoderma and Arabidopsis

A reference genome was generated by concatenating the *A. thaliana* TAIR10.47(ftp://ftp.ensemblgenomes.org/pub/plants/release-51/fasta/arabidopsis_thaliana/dna/) and *T. atroviride* TRIAT_v2.0.47(ftp://ftp.ensemblgenomes.org/pub/plants/release-51/fasta/arabidopsis_thaliana/dna/) genome versions, by custom shell scripts. This allowed us for searching reads that could have an ambiguous origin region (sRNA producing *loci* that are conserved between the two organisms) and make equal conditions for the two organism genomes at the time of read mapping (Bermúdez-Barrientos et al., 2019). Following the same methodology, a reference transcriptome was generated by concatenating the TAIR10.47 (ftp://ftp.ensemblgenomes.org/pub/plants/release-51/fasta/arabidopsis_thaliana/cdna/Arabidopsis_thaliana.TAIR10.cdna.all.fa.gz) and TRIAT_v2.0.47 (ftp://ftp.ensemblgenomes.org/pub/fungi/release-51/fasta/fungi_ascomycota1_collection/trichoderma_atroviride_imi_206040_gca_000171015/cdna/Trichoderma_atroviride_imi_206040_gca_000171015.TRIAT_v2.0.cdna.all.fa.gz) cDNA (complementary DNA) versions.

Annotation of sRNA producing loci

Reads were mapped to the reference genome using Bowtie 1.0 (Langmead et al., 2009) with the function bowtie according to the following parameters: "-q -v 2 -p 6 -S -a -m 500 --best –strata". From the results, 3 different kinds of mapping were considered: 1) unique mapping: reads that only mapped to a specific region of an organism; 2) multimapping: reads mapping on more than one region in the same organism; 3) ambiguous mapping: reads mapping on regions in both organisms. The mapped reads were used to predict sRNA-producing *loci* using ShortStack (Axtell, 2013), MapMi (Guerra-Assunção & Enright, 2010), and Rfam (Kalvari et al., 2021). Predictions of tRNAs and rRNAs producing *loci* were also made using the tRNAscan-SE (Lowe & Chan, 2016) and RNAmmer (Lagesen et al., 2007) to associate them to possible tRFs or rRFs. All the annotation process was performed implementing custom shell scripts.

Analysis of sRNAs with statistical abundance changes through the time of interaction

The reads mapped to ncRNA annotated regions were quantified using custom R scripts using the R package GenomicFeatures (Lawrence et al., 2013). Reads abundances were normalized to counts per million (CPM) values and filtered considering *loci* with a minimum of 4 CPMs in at least two libraries for the differential expression analysis. The differential expression analysis was performed using edgeR (M. D. Robinson et al., 2009) to identify ncRNA *loci* with statistical abundance changes through time. This was performed using an ANOVA test through the function glmLRT using default parameters. The ncRNA *loci* were considered with statistical abundance changes if they have an FDR ≤ 0.1.

The ncRNAs *loci* with statistical abundance changes across time were then clustered, according to their CPM correlation through time, into 4 clusters using the function kmeans from the R stats package (R Core Team, 2021), with default parameters. A clustering by k-means is an unsupervised learning algorithm used to find the abundance similarity profiles between ncRNAs in the data. Then, a top of 8 sRNAs (4 for *A. thaliana* and for 4 *T. atroviride*, respectively) with the lowest FDR values were obtained from every cluster to be considered as representatives of that cluster.

Splitting of sRNAs in k-mers

Due that targeting rules are poorly understood for tRFs and rRFs, and a previous report showed that tRFs of ~30 nt long (tRHs) can interact with their targets by their 3' or center regions (Jehn et al., 2020). In the case of representative sRNAs ranging 30 to 50 nt long: we performed, a manual selection of a possible interacting region through the genome viewer tool IGV 2.8.13 (Robinson et al., 2011) if there was an observation of inconsistencies in nucleotide composition. At the same time, a fragmentation of the sequence into k-mers of 22 nt (or the size of the manually selected region) covering the 5', center and 3' sections was performed by custom R scripts to cover other potential target interacting regions.

sRNAs target prediction

Target prediction analyses were done by psRNATarget (Dai et al., 2018) using the chosen fragment sequences and the concatenated reference transcriptome, modifying the length of the scoring region for complementary analysis (HSP size) parameter according to the actual sRNAs length. A scoring schema, that considers the expectation, type of inhibition, and site accessibility, was created to qualify the predicted targets using custom-made R scripts.

GO enrichment analysis

Custom-made R scripts were generated following hypergeometric tests to determine target GO enrichment for each representative cluster of sRNAs. Then, GO enrichment p-values were compared against random sequences of each sRNA that preserved their original nucleotide composition, to discard possible noise attribution in sRNA regulation of a given functional process.

sRNAs target candidate's analysis

Putative sRNAs target expressions were searched against a count matrix of *Trichoderma-Arabidopsis* interaction mRNA libraries (48-, 72-, and 96 hcc). Target's mean expression changes through time were compared against their corresponding sRNA mean abundances through Spearman correlation tests using the cor.test function from the R stats package (R Core Team, 2021), with modified parameter “method = spearman”. The top 2 targets per sRNA were selected based on their prediction scores, correlations, and correlation p-values for candidate selection.

Additionally, those targets with correlation p-values ≤ 0.1 were searched in the STRING database through the STRINGdb R package (Szklarczyk et al., 2019) using the function string_db$map with default parameters. Top 10 STRING predicted interacting proteins were used to deduce biological function for those targets candidates without GO BP information available.

Differential expression analysis and clustering of genes

Similar analysis processing was performed as with sRNAs abundances. Reads derived from the count matrix mRNA-seq data were normalized to CPM and filtered considering 4 CPMs in at least three libraries. Then the differential expression analysis was performed using edgeR (Robinson et al., 2009) through an ANOVA test. Genes were considered deferentially expressed (DE) if they have an FDR ≤ 0.1. DE genes were clustered by k-means clustering, according to their CPM correlation, into 8 clusters for *A. thaliana* genes and 6 for *T. atroviride*. In addition, the top predicted target candidates were clustered inside these clusters.

sRNA-gene co-expression network analysis

Expression or abundance from the cluster representatives of the sRNAseq and mRNA-seq datasets, as well as the selected candidate targets, were used for the generation of the sRNA-gene co-expression network. Correlation between the abundances was performed through Spearman correlation tests and interactions (correlations) with p-values ≤ 0.1 were selected. Generation, analysis and visualization of the network was performed with the igraph R package (Csardi & Nepusz, 2006) through the function graph_from_data_frame using default parameters. The network considered absolute correlation values as the edge weight feature. Communities of interaction were determined by the random walk clusterization algorithm (Pons & Latapy, 2005), using the function cluster_walktrap with default parameters, provided by the igraph package.

References

Axtell, M. J. (2013). ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA*, *19*(6), 740–751. https://doi.org/10.1261/rna.035279.112

Brilli, M., Asquini, E., Moser, M., Bianchedi, P. L., Perazzolli, M., & Si-Ammour, A. (2018). A multi-omics study of the grapevine-downy mildew (Plasmopara viticola) pathosystem unveils a complex protein coding-A nd noncoding-based arms race during infection. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-19158-8

Cai, Q., Qiao, L., Wang, M., He, B., Lin, F. M., Palmquist, J., Huang, S. Da, & Jin, H. (2018). Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes. *Science*, *360*(6393), 1126–1129. https://doi.org/10.1126/science.aar4142

Carreras-Villaseñor, N., Esquivel-Naranjo, E. U., Villalobos-Escobedo, J. M., Abreu-Goodger, C., & Herrera-Estrella, A. (2013). The RNAi machinery regulates growth and development in the filamentous fungus *Trichoderma atroviride*. *Molecular Microbiology*, *89*(1), 96–112. https://doi.org/10.1111/mmi.12261

Chen, H., Kobayashi, K., Miyao, A., Hirochika, H., Yamaoka, N., & Nishiguchi, M. (2013). Both OsRecQ1 and OsRDR1 Are Required for the Production of Small RNA in Response to DNA-Damage in Rice. *PLoS ONE*, *8*(1), e55252. https://doi.org/10.1371/journal.pone.0055252

Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., Feng, G. H., Peng, H., Zhang, X., Zhang, Y., Qian, J., Duan, E., Zhai, Q., & Zhou, Q. (2016). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*, *351*(6271), 397–400. https://doi.org/10.1126/science.aad7977

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Sy*, 1695. https://igraph.org

Dai, X., Zhuang, Z., & Zhao, P. X. (2018). PsRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Research*, *46*(W1), W49–W54. https://doi.org/10.1093/nar/gky316

Davis, M. P. A., van Dongen, S., Abreu-Goodger, C., Bartonicek, N., & Enright, A. J. (2013). Kraken: A set of tools for quality control and analysis of high-throughput sequence data. *Methods*, *63*(1), 41–49. https://doi.org/10.1016/j.ymeth.2013.06.027

De Palma, M., Salzano, M., Villano, C., Aversano, R., Lorito, M., Ruocco, M., Docimo, T., Piccinelli, A. L., D'Agostino, N., & Tucci, M. (2019).

Transcriptome reprogramming, epigenetic modifications and alternative splicing orchestrate the tomato root response to the beneficial fungus *Trichoderma harzianum*. *Horticulture Research*, *6*(1), 5. https://doi.org/10.1038/s41438-018-0079-1

Deng, J., Ptashkin, R. N., Chen, Y., Cheng, Z., Liu, G., Phan, T., Deng, X., Zhou, J., Lee, I., Lee, Y. S., & Bao, X. (2015). Respiratory Syncytial Virus Utilizes a tRNA Fragment to Suppress Antiviral Responses Through a Novel Targeting Mechanism. *Molecular Therapy*, *23*(10), 1622–1629. https://doi.org/10.1038/mt.2015.124

Estrada-Rivera, M., Rebolledo-Prudencio, O. G., Pérez-Robles, D. A., Rocha-Medina, M. del C., González-López, M. del C., & Casas-Flores, S. (2019). *Trichoderma* Histone Deacetylase HDA-2 Modulates Multiple Responses in Arabidopsis. *Plant Physiology*, *179*(4), 1343–1361. https://doi.org/10.1104/pp.18.01092

Ewels, P., Ns Magnusson, M., Lundin, S., & Aller, M. K. (2016). Data and text mining MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature*, *391*(6669), 806–811. https://doi.org/10.1038/35888

Garcia-Silva, M. R., Cabrera-Cabrera, F., Cura Das Neves, R. F., Souto-Padrón, T., De Souza, W., & Cayota, A. (2014). Gene expression changes induced by *Trypanosoma cruzi* shed Microvesicles in mammalian host cells: Relevance of tRNA-derived halves. *BioMed Research International*, *2014*. https://doi.org/10.1155/2014/305239

Gebetsberger, J., Wyss, L., Mleczko, A. M., Reuther, J., & Polacek, N. (2017). A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biology*, *14*(10), 1364–1373. https://doi.org/10.1080/15476286.2016.1257470

González-Pérez, E., Ortega-Amaro, M. A., Salazar-Badillo, F. B., Bautista, E., Douterlungne, D., & Jiménez-Bremont, J. F. (2018). The Arabidopsis-Trichoderma interaction reveals that the fungal growth medium is an important factor in plant growth induction. *Scientific Reports*, *8*(1), 1–14. https://doi.org/10.1038/s41598-018-34500-w

Guerra-Assunção, J. A., & Enright, A. J. (2010). MapMi: Automated mapping of microRNA *loci*. *BMC Bioinformatics*, *11*(1), 133. https://doi.org/10.1186/1471-2105-11-133

Guler, N. S., Pehlivan, N., Karaoglu, S. A., Guzel, S., & Bozdeveci, A. (2016). *Trichoderma atroviride* ID20G inoculation ameliorates drought stress-induced damages by improving antioxidant defence in maize seedlings. *Acta Physiologiae Plantarum*, *38*(6), 1–9. https://doi.org/10.1007/s11738-016-2153-3

Guo, Q., Liu, Q., A. Smith, N., Liang, G., & Wang, M.-B. (2016). RNA Silencing in Plants: Mechanisms, Technologies and Applications in Horticultural Crops. *Current Genomics*, *17*(6), 476–489. https://doi.org/10.2174/1389202917666160520103117

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, *28*(1), 100. https://doi.org/10.2307/2346830

Henderson, I. R., Zhang, X., Lu, C., Johnson, L., Meyers, B. C., Green, P. J., & Jacobsen, S. E. (2006). Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nature Genetics*, *38*(6), 721–725. https://doi.org/10.1038/ng1804

Hou, Y., Zhai, Y., Feng, L., Karimi, H. Z., Rutter, B. D., Zeng, L., Choi, D. S., Zhang, B., Gu, W., Chen, X., Ye, W., Innes, R. W., Zhai, J., & Ma, W. (2019). A Phytophthora Effector Suppresses Trans-Kingdom RNAi to Promote Disease Susceptibility. *Cell Host and Microbe*, *25*(1), 153-165.e5. https://doi.org/10.1016/j.chom.2018.11.007

Huang, X. xu, Zhao, S. man, Zhang, Y. ying, Li, Y. jie, Shen, H. nuo, Li, X., & Hou, B. kai. (2021). A novel UDP-glycosyltransferase 91C1 confers specific herbicide resistance through detoxification reaction in Arabidopsis. *Plant Physiology and Biochemistry*, *159*, 226–233. https://doi.org/10.1016/j.plaphy.2020.12.026

Ivanov, P., Emara, M. M., Villen, J., Gygi, S. P., & Anderson, P. (2011). Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. *Molecular Cell*, *43*(4), 613–623. https://doi.org/10.1016/j.molcel.2011.06.022

Jehn, J., Treml, J., Wulsch, S., Ottum, B., Erb, V., Hewel, C., Kooijmans, R. N., Wester, L., Fast, I., & Rosenkranz, D. (2020). 5′ tRNA halves are highly expressed in the primate hippocampus and might sequence-specifically regulate gene expression. *RNA*, *26*(6), 694–707. https://doi.org/10.1261/rna.073395.119

Jöchl, C., Rederstorff, M., Hertel, J., Stadler, P. F., Hofacker, I. I., Schrettl, M., Haas, H., & Hüttenhofer, A. (2008). Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein synthesis. *Nucleic Acids Research*, *36*(8), 2677–2689. https://doi.org/10.1093/nar/gkn123

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. In *Nature* (Vol. 444, Issue 7117, pp. 323–329). Nature Publishing Group. https://doi.org/10.1038/nature05286

Kalvari, I., Nawrocki, E. P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., Rivas, E., Eddy, S. R., Finn, R. D., Bateman, A., & Petrov, A. I. (2021). Rfam 14: Expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research*, *49*(D1), D192–D200. https://doi.org/10.1093/nar/gkaa1047

Kapulnik, Y., Volpin, H., Itzhaki, H., Ganon, D., Galili, S., David, R., Shaul, O., Elad, Y., Chet, I., & Okon, Y. (1996). Suppression of defence responses in mycorrhizal alfalfa and tobacco roots. *New Phytologist*, *133*(1), 59–64. https://doi.org/10.1111/j.1469-8137.1996.tb04341.x

Kleifeld, O., & Chet, I. (1992). *Trichoderma harzianum*-interaction with plants and effect on growth response. *Plant and Soil*, *144*(2), 267–272. https://doi.org/10.1007/BF00012884

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100–3108. https://doi.org/10.1093/nar/gkm160

Lalande, S., Merret, R., Salinas-Giegé, T., & Drouard, L. (2020). Arabidopsis tRNA-derived fragments as potential modulators of translation. *RNA Biology*, *17*(8), 1137–1148. https://doi.org/10.1080/15476286.2020.1722514

Lambert, M., Benmoussa, A., & Provost, P. (2019). Small Non-Coding RNAs Derived From Eukaryotic Ribosomal RNA. *Non-Coding RNA*, *5*(1), 16. https://doi.org/10.3390/ncrna5010016

Lamonte G., Philip N., Reardon J., Lacsina J.R., Majoros W., Chapman L, Thornburg C.D., Telen M.J., Ohler U., Nicchitta C.V., Haystead T., and Chi J.T. (2012). Translocation of sickle cell erythrocyte MicroRNAs into Plasmodium falciparum inhibits parasite translation and contributes to malaria resistance. Cell Host Microbe 12: 187–199, doi:10.1016/j.chom.2012.06.007

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Locati, M. D., Pagano, J. F. B., Abdullah, F., Ensink, W. A., Van Olst, M., Van Leeuwen, S., Nehrdich, U., Spaink, H. P., Rauwerda, H., Jonker, M. J., Dekker, R. J., & Breit, T. M. (2018). Identifying small RNAs derived from maternal-and somatic-type rRNAs in zebrafish development. *Genome*, *61*(5), 371–378. https://doi.org/10.1139/gen-2017-0202

Loher, P., Telonis, A. G., & Rigoutsos, I. (2017). MINTmap: Fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data. *Scientific Reports*, *7*(1), 1–20. https://doi.org/10.1038/srep41184

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, *44*(W1), W54–W57. https://doi.org/10.1093/nar/gkw413

Luo, S., He, F., Luo, J., Dou, S., Wang, Y., Guo, A., & Lu, J. (2018). Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response. *Nucleic Acids Research*, *46*(10), 5250–5268. https://doi.org/10.1093/nar/gky189

Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E. G. J., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., Salamov, A., Shapiro, H. J., Wuyts, J., Blaudez, D., Buée, M., Brokstein, P., Canbäck, B., Cohen, D., Courty, P. E., … Grigoriev, I. V. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*, *452*(7183), 88–92. https://doi.org/10.1038/nature06556

Mewalal, R., Yin, H., Hu, R., Jawdy, S., Vion, P., Tuskan, G. A., Tacon, F. Le, Labbé, J. L., & Yang, X. (2019). Identification of Populus small RNAs responsive to mutualistic interactions with mycorrhizal fungi, *Laccaria bicolor* and *Rhizophagus irregularis*. *Frontiers in Microbiology*, *10*(MAR), 515. https://doi.org/10.3389/fmicb.2019.00515

Oh, S. U., Yun, B. S., Lee, S. J., Kim, J. H., & Yoo, I. D. (2002). Atroviridins A~C and neoatroviridins A~D, novel peptaibol antibiotics produced by Trichoderma atroviride F80317. I. Taxonomy, fermentation, isolation and biological activities. *Journal of Antibiotics*, *55*(6), 557–564. https://doi.org/10.7164/antibiotics.55.557

Pieterse, C. M. J., Zamioudis, C., Berendsen, R. L., Weller, D. M., Van Wees, S. C. M., & Bakker, P. A. H. M. (2014). Induced systemic resistance by beneficial microbes. *Annual Review of Phytopathology*, *52*, 347–375. https://doi.org/10.1146/annurev-phyto-082712-102340

Plett, J. M., Kemppainen, M., Kale, S. D., Kohler, A., Legué, V., Brun, A., Tyler, B. M., Pardo, A. G., & Martin, F. (2011). A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Current Biology*, *21*(14), 1197–1203. https://doi.org/10.1016/j.cub.2011.05.033

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3733 LNCS*, 284–293. https://doi.org/10.1007/11569596_31

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Ramírez-Valdespino, C. A., Casas-Flores, S., & Olmedo-Monfil, V. (2019). Trichoderma as a model to study effector-like molecules. In *Frontiers in Microbiology* (Vol. 10, Issue MAY, p. 1030). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2019.01030

Ren, B., Wang, X., Duan, J., & Ma, J. (2019). Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. *Science*, *365*(6456), 919–922. https://doi.org/10.1126/science.aav8907

Roberto Bermúdez-Barrientos, J., Ramírez-S, O., Wang-Ngai Chow, F., Buck, A. H., & Abreu-Goodger, C. (2019). Disentangling sRNA-Seq data to study RNA communication between species. *Nucleic Acids Research*, *48*(4), 21. https://doi.org/10.1093/nar/gkz1198

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. In *Nature Biotechnology* (Vol. 29, Issue 1, pp. 24–26). Nature Publishing Group. https://doi.org/10.1038/nbt.1754

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rocha-Ramírez, V., Omero, C., Chet, I., Horwitz, B. A., & Herrera-Estrella, A. (2002). *Trichoderma atroviride* G-protein α-subunit gene tga1 is involved in mycoparasitic coiling and conidiation. *Eukaryotic Cell*, *1*(4), 594–605. https://doi.org/10.1128/EC.1.4.594-605.2002

Salas-Marina, M. A., Silva-Flores, M. A., Uresti-Rivera, E. E., Castro-Longoria, E., Herrera-Estrella, A., & Casas-Flores, S. (2011). Colonization of Arabidopsis roots by *Trichoderma atroviride* promotes growth and enhances systemic disease resistance through jasmonic acid/ethylene and salicylic acid pathways. *European Journal of Plant Pathology*, *131*(1), 15–26. https://doi.org/10.1007/s10658-011-9782-6

Selin, C., de Kievit, T. R., Belmonte, M. F., & Fernando, W. G. D. (2016). Elucidating the role of effectors in plant-fungal interactions: Progress and challenges. In *Frontiers in Microbiology* (Vol. 7, Issue APR, p. 600). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2016.00600

Shahid, S., Kim, G., Johnson, N. R., Wafula, E., Wang, F., Coruh, C., Bernal-Galeano, V., Phifer, T., Depamphilis, C. W., Westwood, J. H., & Axtell, M. J. (2018). MicroRNAs from the parasitic plant *Cuscuta campestris* target host messenger RNAs. *Nature*, *553*(7686), 82–85. https://doi.org/10.1038/nature25027

Shapulatov, U. M., Buriev, Z. T., Ulloa, M., Saha, S., Devor, E. J., Ayubov, M. S., Norov, T. M., Shermatov, S. E., Abdukarimov, A., Jenkins, J. N., & Abdurakhmonov, I. Y. (2016). Characterization of Small RNAs and Their Targets from *Fusarium oxysporum* Infected and Noninfected Cotton Root Tissues. *Plant Molecular Biology Reporter*, *34*(3), 698–706. https://doi.org/10.1007/s11105-015-0945-z

Silvestri, A., Fiorilli, V., Miozzi, L., Accotto, G. P., Turina, M., & Lanfranco, L. (2019). In silico analysis of fungal small RNA accumulation reveals putative plant mRNA targets in the symbiosis between an arbuscular mycorrhizal fungus and its host plant. *BMC Genomics*, *20*(1), 1–18. https://doi.org/10.1186/s12864-019-5561-0

Su, Z., Kuscu, C., Malik, A., Shibata, E., & Dutta, A. (2019). Angiogenin generates specific stress-induced tRNA halves and is not involved in tRF-3-mediated gene silencing. *Journal of Biological Chemistry*, *294*(45), 16930–16941. https://doi.org/10.1074/jbc.RA119.009272

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Takahashi, S. (2015). Positive and negative regulators of the metallothionein gene (Review). In *Molecular Medicine Reports* (Vol. 12, Issue 1, pp. 795–799). Spandidos Publications. https://doi.org/10.3892/mmr.2015.3459

Thompson, D. M., Lu, C., Green, P. J., & Parker, R. (2008). tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA*, *14*(10), 2095–2103. https://doi.org/10.1261/rna.1232808

Thompson, D. M., & Parker, R. (2009). Stressing Out over tRNA Cleavage. In *Cell* (Vol. 138, Issue 2, pp. 215–219). Elsevier B.V. https://doi.org/10.1016/j.cell.2009.07.001

Ward, J. L., Forcat, S., Beckmann, M., Bennett, M., Miller, S. J., Baker, J. M., Hawkins, N. D., Vermeer, C. P., Lu, C., Lin, W., Truman, W. M., Beale, M. H., Draper, J., Mansfield, J. W., & Grant, M. (2010). The metabolic transition during disease following infection of Arabidopsis thaliana by *Pseudomonas*

syringae pv. tomato. *Plant Journal*, *63*(3), 443–457. https://doi.org/10.1111/j.1365-313X.2010.04254.x

Weiberg, A., Wang, M., Lin, F. M., Zhao, H., Zhang, Z., Kaloshian, I., Huang, H. Da, & Jin, H. (2013). Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science*, *342*(6154), 118–123. https://doi.org/10.1126/science.1239705

Xie, Y., Yao, L., Yu, X., Ruan, Y., Li, Z., & Guo, J. (2020). Action mechanisms and research methods of tRNA-derived small RNAs. In *Signal Transduction and Targeted Therapy* (Vol. 5, Issue 1, pp. 1–9). Springer Nature. https://doi.org/10.1038/s41392-020-00217-4

Zamioudis, C., & Pieterse, C. M. J. (2012). Modulation of host immunity by beneficial microbes. In *Molecular Plant-Microbe Interactions* (Vol. 25, Issue 2, pp. 139–150). The American Phytopathological Society . https://doi.org/10.1094/MPMI-06-11-0179

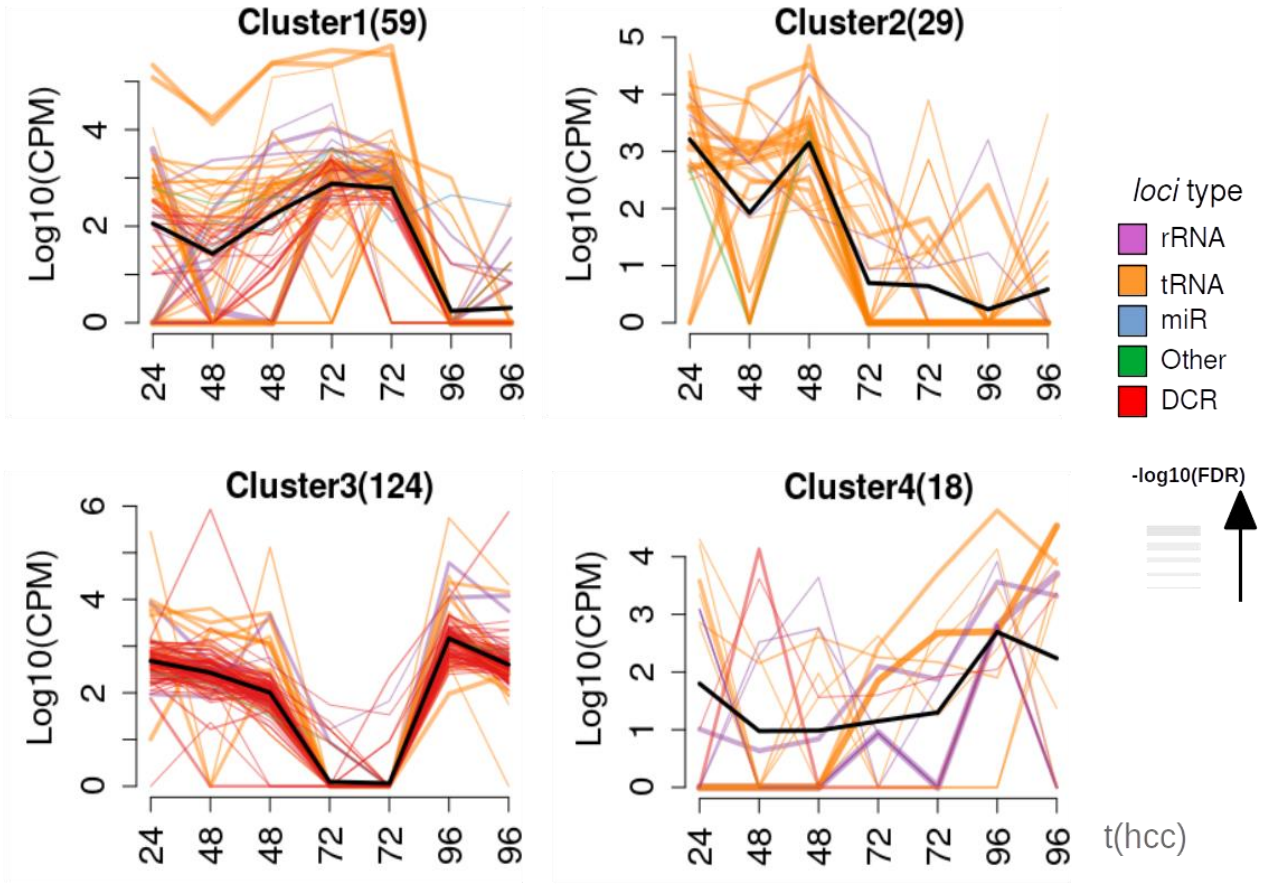
Zhang, C. (2009). Novel functions for small RNA molecules. In *Current Opinion in Molecular Therapeutics* (Vol. 11, Issue 6, pp. 641–651). NIH Public Access. /pmc/articles/PMC3593927/

Zhang, T., Zhao, Y. L., Zhao, J. H., Wang, S., Jin, Y., Chen, Z. Q., Fang, Y. Y., Hua, C. L., Ding, S. W., & Guo, H. S. (2016). Cotton plants export microRNAs to inhibit virulence gene expression in a fungal pathogen. *Nature Plants*, *2*(10), 1–6. https://doi.org/10.1038/nplants.2016.153

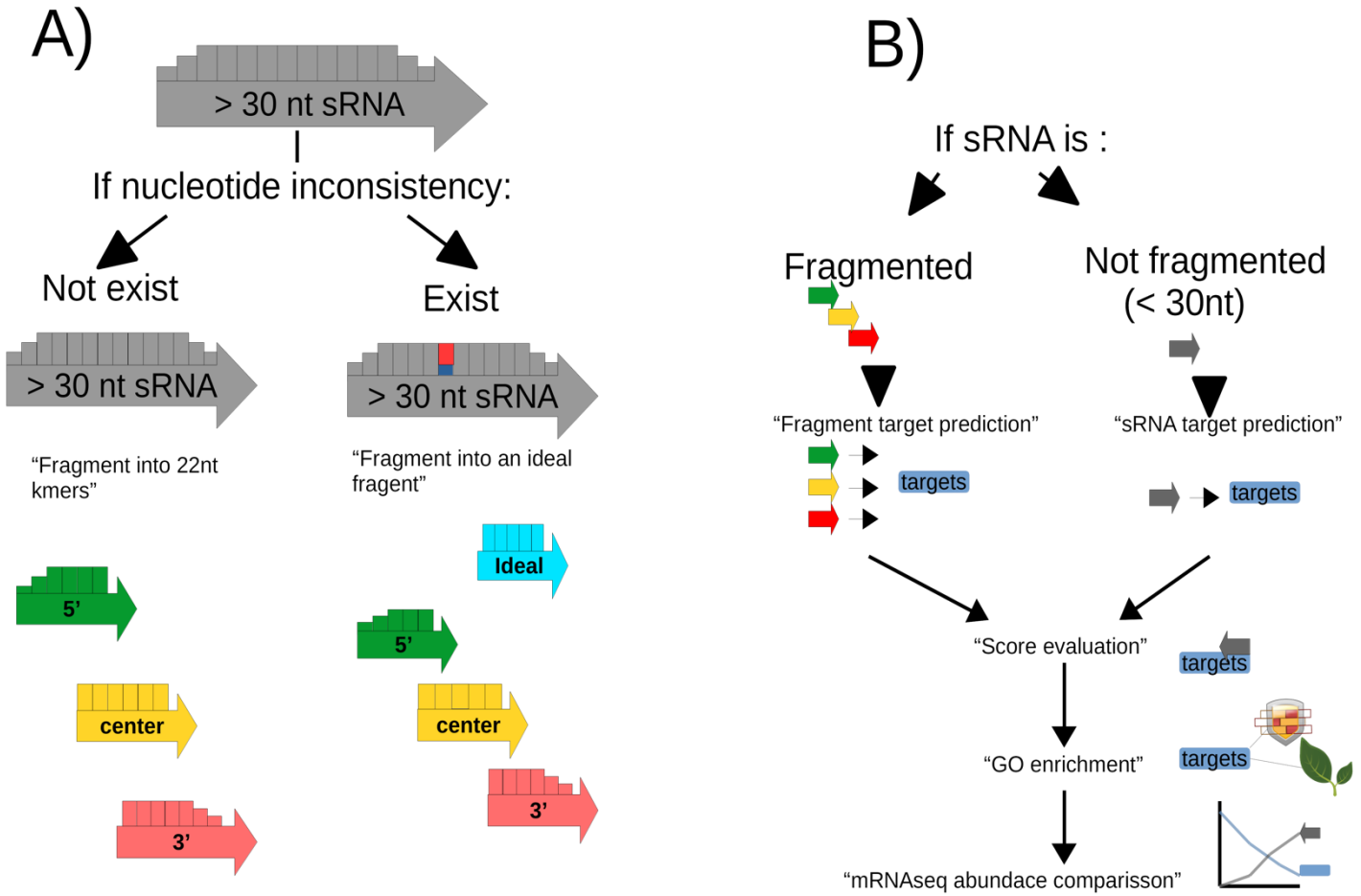
Zhang, P., Wu, W., Chen, Q., & Chen, M. (2019). Non-Coding RNAs and their Integrated Networks. In *Journal of integrative bioinformatics* (Vol. 16, Issue 3). NLM (Medline). https://doi.org/10.1515/jib-2019-0027

Zhao, J., Devaiah, S. P., Wang, C., Li, M., Welti, R., & Wang, X. (2013). Arabidopsis phospholipase Dβ1 modulates defense responses to bacterial and fungal pathogens. New Phytologist, 199(1), 228–240. https://doi.org/10.1111/nph.12256

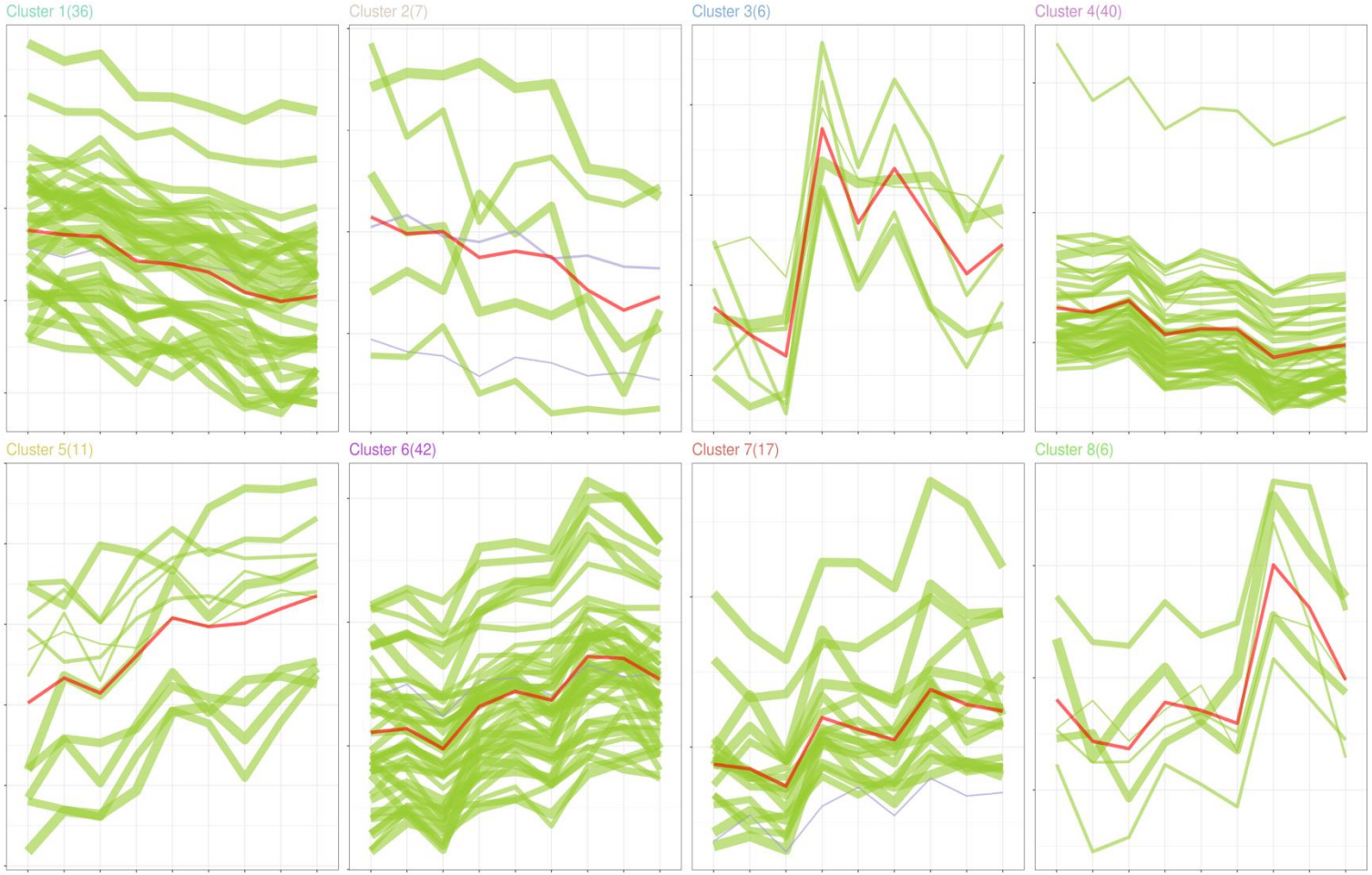
Supplementary information



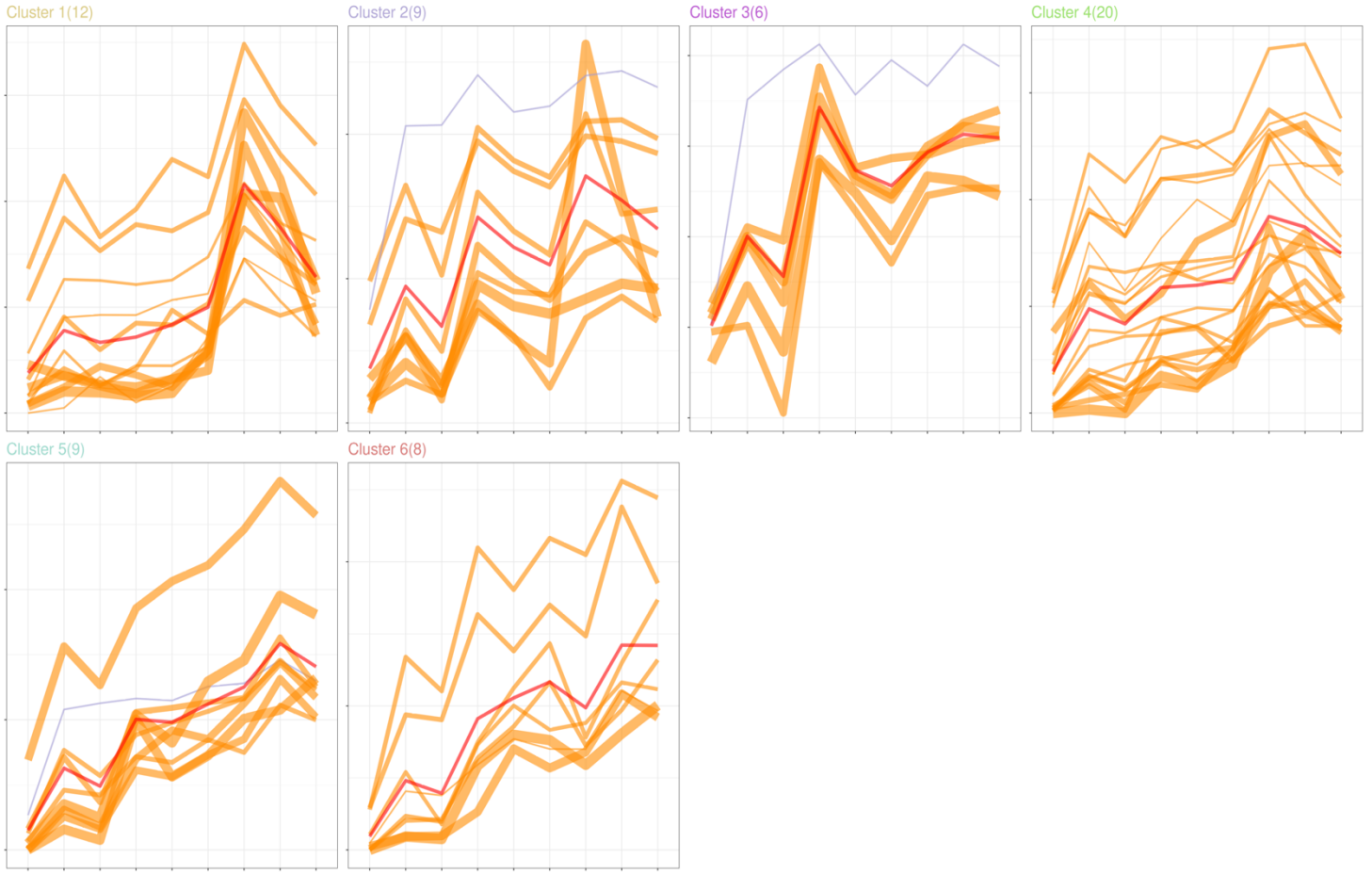
Supplementary Fig. 1. Dissimilar abundant sRNAs were grouped into four clusters (each with a representative pattern). The abundances through time of dissimilar abundant *loci* that compose the 4 clusters and its representative cluster pattern (black line). Each sRNA *loci* is represented as a line, time analyzed is shown in the x-axis and their abundance, in $\log_{10}$(CPM), in the y-axis. The color of the line represents the *loci* type of each sRNAs and the width of their statistical -$\log_{10}$(FDR) values from the DE analysis.



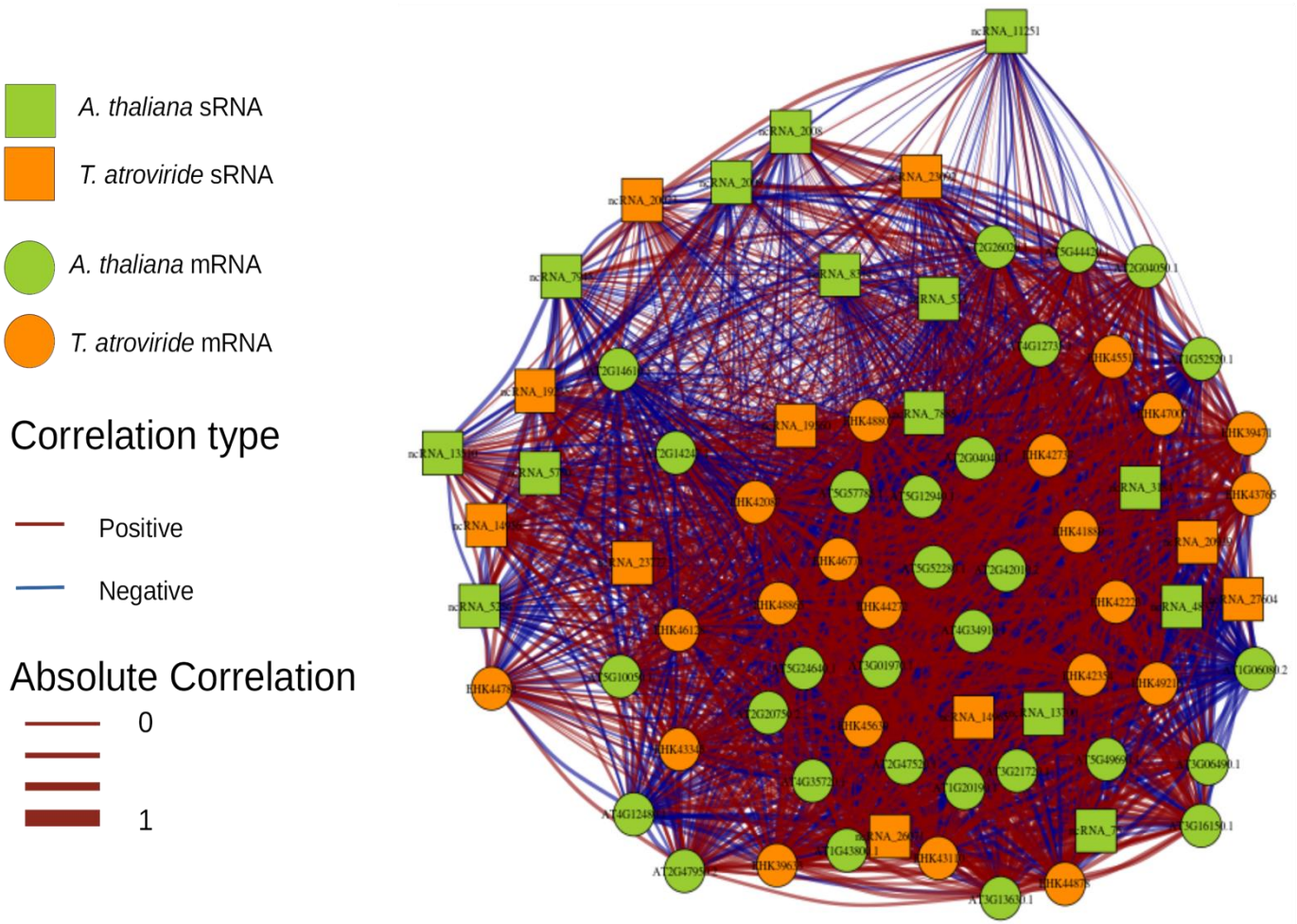
Supplementary Fig. 2. sRNA k-mer splitting and prediction analyses workflow.(A) sRNAs longer than 30nt were fragmented into 5', center, and 3' k-mers of 22 nt in length. In case of detecting nucleotide inconsistencies in the sequence composition, a manually selected (ideal) fragment was established, and fragment sizes were adjusted to ideal fragment size. (B) Subsequently, complete sRNAs, or sRNAs fragments, sequences were given to psRNATarget to predict possible targets and subsequently evaluate their regulation and implication in *Trichoderma-Arabidopsis* interaction.



Supplementary Fig. 3. Clusters of DE genes from *A. thaliana*. The expression through time of the 160 DE genes that compose the 8 clusters (green), its representative cluster pattern (red line), and *A. thaliana* target candidates (purple).



Supplementary Fig. 4. Clusters of DE genes from *T. atroviride*. The expression through time of the 60 DE genes that compose the 6 clusters (orange), its representative cluster pattern (red line), and *T. atroviride* target candidates (purple).



Supplementary Fig. 5. sRNA-gene co-expression network was composed by 2,701 connections. Total interaction network from the expression/abundance correlations between 50 gene cluster representatives, 28 sRNAs cluster representatives and top 8 sRNA target candidates.