UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Biológicas e Ingeniería

TÍTULO: *In silico* Approaches for Prediction of Protein-Protein Interactions Between *Ralstonia solanacearum* GMI1000 and *Solanum lycopersicum*.

Trabajo de integración curricular presentado como requisito para la obtención del título de Bióloga

Autora:

Katlheen Nayade Sarmiento Fajardo

Tutor:

Ph.D. José Antonio Castillo Morales

Urcuquí, septiembre 2020

Urcuquí, 23 de octubre de 2020

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA
CARRERA DE BIOLOGÍA
ACTA DE DEFENSA No. UITEY-BIO-2020-00031-AD

A los 23 días del mes de octubre de 2020, a las 10:00 horas, de manera virtual mediante videoconferencia, y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.
Miembro No Tutor	Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.
Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.

El(la) señor(ita) estudiante **SARMIENTO FAJARDO, KATLHEEN NAYADE**, con cédula de identidad No. **1207468644**, de la **ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA**, de la Carrera de **BIOLOGÍA**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SO-37-No.438-2014**, realiza a través de videoconferencia, la sustentación de su trabajo de titulación denominado: **In silico Approaches for Prediction of Protein-Protein Interactions between Ralstonia solanacearum GMI1000 and Solanum lycopersicum**, previa a la obtención del título de **BIÓLOGO/A**.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación a través de videoconferencia, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Presidente Tribunal De Defensa	Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.	9,9
Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.	10,0
Miembro Tribunal De Defensa	Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.	10,0

Lo que da un promedio de: **10 (Diez punto Cero)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

Certifico que *en cumplimiento del Decreto Ejecutivo 1017 de 16 de marzo de 2020, la defensa de trabajo de titulación (o examen de grado modalidad teórico práctica) se realizó vía virtual, por lo que las firmas de los miembros del Tribunal de Defensa de Grado, constan en forma digital.*

KATLHEEN NAYADE SARMIENTO FAJARDO
Firmado digitalmente por KATLHEEN NAYADE SARMIENTO FAJARDO
Fecha: 2020.12.01 22:20:58 -05'00'

SARMIENTO FAJARDO, KATLHEEN NAYADE
Estudiante

FRANCISCO JAVIER ALVAREZ BOTAS
Digitally signed by FRANCISCO JAVIER ALVAREZ BOTAS
Date: 2020.11.16 20:01:51 -05'00'

Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.
Presidente Tribunal de Defensa

CERTIFICACIÓN ELECTRÓNICA
Firmado Digitalmente por: JOSE ANTONIO CASTILLO MORALES
Hora oficial Ecuador: 01/12/2020 16:53

Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.
Tutor

Firmado electrónicamente por:
NELSON
FRANCISCO
SANTIAGO VISPO

Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.
Miembro No Tutor

Firmado electrónicamente por:
KARLA
ESTEFANIA
ALARCON FELIX

ALARCON FELIX, KARLA ESTEFANIA
Secretario Ad-hoc

Autoría

Yo, **KATLHEEN NAYADE SARMIENTO FAJARDO**, con cédula de identidad 1207468644, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autor (a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, septiembre 2020.

Katlheen Nayade Sarmiento Fajardo

CI: 1207468644

Autorización de publicación

Yo, **KATLHEEN NAYADE SARMIENTO FAJARDO**, con cédula de identidad 1207468644, cedo a la Universidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, septiembre 2020.

Katlheen Nayade Sarmiento Fajardo

CI: 1207468644

Dedication

To my family, who are my support, guide, and motivation to overcome adversity and grow every day.

Katlheen Nayade Sarmiento Fajardo

Acknowledgements

To my thesis advisor Ph.D. José A. Castillo, who encouraged and supported me during the completion of this work.

To my family, friends, and teachers who have forged me into the person I am today with their patience and love.

Katlheen Nayade Sarmiento Fajardo

Resumen

Ralstonia solanacearum es una bacteria patógena de plantas conocida por su letalidad a nivel global, cuyo huésped por excelencia es *Solanum lycopersicum*, comúnmente llamado tomate, alimento originario de Sur América que genera anualmente ingresos importantes para el sector agrícola. *R. solanacearum* ingresa a la planta a través de las raíces e inicia el proceso patogénico al activar la secreción de proteínas especializadas llamadas efectores de tipo III (T3E). Una vez que los T3E de *R. solanacearum* realizan su actividad específica dentro de las células vegetales al interaccionar con proteínas de la planta, este patógeno genera marchitez y posteriormente la muerte de la planta. Por tanto, la infección de este patógeno significa un riesgo económico importante, ocasionando pérdidas de hasta un 50% en la producción anual de tomate. El presente trabajo se enfoca en inferir interacciones proteína-proteína (PPIs) entre los T3E de *R. solanacearum* GMI1000, una cepa que típicamente infecta tomate, y proteínas de tomate. El proceso patogénico consiste en gran medida en interacciones exitosas de proteínas, por esto el estudio de las PPIs permite deducir las funciones que cumplen las proteínas, sus posibles complejos y redes de interacción. Asimismo, aportan al entendimiento de la patogenicidad de *R. solanacearum*. En este trabajo, primero, se emplearon dos enfoques *in silico*, el método Interolog y el método basado en Dominios, obteniendo como resultado 21557 y 13615 PPIs, para el primer y segundo enfoque respectivamente. Posteriormente, se aceptaron como verificadas aquellas interacciones que estuvieran presentes en ambos métodos, alcanzando un total de 12261 posibles PPIs. Adicionalmente, se descubrió que los efectores RipG1 hasta RipG7 comparten sus interacciones, permitiendo deducir que los T3E, cuya familia o función sea similar, pueden interactuar con las mismas proteínas. Finalmente, se realizó un análisis de ontología de genes para conocer las funciones que desempeñan las proteínas de tomate interactuantes. Estos resultados probaron que, la mayoría de T3E interactúan con proteínas que se interrelacionan con sitios específicos de otras moléculas, que actúan como catalizadores o llevan a cabo procesos celulares.

Palabras clave: PPIs, *Ralstonia solanacearum*, *Solanum lycopersicum*, T3E, Interolog, Dominio, GMI1000.

Abstract

Ralstonia solanacearum is a plant pathogenic bacterium known for its lethality worldwide, whose host par excellence is *Solanum lycopersicum*, commonly called tomato; a crop native to South America that generates high annual revenues for the agricultural sector. *R. solanacearum* enters the plant through the roots and initiates the pathogenic process by activating the secretion of specialized proteins called type III effectors (T3E). Once the T3E of *R. solanacearum* performs its specific activity within plant cells by interacting with plant proteins, this pathogen generates wilting symptoms and subsequently the plant death. Therefore, this pathogen's infection means a significant economic risk, causing losses of up to 50% in the annual production of tomato. The present work focuses on predicting protein-protein interactions (PPIs) between the T3E of *R. solanacearum* GMI1000, a strain that typically infects tomato, and tomato proteins. The pathogenic process consists mainly of successful protein interactions, so the study of PPIs allows us to deduce the functions performed by proteins, their possible complexes, and interaction networks. They also contribute to the understanding of the pathogenicity of *R. solanacearum*. In this work, two *in silico* approaches were used, the Interolog method and the Domain-based method. The results were 21557 and 13615 PPIs for the first and second approaches, respectively. Subsequently, those interactions that were present in both methods were accepted as verified, reaching a total of 12261 possible PPIs. Additionally, it was discovered that RipG1 to RipG7 effectors share their interactions, allowing us to deduce that T3E, whose family or function is similar, can interact with the same plant proteins. Finally, a gene ontology analysis was carried out to know the functions performed by the interacting tomato proteins. These results proved that most T3E interact with proteins that interrelate with other molecules' specific sites, which act as catalysts or carry out cellular processes.

Keywords: PPIs, *Ralstonia solanacearum*, *Solanum lycopersicum*, T3E, Interolog, Domain, GMI1000.

Contents

List of Figures

List of Tables

List of Acronyms

RSSC	*Ralstonia solanacearum* Species Complex
EPS	Extracellular Polysaccharides
T3SS	Type III Secretion System
RLS	Root Lateral Structures
T3E	Type III Effectors
PPI	Protein-Protein Interaction
PHI	Pathogen-Host Interaction
Y2H	Yeast Two Hybrid
cytoY2H	Cytosolic Yeast Two-Hybrid System
MS	Mass Spectrometry
RF	Random Forest
SVM	Support Vector Machine
NCBI	National Center for Biotechnology Information
DIP	Database of Interacting Proteins
HMM	Hidden Markov Models
3did	Three-Dimensional Interacting Domains
GO	Gene Ontology
BP	Biological Process
MF	Molecular Function

CC Cellular Component

1 Introduction

Ralstonia solanacearum is one of the most dangerous bacterial plant pathogens worldwide. This soil-borne Gram-negative bacterium causes wilt disease in plants. Its cycle begins when it enters through its plant host secondary roots or damaged tissue and then takes control of the xylem vessels, producing plants to wilt in the initial stages, ultimately causing death [1]. Around 200 families of plants are affected by this pathogen, including prominent crop families like tomatoes, potatoes, bananas, and plantains [2].

Solanum lycopersicum, also known as tomato, is one of the most representative crops in Latin America and worldwide. Native to South America, this vegetable was domesticated since ancient times until it became one of the main economic pillars of agriculture. Additionally, it has biotic and abiotic resistance, which facilitated its spread to other continents [3]. However, it is one of the principal targets of pathogens like *R. solanacearum*, which causes concern since tomato crops' death means significant economic losses [4].

Due to its high economic importance, many studies have been carried out to comprehend and avoid *R. solanacearum* infection on tomato. However, these studies require complementary information to help improve the understanding of these organisms' host-pathogen relationships. Currently, the community has focused on the analysis and prediction of protein-protein interactions. These interactions offer the opportunity of knowing the possible proteins involved in the pathogenicity of *R. solanacearum* [5].

To predict these types of interactions, some laboratory techniques have been used for years, providing outstanding results. Nevertheless, more rapid and effective methods are required, capable of handling large amounts of information and in many occasions that do not require a physical space to take place. For this reason, computational methods, called *in silico*, are practiced more frequently today. Among the most used methods are

those based on homology and domains of proteins [6]. These methods are of utmost importance since discovering the protein network or pathogen interactome could prevent their appearance in crops of great concern.

1.1 Problem Statement

Because *S. lycopersicum* is one of the essential foods in agriculture and represents million-dollar profits annually, it is crucial to identify the pathogens that can damage it to avoid significant economic losses [3]. For years, *R. solanacearum* GMI1000 has attacked tomato crops, allowing this strain to perfectly know this plant's genome, resulting in a high infection rate. Thus, it is fundamental to know how the pathogen-host interaction of these species works [7].

Due to the high incidence of wilting infection in tomatoes, experimental studies have been carried out to understand the behavior of this pathogen in this crop. Further, the role played by the genetic resistance of the tomato against *R. solanacearum* has been investigated, finding that very few varieties show slight resistance. Therefore, most commercial varieties of tomato are susceptible to this pathogen. Additionally, genetic engineering improvements have been made to expand tomato cultivars' productivity and their response to lethal pathogens [8].

Despite these studies and the notable improvements in stress tolerance that tomato crops have undergone, along with chitosan use to improve resistance to pathogens [8,9], there is still insufficient information about how the pathogenic mechanism works and the role of the *R. solanacearum* effector proteins in virulence and tomato plants. Therefore, it is necessary to find alternative methods that provide information from a different perspective that can help dissect the pathogenic mechanism at the molecular level.

The *in silico* analysis and prediction of protein-protein interactions have made it possible for some years to achieve significant progress in the study of pathogens and hosts, obtaining excellent results in species affected by *R. solanacearum*, such as *Arabidopsis thaliana* [10]. This work will provide information on protein-protein interactions between the *R. solanacearum* GMI1000 effector proteins and *S. lycopersicum* genome's proteins. Moreover, using the tomato best-described strain of *R. solanacearum* allows us to analyze the natural system of interaction between the two

species. It also creates an opportunity for upcoming studies about this bacterium's pathogenicity and the roles that tomato proteins fulfill in these interactions.

1.2 Objectives

To implement *in silico* approaches for the prediction of protein-protein interactions between *Ralstonia solanacearum* GMI1000 and *Solanum lycopersicum*, the following objectives have been established in this work.

1.2.1 General Objective

To infer protein-protein interactions between the Type Three Effector proteins of *Ralstonia solanacearum* GMI1000 and the *Solanum lycopersicum* genome's proteins using *in silico* methods.

1.2.2 Specific Objectives

- To apply and compare the results of two *in silico* methods in predicting protein-protein interactions.
- To find interactions present in both methods to guarantee the robustness of the results obtained.
- To show confirmed interactions and analyze possible reasons for such interactions.

2 Literature Review

2.1 Introduction

This chapter introduces fundamental concepts for understanding this research. Firstly, an introduction to plant bacterial pathogens is provided. Then, the bacteria *R. solanacearum* is discussed in detail. Besides, its most relevant strain and its type III effectors are mentioned. Further, the main characteristics of *S. lycopersicum*, crop of high relevance in the agricultural industry, are mentioned. Also, a summary of protein-protein interactions is presented, accompanied by a concise description of the most outstanding laboratory and *in silico* methods used. Besides, a brief description of relevant databases utilized for computational studies is provided. Finally, a concise presentation of the GO terms is presented.

2.2 Plant Bacterial Pathogens

Some plants have large deposits of nutrients in their internal structure; for this reason, they are targets of bacteria that can enter the apoplast through structures such as stomata due to their convenient size. Many of these bacteria attack root tissues modifying natural systems in their developmental stage, helping to deceive and infect the plant. Within Gram-negative bacteria, the ability to poison the plant's apoplasts' interior has been demonstrated with high efficiency [11].

Additionally, it is proven that bacterial pathogens can induce programmed cell death or apoptosis to combat a target plant's defense system response [12]. To carry out this, they have virulence factors, which interact with components that cause the host's cell death, and to destroy or manipulate their cells. These virulence factors can be hormones, enzymes, or toxins. Many pathogens use a type III protein secretion system (T3SS) to cancel the plant's defense mechanism [13].

Even though plants have developed specific and general defense responses to counteract pathogen attacks, some plant species are more likely than others to become infected, proving that pathogens have evolved to infect specific locations and species. Likewise, they can interfere in the regulation of transcription factors whose function is to control cell conservation. However, this depends on the plant's defense system and how easy it is to damage it [12].

Consequently, in a study performed by Leonard, and collaborators [14] it was concluded that the percentage of success with which plant pathogens enter the host to colonize it and carry out a state of early and late virulence, depends on the environmental conditions and the stimuli that the plant provides. These stimuli can be variation in pH, oxidative stress, and defense signals. Equally important, not all bacterial pathogens have the same infection rate, some are more harmful and lethal. Among the most discussed are: *Pseudomonas syringae, Ralstonia solanacearum* (one of the most lethal pathogenic bacteria), *Agrobacterium tumefaciens, Xanthomonas oryzae, Xanthomonas campestris, Xanthomonas axonopodis*, among others [15].

2.3 *Ralstonia solanacearum*

R. solanacearum is one of the most recognized bacterial plant pathogens worldwide. This β-proteobacterium is soil-borne and causes wilt disease, whose main effect is to end the plant's life [14]. Furthermore, it is endemic to tropical and subtropical regions, although some strains are thought to have adapted to temperate climates to attack specific hosts [16]. The *R. solanacearum* species complex (RSSC) is taxonomically divided into three species that contain four phylotypes: *R. pseudosolanacearum* (phylotypes I and III), *R. solanacearum* (phylotype II) and *R. syzygii* (phylotype IV) [17], these phylotypes are evolutionarily different lineages [18]. Likewise, they vary significantly in their hosts, origin in terms of geography, and their pathogenicity [15].

Phylotype I encompasses species originated mostly from Asia; meanwhile, phylotype II includes species belonging to the American continent; phylotype III has species from Africa, and phylotype IV is known for its diversity of species from Australia, Japan, and Indonesia [19]. The host range that attacks this bacterium is extensive since it encompasses around 50 families of plants, which include the potato, tomato, tobacco and eggplant representing nightshades, some legumes such as peanuts, certain

monocotyledons are also affected especially bananas, some trees, and shrubs such as eucalyptus and cassava [16].

Its attack method begins in the soil where it manages to penetrate the xylem vessels through the roots. Immediately the flow of water will drive the bacteria from the roots to the plant's shoots. Once inside the xylem vessels, *R. solanacearum* generates large amounts of extracellular polysaccharides (EPS), creating a vascular obstruction preventing water passage through the entire system causing wilt [7]; Figure 1 illustrates this process. Additionally, the bacterium uses its type III secretion system (T3SS) to favor infection by hijacking the host plant's cellular mechanism, guaranteeing a high degree of virulence [14,20].

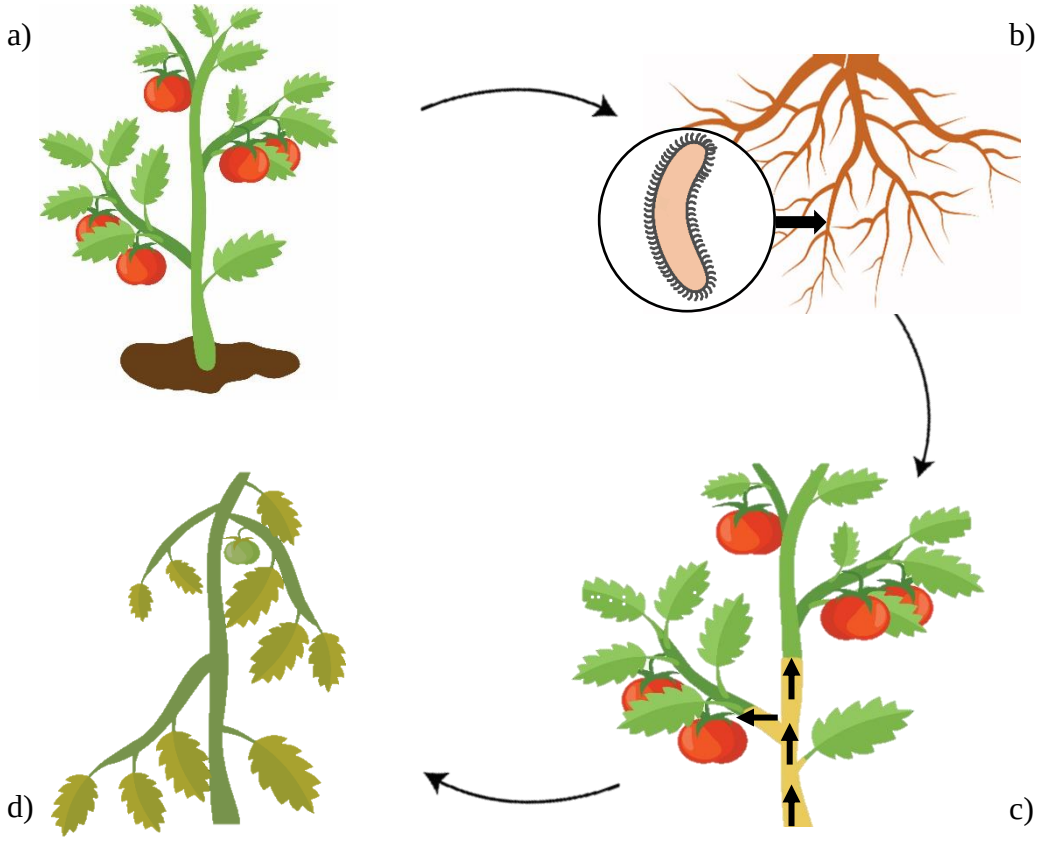


Figure 1. *Ralstonia solanacearum* pathogenic cycle in tomato plant. a) Developed tomato plant, b) Attachment of *R. solanacearum* to the root tissues of the plant, c) Spreading of the bacteria through the xylem vessels and beginning of wilt, d) Plant wilt and death.

Due to its high pathogenicity, lethality, vast geographical distribution, and variety of target hosts, this bacterium has a tremendous economic impact since it generates

significant losses worldwide. Developing countries and those from which *R. solanacearum* is endemic are the most affected due to damages in their cultivars since agriculture is their main economic activity. Another significant disadvantage for the agricultural sector is that this bacterium is hardly eradicated because it can be present for years in soil and water. Likewise, it can use some hosts as reservoirs, causing them to be asymptomatic [15]. The main symptoms that this bacterium generates can be sudden wilt in the whole plant, some cross-sections can show a high presence of bacterial exudates inside the stem, banana or potato plantations, and visible effects in fruits or tubers [19].

2.4 GMI1000 Strain

GMI1000 strain belongs to the phylotype I of *R. solanacearum*, and its pathogenic effects are widely investigated. It was sampled from tomato in French Guyana years ago [21] and it is considered a study model for two main reasons: first, it was the first strain to be sequenced entirely, and second, because 71 of the 102 *R. solanacearum's* effector proteins secreted by the T3SS are present in this strain. Additionally, GMI1000 is thought to be the cause of wilt in several other solanaceous plants [22]. A notable feature of this strain is its ability to induce the growth of root lateral structures (RLS) in the infected plant; these new structures are perfect sites for colonization and multiplication of bacteria. Although RLS are not essential to invade the host vascular system, they are crucial in the rhizosphere-related phases within the *R. solanacearum* life cycle [23].

2.5 Type III Effectors

Although the infection process of *R. solanacearum* is not fully understood, type III effectors (T3E) are critical to the pathogenicity of many bacteria, including *R. solanacearum*. These effectors are proteins secreted by the type III secretion system (T3SS) [24] and are thought to play a significant role in the host's interactions and pathogenesis [25]. The T3SS also fulfills the function of delivering the effectors into the cytosol of plant cells where they start their pathogenic activity [26,27]. This activity consists of creating an adequate environment within the host plant so that the bacteria can colonize it; for this, the effectors must switch off the plant's immune system and change both its metabolism and physiology [28]. Among the bacterial pathogenicity, studying the repertoire of the T3E, either if they are present in a specific strain or diverse species,

is of utmost importance. *R. solanacearum* GMI1000 is one of the most studied tomato's pathogens [29]. It was discovered that around 45% of its confirmed effector proteins could be found in other bacterial pathogens that attack plants or animals, the reason why they are considered ancient and conserved between species [25]. Table 1 shows the T3E of *R. solanacearum* present in the GMI1000 strain, their classification according to eleven families, and if available, each one counts with a description. Additionally, it is detailed if there is one copy (OK) or several copies (MULTI) of the effector in the genome of GMI1000.

Table 1. T3E of *R. solanacearum* present in the GMI1000 strain [30][25]

Family	T3E Family Name	Description/alternate name	GMI1000
1	RipA1	AWR1	OK
	RipA2	AWR2	OK
	RipA3	AWR3	OK
	RipA4	AWR4	OK
	RipA5	AWR5	OK
	RipB	Inosine-uridine nucleoside N-ribohydrolase	OK
	RipC1	HAD-like phosphatase	OK
	RipD	AvrPphD	OK
	RipE1	AvrPphE	OK
2	RipF1	(PopF1) T3SS translocator	MULTI
	RipG1	F-box LRR protein GALA1	OK
	RipG2	F-box LRR protein GALA2	OK
	RipG3	F-box LRR protein GALA3	OK
	RipG4	F-box LRR protein GALA4	OK
	RipG5	F-box LRR protein GALA5	OK
	RipG6	F-box LRR protein GALA6	OK
	RipG7	F-box LRR protein GALA7	OK
3	RipH1	HLK1	OK
	RipH2	HLK2	OK
	RipH3	HLK3	OK
	RipI		OK
	RipJ	Putative acetyltransferase	OK
	RipL	Pentatricopeptide Repeats	OK
	RipM		OK
4	RipN	Nudix hydrolase	OK
	RipO1	HopG1	OK
	RipP1	(PopP1), putative acetyltransferase	OK
	RipP2	(PopP2), Acetyltransferase	OK
	RipQ	HopAA1	OK
	RipR	HopR1	OK
	RipS1	SKWP1	OK

	RipS2	SKWP2	OK
5	RipS3	SKWP3	OK
	RipS4	SKWP4	OK
	RipS5	SKWP5	OK
	RipS6	SKWP6	OK
	RipT	Cysteine protease	OK
	RipU		OK
	RipV1	Ubiquitin ligase domain	OK
6	RipW	Harpin with pectate lyase domain	OK
	RipX	(PopA), Harpin	OK
	RipY	Ankyrin Repeats	OK
	RipZ		OK
	RipAA	(AvrA)	OK
	RipAB	(PopB), NLS harboring protein	OK
	RipAC	(PopC), LRR domain	OK
	RipAD		OK
	RipAE	Putative acetyltransferase	OK
	RipAF1	Putative ADP-ribosyltransferase	OK
7	RipAG		OK
	RipAH		OK
	RipAI		OK
	RipAJ		OK
	RipAK		OK
	RipAM		OK
	RipAN		OK
	RipAO		OK
8	RipAQ		OK
	RipAR	Ubiquitin ligase domain	OK
	RipAS		OK
	RipAU		OK
	RipAV	Coiled-coil domain	OK
	RipAW	Ubiquitin ligase domain	OK
	RipAX1	HopH1	OK
	RipAX2	HopH1	OK
9	RipAY		OK
	RipAZ1		OK
10	RipBJ		OK
	RipBO	[FORMER Hyp16]	OK
11	RipTAL	Transcription activator-like protein	OK
	RipTPS	Trehalose-phosphate synthase	OK

2.6 *Solanum lycopersicum*

The domesticated tomato called *Solanum lycopersicum* belongs to a large and diverse family, *Solanaceae*, which contains more than 3,000 species [31], but it also belongs to the *Lycopersicon* clade [32]. Although it is native to the Andean region, it was imported to Europe in the 16th century, and after this, its distribution was possible in many habitats worldwide [8]. Additionally, the environment in which develops ranges from places that are at sea level, high altitudes (3300 m), or places with very rainy or arid climates. Thus, the tomato has become one of the most common food worldwide and is situated among the first places of production, reaching more than 100 tons per year, generating income of more than $ 1.6 billion globally [32]. Its consumption rate is very high since its preventive properties are attributed to cardiovascular diseases and cancer and its delicious taste [33]. Therefore, demonstrating the importance of this food in agriculture and the economy.

For the reasons mentioned above, the tomato is considered a great study model that has been analyzed for years. It has characteristics such as sympodial shoots, compound leaves, morphology, and resistance that protect it from diseases, besides a tremendous phenotypic variation generated over time, making it perfect for studying its genetics [3]. Also, characteristics such as being a simple diploid, having twelve highly differentiated chromosomes, a genome with both molecular and conventional markers, and a structure that allows a high number of mutations, make the tomato an ideal plant model [34].

In addition, it is of great concern that *S. lycopersicum* is one of the main targets of *R. solanacearum,* a pathogen known for affecting the production of solanaceous crops in regions whose climate is temperate, tropical or subtropical, causing bacterial wilt [35,36] and losses of up to 50% of annual tomato production [37]. This pathogen's lethality is more exhibited in young tomato plants, which tend to die quickly, while older plants show signs of wilting on their younger leaves until the entire plant eventually dies [38]. However, one of the solutions proposed to stop the infection of *R. solanacearum* is to genetically modify tomato crops, allowing them to be resistant to the attacks of this deadly pathogen. While most tomato varieties are susceptible to *R. solanacearum,* there is one variety, Hawaii 7996, that shows natural resistance and could be used as a base to develop possible defense mechanisms against this pathogenic bacteria [39].

2.7 Protein-Protein Interactions

Protein-protein interactions (PPIs) are related to essentially biological processes [40], which, when identified, are of great help in determining the cellular functions the PPIs perform [41]. By interacting with each other, the proteins form complexes, demonstrating that in PPIs, there is no random contact between two or more proteins. On the contrary, these interactions are regulated by cell states, signals, or stimuli, proving that they depend on various factors and the presence of other proteins called "interactors" [42]. PPIs play varied roles, and as a way to carry them out, they interact with each other forming a network or interactome [43,44]. Therefore, knowing the possible interactions between some proteins can give us a general idea of the network they belong to [45].

Additionally, PPIs can modify certain enzymes' interactions with their substrates, the specificity that interacting proteins have for their substrates, and inactivate proteins [46,47]. Studies have shown that interacting protein often generates similar diseases after suffering a mutation, which presumes that these interactions could be used to predict genes that produce diseases [48]. Likewise, many studies have focused on analyzing protein-protein interactions of a single organism (intra-species PPI prediction). However, nowadays, the main objective is to study interactions between different organisms (inter-species PPI prediction) being the pathogen-host interaction (PHI) the most striking for investigating. PHI studies are of utmost importance since achieving a better understanding of this mechanism could develop therapeutic or preventive techniques [49]. Over time, very prominent methods have been used to identify protein interactions; these methods can be performed in the laboratory or using computational tools, that is, *in silico*.

One of the most widely used laboratory methods is the Yeast Two-Hybrid (Y2H) Assay, which is based on the concept of site-specific transcriptional activators, wherein two proteins that are analyzed to verify a possible interaction are expressed as "hybrids" in the yeast [50]. The principle of this technique is that the interaction of two proteins must unite the activation domains. Thus, one of the proteins must bind to a DNA-binding domain while the other protein binds to a transcriptional activator domain, in this way the DNA-binding domain acts as an identifier of the activator when searching for genes that will be expressed and the activator domain detects proteins from the transcriptional tools that will give way to transcription. If the two domain-carrying proteins interact, they not only create an effective activator but also demonstrate that there is a relationship between the two [51]. Moreover, this technique is used for finding PPIs between membrane

proteins, but for the analysis of cytosolic proteins, an alternative method called Cytosolic Yeast Two-Hybrid System (cytoY2H) can be utilized. This method is based on the Split Ubiquitin technique and can be applied to proteins that are difficult to study using Y2H [52].

Additionally, the Mass Spectrometry (MS) of Purified Complexes technique whose use was and still is fundamental, is based on tagged individual proteins that serve as a "hook" for the purification of protein complexes biochemically and then through the use of mass spectrometry the new proteins are identified [53]. Similarly, genetic interactions have been used to recognize possibly related proteins; for this, it is necessary to locate two lethal genes created after undergoing a mutation, resulting in a lethal synthetic interaction. Consequently, these genes are assumed to encode proteins that could interact with each other [53,54]. Even though laboratory practices provide a significant contribution to the study of protein interactions, in recent years, *in silico* techniques have gained strength because they offer a complementary point of view to already developed methods. In the following section, some of the most important computational methods within the study of PPIs are mentioned.

2.8 *In silico* Approaches

Conventional laboratory techniques for determining PPIs can face difficulties like having a high probability of error, a high cost of experimentation, or not applying to all types of organisms. For this reason, computational or *in silico* methods are currently used, allowing more work to be done efficiently, fast, and on a larger scale [6]. *In silico* techniques are usually based on available information of known protein sequences, structures, and interactions, using techniques such as machine learning, where existing data is required for training. Also, methods such as transfer learning are applied; here, it is necessary to use complex neural networks to predict interactions [49].

However, other techniques focus on improving biological characteristics like homology besides the structure or function of proteins. However, both approaches' complementary use has recently been proposed, thus achieving more robust and accurate results [55]. Among the main approaches used is machine learning, which utilizes available PPI data to train and classify possible interactions and non-interactions of protein pairs. Classification algorithms such as Random Forest (RF) and Support Vector

Machine (SVM) are used in this technique, both of which have shown excellent results in significant and challenging data processing studies. However, this method has only shown positive results when using pathogen systems whose research and understanding are complete, leaving aside certain systems that lack information [49].

Then, the structure-based method consists of a set of pathogenic and host proteins that serve to identify similarities between pathogen and host proteins based on known structure or interactions. Sequence matching procedures are used for this method, and its most considerable disadvantage is that on certain occasions finding similarities between pathogen and host proteins is not assured for all pathogens [49]. Other highly acclaimed approaches are those based on homology and domains. Regarding homology, this method's rationale is to find conserved interactions between a pair of proteins that have interacting homologs in other species. For its part, the domain technique is based on the mediator's role that a domain fulfills in the interactions of a protein. In this way, it is possible to identify the possible PPIs of an organism through its domains. Both methods have shown excellent results when used in multiple studies to discover PHIs [49].

Sections 3.4 and 3.5 (see below) provide a complete description of these two approaches, and for this reason, this section does not provide detailed concepts. Finally, to apply the methodologies mentioned here and others available, it is necessary to collect information from our organism of interest. For this, some databases have a large number of verified interactions, experimentally proven or computationally predicted. These databases usually belong to expert organizations, and their information can be reliably used in a study [56]. Some of the most widely used databases are presented in Table 2 below.

Table 2. Widely used databases for Protein-Protein Interactions search.

Database Abbreviation	Database Objective	Publication Year	URL
KEGG [57]	Understanding the high-level functions of the biological system	2000	https://www.genome.jp/kegg/
DIP [58]	Experimentally verified PPIs	2002	https://dip.doe-mbi.ucla.edu/dip/Main.cgi
MIPS [59]	Compilation of curated PPI data	2002	http://mips.gsf.de
PDBsum [60]	3D structures collect in the Protein Data Bank (PDB)	2005	www.ebi.ac.uk/pdbsum/

iRefIndex [61]	Consolidated index to search for interactions and redundant PPIs.	2008	https://irefindex.vib.be//
HPRD [62]	Human PPIs	2009	http://www.hprd.org/
ProPrInt [63]	Predicts physical or functional interactions	2010	http://crdd.osdd.net/raghava/proprint/index.html
IntAct [64]	Molecular interactions	2012	https://www.ebi.ac.uk/intact/
MINT [65]	Experimentally verified PPIs	2012	https://mint.bio.uniroma2.it/
BioGrid [66]	Physical or genetic interactions of model organisms and humans.	2017	https://thebiogrid.org/
STRING [67]	Analysis of protein-protein interaction networks	2017	https://string-db.org/cgi/input.pl

2.9 Gene Ontology

Gene ontology (GO) is a well-known knowledgebase that provides information about the function of genes and their products. Its main objective is to provide generalized and organized information that contains the appropriate vocabulary to describe the roles played by an organism's genes and thus be understood and known in a standard way by the entire scientific community [68]. The information is structured in such a way that it is understandable and robust, so it can be used in computational studies, which are very useful for modern biology analyzes. This knowledgebase classifies gene functions (GO terms) into three main categories: Biological Process, Molecular Function, and Cellular Components. However, these categories are divided into subcategories that also have their division and could be extended at different classification levels depending on the multiple functions that the same gene can fulfill [69].

The first main category is the Biological Process, which refers to the participation of a gene or its product to meet a biological objective. Also, chemical or physical transformations can be used to accomplish a process and obtain a purposed result. The second category is Molecular Function, which encompasses the biochemical activities about a gene's product without mentioning the specific place where they occur. Finally, the Cellular Components of a gene referred to the place in the cell where a gene's product is activated. However, this last category is indicated to eukaryotic cells and may not be available to all organisms, which may be the case to other subcategories. Additionally, an important point that should be highlighted is that these GO terms facilitate the understanding that a gene or its protein can perform various functions and processes,

whether of a cellular or molecular type, demonstrating that interactions between different proteins can be carried out in different places of a cell [68]. Figure 2 shows the GO chart of a *S. lycopersicum* protein, where it can be appreciated that it has a molecular function as its central category, followed by its subcategories.

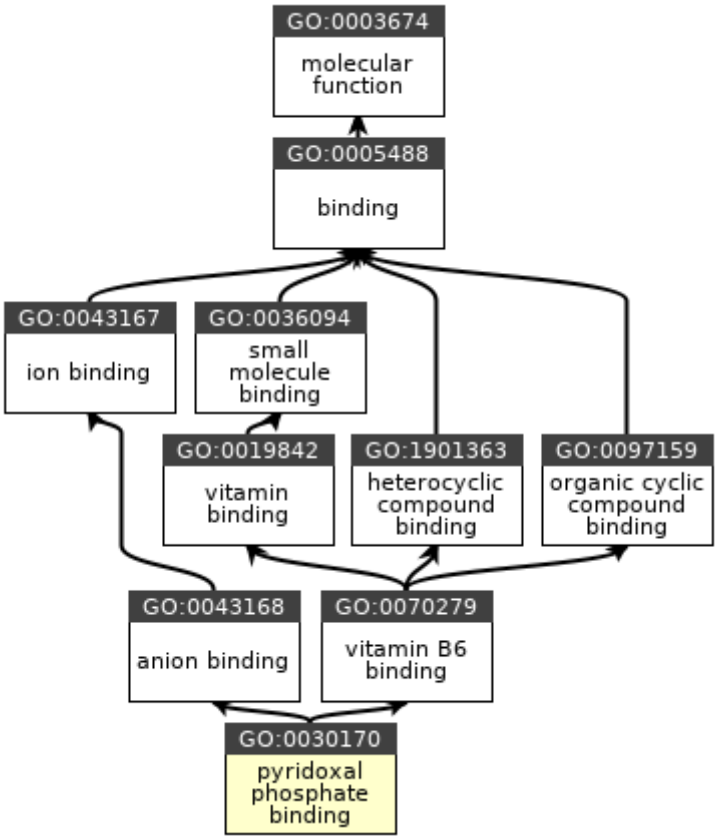


Figure 2. GO chart of the protein gamma aminobutyrate transaminase 3, chloroplastic [*Solanum lycopersicum*]

2.10 Summary

This chapter briefly introduced the primary points about plant bacterial pathogens. Besides, the main characteristics of one of the most lethal pathogenic bacteria around the world, *R. solanacearum*, were described in great detail. Also, its principal targets and the differences between its strains were briefly explained. Likewise, the most investigated strain of *R. solanacearum*, GMI1000, was discussed concisely. This strain contains the majority of the type III effectors, making it ideal for both *in vivo* and *in silico* studies. Consequently, the necessary information about these type III effectors was provided because they are essential for the bacterium's pathogenesis. Further, information was

brought on one of the primary plants affected by this bacterium, *S. lycopersicum*. Additionally, an introduction to protein-protein interactions was presented along with the most used methods to identify them and many useful databases currently available. Finally, the objective and most significant categories of gene ontology were mentioned.

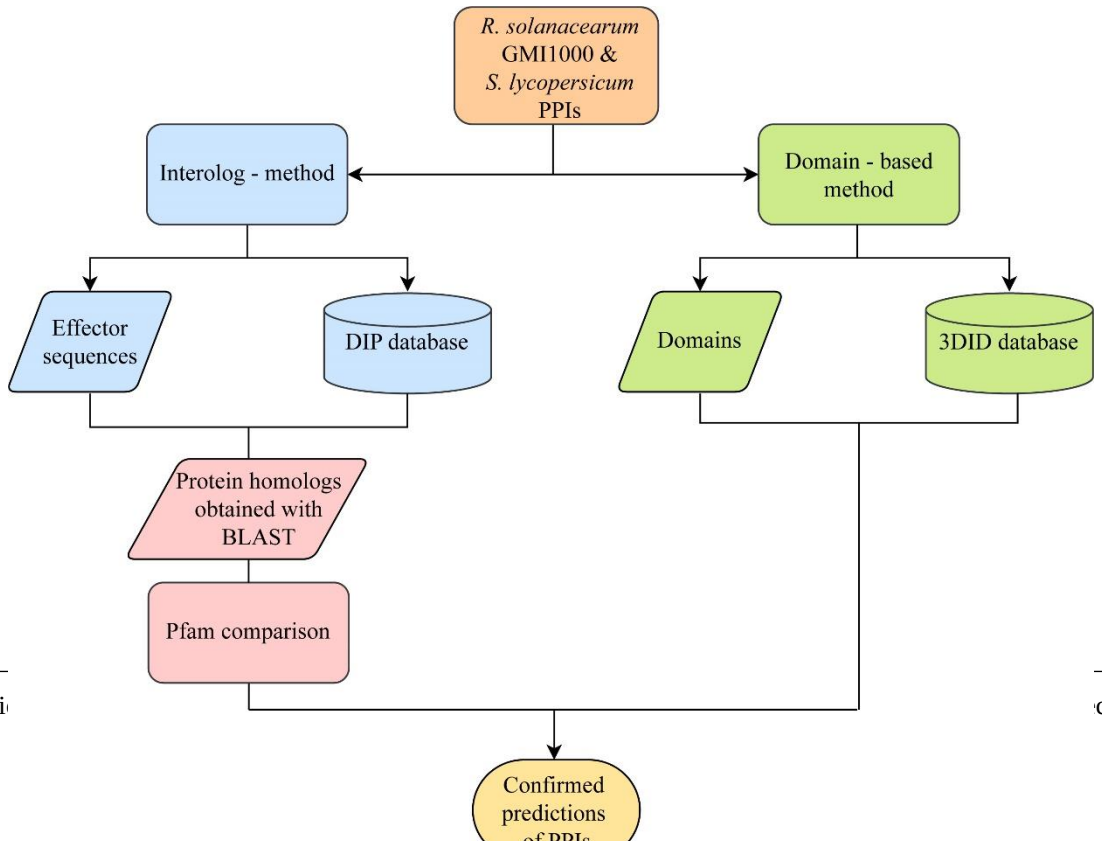
3 Methodology

3.1 Introduction

This chapter details the process followed to predict possible protein-protein interactions between the T3E of *R. solanacearum* GMI1000 and proteins of *S. lycopersicum* genome. First, there is a description of the steps for obtaining the sequences of both organisms. It is then explained in depth about the *in silico* techniques and databases used here to obtain and corroborate the possible interactions. Besides, information is provided regarding the process for the gene ontology analysis.

3.2 General Workflow of the Process

Figure 3 shows an outline of the general process followed in this work; each part is detailed in the sections described below.



XXXXXX

Figure 3. Workflow of the process.

3.3 *Ralstonia solanacearum* GMI1000 Type III Effectors and *Solanum lycopersicum* Sequences

The sequences of 71 *R. solanacearum* type three effectors (T3E) present in GMI1000 strain, were obtained to start this study. The sequences of these effectors were recovered from https://iant.toulouse.inra.fr/bacteria/annotation/site/prj/T3Ev3/. This page, called RALSTO T3E, belongs to the Laboratory of Plant-Microbe Interactions (France) and offers updated and recognized information about the different strains of this bacterium and the T3E present in them [30]. Regarding the protein sequences of *S. lycopersicum*, the whole genomic set of proteins was obtained from https://www.ncbi.nlm.nih.gov/genome/?term=Solanum+lycopersicum, the National Center for Biotechnology Information (NCBI) in FASTA format. NCBI is continuously updating its genomic and biomedical information, with many species and easily accessible formats [70]. Sections 3.4 and 3.5 detail the processes in which these sequences were used.

3.4 Interolog Method

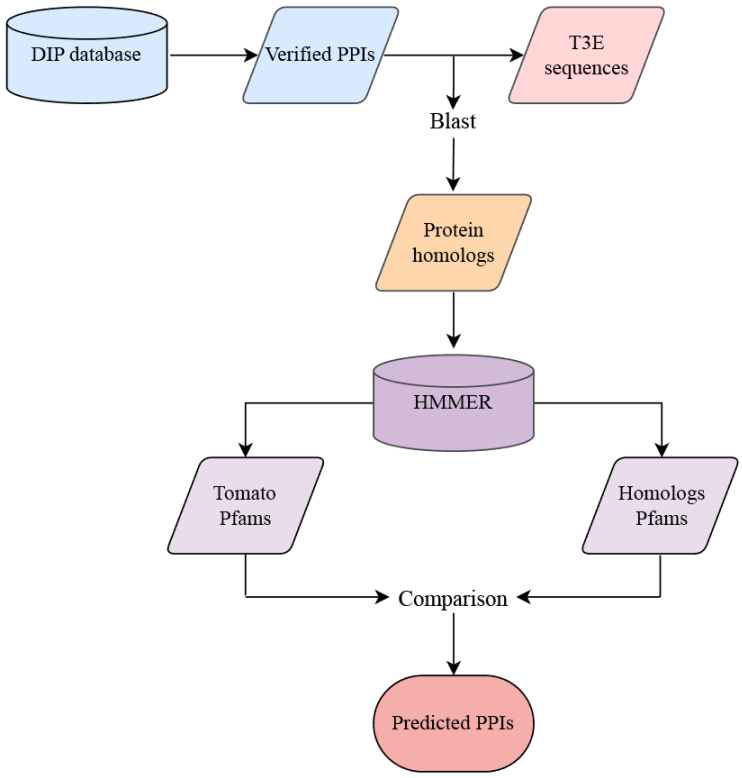


Figure 4. Interolog method flowchart.

This approach is based on the rationale that proteins with interacting homologs in other species must have conserved interactions over time; these interactions are called "interologs" [71]. In other words, there are protein pairs that could interact in various organisms [72]. A series of steps were carried out to utilize this approach for a PPI prediction using the organisms mentioned; these are shown in Figure 4 and explained below. First, experimentally verified PPIs were collected; for this, the entire Database of Interacting Proteins (DIP) was downloaded from https://dip.doe-mbi.ucla.edu/dip/Download.cgi. The data that DIP offers are verified empirically and by experts, although on certain occasions, computational approaches are used to complement the information [73]. Consequently, each T3E sequence was blasted against the DIP database; thus, all the results that met an e-value ≤ 0.001, which is statistically reliable, were accepted as homologs of some effectors.

Then, from these homologous protein sequences, the accession number Pfam was found. Pfam is a website and database that groups proteins in families and domains, and since a protein family is descended from a common ancestor, that is, they are homologous, the proteins that belong to it can have similar functions and sequences. Additionally, this website utilizes search criteria based on the Hidden Markov Models (HMM) and their

alignments [74,75]. The probabilistic models known as HMMs profiles are in charge of obtaining statistical homology inferences [76]. Therefore, the Pfam page https://pfm.xfam.org/ and the HMMER software http://hmmer.org/ were used to obtain the Pfams numbers. Once the Pfams were collected, they were compared with the Pfams of the tomato proteins, obtained from the HMMER page. Finally, if there was a homologous protein that shared the same Pfam accession number as any tomato protein, it was concluded that there was an interaction between the T3E it represented and the *S. lycopersicum* protein.

3.5 Domain-Based Method

Briefly, protein domains are considered the basis of protein-protein interactions, and for this reason, this approach is one of the most extensively used [77]. Unlike other existing methods, this approach, when used to predict PPIs, utilizes the information provided by protein domains as the only source [78], suggesting that if a pair of proteins has a pair of interacting domains, they may interact with each other [49]. This approach was used to verify the results obtained above, as shown in Figure 5 and described below.

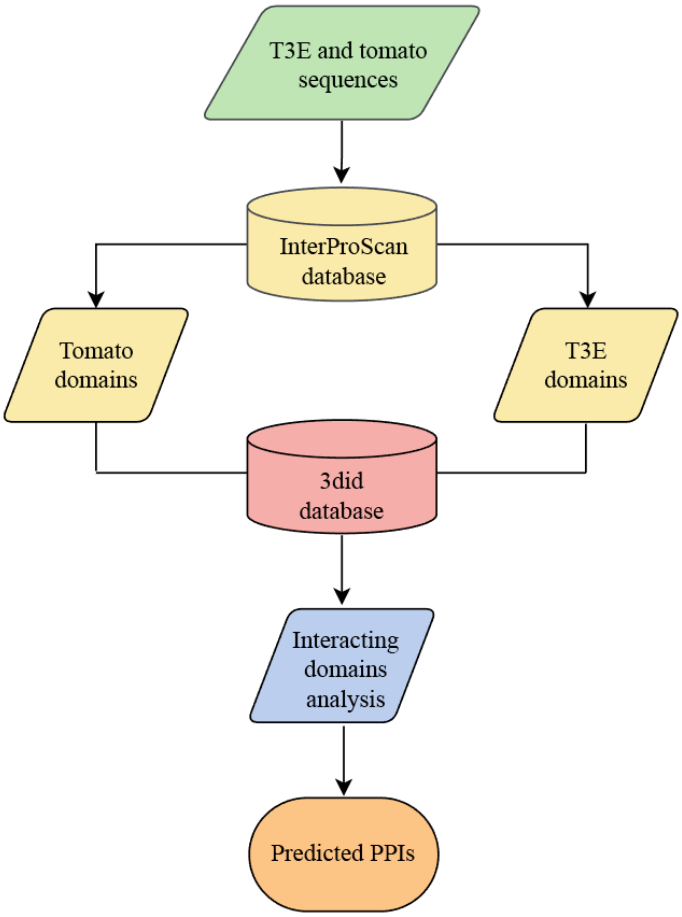


Figure 5. Domain method flowchart.

First, it was required to obtain the domains of the protein sequences of *S. lycopersicum* and the T3E. For this, it was necessary to download the InterProScan database from http://www.ebi.ac.uk/interpro/download/. InterProScan is a free database that compiles information from the InterPro consortium, which provides data and analysis of protein and DNA sequences for educational use [79]. After obtaining the domains, we searched the 3did database https://3did.irbbarcelona.org/. The database of three-dimensional interacting domains, or 3did, contains templates for domain-domain interactions and peptide-domain interactions. It has a simple search engine through which the name of a domain and the Pfam accession number of a protein, can be entered [80]. This database verified that the T3E and tomato domains interact with each other, confirming that there are PPIs between both organisms.

After obtaining the results of this method, they were compared with the first method; this comparison aimed to corroborate the PPIs found. If a PPI was present in both approaches, it was considered as a verified PPI.

3.6 Gene Ontology Classification

To present the results obtained from a different point of analysis, GO terms were used to classify the protein-protein interactions collected according to the T3E to which they belong and the tomato's functions proteins with which they interact. For this, the main GO terms to which the tomato proteins belonged were achieved through the InterProScan database. Then, the subcategories in which said proteins could participate were found through the QuickGO page https://www.ebi.ac.uk/QuickGO/. This page contains all the available information about GO terms and has multiple filters to ensure an exact and straightforward search [81]. It should be mentioned that to simplify the analysis, only the second level subcategories were used in this work. Finally, after complying with the categories and subcategories in which the *S. lycopersicum* proteins participate, they were represented in an organized manner.

3.7 Summary

In this chapter, it is presented the methodology followed to discover possible PPIs between the T3E of *R. solanacearum* GMI1000 and the tomato proteins. For this, two computational approaches were used, the Interolog method and the Domain-Based method. Firstly, the sequences of the T3E and the tomato proteins were downloaded; then to carry out the first method, two main tools were necessary, the DIP database and the Pfam database that works in collaboration with the HMMER software. In this way, PPIs were found based on protein homology. Later, for the second method, the domains of the T3E and the tomato proteins were required. For this, the InterProScan database was used. Thus, these domains were entered in the 3did database, which provided information on the interactions between both organisms' domains. Finally, the confirmed interactions were classified following the distribution of their GO terms.

4 Results

4.1 Introduction

This chapter concisely shows the results obtained during this work. First, the interactions and other results achieved by applying the Interolog method are detailed. Then, there is a description of the results obtained using the Domain-Based method. Further, the PPIs that were verified using both methods are presented. Lastly, a brief description is provided about the results found in the GO analysis of tomato's proteins.

4.2 Interolog Predicted Protein-Protein Interactions

After blasting the T3E sequences used as a query against the DIP database, 491 homologous sequences were obtained; they belonged to 18 effectors of the 71 initially tested (Table 3). The effectors that did not show results were discarded from the analysis since they did not have reliable protein homologs, meaning that they did not comply with the acceptance e-value ≤ 0.001. Subsequently, the Pfam accession numbers of the 490 protein homologs were found (Table 4). One of the effectors, RipE1, was discarded from the study since its homologous sequence did not have a Pfam accession number.

Table 3. Homologous sequences obtained from blasting T3E against DIP database.

Gene ID	T3E Family Name	Number of homologous sequences in DIP database
RSp0875	RipAC	84
RSc0321	RipAE	1
RSp0822	RipAF1	1
RSc3369	RipE1	1
RSp0914	RipG1	62
RSp0672	RipG2	50
RSp0028	RipG3	24
RSc1800	RipG4	50
RSc1801	RipG5	50
RSc1356	RipG6	58
RSc1357	RipG7	11
RSc2132	RipJ	1
RSp0193	RipL	2
RSc0826	RipP1	2

RSc0868	RipP2	2
RSp0731	RipTPS	6
RSc1349	RipV1	1
RSc0257	RipY	85

Table 4. Pfam accession numbers of the homologous sequences of tomato per T3E.

Gene ID	T3E Family Name	Pfams
RSp0875	RipAC	PF00626, PF01582, PF13516, PF00791, PF03106, PF00560, PF00001, PF07714, PF12468, PF00481, PF13676, PF13855, PF08509, PF01462, PF07679, PF00069, PF13306, PF12534, PF18837, PF12661, PF00595, PF12799, PF00211, PF00054, PF03372, PF00931, PF00008, PF00531, PF12369, PF10428, PF07725, PF08263, PF10461, PF12796, PF16095, PF14496, PF01463
RSc0321	RipAE	PF03421
RSp0822	RipAF1	PF09143
RSp0914	RipG1	PF13943, PF01582, PF13516, PF14484, PF00560, PF00001, PF07714, PF13676, PF13855, PF01462, PF07679, PF00069, PF13306, PF18837, PF00646, PF07834, PF17779, PF00619, PF02758, PF17968, PF17888, PF12799, PF16000, PF08263, PF13553, PF05729, PF01463, PF17776
RSp0672	RipG2	PF13943, PF01582, PF13516, PF14484, PF00560, PF00001, PF07714, PF13676, PF00481, PF13855, PF08509, PF01462, PF07679, PF00069, PF13306, PF18837, PF00646, PF07834, PF17779, PF00619, PF02758, PF12799, PF17888, PF00211, PF16000, PF08263, PF13553, PF05729, PF17776
RSp0028	RipG3	PF13943, PF01582, PF02758, PF13516, PF01462, PF14484, PF00069, PF08263, PF13553, PF05729, PF17776, PF00560, PF07714, PF00646, PF17779, PF13676, PF13855, PF00619
RSc1800	RipG4	PF13943, PF01582, PF13516, PF14484, PF00560, PF00001, PF07714, PF13676, PF13855, PF01462, PF00069, PF13306, PF18837, PF00646, PF07834, PF17779, PF00619, PF02758, PF17968, PF12799, PF17888, PF16000, PF08263, PF13553, PF05729, PF01463, PF17776
RSc1801	RipG5	PF13943, PF01582, PF13516, PF14484, PF00560, PF00001, PF07714, PF13676, PF13855, PF01462, PF00069, PF18837, PF00646, PF07834, PF17779, PF00619, PF02758, PF17968, PF12799, PF08263, PF05729, PF01463, PF17776

RSc1356	RipG6	PF13943, PF01582, PF13516, PF14484, PF00560, PF00001, PF07714, PF13676, PF13855, PF01462, PF07679, PF00069, PF13306, PF18837, PF00646, PF07834, PF17779, PF00619, PF02758, PF17968, PF08263, PF13553, PF05729, PF18831, PF01463, PF17776
RSc1357	RipG7	PF13516, PF01462, PF00069, PF08263, PF05729, PF17776, PF00560, PF00001, PF00646, PF17779, PF13855, PF00619
RSc2132	RipJ	PF03421
RSp0193	RipL	PF13041, PF13812, PF12854, PF01535
RSc0826	RipP1	PF03421
RSc0868	RipP2	PF03421
RSp0731	RipTPS	PF02358, PF00982
RSc1349	RipV1	PF14496, PF12468
RSc0257	RipY	PF16705, PF16600, PF06479, PF00644, PF16179, PF00373, PF00791, PF00554, PF01363, PF07714, PF07647, PF00023, PF05033, PF01237, PF00069, PF00651, PF14835, PF00169, PF00640, PF15808, PF00018, PF02204, PF00887, PF12075, PF00856, PF00013, PF14604, PF00520, PF00531, PF07653, PF13606, PF17809, PF00536, PF01412, PF12796, PF16553, PF16632, PF07525

Then, a comparison was made between the Pfam numbers recovered from the T3E homologs and the tomato proteins. In this way, 21557 possible PPIs were obtained for 11 T3E, summarized in Table 5.

Table 5. Predicted PPIs between the T3E of *R. solanacearum* GMI1000 and *S. lycopersicum* genome using the interolog method.

Gene ID	T3E Family Name	Interacting tomato proteins	Pfams
RSp0875	RipAC	2470	PF12796, PF00307, PF13306, PF12799, PF00008, PF16095, PF13855, PF03106, PF03098, PF00595, PF13676, PF08263, PF03372, PF00069, PF00560, PF00626, PF13516, PF00931, PF01582, PF00481, PF07714
RSp0914	RipG1	2260	PF05729, PF13943, PF13306, PF12799, PF00008, PF00646, PF13855, PF13676, PF08263, PF00069, PF00560, PF13516, PF01582, PF07714
RSp0672	RipG2	2421	PF05729, PF13943, PF13306, PF12799, PF00646, PF13855, PF13676, PF08263, PF00069, PF12937, PF00560, PF13516, PF01582, PF00481, PF07714

RSp0028	RipG3	2254	PF00560, PF08263, PF00646, PF13516, PF05729, PF13855, PF13943, PF00069, PF01582, PF13676, PF07714
RSc1800	RipG4	2280	PF05729, PF13943, PF13306, PF12799, PF00646, PF13855, PF13676, PF08263, PF00069, PF12937, PF00560, PF13516, PF01582, PF07714
RSc1801	RipG5	2260	PF05729, PF13943, PF13306, PF12799, PF00646, PF13855, PF13676, PF08263, PF00069, PF00560, PF13516, PF01582, PF07714
RSc1356	RipG6	2274	PF05729, PF13943, PF13306, PF00646, PF13855, PF13676, PF08263, PF00069, PF12937, PF00560, PF13516, PF01582, PF07714
RSc1357	RipG7	2202	PF00560, PF08263, PF00646, PF13516, PF05729, PF13855, PF00069
RSp0193	RipL	974	PF17177, PF13812, PF13041, PF01535, PF12854
RSp0731	RipTPS	35	PF02358, PF00982
RSc0257	RipY	2127	PF01237, PF12796, PF00373, PF18346, PF13606, PF00644, PF01412, PF07653, PF00856, PF00169, PF13637, PF00569, PF00023, PF00013, PF00651, PF01529, PF00018, PF07647, PF00069, PF01363, PF02204, PF17820, PF06479, PF00536, PF00887, PF00520, PF05033, PF14604, PF07714

The GMI1000 effector protein that has the most interactions with the tomato genome is RipAC, with 2470 interactions, while the effector that showed the least interactions was RipTPS, with a total of 35 interactions. Some effectors share a certain amount of their interactions with other effectors. RipG1 and RipG5 were shown to be the only pair of T3E to have the same interactions with the tomato genome, but they also share interactions with the RipG3 (Table 6). This similarity of results between the two effector proteins could be because the proteins belong to the second family of T3E, which comprises the effector proteins RipE2, RipF1, RipF2, RipG1 through RipG7. It is worth mentioning that all the RipG1 through RipG7 proteins are found in this approach's results. However, the other T3E proteins shared a minimum number of interactions between them, especially RipTPS and RipL proteins, which show very different interactions with tomato proteins than the other effectors. Likewise, the RipAC and RipY proteins belong to family six, while the RipL and RipTPS proteins that also participate in these interactions belong to families three and eleven, respectively.

Table 6. Comparison of PPIs between the T3E of GMI1000 strain, interolog method.

T3E 1	T3E 2	T3E1 PPIS	Share PPIs	T3E2 PPIs
RipAC	RipG1	670	1800	460
	RipG2	529	1941	480
	RipG3	676	1794	460
	RipG4	670	1800	480
	RipG5	670	1800	460
	RipG6	676	1794	480
	RipG7	727	1743	459
	RipL	2470	0	974
	RipTPS	2470	0	35
RipG1	RipG2	0	2260	161
	RipG3	6	2254	0
	RipG5	0	2260	0
	RipG6	6	2254	20
	RipG7	58	2202	0
	RipL	2260	0	974
	RipTPS	2260	0	35
RipG2	RipG6	147	2274	0
	RipG7	219	2202	0
	RipL	2421	0	974
	RipTPS	2421	0	35
RipG3	RipG2	0	2254	167
	RipG6	0	2254	20
	RipG7	52	2202	0
	RipL	2254	0	974
	RipTPS	2254	0	35
RipG4	RipG1	20	2260	0
	RipG2	0	2280	141
	RipG3	26	2254	0
	RipG5	20	2260	0
	RipG6	6	2274	0
	RipG7	78	2202	0
	RipL	2280	0	974
	RipTPS	2280	0	35
RipG5	RipG2	0	2260	161
	RipG3	6	2254	0
	RipG6	6	2254	20
	RipG7	58	2202	0
	RipL	2260	0	974
	RipTPS	2260	0	35
RipG6	RipG7	72	2202	0
	RipL	2274	0	974
	RipTPS	2274	0	35
RipG7	RipL	2202	0	974

	RipTPS	2202	0	35
RipTPS	RipL	35	0	974
RipY	RipAC	465	1662	808
	RipG1	596	1531	729
	RipG2	596	1531	890
	RipG3	596	1531	723
	RipG4	596	1531	749
	RipG5	596	1531	729
	RipG6	596	1531	743
	RipG7	607	1520	682
	RipL	2127	0	974
	RipTPS	2127	0	35

4.3 Domain-Based Predicted Protein-Protein Interactions

Regarding the results using this approach, after downloading the InterProScan database, domains were obtained for 34041 of the whole *S. lycopersicum* sequences. Concerning the sequences of the T3E, domains were collected for 36 T3E. Tomato and effectors sequences whose domains could not be found were discarded from the study since domains were necessary to continue searching for interactions.

Subsequently, the interactions of the T3E domains were searched using the 3did database. In this way, 128 interacting domains were found; they belong to 26 effectors (Table 7).

Table 7. T3E interacting domains obtained in 3did.

Gene ID	T3E Family Name	Interacting Domains
RSp0875	RipAC	ANATO, BclA_C, DUF3439, E1_DerP2_DerF2, EPF, Flagellin_C, Flagellin_N, Furin-like_2, I-set, IL6, Ig_3, Internalin_N, LRRCT, LRRNT, LRRNT_2, LRR_1, LRR_12, LRR_4, LRR_5, LRR_6, LRR_8, Laminin_N, Lys, OLF, Reprolysin, Spaetzle, TGF_beta, TGFb_propeptide, TPKR_C2, TSP_1, Trypsin, V-set, VWA, ZNRF_3_ecto, ig
RSc0321	RipAE	Acetyltransf_14, WRKY
RSp1236	RipAR	NEL
RSp1475	RipAW	NEL
RSp1022	RipAY	ChaC
RSc0245	RipB	IU_nuc_hydro

RSp0914	RipG1	EPF, LRRNT_2, LRR_1, LRR_4, LRR_6, LRR_8, LRR_RI_capping, Pkinase, RnaseA, ANATO, BclA_C, DUF3439, E1_DerP2_DerF2, Flagellin_C, Flagellin_N, Furin-like_2, I-set, IL6, Ig_3, Internalin_N, LRRCT, LRRNT, LRR_12, LRR_5, Laminin_N, Lys, OLF, Reprolysin, Spaetzle, TGF_beta, TGFb_propeptide, TPKR_C2, TSP_1, Trypsin, V-set, VWA, ZNRF_3_ecto, ig
RSp0672	RipG2	EPF, LRRNT_2, LRR_1, LRR_4, LRR_6, LRR_8, LRR_RI_capping, Pkinase, RnaseA, ANATO, BclA_C, DUF3439, E1_DerP2_DerF2, Flagellin_C, Flagellin_N, Furin-like_2, I-set, IL6, Ig_3, Internalin_N, LRRCT, LRRNT, LRR_12, LRR_5, Laminin_N, Lys, OLF, Reprolysin, Spaetzle, TGF_beta, TGFb_propeptide, TPKR_C2, TSP_1, Trypsin, V-set, VWA, ZNRF_3_ecto, ig
RSp0028	RipG3	EPF, LRRNT_2, LRR_1, LRR_4, LRR_6, LRR_8, LRR_RI_capping, Pkinase, RnaseA
RSc1800	RipG4	EPF, LRRNT_2, LRR_1, LRR_4, LRR_6, LRR_8, LRR_RI_capping, Pkinase, RnaseA, ANATO, BclA_C, DUF3439, E1_DerP2_DerF2, Flagellin_C, Flagellin_N, Furin-like_2, I-set, IL6, Ig_3, Internalin_N, LRRCT, LRRNT, LRR_12, LRR_5, Laminin_N, Lys, OLF, Reprolysin, Spaetzle, TGF_beta, TGFb_propeptide, TPKR_C2, TSP_1, Trypsin, V-set, VWA, ZNRF_3_ecto, ig
RSc1801	RipG5	ANATO, BclA_C, DUF3439, E1_DerP2_DerF2, EPF, Flagellin_C, Flagellin_N, Furin-like_2, I-set, IL6, Ig_3, Internalin_N, LRRCT, LRRNT, LRRNT_2, LRR_1, LRR_12, LRR_4, LRR_5, LRR_6, LRR_8, Laminin_N, Lys, OLF, Reprolysin, Spaetzle, TGF_beta, TGFb_propeptide, TPKR_C2, TSP_1, Trypsin, V-set, VWA, ZNRF_3_ecto, ig, LRR_RI_capping, Pkinase, RnaseA
RSc1356	RipG6	ANATO, BclA_C, DUF3439, E1_DerP2_DerF2, EPF, Flagellin_C, Flagellin_N, Furin-like_2, I-set, IL6, Ig_3, Internalin_N, LRRCT, LRRNT, LRRNT_2, LRR_1, LRR_12, LRR_4, LRR_5, LRR_6, LRR_8, Laminin_N, Lys, OLF, Reprolysin, Spaetzle, TGF_beta, TGFb_propeptide, TPKR_C2, TSP_1, Trypsin, V-set, VWA, ZNRF_3_ecto, ig, LRR_RI_capping, Pkinase, RnaseA
RSc1357	RipG7	FBA, Skp1, ubiquitin, EPF, LRRNT_2, LRR_1, LRR_4, LRR_6, LRR_8, LRR_RI_capping, Pkinase, RnaseA
RSc2132	RipJ	Acetyltransf_14, WRKY
RSp0193	RipL	PPR, PPR_1, PPR_2, PPR_3
RSp1130	RipN	53-BP1_Tudor, CTP_transf_like, DAP_epimerase, DCP1, DCP2, His_Phos_1, NUDIX, NUDIX-like, Nudix_N, Nudix_N_2, PNRC
RSc0826	RipP1	Acetyltransf_14, WRKY

RSc0868	RipP2	Acetyltransf_14, WRKY
RSc3401	RipS1	HD_4, RelA_SpoT
RSp0930	RipS3	HD_4, RelA_SpoT
RSc3212	RipT	Peptidase_C58
RSc1815	RipTAL	TAL_effector
RSp0731	RipTPS	Glyco_transf_20
RSc1349	RipV1	NEL
RSc2775	RipW	Pectate_lyase
RSc0257	RipY	AAA_lid_3, ACR_tran, APH, Activator_LAG-3, Adeno_knob, Ank, Ank_2, Ank_3, Ank_4, Ank_5, Arf, ArfGap, Arm, B, BTD, Bcl-2, C1-set, CC2-LZ, CENP-T_C, CoA_binding_2, Cob_adeno_trans, DSPc, EF-hand_1, EF-hand_14, EF-hand_7, EpoR_lig-bind, Ets, F-actin_cap_A, FAT, F_actin_cap_B, Fic, Fz, GFP, GF_recep_IV, GalBD_like, Glutaminase, I-set, IL13, IL4, KH_2, LAG1-DNAbind, LIM, LIM_bind, Lys, MamL-1, P53, Patatin, Peptidase_C1, Peptidase_C14, Peptidase_S13, Pkinase, RHD_dimer, Ras, Recep_L_domain, SBP_bac_1, SBP_bac_8, SH3_1, SH3_9, SOCS_box, Sema, TIG, TRP_2, Tubulin, fn3

Later, some domains obtained in InterProScan were not found in the 3did database; therefore, they were discarded from the study. Thus, a comparison was performed between the previously obtained interactive domains and the tomato domains. If a tomato domain was present in the list of domains with which a T3E protein interacted, it was assumed that both interacted. A total of 13615 possible PPIs were obtained, in which 20 T3E participate (Table 8).

Table 8. Predicted PPIs between the T3E of *R. solanacearum* GMI1000 and *S. lycopersicum* genome using the domain-based method.

Gene ID	T3E Family Name	Interacting tomato proteins	Pfams
RSp0875	RipAC	537	PF02221, PF00069, PF07714, PF08263, PF00560, PF13855, PF12799, PF13516, PF12819, PF13306, PF15102, PF11721, PF01582, PF00931, PF13676, PF14580, PF00646, PF16095, PF12937, PF13943, PF18511, PF00240, PF18052, PF00092, PF13519, PF13768, PF17123, PF13639, PF00097, PF14624, PF05762, PF13923, PF12861
RSc0321	RipAE	101	PF03106, PF10533, PF04500
RSp1022	RipAY	4	PF04752
RSc0245	RipB	19	PF01156

RSp0914	RipG1	1375	PF00069, PF07714, PF08263, PF00560, PF13855, PF12799, PF13516, PF12819, PF13306, PF15102, PF11721, PF01582, PF00931, PF13676, PF14580, PF00646, PF16095, PF12937, PF13943, PF18511, PF00240, PF18052, PF06293, PF14531, PF03822, PF13426, PF08447, PF00989, PF02149, PF14593, PF00036, PF13499, PF13405, PF13202, PF13833, PF10591, PF01476, PF12330, PF01657, PF00627, PF01163, PF17667, PF00433, PF00139, PF18483, PF00481, PF14380, PF01453, PF00954, PF08276, PF00582, PF04564, PF18346, PF13637, PF12796, PF00023, PF13606, PF13445, PF13639, PF00097, PF12202, PF12260, PF13947, PF06479, PF01683, PF01011, PF00027, PF13540, PF08311, PF07645, PF00024, PF14381, PF05773, PF13393, PF12745, PF03129, PF01636, PF02985, PF00400, PF00008, PF01179, PF02728, PF02727, PF11883, PF02221, PF00092, PF13519, PF13768, PF17123, PF14624, PF05762, PF13923, PF12861
RSp0672	RipG2	1375	PF00069, PF07714, PF08263, PF00560, PF13855, PF12799, PF13516, PF12819, PF13306, PF15102, PF11721, PF01582, PF00931, PF13676, PF14580, PF00646, PF16095, PF12937, PF13943, PF18511, PF00240, PF18052, PF06293, PF14531, PF03822, PF13426, PF08447, PF00989, PF02149, PF14593, PF00036, PF13499, PF13405, PF13202, PF13833, PF10591, PF01476, PF12330, PF01657, PF00627, PF01163, PF17667, PF00433, PF00139, PF18483, PF00481, PF14380, PF01453, PF00954, PF08276, PF00582, PF04564, PF18346, PF13637, PF12796, PF00023, PF13606, PF13445, PF13639, PF00097, PF12202, PF12260, PF13947, PF06479, PF01683, PF01011, PF00027, PF13540, PF08311, PF07645, PF00024, PF14381, PF05773, PF13393, PF12745, PF03129, PF01636, PF02985, PF00400, PF00008, PF01179, PF02728, PF02727, PF11883, PF02221, PF00092, PF13519, PF13768, PF17123, PF14624, PF05762, PF13923, PF12861

RSp0028	RipG3	1365	PF00069, PF07714, PF08263, PF00560, PF13855, PF12799, PF13516, PF12819, PF13306, PF15102, PF11721, PF01582, PF00931, PF13676, PF14580, PF00646, PF16095, PF12937, PF13943, PF18511, PF00240, PF18052, PF06293, PF14531, PF03822, PF13426, PF08447, PF00989, PF02149, PF14593, PF00036, PF13499, PF13405, PF13202, PF13833, PF10591, PF01476, PF12330, PF01657, PF00627, PF01163, PF17667, PF00433, PF00139, PF18483, PF00481, PF14380, PF01453, PF00954, PF08276, PF00582, PF04564, PF18346, PF13637, PF12796, PF00023, PF13606, PF13445, PF13639, PF00097, PF12202, PF12260, PF13947, PF06479, PF01683, PF01011, PF00027, PF13540, PF08311, PF07645, PF00024, PF14381, PF05773, PF13393, PF12745, PF03129, PF01636, PF02985, PF00400, PF00008, PF01179, PF02728, PF02727, PF11883
RSc1800	RipG4	1375	PF00069, PF07714, PF08263, PF00560, PF13855, PF12799, PF13516, PF12819, PF13306, PF15102, PF11721, PF01582, PF00931, PF13676, PF14580, PF00646, PF16095, PF12937, PF13943, PF18511, PF00240, PF18052, PF06293, PF14531, PF03822, PF13426, PF08447, PF00989, PF02149, PF14593, PF00036, PF13499, PF13405, PF13202, PF13833, PF10591, PF01476, PF12330, PF01657, PF00627, PF01163, PF17667, PF00433, PF00139, PF18483, PF00481, PF14380, PF01453, PF00954, PF08276, PF00582, PF04564, PF18346, PF13637, PF12796, PF00023, PF13606, PF13445, PF13639, PF00097, PF12202, PF12260, PF13947, PF06479, PF01683, PF01011, PF00027, PF13540, PF08311, PF07645, PF00024, PF14381, PF05773, PF13393, PF12745, PF03129, PF01636, PF02985, PF00400, PF00008, PF01179, PF02728, PF02727, PF11883, PF02221, PF00092, PF13519, PF13768, PF17123, PF14624, PF05762, PF13923, PF12861

RSc1801	RipG5	1375	PF02221, PF00069, PF07714, PF08263, PF00560, PF13855, PF12799, PF13516, PF12819, PF13306, PF15102, PF11721, PF01582, PF00931, PF13676, PF14580, PF00646, PF16095, PF12937, PF13943, PF18511, PF00240, PF18052, PF00092, PF13519, PF13768, PF17123, PF13639, PF00097, PF14624, PF05762, PF13923, PF12861, PF06293, PF14531, PF03822, PF13426, PF08447, PF00989, PF02149, PF14593, PF00036, PF13499, PF13405, PF13202, PF13833, PF10591, PF01476, PF12330, PF01657, PF00627, PF01163, PF17667, PF00433, PF00139, PF18483, PF00481, PF14380, PF01453, PF00954, PF08276, PF00582, PF04564, PF18346, PF13637, PF12796, PF00023, PF13606, PF13445, PF12202, PF12260, PF13947, PF06479, PF01683, PF01011, PF00027, PF13540, PF08311, PF07645, PF00024, PF14381, PF05773, PF13393, PF12745, PF03129, PF01636, PF02985, PF00400, PF00008, PF01179, PF02728, PF02727, PF11883
RSc1356	RipG6	1375	PF02221, PF00069, PF07714, PF08263, PF00560, PF13855, PF12799, PF13516, PF12819, PF13306, PF15102, PF11721, PF01582, PF00931, PF13676, PF14580, PF00646, PF16095, PF12937, PF13943, PF18511, PF00240, PF18052, PF00092, PF13519, PF13768, PF17123, PF13639, PF00097, PF14624, PF05762, PF13923, PF12861, PF06293, PF14531, PF03822, PF13426, PF08447, PF00989, PF02149, PF14593, PF00036, PF13499, PF13405, PF13202, PF13833, PF10591, PF01476, PF12330, PF01657, PF00627, PF01163, PF17667, PF00433, PF00139, PF18483, PF00481, PF14380, PF01453, PF00954, PF08276, PF00582, PF04564, PF18346, PF13637, PF12796, PF00023, PF13606, PF13445, PF12202, PF12260, PF13947, PF06479, PF01683, PF01011, PF00027, PF13540, PF08311, PF07645, PF00024, PF14381, PF05773, PF13393, PF12745, PF03129, PF01636, PF02985, PF00400, PF00008, PF01179, PF02728, PF02727, PF11883

RSc1357	RipG7	1474	PF01466, PF03931, PF09668, PF13975, PF00240, PF13650, PF00627, PF11976, PF08284, PF09280, PF13855, PF01020, PF14560, PF01599, PF00454, PF03031, PF12230, PF01805, PF00632, PF02179, PF12157, PF00439, PF09247, PF15288, PF06424, PF14559, PF13181, PF13881, PF00443, PF13423, PF00069, PF07714, PF08263, PF00560, PF12799, PF13516, PF12819, PF13306, PF15102, PF11721, PF01582, PF00931, PF13676, PF14580, PF00646, PF16095, PF12937, PF13943, PF18511, PF18052, PF06293, PF14531, PF03822, PF13426, PF08447, PF00989, PF02149, PF14593, PF00036, PF13499, PF13405, PF13202, PF13833, PF10591, PF01476, PF12330, PF01657, PF01163, PF17667, PF00433, PF00139, PF18483, PF00481, PF14380, PF01453, PF00954, PF08276, PF00582, PF04564, PF18346, PF13637, PF12796, PF00023, PF13606, PF13445, PF13639, PF00097, PF12202, PF12260, PF13947, PF06479, PF01683, PF01011, PF00027, PF13540, PF08311, PF07645, PF00024, PF14381, PF05773, PF13393, PF12745, PF03129, PF01636, PF02985, PF00400, PF00008, PF01179, PF02728, PF02727, PF11883
RSc2132	RipJ	101	PF03106, PF10533, PF04500
RSp0193	RipL	974	PF13041, PF01535, PF12854, PF13812, PF17177, PF14432, PF00076, PF16953, PF11977, PF10037, PF00571, PF00637, PF02889, PF00270, PF04851, PF00271, PF13431, PF03161, PF00265
RSp1130	RipN	99	PF01467, PF01678, PF06058, PF05026, PF00293, PF00300, PF01591, PF00686, PF13671, PF14803, PF18290, PF09296, PF15916, PF13869, PF03571
RSc0826	RipP1	101	PF03106, PF10533, PF04500
RSc0868	RipP2	101	PF03106, PF10533, PF04500
RSc3401	RipS1	10	PF13328, PF04607, PF01966, PF02824, PF00036, PF13202, PF13499, PF13405
RSp0930	RipS3	10	PF13328, PF04607, PF01966, PF02824, PF00036, PF13202, PF13499, PF13405
RSp0731	RipTPS	21	PF00982, PF02358, PF08282

RSc0257	RipY	1823	PF00004, PF16450, PF17862, PF07728, PF07724, PF05496, PF01434, PF06068, PF02359, PF02933, PF09336, PF04212, PF16725, PF06480, PF01057, PF09262, PF00439, PF01695, PF00498, PF00308, PF13771, PF13832, PF01636, PF00441, PF02770, PF02771, PF08028, PF12796, PF13857, PF13637, PF00023, PF13606, PF13962, PF12313, PF11900, PF00651, PF13920, PF00520, PF11834, PF00027, PF07885, PF03859, PF00612, PF01833, PF07714, PF00069, PF00415, PF18044, PF00642, PF18346, PF13445, PF13639, PF00097, PF17830, PF01529, PF00887, PF00635, PF16746, PF01412, PF03114, PF00169, PF14244, PF00025, PF08477, PF00071, PF01926, PF09439, PF04670, PF00168, PF00514, PF01749, PF13513, PF16186, PF13646, PF02985, PF04564, PF12937, PF00646, PF00225, PF16796, PF04826, PF01734, PF15511, PF02969, PF02629, PF00549, PF13607, PF00782, PF09192, PF00626, PF01331, PF03919, PF16561, PF10409, PF13350, PF00036, PF13405, PF13833, PF13202, PF07992, PF00070, PF13499, PF10591, PF14531, PF12763, PF06293, PF17958, PF00404, PF01699, PF00153, PF08976, PF04607, PF13328, PF01267, PF02259, PF00454, PF11865, PF08771, PF02260, PF08064, PF15785, PF01115, PF07650, PF02421, PF03029, PF00189, PF00412, PF12315, PF01803, PF11815, PF00112, PF08246, PF03051, PF00396, PF08127, PF08263, PF00560, PF13855, PF12799, PF03822, PF13426, PF08447, PF00989, PF02149, PF14593, PF01476, PF12330, PF01657, PF13516, PF00627, PF01163, PF17667, PF00433, PF00139, PF18483, PF00481, PF14380, PF01453, PF00954, PF08276, PF00582, PF12202, PF12260, PF13947, PF06479, PF13306, PF01683, PF01011, PF13540, PF15102, PF11721, PF08311, PF07645, PF00024, PF12819, PF14381, PF05773, PF13393, PF12745, PF03129, PF00400, PF00008, PF01179, PF02728, PF02727, PF11883, PF00910, PF08356, PF08355, PF10199, PF00009, PF01504, PF02493, PF14604, PF00018, PF07653, PF00091, PF12327, PF03953

Moreover, of 13615 predicted PPIs, 1823 correspond to the effector protein RipY, occupying the first place. On the contrary, the protein that showed the least interaction was RipAY with four interactions. In the results of this method, it can be observed that some pairs of effectors proteins (RipAE-RipP1; RipAE-RipP2; RipAE-RipJ; RipP1-RipP2; RipP1-RipJ; RipP2-RipJ; RipS1-RipS3) have the same interactions with the

tomato genome despite belonging to different T3E families (Table 9). These results could be since RipAE, RipJ, RipP1, and RipP2 are putative acetyltransferase proteins. Likewise, the RipS1 and RipS3 proteins share the same protein-protein interactions as they both are SKWP proteins. Also, some other pairs of T3E shared interactions, RipG1-RipG2; RipG1-RipG4; RipG1-RipG5; RipG1-RipG6; RipG2-RipG4; RipG2-RipG5; RipG2-RipG6; RipG4-RipG5; RipG4-RipG6; RipG5-RipG6. As mentioned above, they belong to the same family of proteins, and they are F-box LRR proteins; therefore, they can have the same interactions [30]. However, it should be emphasized that the other proteins of this family, RipG3 and RipG7, that also participate in this approach do not have the same interactions but instead have different ones. RipAE, RipAY, RipB, RipJ, RipN, RipP1, RipP2, RipS1, RipS3 proteins are part of the results of this approach but were not found in the results obtained with the first one, unlike the RipAC, RipG1 through RipG7, RipL, RipTPS, and RipY proteins, which are part of the interactions found with the Interolog method.

Table 9. Comparison of PPIs between the T3E of GMI1000 strain, domain-based method.

T3E 1	T3E 2	T3E1 PPIS	Share PPIs	T3E2 PPIs
RipAC	RipAY	537	0	4
	RipG1	0	537	838
	RipN	537	0	99
	RipS3	537	0	10
RipAE	RipAC	101	0	537
	RipAY	101	0	4
	RipG1	101	0	1375
	RipG2	101	0	1375
	RipG3	101	0	1365
	RipG4	101	0	1375
	RipG5	101	0	1375
	RipG6	101	0	1375
	RipG7	101	0	1474
	RipJ	0	101	0
	RipL	101	0	974
	RipN	101	0	99
	RipP1	0	101	0
	RipP2	0	101	0
	RipS1	101	0	10
	RipS3	101	0	10
	RipTPS	101	0	21
RipAY	RipN	4	0	99
RipB	RipAC	19	0	537

	RipAE	19	0	101
	RipAY	19	0	4
	RipG1	19	0	1375
	RipG2	19	0	1375
	RipG3	19	0	1365
	RipG4	19	0	1375
	RipG5	19	0	1375
	RipG6	19	0	1375
	RipG7	19	0	1474
	RipJ	19	0	101
	RipL	19	0	974
	RipN	19	0	99
	RipP1	19	0	101
	RipP2	19	0	101
	RipS1	19	0	10
	RipS3	19	0	10
	RipTPS	19	0	21
	RipY	19	0	1823
RipG1	RipAY	1375	0	4
	RipN	1375	0	99
	RipS3	1375	0	10
RipG2	RipAC	838	537	0
	RipAY	1375	0	4
	RipG1	0	1375	0
	RipN	1375	0	99
	RipS3	1375	0	10
	RipTPS	1375	0	21
RipG3	RipAC	838	527	10
	RipAY	1365	0	4
	RipG1	0	1365	10
	RipG2	0	1365	10
	RipL	1365	0	974
	RipN	1365	0	99
	RipS3	1365	0	10
	RipTPS	1365	0	21
RipG4	RipAC	838	537	0
	RipAY	1375	0	4
	RipG1	0	1375	0
	RipG2	0	1375	0
	RipG3	10	1365	0
	RipG5	0	1375	0
	RipJ	1375	0	101
	RipL	1375	0	974
	RipN	1375	0	99
	RipS1	1375	0	10

	RipS3	1375	0	10
	RipTPS	1375	0	21
RipG5	RipAC	838	537	0
	RipAY	1375	0	4
	RipG1	0	1375	0
	RipG2	0	1375	0
	RipG3	10	1365	0
	RipJ	1375	0	101
	RipL	1375	0	974
	RipN	1375	0	99
	RipS1	1375	0	10
	RipS3	1375	0	10
	RipTPS	1375	0	21
RipG6	RipAC	838	537	0
	RipAY	1375	0	4
	RipG1	0	1375	0
	RipG2	0	1375	0
	RipG3	10	1365	0
	RipG4	0	1375	0
	RipG5	0	1375	0
	RipG7	10	1365	109
	RipJ	1375	0	101
	RipL	1375	0	974
	RipN	1375	0	99
	RipS1	1375	0	10
	RipS3	1375	0	10
	RipTPS	1375	0	21
RipG7	RipAC	947	527	10
	RipAY	1474	0	4
	RipG1	109	1365	10
	RipG2	109	1365	10
	RipG3	109	1365	0
	RipG4	109	1365	10
	RipG5	109	1365	10
	RipJ	1474	0	101
	RipL	1474	0	974
	RipN	1474	0	99
	RipS1	1474	0	10
	RipS3	1474	0	10
	RipTPS	1474	0	21
RipJ	RipAC	101	0	537
	RipAY	101	0	4
	RipG1	101	0	1375
	RipG2	101	0	1375
	RipG3	101	0	1365

	RipL	101	0	974
	RipN	101	0	99
	RipS1	101	0	10
	RipS3	101	0	10
	RipTPS	101	0	21
RipL	RipAC	974	0	537
	RipAY	974	0	4
	RipG1	974	0	1375
	RipG2	974	0	1375
	RipN	974	0	99
	RipS3	974	0	10
	RipTPS	974	0	21
RipP1	RipAC	101	0	537
	RipAY	101	0	4
	RipG1	101	0	1375
	RipG2	101	0	1375
	RipG3	101	0	1365
	RipG4	101	0	1375
	RipG5	101	0	1375
	RipG6	101	0	1375
	RipG7	101	0	1474
	RipJ	0	101	0
	RipL	101	0	974
	RipN	101	0	99
	RipP2	0	101	0
	RipS1	101	0	10
	RipS3	101	0	10
	RipTPS	101	0	21
RipP2	RipAC	101	0	537
	RipAY	101	0	4
	RipG1	101	0	1375
	RipG2	101	0	1375
	RipG3	101	0	1365
	RipG4	101	0	1375
	RipG5	101	0	1375
	RipG6	101	0	1375
	RipG7	101	0	1474
	RipJ	0	101	0
	RipL	101	0	974
	RipN	101	0	99
	RipS1	101	0	10
	RipS3	101	0	10
	RipTPS	101	0	21
RipS1	RipAC	10	0	537
	RipAY	10	0	4

	RipG1	10	0	1375
	RipG2	10	0	1375
	RipG3	10	0	1365
	RipL	10	0	974
	RipN	10	0	99
	RipS3	0	10	0
	RipTPS	10	0	21
RipS3	RipAY	10	0	4
	RipN	10	0	99
RipTPS	RipAC	21	0	537
	RipAY	21	0	4
	RipG1	21	0	1375
	RipN	21	0	99
	RipS3	21	0	10
RipY	RipAC	1626	197	340
	RipAE	1823	0	101
	RipAY	1823	0	4
	RipG1	788	1035	340
	RipG2	788	1035	340
	RipG3	788	1035	330
	RipG4	788	1035	340
	RipG5	788	1035	340
	RipG6	788	1035	340
	RipG7	788	1035	439
	RipJ	1823	0	101
	RipL	1823	0	974
	RipN	1823	0	99
	RipP1	1823	0	101
	RipP2	1823	0	101
	RipS1	1822	1	9
	RipS3	1822	1	9
	RipTPS	1823	0	21

4.4 Confirmed Protein-Protein Interactions

Since some of the PPIs predicted in this study can be false positives, the two approaches utilized in this work were compared to corroborate the results obtained. After using the first approach, 21557 possible PPIs were discovered, while with the second approach, 13615 possible interactions were found. The comparison of results from both methods found 12261 confirmed PPIs belonging to 11 T3E (Table 10).

Table 10. Confirmed predicted PPIs between the T3E of *R. solanacearum* GMI1000 and *S. lycopersicum* genome using both approaches.

Gene ID	T3E Family Name	Interacting tomato proteins	Pfams
RSp0875	RipAC	527	PF13306, PF12799, PF16095, PF13855, PF13676, PF08263, PF00069, PF00560, PF13516, PF00931, PF01582, PF07714
RSp0914	RipG1	1365	PF13943, PF13306, PF12799, PF00008, PF00646, PF13855, PF13676, PF08263, PF00069, PF00560, PF13516, PF01582, PF07714
RSp0672	RipG2	1365	PF13943, PF13306, PF12799, PF00646, PF13855, PF13676, PF08263, PF00069, PF12937, PF00560, PF13516, PF01582, PF00481, PF07714
RSp0028	RipG3	1360	PF00560, PF08263, PF00646, PF13516, PF13855, PF13943, PF00069, PF01582, PF13676, PF07714
RSc1800	RipG4	1365	PF13943, PF13306, PF12799, PF00646, PF13855, PF13676, PF08263, PF00069, PF12937, PF00560, PF13516, PF01582, PF07714
RSc1801	RipG5	1365	PF13943, PF13306, PF12799, PF00646, PF13855, PF13676, PF08263, PF00069, PF00560, PF13516, PF01582, PF07714
RSc1356	RipG6	1360	PF13943, PF13306, PF00646, PF13855, PF13676, PF08263, PF00069, PF12937, PF00560, PF13516, PF01582, PF07714
RSc1357	RipG7	1360	PF00560, PF08263, PF00646, PF13516, PF13855, PF00069
RSp0193	RipL	973	PF17177, PF13812, PF13041, PF01535, PF12854
RSp0731	RipTPS	21	PF02358, PF00982
RSc0257	RipY	1200	PF12796, PF18346, PF13606, PF01412, PF07653, PF00169, PF13637, PF00023, PF00651, PF01529, PF00018, PF00069, PF06479, PF00887, PF00520, PF14604, PF07714

The T3E that showed the most interactions with the tomato genome were RipG1, RipG2, RipG4, and RipG5, with 1365 interactions each. The effector protein that showed the least interactions with the tomato genome was RipTPS with 21 interactions. Also, as occurred in both methods, some pairs of the F-box LRR proteins have the same interactions, these pairs of effectors are RipG1-RipG2; RipG1-RipG4; RipG1-RipG5; RipG2-RipG4; RipG2-RipG5; RipG3-RipG6; RipG3-RipG7; RipG4-RipG5; RipG6-RipG7 (Table 11). They all belong to the second family of the T3E and are the same type

of protein, which could explain their resemblance. Similarly, the RipG4 and RipG5 proteins share most of their interactions. However, some other proteins do not share interactions with the rest of the effector proteins, such as RipTPS and RipL. RipTPS belong to the last family of T3E and is a trehalose-phosphate synthase protein, whereas RipL falls into the third family, and is a pentatricopeptide repeats [30]. Finally, the interactions obtained from the RipAC and RipY proteins are very diverse, but they share some of their interactions with all of the effector proteins less RipTPS and RipL.

Table 11. Comparison of confirmed PPIs between the T3E of GMI1000 strain.

T3E 1	T3E 2	T3E1 PPIS	Share PPIs	T3E2 PPIs
RipAC	RipG1	0	527	838
	RipG2	0	527	838
	RipG3	5	522	838
	RipG4	0	527	838
	RipG5	0	527	838
	RipG6	5	522	838
	RipG7	5	522	838
	RipL	527	0	973
	RipTPS	527	0	21
RipG1	RipG5	0	1365	0
	RipG2	0	1365	0
	RipG3	5	1360	0
	RipG6	5	1360	0
	RipG7	5	1360	0
	RipL	1365	0	973
	RipTPS	1365	0	21
RipG2	RipG6	5	1360	0
	RipG7	5	1360	0
	RipL	1365	0	973
	RipTPS	1365	0	21
RipG3	RipG2	0	1360	5
	RipG6	0	1360	0
	RipG7	0	1360	0
	RipL	1360	0	973
	RipTPS	1360	0	21
RipG4	RipG1	0	1365	0
	RipG5	0	1365	0
	RipG2	0	1365	0
	RipG3	5	1360	0
	RipG6	5	1360	0
	RipG7	5	1360	0
	RipL	1365	0	973
	RipTPS	1365	0	21

RipG5	RipG3	5	1360	0
	RipG2	0	1365	0
	RipG6	5	1360	0
	RipG7	5	1360	0
	RipL	1365	0	973
	RipTPS	1365	0	21
RipG6	RipG7	0	1360	0
	RipL	1360	0	973
	RipTPS	1360	0	21
RipG7	RipL	1360	0	973
	RipTPS	1360	0	21
RipTPS	RipL	21	0	973
RipY	RipAC	1003	197	330
	RipG1	165	1035	330
	RipG2	165	1035	330
	RipG3	165	1035	325
	RipG4	165	1035	330
	RipG5	165	1035	330
	RipG6	165	1035	325
	RipG7	165	1035	325
	RipL	1200	0	973
	RipTPS	1200	0	21

4.5 Gene Ontology Analysis

Since the verified interactions are the main result of this work, an analysis based on gene ontology was made regarding the tomato proteins that interact with each T3E to clearly and easily show results. For this, interacting tomato proteins were classified according to the GO terms they belong to, demonstrating the tomato proteins' main functions, as shown in the bar plots below. Each bar plot has a Y-axis that represents the GO terms (functions) of the tomato proteins, and the X-axis represents the percentage of interacting tomato proteins that fulfill each function. Additionally, the GO terms provided in each bar plot are group according to the principal GO categories (Biological Process: BP, Molecular Function: MF, Cellular Component: CC), the groups are indicated on each graph.

The first bar plot (Figure 6) summarizes the confirmed PPIs obtained in this study, representing the functions (Go terms) of all the tomato proteins that interact with all the T3E. It also shows that 33.85% of the tomato proteins have the function of a ligand that interacts with other molecule's specific sites [81]. In contrast, 22.60% carry out a catalytic

activity. Then 20.50% of interacting proteins perform metabolic processes, and the rest of the tomato proteins (≈ 1.38%) are involved in biological and molecular regulation, reproductive processes, transducers, localization, transporters, and cellular, anatomical entities. This last term is a cellular component that refers to a cellular organism that can be either a material or an immaterial entity with a granularity bigger than a protein complex but smaller than an anatomical system [82]. It is important to remark that most of the interacting tomato proteins carry out biological processes.

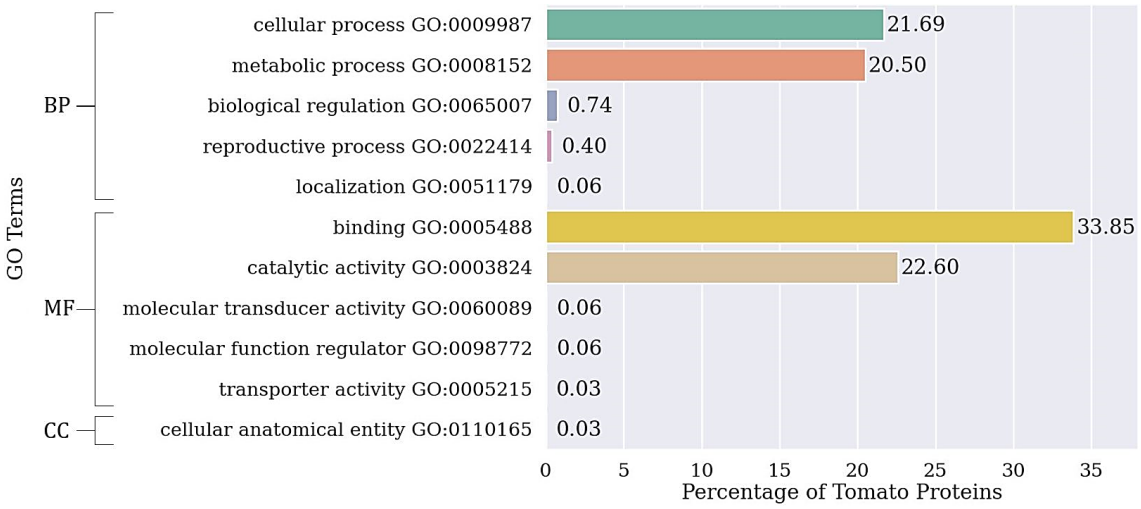


Figure 6. Confirmed PPIs bar plot.

Later, the following figures (7 to 17) are the ones that demonstrate the GO classification of the interacting tomato proteins, in a manner that the interactions can be appreciated per T3E. Figure 7 shows that the RipAC has 46.60% of interacting tomato proteins whose primary function is binding. Also, cellular, catalytic, and metabolic processes have ≈ 18% each, whereas biological regulation and molecular transducer activities occupy less than 1%. Figure 8 demonstrates similar results that RipAC, 31.96% of the tomato proteins that interact with RipG1 perform binding activities, catalytic activity occupies a 23.29% while cellular and metabolic processes are the 22.37% and 21.13%, respectively. Biological regulation and molecular transducer activity represent 0.78% and 0.05%, whereas the reproductive process perform a 0.42%. It is essential to notice that the following bar plots represent the effector proteins RipG2 through RipG7 (Figures 9-14), and they have almost the same interactions with tomato proteins as RipG1, which results in their GO terms being the same with approximately same percentages.

Moreover, Figure 15, which corresponds to RipL, shows different results; 98.78% of its interacting tomato proteins have a binding function, and catalytic activities,

localization, and cellular processes occupy 0.87%, 0.17%, and 0.17%, respectively. Furthermore, RipTPS results showed in Figure 16 indicate that 41.67% of the interacting tomato proteins accomplish a catalytic activity, and both cellular and metabolic processes represent 29.17%, each. Finally, results from Figure 17 demonstrate that a 29.88% of the tomato proteins with which RipY interacts do a binding activity, a 23.54% correspond to catalytic activities, a 22.57% carry out cellular processes while another 21.13% executes metabolic processes. With less than 1% each, other functions such as biological regulation, molecular function regulators, localization, reproductive processes, transporter activities, cellular, anatomical entity, and molecular transducer are also represented in this graph.

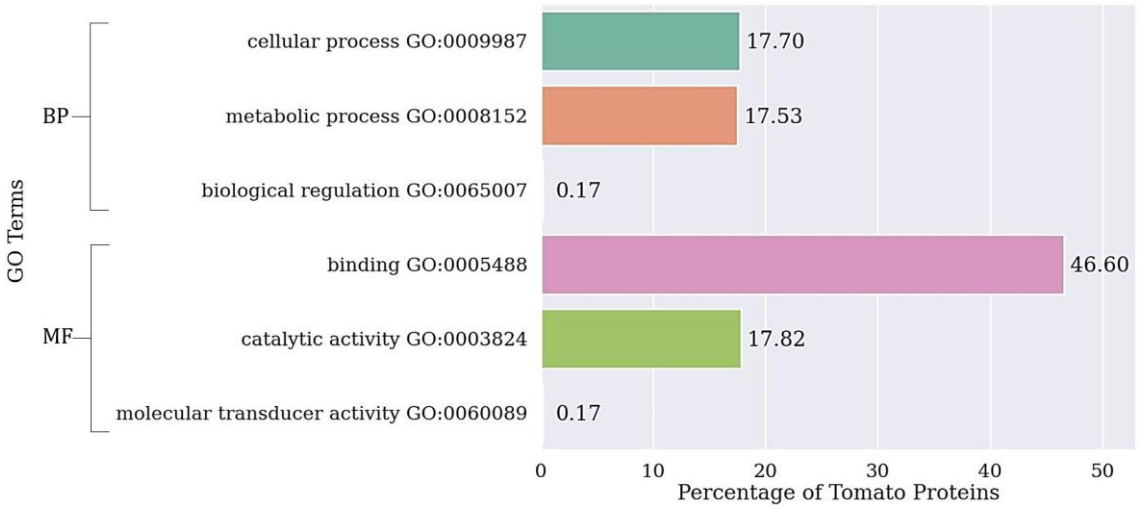


Figure 7. RipAC bar plot.

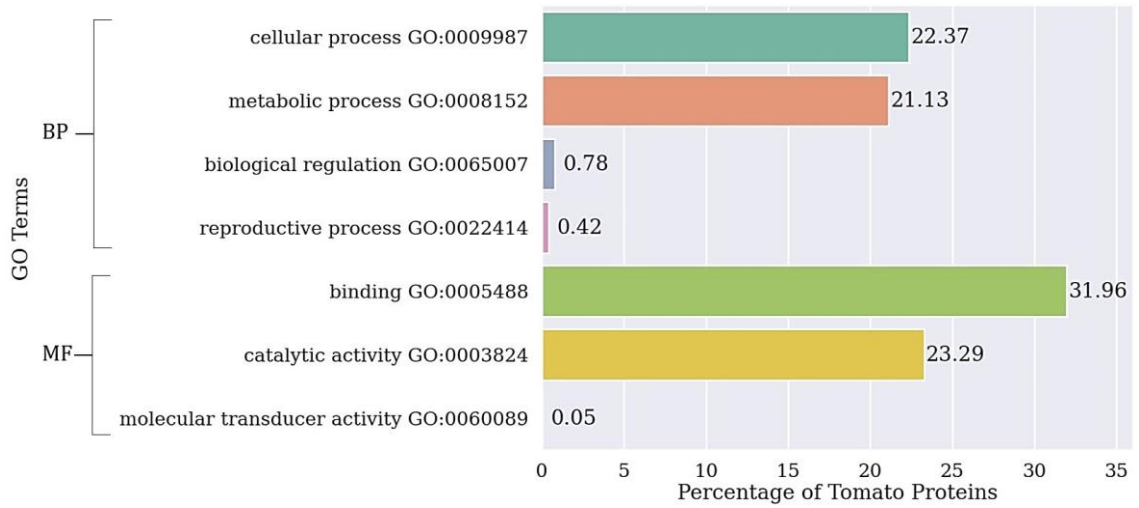


Figure 8. RipG1 bar plot.

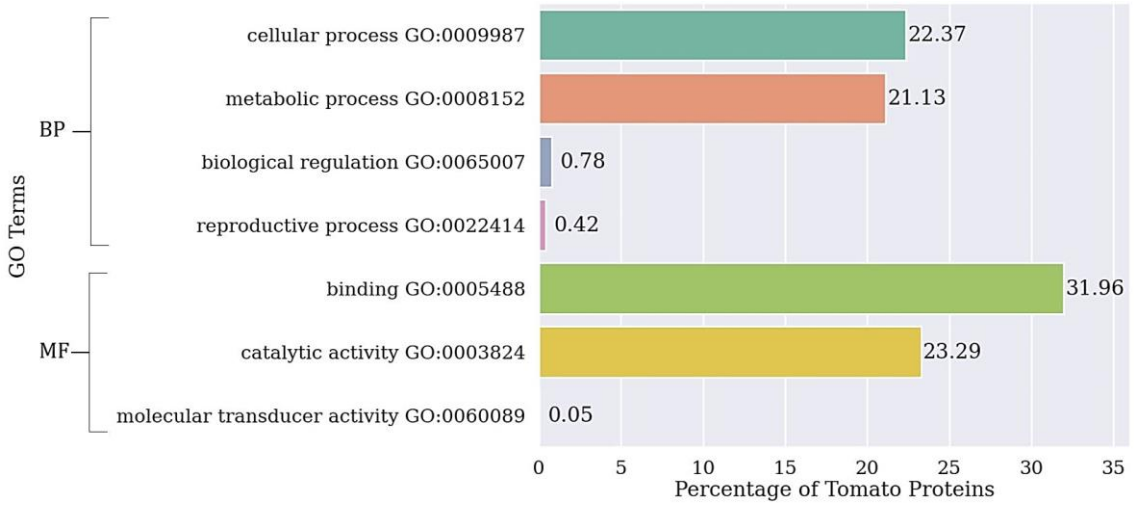


Figure 9. RipG2 bar plot.

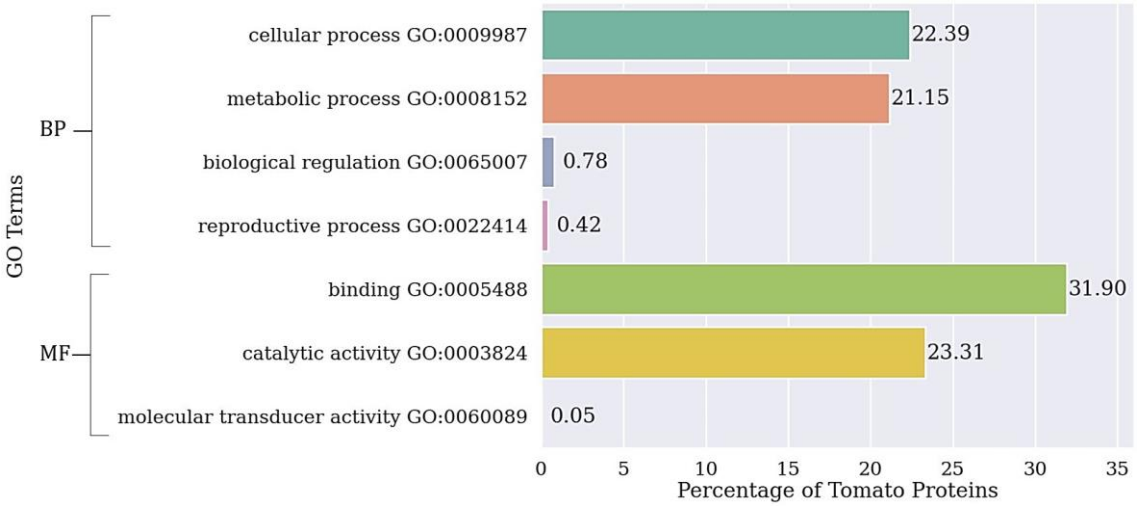


Figure 10. RipG3 bar plot.

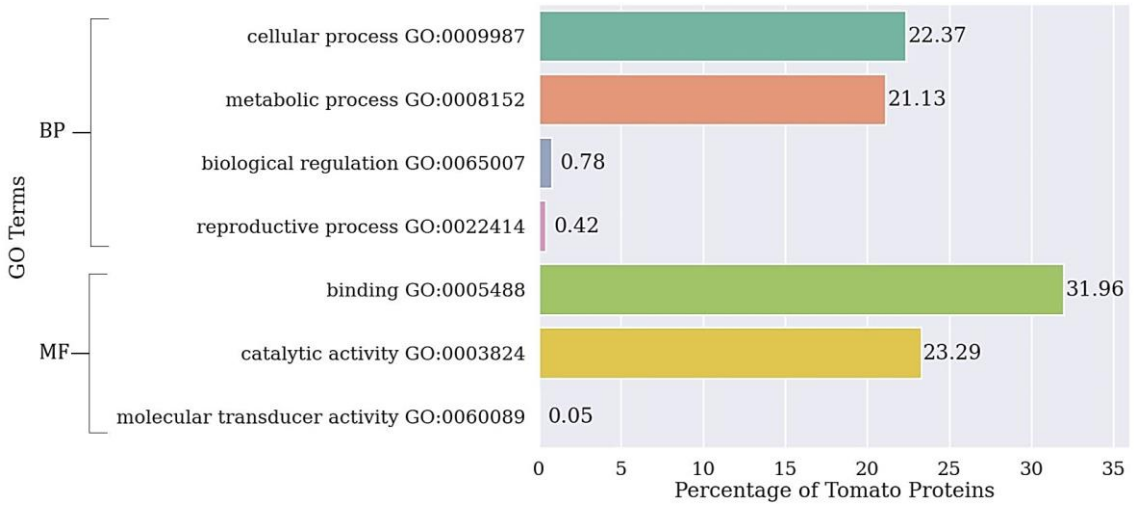


Figure 11. RipG4 bar plot.

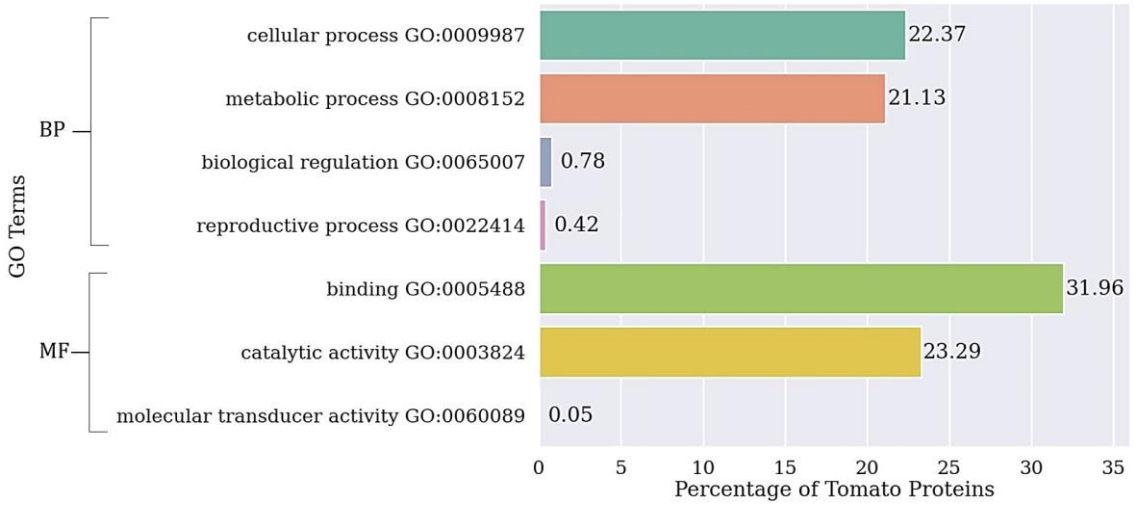


Figure 12. RipG5 bar plot.

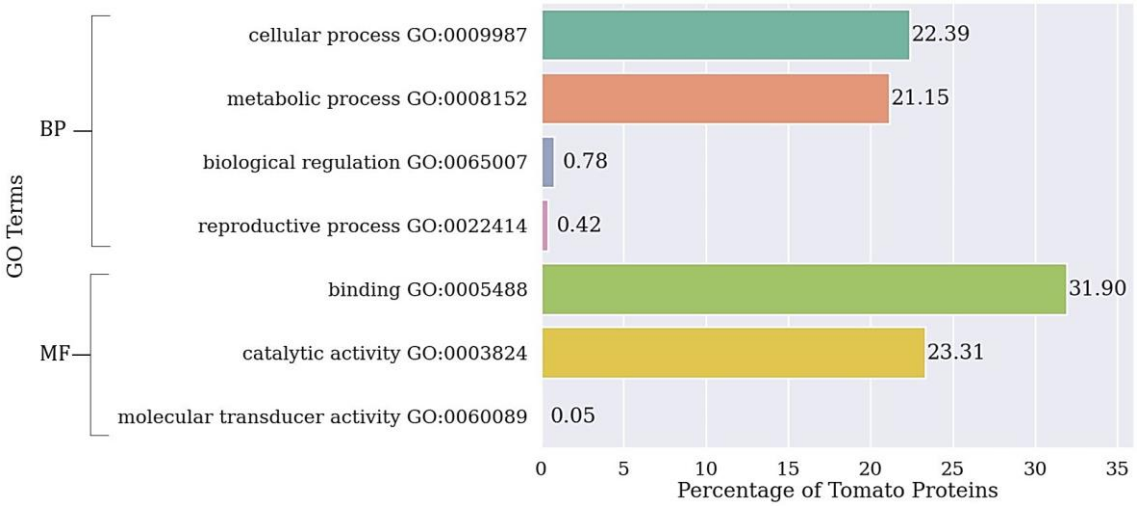


Figure 13. RipG6 bar plot.

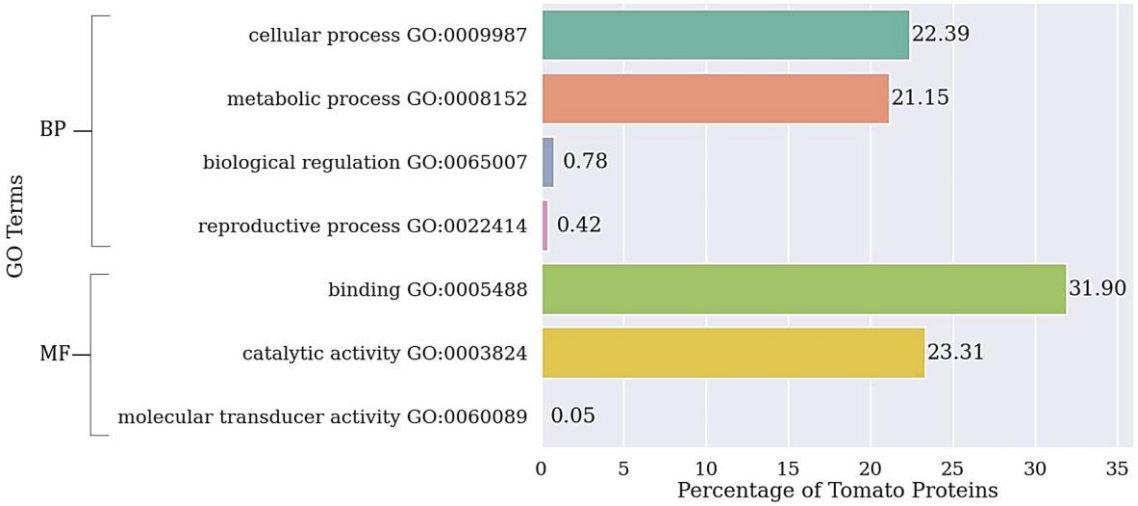


Figure 14. RipG7 bar plot.

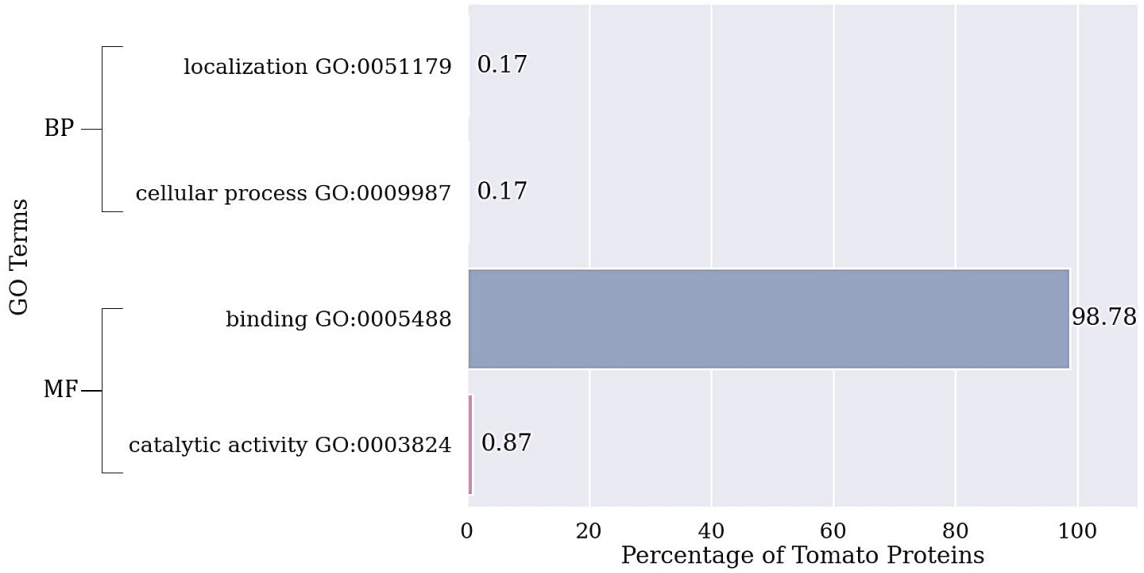


Figure 15. RipL bar plot.

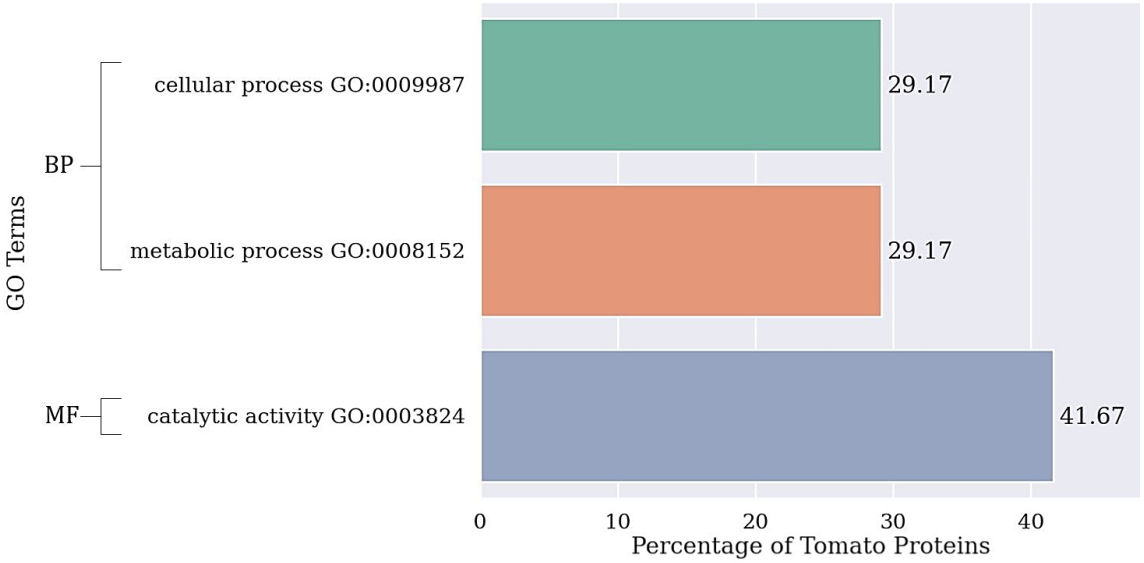


Figure 16. RipTPS bar plot.

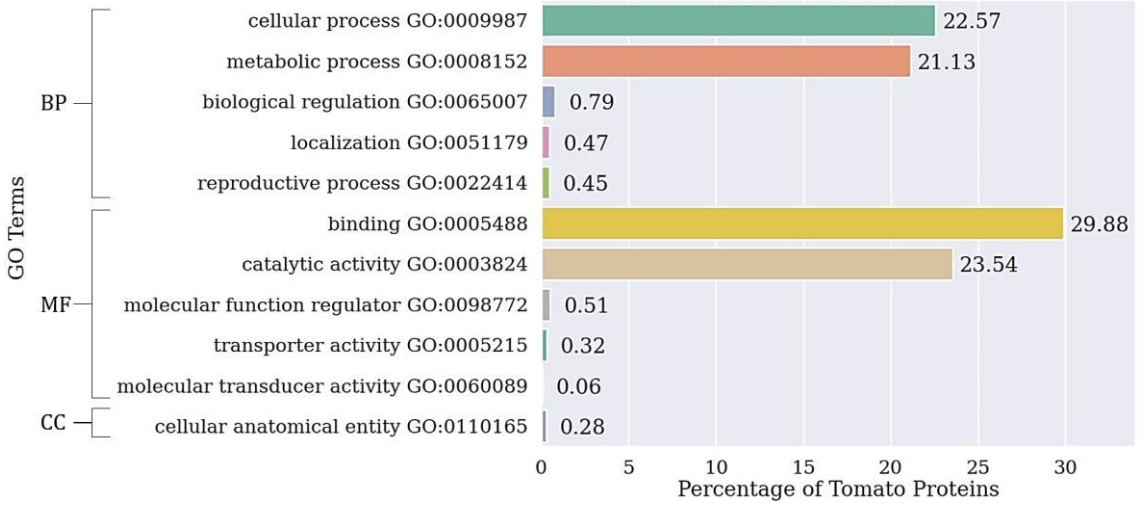


Figure 17. RipY bar plot.

4.5 Summary

This chapter shows the results obtained after applying two *in silico* approaches, the Interolog method and the Domain-Based method, to determine PPIs between the T3Es of *R. solanacearum* GMI1000 and *S. lycopersicum's* genome. A total of 35172 possible PPIs were found using both methods. Of this total, 12261 PPIs from 11 T3Es are considered confirmed, since they are the interactions present in both techniques. Also, a series of graphs that show both overall and individual results according to a GO analysis are presented.

5 Discussion

The sequence homology method has been used for many years to find protein-protein interactions, obtaining good results in applied studies. These excellent results are because the information necessary to carry it out is not difficult to find, as long as a database contains information about known and verified interactions regarding many organisms. Furthermore, the simplicity of its rationale that a protein interaction can be conserved in related species has allowed host-pathogen, that is, inter-species investigations to be conducted. In this way, it has been possible to expand a field of study where only the same individual's protein interactions were investigated [83,84]. Likewise, the domain-based method provides a significant contribution to the study of PPIs. This method is based on the role of protein domains and how sharing the same function can determine the interaction between a pair of them [85], which facilitates the search for interactions by knowing the domains of a pair of target proteins. Both methods help analyze large amounts of information, which means they are robust. Their false positive rate is low, given the reliability of the data used [49].

However, both methods have essential limitations to mention. In the case of the interolog method, relying only on protein homology may not be conclusive. Thus, different techniques should be carried out to corroborate the results obtained with this method [49], which is this study's case. Besides, very low e-values must be taken into account to ensure a high degree of confidence in the results achieved; otherwise, homologs that do not provide accurate results to the study could be used, increasing the false positives rate. On the side of the domain-based method, even though it is more precise and handles more specific information like protein domains, not enough information can be found about a specific organism, this depends on how well the species is known and whether proteins have clear/known domains. The difficulty in finding data in poorly studied organisms is also a limitation in the interolog method and many other *in silico* processes.

Despite their limitations and due to their results veracity, both approaches have been used in various studies. Li and collaborators [10] used both, methods to predict PPIs between *R. solanacearum* and *A. thaliana*, obtaining 3074 potential protein interactions

between both organisms. However, the author mentions that due to the lack of information in specific sources, the network of interactions is not complete, but it is of great importance for future studies of these species since antibacterial drugs can be designed from these results. In another study by Sahu [86], both methods were also used to predict the interactome between *Arabidopsis* and *Pseudomonas syringae*. In this work, approximately 11,000 possible interactions between these organisms were obtained, ensuring that the results represent an advance in understanding the host defense mechanism against the virulence of the pathogen. Another study by Lee et al. [87] aimed at predicting protein-protein interactions using only a method based on orthologs. They demonstrated that this method is also useful in studying host-pathogen interactions applied in humans and *P. falciparum* and not only plant-pathogen interactions. At the same time, Zhou et al. [88] carried out the study, whose primary focus was the prediction of PPIs among *H. sapiens-M. tuberculosis* H37Rv, the interolog method was used to obtain 1005 possible interactions. Furthermore, their results demonstrated that these *in silico* methods could be used for medical research purposes. These studies and many others not mentioned here prove the relevance of these techniques. Thus, as they are 100% valid and provide robust results, they represent a starting point for *in vivo* analysis to be performed from these data.

Since both methods were used in this study, it is necessary to mention specific points that must be improved or considered about the process followed here. Regarding the interolog method, not all the T3E had homologous sequences when blasting against the DIP database because the results had to comply with an e-value ≤ 0.001 to ensure its veracity. Therefore, not all effector proteins were used in this study. Consequently, some of the homologs obtained were eliminated since they belonged to *homo sapiens*; thus, they were considered false positives and were discarded from the analysis. Then, in the case of the Pfam accession number search, a few homologs and tomato proteins were discarded from the process since they did not have a Pfam number. In the case of the domain-based method, some tomato and T3E proteins did not possess known domains. Therefore, they were removed from the process. When comparing interacting domains carried out in 3did, some proteins were not included in the database; for this reason, it was assumed that they have no interactions with any known domain.

Concerning the results obtained, 12261 possible PPIs were accepted as verified, meaning that they were present in the results of both approaches used in this work. Also,

11 T3E participate in these interactions. After analyzing the confirmed PPIs, it was noticeable that some proteins, which belong to the second family of T3E, RipG1 to RipG7, share almost all their interactions. These proteins, known as GALAs, were found in an ancestral strain and evolved to the point where they are considered the basis for *R. solanacearum's* pathogenicity [89]. Therefore, and because they come from a common ancestor, it could be assumed that their similarity in terms of PPIs is logical, leading us to suggest that effector proteins which form part of the same family are most likely to have identical PPIs with an organism. In the same manner, when taking into account the results obtained from the domain-based method, RipAE, RipJ, RipP1, and RipP2, which are putative acetyltransferase proteins, demonstrated to have equal interactions despite belonging to different T3E families. Also, RipS1 and RipS3 share the same protein-protein interactions, and they both are SKWP proteins. In other words, it can be proposed that these effectors share their interactions due to a similarity in function regardless of their T3E families. To illustrate these hypotheses, from the confirmed PPIs, the effector proteins RipL and RipTPS did not share a single interaction with the rest of the T3E or between them, and they are both distinct types of proteins that belong to different families. Finally, the results discovered after conducting the gene ontology analysis showed that most of the tomato proteins, 33.85%, with which T3Es interact, are ligands that interrelate with specific sites on other molecules. Additionally, another 22.60% perform catalytic activities; 20.50% are in charge of metabolic processes while the rest of the interacting proteins are either involved in regulation processes, location or are cellular entities.

6 Conclusions and Future Prospects

In this study, we aimed at finding the possible protein-protein interactions (PPIs) between the type three effector (T3E) proteins of *R. solanacearum* GMI1000 and the proteins belonging to the *S. lycopersicum* genome, also called tomato. *R. solanacearum* is one of the deadliest pathogenic bacteria worldwide; it mainly attacks tomato crops, generating significant economic losses worldwide. Thus, it was necessary to find the possible PPI network between both organisms, to know the main biological functions that lead to a host-pathogen interaction and pathogenesis. Since there are already proven experimental methods for obtaining PPIs, two *in silico* approaches known for their exceptional performance were used for this work, the interolog method and the domain-based method.

After using the interolog method, 21,557 possible interactions were obtained with 11 T3E, while using the domain-based method, a total of 13,615 possible interactions were discovered, with 20 T3E. Both methods demonstrated high performance when processing the data. However, to ensure robust data, it was necessary to corroborate the discovered PPIs. The comparison of results from both approaches allowed obtaining a total of 12261 tomato interacting proteins with 11 T3E. From these verified interactions, the effectors RipG1 to RipG7 were found to share almost the same interactions with tomato proteins. In contrast, the effectors RipTPS and RipL were shown to interact with tomato proteins that are not associated with other T3Es.

For a better understanding of the roles of the tomato proteins with which the T3E interact, a gene ontology-based analysis was performed. As a result, 33.85% of tomato proteins fulfill binding functions, while 22.60% carry out catalytic activities, another 21.69% perform cellular processes, 20.50% carry out metabolic processes, and about 1.38% achieve processes such as biological regulation, molecular transducer, localization, among others. The possible reason assumed in this study for some effectors to have the same PPIs is that they belong to the same family; therefore, they can interact with the same proteins. Likewise, it is believed that if two effector proteins have similar functions, they will interact partially or totally with the same tomato proteins, despite belonging to different T3E families.

This study presented a bioinformatics application in a real case on a pathogen's strain dedicated to attacking tomato crops, which generates food losses and can harm the country's economy. The results obtained here could guide future research that wishes to fully understand the pathogenicity of *R. solanacearum* and the functions that specific tomato proteins play in this process. Additionally, it is shown that *in silico* methods are of high relevance within modern studies whose objective is to efficiently handle large amounts of genomic data, providing robust, reliable results. Finally, it is hoped that this work will serve as an inspiration to use other *in silico* methods to discover and study PPIs. Additionally, techniques like Yeast Two-Hybrid and Mass Spectrometry could be of great help to corroborate the results presented in this study.

References

[1] P. M. Pradhanang, M. T. Momol, S. M. Olson, and J. B. Jones, "Effects of plant essential oils on *Ralstonia solanacearum* population density and bacterial wilt incidence in tomato," *Plant Dis.*, vol. 87, no. 4, pp. 423–427, 2003.

[2] J. Tans-kersten, H. Huang, C. Allen, J. Tans-kersten, H. Huang, and C. Allen, "*Ralstonia solanacearum* Needs Motility for Invasive Virulence on Tomato," *J. Bacteriol.*, vol. 183, no. 12, pp. 3597–3605, 2001.

[3] S. Kimura and N. Sinha, "Tomato (*Solanum lycopersicum*): A model fruit-bearing crop," *Cold Spring Harb. Protoc.*, vol. 3, no. 11, 2008.

[4] D. R. Zeiss, M. I. Mhlongo, F. Tugizimana, P. A. Steenkamp, and I. A. Dubery, "Metabolomic profiling of the host response of tomato (*Solanum lycopersicum*) following infection by *Ralstonia solanacearum*," *Int. J. Mol. Sci.*, vol. 20, no. 16, 2019.

[5] E. M. Marcotte, M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg, "Detecting protein function and protein-protein interactions from genome sequences," *Science (80-.).*, vol. 285, no. 5428, pp. 751–753, 1999.

[6] L. Skrabanek, H. K. Saini, G. D. Bader, and A. J. Enright, "Computational prediction of protein-protein interactions," *Mol. Biotechnol.*, vol. 38, no. 1, pp. 1–17, 2008.

[7] D. Caldwell, B. S. Kim, and A. S. Iyer-Pascuzzi, "*Ralstonia solanacearum* differentially colonizes roots of resistant and susceptible tomato plants," *Phytopathology*, vol. 107, no. 5, pp. 528–536, 2017.

[8] A. Gerszberg, K. Hnatuszko-Konka, T. Kowalczyk, and A. K. Kononowicz, "Tomato (*Solanum lycopersicum* L.) in the service of biotechnology," *Plant Cell. Tissue Organ Cult.*, vol. 120, no. 3, pp. 881–902, 2015.

[9] L. M. Kiirika, F. Stahl, and K. Wydra, "Phenotypic and molecular characterization of resistance induction by single and combined application of chitosan and silicon

in tomato against *Ralstonia solanacearum*," *Physiol. Mol. Plant Pathol.*, vol. 81, pp. 1–12, 2013.

[10] Z. G. Li, F. He, Z. Zhang, and Y. L. Peng, "Prediction of protein-protein interactions between *Ralstonia solanacearum* and *Arabidopsis thaliana*," *Amino Acids*, vol. 42, no. 6, pp. 2363–2371, 2012.

[11] J. R. Alfano and A. Collmer, "Bacterial pathogens in plants: Life up against the wall," *Plant Cell*, vol. 8, no. 10, pp. 1683–1698, 1996.

[12] Y. Weinrauch and A. Zychlinsky, "The Induction of Apoptosis by Bacterial Pathogens," *Annu. Rev. Microbiol.*, vol. 53, pp. 155–187, 1999.

[13] D. L. Arnold and R. W. Jackson, "Bacterial genomes: Evolution of pathogenicity," *Curr. Opin. Plant Biol.*, vol. 14, no. 4, pp. 385–391, 2011.

[14] S. Leonard, F. Hommais, W. Nasser, and S. Reverchon, "Plant–phytopathogen interactions: bacterial responses to environmental and plant stimuli," *Environ. Microbiol.*, vol. 19, no. 5, pp. 1689–1716, 2017.

[15] J. Mansfield *et al.*, "Top 10 plant pathogenic bacteria in molecular plant pathology," *Mol. Plant Pathol.*, vol. 13, no. 6, pp. 614–629, 2012.

[16] S. Genin and C. Boucher, "*Ralstonia solanacearum*: Secrets of a major pathogen unveiled by analysis of its genome," *Mol. Plant Pathol.*, vol. 3, no. 3, pp. 111–118, 2002.

[17] J. A. Castillo, H. Secaira-Morocho, S. Maldonado, and K. N. Sarmiento, "Diversity and Evolutionary Dynamics of Antiphage Defense Systems in *Ralstonia solanacearum* Species Complex," *Front. Microbiol.*, vol. 11, no. May, 2020.

[18] J. A. Castillo and J. T. Greenberg, "Evolutionary dynamics of *Ralstonia solanacearum*," *Appl. Environ. Microbiol.*, vol. 73, no. 4, pp. 1225–1238, 2007.

[19] N. Peeters, A. Guidot, F. Vailleau, and M. Valls, "*Ralstonia solanacearum*, a widespread bacterial plant pathogen in the post-genomic era," *Mol. Plant Pathol.*, vol. 14, no. 7, pp. 651–662, 2013.

[20] T. M. Lowe-Power, D. Khokhani, and C. Allen, "How *Ralstonia solanacearum* Exploits and Thrives in the Flowing Plant Xylem Environment," *Trends Microbiol.*, vol. 26, no. 11, pp. 929–942, 2018.

[21] LIPM Bioinformatics Platform, "*Ralstonia solanacearum* GMI1000, MOLK2 and IPO1609 A phytopathogenic bacterium with a wide host range," 2020. [Online]. Available: https://iant.toulouse.inra.fr/bacteria/annotation/cgi/ralso.cgi.

[22] M. Poueymiro *et al.*, "Two type III secretion system effectors from *Ralstonia solanacearum* GMI1000 determine host-range Specificity on Tobacco," *Mol. Plant-Microbe Interact.*, vol. 22, no. 5, pp. 538–550, 2009.

[23] L. Zolobowska and F. Van Gijsegem, "Induction of lateral root structure formation on petunia roots: A novel effect of GMI1000 *Ralstonia solanacearum* infection impaired in Hrp mutants," *Mol. Plant-Microbe Interact.*, vol. 19, no. 6, pp. 597–606, 2006.

[24] J. T. Greenberg and B. A. Vinatzer, "Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells," *Curr. Opin. Microbiol.*, vol. 6, no. 1, pp. 20–28, 2003.

[25] M. Poueymiro and S. Genin, "Secreted proteins from *Ralstonia solanacearum*: a hundred tricks to kill a plant," *Curr. Opin. Microbiol.*, vol. 12, no. 1, pp. 44–52, 2009.

[26] B. K. Coombes, "Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria," *Trends Microbiol.*, vol. 17, no. 3, pp. 89–94, 2009.

[27] L. Deslandes and S. Genin, "Opening the *Ralstonia solanacearum* type III effector tool box: Insights into host cell subversion mechanisms," *Curr. Opin. Plant Biol.*, vol. 20, pp. 110–117, 2014.

[28] C. R. R. Sabbagh *et al.*, "Pangenomic type III effector database of the plant pathogenic *Ralstonia spp.*," *PeerJ*, vol. 2019, no. 8, pp. 1–21, 2019.

[29] N. S. Coll and M. Valls, "Current knowledge on the *Ralstonia solanacearum* type III secretion system," *Microb. Biotechnol.*, vol. 6, no. 6, pp. 614–620, 2013.

[30] LIPM Bioinformatic Team, "RALSTO T3E," 2015. [Online]. Available: https://iant.toulouse.inra.fr/bacteria/annotation/site/prj/T3Ev3/.

[31] S. Knapp and I. E. Peralta, "The Tomato and Its Botanical Relatives," pp. 7–21, 2016.

[32] M. Çalişkan, *Genetic Diversity in Plants*. Croatia, 2012.

[33] M. Dorais, D. L. Ehret, and A. P. Papadopoulos, "Tomato (*Solanum lycopersicum*) health components: From the seed to the consumer," *Phytochem. Rev.*, vol. 7, no. 2, pp. 231–250, 2008.

[34] C. M. Rick and J. I. Yoder, "Classical and molecular genetics of tomato: highlights and perspectives.," *Annu. Rev. Genet.*, vol. 22, pp. 281–300, 1988.

[35] I. Buddenhagen and A. Kelman, "Biological and physiological aspects of bacterial wilt caused by *Pseudomonas solanacearum*," *Annu. Rev. Phytopathol*, pp. 2:203-230, 1964.

[36] A. C. Hayward, "Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*," *Annu. Rev. Phytopathol*, pp. 29:65-87, 1991.

[37] E. Ortega, "La Mancha Bacteriana del Tomate *Ralstonia solanacearum*, sus Características y Manejo Integrado," Universidad Tecnológica Oteima, 2015.

[38] S. Mandal, I. Kar, A. K. Mukherjee, and P. Acharya, "Elicitor-induced defense responses in *Solanum lycopersicum* against *Ralstonia solanacearum*," *Sci. World J.*, vol. 2013, 2013.

[39] S. G. Kim *et al.*, "Evaluation of resistance to *Ralstonia solanacearum* in tomato genetic resources at seedling stage," *Plant Pathol. J.*, vol. 32, no. 1, pp. 58–64, 2016.

[40] S. Jones and J. M. Thornton, "Principles of protein-protein interactions," *Proc. Natl. Acad. Sci. USA*, vol. 93, no. January, pp. 13–20, 1996.

[41] R. Jansen *et al.*, "A Bayesian Networks Approach for Predicting Protein-Protein Interactions from Genomic Data," *Am. Assoc. Adv. Sci.*, vol. 302, pp. 449–453,

2003.

[42] A. G. Ngounou Wetie, I. Sokolowska, A. G. Woods, U. Roy, K. Deinhardt, and C. C. Darie, "Protein-protein interactions: Switch from classical methods to proteomics and bioinformatics-based approaches," *Cell. Mol. Life Sci.*, vol. 71, no. 2, pp. 205–228, 2014.

[43] A. Grigoriev, "On the number of protein-protein interactions in the yeast proteome," *Nucleic Acids Res.*, vol. 31, no. 14, pp. 4157–4161, 2003.

[44] S. Kerrien *et al.*, "IntAct - Open source resource for molecular interaction data," *Nucleic Acids Res.*, vol. 35, no. SUPPL. 1, pp. 561–565, 2007.

[45] M. Deng, S. Mehta, F. Sun, and T. Chen, "Inferring domain-domain interactions from protein-protein interactions," *Genome Res.*, vol. 12, no. 10, pp. 1540–1548, 2002.

[46] T. Pawson, "Protein modules and signalling networks," *Nature*, vol. 373, no. 6515. pp. 573–580, 1995.

[47] M. Pellegrini, D. Haynor, and J. M. Johnson, "Protein interaction networks," *Protein Interact. networks*, pp. 239–249, 2004.

[48] M. Oti, B. Snel, M. A. Huynen, and H. G. Brunner, "Predicting disease genes using protein-protein interactions," *J. Med. Genet.*, vol. 43, no. 8, pp. 691–698, 2006.

[49] E. Nourani, F. Khunjush, and S. Durmus, "Computational approaches for prediction of pathogen-host protein-protein interactions," *Front. Microbiol.*, vol. 6, no. FEB, pp. 1–10, 2015.

[50] S. Fields and O. K. Song, "A novel genetic system to detect protein-protein interactions," *Nature*, vol. 340, no. 6230, pp. 245–246, 1989.

[51] C. R. Montgomery, "Protein-protein interactions (PPIs): Types, methods for detection and analysis," *Protein-Protein Interact. Types, Methods Detect. Anal.*, vol. 59, no. 1, pp. 1–165, 2016.

[52] N. Möckli *et al.*, "Yeast split-ubiquitin-based cytosolic screening system to detect

interactions between transcriptionally active proteins," *Biotechniques*, vol. 42, no. 6, pp. 725–730, 2007.

[53] C. von Mering *et al.*, "Comparative assessment of large-scale data sets of protein-protein interactions," *Nature*, vol. 417, no. 6887, pp. 399–403, 2002.

[54] A. H. Y. Tong *et al.*, "Systematic genetic analysis with ordered arrays of yeast deletion mutants," *Science (80-.).*, vol. 294, no. 5550, pp. 2364–2368, 2001.

[55] J. W. Chang, Y. Q. Zhou, M. T. Ul Qamar, L. L. Chen, and Y. D. Ding, "Prediction of protein–protein interactions by evidence combining methods," *Int. J. Mol. Sci.*, vol. 17, no. 11, 2016.

[56] H. Tanwar and C. George Priya Doss, *Computational Resources for Predicting Protein–Protein Interactions*, 1st ed., vol. 110. Elsevier Inc., 2018.

[57] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 27, no. 1, pp. 29–34, 1999.

[58] I. Xenarios, Ł. Salwínski, X. J. Duan, P. Higney, S. M. Kim, and D. Eisenberg, "DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions," *Nucleic Acids Res.*, vol. 30, no. 1, pp. 303–305, 2002.

[59] H. W. Mewes *et al.*, "MIPS: A database for genomes and protein sequences," *Nucleic Acids Res.*, vol. 30, no. 1, pp. 31–34, 2002.

[60] R. A. Laskowski, V. V. Chistyakov, and J. M. Thornton, "PDBsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids," *Nucleic Acids Res.*, vol. 33, no. DATABASE ISS., pp. 266–268, 2005.

[61] S. Razick, G. Magklaras, and I. M. Donaldson, "iRefIndex: A consolidated protein interaction database with provenance," *BMC Bioinformatics*, vol. 9, pp. 1–19, 2008.

[62] T. S. Keshava Prasad *et al.*, "Human Protein Reference Database - 2009 update," *Nucleic Acids Res.*, vol. 37, no. SUPPL. 1, pp. 767–772, 2009.

[63] M. Rashid, S. Ramasamy, and G. P.S. Raghava, "A Simple Approach for Predicting Protein-Protein Interactions," *Curr. Protein Pept. Sci.*, vol. 11, no. 7, pp. 589–600, 2011.

[64] S. Kerrien *et al.*, "The IntAct molecular interaction database in 2012," *Nucleic Acids Res.*, vol. 40, no. D1, pp. 841–846, 2012.

[65] L. Licata *et al.*, "MINT, the molecular interaction database: 2012 Update," *Nucleic Acids Res.*, vol. 40, no. D1, pp. 857–861, 2012.

[66] A. Chatr-Aryamontri *et al.*, "The BioGRID interaction database: 2017 update," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D369–D379, 2017.

[67] D. Szklarczyk *et al.*, "The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D362–D368, 2017.

[68] M. Ashburner *et al.*, "Gene Ontology: tool for the unification of biology," *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, 2000.

[69] S. Carbon *et al.*, "The Gene Ontology Resource: 20 years and still GOing strong," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D330–D338, 2019.

[70] U. S. N. L. of M. National Center for Biotechnology Information, "National Center for Biotechnology Information (NCBI)." [Online]. Available: https://www.ncbi.nlm.nih.gov/.

[71] Z. Ding and D. Kihara, "Computational Methods for Predicting Protein-Protein Interactions Using Various Protein Features," *Current protocols in protein science*, vol. 93, no. 1. p. e62, 2018.

[72] A. J. M. Walhout *et al.*, "Protein Interaction Mapping in *C. elegans* Using Proteins Involved in Vulval Development," *Science (80-.).*, vol. 287, no. 5450, pp. 116–122, 2000.

[73] J. M. Urquiza Ortiz, "Nuevos Métodos de Predicción de Interacción de Proteína-Proteína Utilizando Sistemas Inteligentes en Bases de Datos de Proteómica," University of Granada, 2011.

[74] F. Jelinek, L. R. Bahl, and R. L. Mercer, “Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech,” *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 250–256, 1975.

[75] S. El-Gebali *et al.*, “The Pfam protein families database in 2019,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D427–D432, 2019.

[76] R. D. Finn *et al.*, “Pfam: The protein families database,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. 222–230, 2014.

[77] X. M. Zhao and L. Chen, “Domain-domain interaction identification with a feature selection approach,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5265 LNBI, pp. 178–186, 2008.

[78] T. Nguyen and T. Ho, “An Integrative Domain-Based Approach to Predicting Protein – Protein Interactions,” vol. 6, no. 6, pp. 1115–1132, 2008.

[79] E. Quevillon *et al.*, “InterProScan: Protein domains identifier,” *Nucleic Acids Res.*, vol. 33, no. SUPPL. 2, pp. 116–120, 2005.

[80] Structural Bioinformatics and Network Biology Group, “3did,” 2015. [Online]. Available: https://3did.irbbarcelona.org/.

[81] D. Binns *et al.*, “QuickGO: a web-based tool for Gene Ontology searching,” *Proteomics*, vol. 8, no. 23–24, p. n/a-n/a, 2008.

[82] M. The Jackson Laboratory, Bar Harbor, “Mouse Genome Informatics Web Site,” 2020. [Online]. Available: http://www.informatics.jax.org/vocab/gene_ontology.

[83] J. G. Kim *et al.*, “Predicting the interactome of *Xanthomonas oryzae* pathovar oryzae for target selection and DB service,” *BMC Bioinformatics*, vol. 9, pp. 1–6, 2008.

[84] J. Soyemi, I. Isewon, J. Oyelade, and E. Adebiyi, “Inter-Species/Host-Parasite Protein Interaction Predictions Reviewed,” *Curr. Bioinform.*, vol. 13, no. 4, pp. 396–406, 2018.

[85] T. Ideker and R. Sharan, “Protein networks in disease,” *Genome Res.*, vol. 18, no.

4, pp. 644–652, 2008.

[86] S. S. Sahu, T. Weirick, and R. Kaundal, "Predicting genome-scale *Arabidopsis-Pseudomonas syringae* interactome using domain and interolog-based approaches," *BMC Bioinformatics*, vol. 15, no. 11, pp. 1–8, 2014.

[87] S. A. Lee *et al.*, "Ortholog-based protein-protein interaction prediction and its application to inter-species interactions," *BMC Bioinformatics*, vol. 9, no. SUPPL. 12, pp. 1–9, 2008.

[88] H. Zhou *et al.*, "Stringent homology-based prediction of *H. sapiens-M. tuberculosis* H37Rv protein-protein interactions," *Biol. Direct*, pp. 1–30, 2014.

[89] P. Remigi, M. Anisimova, A. Guidot, S. Genin, and N. Peeters, "Functional diversification of the GALA type III effector family contributes to *Ralstonia solanacearum* adaptation on different plant hosts," *New Phytol.*, vol. 192, no. 4, pp. 976–987, 2011.