TERRA LATINOAMERICANA

Phytophthora-Responsive Proteins in Cacao: Comparative Bioinformatic Analysis for Biological Control

Proteínas Asociadas a la Respuesta a Phytophthora en Cacao: Análisis Bioinformático Comparativo para el Control Biológico

Danna Belén Castillo-Quijije[1], Ángel Virgilio Cedeño-Moreira[2], María Eugenia Romero-Román[2], and Ronald Oswaldo Villamar-Torres[2‡]

[1] Universidad Técnica Estatal de Quevedo, Carrera de Agronomía, Facultad de Ciencias Agrarias y Forestales. Av. Quito Km 1.5, Vía a Santo Domingo de los Tsáchilas. Quevedo, Los Ríos, Ecuador; (D.B.C.Q.).
[2] Universidad Técnica Estatal de Quevedo, Facultad de Ciencias Pecuarias y Biológicas, Carrera Agropecuaria. Vía Quevedo-El Empalme km 7, Campus "La María". 120501 Quevedo, Los Ríos, Ecuador; (A.V.C.M.), (M.E.R.R.), (R.O.V.T.).
‡ Corresponding author: rvillamart@uteq.edu.ec

SUMMARY

Phytophthora spp. is one of the most important phytosanitary threat to cacao agroforestry systems, causing yield losses exceeding 40%. This study conducted a comparative bioinformatic analysis to identify proteins associated with resistance mechanisms in *Arabidopsis thaliana*, *Theobroma cacao L.*, and *Herrania umbratica*. From an initial dataset of 91 *Arabidopsis* sequences, 20 genes were identified as potentially involved in plant defense. Protein sequences were analyzed using BLASTX, selecting those with E-values ≤ 1e-10, adequate query coverage, and identity values above 65%. These proteins were then compared against *T. cacao* and *H. umbratica* genomes using BLASTP, enabling detection of functionally relevant homologs. Ortholog analysis with OrthoVenn3 revealed 13 conserved proteins shared among the three species, including lectin-like receptors, transcription factors (bZIP60, NAC089), stress-related proteins (PHOS32), and key immune signaling components such as FLS2, RBK1, and RPH1. Phylogenetic analyses highlighted strong evolutionary proximity between *Theobroma cacao* L. and *Herrania umbratica*, while *Arabidopsis* showed expected divergence but retained conserved immune-associated pathways. These conserved proteins represent promising candidates for functional validation and potential application in cacao breeding programs aimed at enhancing resistance to *Phytophthora spp*. Overall, this study demonstrates the value of *in silico* comparative approaches to accelerate the discovery of defense-related genes and support the development of more resilient and sustainable cacao production systems.

Index words: *comparative bioinformatics, orthologous proteins, resistance to oomycetes, transcription factors, virulence factors.*

RESUMEN

Phytophthora spp. representa una de las principales amenazas fitosanitaria para los sistemas agroforestales de cacao, provocando pérdidas superiores al 40%. Este estudio realizó un análisis comparativo bioinformático con el fin de identificar proteínas asociadas a mecanismos de resistencia en *Arabidopsis thaliana*, *Theobroma cacao* L. y *Herrania umbratica*. A partir de 91 secuencias iniciales de *Arabidopsis* se identificaron 20 genes vinculados con la defensa vegetal. Las secuencias proteicas se analizaron mediante BLASTX, seleccionándose aquellas con valores E ≤ 1e-10, coberturas adecuadas e identidades superiores al 65%. Posteriormente, estas proteínas fueron comparadas con los genomas de *T. cacao* y *H. umbratica* mediante BLASTP, lo que permitió detectar homólogos funcionalmente relevantes. El análisis

Recommended citation:

Castillo-Quijije, D. B., Cedeño-Moreira, A. V., Romero-Román, M. E., & Villamar-Torres, R. O., (2026). *Phytophthora*-Resistance–Related Proteins in Cacao: Bioinformatic Analysis for Biological Control. *Terra Latinoamericana, 44*, 1-14. e2472. https://doi.org/10.28940/terralatinoamericana.v44i.2472

Received: December 5, 2025.
Accepted: March 23, 2026.
Article, Volume 44.
June, 2026.

Section Editor:
Dr. Fernando Abasolo Pacheco

ortólogo con OrthoVenn3 mostró 13 proteínas conservadas entre las tres especies, destacando receptores tipo lectina, factores de transcripción (bZIP60, NAC089), proteínas de estrés (PHOS32) y componentes clave de señalización como FLS2, RBK1 y RPH1. Los análisis filogenéticos evidenciaron una estrecha relación evolutiva entre *T. cacao y H. umbratica*, junto con divergencia esperada respecto a *Arabidopsis*, aunque con conservación funcional de rutas inmunes esenciales. Estas proteínas conservadas representan candidatas clave para estudios funcionales y su potencial aplicación en programas de mejoramiento genético orientados a fortalecer la resistencia del cacao frente a *Phytophthora spp*. Los resultados confirman el valor del enfoque *in silico* para acelerar la identificación de genes defensivos y promover sistemas de producción más resilientes y sostenibles.

Palabras clave: *bioinformática comparativa, proteínas ortólogas, resistencia a oomicetos, factores de transcripción, factores de virulencia*.

INTRODUCTION

Theobroma cacao L**.**, an indigenous tree of the South American tropical rainforest, originates from the lower eastern equatorial slopes of the Andes (Vega-Jarquín, 2017). Currently, its cultivation has expanded to low-altitude tropical regions around the world, becoming an important commercial crop for millions of small-scale farmers (Argout *et al*., 2008). In Ecuador, especially in the province of Los Ríos, cacao is a pillar of the agricultural economy, standing out for its production of fine-flavor cacao and the CCN51 variety (Carranza-Quimi, Castro, Risco, and Cabezas, 2020).

Ecuadorian beans possess sensory characteristics that grant them an international competitive advantage. However, their production is threatened by diseases primarily black pod rot caused by *Phytophthora* spp., which can reduce yields by more than 40% (Argout *et al*., 2008; Guerrero, Vera, Acosta, Mónaco y Palma, 2026), generating annual losses of up to USD 2 billion in dry beans (Baruah, Ali, Shao, Lary, and Bailey, 2022). This pathogen is an obstacle to sustainable cacao production, causing global losses of nearly 30% and up to 10% of tree mortality. These conditions can worsen in the presence of factors such as high humidity in both soil and environment (Abad, Cevallos, Montealegre, and Romero, 2021; Rodríguez and Vera, 2015).

Phytophthora spp. is a filamentous eukaryotic pathogen of major agricultural and economic impact, responsible for multi-billion-dollar losses in global agriculture each year (Hou *et al*., 2019). Some studies suggest that resistance to one species of *Phytophthora* may confer cross-tolerance to other species within the same genus (Ali *et al*., 2017). In this context, the use of *Arabidopsis thaliana* as a model plant—given its well-characterized genetic system offers an opportunity to study plant resistance mechanisms (Irion and Nusslein, 2022). Thus, the existence of orthologous genes or conserved mechanisms suggests a possible functional similarity between Arabidopsis and *T.* cacao (Bomfim-Rego, Mora, Pirovani, Luz, and Corrêa, 2022). Identifying key proteins in Arabidopsis could therefore open new avenues for strengthening cacao resistance to *Phytophthora* spp.

In this regard, bioinformatics plays an essential role by enabling the analysis of genomic and protein sequences associated with pathogen responses. This discipline has facilitated gene annotation and the discovery of proteins related to tolerance in various plant species (Capote-Maínez and Hernández, 2021). Its application in *Arabidopsis thaliana* can reveal valuable information which, when compared with *T. cacao* and *Herrania umbratica*, a species closely related to *T. cacao* (Tarigan, Maharijaya, and Izzah, 2021), may be used to develop biotechnological strategies for genetic improvement.

This research aims to identify protein sequences associated with resistance to *Phytophthora* spp. in *Arabidopsis thaliana* using bioinformatic tools, and to evaluate their potential application for improving tolerance in *T. cacao*. By integrating plant biology and bioinformatics, this study seeks to contribute to a promising strategy to address the challenges posed by *Phytophthora* spp. Moving toward a more resilient and sustainable cacao production system reducing the use of agrochemicals harmful to both human health and the planet is the current challenge.

MATERIALS AND METHODS

This research was based on a bioinformatic comparative analysis of proteins associated with defense mechanisms against *Phytophthora* spp., using sequences from *Arabidopsis thaliana*, *T. cacao*, and *Herrania umbratica*. Public databases and bioinformatic tools were used to find, compare, and analyze orthologous protein sequences to infer potential resistance mechanisms in *T. cacao*.

Data Collection

Protein sequences from *T. cacao* and *Herrania umbratica* were downloaded from the NCBI Datasets repository, selecting the most complete and up-to-date available assemblies. Files were downloaded in FASTA format, suitable for bioinformatic analyses. Sequences associated with the response to *Phytophthora spp*. in *Arabidopsis thaliana* were obtained through an advanced search in NCBI, using the query syntax: ((“*Phytophthora*”[Organism] OR *Phytophthora* [All Fields]) AND (“*Arabidopsis thaliana*”[Organism] OR *Arabidopsis thaliana* [All Fields])) AND alive[prop]. Relevant nucleotide sequences were also downloaded in FASTA format.

Protein Sequence Similarity Analysis

A total of 91 nucleotide sequences from *Arabidopsis thaliana* were analyzed using BLASTX (NCBI, 2025) to identify homologous proteins related to defense against *Phytophthora spp*. Sequences with low E-values (≤ 1e-10) (Zaru, Orchard, and UniProt Consortium, 2023), identity ≥ 65%, and acceptable coverage were prioritized. Only sequences meeting these criteria were retained, resulting in a subset of 20 candidate proteins associated with defense-related functions. These sequences were subsequently used as references for further analyses. The resulting sequences were used as references for subsequent analyses. Next, these proteins were subjected to BLASTP to identify similar sequences in *T. cacao*, using *T. cacao* as the comparison organism. The selected candidate proteins showed support from ESTs (Expressed Sequence Tags), reinforcing their functional relevance despite the lack of strict identity or coverage criteria applied at this stage (Magar *et al*., 2022).

Comparison of Orthologous Protein Sequences

The OrthoVenn3 tool (Sun, Lu, Luo, Bie, Xu y Wang, 2023) was used to perform orthologous cluster analyses among the studied species. Protein sequences were grouped into FASTA files by species. Three comparative analyses were conducted:

Arabidopsis thaliana vs. *Theobroma cacao L*. and *Herrania umbratica*

Arabidopsis thaliana vs. *Theobroma cacao L*.

Theobroma cacao L. vs. *Herrania umbratica*

In all cases, an E-value threshold of 1e-10 was applied. A total of 13 orthologous clusters were identified through comparison, from which 11 clusters of interest were selected based on their functional annotation and potential association with defense mechanisms.

Intracluster Phylogeny of Proteins Involved in Pathogen Defense Mechanisms

Proteins contained within the 11 selected clusters were aligned using Clustal Omega (Sievers *et al*., 2011) to construct phylogenetic trees. For the phylogenetic analysis, the FASTA file was uploaded to Clustal Omega, the input type was set to “protein,” and the output format was configured as “ClustalW with character counts.” The resulting trees allowed the inference of evolutionary relationships and conservation patterns among orthologous proteins from the species studied.

RESULTS AND DISCUSSION

From the initial search, 91 nucleotide sequences were identified, of which 20 genes in *Arabidopsis thaliana* were potentially linked to defense mechanisms against pathogen attack, and particularly to possible resistence to *Phytophthora spp*. These genes represent a key starting point for understanding the molecular pathways involved in the immune response of this model species.

Sequence Similarity Analysis of Proteins

Search for protein sequence similarity related to the response to *Phytophthora spp.* in *Arabidopsis thaliana*. Based on the nucleotide sequences identified, translations were obtained using BLASTX, and their corresponding protein sequences were analyzed. Only those proteins showing identity values above 65%, high coverage percentages, and extremely low E-values were selected, prioritizing sequences with previously reported functions associated with defense. Sequences without significant similarity or with incorrect annotations were discarded. Finally, a set of candidate proteins with high similarity and functional relevance was selected for subsequent analyses (Table 1).

Search for protein sequence similarity involved in defense mechanisms against *Phytophthora spp.* in *Theobroma cacao L.* The selected proteins from *A. thaliana* were used as reference sequences in a BLASTP analysis against the protein database of *T. cacao*. Using this species as the organism of comparison optimized the sensitivity of the search and allowed the detection of homologs with potentially conserved functions in defense against *Phytophthora spp*. The sequences with the highest similarity values are presented in Table 2.
The identification of PHOS32 and PHOS34 from *Arabidopsis thaliana* mapping to the same predicted protein in *T. cacao* (XP_017980543.1) likely reflects a paralogous relationship within *Arabidopsis*. Gene duplication events in the *Arabidopsis* genome may have generated multiple paralogs belonging to the universal stress protein family, which correspond to a single orthologous protein in cacao. Such many-to-one orthology relationships are commonly observed in comparative genomic analyses due to lineage-specific gene duplication events (Tegenfeldt *et al*., 2025).

Comparison of Orthologous Protein Sequences

Comparative orthology analyses were conducted between the defense-related proteins identified in *Arabidopsis thaliana* and the protein sequences of *T. cacao* and *Herrania umbratica* using the OrthoVenn3 software. This tool enabled the generation of orthologous clusters based on high-quality alignments (E-value ≤ 1e-10), which were visualized through Venn diagrams.

Table 1 Selection of specific protein sequences involved in the response to *Phytophthora* spp. in *Arabidopsis thaliana*.

N°	Gen	Protein	Organism	Query Cover	E value	Per. Ident	Accession	Amino acids
1	LecRK-I.9	Concanavalin A-like lectin protein kinase family protein	*Arabidopsis thaliana*	60%	0.0	100.00%	NP_200838.1	718
2	BZIP60	Basic region/leucine zipper motif 60	*Arabidopsis thaliana*	44%	1.00E-76	100.00%	NP_174998.1	295
3	NAC089	NAC domain containing protein 89	*Arabidopsis thaliana*	37%	1.00E-100	100.00%	NP_001330322.1	241
4	AT5G65600	Concanavalin A-like lectin protein kinase family protein	*Arabidopsis thaliana*	76%	0.0	100.00%	NP_201363.1	675
5	PHOS32	Adenine nucleotide alpha hydrolases-like superfamily protein	*Arabidopsis thaliana*	25%	7.00E-106	100.00%	NP_001332520.1	250
6	AT1G22810	Integrase-type DNA-binding superfamily protein	*Arabidopsis thaliana*	53%	7.00E-106	100.00%	NP_173695.1	144
7	CLT2	CRT (chloroquine-resistance transporter)-like transporter 2	*Arabidopsis thaliana*	53%	6.00E-91	100.00%	NP_001190820.1	431
8	PHOS34	Adenine nucleotide alpha hydrolases-like superfamily protein	*Arabidopsis thaliana*	30%	8.00E-105	100.00%	NP_001328092.1	259
9	FLS2	Leucine-rich receptor-like protein kinase family protein	*Arabidopsis thaliana*	81%	0.0	97.63%	NP_001330009.1	1173
10	UMAMIT36	Nodulin MtN21 /EamA-like transporter family protein	*Arabidopsis thaliana*	44%	1.00E-85	94.44%	NP_177183.2	375
11	RBK1	ROP binding protein kinases 1	*Arabidopsis thaliana*	53%	2.00E-151	90.08%	NP_568231.1	467
12	AT5G43930	Transducin family protein / WD-40 repeat family protein	*Arabidopsis thaliana*	62%	0.0	86.87%	NP_001190467.1	752
13	RPH1	Resistance to Phytophthora 1	*Arabidopsis thaliana*	35%	5.00E-67	88.10%	NP_001189776.1	201
14	CLT3	CRT (chloroquine-resistance transporter)-like transporter 3	*Arabidopsis thaliana*	44%	1.00E-106	66.33%	NP_001078575.1	452

Table 2 Protein sequences in *Theobroma cacao* L. homologous to protein sequences in *Arabidopsis thaliana*.

N°	Symbol	Description	Scientific Name	Query Cover	E value	Per. Ident	Accession	Amino acids
1	RPH1	PREDICTED: uncharacterized protein LOC18613423	*Theobroma cacao*	72%	6.00E-86	82.88%	XP_007050706.2	229
2	FLS2	Hypothetical protein QUC31_012664	*Theobroma cacao*	98%	0.0	55.69%	KAK6268504.1	1167
3	LecRK-I.9	PREDICTED: L-type lectin-domain containing receptor kinase I.8	*Theobroma cacao*	85%	0.0	50.96%	XP_017982159.1	675
4	BZIP60	PREDICTED: bZIP transcription factor 60	*Theobroma cacao*	87%	3.00E-39	44.93%	XP_017981823.1	315
5	NAC089	PREDICTED: NAC domain-containing protein 89	*Theobroma cacao*	91%	7.00E-43	37.27%	XP_007015733.2	407
6	AT5G65600	Concanavalin A-like lectin protein kinase family protein	*Theobroma cacao*	98%	0.0	53.04%	EOX91877.1	818
7	PHOS32	PREDICTED: universal stress protein PHOS32 isoform X1	*Theobroma cacao*	89%	2.00E-101	74.36%	XP_017980543.1	222
8	AT1G22810	AP2/ERF domain - like 10	*Theobroma cacao*	90%	1.00E-47	53.42%	WRX30759.1	153
9	AT5G43930	Transducin family protein / WD-40 repeat family protein isoform 3	*Theobroma cacao*	94%	0.0	55.10%	EOY06258.1	715
10	CLT2	Hypothetical protein SCA6_008887	*Theobroma cacao*	86%	2.00E-162	68.18%	KAK6245797.1	425
11	UMAMIT36	Hypothetical protein SCA6_007450	*Theobroma cacao*	99%	2.00E-165	62.47%	KAK6242061.1	368
12	RBK1	ROP binding protein kinases 1, putative	*Theobroma cacao*	94%	0.0	63.88%	EOX91923.1	464
13	PHOS34	PREDICTED: universal stress protein PHOS32 isoform X1	*Theobroma cacao*	90%	4.00E-98	71.67%	XP_017980543.1	222
14	CLT3	CRT-like transporter 3 isoform 1	*Theobroma cacao*	89%	4.00E-147	62.80%	EOY32881.1	447

Comparative analysis: *Arabidopsis thaliana* vs. *Theobroma cacao L.* and *Herrania umbratica*

A total of 14 proteins from *Arabidopsis thaliana* associated with pathogen defense were compared. Thirteen orthologous proteins were identified as shared among the three species. No exclusive sequences were detected as shared only between *Arabidopsis thaliana* and *Theobroma cacao*, nor between *Arabidopsis thaliana* and *Herrania umbratica*. However, *Theobroma cacao* and *Herrania umbratica* share 17 156 sequences, demonstrating a high degree of similarity between these closely related species (Figure 1).

This analysis confirmed the presence of 13 shared orthologous sequences between *Arabidopsis thaliana* and *T. cacao*., indicating strong conservation of defense-associated proteins in both species. Following this analysis, only the small group of 13 proteins was evaluated, and similar to the findings reported by Castro-Bustos (2023[1]), no exclusive proteins were found outside this set, reinforcing the relevance of these sequences as central elements in the conserved immune response.

Analyses between *Arabidopsis thaliana* and *T. cacao*, 13 shared orthologous proteins were identified, with no proteins exclusive to either species within the analyzed dataset (Figure 2). The presence of these conserved proteins suggests that several defense-related mechanisms identified in *Arabidopsis thaliana* may also be present in *T. cacao*. This functional conservation supports the use of *Arabidopsis* as a reference model for identifying candidate genes associated with defense responses in cacao. The conservation of orthologous proteins across plant species is frequently associated with key immune signaling pathways involved in pathogen perception and downstream defense responses (Ding *et al*., 2022). Therefore, the identification of these shared proteins provides a useful framework for exploring resistance-related mechanisms in cacao through comparative bioinformatic approaches.

[1] Castro-Bustos, S. (2023). *Caracterización de la interacción de la proteína AtGRDP2 de Arabidopsis thaliana con proteínas implicadas en procesos post-transcripcionales* (Tesis para obtener el grado de Doctora en Ciencias en Biología Molecular). San Luis Potosí, S. L. P., México: Instituto Potosino de Investigación Científica y Tecnológica, A.C.

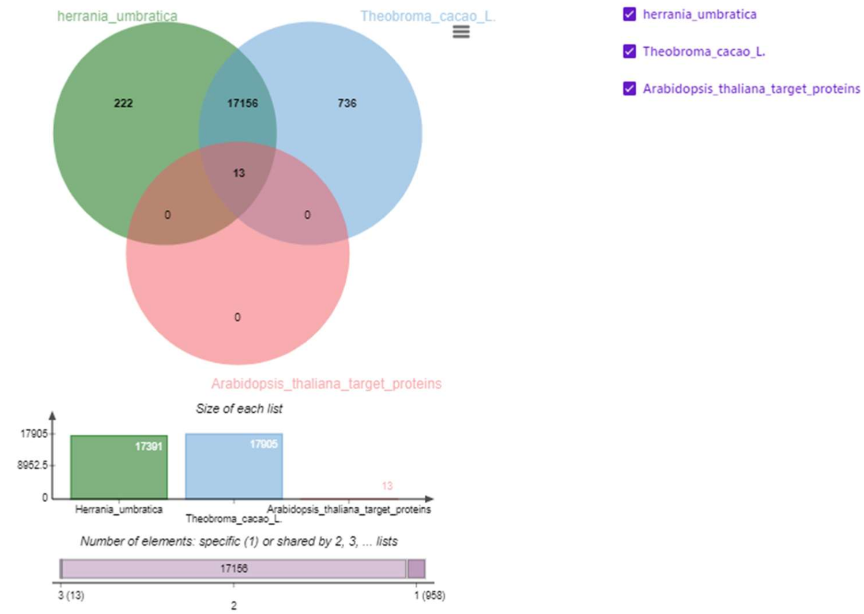

Figure 1. Venn diagram of the first comparative analysis between the studied species: *Arabidopsis thaliana*, *Theobroma cacao* L. and *Herrania umbratica.*

Comparative Analysis: *Theobroma cacao* L. vs. *Herrania umbratica*

In the specific analysis of *T. cacao* proteins homologous to those identified in *Arabidopsis thaliana* and compared with the *Herrania umbratica* genome, 13 shared proteins were identified, with no proteins exclusive to *T. cacao* found within this group. Although this set represents a small percentage of the total proteins identified in *Herrania umbratica*, its presence suggests the functional conservation of specific defense pathways between these species of the same genus (Figure 3). This coincidence could indicate that key mechanisms associated with pathogen perception and signal transduction remain evolutionarily stable among closely related species.

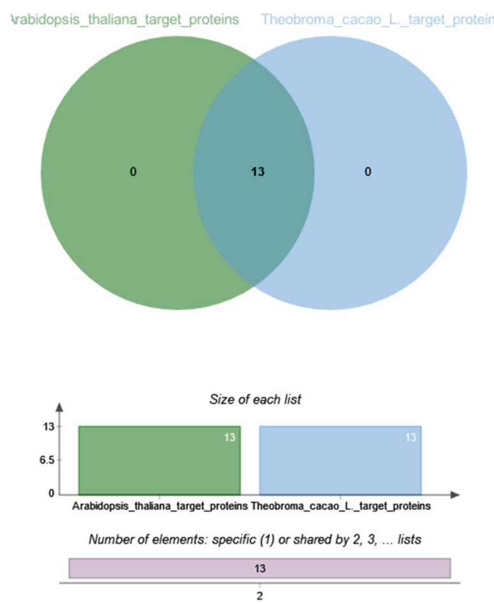

Figure 2. Venn diagram of the first comparative analysis between the studied species: *Arabidopsis thaliana*, *Theobroma cacao* L.

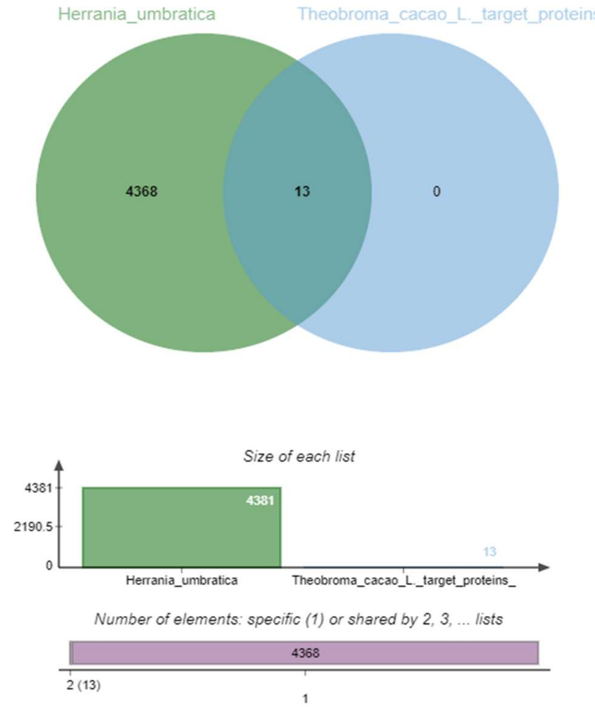

Figure 3. Venn diagram of the comparative analysis between the studied species: *Theobroma cacao* L. and *Herrania umbratical*.

In fact, studies in tomatoes show that various proteins linked to defense signaling pathways, particularly MAPKs, exhibit similar conservation patterns, reinforcing this hypothesis, that these regulatory modules fulfill essential and highly preserved functions in higher plants (Téllez de Pablos, 2021[2]).

Common Proteins in Comparisons

In comparisons between *Arabidopsis thaliana*, *T. cacao*, and *Herrania umbratica*, several conserved proteins were identified that could represent key components in defense mechanisms against *Phytophthora* spp. Among the proteins present in the three comparative analyses, the following stand out: WAT1-related protein At1g70260, Universal stress protein PHOS32, bZIP transcription factor 60, Receptor-like cytosolic serine/threonine-protein kinase RBK1, Protein CLT2 (chloroplast), NAC domain-containing protein 89 (NAC089), Protein RESISTANCE TO *PHYTOPHTHORA* 1 (RPH1, chloroplast), and the LRR-like receptor FLS2. Their conservation suggests an evolutionarily established function in the response to biotic stress (Table 3).

Some key regulators of the endoplasmic reticulum stress response linked to plant immunity or involved in programmed cell death induced by *Phytophthora* signals (Ai *et al*., 2021) suggest that cellular homeostasis and internal stress responses are also part of the conserved defense system. Taken together, these results indicate that the three species share core signaling components for the recognition, transduction, and management of infection-associated stress.

Intracluster Phylogeny of Proteins Involved in Pathogen Defense Mechanisms

Phylogenetic analysis of protein sequences compared between *Arabidopsis thaliana*, *T.* cacao, and *Herrania umbratica* revealed clusters that reflect both conserved and divergent evolutionary relationships among these species. In the first analysis (Figure 4), clades were identified as group orthologous proteins with varying levels of divergence. For example, the *Arabidopsis thaliana* protein NP_001330322.1 showed high divergence (0.30724) from orthologous proteins in *T.* cacao and *Herrania umbratica*, while the latter showed low evolutionary distance from each other (<0.042), indicating recent evolution. Other clades, such as those grouped by NP_173695.1, NP_001190820.1, and NP_174998.1, reflect similar patterns of moderate conservation or pronounced divergence.

[2] Téllez de Pablos, J. (2021). *La proteína quinasa YODA regula respuestas de inmunidad y la resistencia a patógenos de Arabidopsis thaliana y tomate (Solanum lycopersicum)*. Tesis para obtener el grado de Doctor. Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas. Universidad Politécnica de Madrid. Disponible en https://oa.upm.es/68842/

Table 3. Comparison of proteins obtained in clusters related to pathogen tolerance.

Comparative Analysis 1			Comparative Analysis 2			Comparative Analysis 3		
Swiss-Prot Hit	N° proteins	Protein	Swiss-Prot Hit	N° proteins	Proteins	Swiss-Prot Hit	N° proteins	Proteins
Q9LXA5	7	L-type lectin-domain containing receptor kinase IX.1	Q9LSL5	2	L-type lectin-domain containing receptor kinase IX.2	Q9LXA5	5	L-type lectin-domain containing receptor kinase IX.1
F4I5D5	6	WAT1-related protein At1g70260	F4I5D5	2	WAT1-related protein At1g70260	F4I5D5	4	WAT1-related protein At1g70260
Q8VYN9	5	Universal stress protein PHOS32	Q8VYN9	3	Universal stress protein PHOS32	Q8VYN9	2	Universal stress protein PHOS32
Q9C7S0	5	bZIP transcription factor 60	Q9C7S0	2	bZIP transcription factor 60	Q9C7S0	3	bZIP transcription factor 60
Q8H1D6	5	Receptor-like cytosolic serine/threonine-protein kinase RBK1	Q8H1D6	2	Receptor-like cytosolic serine/threonine-protein kinase RBK1	Q8H1D6	4	Receptor-like cytosolic serine/threonine-protein kinase RBK1
A1L4X0	3	Protein CLT2, chloroplastic	A1L4X0	2	Protein CLT2, chloroplastic	A1L4X0	2	Protein CLT2, chloroplastic
Q9ZU82	3	Protein RESISTANCE TO PHYTOPHTHORA 1, chloroplastic	Q9ZU82	2	Protein RESISTANCE TO PHYTOPHTHORA 1, chloroplastic	Q9ZU82	2	Protein RESISTANCE TO PHYTOPHTHORA 1, chloroplastic
Q94F58	3	NAC domain-containing protein 89	Q94F58	2	NAC domain-containing protein 89	Q94F58	2	NAC domain-containing protein 89
Q9FL28	3	LRR receptor-like serine/threonine-protein kinase FLS2	Q9FL28	2	LRR receptor-like serine/threonine-protein kinase FLS2	Q9FL28	2	LRR receptor-like serine/threonine-protein kinase FLS2
Q9M1G3	3	Probable L-type lectin-domain containing receptor kinase I.6	Q9LSR8	2	L-type lectin-domain containing receptor kinase I.9	Q9M1G3	2	Probable L-type lectin-domain containing receptor kinase I.6
Q9C9I8	3	Ethylene-responsive transcription factor ERF020	O80542	2	Ethylene-responsive transcription factor ERF019	O80542	2	Ethylene-responsive transcription factor ERF019

In the second analysis (Figure 5), which exclusively compares *Arabidopsis thaliana* and *T.* cacao, greater diversity in evolutionary relationships was observed. Some clades, such as those grouped by NP_001190820.1 and KAK6245797.1, reflect high conservation (divergence ~0.18), while others, such as those containing NP_173695.1, exhibit older divergences (>0.4). Recent gene duplications are evident in T. cacao, as in WRX30759.1 and EOX91923.1, with divergences close to zero. Internal distances within clades range from high to moderate conservation.

In the third analysis (Figure 6), between *T. cacao* and *Herrania umbratica*, high conservation between pairs of orthologous proteins was evident. Notable examples include XP_021283978.1 and XP_021283979.1 in *Herrania umbratica* with no divergence, and their orthologs in *Theobroma cacao* with divergences <0.04. Subclades with recent gene duplications and low evolutionary divergence were identified, such as those including XP_021278361.1, XP_021288329.1, and XP_021284387.1 for *Herrania umbratica* and their counterparts in *T. cacao*. Internal distances generally reinforce high conservation between homologous proteins.

In this study, 91 nucleotide sequences were identified in *Arabidopsis thaliana*, of which 20 showed potential involvement in defense mechanisms against *Phytophthora* spp. Using BLASTX similarity analyses, these sequences were translated and found to exhibit high identity percentages (greater than 65%, reaching up to 100%), suggesting functional homology with previously characterized genes. Extremely low E-values support the relevance of these matches, confirming the accuracy of the methodology used. These findings are consistent with previous studies in *T. cacao* and other species that have employed BLAST as a key tool for identifying conserved defense-related genes (Mucherino-Munoz, de Melo, Santana, Luz, and Corrêa, 2021; Winters *et al*., 2024).

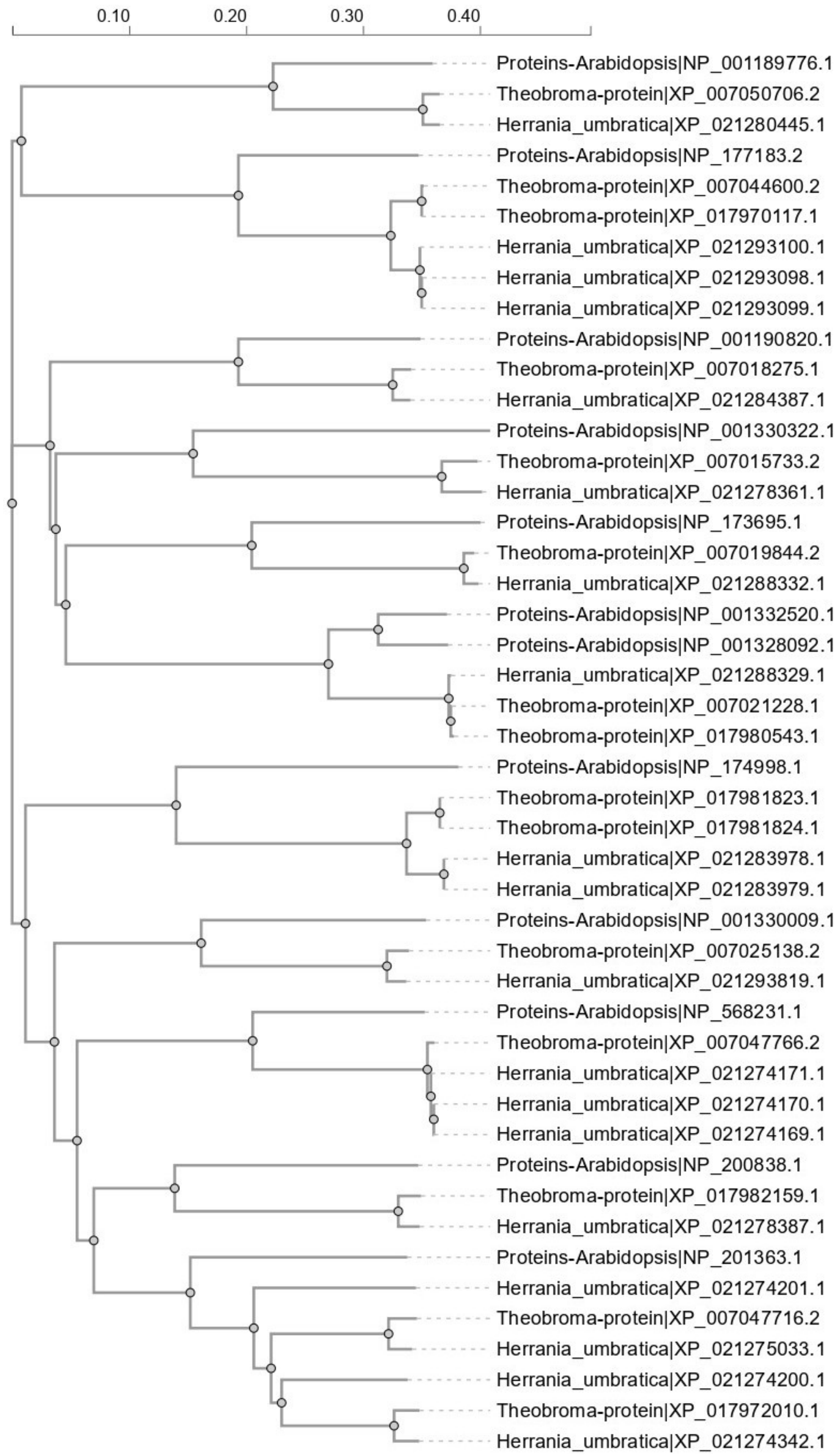

Figure 4. Intra-cluster phylogeny of the first comparative analysis: *Arabidopsis thaliana* vs. *Theobroma cacao* L. and *Herrania umbratica*.

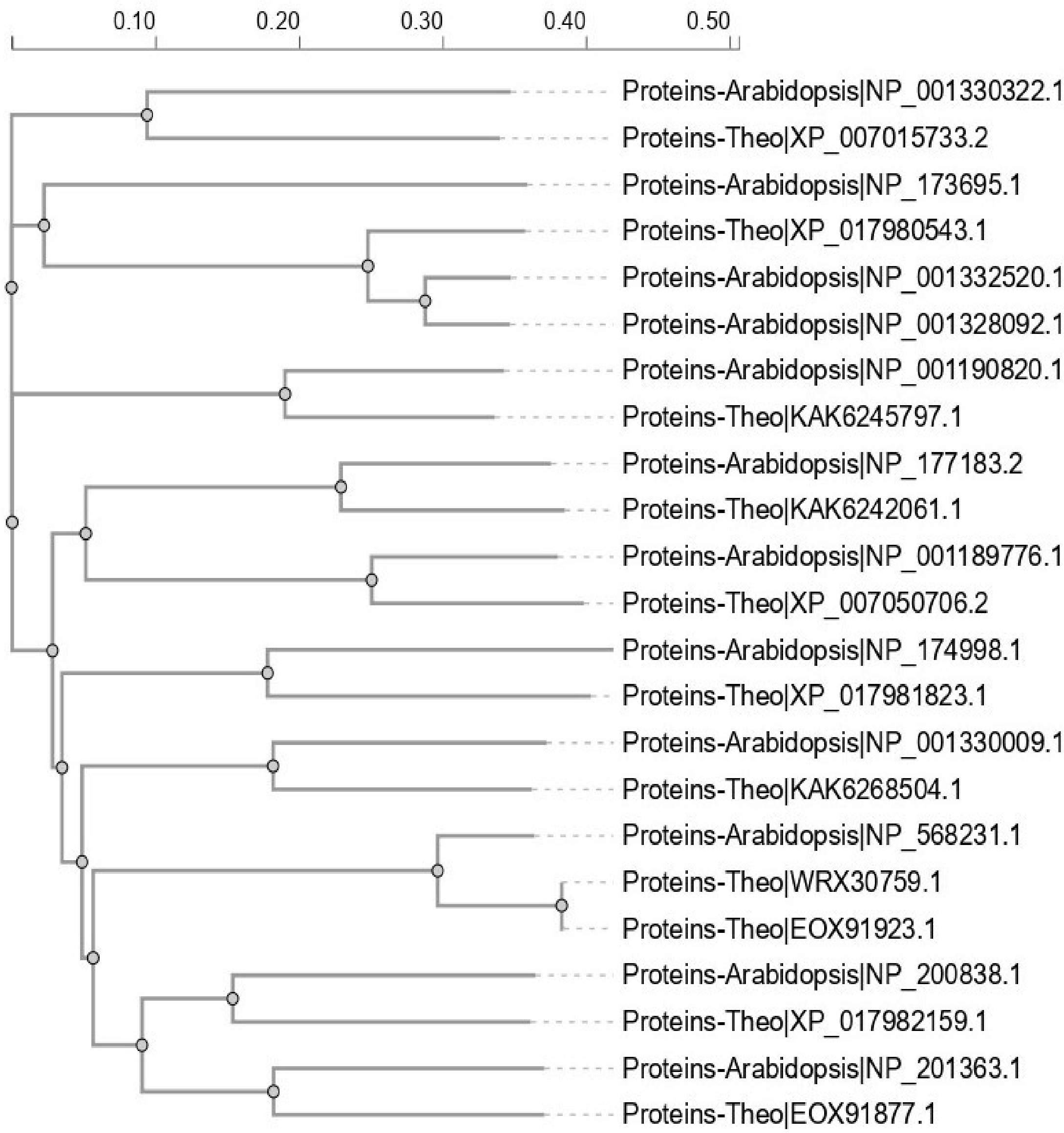

Figure 5. Intra-cluster phylogeny of second comparative analysis: *Arabidopsis thaliana* vs. *Theobroma cacao* L.

The similarity analysis between *Arabidopsis thaliana* and *T. cacao* enabled the identification of relevant orthologs, with coverages ranging from 72% to 99% and E-values that reinforce confidence in their biological significance. The observed identity values, although variable (37%–88.10%), reflect both functional similarity and evolutionary divergence, which is expected between distantly related species. Tools such as BLAST also allowed the detection of homologies with proteins from *Phytophthora* nicotianae and other pathogenic species (Armitage *et al*. 2018; Yuan *et al*. 2021).

The comparison through ortholog clusters among *Arabidopsis thaliana*, *T. cacao*, and *Herrania umbratica* yielded 13 shared proteins, showing evolutionary conservation in genes associated with responses to oomycetes. Among them, the WAT1-related protein At1g70260, Resistance to *Phytophthora* 1 (RPH1), and the kinase RBK1 stand out, all of which take part in plant defense pathways. The WAT1 protein, for instance, negatively regulates tolerance to biotrophic pathogens; its silencing enhances resistance, while its overexpression reduces it (Gao *et al*. 2023; Pan *et al*. 2016). WAT1 in defense responses may still be relevant in the context of *Phytophthora* infection, particularly during the early stages of host colonization when the pathogen behaves in a biotrophic manner (Truman, Bennett, Kubigsteltig, Turnbull, and Grant, 2013). RBK1 functions as a receptor-like kinase that becomes activated during infection, triggering immune signaling cascades (Ménesi, Klement, Ferenc, and Fehér, 2021; Molendijk *et al*. 2008).

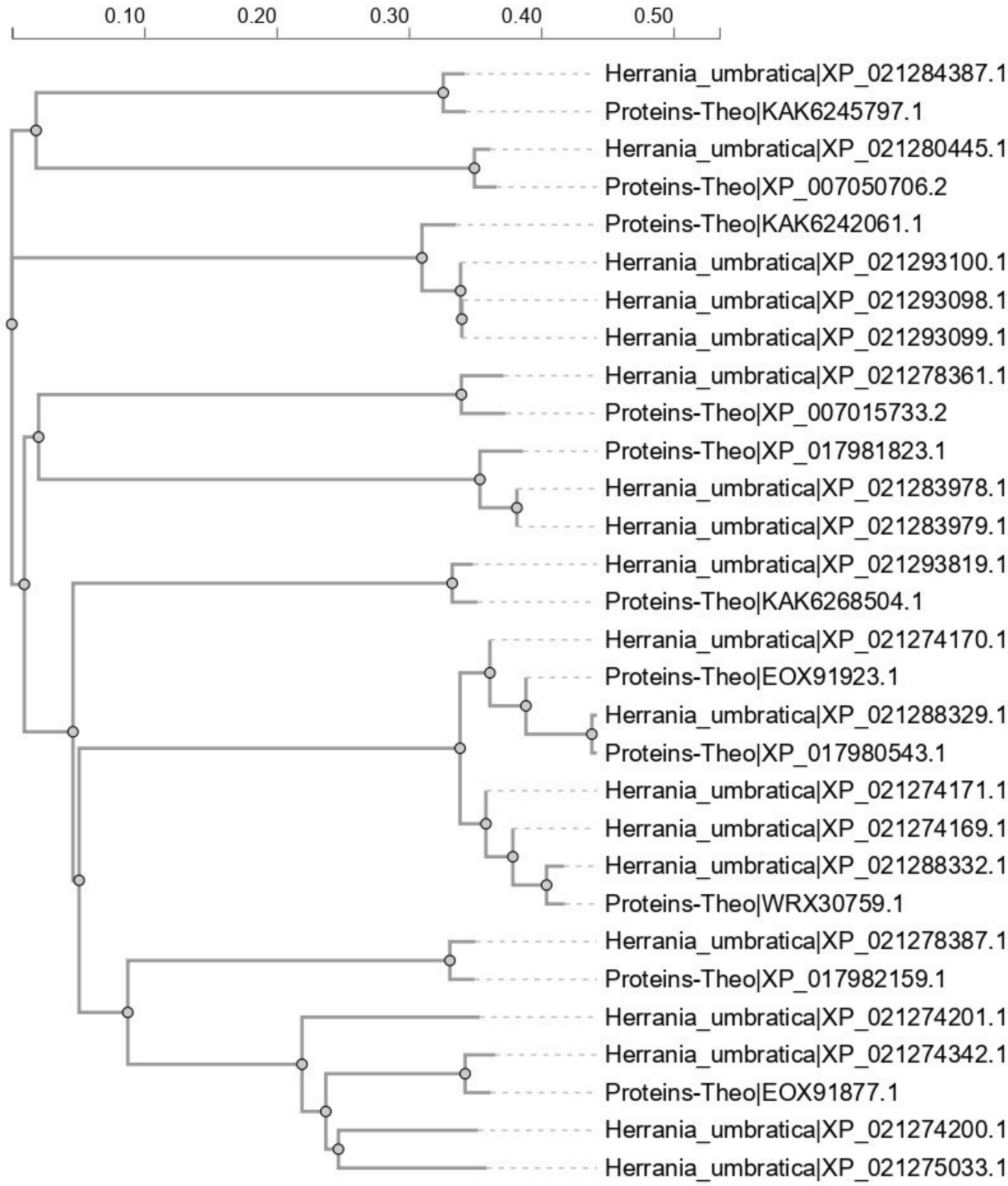

Figure 6. Intra-cluster phylogeny of the third Comparative Analysis: *Theobroma cacao* L. vs. *Herrania umbratical*.

In turn, RPH1 is localized in chloroplasts, where it participates in the production of reactive oxygen species (ROS) and in the regulation of defense-related gene expression; its absence increases susceptibility to *Phytophthora brassicae* (Belhaj, Lin, and Mauch 2009).

The phylogenetic analysis provided a complementary evolutionary perspective. Clades were observed contain highly conserved orthologous proteins between *T. cacao* and *Herrania umbratica*, which were more divergent relative to *Arabidopsis thaliana*. For example, the cluster of the NAC089 protein showed an evolutionary separation between *A. thaliana* and the Malvaceae species. NAC089 is crucial for regulating programmed cell death (PCD) under endoplasmic reticulum stress and is essential for containing *Phytophthora* infections (Ai *et al*., 2021; Yang *et al*. 2014).

Another relevant case is CLT2, a chloroplastic protein involved in glutathione homeostasis, essential for antioxidant defense and ROS detoxification. The lower divergences between its orthologs in *T. cacao* and *Arabidopsis thaliana* suggest strong functional conservation (Maughan *et al*. 2010; Noctor *et al*. 2012).

In contrast, some *Arabidopsis thaliana* proteins, such as NP_174998.1, showed significant divergence, reflecting species-specific adaptations. On the other hand, *Herrania umbratica* proteins XP_021283978.1 and XP_021283979.1 exhibit recent gene duplication and high similarity to XP_017981823.1 from *T. cacao*, within the cluster of the bZIP60 transcription factor. This factor participates in the response to misfolded proteins in the endoplasmic reticulum during infection, helping manage cellular damage induced by *Phytophthora* (Iwata, Fedoroff, and Koizumi, 2009; Qiang *et al*. 2021).

Although the present study provides insights into proteins potentially involved in resistance mechanisms against *Phytophthora* spp., it is important to recognize the limitations of in silico analyses. The identification of candidate proteins was based on sequence similarity and orthology inference, which represent computational predictions rather than direct evidence of biological function. Therefore, the roles proposed for these proteins should be experimentally validated through functional studies such as gene expression analyses or transgenic approaches in *T.* cacao (Gabaldón and Koonin, 2013).

Nevertheless, it is noteworthy that cacao genotypes exhibit cross-resistance to different *Phytophthora* species, suggesting the existence of shared immune mechanisms that act broadly against this group of pathogens (Ali *et al*. 2017). This property reinforces the importance of studying conserved genes, not only to understand immune responses in cacao but also for their potential application in breeding programs for other species (Jazayeri, Cruzatty, and Villamar, 2019; Villamar-Torres, Oviedo, Torres, Zambrano, and Jazayeri, 2022; Villamar-Torres *et al*., 2024, 2025).

CONCLUSIONS

The sequence similarity analysis using BLAST revealed high identity among proteins, with values exceeding 65% and exact matches of 100%, supporting their biological relevance. The comparative study among *Arabidopsis thaliana*, *Theobroma cacao*, and *Herrania umbratica* allowed the identification of 13 key orthologous proteins, highlighting significant evolutionary conservation against pathogens such as *Phytophthora* spp. Among these, the roles of proteins such as WAT1-related, Resistance to *Phytophthora* 1 (RPH1), and RBK1 in plant immune responses stand out.

The phylogenetic analysis complemented these findings by revealing both divergence and functional conservation among species, suggesting species-specific adaptations to different types of biotic stress. Proteins such as NAC089 and CLT2 showed high structural and functional conservation, while bZIP60 was identified as an essential regulator in the adaptive response to endoplasmic reticulum stress during infection. These results provide valuable insights for future breeding strategies aimed at strengthening resistance to *Phytophthora* in cultivated species.

ETHICS STATEMENT

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF SUPPORTING DATA

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FINANCING

Not applicable.

AUTHORS' CONTRIBUTION

Conceptualization: D.B.C.Q. and R.O.V.T.; Methodology: D.B.C.Q.; Validation: A.V.C.M. and R.O.V.T.; Formal analysis: M.E.R.R.; Investigation: D.B.C.Q.; Data curation: D.B.C.Q. and R.O.V.T.; Writing – original draft preparation: D.B.C.Q. and R.O.V.T.; Writing – review and editing: D.B.C.Q., R.O.V.T., A.V.C.M., and M.E.R.R.; Visualization: M.E.R.R.; Supervision: R.O.V.T.

ACKNOWLEDGMENTS

Thanks to the State Technical University of Quevedo for partially supporting the financing of this research and to the Master's Program in Agricultural Biotechnology at UTEQ.

REFERENCES

Ai, G., Zhu, H., Fu, X., Liu, J., Li, T., Cheng, Y., ... & Jing, M. (2021). Phytophthora infection signals-induced translocation of NAC089 is required for endoplasmic reticulum stress response-mediated plant immunity. *The Plant Journal, 108*(1), 67-80. https://doi.org/10.1111/TPJ.15425

Abad, K. L. B., Cevallos, H. V., Montealegre, V. J. G., & Romero, H. C. (2021). Análisis de las exportaciones del cacao ecuatoriano en grano en el periodo 2008 al 2018. *Revista Metropolitana de Ciencias Aplicadas, 4*(1), 147-155.

Ali, S. S., Shao, J., Lary, D. J., Strem, M. D., Meinhardt, L. W., & Bailey, B. A. (2017). Phytophthora megakarya and *P. palmivora*, causal agents of black pod rot, induce similar plant defense responses late during infection of susceptible cacao pods. *Frontiers in Plant Science, 8*, 169. https://doi.org/10.3389/fpls.2017.00169

Argout, X., Fouet, O., Wincker, P., Gramacho, K., Legavre, T., Sabau, X., ... & Lanaud, C. (2008). Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. *BMC Genomics, 9*(1), 512. https://doi.org/10.1186/1471-2164-9-512

Armitage, A. D., Lysøe, E., Nellist, C. F., Lewis, L. A., Cano, L. M., Harrison, R. J., & Brurberg, M. B. (2018). Bioinformatic characterisation of the effector repertoire of the strawberry pathogen Phytophthora cactorum. *PloS One, 13*(10), e0202305. https://doi.org/10.1371/journal.pone.0202305

Baruah, I. K., Ali, S. S., Shao, J., Lary, D., & Bailey, B. A. (2022). Changes in gene expression in leaves of cacao genotypes resistant and susceptible to *Phytophthora palmivora* infection. *Frontiers in Plant Science, 12*, 780805. https://doi.org/10.3389/fpls.2021.780805

Belhaj, K., Lin, B., & Mauch, F. (2009). The chloroplast protein RPH1 plays a role in the immune response of Arabidopsis to Phytophthora brassicae. *The Plant Journal, 58*(2), 287-298. https://doi.org/10.1111/j.1365-313X.2008.03779.x

Bomfim-Rego, A. P, Mora-Ocampo, I. Y., Pirovani, C. P., Luz, E. D. M. N., & Corrêa, R. X. (2022). Protein level defense responses of Theobroma cacao interaction with Phytophthora palmivora. *Frontiers in Agronomy, 4*, 836360. https://doi.org/10.3389/FAGRO.2022.836360/BIBTEX

Capote-Maínez, N., & Hernández-Fort, C. (2021). *Ómicas y fitopatología*. Madrid, España: Sociedad Española de Fitopatología.

Carranza-Quimi, W. D., Castro, M. B. A., Risco, G. S. C., & Cabezas, Y. K. P. (2020). Evaluación socioeconómica del cultivo de cacao (*Theobroma cacao* L.) en la zona norte de la Provincia de los Ríos. *Journal of Business and Entrepreneurial Studies, 4*(2), 96-106.

Ding, L. N., Li, Y. T., Wu, Y. Z., Li, T., Geng, R., Cao, J., ... & Tan, X. L. (2022). Plant disease resistance-related signaling pathways: recent progress and future prospects. *International Journal of Molecular Sciences, 23*(24), 16200. https://doi.org/10.3390/ijms232416200

Gabaldón, T., & Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics, 14*(5), 360-366.

Gao, X.-X., Tang, Y.-L., Shi, Q.-Y., Wei, Y.-S., Wang, X.-X., Shan, W.-X., ... & Qiang, X.-Y. (2023). Vacuolar processing enzyme positively modulates plant resistance and cell death in response to Phytophthora parasitica infection. *Journal of Integrative Agriculture, 22*(5), 1424-1433. https://doi.org/10.1016/J.JIA.2022.08.124

Guerrero, R., Vera-Saltos, D., Acosta-Farías, M., Mónaco, C., & Palma, R. (2026). Reduction of the Survival of the Soilborne Pathogen *Phytophthora palmivora* on *Theobroma cacao* Pods Through the Application of *Trichoderma* spp. *Terra Latinoamericana, 44,* 1-11. https://doi.org/10.28940/terralatinoamericana.v44i.2435

Hou, Y., Zhai, Y. I., Feng, L. I., Karimi, H. Z., Rutter, B. D., Zeng, L., ... & Ma, W. (2019). A Phytophthora effector suppresses trans-kingdom RNAi to promote disease susceptibility. *Cell host & Microbe, 25*(1), 153-165. https://doi.org/10.1016/j.chom.2018.11.007

Irion, U., & Nüsslein-Volhard, C. (2022). Developmental genetics with model organisms. *Proceedings of the National Academy of Sciences, 119*(30), e2122148119. https://doi.org/10.1073/PNAS.2122148119

Iwata, Y., Fedoroff, N. V., & Koizumi, N. (2008). Arabidopsis bZIP60 is a proteolysis-activated transcription factor involved in the endoplasmic reticulum stress response. *The Plant Cell, 20*(11), 3107-3121. https://doi.org/10.1105/tpc.108.061002

Jazayeri, S. M., Cruzatty, L. C., & Villamar-Torres, R. (2019). Genomic comparison among three Arabidopsis species revealed heavy metal responsive genes. *JAPS: Journal of Animal & Plant Sciences, 29*(2), 1.

Magar, N. D., Shah, P., Harish, K., Bosamia, T. C., Barbadikar, K. M., Shukla, Y. M., ... & Sundaram, R. M. (2022). Gene expression and transcriptome sequencing: Basics, analysis, advances. In F. Uchiumi (Ed.). *Gene expression*. London, UK: IntechOpen. https://doi.org/10.5772/INTECHOPEN.105929

Maughan, S. C., Pasternak, M., Cairns, N., Kiddle, G., Brach, T., Jarvis, R., ... & Cobbett, C. S. (2010). Plant homologs of the Plasmodium falciparum chloroquine-resistance transporter, Pf CRT, are required for glutathione homeostasis and stress responses. *Proceedings of the National Academy of Sciences, 107*(5), 2331-2336. https://doi.org/10.1073/pnas.0913689107

Ménesi, D., Klement, É., Ferenc, G., & Fehér, A. (2021). The Arabidopsis Rho of plants GTPase ROP1 is a potential calcium-dependent protein kinase (CDPK) substrate. *Plants, 10*(10), 2053. https://doi.org/10.3390/plants10102053

Molendijk, A. J., Ruperti, B., Singh, M. K., Dovzhenko, A., Ditengou, F. A., Milia, M., ... & Palme, K. (2008). A cysteine-rich receptor-like kinase NCRK and a pathogen-induced protein kinase RBK1 are Rop GTPase interactors. *The Plant Journal, 53*(6), 909-923. https://doi.org/10.1111/j.1365-313X.2007.03384.x

Mucherino-Munoz, J. J., de Melo, C. A. F., Santana Silva, R. J., Luz, E. D. M. N., & Corrêa, R. X. (2021). Structural and Functional Genomics of the Resistance of Cacao to *Phytophthora* palmivora. *Pathogens, 10*(8), 961. https://doi.org/10.3390/pathogens10080961

NCBI (National Center for Biotechnology Information). (2025). *Basic Local Alignment Search Tool (BLAST)*. Bethesda, MD, USA: National Library of Medicine, National Institutes of Health.

Noctor, G., Mhamdi, A., Chaouch, S., Han, Y. I., Neukermans, J., Marquez-Garcia, B. E. L. E. N., ... & Foyer, C. H. (2012). Glutathione in plants: an integrated overview. *Plant, Cell & Environment, 35*(2), 454-484. https://doi.org/10.1111/J.1365-3040.2011.02400.X

Pan, Q., Cui, B., Deng, F., Quan, J., Loake, G. J., & Shan, W. (2016). RTP 1 encodes a novel endoplasmic reticulum (ER)-localized protein in Arabidopsis and negatively regulates resistance against biotrophic pathogens. *New Phytologist, 209*(4), 1641-1654. https://doi.org/10.1111/nph.13707

Qiang, X., Liu, X., Wang, X., Zheng, Q., Kang, L., Gao, X., ... & Shan, W. (2021). Susceptibility factor RTP1 negatively regulates Phytophthora parasitica resistance via modulating UPR regulators bZIP60 and bZIP28. *Plant Physiology, 186*(2), 1269-1287. https://doi.org/10.1093/PLPHYS/KIAB126

Rodríguez-Polanco, E., & Vera-Rodríguez, A. G. (2015). Identificación y manejo de la pudrición parda de la mazorca (*Phytophthora* sp.) en cacao. Bogotá, Colombia: Corporación Colombiana de Investigación Agropecuaria (CORPOICA). ISBN: 978-958-740-197-4

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular *Systems Biology, 7*(1), 539. https://doi.org/10.1038/msb.2011.75

Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., & Wang, Y. (2023). OrthoVenn3: An integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Research, 51*, 397-403. https://doi.org/10.1093/nar/gkad313

Tarigan, R., Maharijaya, A., & Izzah, N. K. (2021). Snap markers derived from catalase-1 gene sequence used for black pod disease resistance in cacao (*Theobroma cacao* L.). *SABRAO Journal of Breeding and Genetics, 53*(3), 510-526.

Tegenfeldt, F., Kuznetsov, D., Manni, M., Berkeley, M., Zdobnov, E. M., & Kriventseva, E. V. (2025). OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Research, 53*(1), D516-D522. https://doi.org/10.1093/nar/gkae987

Truman, W., Bennett, M. H., Kubigsteltig, I., Turnbull, C., & Grant, M. (2010). Arabidopsis auxin mutants are compromised in systemic acquired resistance and exhibit aberrant accumulation of various indolic compounds. *Plant Physiology, 152*(3), 1562–1573. https://doi.org/10.1104/pp.109.152173

Vega-Jarquín, C. (2017). Un nuevo inicio: sistemas agroforestales con cacao, un legado biocultural para construir el futuro. La *Calera, 17*(29), 87-98.

Villamar-Torres, R., Oviedo-Bayas, B., Torres-Navarreete, Y., Zambrano-Vega, C., & Jazayeri, S. M. (2022). Comparative computational genomic analysis between wild and domesticated species of cacao and coffee. *Journal of Pharmaceutical Negative Results, 13*(3), 535-551. https://doi.org/10.47750/pnr.2022.13.03.082

Villamar-Torres, R. O., Oviedo-Bayas, B., Mestanza-Uquillas, C. A., Guerrero-Chuez, R., Santos, M. D. A., Ghafoor, S. M. H. A., ... & Jazayeri, S. M. (2024). Herrania species: A potential genetic resource for cocoa breeding revealed by a comparative bioinformatic study among close species. *Bragantia, 84*, e20230298.

Villamar-Torres, R. O., Mestanza Uquillas, C. A., Chévez-Vera, H. D., Heredia-Pinos, M. R., Viot, C., & Jazayeri, S. M. (2024). A computational analysis revealed BES1 transcription factor and β-amylase as crosstalk elements in Upland cotton species (*Gossypium* sp.). *Scientia Agropecuaria, 15*(3), 449-60. https://doi.org/10.17268/sci.agropecu.2024.033

Winters, N. P., Wafula, E. K., Knollenberg, B. J., Hämälä, T., Timilsena, P. R., Perryman, M., ... & Guiltinan, M. J. (2024). A combination of conserved and diverged responses underlies Theobroma cacao's defense response to *Phytophthora palmivora*. *BMC Biology, 22*(1), 38. https://doi.org/10.1186/s12915-024-01831-2

Yang, Z. T., Wang, M. J., Sun, L., Lu, S. J., Bi, D. L., Sun, L., ... & Liu, J. X. (2014). The membrane-associated transcription factor NAC089 controls ER-stress-induced programmed cell death in plants. *PLoS Genetics, 10*(3), e1004243. https://doi.org/10.1371/journal.pgen.1004243

Yuan, X. L., Zhang, C. S., Kong, F. Y., Zhang, Z. F., & Wang, F. L. (2021). Genome analysis of *Phytophthora nicotianae* JM01 provides insights into its pathogenicity mechanisms. *Plants, 10*(8), 1620. https://doi.org/10.3390/plants10081620

Zaru, R., Orchard, S., & UniProt Consortium. (2023). UniProt tools: BLAST, align, peptide search, and ID mapping. *Current Protocols, 3*(3), e697. https://doi.org/10.1002/cpz1.697