



Ciencia y Tecnología

Secretaría de Ciencia, Humanidades, Tecnología e Innovación



Centro de Investigación Científica de Yucatán, A.C.

Posgrado en Ciencias Biológicas

ANÁLISIS *IN SILICO* DE EFECTORES DE
FITOPLASMAS, UN ACERCAMIENTO AL
FITOPLASMA QUE ATACA EL COCOTERO

Tesis que presenta

SARA ELENA VILA LUNA

En opción al título de

DOCTOR EN CIENCIAS

(Ciencias Biológicas: *Opción Biotecnología*)

Mérida, Yucatán, México

2025

CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.
POSGRADO EN CIENCIAS BIOLÓGICAS

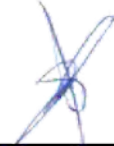


RECONOCIMIENTO

Por medio de la presente, hago constar que el trabajo de tesis de Sara Elena Vila Luna titulado **Análisis *in silico* de efectores de fitoplasmas, un acercamiento al fitoplasma que ataca el cocotero**, fue realizado en la unidad de Biotecnología, en la línea de Biotecnología de microorganismos, en el laboratorio de cocotero del Centro de Investigación Científica de Yucatán, A.C. bajo la dirección de la Dra. Blondy Beatriz Canto Canché y la codirección del Dr. Luis Alfonso Sáenz Carbonell, dentro de la opción de Biotecnología, perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Este trabajo sienta las bases para el control del fitoplasma causante del Amarillamiento letal, enfermedad que causa graves estragos en la economía de la población asociada al cultivo de cocotero, al permitir identificar las proteínas involucradas en la virulencia de este patógeno, facilitando la comprensión de los mecanismos de infección y en consecuencia el control de esta enfermedad.

Atentamente

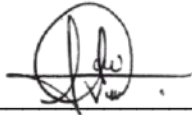


Dr. José Luis Hernández-Stefanoni
Director de Docencia

Mérida, Yucatán, México, a 18 de agosto de 2025

DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en la sección de Materiales y Métodos, los Resultados y Discusión de este documento proviene de las actividades de investigación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que a razón de lo anterior y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica de Yucatán, A.C., y en el mismo tenor, reconozco que si derivaren de este trabajo productos intelectuales o desarrollos tecnológicos, en lo especial, estos se registrarán en todo caso por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.

Firma  _____
Sara Elena Vila Luna

Este trabajo se llevó a cabo en la Unidad de **Biotecnología** del Centro de Investigación Científica de Yucatán, A.C., y forma parte del proyecto titulado “Predicción de efectores no canónicos: trascendiendo los límites de la efectorómica” (FOP16-2021-01 No. 320993) bajo la dirección de la **Dra. Blondy Beatriz Canto Canché y el Dr. Luis Sáenz Carbonell**.

AGRADECIMIENTOS

Agradezco a La Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por la beca otorgada (550897).

A mis asesores la Dra. Blondy Canto Canché y el Dr. Luis Sáenz Carbonell por su orientación, asesoramiento y apoyo para la realización de este trabajo

A los miembros de mi comité tutorial el Dr. Felipe Sánchez Teyer y el Dr. Roberto Vázquez Euán por sus aportaciones para mejorar mi trabajo de tesis

A los miembros de mi comité predoctoral y comité revisor de tesis el Dr. Felipe Sánchez Teyer, el Dr. Roberto Vázquez Euán, el Dr. Ignacio Islas Flores, la Dra. Marcela Gamboa, el Dr. Jairo Cristóbal Alejo y la Dra. Ileana Echevarría Machado

A la Dra. Gisel Carreón Anguiano por orientarme en el estudio in silico para la realización de este proyecto de tesis y sus importantes aportaciones para esta investigación

Al M. en C. Iván Córdova Lara por el asesoramiento, paciencia y apoyo técnico en la parte experimental para poder llevar a cabo este trabajo

Al M. en C. Bartolomé Chí Manzanero Q.E.P.D. por sus aportaciones y orientación en cada seminario

A mis profesores por el conocimiento transmitido en las asignaturas de mi doctorado

Al Centro de Investigación Científica de Yucatán por las instalaciones brindadas para la realización de este trabajo

LISTA DE LOS PRODUCTOS GENERADOS

Artículos.

Novel Insights into Phytoplasma Effectors

Carreón-Anguiano, K. G., Vila-Luna, S. E., Sáenz-Carbonell, L. y Canto-Canché, B. (2023). Novel Insights into Phytoplasma Effectors. *Horticulturae*, 9(11), 1228. doi: 10.3390/horticulturae9111228. Manuscrito publicado.

PhyEffector, the First Algorithm That Identifies Classical and Non-Classical Effectors in Phytoplasmas.

Carreón-Anguiano KG, Vila-Luna SE, Sáenz-Carbonell L, Canto-Canche B. (2023). PhyEffector, the First Algorithm That Identifies Classical and Non-Classical Effectors in Phytoplasmas. *Biomimetics (Basel)*, 8(7):550. doi: 10.3390/biomimetics8070550. Manuscrito publicado.

The draft genome of 'Candidatus Phytoplasma palmae' strain LY-C2, the phytoplasma associated with coconut lethal yellowing disease, reveals insights into its biological characteristics

Vila-Luna SE, Carreón-Anguiano KG, Córdova-Lara I, Sáenz-Carbonell L, Oropeza-Salín C, Canto-Canché B. (2025). The draft genome of 'Candidatus Phytoplasma palmae' strain LY-C2, the phytoplasma associated with coconut lethal yellowing disease, reveals insights into its biological characteristics. *World Journal of Microbiology and Biotechnology*, 41(7):242. doi: 10.1007/s11274-025-04418-3. Manuscrito publicado.

Review. Genetics characteristics of phytoplasmas

Vila-Luna, Sara E.; Canto-Canche, Blondy; Cordova-Lara, Iván; Sáenz-Carbonell, Luis. (2024). Genetics characteristics of phytoplasmas. *Phytopathogenic Mollicutes*, 14(1). doi: 10.5958/2249-4677.2024.00001.5. Manuscrito publicado.

Review. Algorithms for Effector Prediction in Plant Pathogens and Pests: Achievements and Current Challenges.

Santos-Briones, C. D. I., Carreón-Anguiano, K. G., Vila-Luna, S. E., Todd, J. N. A., Islas-Flores, I., Sáenz-Carbonell, L., Gamas-Trujillo, P. A. y Canto-Canché, B. (2024). Algorithms for Effector Prediction in Plant Pathogens and Pests: Achievements and Current Challenges. *Microbiology Research*, 15(4), 2162-2183. doi: 10.3390/microbiolres15040145. Manuscrito publicado.

Capítulo de libro. Libro de divulgación científica para jóvenes

De los Santos Briones, C.; Gómez Tah, J. R.; Vila Luna, S. E.; Todd, J. N. A.; Carreón Anguiano, K. G. (2024). Biotecnología para todos y todas: usos cotidianos. En E. I. Espana Gamboa, & L. O. Polanco Vasquez (Comps.), *Descubriendo mi Talento 2023* (1era. Ed., pp. 90-100). Centro de Investigacion Cientifica de Yucatan, A.C.

Participación en Congresos.

XXVI Congreso Internacional y LI Congreso Nacional de la Sociedad Mexicana de Fitopatología A.C. 1^{er} Encuentro nacional de estudiantes de postgrado en salud de plantas. 2024. Realizado por la Sociedad Mexicana de Fitopatología y la Universidad Autónoma Chapingo. Ponencia

XII Congreso de Biotecnología y Bioingeniería del Sureste y III Simposio aprovechamiento integral de pescados, mariscos y biomasa de algas. 2024. Realizado por la sociedad Mexicana de Biotecnología y Bioingeniería Delegación Sur Sureste. Cartel

1er Congreso Nacional de Biotecnología Vegetal y Nanotecnología 2025. Realizado por la Universidad Nacional Autónoma de México (UNAM) y la Escuela Nacional de Estudios Superiores, Unidad León. Ponencia

Reconocimiento

Segundo lugar en la modalidad cartel con el trabajo “Caracterización de efectores de fitoplasmas”. En el XII Congreso de Biotecnología y Bioingeniería del Sureste y III Simposio aprovechamiento integral de pescados, mariscos y biomasa de algas. 2024. Realizado por la sociedad Mexicana de Biotecnología y Bioingeniería Delegación Sur Sureste.

Participación en actividades de retribución social

Talento CICY-Preparatoria 2022. Curso: Los habitantes invisibles ¿Amigos o enemigos? Realizado por el Centro de Investigación científica de Yucatán (CICY). Participación en la impartición del curso

Webinario: "Efectores en la fitopatología y más allá" como parte del proyecto "Trascendiendo los límites de la efectorómica" de la convocatoria Ciencia de frontera, modalidad: paradigmas y controversias de la ciencia 2022. Realizado por el Centro de Investigación científica de Yucatán (CICY). Participación como conferencista.

Semana Nacional del Conocimiento, Yucatán 2022: Ciencias Básicas para el Desarrollo Sostenible. Actividad "Los efectores y el mundo microbiano". Realizado por el Gobierno del Estado de Yucatán a través de la Secretaría de Investigación, Innovación y Educación Superior (SIIES). Participación como ponente.

Talento CICY–Secundaria 2023. Curso: "Biotecnología para todos: usos cotidianos". Realizado por el Centro de Investigación científica de Yucatán (CICY). Participación en la impartición del curso

Evento de divulgación CICY Casa Abierta 2023 / 2024. Actividad Los fitoplasmas: patógenos de plantas. Realizado por el Centro de Investigación científica de Yucatán (CICY). Participación como expositor

DEDICATORIAS

A Dios. ¿Qué pagaré a Jehová por todos sus beneficios hacia mí? (Salmo 116:12).
“Bendice, alma mía, a Jehová, y no olvides ninguno de sus beneficios”
(Salmo 103:2).

A mi familia, por su amor y motivación para seguir adelante con mi carrera profesional.

A mis amigos, por su comprensión y apoyo en los momentos que más necesite.

ÍNDICE

LISTADO DE FIGURAS	vii
LISTADO DE TABLAS.....	ix
ABREVIATURAS	xi
RESUMEN.....	xii
ABSTRACT	xiv
INTRODUCCIÓN.....	1
CAPÍTULO I	3
ANTECEDENTES	3
1. COCOTERO	3
1.1 AMARILLAMIENTO LETAL DEL COCOTERO (ALC)	5
1.2 FITOPLASMAS	7
1.2.1 GENOMAS DE FITOPLASMAS	10
1.2.2 FITOPLASMAS-PLANTAS	13
1.2.3 FITOPLASMAS E INSECTO VECTOR	15
1.2.4 MODO DE TRANSMISION DE LOS FITOPLASMAS.....	16
1.3 EFECTORES.....	18
1.3.1 PAPEL DE LOS EFECTORES Y SU RECONOCIMIENTO POR PARTE DE LAS PLANTAS.....	18
1.3.2 EFECTORES IDENTIFICADOS EN LOS FITOPLASMAS	20

1.4 UNIDADES POTENCIALMENTE MÓVILES (PMU)	29
HIPÓTESIS	32
JUSTIFICACIÓN	32
OBJETIVO GENERAL	33
OBJETIVOS ESPECÍFICOS	33
ESTRATEGIA EXPERIMENTAL	34
CAPÍTULO II	35
ARTÍCULO. NOVEL INSIGHTS INTO PHYTOPLASMA EFFECTORS.....	35
2.1. INTRODUCTION	36
2.2. MATERIALS AND METHODS.....	39
2.2.1. PROTEIN DATASET	39
2.2.2. IN SILICO CHARACTERIZATION OF PHYTOPLASMA EFFECTORS.....	40
2.2.3 GENE ONTOLOGY DISTRIBUTION AND FUNCTIONAL ANNOTATION	40
2.2.4 CLASSIFICATION OF EFFECTORS IN TRIBES BASED ON MOTIFS	41
2.2.5 SEARCH FOR SHORT LINEAR MOTIFS IN TRUE PHYTOPLASMA EFFECTORS	41
2.3. RESULTS	42
2.3.1 PROTEIN DATABASES.....	42
2.3.2 CHARACTERIZATION OF PHYTOPLASMA EFFECTORS.....	45
2.3.3 FUNCTIONAL CATEGORIES OF THE PHYTOPLASMA EFFECTORS.....	46
2.3.4 FUNCTIONAL DOMAINS IN PHYTOPLASMA EFFECTORS	48

2.3.5 IDENTIFICATION OF EFFECTOR TRIBES IN PHYTOPLASMAS: CLASSIFICATION BY PROTEIN MOTIFS.....	49
2.3.6 SHORT LINEAR MOTIFS IN PHYTOPLASMA EFFECTORS.....	56
2.4. DISCUSSION	60
CAPÍTULO III	71
ARTÍCULO 2. PHYEFFECTOR, THE FIRST ALGORITHM THAT IDENTIFIES CLASSICAL AND NON-CLASSICAL EFFECTORS IN PHYTOPLASMAS	71
3.1 INTRODUCTION	72
3.2 MATERIALS AND METHODS.....	74
3.2.1. CREATION OF DATABASES	74
3.2.2 IN SILICO CHARACTERIZATION OF EFFECTORS FROM PHYTOPLASMAS....	75
3.2.3 CHARACTERIZATION OF DIFFERENT PIPELINES TO PREDICT EFFECTORS IN PHYTOPLASMAS: CONSTRUCTION OF PHYEFFECTOR ALGORITHM.....	75
3.2.4 VALIDATION OF PHYEFFECTOR ALGORITHM	77
3.3 RESULTS.....	80
3.3.1 CONSTRUCTION OF POSITIVE DATASET	80
3.3.2 CHARACTERIZATION OF PHYTOPLASMA EFFECTORS.....	81
3.3.3 COMPARISON OF MULTIPLE PIPELINES TO IDENTIFY PHYTOPLASMA EFFECTORS	82
3.3.4 PHYEFFECTOR PIPELINE	86
3.3.5 PHYEFFECTOR PERFORMANCE: PREDICTION OF EFFECTORS ON A TESTING DATASET AND ON PHYTOPLASMA GENOMES AND COMPARISON WITH	

LITERATURE	86
3.4 DISCUSSION	91
CAPÍTULO IV	97
SECUENCIACIÓN PARCIAL DEL GENOMA DE ‘ <i>Ca. P. palmae</i> ’, FITOPLASMA ASOCIADO AL ALC	97
4.1 INTRODUCCIÓN	97
4.2 MATERIALES Y MÉTODOS	99
4.2.1. MATERIAL VEGETAL Y EXTRACCIÓN DE ADN	99
4.2.2. SECUENCIACIÓN Y ANOTACIÓN DEL GENOMA DE ALC	101
4.2.3 ANÁLISIS COMPARATIVO CON DIFERENTES GENOMAS DE FITOPLASMAS	102
4.2.4 IDENTIFICACIÓN Y CARACTERIZACIÓN COMPUTACIONAL DE EFECTORES PUTATIVOS EN ‘ <i>Ca. P. palmae</i> ’-ALC	102
4.3 RESULTADOS	103
4.3.1 ENSAMBLAJE DEL GENOMA	103
4.3.2 ANOTACIÓN FUNCIONAL DEL GENOMA PARCIAL DE ‘ <i>Ca. P. palmae</i> ’-ALC	104
4.3.2.1 METABOLISMO ENERGÉTICO	106
4.3.2.2 BIOSÍNTESIS DE PROTEÍNAS	107
4.3.2.3 METABOLISMO DE PURINAS Y PIRIMIDINAS	108
4.3.2.4 ESTRÉS OXIDATIVO	108
4.3.2.5 MAQUINARIA SECRETORA	108

4.3.3 IDENTIFICACIÓN COMPUTACIONAL DE EFECTORES EN ‘ <i>Ca. P. palmae</i> ’-ALC	109
4.3.4 SIMILITUDES Y DIFERENCIAS DEL GENOMA DE ‘ <i>Ca. P. palmae</i> ’-ALC CON OTROS GENOMAS DE FITOPLASMAS	116
4.4 DISCUSIÓN.....	118
CAPÍTULO V	125
CARACTERIZACIÓN DE DOS EFECTORES DE FITOPLASMA DE ALC	125
5.1 INTRODUCCIÓN	125
5.2 MATERIALES Y MÉTODOS	126
5.2.1. OBTENCIÓN DEL MATERIAL VEGETAL	126
5.2.2 EXTRACCIÓN DE ADN PARA LA EVALUACIÓN DE LA PRESENCIA DE FITOPLASMAS	126
5.2.3 CONFIRMACIÓN DE PRESENCIA DE FITOPLASMAS MEDIANTE PCR EN TIEMPO REAL (qPCR)	127
5.2.4 EXTRACCIÓN DE ARN.....	127
5.2.5 DETERMINACIÓN DE LA CANTIDAD DE ARN Y VISUALIZACIÓN	129
5.2.6 LIMPIEZA DEL ARN MEDIANTE DIGESTIÓN ENZIMÁTICA DEL ADN	130
5.2.7 SÍNTESIS DEL ADN COMPLEMENTARIO (ADNc)	130
5.2.8 DISEÑO DE CEBADORES ESPECÍFICOS PARA EFECTORES DEL FITOPLASMA DEL ALC	130
5.2.9 CARACTERIZACIÓN DE DOS EFECTORES CANDIDATOS DE ALC.....	131
5.2.10 AMPLIFICACIÓN DE EFECTORES DEL FITOPLASMA DE ALC MEDIANTE PCR	

DE PUNTO FINAL.....	132
5.2.11 PURIFICACIÓN DE LAS BANDAS AMPLIFICADAS Y ENVÍO A SECUENCIACIÓN	132
5.3 RESULTADOS	133
5.3.1. CONFIRMACIÓN DE FITOPLASMAS EN PALMAS DE <i>Cocos nucifera</i> L.	133
5.3.2 OBTENCIÓN DE ARN EN TALLO, RAÍCES Y HOJAS EN PALMAS CON ALC	134
5.3.3 SELECCIÓN DE EFECTORES CANDIDATOS IDENTIFICADOS EN EL GENOMA DE ‘ <i>Ca. P. palmae</i> ’ PARA EL DISEÑO DE CEBADORES	136
5.3.4 CARACTERIZACIÓN DE 2 EFECTORES CANDIDATOS SELECCIONADOS DE ALC	138
5.3.5 AMPLIFICACIÓN DE EFECTORES DE FITOPLASMA DE ALC	140
5.3.6 SECUENCIAS DE FRAGMENTOS AMPLIFICADOS CON LAS MUESTRAS P1-T Y P2-H	141
5.3.7 CONFIRMACIÓN DE LA AMPLIFICACIÓN DE EFECTORES DE ‘ <i>Ca. P. palmae</i> ’ EN PALMAS CON ALC	143
5.4 DISCUSIÓN.....	145
CAPÍTULO VI.....	149
DISCUSIÓN, CONCLUSIONES GENERALES Y PERSPECTIVAS.....	149
6.1. DISCUSIÓN.....	149
6.2. CONCLUSIONES GENERALES	156
6.3. PERSPECTIVAS	157

BIBLIOGRAFÍA.....	158
--------------------------	------------

LISTADO DE FIGURAS

Figura 1.1 El cocotero “árbol de la vida” y sus productos finales.....	4
Figura 1.2 Síndrome del ALC en orden de aparición.....	6
Figura 1.3 Fitoplasmas en el floema de una planta infectada.....	8
Figura 1.4 Árbol filogenético de algunas especies de fitoplasmas basados en el gen 16SrRNA.	9
Figura 1.5 Síntomas causados por fitoplasmas en distintas especies vegetales.....	15
Figura 1.6 Modo de transmisión del fitoplasma e insecto vector..	17
Fig. 1.7. Síntomas de enanismo y escoba de bruja que causa el efector TENGU.....	22
Fig. 1.8. Efector SAP11 detectado en los núcleos de células de tricoma en secciones de hojas de áster de China infectadas por AY-WB.	24
Fig. 1.9. Efecto del efector SAP54 sobre el desarrollo normal de las flores de <i>A. thaliana</i>	25
Figura 1.10. Modelo de inducción de escoba de bruja por SAP05..	27
Figura 1.11. Modelo de inducción de la filodia por el efector PHYL1.....	28
Figura 1.12. Unidades potencialmente móviles en el cromosoma de AY-WB (PMU1).	30
Figura 1.13 Diagrama de las estrategias experimentales.....	34
Figure 2.1 Gene Ontology classification of the phytoplasma effectors.	47
Figure 2.2 InterProScan functional domains found in phytoplasma effectors.	48
Figure 2.3 Top 50 motifs found in the positive dataset of phytoplasma effectors.....	51

Figure 2.4. Schematic representation of the organization of the motifs in the different effector tribes (effector families) in phytoplasmas..	52
Figure 2.5 Short linear motifs found in phytoplasma effectors..	56
Figure 3.1 PhyEffector workflow for the prediction of effectors in phytoplasma proteome..	76
Figura 4.1 Representación circular del genoma parcial de 'Ca. P. palmae' fitoplasma del ALC..	104
Figura 4.2 Mapa de clasificación del análisis de anotación funcional de ontología génica (GO) del genoma parcial de 'Ca. P. palmae'-ALC.....	105
Figura 4.3 Mapa de clasificación del análisis de anotación de la Enciclopedia de Kioto de Genes y Genomas (KEGG) de 'Ca. P. palmae'-ALC.....	106
Figura 4.4 Motivos <i>de novo</i> (MEME) encontrados exclusivamente en los efectores predichos de 'Ca. P. palmae'-ALC.	114
Figura 4.5 Comparación circular en BLASTN del genoma de 'Ca. P. palmae'-ALC con otros doce genomas de fitoplasmas..	117
Figura 5.1 Palmas de <i>Cocos nucifera</i> L. colectadas en Telchac Puerto.....	134
Figura 5.2 Visualización del ARN en gel de agarosa al 1 % en muestras de tallo, hojas y raíz en palmas con ALC.....	135
Figura 5.3 Observación del ARN (sin trazas de ADN).....	136
Figura 5.4 Secuencias de proteínas de los efectores candidatos del fitoplasma de ALC..	139
Figura 5.5 Amplificación de efectores de 'Ca. P. palmae' en palmas con ALC en gel de agarosa al 1.5 % con ADNc.	141
Figura 5.6 Alineamiento con ClustalW del efector candidato de 'Ca. P. palmae' con la	

secuencia de la palma 1 con la muestra de tallo (P1-T).....	143
---	-----

Figura 5.7 Alineamiento con ClustalW del efector candidato de 'Ca. P. palmae' con la secuencia de la palma 2 con la muestra de hojas (P2-H).	144
---	-----

LISTADO DE TABLAS

Tabla 1.1 Síndrome en la palma de coco con ALC de acuerdo con la categoría correspondiente	5
--	---

Tabla 1.2 Genomas completos de fitoplasmas.....	12
---	----

Tabla 1.3 Algunas proteínas efectoras y su función reportadas en estudios de interacción entre patógenos y plantas hospedantes.....	18
---	----

Tabla 1.4 Lista de los efectores de fitoplasmas más conocidos actualmente	21
---	----

Table 2.1 List of phytoplasma effectors used in this work.	42
---	----

Table 2.2 Characterization of phytoplasma effectors.....	45
--	----

Table 2.3 Phytoplasma effector tribes and pattern distribution of effector members among phytoplasmas.....	53
--	----

Table 2.4 The top 10 ELM motifs found in phytoplasma effectors.	59
--	----

Table 3.1 List of true phytoplasma effectors used in the present work.....	78
--	----

Table 3.2 Analyses of phytoplasma effectors features.	81
--	----

Table 3.3 Comparison of SignalP v4.1 and SignalP v5.0, in Gram-positive mode, in different pipelines for the prediction of Phytoplasma effectors.	82
--	----

Table 3.4 Comparison of SignalP v4.1 and SignalP v5.0, in Gram-negative mode, in different pipelines for the prediction of Phytoplasma effectors.	83
--	----

Table 3.5 Prediction of Phytoplasma effectors using different pipelines on a testing dataset.	
---	--

.....	87
Table 3.6 Effectors prediction by PhyEffector on phytoplasma genomes.	89
Tabla 4.1 Efectores putativos predichos en 'Ca. P. palmae'-ALC	110
Tabla 4.2 Motivos ELM encontrados en los efectores putativos de 'Ca. P. palmae'-ALC.	115
Tabla 4.3 Comparación del genoma de 'Ca. P. palmae'-ALC (16SrIV-A) y genomas de otros fitoplasmas	116
Tabla 5.1 Confirmación (qPCR) y cuantificación de fitoplasmas en palmas de cocotero.....	134
Tabla 5.2 Concentración de ARN obtenida en tallo, raíces y hojas en palmas con ALC..	135
Tabla 5.3 Efectores candidatos identificados en el genoma de 'Ca. P. palmae'	137
Tabla 5.4 Cebadores (iniciadores) específicos diseñados para cada secuencia de efectores candidatos identificados en el genoma de 'Ca. P. palmae'.	138
Tabla 5.5 Análisis <i>in silico</i> de dos efectores candidatos de 'Ca. P. palmae', fitoplasma del ALC.	140
Tabla 5.6 Secuencia de los fragmentos amplificados con la muestra de la palma 1 con tallo (P1-T) y de la palma 2 con hojas (P2-H).	142

ABREVIATURAS

aa.: aminoácidos

ADN: ácido desoxirribonucleico

AJ: ácido jasmónico

ALC: amarillamiento letal del
cocotero

AP: proliferación del manzano

ARN: ácido ribonucleico

ARNt: ARN de transferencia

ARNr: ARN ribosómico

AVR: avirulenta

AY-WB: amarillamiento del aster
escoba de bruja

Ca.: Candidatus

Ca. P.: Candidatus fitoplasmas

DAPI: 4',6-diamidino-2-fenilindol

DTM: dominio transmembrana

ETI: resistencia desencadenada
por efectores

HR: respuesta hipersensible

kDa: KiloDaltons

LYD: enfermedad de tipo amarillamiento

NLS: señal de localización nuclear

MET: microscopia electrónica de
transmisión

MEB: microscopia electrónica de
barrido

OY-M: amarillamiento de la cebolla

PAMPs: patrones moleculares
asociados a patógenos

Primers: cebadores o iniciadores

PS: péptido señal

PMU: unidad potencialmente móvil

R: resistencia

SLY: amarillamiento letal de la
fresa

SST3: sistema de secreción tipo III

SST4: sistema de secreción tipo IV

TLE: tamaño límite de exclusión

TPPD: Declive de la palma Phoenix de
Texas (cepa ACPD)

RESUMEN

La palma de cocotero es una especie vegetal de gran interés económico, sin embargo, este cultivo está amenazado por una enfermedad denominada amarillamiento letal del cocotero (ALC) causado por fitoplasmas. Los fitoplasmas son patógenos obligados que habitan en el floema de las plantas, estos procariotas son capaces de secretar proteínas efectoras que se diseminan libremente hacia tejidos distantes, causando modificaciones en la planta y de esta manera garantizan su dispersión dentro del hospedante. Aunque se conoce que estos efectores son causantes de las alteraciones en las plantas, solamente se han validado 21 de ellos; muchas de estas proteínas efectoras se han identificado mediante el uso de distintos programas computacionales, por lo que los estudios *in silico* representan una herramienta útil para su búsqueda.

En este trabajo se realizó un análisis de caracterización *in silico* para evidenciar rasgos únicos (firmas) en los efectores de fitoplasmas, como motivos conservados. Al mismo tiempo se desarrolló una herramienta bioinformática que ayuda a identificar efectores en fitoplasmas. Por otro lado, se realizó el enriquecimiento de ADN del fitoplasma del ALC y secuenciación del genoma del patógeno. Adicionalmente, se diseñaron cebadores específicos con dos secuencias de efectores obtenidas del genoma del fitoplasma del ALC y se utilizaron para amplificar sus efectores.

Los resultados de la caracterización *in silico* mostraron que, con base en la organización de los motivos *de novo* encontrados, los fitoplasmas en general se podrían clasificar en 15 tribus o familias. Las tribus 1, 6 y 7 están ampliamente distribuidas, sugiriendo que provienen de un ancestro común y que fueron heredados por transferencia vertical; mientras que otras tribus solo se encuentran en fitoplasmas particulares, tal es el caso de las tribus 2, 3, 4 y 5 con una distribución discreta. Más aún, se observó que los efectores de fitoplasmas sobresalen por contener >25 residuos de lisina (~50 %), leucina (~20 %) y asparagina (~15 %), así como como nulo o escaso contenido (<3 residuos) de cisteína (~80 %) y triptófano (~60 %) en la mayoría de sus secuencias; además que poseen motivos SLiM o ELM (motivos lineales cortos o eucariotas), imitando los motivos de su hospedante y probablemente contribuyendo a su virulencia. La creación del algoritmo PhyEffector es otra contribución de este trabajo que permitirá una búsqueda más amplia de

estas proteínas efectoras, al poder identificar efectores clásicos (canónicos) y no-clásicos (no-canónicos) en los genomas de fitoplasmas.

La secuenciación y análisis del genoma parcial de '*Ca. P. palmae*' asociado a la enfermedad del ALC, mostró que éste es pequeño, de 434,453 pb y contiene 54 efectores, localizados en su mayor parte fuera de regiones PMUs, algunos de los cuales pertenecen a las tribus 1, 7 y 10. Los resultados del genoma parcial de '*Ca. P. palmae*'-ALC mostraron que este fitoplasma posee vías metabólicas involucradas en glucólisis, el metabolismo de nucleótidos, acetogénesis y el metabolismo de los glicerofosfolípidos. Además de genes involucrados en el metabolismo de las purinas y pirimidinas; así como la identificación de los genes *secA* y *secY*, importantes en el sistema de secreción. Sin embargo, aún falta determinar el genoma completo y saber con exactitud las vías completas de este genoma. Cuando se analizó la expresión de sus efectores, a través del diseño de cebadores específicos con dos secuencias de efectores candidatos identificados en su genoma, se pudo detectar la expresión de un efector en palmas infectadas.

Los resultados de este trabajo sientan las bases para comprender como los efectores de fitoplasmas intervienen en la virulencia de estos patógenos y entender mejor la enfermedad del ALC. La identificación de los efectores candidatos es el primer paso para conocer las moléculas de virulencia que posee el genoma del fitoplasma de ALC. Este trabajo permitió la amplificación de dos muestras infectadas, mediante la técnica de PCR, con cebadores diseñados a partir de los efectores candidatos, confirmando que la secuencia amplificada corresponde a uno de los efectores predichos en el genoma, lo que sugiere la probabilidad de ser un efector de ALC.

ABSTRACT

The coconut palm is a plant species of great economic interest. However, this crop is threatened by a disease called Coconut lethal yellowing (CLY) caused by phytoplasmas. Phytoplasmas are obligate pathogens that inhabit the phloem of plants, these prokaryotes are capable of secreting effector proteins that spread freely to other distant tissues, causing modifications in the plant and ensuring their dispersion within the host. Although it is known that these effectors are responsible for alterations in plants, only 21 effectors have been validated. Many of these effector proteins have been identified with the support of various computer programs, making *in silico* studies a useful tool for their discovery.

In this work, an *in silico* characterization analysis was performed to search to highlight unique features (signatures) in phytoplasma effectors, such as conserved motifs. At the same time, a bioinformatics tool was developed to help identify effectors in phytoplasmas. DNA enrichment of the CLY phytoplasma and sequencing of the pathogen's genome were carried out. Additionally, specific primers were designed with two effector sequences obtained from the CLY phytoplasma genome and were used to amplify its effectors.

The results of the *in silico* characterization showed that based on the organization of the *de novo* motifs found, the phytoplasmas in general could be classified into 15 tribes or families. Tribes 1, 6 and 7 were widely distributed, suggesting they come from a common ancestor and were inherited by vertical transfer, while other tribes had a discrete distribution. Additionally, phytoplasma effectors were observed to contain >25 residues of lysine (~50 %), leucine (~20 %), and asparagine (~15 %), as well as null or scarce content (<3 residues) of cysteine (~80 %) and tryptophan (~60 %) in most of their sequences; they also possess SLiM or ELM motifs (short linear motifs or eukaryotic), imitating host motifs and probably contributing to their virulence. The creation of the PhyEffector algorithm is another contribution of this work that will allow a broader search for these effector proteins, by identifying classical (canonical) and non-classical (non-canonical) effectors in phytoplasma genomes.

The sequencing and analysis of the partial genome of 'Ca. *P. palmae*' associated with CLY disease, showed that it is small of 434,453 bp and contain 54 effectors, mostly

located outside of PMU regions, some of which belong to tribes 1, 7, and 10. The results of the partial genome of '*Ca. P. palmae*'-CLY showed that this phytoplasma has metabolic pathways involved in glycolysis, nucleotide metabolism, acetogenesis, and glycerophospholipid metabolism. In addition to genes involved in purine and pyrimidine metabolism, as well as the identification of the *secA* and *secY* genes, important in the secretion system. However, the complete genome and the exact complete pathways of this genome have yet to be determined. When the expression of its effectors was analyzed, through the design of specific primers with two candidate effector sequences identified in its genome, the expression of one effector was detected in infected palms.

The results of this work lay the foundation for understanding how phytoplasma effectors intervene in the virulence of these pathogens and better understand CLY disease. The identification of candidate effectors is the first step for know the virulence molecules that the CLY phytoplasma genome has. This work enabled the amplification of two infected samples using the PCR technique with primers designed from the candidate effectors, confirming that the amplified sequence corresponds to one of the predicted effectors in the genome, which suggests the probability of being a CLY effector.

INTRODUCCIÓN

A nivel mundial, el cocotero (*Cocos nucifera* L.), representa una de las especies vegetales de mayor interés comercial debido a los múltiples productos que de este cultivo se obtienen (Granados y López, 2002; Limones y Fernandez, 2016). Sin embargo, *C. nucifera* es atacado por una enfermedad conocida como Amarillamiento letal del cocotero (ALC) que produce daños fundamentales a las palmas y causa importantes pérdidas de producción (Oropeza *et al.*, 2020). En la península de Yucatán, el ALC ha causado la muerte de miles de palmas de cocotero, disminuyendo la superficie de este cultivo. El ALC es causado por el fitoplasma que pertenece al grupo 16SrIV-A, '*Candidatus* Phytoplasma palmae' (Ramos-Hernández *et al.*, 2018).

Los fitoplasmas son capaces de secretar un tipo de proteínas o moléculas llamadas efectores, los cuales mejoran la capacidad de infectar a las plantas (Padilla-Ramos *et al.*, 2019). Sin embargo, aún son pocos los efectores conocidos que se relacionan con los fitoplasmas (Singh *et al.*, 2019; Sugio *et al.*, 2011a), por lo que la bioinformática representa una herramienta útil para realizar minería de datos en las secuencias de fitoplasmas disponibles en las bases de datos.

Por otro lado, no se conocía ningún efector del grupo 16SrIV de fitoplasmas, por lo cual la estrategia de este trabajo comprende una fase de análisis computacionales y luego una fase de aplicación de técnicas moleculares. Mediante el análisis *in silico* para la identificación de características conservadas como motivos o dominios entre los efectores de fitoplasmas, se encontraron motivos *de novo* que permitieron clasificarlos en 15 tribus, se observó un escaso o nulo contenido de los aminoácidos cisteína y triptófano y la presencia de SLiMs; además se logró la creación de una herramienta bioinformática para la identificación de estos efectores.

Se realizó la técnica de ADN-enriquecido lo que permitió la secuenciación del genoma parcial del fitoplasma causante de ALC con la identificación de sus efectores. Adicionalmente, se realizó un análisis de expresión que permitió la amplificación de uno de los efectores del fitoplasma de ALC. Este trabajo establece las bases para profundizar en la comprensión de este importante patógeno.

CAPÍTULO I

ANTECEDENTES

1. COCOTERO

El cocotero o palma de coco (*Cocos nucifera* L.) es originario de Indo-Malasia en el Pacífico Occidental (Limonés y Fernández, 2016), pertenece a la familia Arecaceae, subfamilia Arecoideae (Palmae) del género de las palmeras; comprende alrededor de 200 géneros y cerca de 2,500 especies (Valois-Cuesta *et al.*, 2013; Delucchi y Hurrell, 2008; Granados y López, 2002). Es una palma de hasta 30 m de alto con hojas de 1.8 a 6 m de largo, un tronco robusto de ~30 centímetros de diámetro, flores amarillas y fruto de 20 a 30 cm de diámetro de color verde a pardo rojizo, compuesto de tres capas: epicarpio (brillante), mesocarpio (fibroso) y endocarpio (lignificado) y que contienen en el interior de la semilla una porción carnosa (endospermo sólido) y un líquido denominado agua de coco (endospermo líquido) (CONABIO, 2021; Limonés y Fernández, 2016; Granados y López, 2002).

El cocotero crece en climas con una temperatura promedio superior a 20°C, de suelos tipo arenoso, con presencia de materia orgánica y pH entre 5 y 8. La profundidad mínima del suelo para su óptimo desarrollo radicular debe ser de 80 a 100 cm. (Granados y López, 2002). Se le conoce como el “árbol de la vida o árbol de los mil usos” porque cada parte de esta especie vegetal se utiliza para la obtención de distintos productos; por ejemplo, el tallo es usado para la fabricación de diferentes materiales de construcción para muebles, postes, etc.; las hojas se utilizan como techos para viviendas, en la fabricación de sombreros y otro tipo de artesanías; el fruto de la palma de coco es utilizado como alimento y bebida; el aceite de coco es utilizado con fines comestibles, así como en la fabricación de productos cosméticos como shampoo, jabones, entre otros usos (Figura 1.1). El principal objetivo industrial de la palma de coco es la producción de copra, ya que contiene aproximadamente 63 % de aceite (CONABIO, 2021; Limonés y Fernández, 2016; Granados, y López, 2002).



Figura 1.1 El cocotero “árbol de la vida” y sus productos finales.
(Granados Sánchez y López Ríos, 2002).

La palma de coco es una especie de distribución pantropical, se encuentra cultivado en Filipinas, Malasia, Indonesia, India, Oceanía y las Antillas (Milo, 2001). Indonesia, Filipinas e India contribuyen al 82 % de la producción mundial, mientras que, en Latinoamérica, México y Brasil son los principales países productores de este cultivo. México aporta el 2 % de la producción mundial, donde el estado de Guerrero ocupa el primer lugar en la producción de coco (SADER, 2022; FAOSTAT, 2020; Planeación Agrícola Nacional 2017-2030), cultivado en distintos estados del país desde las costas del Pacífico hasta el Atlántico (SAGARPA, 2017; Granados y López, 2002; Zizumbo *et al.*, 1993). El estado de Yucatán aporta un 9 % a nivel nacional a la producción, entre los que se encuentran los municipios de Celestún, Progreso, Río Lagartos, Telchac Puerto, Ticul, etc. (Benítez-Arellano y Sosa-Alcaraz, 2017).

Sin embargo, la palma de coco es afectada por distintas plagas y enfermedades que ponen en riesgo la producción de este cultivo, como la enfermedad del amarillamiento letal del

cocotero (CLY, por sus siglas en inglés). En América Latina esta enfermedad ha causado un impacto significativo en el cocotero (Ramos-Hernández *et al.*, 2018).

1.1 AMARILLAMIENTO LETAL DEL COCOTERO (ALC)

El amarillamiento letal es la enfermedad más importante del cocotero a nivel mundial. En México se detectó a finales de los años 70's del siglo pasado, con pérdidas de producción de al menos el 80 % de las plantaciones de cocotero, en la Península de Yucatán se reportó en 1982, además de confirmarse en otros estados como: Tabasco, Guerrero y Oaxaca (Ramos-Hernández, *et al.*, 2018; Góngora-Canul *et al.*, 2004).

El ALC es causado por '*Candidatus Phytoplasma palmae*' perteneciente al subgrupo 16SrIV-A y se transmite por el insecto *Haplaxius (Myndus) crudus*, conocido comúnmente como Chicharrita (Ramos-Hernández *et al.*, 2018). El grupo 16SrIV se divide en 6 subgrupos de cepas nombrados como A, B, C, D, E y F (Oropeza *et al.*, 2010) de los cuales el 16SrIV-A y 16SrIV-D han sido reportados en Yucatán (Vázquez-Euán *et al.*, 2011).

Cuando el cocotero es infectado por los fitoplasmas permanece en periodo asintomático llamado período de incubación (4-15 meses). Los síntomas visuales característicos del ALC en orden de aparición son: caída prematura de los frutos, seguido por necrosis de las inflorescencias (de parcial a extendida); posteriormente aparece el amarillamiento de las hojas, progresando desde las hojas inferiores hacia arriba, a medida que avanza la enfermedad, estas hojas cambian a un color café y se secan. Por último, el tronco queda desnudo, como consecuencia del colapso de las hojas y el cogollo (Tabla 1.1 y Figura 1.2); la planta muere en un máximo de seis meses después de que aparecen los síntomas visuales (Ramos-Hernández *et al.*, 2018; Oropeza *et al.*, 2010).

Tabla 1.1 Síndrome en la palma de coco con ALC de acuerdo con la categoría correspondiente

Categoría	Síntomas
Asintomático	Período de incubación (sin síntomas visuales)
Primario	Caída de frutos Necrosis parcial de las inflorescencias Necrosis extendida de dos o más inflorescencias

Intermedio (Amarillamiento)	Amarillamiento en las hojas maduras (desde abajo) Amarillamiento en las hojas jóvenes (hacia arriba)
Avanzado	Hojas amarillas cambian a un color café y se secan Hoja bandera muerta Palma en forma de poste telefónico

Fuente: adaptada de Oropeza *et al.* (2010)



Figura 1.2 Síndrome del ALC en orden de aparición. (A) Caída prematura de frutos. (B) Ennegrecimiento de inflorescencias nuevas. (C) Ennegrecimiento de inflorescencias después de emerger la espata. (D) Amarillamiento de las hojas más viejas inferiores. (E) Amarillamiento en hojas medias. (F) Amarillamiento en hojas jóvenes superiores y necrosis. (G) Finalmente el tronco desnudo sin cogollo (Oropeza *et al.*, 2010).

Las especies vegetales activan distintos mecanismos de defensa, que incluyen reacciones químicas y bioquímicas, que les permite reconocer agentes extraños para evitar la invasión y proliferación de diferentes microorganismos (Sautua y Carmona, 2021; Burbano-Figueroa, 2020; Azcón-Bieto y Talón, 2013). La respuesta de resistencia de las plantas se clasifica en dos líneas principales de defensa durante la interacción planta-patógeno: el reconocimiento de patrones moleculares asociados a patógenos (PAMPs) y la resistencia

desencadenada por efectores (ETI) (Burbano-Figueroa, 2020; Torres Pacheco *et al.*, 2012; Dodds y Rathjen, 2010; Jones y Dangl, 2006). Es importante considerar que para que se logre desarrollar una enfermedad en una planta deben complementarse distintos factores, denominado como el “Triángulo de la enfermedad”, modelo que explica que deben existir tres condiciones favorables que propicien la enfermedad: el agente patógeno, un hospedante susceptible (planta) y las condiciones ambientales adecuadas (Scholthof, 2007).

Entre los patógenos capaces de suprimir la respuesta de resistencia en diferentes especies vegetales invadiendo y multiplicándose en distintos tipos de hospedante se encuentran los fitoplasmas.

1.2 FITOPLASMAS

Los fitoplasmas fueron descubiertos por primera vez en 1967, mediante microscopía electrónica en plantas con amarillamiento (Doi *et al.*, 1967). Se determinó que son miembros de la clase mollicutes, los cuales basados en la secuencia del gen 16S ARNr se dedujo que provienen de un ancestro bacteriano gram-positivo, similar a *Clostridium*, del linaje lactobacillus (Lee *et al.*, 2000); los estudios filogenéticos sugieren que el ancestro común de los fitoplasmas es *Acholeplasma laidlawii* (Bertaccini *et al.*, 2019; Lee *et al.*, 2000) al mostrar una relación cercana con esta bacteria gram-positiva (Hoischen *et al.*, 1993; Tully *et al.*, 1993).

Estos microorganismos patógenos se encuentran dentro del género ‘*Candidatus Phytoplasma*’ (‘Ca. P.’). El término de ‘*Candidatus*’ se le asignó por pertenecer al grupo de organismos que no puede ser cultivados *in vitro* (Reveles-Torres *et al.*, 2017). Pueden observarse por microscopía mediante cortes longitudinales a través del floema de una planta infectada (Figura 1.3), mediante microscopía electrónica de fluorescencia con DAPI (4',6-diamidino-2-fenilindol) o mediante microscopía electrónica de transmisión (MET) y de barrido (MEB) (Álvarez-Sánchez *et al.*, 2019).

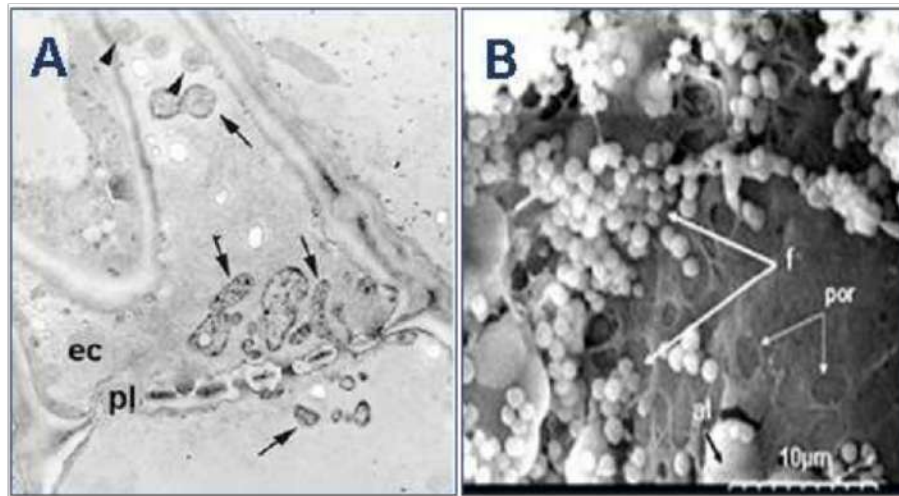


Figura 1.3 Fitoplasmas en el floema de una planta infectada. (A) Forma pleomórfica de fitoplasmas (flechas) en los elementos cribosos (ec) cerca de plasmodesmos (pl) mediante MET. (B) Fitoplasmas (f) con respecto a los poros de tubos cribosos (por) mediante MEB. Basado en Álvarez-Sánchez *et al.* (2019) y Douriet (2011).

Los fitoplasmas pertenecientes a la clase *Mollicutes* (Fletcher y Wayadanda, 2002), son organismos pleomórficos con un diámetro entre 0.1 a 1 µm, que se mueven libremente a través de los elementos cribosos de las plantas (Douriet, 2011; Leyva-López *et al.*, 2006), rodeados solamente por la membrana plasmática; una cualidad importante es que no tienen pared celular, particularidad que le confiere características citológicas y moleculares inusuales, concediéndoles sensibilidad a la lisis, resistencia a la penicilina u otras sustancias antibacterianas, mientras que son sensibles a las tetraciclinas (Ferriol Marchena *et al.*, 2016; Fletcher y Wayadanda, 2002). Poseen genomas entre 530-1350 kb, con características particulares como contenido de 23-31 % de guanina y citosina, lo que representa un contenido bajo en comparación con lo que contienen otros procariontes (25-80 %) (Oropeza *et al.*, 2010; Leyva-López *et al.*, 2006). Son parásitos obligados de plantas e insectos vectores, y por lo que pueden sobrevivir y multiplicarse en el floema de las plantas o en la hemolinfa de los insectos, reproduciéndose por gemación o fisión binaria (Ferriol *et al.*, 2016; Fletcher y Wayadanda, 2002).

Los fitoplasmas se clasifican en 37 grupos basados en su porcentaje de identidad, del 98.65 % con el gen de ARN ribosomal 16S (16S rRNA, de acuerdo con sus siglas en inglés), encontrándose desde el grupo 16SrI al grupo 16SrXXXIX (Wei y Zhao, 2022). En la figura 1.4 se muestra el ejemplo de un árbol filogenético basados en el gen 16S rRNA de diferentes fitoplasmas y el grupo al que pertenecen (Cho *et al.*, 2019), separados por colores, se observa dentro del cuadro amarillo a '*Ca. P. palmae*' y las especies de fitoplasmas cercanas a éste.

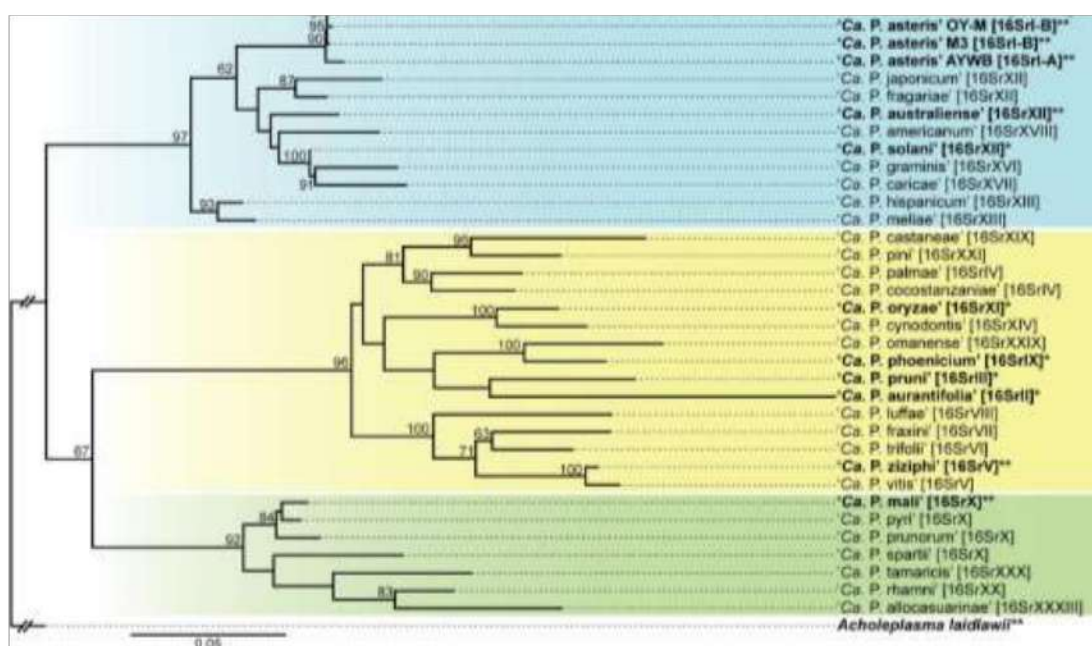


Figura 1.4 Árbol filogenético de algunas especies de fitoplasmas basados en el gen 16SrRNA. Se observa la relación cercana que existe entre las diferentes especies de '*Candidatus Phytoplasma*'. (Cho *et al.*, 2019).

A pesar de ser microorganismos que no se han podido cultivar *in vitro* se ha logrado secuenciar los genomas completos de varios fitoplasmas. Entre las primeras especies reportadas se encuentran el genoma de '*Ca. P. asteris*', responsable del amarillamiento de la cebolla línea M (OY-M, de acuerdo con sus siglas en inglés) y '*Ca. P. asteris*' del amarillamiento del aster-escoba de bruja (AY-WB, por sus siglas en inglés), entre otros genomas que a continuación se describen.

1.2.1 GENOMAS DE FITOPLASMAS

Los avances en las técnicas moleculares han permitido la secuenciación de un mayor número de genomas de fitoplasmas. Actualmente, se han depositado al menos 29 genomas completos en GenBank (<https://www.ncbi.nlm.nih.gov/home/genomes/>), algunos de los cuales no se han publicado. Los primeros genomas ensamblados y más conocidos pertenecen a los grupos I, II, V, VIII, X y XII, de acuerdo con la clasificación del gen 16Sr (Tabla 1.2); se describen a continuación los primeros genomas ensamblados.

El primer genoma reportado fue el del fitoplasma del amarillamiento de la cebolla línea M (OY-M, Onion yellows phytoplasma), el cual pertenece al grupo 16SrI-B (Oshima *et al.*, 2004). Consiste en un cromosoma circular de aproximadamente 860 Kb con un contenido de guanina y citosina (GC) de 28 %, además de la presencia de dos plásmidos de 5 y casi 4 Kb respectivamente, con 25 y 24 % de GC, respectivamente, además de genes (18 % del genoma) asociados a regiones denominadas Unidades Potencialmente Móviles o PMUs.

El segundo genoma ensamblado fue reportado por Bai *et al.* (2006), también pertenece a 'Ca. *P. asteris*'; este es agente causal del amarillamiento del aster-escoba de bruja (AYWB, Aster yellows witches'-broom phytoplasma), una enfermedad transmitida por *Macrosteles quadrilineatus*. AYWB pertenece al grupo 16SrI-A, posee un cromosoma circular de ~700 Kb con 27 % de GC, el cual es más pequeño que el genoma de OY-M; además de cuatro regiones PMUs (PMU1 - PMU4). AYWB; contiene el doble de plásmidos que OY, tres de ellos con tamaños de 4 o cerca de 4 Kb y un cuarto plásmido de 5 Kb con contenidos de GC que oscilan entre 21.8 y 25.6 %.

El tercer genoma secuenciado se encuentra dentro de los genomas más pequeños y con menor contenido de GC, fue reportado por Kube *et al.* (2008) y corresponde a 'Ca. *P. mali*' cepa AT, primero del grupo 16SrX; éste es agente causal de la proliferación del manzano (AP, apple proliferation). A diferencia de los genomas reportados hasta el momento con cromosomas circulares, éste posee un cromosoma lineal, ya que se observó la presencia de repeticiones invertidas terminales o TIRs, característicos de los cromosomas lineales. Tiene un tamaño cercano a 602 Kb y un contenido de GC de 21.4 %, así como la presencia de dos regiones putativas de PMUs.

Tran-Nguyen *et al.* (2008), lograron ensamblar el cuarto genoma de '*Ca. P. australiense*' aislado de PAa (subgrupo *tuf*-Australia I, *rp*-A) que se encuentra estrechamente relacionado con el fitoplasma asteris; este fitoplasma se asocia con distintas enfermedades en Australia y Nueva Zelanda, como el amarillamiento de los viñedos australianos (AUSGY, Australian grapevine yellows), la muerte regresiva de la papaya, el amarillo letal de la fresa (SLY, Strawberry lethal yellows), etc. '*Ca. P. australiense*' pertenece al grupo 16SrXII, posee un cromosoma circular de más de 879 Kb, siendo 18, 693 pb más grande que '*Ca. P. asteris*' cepa OY-M; su cromosoma contiene un 27 % de GC, además de un plásmido de 3.7 Kb y con 5 PMUs que representan el 12.1 % de su genoma.

Más tarde Andersen *et al.* (2013) publicaron el quinto genoma de '*Ca. P. australiense*', cepa del amarillamiento letal de la fresa (SLY, Strawberry lethal yellows). Posee un cromosoma circular, el más grande reportado hasta el momento, con un tamaño aproximado a 960 Kb con 27 % de GC; se pudieron identificar 14 clusters de genes repetidos que corresponden a PMUs, representando 39 % del genoma de SLY, además de un plásmido de 3.6 Kb.

El sexto genoma fue publicado por Orlovskis *et al.* (2017) y corresponde a '*Ca. P. asteris*' del enanismo arbustivo del maíz (MBSP M3, Maize bushy stunt phytoplasma) del grupo 16SrI-B; este fitoplasma afecta la cosecha del maíz en países latinoamericanos. Siendo el tercer genoma del fitoplasma asteris, representa el genoma más pequeño de todos los reportados hasta el momento, con un cromosoma circular de ~576 Kb y un contenido de 28.5 % de GC, identificando genes semejantes de regiones PMUs.

Wang *et al.* (2018a) describieron el séptimo genoma de '*Ca. P. ziziphi*', fitoplasma de la escoba de bruja del jujube o dátíl rojo (JWB, jujube witches'-broom), el primero del grupo V (grupo del amarillamiento del olmo); este fitoplasma posee un cromosoma circular de ~750 Kb y un contenido de 23.3 % de GC, además de cuatro regiones PMUs que fueron identificadas en su genoma.

Desde el 2004 solamente se pudieron secuenciar con dificultad estos siete genomas; actualmente con las nuevas tecnologías se han podido ensamblar otros genomas. Entre los más recientes se encuentra el genoma de '*Ca. P. asteris*' cepa De Villa, del grupo 16SrI, responsable de la escoba de bruja en '*Catharanthus roseus*' que posee un cromosoma circular de ~600 Kb (Coetzee *et al.*, 2019). Tan *et al.* (2021a) reportaron a '*Ca. P.*

aurantifolia', primer fitoplasma del grupo 16SrII (EpWB, '*Echinacea purpurea*' witches-broom), posee un cromosoma circular de ~635 Kb, un plásmido (4 Kb) y una región PMU. Posteriormente, Huang *et al.* (2022) reportaron el genoma de 'Ca. *P. luffae*' cepa NCHU2019, del grupo VIII, responsable de la escoba de bruja de la Luffa egipcia (*Luffa aegyptiaca*), tiene un cromosoma circular de ~769 Kb y la presencia de 13 regiones PMUs. El onceavo genoma completo pertenece a 'Ca. *P. vitis*' cepa CH (Debonneville *et al.*, 2022) y representa el segundo fitoplasma del grupo V, responsable de la enfermedad denominada Flavescencia dorada (FD, Flavescence Dorée) de la vid, posee un cromosoma circular de ~654 Kb y es el único genoma hasta ahora reportado que carece de regiones PMUs. En la tabla 1.2 se muestran los genomas completos ensamblados descritos en este trabajo.

Tabla 1.2 Genomas completos de fitoplasmas

Fitoplasma	Enfermedad	Grupo 16S rRNA	Tamaño (pb)	Proteínas (NCBI)	Efectores predichos	Referencia	Acceso (NCBI)
'Ca. Phytoplasma asteris' OY-M	Amarillamiento de la cebolla- cepa M (OY-M)	I-B	853,092	752	30 (Hoshi <i>et al.</i> , 2009)	Oshima <i>et al.</i> , 2004	AP006628
'Ca. Phytoplasma asteris' AY-WB	Amarillamiento del Aster escoba de bruja (AY-WB)	I-A	706,569	568	56 (Bai <i>et al.</i> , 2009)	Bai <i>et al.</i> , 2006	CP000061
'Ca. Phytoplasma mali' AT	Proliferación del manzano (AP)	X-A	601,943	498	---	Kube <i>et al.</i> , 2008	CU469464
'Ca. Phytoplasma australiense' PAa	Asociado con Amarillamiento de los viñedos australianos (AUSGY), etc.	XII-B	879,324	570	---	Tran-Nguyen <i>et al.</i> , 2008	AM422018
'Ca. Phytoplasma australiense'	Amarillamiento letal de la fresa (SLY)	XII	959,779	876	---	Andersen <i>et al.</i> , 2013	CP002548
'Ca. Phytoplasma asteris'	Enanismo arbustivo del maíz (MBSP M3)	I	576,118	503	36 (Orlovskis <i>et al.</i> , 2017)	Orlovskis <i>et al.</i> , 2017	CP015149
'Ca. Phytoplasma ziziphi'	Escoba de bruja del jujube o dátil rojo (JWB)	V	750,803	662	28 (Wang <i>et al.</i> , 2018a)	Wang <i>et al.</i> , 2018 ^a	CP025121
'Ca. Phytoplasma asteris' De Villa	Escoba de bruja en <i>Catharanthus roseus</i> cepa De Villa	I	600,116	534	---	Coetzee <i>et al.</i> , 2019	CP035949
'Ca. Phytoplasma aurantifolia'	Escoba de bruja de <i>Echinacea purpurea</i> (EpWB)	II	635,584	471	27 (Tan <i>et al.</i> , 2021)	Tan <i>et al.</i> , 2021	CP040925
'Ca. Phytoplasma luffae'	Escoba de bruja de la luffa egipcia	VIII	769,143	716	---	Huang <i>et al.</i> , 2022	CP054393
'Ca. Phytoplasma vitis' cepa CH	Flavescence dorée phytoplasma (FDp)	V	654,223	498	17 (Debonneville <i>et al.</i> , 2022)	Debonneville <i>et al.</i> , 2022	CP097583

Los genomas de fitoplasmas publicados hasta el momento son menores de 1000 kb, siendo el más pequeño el genoma de 'Ca. *P. asteris*' MBSP (~576 Kb) y el más grande el de 'Ca.

P. australiense SLY (~959 Kb). La mayoría posee un cromosoma circular, con excepción de 'Ca. *P. mali*' cuyo cromosoma es lineal (Kube *et al.*, 2008). También es común, la presencia de plásmidos (Tan *et al.*, 2021; Coetzee *et al.*, 2019; Andersen *et al.*, 2013; Tran-Nguyen *et al.*, 2008; Bai *et al.*, 2006; Oshima *et al.*, 2004). Además, estos patógenos se caracterizan por poseer genomas reducidos, pero ricos en regiones repetidas denominadas PMUs.

Los fitoplasma solamente contienen genes para funciones celulares básicas como la replicación, transcripción, traducción y translocación de proteínas del ADN (Oshima *et al.*, 2013). Entre los pocos genes conservados están los genes que codifican para la síntesis de folato, sacarosa fosforilasa, glucanasa, para proteínas similares a la hemolisina y genes glicolíticos (Tran-Nguyen *et al.*, 2008; Bai *et al.*, 2006; Oshima *et al.*, 2004). Además de la presencia de muchos genes que codifican sistemas transportadores, importantes para la captación de moléculas de la célula hospedante (Andersen *et al.*, 2013; Oshima *et al.*, 2013; Oshima *et al.*, 2004).

Conocer el genoma de estos patógenos puede ayudar a comprender mejor el modo de interacción con su hospedante, así como conocer los cambios que ejercen en la especie vegetal infectada, síntomas que causan, entre otros, tema que ha sido abordado por diversos reportes.

1.2.2 FITOPLASMAS-PLANTAS

Los fitoplasmas causan en las plantas que infectan un desbalance en las fitohormonas, las cuales tienen importantes funciones en el crecimiento, desarrollo y defensa vegetal (Asif *et al.*, 2022; Bari y Jones, 2009), e inducen distintas deformaciones y sobrecrecimientos (Dermastia, 2019). El ácido salicílico (SA), ácido jasmónico (JA) y etileno son las principales fitohormonas responsables de la resistencia vegetal. Se ha observado un aumento del SA y de sus conjugados en nervaduras y hojas en plantas infectadas; la mayoría de los genes biosintéticos para Jasmonatos también se encuentran alterados, por lo que durante una infección hay un incremento temprano en los niveles del JA, pero después disminuyen. Otra

fitohormona que interviene en la defensa vegetal es el etileno, ya que se sobreexpresan los genes que intervienen en su biosíntesis (Dermastia, 2019).

Los fitoplasmas causan diferentes síntomas, entre los cuales se incluyen el amarillamiento, enrojecimiento de hojas, necrosis del floema, filodia (transformación de órganos florales en tejidos semejantes a hojas), escoba de bruja (desarrollo de numerosos brotes), virescencia (coloración verde de los órganos florales) y retraso del crecimiento (enanismo); síntomas que varían con el fitoplasma y con el estado de infección en que se encuentre la planta (Figura 1.5), siendo algunas especies tolerantes o resistentes, estas plantas podrán tener infecciones asintomáticas o mostrar síntomas leves (Camarena-Gutiérrez y De la Torre-Almaraz, 2008; Davis y Sinclair, 1998; Lee *et al.*, 2000).

El amarillamiento, se produce porque estos patógenos obstruyen los tubos cribosos del floema, afectan su función e interfieren en el transporte de carbohidratos, así como una disminución en el contenido de clorofila, carotenoides y proteína (Oropeza *et al.*, 2010; Lepka *et al.*, 1999). Estos cambios morfológicos son consecuencia de las alteraciones que causan estos patógenos sobre las fitohormonas, a través de sus proteínas efectoras secretadas (Dermastia, 2019; Singh *et al.*, 2019), se ha demostrado que algunas proteínas efectoras de los fitoplasmas alteran los factores de transcripción con dominio MADS, interfiriendo en los genes homeóticos florales (Iwabuchi *et al.*, 2020; MacLean *et al.*, 2014).



Figura 1.5 Síntomas causados por fitoplasmas en distintas especies vegetales. (A) Filodia en plantas áster (margarita). (B) Amarillamiento de las hojas en palmas de coco. (C) Escoba de bruja en pino. (D) Enrojecimiento de las hojas en el maíz. (E) Virescencia en vinca rosa. Camarena-Gutiérrez y De la Torre-Almaraz (2008).

Además de parasitar las plantas, estos patógenos también son capaces de sobrevivir en insectos, los cuales son los vectores de los fitoplasmas y favorecen la diseminación de estos a otros individuos vegetales de la misma u otras especies.

1.2.3 FITOPLAMAS E INSECTO VECTOR

Los fitoplasmas son parásitos obligados por lo que tienen muchos requerimientos nutricionales que dependen del hospedante. Son transmitidos por insectos vectores chupadores de savia, que pertenecen al orden de los Hemípteros, principalmente provenientes de la familia de los cicadélidos, fulgoromorfos y psílidos (Trivellone y Dietrich, 2021; Weintraub y Beanland, 2006; Lee *et al.*, 2000). Estos patógenos pueden ser también

transmitidos por varias especies diferentes de insectos y a su vez una sola especie de vector puede transmitir dos o más fitoplasmas, habiendo la posibilidad de encontrar vectores infectados con múltiples cepas de fitoplasmas (Trivellone y Dietrich, 2021; Weintraub y Beanland, 2006). La especie *Haplaxius (Myndus) crudus* (Hemiptera: Cixiidae), conocido comúnmente como chicharrita, es el principal transmisor del ALC y se encuentra ampliamente distribuido en los países tropicales de América Latina y el Caribe (Álvarez *et al.*, 2019).

Los fitoplasmas pasan el invierno en vectores infectados y se diseminan en la primavera siguiente (Lee *et al.*, 2000), tanto a través de las ninfas como de los adultos (Hogenhout *et al.*, 2008; Weintraub y Beanland, 2006). Hay evidencia de una interacción beneficiosa insecto-fitoplasma, aumentando la longevidad y el número de crías del insecto cuando se alimentan de plantas infectadas (Bertaccini *et al.*, 2019). Los insectos transmisores de fitoplasmas pueden clasificarse según su gama de hospedantes en polífagos (se alimentan de especies de distintas familias u órdenes), oligófagos (se alimentan de especies de una misma familia o familias afines) y monófagos (se alimentan de especies de un mismo género) (Lee *et al.*, 2000).

1.2.4 MODO DE TRANSMISION DE LOS FITOPLASMAS

Los vectores encargados de la transmisión de los fitoplasmas se caracterizan por tener piezas bucales alargadas o estiletes, que les permite perforar y succionar la savia de las plantas (Aguin-Pombo y Bourgoin, 2012). Cuando estos insectos se alimentan del floema de una planta infectada por fitoplasmas, el insecto adquiere el patógeno, proceso denominado adquisición por alimentación, en el cual los fitoplasmas entran al intestino del insecto, de ahí a la hemolinfa y otros órganos del insecto como las glándulas salivales; posteriormente este insecto vector es capaz de transmitir el fitoplasma a otras especies vegetales, proceso denominado inoculación por alimentación; la planta sana es inoculada, después de que infecta y transcurre un periodo de incubación aparecen los síntomas característicos de esta enfermedad (figura 1.6) (Sugio *et al.*, 2011a; Weintraub y Beanland, 2006).

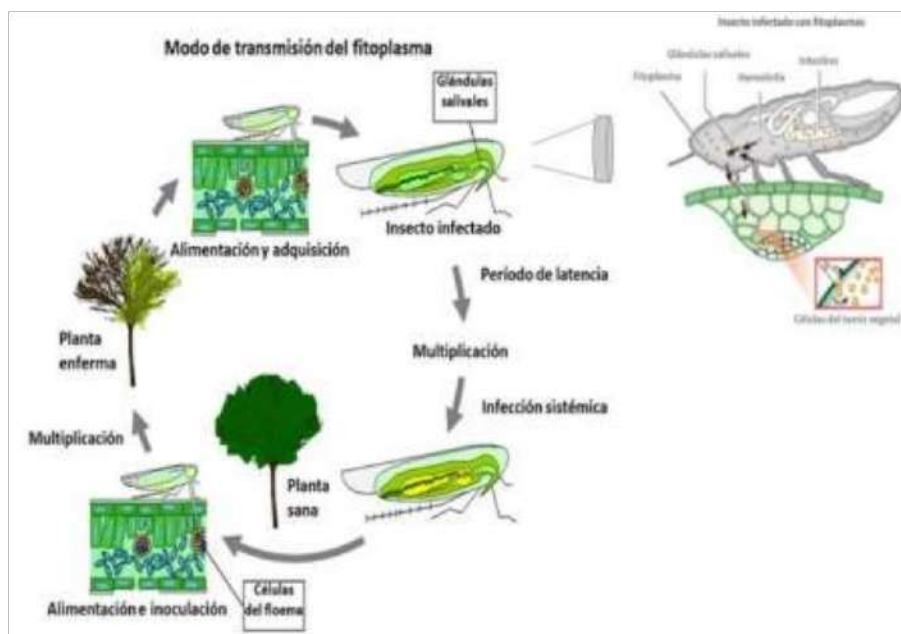


Figura 1.6 Modo de transmisión del fitoplasma e insecto vector. En la imagen de la izquierda se observa el ciclo de vida del fitoplasma, momento en el cual el insecto perfora la hoja de la planta para alimentarse y adquiere al patógeno (alimentación y adquisición), ocurre la multiplicación e infección sistémica del fitoplasma en el insecto, posteriormente el insecto vector al alimentarse de una planta sana inocula al patógeno (alimentación e inoculación), transcurre un periodo de incubación y aparecen los síntomas característicos por fitoplasmas. En la imagen superior derecha se observa la diseminación del patógeno a través de los diferentes órganos del insecto. Basado en Alonso (2011) y Sugio *et al.* (2011a).

Un insecto vector puede portar varias cepas de fitoplasma, por lo tanto, una sola especie de vector es capaz de transmitir varios fitoplasmas al hospedante, permitiendo que los fitoplasmas sean capaces de infectar a cientos de plantas alrededor del mundo, provocando al menos 700 enfermedades (Weintraub y Beanland, 2006).

Los fitoplasmas secretan proteínas llamadas efectores mediante los cuales modulan la fisiología vegetal para facilitar su multiplicación en la planta (Padilla-Ramos *et al.*, 2019; Rashid *et al.*, 2018; Lo Presti *et al.*, 2015).

1.3 EFECTORES

1.3.1 PAPEL DE LOS EFECTORES Y SU RECONOCIMIENTO POR PARTE DE LAS PLANTAS

Un efector se define como una proteína o molécula pequeña (metabolito, enzima, etc.) secretada por el patógeno que se une selectivamente a una proteína y regula su actividad biológica; actúa como ligando y modifican la estructura y función de la célula del hospedante, lo que favorece la colonización del patógeno (Hogenhout *et al.*, 2009; Singh *et al.*, 2019). Un efector puede ejercer un efecto negativo o virulencia (suprimir los mecanismos de defensa de la planta) o tener un efecto positivo o avirulencia para el hospedante (activa la respuesta de defensa de la planta) (Hogenhout *et al.*, 2009; Lo Presti *et al.*, 2015).

Los efectores son fundamentales en la interacción patógeno-hospedante, en donde el patógeno depende metabólica y evolutivamente de su hospedante. La interacción ocurre por medio de proteínas efectoras y que son las responsables del éxito en diferentes tipos de patógenos (Padilla-Ramos *et al.*, 2019). Estas proteínas efectoras se translocan en las células del hospedante e interaccionan con moléculas involucradas en las funciones celulares normales, lo que causan diversos efectos como anular la capacidad de autofagia del hospedante, cambios en su morfología, supresión de la muerte celular programada, entre otros. En la tabla 1.3 se muestran ejemplos de distintos efectores microbianos y la función que ejercen en el hospedante (Padilla-Ramos *et al.*, 2019).

Tabla 1.3 Algunas proteínas efectoras y su función reportadas en estudios de interacción entre patógenos y plantas hospedantes (Revisado en Padilla-Ramos *et al.*, 2019)

Organismo	Efector (proteína)	Función en el hospedero (Objetivos operativos)	Referencia
<i>Cladosporium fulvum</i> (hongo)	Ecp2	Reconocimiento de secuencias específicas	(Dagdas <i>et al.</i> , 2016)
<i>Candidatus Phytoplasma trifolii</i> (Bacteria)	SAP54	Cambios en la morfología de la planta, inducen fenotipos adaptativos para los vectores (filodia)	(Amselem <i>et al.</i> , 2015)
<i>Fusarium oxysporum</i> (hongo)	Avr2	Interfieren de forma específica a mecanismos de defensa conservado para la resistencia contra genes de virulencia de patógenos	(McCann, 2016)

Tabla 1.3 Algunas proteínas efectoras y su función reportadas en estudios de interacción entre patógenos y plantas hospedantes (Revisado en Padilla-Ramos *et al.*, 2019)

Organismo	Efector (proteína)	Función en el hospedero (Objetivos operativos)	Referencia
<i>Phytophthora infestans</i> (oomiceto)	PexRD54	Suprime la capacidad de autofagia selectiva en el hospedero, las plantas usan la autofagia para protegerse contra los patógenos	(Washington <i>et al.</i> , 2016)
<i>Pseudomonas syringae</i> (bacteria)	AvrRpt2	Actividad proteolítica contra al menos cinco proteínas de Arabidopsis, incluyendo el regulador de defensa negativo RIN4	(Chisholm <i>et al.</i> , 2006)
<i>Xanthomonas vesicatoria</i> (bacteria)	AvrBs3	Actúa como un activador transcripcional, uniéndose a una secuencia promotora	(Römer <i>et al.</i> , 2007)
<i>Pseudomonas syringae</i> (bacteria)	HopAF1	Suprime la resistencia de la planta, este efector se encuentra en los genomas de muchas bacterias de este tipo y posiblemente esté relacionado con la proteína deamidasa	(Hughes y Banfield, 2014)
<i>Phytophthora sojae</i> (oomiceto)	Avr1b	Suprime la muerte celular programada	(Dou <i>et al.</i> , 2008)
<i>Xanthomonas vesicatoria</i> (bacteria)	AvrBs3	Hipertrofia celular. Se cree que estas lesiones cancerosas facilitan la liberación bacteriana del tejido infectado y aumentan su diseminación	(Kay <i>et al.</i> , 2007)

Los efectores se dividen en dos grupos: efectores apoplásticos y efectores citoplasmáticos. Los efectores apoplásticos son secretados en los espacios extracelulares de las células del hospedante, mientras que los efectores citoplasmáticos son secretados dentro de las células del hospedante mediante un dominio de translocación que se encuentra generalmente después del péptido señal (Guzmán, 2017; Martínez, 2015; Hogenhout *et al.*, 2009).

Para poder transportar estas proteínas dentro de la célula del hospedante, los patógenos utilizan distintos sistemas o estructuras; generalmente los hongos y oomicetos secretan sus efectores mediante haustorios (Padilla-Ramos *et al.*, 2019; Martínez, 2015). Sin embargo, en las bacterias se han descrito siete sistemas de secreción (SST1-SST7), los cuales les permiten transportar sus proteínas al espacio extracelular y mantener una relación simbiótica o de parasitismo con el hospedante.

Los sistemas de secreción de tipo III, IV y VI (SST3, SST4 y SST6) se conocen como “nanojeringas” o “inyectisomas”, los cuales poseen una estructura que forma un conducto capaz de transportar proteínas efectoras directamente de la célula donadora (bacteria) a la célula receptora (hospedante) mediante un mecanismo de un solo paso, por lo que son Sec-independientes (Bernal-Bayard, 2010a; Cabañas y Huerta, 2014). En los otros

sistemas el mecanismo ocurre en dos pasos, Sec-dependientes, en el cual las proteínas primero necesitan de una ruta general secretora para translocarse de la membrana interna al periplasma; después la proteína atraviesa la membrana externa logrando ser secretada. Generalmente utilizan la vía Sec, vía específica para la secreción de proteínas no plegadas y en algunos casos la vía Tat (twin arginine translocation system) específica para la secreción de proteína plegadas (Bernal-Bayard, 2010a; De Paz Fernández, 2010; Green y Mecsas, 2016).

Los fitoplasmas poseen dos sistemas de secreción, el sistema YidC y el sistema Sec, el primero para la integración de proteínas a la membrana y el segundo para la integración y secreción de proteínas en el citoplasma de la célula hospedante (Oshima *et al.*, 2013).

1.3.2 EFECTORES IDENTIFICADOS EN LOS FITOPLASMAS

Los fitoplasmas son fitopatógenos capaces de infectar plantas y sus insectos vectores, secretan efectores de virulencia para inducir diferentes modificaciones en el hospedante (Rashid *et al.*, 2018; Bulgari *et al.*, 2011; Hogenhout *et al.*, 2009). Sin embargo, aunque los fitoplasmas son patógenos que en las plantas solamente se encuentran en el floema, sus proteínas efectoras se han encontrado en otros tejidos (Rashid *et al.*, 2018; Hoshi *et al.*, 2009). Se ha reportado que estas proteínas efectoras poseen tamaños <40 kDa y probablemente son transportados a través de los plasmodesmos; en las células vegetales el tamaño límite de exclusión (TLE) que conecta a las células cribosas y las células acompañantes son >67 kDa, mientras que entre otras células el TLE oscila entre 10 a 50 kDa (Sugio *et al.*, 2011a), lo que explicaría porque son capaces de circular fácilmente dentro de las células vegetales.

Las proteínas efectoras de fitoplasmas más descritas al momento son: TENGU, SAP11, SAP54/PHYL1 y SAP05, las cuales ocasionan cambios morfológicos en las especies vegetales, debido a que actúan sobre distintos factores de transcripción involucrados en el correcto desarrollo de la planta (Tabla 1.4). En consecuencia, las plantas quedan en su fase vegetativa, incapaces de reproducirse, convirtiéndose únicamente en hospedante para los

fitoplasmas, razón por lo cual han sido descritos que convierten a las plantas en zombies (Yong, 2014).

Tabla 1.4 Lista de los efectores de fitoplasmas más conocidos actualmente.
(Tabla adaptada de Huang *et al.*, 2021; MacLean *et al.*, 2014; Maejima *et al.*, 2014; Hoshi *et al.*, 2009; Sugio *et al.*, 2011b).

Efectores de fitoplasmas					
Efactor	Tamaño	Identificado	Síntomas	Altera	Fitohormonas Afectadas
TENGU	4.5 kDa	OY (Hoshi <i>et al.</i> , 2009)	Escoba de bruja y enanismo	ARFs	↓ niveles Auxinas y JA
SAP11	9 kDa	AY-WB (Sugio <i>et al.</i> , 2011b)	Escoba de bruja	TCPs	↓ JA
SAP54	10.7 kDa	AY-WB (MacLean <i>et al.</i> , 2014)	Filodia	MTFs	No descrito
PHYL1	10.6 kDa	OY-W (Maejima <i>et al.</i> , 2014)	Filodia	MADS	No descrito
SAP05	12.3 kDa	AY-WB (Huang <i>et al.</i> , 2021)	Escoba de bruja	SPLs, GATAs	No descrito

TENGU: el primer efector de fitoplasma descrito fue el inductor de Tengu-su (TENGU); se denominó TENGU porque en Japón provoca la enfermedad de Tengu-su (nido de Tengu) o síntoma de escoba de bruja, caracterizado por el desarrollo de numerosas ramas pequeñas y cortas que parecen el nido de Tengu, un mítico duende japonés que vive en las montañas y vuela por el cielo (Hoshi *et al.*, 2009). La preproteína TENGU es de 70 aminoácidos (aa) y posee un péptido señal (PS) de 32 aa en su extremo N-terminal, el cual es eliminado al momento de ser secretada por el fitoplasma, generando una proteína madura de 38 aa de longitud y 4.5 kDa.; este efector altera la morfología de la planta infectada, disminuye los niveles de auxinas y del ácido jasmónico, dos fitohormonas involucradas para la floración y la respuesta de defensa en las plantas, resultando en esterilidad de la especie vegetal (Sugawara *et al.*, 2013; Minato *et al.*, 2014; Hoshi *et al.*, 2009).

TENGU se encontró dentro de las células del meristemo apical y en los brotes axilares de especies vegetales infectadas con fitoplasmas causantes del amarillamiento de la cebolla (por sus siglas en inglés OY de Onion-Yellow). Expresado en *Arabidopsis thaliana* y *Nicotiana benthamiana* indujo enanismo y síntomas de escoba de bruja, típicos de una

infección por fitoplasmas (Hoshi *et al.*, 2009). TENGU afecta los factores de transcripción ARF6 y ARF8 los cuales regulan la respuesta hacia auxinas (Minato *et al.*, 2014); durante la floración, ambos se encuentran involucrados en la maduración del estambre y el gineceo dependiente del ácido jasmónico (Nagpal *et al.*, 2005). En la figura 1.7 se observan los síntomas característicos del efector TENGU.



Fig. 1.7. Síntomas de enanismo y escoba de bruja que causa el efector TENGU (centro y derecha) comparado con el control (izquierda) en plantas transgénicas de *A. thaliana*. Fotografía tomada de Hoshi *et al.* (2009).

Los efectores SAPs fueron identificados en el genoma del fitoplasma causante del amarillamiento del áster-escoba de bruja o AY-WB (de acuerdo con sus siglas en inglés); por ser proteínas secretadas del AY-WB se les denominó SAP (secreted AY-WB proteins), se lograron identificar 56 proteínas efectoras candidatas (Bai *et al.*, 2009). Los efectores SAP que más se han estudiado son SAP11, SAP54 y SAP05; éstos han sido caracterizados funcionalmente encontrándose que interfieren en el desarrollo normal de la planta y que alteran su proceso de floración. A continuación, se presenta una descripción de cada uno.

SAP11: fue el primer efector SAP de virulencia caracterizado; es una proteína pequeña que antes de ser secretada posee una longitud de 121 aa y un PS en el extremo N-terminal de 31 aa; además posee una señal de localización nuclear (NLS) bipartita de 18 aa (entre la posición 40-57) que lo dirige al núcleo de las células vegetales. En el núcleo se genera una proteína madura (sin péptido señal) de 90 aa (Bai *et al.*, 2009) con un peso molecular de 9 kDa (Sugio *et al.*, 2011b). Este efector se caracteriza por producir el síntoma de escoba de

bruja y disminución en la síntesis de ac. Jasmónico (AJ) (Lamilla-Monje y Franco-Lara, 2019; Rashid *et al.*, 2018; Sugio *et al.*, 2011b); una fitohormona indispensable para el crecimiento y fisiología vegetal, así como la regulación de la respuesta de defensa frente a patógenos, plagas y heridas mecánicas (Li *et al.*, 2022; Ruan *et al.*, 2019; Eng, 2008).

SAP11 se une y desestabiliza factores de transcripción TCP (TEOSINTE BRANCHED/CYCLOIDEA/PROLIFERATING CELL FACTOR), particularmente de la clase II CYCLOIDEA/ TEOSINTE BRANCHED-1 (CYCTB1) y CINCINNATA (CIN-TCP), lo que ocasiona una disminución en la expresión de la LIPOXYGENASA (LOX) y de la producción de Jasmonatos (Pecher *et al.*, 2019; Sugio *et al.*, 2014, Sugio *et al.*, 2011b). La familia Teosinte branched1/Cinninnata/proliferating cell factor (TCP) son un grupo de genes específicos de plantas, que codifican factores de transcripción con dominio TCP conservado; la familia está formada por cuatro proteínas que intervienen en diferentes procesos durante el crecimiento y desarrollo vegetal, como es el control floral, el desarrollo de las hojas, la dominancia apical, la germinación, la señalización hormonal, etc (Gastaldi, 2020; González-Grandío y Cubas, 2016). La lipoxigenasa es una enzima encargada de convertir el ac. α -linolénico en ac. 13- hidroxiperoxido-linolénico, un paso importante en la ruta de síntesis de jasmonatos.

La expresión de SAP11 lleva a la acumulación de esta proteína efectora en el núcleo de diferentes células, incluyendo células del mesófilo y tricomas de plantas de *N. benthamiana*, y causan cambios morfológicos como sobreproducción de células inmaduras, hojas grandes y arrugadas (Pecher *et al.*, 2019; Sugio *et al.*, 2011b). En la figura 1.8 se observa la localización de SAP11 en el núcleo de una célula vegetal (tomada de Bai *et al.*, 2009). El gen que codifica esta proteína efectora se encuentra dentro de regiones llamadas “unidades potencialmente móviles”, PMU de acuerdo con sus siglas en inglés (Bai *et al.*, 2009).

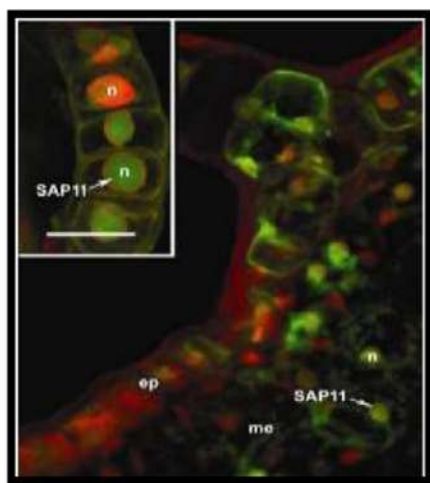


Fig. 1.8. Efecto SAP11 (conjugado con un anticuerpo fluorescencia verde) detectado en los núcleos (n) de células de tricoma en secciones de hojas de áster de China infectadas por AY-WB; (ep) epidermis y (me) el mesófilo. Foto tomada de Bai *et al.* (2009).

SAP54: Este efector de virulencia se caracteriza por provocar una condición llamada filodia, que son estructuras semejantes a hojas en donde deberían estar las flores (MacLean *et al.*, 2011); este efector tiene una longitud de 124 aa con la presencia de un PS de aproximadamente 34 aa. (Bai *et al.*, 2009). Esta proteína efectora es responsable de uno de los cambios morfológicos causados por los fitoplasmas de AY-WB. MacLean *et al.* (2011) generaron líneas transgénicas de *A. thaliana* y demostraron que SAP54 altera el desarrollo floral (Figura 1.9), ocasionando que muchas de las plantas sean estériles o produzcan semillas que no logran germinar.

SAP54 degrada factores de transcripción que poseen dominio MADS (MTF, por sus siglas en inglés), como SEPALLATA3 y APETALA1, mediante la interacción con proteínas de la familia RADIATION SENSITIVE23 (RAD23), ocasionando cambios importantes en el desarrollo floral (MacLean *et al.*, 2014). Los miembros de la familia MADS-box codifican factores de transcripción involucrados en importantes procesos en los eucariotas; en las plantas, estos factores de transcripción son indispensables para la identidad del órgano floral, de acuerdo con el modelo ABCE de desarrollo floral, se necesita la intervención de diversos genes homeóticos. La familia MADS-box se caracteriza por la presencia de un

dominio conservado de unión a ADN denominado dominio MADS; su nombre proviene de las iniciales de los genes de *Saccharomyces cerevisiae*, MINICHROMOSOME MAINTENANCE1 (MCM1), AGAMOUS (AG) en *A. thaliana*, DEFICIENS (DEF) en *Anthirrinum majus* y SERUM RESPONSE FACTORS (SRF) en *Homo sapiens*. La familia MADS-box se divide en dos subfamilias, tipo I y tipo II, las cuales se dividen en distintos grupos; en particular los genes de la subfamilia II interviene en varias etapas del desarrollo floral, determinando el tiempo de floración, la identidad del meristemo floral, identidad del órgano floral, formación del fruto y pigmentación de la semilla, por ejemplo, el gen APETALA1 (AP1) interviene en la identidad del meristemo y del órgano floral, mientras que SEPALLATA3 (SEP3) en la identidad del órgano floral (Curto-Prieto, 2013; Theißen *et al.*, 2016; Saedler *et al.*, 2001; Theissen *et al.*, 2000). Por lo tanto, al degradar este tipo de factores de transcripción, SAP54 afecta el adecuado desarrollo de la planta y ocasiona los síntomas morfológicos característicos causados por los fitoplasmas.



Fig. 1.9. Efecto del efector SAP54 sobre el desarrollo normal de las flores de *A. thaliana* (control) comparadas con plantas agro-infiltradas con este efector. Foto tomada de MacLean *et al.* (2011).

SAP05: Recientemente se reportó esta nueva proteína efectora SAP, caracterizada por Huang *et al.* (2021). Estos autores observaron que líneas transgénicas de *A. thaliana* que expresan SAP05 muestran cambios fenotípicos característicos por fitoplasmas, ocasionando el síntoma semejante a la escoba de bruja, retraso en el crecimiento, proliferación de brotes y esterilidad. SAP05, antes de ser secretado, posee un peso molecular de 12.3 kDa y una longitud de 135 aa. que incluye un PS de 32 aa (Bai *et al.*,

2009). Se ha determinado que es capaz de causar el síntoma de escoba de bruja mediante la unión y degradación de dos distintas familias de factores de transcripción, la familia SPL y la familia GATA.

Las proteínas generalmente son degradadas por el proteasoma 26S, las cuales se marcan previamente con una molécula llamada ubiquitina, pero el efector fitoplasmático SAP05 utiliza una vía independiente de ubiquitina, este se enlaza al receptor de ubiquitina RPN10 del proteasoma para mediar la degradación de los factores de transcripción GATA y SPL, formando un complejo ternario SPL/GATA-SAP05-RPN10 (Huang *et al.*, 2021). Los factores de transcripción de unión a secuencias GATAs (Guanina-Adenina-Timina) son proteínas que poseen uno o más dominios conservados tipo dedos de zinc o zinc fingers, de unión a ADN, que están involucrados en la regulación de la transcripción en respuesta a estímulos de luz y regulan la expresión de genes homeóticos florales en *Arabidopsis* (Behringer y Schwechheimer, 2015; Zhang *et al.*, 2013). Mientras que los SPLs (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE) son una familia de factores de transcripción que se caracterizan por tener una región altamente conservada de 76 aminoácidos denominada dominio SBP; son proteínas importantes para la transición de la fase vegetativa a la fase reproductiva en las plantas, interviniendo en el tiempo de floración, ramificación y tasa de iniciación de hojas (Xu *et al.*, 2016; Preston y Hileman, 2013; Wang *et al.*, 2009). En la figura 1.10 se observa un modelo de como se produce la alteración morfológica por SAP05 (Huang *et al.*, 2021).

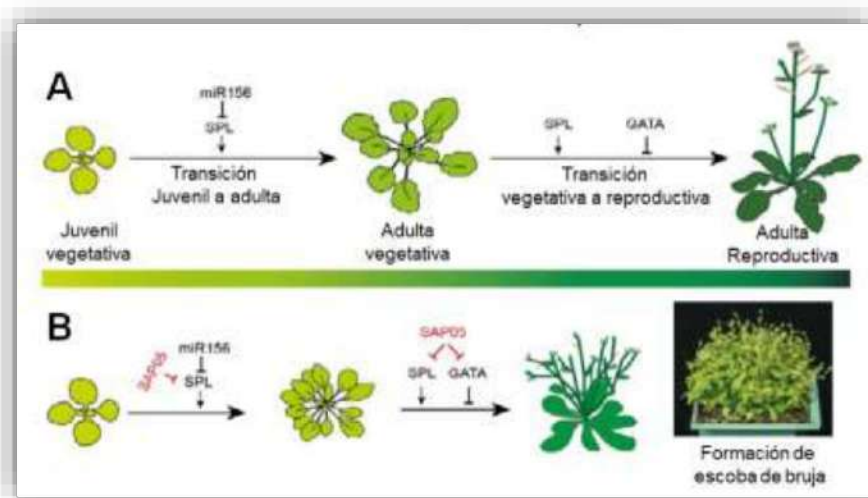


Figura 1.10. Modelo de inducción de escoba de bruja por SAP05. (A) Etapas normales en una planta sana. (B) Desestabilización de los factores de transcripción SPL y GATA por el efector SAP05 y consecuente formación del síntoma de escoba de bruja. Tomado de Huang *et al.* (2021).

PHYL1: el efector fitoplasmático que causa síntomas de filodia 1 (PHYL1) fue identificado en el genoma de OY-W (fitoplasma del amarillamiento de la cebolla, de acuerdo con sus siglas en inglés); es un homólogo de SAP54, con 88 % de identidad. Se encuentra dentro de la denominación de “filógenos”, así como SAP54 por ser miembros de la familia de genes inductores de filodia (Maejima *et al.*, 2014). Este efector posee una longitud de 125 aa con un PS en la región N-terminal de 34 aa, que es cortado al momento de ser secretado, dando origen a una proteína madura de 91 aa y 10.6 kDa. PHYL1 degrada las proteínas homeóticas florales con dominio MADS, como son SEPALLATA3 (SEP3), APETALA1 (AP1) y CAULIFLOWER (CAL), lo que ocasiona la inhibición de sus funciones durante el proceso de floración (Maejima *et al.*, 2014).

La floración es un proceso complejo en donde participan múltiples factores endógenos y exógenos; se puede sintetizar en tres etapas fundamentales: 1. transición de la fase vegetativa a la fase reproductiva, 2. establecimiento y mantenimiento del meristemo floral y 3. desarrollo de los órganos florales (Azcón-Bieto y Talón, 2013; Alvarez-Buylla *et al.*, 2010; Taiz y Zeiger, 2002). El proceso de identidad floral o desarrollo de los órganos florales (etapa 3) se puede comprender por medio del modelo ABCE (originado a partir del modelo ABC), el cual está compuesto por cuatro clases de genes homeóticos (genes pertenecientes a la clase A, B, C y E) o también llamados genes de identidad floral; los genes ABCE son factores de transcripción pertenecientes de la familia MADS-box. El modelo ABCE explica que la combinación de estas proteínas homeóticas florales o proteínas MADS-box dan lugar al cuarteto del órgano floral: sépalos, pétalos, estambres y carpelos, de tal manera que para originar los sépalos es necesario la combinación de las proteínas que pertenecen dentro de la clase A y la clase E; para originar los pétalos se combinan la clase A, B y E; para los estambres la clase B, C y E, y por último para los carpelos interactúan las de la clase C y E. Dentro de la clase A se encuentran los genes APELATA 1 (AP1) y FRUITFULL (FUL), en la clase B se encuentran APELATA 3 (AP3),

PISTILLATA (PI) y GLOBOSA (GLO), en la clase C están los genes AGAMOUS (AG) y la clase E incluye SEPALLATA1 (SEP1), SEP2, SEP3 y SEP4 (Chen *et al.*, 2019; Mendoza y Encina, 2018; Alvarez-Buylla *et al.*, 2010; Theißen y Saedler, 2001).

En la figura 1.11 se observa el modelo propuesto por Maejima *et al.* (2014) de la inducción de la filodia provocada por PHYL1, en donde se puede observar el modelo ABCE, la interacción de las proteínas MADS-box y el consecuente desarrollo de la filodia.

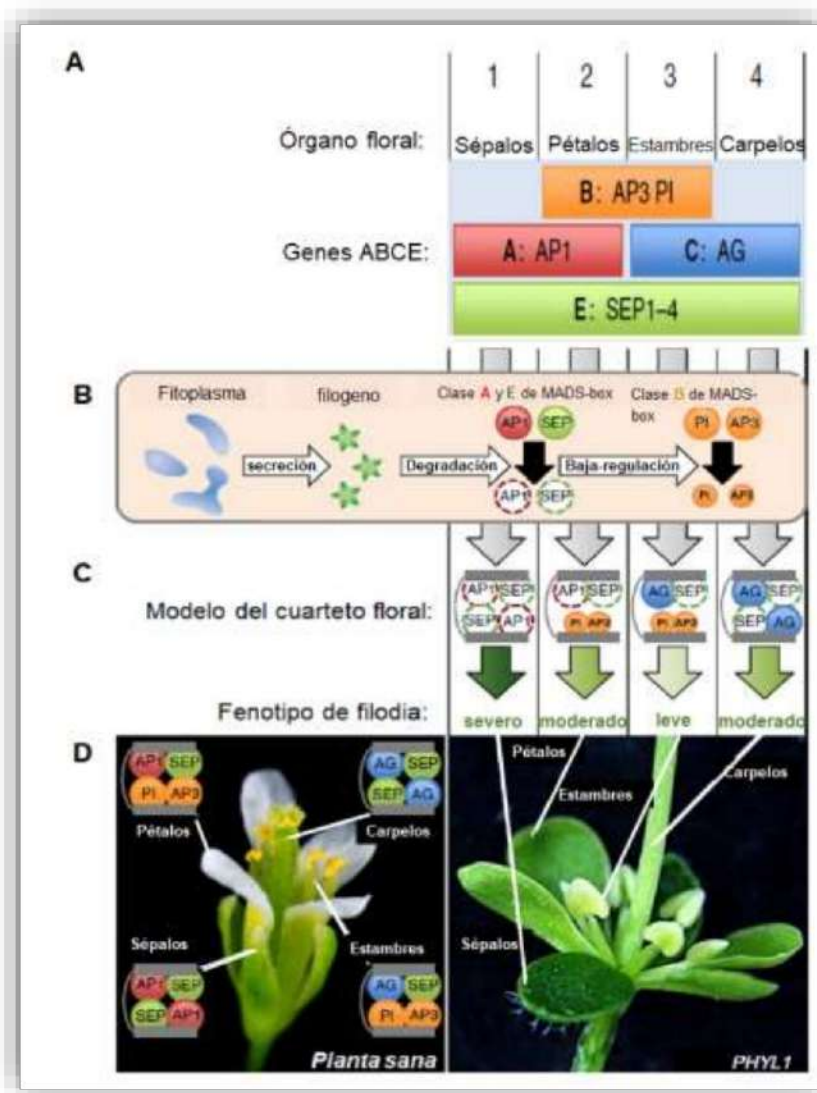


Figura 1.11. Modelo de inducción de la filodia por el efector PHYL1. (A) Modelo ABCE de identidad floral. (B) Degradación de proteínas pertenecientes a la clase A y E por el efector PHYL1. (C) Combinación de

las diferentes clases de proteínas homeóticas florales cuyo daño provoca la alteración del fenotipo; se observa el grado de daño en la formación de cada órgano del cuarteto floral. (D) Comparación de una planta sana (izquierda) con la planta con síntoma de filodia (derecha). Modelo propuesto por Maejima *et al.* (2014).

La mayoría de los efectores que se han caracterizado, así como parte de los efectores candidatos, se encuentran codificados dentro de islas de patogenicidad; estas regiones son denominadas unidades potencialmente móviles PMU (Potential Mobile Unit) (Bai *et al.*, 2006), de las que se hablará a continuación.

1.4 UNIDADES POTENCIALMENTE MÓVILES (PMU)

Varios genes que codifican para proteínas efectoras fitoplasmáticas se han encontrado agrupados en regiones denominadas “Unidades Potencialmente Móviles” (PMUs, por sus siglas en inglés), mosaicos variables (SVMs) o unidades de genes móviles (MUGs) (Lamilla Monje y Franco Lara, 2019; Rashid *et al.*, 2018; Dickinson, 2010; Jomantiene *et al.*, 2007; Bai *et al.*, 2006). Los PMUs son regiones genómicas semejantes a transposones, de aproximadamente 20 Kb; estos elementos móviles son secuencias repetidas de Adenina-Timina (AT) que pueden estar en forma lineal (dentro del cromosoma) o forma circular (extracromosómico) (Dickinson, 2010; Bai *et al.*, 2006). Las regiones PMU comprenden varios genes en un orden conservado y agrupados en tándem o en múltiples repeticiones; entre estos están genes que codifican funciones importantes de la replicación y recombinación del ADN (Dickinson, 2010). Se caracterizan por contener un grupo de genes centrales que incluyen al gen de la transposasa (*tra5*), ADN primasa (*dnaG*), ADN helicasa (*dnaB*), timidilato cinasa (*tmk*), proteasa dependiente de Zn (*hflB*), proteínas reguladoras de unión al ADN HU (*himA*), proteínas de unión al ADN monocatenario (*ssb*) y factores sigma especializados (*sigF*) (Bai *et al.*, 2006).

Dentro de estos elementos móviles se encuentran también genes que codifican para proteínas secretadas, como el efector SAP11 y otras proteínas secretadas candidatas de efectores (Bai *et al.*, 2009). Los PMUs se han encontrado dentro del genoma de diferentes fitoplasmas (Huang *et al.*, 2021; Cho *et al.*, 2019; Andersen *et al.*, 2013; Chung *et al.*, 2013;

Bai *et al.*, 2006), por lo que es probable que estas regiones se encuentren involucradas en la adaptación del fitoplasma a sus diferentes hospedantes y funcionen como islas de patogenicidad. Los PMU correspondientes al cromosoma del AY-WB, denominado PMU1, se han encontrado flanqueados por secuencias de inserción *tra5* (IS) con repeticiones invertidas (IR) de 327 pb (Bai *et al.*, 2006).

En la figura 1.12 se observa la distribución de los genes centrales en PMU1 y los genes que codifican para proteínas secretadas en el interior de estos elementos móviles.

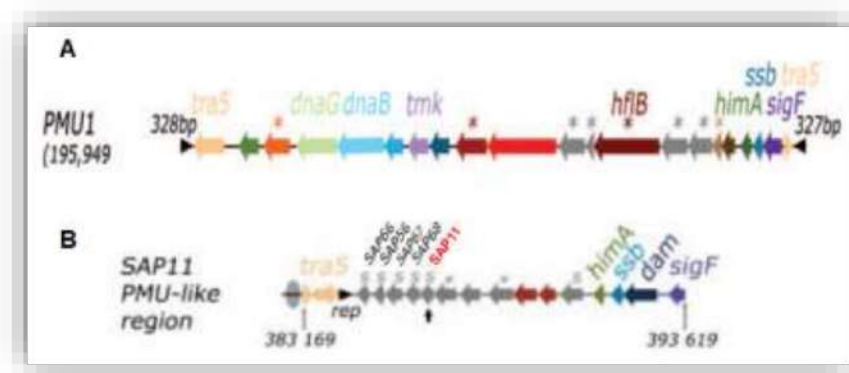


Figura 1.12. Unidades potencialmente móviles en el cromosoma de AY-WB (PMU1). (A) Se observan los genes centrales (marcados en colores) característicos de los PMU. (B) Proteínas secretadas dentro de regiones semejantes a PMU; efector SAP11 (marcado en rojo) y proteínas efectoras candidatas (marcadas en gris). Modificado de Bai *et al.* (2009) y Bai *et al.* (2006).

Con base en todo lo anterior, se resume que el cocotero es una especie de gran interés comercial que es afectada por fitoplasmas, éstos causan importantes pérdidas en su producción. Actualmente existen herramientas para su diagnóstico, se ha estudiado su epidemiología, su método de dispersión y sus hospedantes. Por otro lado, en la fitopatología los efectores ocupan un lugar central para entender los patosistemas, y los mecanismos moleculares que gobiernan la interacción patógeno-hospedero. La efectorómica es un área en desarrollo, y al momento se han validado sólo 21 efectores de algunos fitoplasmas (Wang *et al.*, 2023; Chen *et al.*, 2022; Arricau-Bouvery *et al.*, 2021; Huang *et al.*, 2021; Wang *et al.*, 2017; MacLean *et al.*, 2014; Maejima *et al.*, 2014; Sugio *et al.*, 2014; Boonrod

et al., 2012; Hoshi *et al.*, 2009; Kakizawa *et al.*, 2009), pero ninguno del fitoplasma causal del ALC.

En este trabajo se planteó identificar efectores del fitoplasma responsable del ALC secuenciando su genoma e iniciar el estudio de estos importantes determinantes de patogenicidad y virulencia de este fitoplasma. Hasta el momento, los efectores de fitoplasmas que se han caracterizado, interaccionan con factores de transcripción indispensables en el desarrollo de la planta (Huang *et al.*, 2021; Maejima *et al.*, 2014; MacLean *et al.*, 2011), afectan la biosíntesis o los niveles de fitohormonas involucradas en la defensa vegetal y adecuado desarrollo (Minato *et al.*, 2014; Sugio *et al.*, 2011b), poseen un péptido señal en la región N-terminal (Maejima *et al.*, 2014; Bai *et al.*, 2009; Hoshi *et al.*, 2009) y algunos han mostrado señales de localización nuclear bipartita, como el efector SAP11 (Bai *et al.*, 2009). Al momento son pocos los efectores de fitoplasmas que han sido caracterizados, por lo que es necesario realizar más estudios para conocer mejor sobre estas proteínas efectoras. En el caso de los efectores del fitoplasma responsable del ALC no se conocía ningún efector, por lo que era importante iniciar esta área de estudio.

HIPÓTESIS

El fitoplasma 'Ca. P. palmae' causante del amarillamiento letal del cocotero comparte algunos efectores conservados con otros grupos de fitoplasmas.

JUSTIFICACIÓN

Los fitoplasmas son procariontes, patógenos que causan importantes pérdidas de producción en diferentes cultivos, incluido el cocotero, una especie vegetal de interés comercial a nivel mundial. Estos patógenos son capaces de secretar moléculas especializadas llamadas efectores que causan alteraciones en la planta hospedera e interfieren en los mecanismos de resistencia vegetal. Es crucial identificar estas proteínas efectoras y comprender las interacciones entre el hospedante y el patógeno. Conocer la función de los efectores puede ayudar a identificar sus dianas y eventualmente desarrollar nuevos productos fitosanitarios. A la fecha, se han caracterizado pocos efectores de fitoplasmas, y no se ha descrito ningún efector de fitoplasma perteneciente al grupo 16SrIV. Por lo que en este trabajo se realizaron diferentes estrategias para identificar los efectores en el fitoplasma que causa el amarillamiento letal del cocotero. Estas estrategias incluyeron el uso de herramientas computacionales y técnicas moleculares, para la caracterización *in silico* de los efectores disponibles en otros fitoplasmas, la creación de un algoritmo para su identificación, así como la secuenciación del genoma del fitoplasma del ALC, la identificación de sus efectores y la caracterización de su expresión.

OBJETIVO GENERAL

Identificar efectores en el fitoplasma del ALC, 'Ca. P. palmae'

OBJETIVOS ESPECÍFICOS

- Realizar una caracterización *in silico* de los efectores verdaderos y predichos de fitoplasmas e identificar características comunes entre los diferentes fitoplasmas.
- Identificar efectores comunes entre los diferentes fitoplasmas (potenciales efectores conservados) mediante el análisis *in silico* de efectores verdaderos y predichos.
- Crear un algoritmo para la búsqueda e identificación de efectores de fitoplasmas.
- Identificar los efectores del fitoplasma causante del amarillamiento letal en *Cocos nucifera* L. (ALC) mediante la secuenciación de su genoma.
- Caracterizar la expresión de algunos efectores identificados del fitoplasma causante del amarillamiento letal

ESTRATEGIA EXPERIMENTAL

Para la realización de este trabajo de tesis se llevaron a cabo cuatro estrategias experimentales, como se muestra en la siguiente figura.

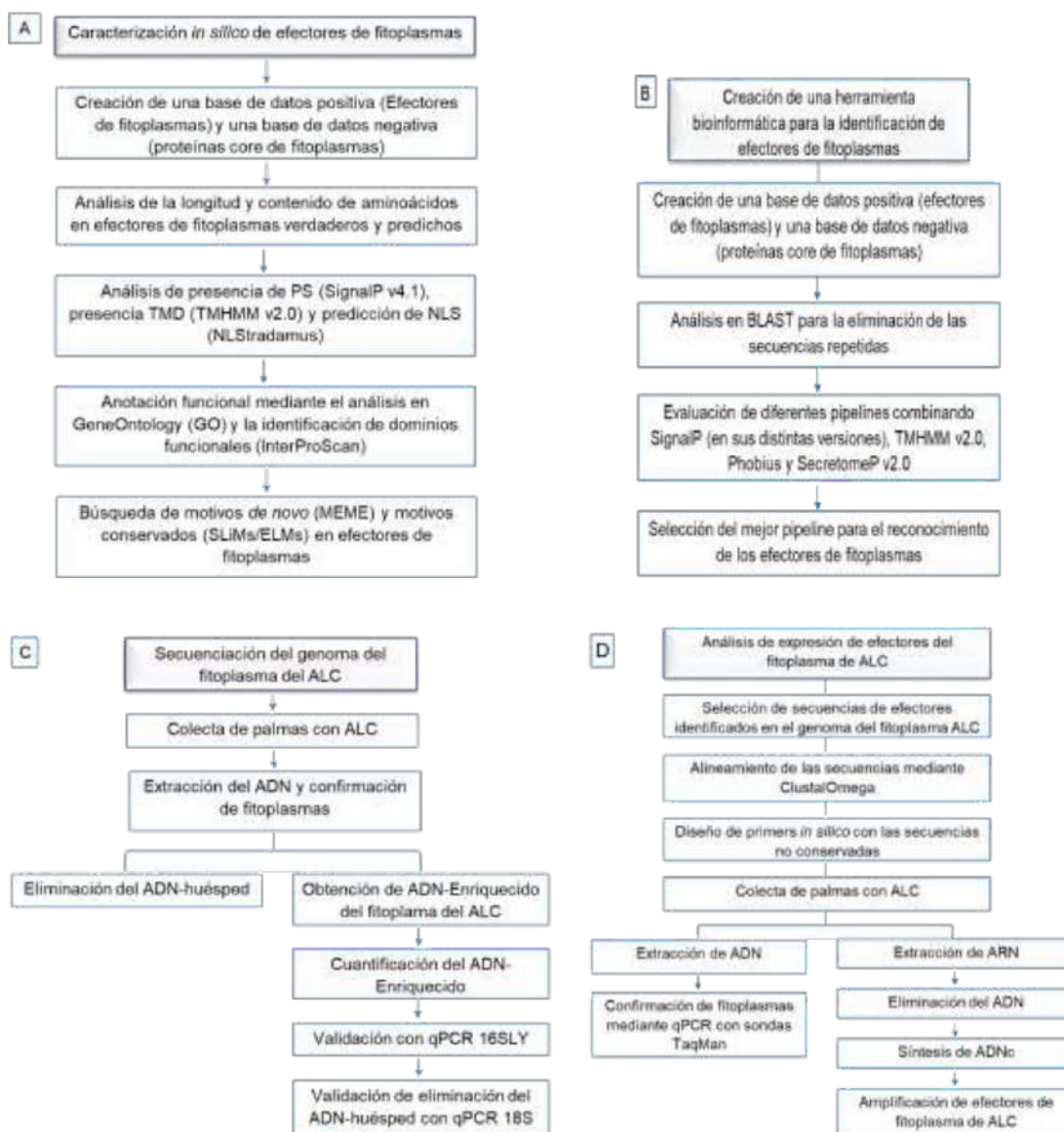


Figura 1.13 Diagrama de las estrategias experimentales. (A) Metodología para la caracterización *in silico* de efectores de fitoplasmas. (B) Metodología para la elaboración de un programa de búsqueda de efectores de fitoplasmas. (C) Esquema para la secuenciación del genoma del fitoplasma de ALC mediante la técnica de ADN-enriquecido. (D) Esquema para el análisis de expresión de efectores de fitoplasma de ALC.

CAPÍTULO II

ARTÍCULO. NOVEL INSIGHTS INTO PHYTOPLASMA EFFECTORS

Autores:

Karla Gisel Carreón-Anguiano⁺, Sara Elena Vila-Luna⁺, Luis Sáenz-Carbonell y Blondy Canto-Canché*

⁺ Estos autores contribuyeron igualmente en este trabajo

Cita: Carreón-Anguiano, K.G.; Vila-Luna, S.E.; Sáenz-Carbonell, L.; Canto-Canché, B. Novel Insights into Phytoplasma Effectors. *Horticulturae*. 2023, 9, 1228. <https://doi.org/10.3390/horticulturae9111228>

Abstract: Effectoromics has become integral to the identification of pathogen targets and/or host-resistant proteins for the genetic improvement of plants in agriculture and horticulture. Phytoplasmas are the causal agents of more than 100 plant diseases in economically important crops such as vegetables, spices, medicinal plants, ornamentals, palms, fruit trees, etc. To date, around 20 effectors in phytoplasmas have been experimentally validated but the list of putative effectors comprises hundreds of different proteins. Very few families (tribes) have been identified based on homology, such as the SAP05-like, SAP11-like, SAP54-like and TENGU-like families. The lack of conservation in amino acid sequences slows the progress of effectoromics in phytoplasmas since many effectors must be studied individually. Here, 717 phytoplasma effector candidates and 21 validated effectors were characterized *in silico* to identify common features. We identified functional domains in 153 effectors, while 585 had no known domains. The most frequently identified domain was the sequence-variable mosaic domain (SVM domain), widely distributed in 87 phytoplasma effectors. Searching for *de novo* amino acid motifs, 50 were found in the phytoplasma effector dataset; 696 amino acid sequences of effectors had at least 1 motif while 42 had no motif at all. These data allowed us to organize effectors into 15 tribes, uncovering, for the first time, evolutionary relationships largely masked by lack of sequence conservation among effectors. We also identified 42 eukaryotic linear motifs (ELMs) in phytoplasma effector sequences. Since the motifs are related to common functions, this novel

organization of phytoplasma effectors may help further advance effectoromics research to combat phytoplasma infection in agriculture and horticulture.

2.1. INTRODUCTION

Phytoplasmas (Kingdom, Bacteria; Phylum, Mycoplasmatota; class, Mollicutes; genus, '*Candidatus* Phytoplasma' or '*Ca.* Phytoplasma') are mycoplasma-like cell wall-less pathogenic microorganisms transmitted by phloem-feeding insect vectors (Oshima *et al.*, 2013; Hogenhout *et al.*, 2008; Bertaccini, 2007). These pathogens reside in plant phloem and cause severe damage to the agriculture and horticulture industries worldwide, with extensive yield losses in economically important crops such as vegetables, spices, medicinal plants, ornamentals, palms, fruit trees, among others (Kirdat *et al.*, 2023). In China, for example, more than 100 phytoplasma diseases have been reported (Wang *et al.*, 2022). Phytoplasmas secrete virulence proteins known as effectors, which interfere with host hormone signaling (Rashid *et al.*, 2018; Sugio *et al.*, 2011) and cause abnormal plant morphologies such as phyllody, virescence, chlorosis, and witches' broom, among other symptoms (Bertaccini, 2022; Hemmati *et al.*, 2021). Integrated management of the main phytoplasma diseases is expected to include the use of resistant plant material (Oliveira *et al.*, 2019) and novel strategies for environmentally friendly insect vector control (Tzec-Simá *et al.*, 2022). Because effectors are essential for the virulence of the pathogen, they are also susceptible targets for the control of phytoplasma-associated diseases (Galetto *et al.*, 2021; Tomkins *et al.*, 2018). They are also suitable tools for the identification of protein targets in the hosts, including the nucleotide-binding leucine-rich repeat (NLR) receptors (also called resistance proteins). Additionally, genetic engineering of NLR receptors can improve host recognition of pathogen effectors, enabling the production of resistant host lines (Cadiou *et al.*, 2023; Zdrzałek *et al.*, 2023).

To date, only around 20 phytoplasma effectors have been experimentally validated, which include SAP05 (Huang *et al.*, 2021), SAP11 (Sugio *et al.*, 2014), SAP54/Phyl1 (MacLean *et al.*, 2014), TENGU (Sugawara *et al.*, 2013; Hoshi *et al.*, 2009), SWP1, SWP11, SWP12, SWP21 (Wang *et al.*, 2018), Zaofeng3 and Zaofeng6 (Chen *et al.*, 2022), and the membrane-bound proteins IdpA (Kakizawa *et al.*, 2009), Imp (Boonrod *et al.*, 2012), VmpA

(Arricau-Bouvery *et al.*, 2021) and Amp (Wang *et al.*, 2023). A number of novel effectors have been reported in the in silico characterization of phytoplasma genomes, for example, 257 hypothetical proteins in ‘*Ca. Phytoplasma solani*’ (Music *et al.*, 2019) or 7 unique effectors in Flavescence dorée phytoplasma (Debonneville *et al.*, 2022). The list of phytoplasmas effectors keeps growing as more phytoplasma genomes are sequenced.

Different strategies have been used to classify effectors in other microbial kingdoms. In bacteria, effectors are classified according to the secretion system (type III, type IV) through which they are exported or translocated from the pathogen cell to the host (Wagner *et al.*, 2022; Noroy *et al.*, 2019). In fungi, effectors that meet certain structural criteria such as small size, high cysteine content, presence of a signal peptide (SP) and absence of transmembrane domains (TMDs) are termed canonical or classical effectors, while those that do not meet some of these criteria are termed noncanonical (Wilson and McDowell, 2022; Carreón-Anguiano *et al.*, 2020; Sperschneider *et al.*, 2018); this classification has helped to expand the size of fungal effectoromes, since the noncanonical effectors that were previously discarded contribute significantly to the overall effectorome (Carreón-Anguiano *et al.*, 2020). Phytoplasma effectors are usually described as small, secreted proteins and like with fungi, they are secreted through the Sec-dependent type II secretion system (Tan *et al.*, 2021; Bai *et al.*, 2009; Music *et al.*, 2019;), although recent developments have challenged this as a bona fide phytoplasma effector characteristic. Recently, six nonclassical effectors (ncSecPs) were identified in ‘*Ca. Phytoplasma ziziphi*’; these effectors lack signal peptides or translocation signals but are secreted through a Sec-independent secretion pathway. Agroinfiltration of these effectors in *Nicotiana benthamiana* Domin suppressed the hypersensitive response (HR) by enhancing the expression of the cell death suppressor genes PR-1 and PR-5 (Gao *et al.*, 2023). Similarly, a few known phytoplasma effectors are also transmembrane proteins, such as Imp (immunodominant membrane protein) (Boonrod *et al.*, 2012) and Amp (antigenic membrane protein) (Wang *et al.*, 2023), demonstrating that the phytoplasma effector definition must evolve from the current terminology, “secreted proteins”. The identification of larger and more complex effectoromes gives rise to further classifications of phytoplasma effector; thus, there is a need to identify effector families based on common characteristics that can further facilitate their organization and investigation. In other kingdoms, effector families, or “tribes”, have been defined based on sequence homology (Stern and Han, 2022; Jones *et al.*, 2021), common domains (Rocafort

et al., 2022; Seong and Krasileva, 2021; Dean, 2011;), or common motifs (Liu *et al.*, 2019; Solé *et al.*, 2012; Jiang *et al.*, 2008), among other criteria. The last two criteria are very useful, as many effectors do not share sequence homology among the organisms of a given microbial kingdom.

The inculturability of phytoplasma in artificial media greatly complicates research for specific cures. Effectoromics could become an effective approach to searching for targets in phytoplasmas. However, currently, only a few families/tribes of effector orthologs have been identified in a range of phylogenetically distant phytoplasmas: SAP11/SWP1 (Chang *et al.*, 2018), SAP05, SAP54/PHYL1 and TENGU (Cho *et al.*, 2020; Wang *et al.*, 2017; Maejima *et al.*, 2014; Sugawara *et al.*, 2013). This lack of conservation slows the progress of functional characterization of phytoplasma effector candidates; most effector candidates must be studied individually and our knowledge about phytoplasma effectors is in its infancy, making functional characterization a challenge.

Using the UNIPROT database, a list of 738 amino acid sequences corresponding to nonredundant phytoplasma effectors was compiled (21 amino acid sequences corresponding to validated effectors and 717 amino acid sequences corresponding to effector candidates). This list was compiled using key words such as “*Candidatus* effectors”, “phytoplasma effectors”, “TENGU” and “SAP effectors”. Since programs that identify high levels of conservation have failed to find families in phytoplasma effectors beyond SAP05, SAP11, SAP54 and TENGU, we analyzed this dataset of 738 amino acid sequences through alternative methods, including Gene Ontology (GO), protein domains and motifs. Amino acid composition was also determined.

The search for de novo motifs enabled us to classify, for the first time, 696 phytoplasma effector candidates into 15 tribes. In other words, protein families that shared the same organization of multiple motifs were identified and grouped. The distribution of the tribes in the phytoplasmas revealed two evolutionary histories for their effectors. The tribes numbered 1, 6 and 7 were widespread in phytoplasmas, suggesting that their precursors occurred in common ancestors and were transmitted by vertical gene transfer, followed by duplication and rapid divergence in their evolution (Russell, 2019; Lipsitch *et al.*, 1996).

Other tribes (such as tribes 2, 3, 4, 5) showed discrete distribution, occurring only in phylogenetically related phytoplasmas; horizontal transmission to distant phytoplasmas was also observed in the members of these tribes.

Forty-two eukaryotic linear motifs, also known as short linear motifs (ELMs, SLiMs), were identified in the phytoplasma effectors. These SLiMs are novel tools that suggest the molecular mechanisms used by phytoplasma effectors during host infection. Phytoplasmas may hijack host targets by mimicry of key motifs contained in the regulators of critical host proteins (Kumar *et al.*, 2022).

The establishment of this new organization of phytoplasma effectors may help deepen our understanding of effectors in the race to combat phytoplasmas in agriculture and horticulture.

2.2. MATERIALS AND METHODS

2.2.1. Protein Dataset

The phytoplasma effector dataset comprises 21 amino acid sequences corresponding to validated phytoplasma effectors taken from the scientific literature, and the list was further complemented by adding the results from a search of the UNIPROT database using the keywords SAP01, SAP02, SAP03, until SAP80, TENGU, phyllody, phyll, anti- genic membrane protein (Amp), immunodominant membrane protein A (IdpA), immun- odominant membrane protein (Imp), and variable membrane protein A (VmpA). All protein sequences were collected, and duplications were eliminated. The final dataset contained amino acid sequences corresponding to 21 true effectors and 717 effector can- didates (total 738 amino acid sequences) and can be found in Table S1A and at https://github.com/Gisel-Carreon/Phytoplasma_effectors.

For a negative control set, 30 amino acid sequences corresponding to phytoplasma core proteins (proteins involved in essential metabolic activities) were randomly selected from validated genomes (draft genomes were excluded). The selected genomes for these core

proteins were the onion yellows phytoplasma OY-M (T00154) (16SrI-B), the aster yellows witches' broom phytoplasma AYWB (T00314) (16SrI-A), the 'Ca. Phytoplasma mali' (T00729) (16SrX), the 'Ca. Phytoplasma australiense' (T00752) (16SrXII), 'Ca. Phytoplasma ziziphi' Jwb-nky (T05675) (16SrV-B), 'Ca. Phytoplasma aurantifolia' (NCHU2014) (16SrII) and 'Ca. Phytoplasma asteris' MBSP (CP015149) (16SrI); these validated genomes were found in the PUBMED database (<https://pubmed.ncbi.nlm.nih.gov/>; accessed on 14 July 2023); protein annotations were obtained from KEGG (<https://www.genome.jp/kegg/genome/>; accessed on 14 July 2023) (Table S1B).

2.2.2. In Silico Characterization of Phytoplasma Effectors

The phytoplasma effectors in the dataset were in silico characterized by amino acid length and amino acid composition (specifically of amino acids Cys, Ser, Leu, Lys, Asn and Trp), using a set of Perl scripts (Carreón-Anguiano *et al.*, 2020). The presence of signal peptides and transmembrane domains were analyzed with SignalP v4.1 (Petersen *et al.*, 2011) and TMHMM v2.0 (Krogh *et al.*, 2001), respectively, both with default parameters, on mature proteins (without signal peptide).

For predicting nuclear localization signal (NLS), we used the NLStradamus program (<http://www.moseslab.csb.utoronto.ca/NLStradamus/>) (Nguyen *et al.*, 2009); the proteins were analyzed using the two-state static HMM method with a threshold score of 0.6.

2.2.3 Gene Ontology Distribution and Functional Annotation

Gene Ontology analysis for mapping biological process (BP), molecular function (MF) and cellular component (CC) were performed on the dataset of 738 amino acid sequences corresponding to phytoplasma effectors against the GO database integrated in the InterProScan v.86.0 tool under default parameters (Paysan-Lafosse *et al.*, 2023; Carbon *et al.*, 2021; Ashburner *et al.*, 2000).

The GO functional classification histogram was plotted using the web server WEGO 2.0 (<https://wego.genomics.cn/>, accessed on 20 August 2023) (Ye *et al.*, 2006). The output file produced was the GO-native format without reference statement.

Functional domain identification was carried out using the InterproScan program v.86.0 in STANDALONE mode, which includes diverse source databases like the PFAM, CDD and SMART modules. We submitted the proteins in FASTA file as input and carried out the analysis under default parameters (Paysan-Lafosse *et al.*, 2023; Blum *et al.*, 2021; Mistry *et al.*, 2021).

2.2.4 Classification of Effectors in Tribes Based on Motifs

The collection of 738 amino acid sequences corresponding to phytoplasma effectors was analyzed for motifs using the “Multiple Expectation-maximization for Motif Elicitation” (MEME) tool version 5.5.4, included in the “MEME suite” online platform (<https://meme-suite.org/meme/>) (Bailey *et al.*, 2015). We used de novo motif discovery in classic mode, with 0 or 1 occurrence per sequence, and maximum 50 motifs as cut off.

The effector sequences that share various motifs were grouped together as a multigene family (tribe). To identify similarities in the patterns of amino acids, online multiple sequences alignments were conducted using the CLUSTALW (GenomeNet) software program (<https://www.genome.jp/tools-bin/clustalw>). Then, to create weighted sequence alignments, the WebLogo online server was utilized (<https://weblogo.berkeley.edu/logo.cgi>), using the CLUSTALW alignment file as input.

To determine whether these motifs were specific to phytoplasma effectors, phytoplasma non-effector core proteins were also screened for these motifs.

2.2.5 Search for Short Linear Motifs in True Phytoplasma Effectors

Short linear motifs (SLiMs, also often referred to as ELMs), were identified by submitting a pair of protein members of different classes of phytoplasma effectors (for example, two

proteins for SAP11, two for Zaofeng effectors, two for ncSecPs, etc). The sequences, in a FASTA file, were submitted to the ELM database (<http://elm.eu.org>, accessed on 21 August 2023) (Kumar *et al.*, 2020). ELM is a database of 3934 manually curated eukaryotic motifs. The analysis was run with default settings and a probability cutoff of 1×10^{-5} .

ELMs found in the amino acid sequences of phytoplasma effectors were searched in the amino acid sequences of phytoplasma core proteins to determine whether they were specific to phytoplasma effector proteins.

2.3. RESULTS

2.3.1 Protein Databases

Table 2.1 shows the list of UNIPROT database ID accessions for the amino acid sequences corresponding to the phytoplasma effectors used in this work.

Table 2.1 List of phytoplasma effectors used in this work.

Classes of Phytoplasma Effectors Description in UNIPROT	Hits	UNIPROT ID *
SAP1-like protein	1	A0A859I9H9
SAP02-like protein	1	A0A859I9K5
SAP05-like protein	5	A0A975ILX0, A0A859I930, A0A0P7KD01, A0A531Y329, A0A0L0MJZ3
SAP06-like protein	2	A0A531Y0U6, A0A975FJD4
SAP08-like protein	2	A0A0P7IRG7, A0A859I9N5
SAP09-like protein	4	A0A531Y0B2, A0A859I9U4, A0A975FJ78, A0A975INP5
SAP11 effector protein	37	A0A081D545, A0A0L0MK48, A0A0N8IAU9, A0A0U2D622, A0A1Q1NH89, A0A1Q1NH90, A0A1Q1NH91, A0A1Q1NH92, A0A1Q1NH98, A0A1Q1NH99, A0A1Q1NHA1, A0A1Q1NHA2, A0A1Q1NHA3, A0A1Q1NHA5, A0A1Q1NHA6, A0A1Q1NHA7, A0A1V0PKN0, A0A1V0PKQ3, A0A1V0PKT7, A0A2P9JTQ0, A0A410HXL4, A0A421NZ11, A0A4Y5R0G3, A0A531Y0A6, A0A531Y154, A0A6M3WSS2, A0A7H9SKF2, A0A7H9SLX7, A0A7H9SP62, A0A7M3UQ86, A0A7M3UQ89, A0A7M3UQ91, A0A7M3UQ93, A0A859I8L9, A0A895HT12, A0A975FJX5, A0A975IND1
SAP19-like protein	4	A0A0P7JJB4, A0A531XZZ7, A0A859I959, A0A975FIG9
SAP20-like protein	1	A0A859IAA9
SAP21-like protein	5	A0A0P7IR19, A0A1V0PKQ1, A0A1V0PKR3, A0A7G3ZPG3, A0A859I965
SAP30-like protein	2	A0A0N8IAS3, A0A531Y0Y5
SAP34-like protein	2	A0A531XZH1, A0A859I9Q8
SAP36-like protein	3	A0A531XZX2, A0A859I9S9, A0A859IA72

SAP37-like protein	1	A0A859I9K5
SAP39-like protein	2	A0A975FJ78 (SAP09/SAP39-like), A0A975INP5 (SAP09/SAP39-like)
SAP40-like protein	2	A0A7G3ZP82, A0A975FIG9 (SAP19/SAP40-like)
SAP42-like protein	4	A0A0P7IR40, A0A531Y098, A0A859IAT1, A0A859IB38
SAP43-like protein	3	A0A531XZZ8, A0A531Y0C0, A0A859IAG2
SAP44-like protein	2	A0A531Y018, A0A859I9J4
SAP45-like protein	2	A0A0P7JJR0, A0A975FIA7
SAP48-like protein	2	A0A0N8IAV0, A0A975FJD4 (SAP06/SAP48-like)
SAP49-like protein	3	A0A0P7J0F0, A0A531Y0U5, A0A859IA30
SAP50-like protein	2	A0A531XYW6, A0A859I9B6
SAP53-like protein	7	A0A531Y0U7, A0A859I8L6, A0A859I8U8, A0A859I9C2, A0A859I9G1, A0A859IA89, A0A859IC30
SAP54-like protein	7	A0A0P7KHL3, A0A385GMC4, A0A4Y5N0H8, A0A531Y175, A0A6G5ZVL2, A0A859I8I1, A0A975FJ37
SAP55-like protein	2	A0A0P7KH40, A0A531Y2Y1
SAP56-like protein	2	A0A531Y0A1, A0A859IAC1
SAP59-like protein	1	A0A859I9X7
SAP61-like protein	2	A0A531Y053, A0A859I9W5
SAP63-like protein	2	A0A531Y0N6, A0A859IAU1
SAP64-like protein	1	A0A859IA25
SAP65-like protein	2	A0A531XZQ7, A0A859I9R5
SAP66-like protein	1	A0A531Y0A4
SAP67-like protein	6	A0A0P7IRG0, A0A1V0PKR5, A0A531Y097, A0A531Y0P6, A0A7G3ZP81, A0A859IA66
SAP68-like protein	2	A0A0P7JJ83, A0A531Y090
Phytoplasma effector causing phyllody symptoms	16	A0A0A8JBX1, A0A0A8JBX6, A0A0A8JBY1, A0A0A8JCF3, A0A0A8JCG0, A0A0A8JCG5, A0A0A8JCY1, A0A0A8JCY8, A0A0A8JCZ4, A0A0A8JD20, A0A0A8JD26, A0A0A8JD31, A0A0A8JD38, A0A0A8JER4, A0A0A8JER9, A0A0A8JES5
Secreted effector protein containing SVM protein	2	A0A0L0MJZ3 (SAP05-like), A0A0L0MK48 (SAP11-like)
Effector causing phyllody symptoms 1 (Phyl-1)	32	A0A0A8JD20, A0A0A8JBX1, A0A0A8JCF3, A0A7G1GAE9, A0A7G1GBC6, A0A7G1G848, A0A7G1G7V6, A0A7G1GAX5, A0A1S9M4S1, A0A7G1GBC7, A0A7G1G7U3, A0A7G1GCL4, A0A7G1G98, A0A0A8JCG5, A0A0A8JES5, A0A7G1GCA2, A0A851HA21, A0A0A8JCY1, A0A0A8JCZ4, A0A7G1GAE5, A0A7G1G7X2, A0A0A8JD26, A0A7G1GCL5, A0A0A8JBX6, A0A7G1G7V7, A0A0A8JER9, A0A7G1GDX1, A0A0A8JCY8, A0A7G1G829, A0A859I980, A0A859IA68, A0A7G1G7U9
SVM family protein	19	A0A7S7FZA6, A0A7S7FZA7, A0A7S7FZT4, A0A7S7G0A6, A0A7S7G133, A0A7S7JMA4, A0A9K3STF5, A0A9K3STG2, A0A9K3STJ3, A0A9K3STL1, A0A9K3VKC1, A0A9K3VQA7, A0A9K3WR63, A0A9K3WR67, A0A9K3WR81, A0A9K3WRI7, A0A9K3WRJ4, A0A9K3WSE9, A0A9K3WT14
Effector protein/putative effector	4	A0A0L0MJZ3 (SAP05-like), A0A0L0MK48 (SAP11-like), A0A0P7KHL3 (SAP54-like), A0A421NZ11 (SAP11-like protein)
Putative phage integrase	20	A0A081D424, A0A081D433, A0A081D4B3, A0A081D563, A0A081D571, A0A081D5B7, A0A081D5P7, A0A081D5U5, A0A0L0MKQ6, A0A0M1MZX0, A0A0M1N0J3, A0A559KJX4, B1V8S0, B1V929, B1V9A9, B1V9R4, B1V9U7, B1VAI1, B1VAZ9, B1VB07
MPEP-jgl1_1	1	F4YY99
PME2-sporadic	1	A0A5J6ED47
PME2ST	1	A0A5J6EF17
Antigenic membrane protein (Amp)	126	Q7M1T6, Q7M1T6, A0A076EAG8, A0A076EAG8, A0A081D5A2, A0A081D5A2, A0A0A6ZJW5, A0A0A6ZJW5, A0A0B5A2B3, A0A0B5A2B3, A0A0G2RL99, A0A0G2RL99, A0A0G3IAH0, A0A0G3IAH0, A0A0G3IAN5, A0A0G3IAN5, A0A0G3IAN8, A0A0G3IAN8, A0A0G3IEB2,

Immunodominant membrane protein A (idpA)	2	A0A3G3BKC1, F8QQ90						
Immunodominant membrane protein (Imp)	241	A0A081D461, A0A0G2SJS4, A0A0G2SJW3, A0A0G2SK05, A0A0G2SK62, A0A0K3AQA8, A0A0K3ASS0, A0A0N8IAV7, A0A167RQ06, A0A1C3K9C2, A0A1C3K9C3, A0A1C3K9D4,						

		G3XGG6, G3XGG9, G3XGI1, G3XGI3, G3XGI7, G3XGJ0, G3XGJ3, G3XGJ6, G3XGJ9, G3XGK5, G3XGL1, G3XGL7, G3XGM0, I3UIB8, I3UIB9, I3UIC0, I3UIC1, I3UIC2, I3UIC3, I3UIC4, I3UIC5, I3UIC6, I3UIC7, I3UIC8, I3UIC9, I3UID0, Q8KWR9, Q9Z4Q4, U5LMX8, U6G028, X2CSY1
Variable membrane protein A (VmpA)	52	A0A0D6DTS2, A0A4E9CZE5, A0A4E9CZG2, A0A4E9CZI4, A0A4E9CZK2, A0A4E9CZL0, A0A4E9D018, A0A4E9D027, A0A4E9D036, A0A4E9D043, A0A4E9D050, A0A4E9D063, A0A4E9D398, A0A4E9D3B5, A0A4E9D3F2, A0A4E9D3J3, A0A4E9D3L5, A0A4E9D3N2, A0A4E9DGY5, A0A4E9DH23, A0A4E9DH31, A0A4E9DH40, A0A4E9DJ74, A0A4E9DJ82, A0A4E9DJ94, A0A4E9DJA2, A0A4E9DJB4, A0A4E9DKV2, A0A4E9DKY4, A0A4E9DKZ5, A0A4E9DL10, A0A4E9DPW9, A0A4E9DQ17, A0A4E9DQ26, A0A4E9DQ36, A0A4E9DQ45, A0A4E9E2M2, A0A4E9E2N4, A0A4E9E2N9, A0A4E9E2P3, A0A4E9E2P9, A0A4E9E2Q4, A0A4E9E2R0, A0A4E9E2R4, A0A4E9E4J3, A0A4E9E4J7, A0A4E9E4L5, A0A4E9E9I3, A0A4E9E9J2, A0A4E9E9J8, A0A4E9E9K0, A0A8B0MFL7
Secreted AYWB protein (SAP)	36	A0A0N8IAS3, A0A0N8IAU9, A0A0N8IAV0, A0A0P7IR19, A0A0P7IR40, A0A0P7IRG7, A0A0P7JJ83, A0A0P7JJB4, A0A0P7JJR0, A0A0P7KH40, A0A0P7KHL3, A0A531XYW6, A0A531XZH1, A0A531XZQ7, A0A531XZX2, A0A531XZZ7, A0A531XZZ8, A0A531Y018, A0A531Y053, A0A531Y090, A0A531Y097, A0A531Y098, A0A531Y0A1, A0A531Y0A4, A0A531Y0C0, A0A531Y0N6, A0A531Y0P6, A0A531Y0U5, A0A531Y0U7, A0A531Y0Y5, A0A531Y154, A0A531Y175, A0A531Y2Y1, A0A531Y329, A0A7G3ZP82, A0A7G3ZPG3
Type III secretion system effector protein Candidatus	50	A0A1E2USP0, A0A1J7D1C9, A0A249DXM8, A0A2D3T254, A0A2D3T9Z4, A0A2D3TD58, A0A2D3TFY7, A0A2U8I5A1, A0A4P2SM26, A0A5K6V4X6, A0A5K6V7I2, A0A6L2ZRP7, A0A9D2KKF3, A0A9E4K1U0, A0A9E4KC36, A0A9E4P750, C4K6C8, C4K8T3, E0WUJ2, G2GX33, G2GYW6, G2J8Z5, I6PE34, W0HMT0, A0A0C1H4D9, A0A0G1JG32, A0A1A9HUH7, A0A1A9HUI4, A0A1A9HUV6, A0A1A9HWR9, A0A1J7C9A4, A0A3A4PR12, A0A3C1SQQ9, A0A3S0UHE5, A0A6L2ZKD9, A0A6L2ZKI3, A0A6M1YMG1, A0A6M1YQX0, A0A6M1YRV2, A0A6M1Z4N3, A0A6M1Z7H7, A0A6M1ZIG1, A0A6M2AFE6, A0A846KW82, A0A924DW95, A0A956SX96, E0WTJ4, I6SXS7, Q6MD78, W6MBI3
Conserved hypothetical protein *	1	Q3LBN8
Uncharacterized protein *	1	U4KNV3
Candidatus hypothetical proteins	6	A0A660HMR4, A0A660HM07, A0A660HM34, A0A660HNE9, A0A660HMI8, A0A660HNK1

* The amino acid sequences are available in Table S1A.

2.3.2 Characterization of Phytoplasma Effectors

To learn more about structural features of the known phytoplasma effectors, the dataset was analyzed for peptide length, amino acids composition, presence of signal peptide, presence of transmembrane domains (Table 2.2), among other characteristics. More than 90% were small peptides, with 400 amino acids or less. Consistent with the current belief, ~70% of the phytoplasma effectors are secreted and have a SP, but ~30% have TMDs, revealing that the non-classical effectors constitute an important subset in phytoplasma effectoromes.

Table 2.2 Characterization of phytoplasma effectors.

Characteristics	Number of Effectors	% of the Total *
≤200 amino acids	518	70.2
201–400 amino acids	156	21.1
401–1000 amino acids	63	8.5

>1000 amino acids	1	0.1
Signal peptide **	527	71.4
TMDs	249	33.8
Nuclear target (NLS)	29	3.9
≥25 Asn residues	109	14.8
≥25 Lys residues	341	46.2
≥25 Leu residues	155	21
≥25 Ser residues	61	8.3
0 Trp residues	225	30.5
1 Trp residue	234	31.7
2 Trp residues	158	21.4
3–7 Trp residues	121	16.4
0 Cys residues	378	51.2
1 Cys residue	214	29
2 Cys residues	86	11.7
3 Cys residues	30	4.05
4 Cys residues	10	1.35
5–8 Cys residues	20	2.7

* Considering a total of 738 amino acid sequences corresponding to phytoplasma effectors.

** Using SignalP v4.1.

In terms of amino acid composition, almost 50% of the phytoplasma effectors are rich in lysine, ~20% are rich in leucine and ~15% in asparagine (>25 residues per sequence), while tryptophan and cysteine are rare amino acids; ~60% of the effectors have zero or one tryptophan and ~80% have zero or one cysteine residue.

2.3.3 Functional Categories of the Phytoplasma Effectors

For the 738 phytoplasma effectors, the Gene Ontology analysis identified annotations for 462 phytoplasma effectors, while 276 had no Gene Ontology annotations. Fourteen GO terms were identified: seven in the category “biological process” (phosphorelay signal transduction system, GO:0000160; regulation of DNA-templated transcription, GO:0006355; proteolysis, GO:0006508; protein targeting, GO:0006605; vesicle-mediated transport, GO:0016192; protein import, GO:0017038; protein refolding, GO:0042026); two in the category “cellular component” (plasma membrane, GO:0005886; membrane, GO:0016020);

and five in the category “molecular function” (DNA binding, GO:0003677; ATP-dependent peptidase activity, GO:0004176; metalloendopeptidase activity, GO:0004222; ATP binding, GO:0005524; ATP-dependent protein-folding chaperone, GO:0140662). Figure 2.1 shows the distribution of GOs.

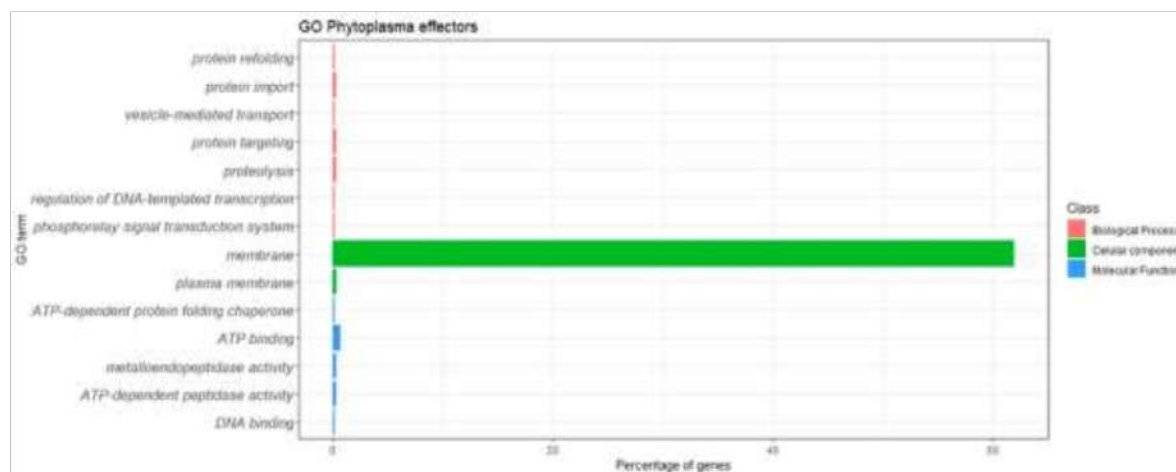


Figure 2.1 Gene Ontology classification of the phytoplasma effectors.

The most represented GO term was GO:0016020, with 457 proteins assigned to the term “membrane”; as expected, this term was assigned to 298 known membrane-bound effectors (IcpA, Imp, VmpA and Amp from different phytoplasmas), but also to a few homologs to PHYL1, SAP01, SAP02/saP37/SAP76, SAP08, SAP09/SAP39, SAP34, SAP36, SAP40, SAP42, SAP48, SAP50, SAP53, SAP65, SAP66, SAP68, PME2ST, PME2 and 1 outer-surface lipoprotein. One hundred and twenty-two membrane effectors had no homology to known effectors; according to the UNIPROT database, six were assigned “putative effector”, four were assigned “effector protein” and one hundred and twelve were assigned “effector”. It was observed in this study that membrane-bound effectors are widely distributed in phytoplasmas, not to specific phytoplasma 16Sr groups.

Most effectors received multiple GO assignments; for example, the effectors A0A531Y2Y1 (SAP55-like protein of periwinkle leaf yellowing phytoplasma) and A0A0P7KH40 (SAP55-like of “*Echinacea purpurea*” witches’ broom phytoplasma) were classified with the GO terms GO:0005524 ATP-binding; GO:0004222 metalloendopeptidase activity; GO:0004176 ATP-dependent peptidase activity; and GO:0006508 proteolysis). Likewise, the effec- tor

A0A0G2SK05 (SecA translocation protein of Napier grass stunt phytoplasma) and A0A0G2SK62 (SecA translocation protein of Hyparrhenia grass white leaf phytoplasma) were assigned the terms GO:0005886 plasma membrane; GO:0017038 protein import; and GO:0006605 protein targeting. The effector B1Q3E7 (molecular chaperonin GroEL of “*Ca. Phytoplasma japonicum*”) was assigned GO:0140662 ATP-dependent protein-folding chaperone; GO:0005524 ATP binding; and GO:0042026 protein refolding.

The phytoplasma effectors retrieved a low number of GO terms, although some effectors had three to five GO terms assigned to them; the functions of the majority of these effectors could not be predicted through GO analysis.

2.3.4 Functional Domains in Phytoplasma Effectors

The use of effector-related domains is an emerging strategy for effector identification in other phytopathogens (Qian *et al.*, 2022; Wood *et al.*, 2020; Zhu *et al.*, 2017). The phytoplasma effectors were analyzed with the program InterProScan version 5.39–77.0, which identified functional domains in 153 effectors, while 585 had no known domains (Figure 2.2).

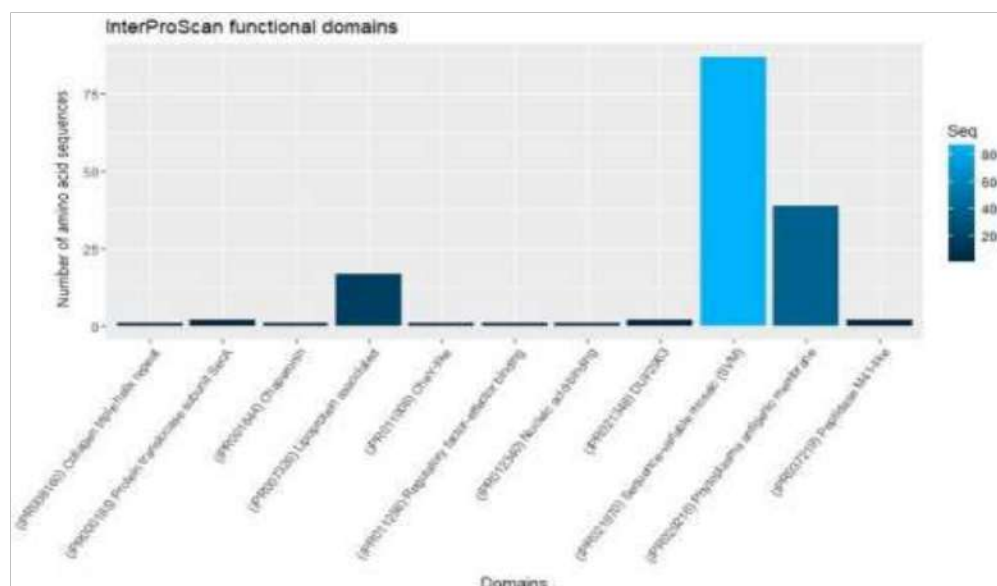
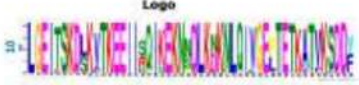


Figure 2.2 InterProScan functional domains found in phytoplasma effectors.

The most frequent domain was the “SVM domain”, widely distributed in 87 phytoplasma effectors such as phyllody 1, Sap01, SAP04, SAP05, SAP11, SAP19, SAP21, SAP30, SAP40, SAP42, SAP43, SAP44, SAP45, SAP48, SAP49, SAP54, SAP56, SAP66, SAP67 and SAP68. The second most frequently found domain was “Phytoplasma antigenic membrane”, present in 39 phytoplasma antigenic membrane proteins (AMP). The third was “Lipoprotein associated”, present in 17 phytoplasma variable membrane proteins A (VmpA). Molecular chaperonin (GroEL) domain is present in the effector B1Q3E7 of ‘*Ca. Phytoplasma japonicum*’. The domain “response regulator transcription factor” was found in the effector A0A847N6X9 of ‘*Ca. Phytoplasma sp*’ and the domain “collagen triple helix repeats” in the effector A0A421NUF2 of ‘*Ca. Phytoplasma solani*’. The domain DUF2963 was found in two effectors of Rapeseed phyllody phytoplasma, A0A859I9K5 (SAP02/SAP37/SAP76-like protein) and A0A859IA25 (SAP64-like protein). The domain “Nucleic acid-binding” in A0A0P7KDF6 of “*Echinacea purpurea*” witches’ broom phytoplasma and the domain peptidase M41-like was identified in A0A0P7KH40 and A0A531Y2Y1, two SAP55-like effectors of “*Echinacea purpurea*” witches’ broom phytoplasma and periwinkle leaf yellowing phytoplasma, respectively.

2.3.5 Identification of Effector Tribes in Phytoplasmas: Classification by Protein Motifs

In an effort to classify the phytoplasma effectors, set 1 (738 effectors) was analyzed using the MEME program to find de novo motif sequences. The search was conducted either on the full set #1 or in the set #1 but lacking the largest classes of effectors such as SAP11 (members) and VmpA (members). Both strategies rendered similar results. Figure 2.3 shows the 50 top amino acid motifs found in the phytoplasma effector dataset; 696 amino acid sequences of effectors have at least 1 motif while 42 have no motif at all.

Logo	Site	E-value	Logo	Site	E-value
	262	3.6×10^{-107}		48	3.26×10^{-109}
	171	1.6×10^{-108}		42	8.36×10^{-104}
	70	2.36×10^{-103}		40	2.56×10^{-108}
	50	9.16×10^{-103}		46	3.4×10^{-80}
	38	6.86×10^{-103}		72	2.6×10^{-94}
	49	1.76×10^{-103}		101	9.2×10^{-71}
	68	2.16×10^{-104}		45	9.7×10^{-70}
	48	4.86×10^{-103}		26	9.2×10^{-70}
	36	4.66×10^{-100}		24	1.7×10^{-111}

Logo	Site	E-value	Logo	Site	E-value
	26	7.4×10^{-40}		13	1.4×10^{-103}
	50	5.7×10^{-100}		21	5.4×10^{-100}
	115	1.4×10^{-275}		24	5.4×10^{-106}
	37	6.0×10^{-107}		46	7.8×10^{-107}
	73	1.7×10^{-114}		14	3.9×10^{-121}
	36	9.8×10^{-93}		10	3.8×10^{-108}
	45	3.6×10^{-121}		75	3.2×10^{-108}
	50	2.1×10^{-100}		26	5.4×10^{-107}
	12	1.1×10^{-103}		14	1.4×10^{-103}

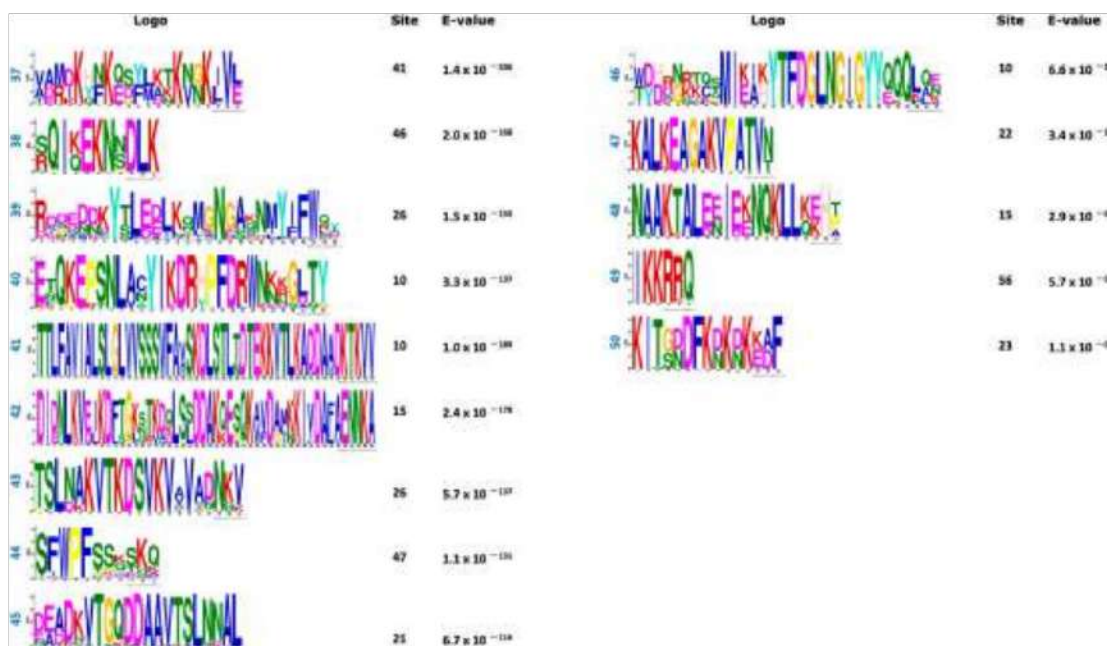


Figure 2.3 Top 50 motifs found in the positive dataset (738 amino acid sequences) of phytoplasma effectors. The sites correspond to the number of effectors that contain that motif. E-value per each motif was included.

Based on the organization of these motifs, 15 effector tribes (families) were distinguished; each tribe consists of members that share various motifs in their sequences. Figure 2.4 shows the schematic representations of the organization of the motifs in the tribes. The WebLogo sequences of these 15 tribes are provided as Figures S1–S15. In some sequences, the gaps between the motifs were narrow and their WebLogo sequences appear as a single continuum sequence instead of various motifs (tribes 3, 5, 11 and 12), but all WebLogos have at least three motifs.

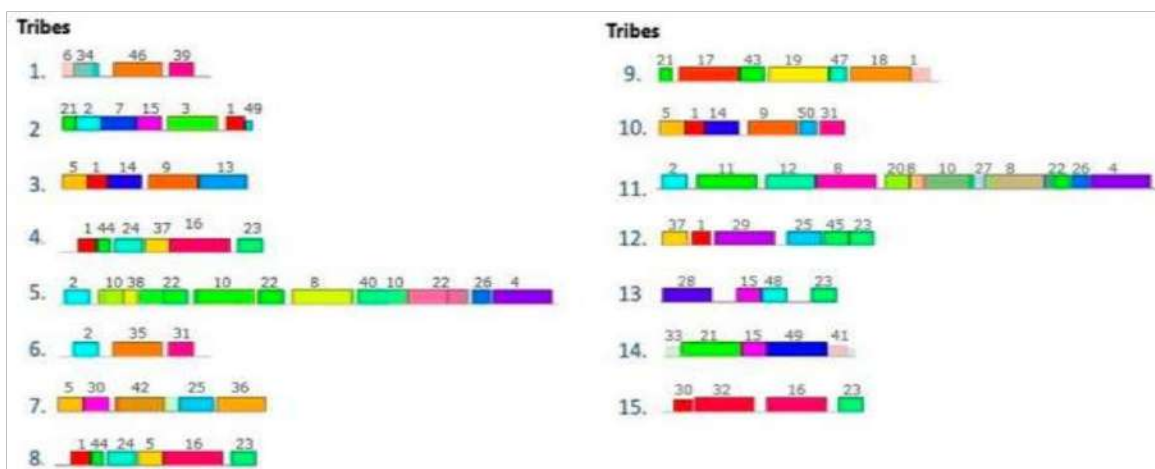


Figure 2.4. Schematic representation of the organization of the motifs in the different effector tribes (effector families) in phytoplasmas. These configurations were found in the positive control dataset (738 amino acid sequences of phytoplasma effectors). Rectangles with the same color correspond to the same motif; the lines correspond to the gaps between the motifs in the amino acid sequences. The different sizes of the rectangles represent the different amino acid lengths of the motifs (different number of amino acids). The numbers above the rectangles correspond to the motifs enlisted in Figure 2.3.

The 15 phytoplasma protein tribes classified 696 effectors (Table 2.3) while 42 effectors were orphan sequences that do not belong to any tribe. Tribe 1 comprises the largest group of effectors (223 members from the set of 738 effectors). This effector tribe was found widely distributed in phytoplasmas with multiple effector members per genome (>6 members), and in some genomes with more than 18 members (Table 2.3). The effector tribes 6 and 7 also have a wide distribution among the phytoplasmas but have a smaller number of members per genome; for example, from family 7, only two or three members per genome were found.

Other tribes were restricted to specific phytoplasmas and very few other phytoplasmas. For example, tribe 2 of '*Ca. Phytoplasma solani*', tribe 3 of elm yellows phytoplasma, tribe 4 of apple proliferation phytoplasma and tribe 5 of alder yellows phytoplasma and Flavescence dorée phytoplasma.

To characterize the effector members of each tribe, their amino acid composition was analyzed. The amino acid asparagine was rich in effectors belonging to all tribes, except the effector members of tribes 2 and 14. The amino acid lysine was abundant in effector members of all tribes, especially tribes 5 and 11, with ~70 Lys per sequence. About 10 serine amino acids per sequence were observed, except in tribes 5 and 11, in which effector members contained >30 Ser per sequence, and in tribes 3 and 10, in which most effector members contained 4–5 Ser per sequence. Tryptophan and cysteine were poor in all effector members in all tribes; some patterns were observed with respect to these amino acids. Tribe 2 members have one Trp and zero Cys; in tribe 3, effector members have one Trp and one Cys, while in tribes 5 and 6, most effector members have neither Trp nor Cys (Table S2).

Table 2.3 Phytoplasma effector tribes and pattern distribution of effector members among phytoplasmas.

Tribe	WebLogo sequence	Number of members	Distribution	Phytoplasmas
1	Figure S1	223	Wide; usually 6-8 members per genome. Some phytoplasmas (' <i>Ca. Phytoplasma phoenicium</i> ', ' <i>Ca. Phytoplasma solani</i> ', ' <i>Echinacea purpurea</i> ' witches'-broom phytoplasma, Loofah witches'-broom phytoplasma) have 18-24 members.	Apple proliferation phytoplasma, Aster yellows witches'-broom phytoplasma (strain AYWB), ' <i>Ca. Phytoplasma aurantifolia</i> ', ' <i>Ca. Phytoplasma fragariae</i> ', ' <i>Ca. Phytoplasma japonicum</i> ', ' <i>Ca. Phytoplasma oryzae</i> ', ' <i>Ca. Phytoplasma phoenicium</i> ', ' <i>Ca. Phytoplasma pini</i> ', ' <i>Ca. Phytoplasma pruni</i> ', ' <i>Ca. Phytoplasma sacchari</i> ', ' <i>Ca. Phytoplasma solani</i> ', ' <i>Ca. Phytoplasma sp.</i> ', ' <i>Ca. Phytoplasma tritici</i> ', ' <i>Ca. Phytoplasma vitis</i> ', ' <i>Ca. Phytoplasma ziziphi</i> ', ' <i>Catharanthus roseus</i> ' aster yellows phytoplasma, ' <i>Chrysanthemum coronarium</i> ' phytoplasma, Clover proliferation phytoplasma, Crotalaria phyllody phytoplasma, ' <i>Cynodon dactylon</i> ' phytoplasma, ' <i>Echinacea purpurea</i> ' witches'-broom phytoplasma, ' <i>Echinacea purpurea</i> ' witches'-broom phytoplasma, Lime witches'-broom phytoplasma, Loofah witches'-broom phytoplasma, Mulberry dwarf phytoplasma, Onion yellows phytoplasma (strain OY-M), ' <i>Parthenium sp.</i> ' phyllody phytoplasma, Paulownia witches'-broom phytoplasma, Periwinkle leaf yellowing phytoplasma, Phytoplasma australiense, Phytoplasma mali (strain AT), Poinsettia branch-inducing phytoplasma, Porcelain vine witches'-broom phytoplasma, Rapeseed phyllody phytoplasma, Rice orange leaf phytoplasma, <i>Ziziphus jujuba</i> witches'-broom phytoplasma
2	Figure S2	74	Restricted; 68 members are from ' <i>Ca. Phytoplasma solani</i> '	' <i>Ca. Phytoplasma phoenicium</i> ', ' <i>Ca. Phytoplasma solani</i> ', ' <i>Ca. Phytoplasma vitis</i> ', ' <i>Ca. Phytoplasma ziziphi</i> ', ' <i>Dodonaea viscosa</i> ' witches'-broom phytoplasma, ' <i>Echinacea purpurea</i> ' witches'-broom phytoplasma, Loofah witches'-broom phytoplasma

3	Figure S3	49	Restricted; 47 members are from Elm yellows phytoplasma	Elm yellows phytoplasma, Alder yellows phytoplasma, Phytoplasma vitis (Flavescence dorée phytoplasma)
4	Figure S4	48	Restricted; 17 members from Apple proliferation phytoplasma. One to five members in the other genomes	Apple proliferation phytoplasma, 'Ca. Phytoplasma aurantifolia', 'Ca. Phytoplasma oryzae', 'Ca. Phytoplasma phoenicium', 'Ca. Phytoplasma pruni', 'Ca. Phytoplasma solani', Crotalaria phyllody phytoplasma, 'Echinacea purpurea' witches'-broom phytoplasma, Faba bean phyllody phytoplasma, Loofah witches'-broom phytoplasma, Loofah witches'-broom phytoplasma, Onion yellows phytoplasma (strain OY-M), 'Parthenium sp.' phyllody phytoplasma, Peanut witches'-broom phytoplasma, Periwinkle leaf yellowing phytoplasma, Phytoplasma mali (strain AT), Rapeseed phyllody phytoplasma, Rice orange leaf phytoplasma, Ziziphus jujuba witches'-broom phytoplasma
5	Figure S5	39	Restricted; 17 members from Alder yellows phytoplasma, and 15 members from Flavescence dorée phytoplasma. One or two members in the other genomes	Alder yellows phytoplasma, Phytoplasma vitis (Flavescence dorée phytoplasma), 'Ca. Phytoplasma pruni', 'Ca. Phytoplasma solani', 'Ca. Phytoplasma sp.', Clover phyllody phytoplasma, 'Echinacea purpurea' witches'-broom phytoplasma, 'Parthenium sp.' phyllody phytoplasma, Peanut witches'-broom phytoplasma, Periwinkle leaf yellowing phytoplasma, Phytoplasma mali (strain AT), Rapeseed phyllody phytoplasma, Rice orange leaf phytoplasma, Ziziphus jujuba witches'-broom phytoplasma
6	Figure S6	37	Wide; few members (1-6) per genome	Apricot aster yellows phytoplasma A-AY, Aster yellows phytoplasma AY2192, 'Ca. Phytoplasma aurantifolia', 'Ca. Phytoplasma phoenicium', 'Ca. Phytoplasma pruni', 'Ca. Phytoplasma sp.', Carrot yellows phytoplasma CA-76, 'Citrus aurantiifolia' phytoplasma, Clover proliferation phytoplasma, Crotalaria phyllody phytoplasma, 'Echinacea purpurea' witches'-broom phytoplasma, Eggplant dwarf phytoplasma ED, Faba bean phyllody phytoplasma, Gladiolus witches'-broom phytoplasma, Leontodon yellows phytoplasma LEO, Lime witches'-broom phytoplasma, Loofah witches'-broom phytoplasma, Peach yellows phytoplasma PYR, Rapeseed phyllody phytoplasma, 'Solanum lycopersicum' phytoplasma, Ziziphus jujuba witches'-broom phytoplasma, 'Ca. Phytoplasma vitis' (Flavescence dorée phytoplasma), Peanut witches'-broom phytoplasma, Periwinkle leaf yellowing phytoplasma, Phytoplasma mali (strain AT), Rapeseed phyllody phytoplasma, Rice orange leaf phytoplasma, Ziziphus jujuba witches'-broom phytoplasma
7	Figure S7	32	Wide; two or three members per genome	Alfalfa witches'-broom phytoplasma, 'Ca. Phytoplasma aurantifolia', 'Ca. Phytoplasma pruni', 'Ca. Phytoplasma sp.', Carrot witches'-broom phytoplasma, Chickpea phyllody phytoplasma, Crotalaria phyllody phytoplasma, 'Cucurbita pepo' phytoplasma, 'Echinacea purpurea' witches'-broom phytoplasma, Eggplant big bud phytoplasma, Faba bean phyllody phytoplasma, Hyparrhenia grass white leaf phytoplasma, Lactuca serriola phytoplasma, Lime witches'-broom phytoplasma, Napier grass stunt phytoplasma, Parsely witches'-broom phytoplasma, 'Parthenium sp.' phyllody

				phytoplasma, Pear decline phytoplasma (Taiwan II), Periwinkle phyllody phytoplasma, Sesame phyllody phytoplasma
8	Figure S8	31	Restricted; 10 members from Apple proliferation phytoplasma, 5 members from 'Ca. Phytoplasma pyri', 3 members from 'Ca. Phytoplasma solani' and one or two members in the other genomes	Apple proliferation phytoplasma, 'Ca. Phytoplasma pruni', 'Ca. Phytoplasma prunorum', 'Ca. Phytoplasma pyri', 'Ca. Phytoplasma solani', 'Ca. Phytoplasma sp.', 'Echinacea purpurea' witches'-broom phytoplasma, Loofah witches'-broom phytoplasma, 'Parthenium sp.' phyllody phytoplasma, Rapeseed phyllody phytoplasma, Rice orange leaf phytoplasma, Tsuabuki witches'-broom phytoplasma, 'Sesamum indicum' phyllody phytoplasma
9	Figure S9	30	Wide; one member per genome	'Brassica napus' phytoplasma, 'Ca. Phytoplasma solani', 'Catharanthus roseus' aster yellows phytoplasma, 'Chrysanthemum coronarium' phytoplasma, Chrysanthemum yellows phytoplasma, 'Echinacea purpurea' witches'-broom phytoplasma, Eggplant dwarf phytoplasma, Hydrangea phyllody phytoplasma, Iceland poppy yellows phytoplasma, 'Lactuca sativa' aster yellows phytoplasma, Lettuce yellows phytoplasma, Maize bushy stunt phytoplasma, Marguerite yellows phytoplasma, Mulberry dwarf phytoplasma, Mulberry yellow dwarf phytoplasma, Onion yellows phytoplasma, Paulownia witches'-broom phytoplasma, Periwinkle leaf yellowing phytoplasma, Phytoplasma sp. AYBG, Porcelain vine witches'-broom phytoplasma, Potato purple top phytoplasma, 'Primula acaulis' yellows phytoplasma, Rapeseed phyllody phytoplasma, Rice orange leaf phytoplasma, Strawberry lethal yellows phytoplasma (CPA) str. NZSb11, Sumac witches'-broom phytoplasma, Tomato yellows phytoplasma
10	Figure S10	30	Restricted; 24 members from Elm yellows phytoplasma; one member in the other genomes	Elm yellows phytoplasma, Clover proliferation phytoplasma, Korean potato witches'-broom phytoplasma, Lime witches'-broom phytoplasma, Loofah witches'-broom phytoplasma, 'Parthenium sp.' phyllody phytoplasma, Phytoplasma vitis (Flavescence dorée phytoplasma),
11	Figure S11	26	Restricted; 18 members from Alder yellows phytoplasma; 1-3 members in the other genomes	Alder yellows phytoplasma, 'Ca. Phytoplasma pruni', 'Ca. Phytoplasma solani', 'Ca. Phytoplasma sp.', Periwinkle leaf yellowing phytoplasma, Rapeseed phyllody phytoplasma
12	Figure S12	20	Restricted; 18 members from Poinsettia branch-inducing phytoplasma; 1-3 members in the other genomes	Poinsettia branch-inducing phytoplasma, 'Ca. Phytoplasma pruni', 'Ca. Phytoplasma solani', 'Ca. Phytoplasma sp.', Periwinkle leaf yellowing phytoplasma, Rapeseed phyllody phytoplasma,
13	Figure S13	20	Restricted; 19 members from 'Ca. Phytoplasma pyri'	'Ca. Phytoplasma pyri', 'Ca. Phytoplasma pruni'
14	Figure S14	19	Restricted; 12 members from 'Ca. Phytoplasma meliae'; 3 members from 'Ca. Phytoplasma solani'; one member in the other genomes	'Ca. Phytoplasma japonicum', 'Ca. Phytoplasma meliae', 'Ca. Phytoplasma solani', 'Echinacea purpurea' witches'-broom phytoplasma, Hyparrhenia grass white leaf phytoplasma, Napier grass stunt phytoplasma,
15	Figure S15	18	Restricted; 14 members from 'Ca. Phytoplasma prunorum'; 1-2 members in the other genomes	'Ca. Phytoplasma oryzae', 'Ca. Phytoplasma prunorum', 'Echinacea purpurea' witches'-broom phytoplasma, Ziziphus jujuba witches'-broom phytoplasma

2.3.6 Short Linear Motifs in Phytoplasma Effectors

Short linear motifs (SLiMs), also known as ELMs (from eukaryotic linear motifs), are short linear peptides that have a specific sequence pattern (3–10 amino acids), which is recognized by interacting domains (Van *et al.*, 2014). Usually, SLiMs are involved in transient key interactions with proteins, DNA or RNA, regulating cell processes such as cell signaling, cell cycle, protein degradation, etc. (Kumar *et al.*, 2022). Many eukaryotic, bacterial and viral pathogens mimic SLiMs present in host cell proteins to hijack cellular processes as part of the infection cycle (Wadie *et al.*, 2022; Sámano-Sánchez and Gibson, 2020; Davey *et al.*, 2011). Therefore, we decided to explore SLiMs by analyzing a subset of 87 phytoplasma effectors in the ELM server (2 per each class in Table 2.1; each class contains at least 1 ID). Forty-two SLiMs were identified (Figure 2.5). In the top 10 most frequently identified motifs was the SLiM `LIG_PDZ_Class_2`, which has the pattern `(VYF)X(VIL)` and is present in diverse SAP effectors; this motif binds to a surface groove of PDZ domains of the target proteins. The PDZ is a ubiquitous motif of 80–90 amino acids found in the signaling proteins of bacteria, yeast, plants, viruses and animals (Ponting, 1997). Proteins containing PDZ motif anchor receptor proteins in the membrane to cytoskeleton, as well as help organize and hold together signaling complexes at the plasma membrane membrane (Christensen *et al.*, 2019; Fanning and Anderson, 1999). `LIG_PDZ_Class_2`-containing effectors may function close to the plasma membrane of the host to fulfill their functions.

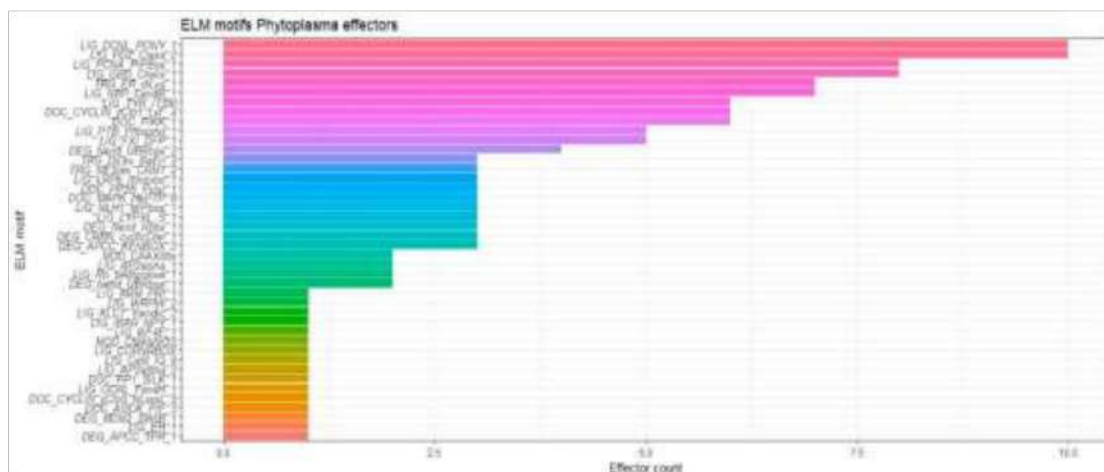


Figure 2.5 Short linear motifs found in phytoplasma effectors. A set of 87 effectors was analyzed, selecting 2 effectors per each class in Table 2.1.

The SLiM `LIG_DCNL_PONY_1` was as frequent as the `LIG_PDZ_Class_2` motif, and it was also present in diverse SAP effectors. The `LIG_DCNL_PONY_1` motif binds the ubiquitin-conjugating enzyme E2 M (UBE2M) and the ubiquitin-conjugating enzyme E2 F (UBE2F), which play diverse biological roles (Kumar *et al.*, 2022). Phytoplasma effectors containing this motif may regulate their targets through proteolysis. Proteolytic activity has been described in the effectors SAP54 (MacLean *et al.*, 2014) and SAP05 (Huang *et al.*, 2021), and may be the mechanism of action of other phytoplasma effectors as well. The `LIG_PCNA_PIPBox_1` motif was identified in eight effectors and is a nuclear localization domain (Kumar *et al.*, 2022). Several phytoplasma effectors, such as SAP11 and SWP1, target nuclear proteins in the host, including transcription factors, and the removal of the nuclear localization signal leads to loss of function (Bai *et al.*, 2023; Strohmayer *et al.*, 2021). The presence of the `LIG_PCNA_PIPBox_1` domain may uncover other effectors that also function in the plant nuclei. In '*Ca. Phytoplasma asteris*', ~7% of the effectorome is predicted to target the plant cell nuclei (Bai *et al.*, 2009).

The SLiM `LIG_GBD_Chelix_1` was found in eight effectors; this motif allows for the recruitment of the actin-regulatory proteins that initiates actin polymerization (Cheng *et al.*, 2008; Mészáros *et al.*, 2021). Polymerization of actin is a common molecular mechanism found in infections by pathogens in different kingdoms (Wang *et al.*, 2022; Mészáros *et al.*, 2021; Choe and Welch, 2016; Upadhyay and Shaw, 2008; Guiney and Lesnick, 2005).

The SLiM `LIG_NRP_CendR_1` was found in seven effectors that include several SAP- type effectors and the membrane antigenic *I*dpA. In humans, this domain binds to the neuropilin b1 domain binding site (Kumar *et al.*, 2022). In addition to the very well-known interaction between the spike receptor protein of SARS-CoV-2 and the human angiotensin-converting enzyme 2 (ACE2), it was recently found that the spike receptor also targets neuropilin b1 (Alnomasy, 2021), evidencing that the neuropilin b1 domain binding site is targetable by pathogens. This motif may be involved in infection by phytoplasmas, but the host target is largely unknown. The next motif was `TRG_ER_diLys_1`, and it was identified in seven effectors of the SAP-type and the membrane antigenic *A*mp and *V*mpA. This motif is an endoplasmic reticulum retrieving signal; it has been found in proteins from humans, rats and yeast. The proteins interact with the WD40 domain and G-beta repeat domain; the former is present in many transcription factors and E3ubiquitin ligase, and the latter in G proteins

(Kumar *et al.*, 2022). Although TRG_ER_dilys_1 has not been described in phytopathogens, it likely participates in phytoplasma pathogenicity.

The SLiM DOC_PIKK_1 was found in six effectors belonging to SAP05, SAP49, SAP54 and the membrane antigenic ldpA. This motif is a docking site for multiple phosphatidylinositol-3 kinase-related kinases (PIKKs) involved in cell cycle DNA damage checkpoints and oxidative stress, and response to DNA damage (Kumar *et al.*, 2022). The DOC_PIKK_1 motif has been identified in the effector proteins CagA and Tir of the bacteria *Legionella pneumophila* and *L. pasculli* (Sampietro *et al.*, 2018), but the precise role of this domain has not been established. The effectors with this domain may prevent DNA repair in the host caused by oxidative stress, or these effectors may inactivate PIKKs and suppress host signaling. The SLiM DOC_CYCLIN_yCib1_LxF_4, present in SAP09, SAP20, SAP49 and SAP61, is a docking site present in substrates and inhibitors of the M-phase cyclins Cib1/2 (Örd *et al.*, 2019). These effectors possibly interfere in the cell cycle of the host cell. The SLiM LIG_TYR_ITSM occurred in SAP43, SAP49 and SAP50. This motif binds to and is regulated by SH2 adaptor molecules, and is critical for the activation and termination of signal transduction pathways (Kumar *et al.*, 2022). The motif LIG_TYR_ITSM was identified in effector proteins of *Corynebacterium diphtheria* (Weerasekera *et al.*, 2018), but its role is unknown. The SLiMs LIG_FXI_DFP_1 and LIG_PTB_Phospho_1 are in the 10th position, each one identified in five effectors. LIG_FXI_DFP_1 was found in SAP39, SAP43 and PME2; this motif is a disulfide-linked dimer, each subunit containing four apple domains (A1-4) and a C-terminal trypsin-like catalytic domain. The DFP binds to the second apple domain of coagulation factor XI and plasma kallikrein heavy chain (Kumar *et al.*, 2020); nanobodies against factor XI apple 3 domain inhibit its protein–protein interaction, evidencing the importance of these domains (Barroeta *et al.*, 2022). These proteins are typically present in mammals and absent in plants, but coagulation factor XI and plasma kallikrein are serine proteases (Li *et al.*, 2023); plant serine proteases play key roles in plant defense (Balakireva and Zamyatin, 2018; Ekchaweng *et al.*, 2017). The phytoplasma effectors containing LIG_FXI_DFP_1 may target some of those proteases. The last motif in the top 10 motifs identified here was LIG_PTB_Phospho_1, present in SAP08, SAP67 and PHY1. The LIG_PTB_Phospho_1 motif binds short peptides with a core Asn-X-X-Tyr motif, phosphorylated on the Tyr residue. To the best of our knowledge, the motif Asn-X-X-Tyr motif has not been described in plants, making it difficult to predict the plant targets for

effectors harboring the `LIG_PTB_Phospho_1` motif, but tyrosine kinases involved in defense responses are possible targets of effectors with this motif (Bender and Zipfel, 2023; Cock *et al.*, 2002; Luan, 2002; Shiu and Bleecker, 2001). Tyrosine phosphorylation plays an important role in plant cell signaling (Luan, 2002), and plant and animal tyrosine kinases share ancestral origin (Shiu and Bleecker, 2001). Table 2.4 corresponds to a summary of the information about the top 10 SLiMs.

Table 2.4 The top 10 ELM motifs found in phytoplasma effectors.

SLiMs	SLiM Pattern	SLiM Description	Number of Effectors Proteins	ID of Effector Proteins
LIG_PDZ_Class_2	...[VLIFY].[ACVILF]\$	The C-terminal class 2 PDZ-binding motif is Classically represented by a pattern such as (VYF)X(VIL)	10	A0A859I9H9, A0A531XZH1, A0A0P7J0F0, A0A531XYW6, A0A859I9B6, A0A0P7KH40, A0A531Y2Y1, A0A859IA25, A0A7S7FZA6, A0A1E2USP0
LIG_DCNL_PONY_1	^M[MIL].[MIL]	DCNL PONY domain binding motif variant based on UBE2M and UBE2F interactions.	10	A0A975FJD4, A0A0P7JJB4, A0A531XZZ7, A0A0N8IAS3, A0A7G3ZP82, A0A975FIG9, A0A531Y098, A0A0P7KHL3, A0A531Y0A1, A0A7S7FZA7
LIG_PCNA_PIPBox_1	[QM].[^FHWY][LIVM][^P][^PF WYMLIV](((FYHL)[FYW])((FYH)[FYWL]))..	The PCNA binding motifs include the PIP-box and APIM motifs, and are found in proteins involved in DNA replication, repair, methylation and cell cycle control.	8	A0A975FJD4, A0A531XZZ7, A0A859IAA9, A0A975FIG9, A0A975FIA7, A0A531Y053, A0A859I9W5, A0A081D424
LIG_GBD_Chelix_1	[ILV][VA][^P][^P][LI][^P][^P][^P][LM]	Amphipathic alpha-helix that binds the GTPase-binding domain (GBD) in WASP and N-WASP.	8	A0A859I9H9, A0A0P7IR40, A0A859I9B6, A0A531Y053, A0A859I9W5, A0A7S7FZA6, A0A1J7D1C9, AYJ01076.1
LIG_NRP_CendR_1	[RK].[0,2]R\$	The CendR motif has a carboxy-terminal arginine, which binds to the neuropilin b1 domain binding site. CendR motifs are either located at the protein C-terminus or are generated by internal cleavage by a polybasic protease, such as furin.	7	A0A531Y0U6, A0A531XZZ7, A0A7G3ZP82, A0A975FIG9, A0A531Y090, A0A3G3BKC1, F8QQ90

TRG_ER_dilys_1	K.{0,1}K.{2,3}\$	ER retention and retrieving signal found at the C-terminus of type I ER membrane proteins (cytoplasmic in this topology). Di-Lysine signal is responsible for COPI-mediated retrieval from post-ER compartments.	7	A0A0L0MK48, A0A0N8IAS3, A0A531Y0U5, Q7M1T6, A0A076EAG8, A0A0D6DTS2, A0A4E9CZE5
DOC_PIKK_1	[DEN][DEN].[2,3][ILMVA][DEN][DEN]L	DOC_PIKK_1 motif is located in the C terminus of Nbs1 and its homologues and interacts with PIKK family members.	6	A0A975ILX0, A0A859I930, A0A0P7J0F0, A0A531Y0U5, A0A385GMC4, A0A3G3BKC1
DOC_CYCLIN_yCib1_Lx_F_4	(P.[KR]L.F)(.N[KR]L.F)(.N.L.F[LMIVFY])	The LxF motif found in budding yeasts serves as a docking site for mitotic cyclin-CDK complexes (M-CDK). It is found in both regulators and mitotic phosphorylation target proteins.	6	A0A859I9U4, A0A859IAA9, A0A531Y0U5, A0A531Y053, A0A859I9W5, A0A7S7FZA6
LIG_TYR_ITSM	.T.(Y)..[IV]	ITSM (immunoreceptor tyrosine-based switch motif). This motif is present in the cytoplasmic region of the CD150 subfamily within the CD2 family and it enables these receptors to bind to and to be regulated by SH2 adaptor.	6	A0A531XZZ8, A0A531Y0C0, A0A531Y0U5, A0A531XYW6, A0A859I9B6, A0A1E2USP0
LIG_FXI_DFP_1	[FYWHIL].DF[PD]	The DFP motif enables binding to the second apple domain of coagulation factor XI (FXI) and plasma kallikrein heavy chain.	5	A0A975FJ78, A0A531XZZ8, A0A531Y0C0, A0A5J6ED47, A0A5J6EFI7
LIG_PTB_Phospho_1	(.[*P].NP.(Y))(.[ILVMFY].N..(Y))	This phosphorylation-dependent motif binds to Shc-like and IRS-like PTB domains. The pTyr is positioned within a highly basic-charged anchoring pocket. A hydrophobic residue -5 (compared to pY) increases the affinity of the interaction.	5	A0A0P7IRG7, A0A859I9N5, A0A1V0PKR5, A0A0A8JBX1, A0A0A8JBX6

2.4. DISCUSSION

Phytoplasmas are diverse pathogens that cause severe problems in agriculture and horticulture. Integrated control management of the main phytoplasma diseases is expected to include early detection (Nair and Manimekalai, 2021), the use of nonantibiotic antimicrobials (Bertaccini, 2021), the use of resistant plant material (Oliveira *et al.*, 2019) and novel strategies for environmentally

friendly insect vector control (Tzec-Simá *et al.*, 2022). In line with these strategies, effectoromics may play a key role in crop protection since effectors are targetable genes (Galletto *et al.*, 2021; Tomkins *et al.*, 2018), or they may be used as “hunter genes” to detect genes encoding resistance proteins (Domazakis *et al.*, 2017; Vleeshouwers *et al.*, 2008). These strategies are already strengthening certain genetic improvement programs (Maidment *et al.*, 2023; Cesari *et al.*, 2022).

Effectoromics is currently a rapidly growing and evolving area. Until recently, the description of effectors in bacteria, fungi and phytoplasmas was “small, TMD-lacking secreted proteins” (Carreón-Anguiano *et al.*, 2020; Music *et al.*, 2019; Bai *et al.*, 2009). It is now clear that noncanonical (or non-classical) effectors that do not meet these criteria also exist (Carreón-Anguiano *et al.*, 2022). In fungi and oomycetes, the description of effectors has been rapidly changing, since validated novel effectors with sizes larger than 400 amino acids or with TMDs, or without a signal peptide, among other non-classical characteristics, have been discovered (Carreón-Anguiano *et al.*, 2022). Most reports have identified effectors in phytoplasma genomes while looking for proteins with signal peptides, and the absence of TMDs. Contrarily, Debonneville *et al.* (2022) recently drew attention to effectors with TMD in *Flavescence dorée* phytoplasma, and Gao *et al.* (2023) demonstrated the existence of secreted effectors that lack a signal peptide (SP) but are secreted through Sec 2-independent secretion pathway in ‘*Ca. Phytoplasma ziziphi*’, evidencing that non-classical effectors also exist in phytoplasmas. Here, the analysis of 738 phytoplasma effectors revealed that non-classical effectors are more common in phytoplasmas than previously believed. Approximately 30% of the known effectors lack SP, ~30% have TMDs and almost 10% are greater than 400 amino acids in length (Table 2.2). This list of phytoplasma effectors comprises membrane-bound proteins such as *IdpA*, *Imp*, *VmpA* and *Amp*. Based on more traditional descriptions of effectors as “no TMD, secreted proteins”, these proteins should be excluded, but it is known that the *Amp* protein interacts with the actin of insect vectors (Rashidi *et al.*, 2015), and recently, Wang *et al.* (2023) showed that the expression of this protein in *Nicotiana tabacum* inhibits plant host defense and promotes infection by the rice orange leaf phytoplasma, validating this protein as a true effector.

Currently, the list of experimentally validated phytoplasma effectors is less than 25, and functional effectoromics is tedious and only slowly advancing. In phytoplasma effectors,

homology has been described only for the effector families/tribes SAP05, SAP11, SAP54 and TENGU; other homology-based families/tribes such as *IdpA*, *Imp*, *VmpA* and *Amp* add to the list, but the number of effector families/tribes still remains very limited. This lack of conservation makes it difficult to progress in the functional characterization of phytoplasma effectors, since each effector must be studied individually. Homology is also usually low or nonexistent among effectors from other taxonomic kingdoms (Rozano *et al.*, 2023; Mak and Thurston, 2021).

The Gene Ontology database is the world's largest source of information on gene function. Unfortunately, the result of the GO analysis for the phytoplasma effectors was not very informative. Fourteen GO terms were retrieved but eight terms were assigned to the same effectors, leaving most effectors unassigned; the Gene Ontology result reflects our lack of knowledge about the functions of most phytoplasma effectors. In total, 462 phytoplasma effectors received a GO assignation; the GO term "membrane" was the largest one, assigned to 457 effectors. Of these 457 effectors, 298 correspond to *Amp*, *Stamp*, *VmpA*, *IdpA* and *Imp*; 37 correspond to SAP-like proteins; and 122 correspond to novel phytoplasma effectors. As Debonneville *et al.* (2022) evidenced, most unknown phytoplasma effectors may be transmembrane proteins. It is expected that in coming times, the phytoplasma effectoromes routinely include transmembrane proteins.

The second strategy to explore possible functions involved the identification of protein domains. The most frequently identified domain was the SVM domain, widely distributed in 87 PHYL1 and SAP phytoplasma effectors. The SVM domain is a signal sequence about 30 amino acids in length, which has been reported as a unique feature of phytoplasmal genome architecture (Jomantiene *et al.*, 2007). The genomes of ancestral relatives of phytoplasmas have no SVM-like structures and it is supposed that this domain arose from ancient phage attacks to phytoplasmas (Wei *et al.*, 2008). The second and third most frequent domains in phytoplasma effectors were restricted to *AMP* and *VmpA*, respectively, and are related with their location on the membrane surface. Some domains suggest the functions of the effectors that contain them, such as "chaperonin (GroEL) domain", "response regulator transcription factor", "Nucleic acid-binding", and "Peptidase M41-like domain". This highlights new opportunities for future research on these effectors.

The classification of effectors in tribes (protein families that share motifs) is useful to accelerate effectoromics (Carreón-Anguiano *et al.*, 2022; He *et al.*, 2020; Lorrain *et al.*, 2015; Saunders *et al.*, 2012), since shared motifs are related to common functions (Espadaler *et al.*, 2006; Polacco and Babbitt, 2006), and it was the third strategy that was followed here. Fifteen families or tribes were distinguished based on the different combinations found of protein (de novo) motifs. The occurrence of these effector tribes in the phytoplasma genomes revealed two evolutionary histories: the tribes 1, 6 and 7 are widely distributed, suggesting they come from the common ancestor of the phytoplasmas and were inherited by vertical transfer while other groups of phytoplasmas do not have effector members in these tribes, suggesting gene loss, which is common in the evolutionary histories of effectors (Latorre *et al.*, 2020; Fouché *et al.*, 2018).

Other tribes (2, 3, 4 and 5) were restricted to particular phytoplasmas; these effector tribes probably arose through the interaction and coevolution with the hosts. For example, the tribe 3 is characteristic of phytoplasmas in group 16SrV: elm yellows phytoplasma genome (16SrV), alder yellows phytoplasma (16SrV-A), 'Ca. *Phytoplasma vitis*' (Flavescence dorée phytoplasma) (16SrV-C); in the alder yellows phytoplasma genome, this tribe has been amplified. Tribe 5 was also found principally in phytoplasmas belonging to the 16SrV phytoplasma group: the alder yellows phytoplasma (16SrV-A), *Ziziphus jujuba* witches' broom phytoplasma (16SrV-B) and Flavescence dorée phytoplasma (16SrV-C), with 17, 2 and 15 effectors per genome, respectively, suggesting that this tribe comes from a common ancestor of phytoplasmas belonging to the phytoplasma 16SrV group and became amplified in certain genomes. One or two members of tribe 5 are present in the genome of phytoplasmas of other 16Sr phytoplasma groups. This tribe probably arose in phytoplasmas of the phytoplasma group 16SrV and arrived at phytoplasmas of other 16Sr phytoplasma groups by horizontal gene transfer (gain of genes), which is also a common mechanism of genome evolution in other organisms for the genomic content of effectors (Tiwari and Bae, 2020; Schweizer *et al.*, 2018; Fouché *et al.*, 2018; Huang and Gogarten, 2006; Omelchenko *et al.*, 2003). Gain and loss of effectors are common in all microbial kingdoms, and drive the patchy/discontinuous distribution typical of an effector's phylogenetic distribution (Morgado and Vicente, 2022; Dillon *et al.*, 2019).

De novo motifs were useful in the organization of nonhomologous effectors, and this classification may help accelerate the discovery of effector functions by studying only a few members per tribe as representatives of that family. In our analysis, de novo motifs did not reveal much information about the effector functions. Known motifs (SLiM or ELM) were identified in the repository of the ELM server, which is a comprehensive database of known experimentally validated motifs (Kumar *et al.*, 2022). Although the ELM acronym means “eukaryotic linear motif”, the search for ELMs is also useful for studies on pathogenic prokaryotes, since pathogens mimic features of critical host proteins to hijack their cell machinery, promoting infection of the host (Tayal *et al.*, 2022; Elkhaily *et al.*, 2022; Garg *et al.*, 2022; Sámano-Sánchez and Gibson, 2020). Effectors that do not share high overall sequence identity, but share motifs, domains or similar tridimensional structures, may share similar functions and are termed “functional orthologs” (Stergiopoulos *et al.*, 2010).

ELM analysis can only be performed on a single protein sequence, which is a tedious task when high-throughput analysis is required. As such, the complete dataset of 738 phytoplasma effector proteins could not be analyzed; instead, we analyzed a subset of 87 amino acid sequences, which represented all classes of phytoplasma effector proteins known to date, and we were able to identify 42 SLiMs. Each of these protein signatures occurs in different classes of effectors, for example, SAPs and Vamp share the SLiM TRG_ER_diLys_1. We revised, in detail, the top 10 SLiMs (those most frequently found in phytoplasma effectors) and found that some SAP effectors may regulate their targets through proteolysis, either binding ubiquitin-conjugating enzyme or docking serine proteases. Proteolysis plays a central role in plant–pathogen interaction; both actors, pathogen and host, use proteases for defense and attack (Figaj *et al.*, 2019; Thomas and Van Der Hoorn, 2018; Pearson *et al.*, 2017). Some phytoplasmas effectors with protease activity are SAP05 and SAP54 (Huang *et al.*, 2021), but the identification of the SLiMs LIG_DCNL_PONY_1 and LIG_FXI_DFP_1 suggests that proteolysis plays a more important role than evidenced in the pathogenicity of phytoplasmas (Kumar *et al.*, 2022). Other known mechanisms of pathogens of other kingdoms were also revealed as probable infection mechanisms in phytoplasmas, for example, actin polymerization by effectors containing the motif LIG_GBD_Chelix_1 (Mészáros *et al.*, 2021; Cheng *et al.*, 2008), or targeting of the host nuclei and transcription factors by effectors with the motif LIG_PCNA_PIPBox_1 (Kumar *et al.*, 2022). Other host targets are more difficult to predict, such as the target for

the SLiM LIG_NRP_CendR_1 that binds to the human neuropilin b1 domain binding site (Kumar *et al.*, 2022), or the motif LIG_PTB_Phospho_1 that binds short peptides with a core Asn-X-X-Tyr motif (Kumar *et al.*, 2022), which has not been found in plant proteins. Other effectors seem to be localized part of the time in the cytosolic and part of the time bound to membrane proteins, approaching the plasma membrane to interact with their targets (Zhang *et al.*, 2022). Other effectors appear to regulate cell cycle or signaling pathways (Mukhtar *et al.*, 2016; Alto and Orth, 2012). Interestingly, some of the top 10 SLiMs have been identified in bacterial effectors such as DOC_PIKK_1 identified in the effector proteins CagA and Tir of the bacteria *L. pneumophila* and *L. pasculli* (Sampietro *et al.*, 2018) and the motif LIG_TYR_ITSM identified in effector proteins of *C. diphtheria* (Weerasekera *et al.*, 2018), reinforcing the potential role of these motifs in phytoplasmas for the successful infection of the host.

Pathogens and hosts are constantly coevolving, and SLiMs/ELMs are becoming increasingly studied in pathogens, as these mimic critical host motifs, allowing mimicry peptides (mimitopes) to sabotage key host processes (Davey *et al.*, 2023; Martins *et al.*, 2023; Goswami *et al.*, 2023; Elkhaily *et al.*, 2022; Irwin *et al.*, 2021; Mondino *et al.*, 2020). Since these SLiMs are abundant and critically important in pathogenesis, they are now being studied as novel drug targets (Alnomasy, 2021; Garg *et al.*, 2022), either to screen for novel molecules (Garg *et al.*, 2022), or by docking-based peptide design that disrupts mimitope–host target interaction (Delaunay and Ha-Duong, 2023; Kazmirchuk *et al.*, 2023; Chang *et al.*, 2022). To the best of our knowledge, this is the first report of motifs (known motifs and de novo motifs) in phytoplasma effectors. Some effectors that share common characteristics may target the same protein in the host, named hub proteins, which play key roles in the host cell (Todd *et al.*, 2022). For example, the effectors IPI-01 and IPI-04 from the oomycete *Phytophthora infestans* are able to interact and interfere with functionality of the RB protein, an NLR immune receptor of *Solanum bulbocastanum* Dunal (Zhao and Song, 2021). In other example, the type III Effectors NopT and NopP of rhizobial bacteria target the hub protein GmPBS1, which is key for nodulation in soybean (Li *et al.*, 2023). The identification of common targets for different members of phytoplasma effector families/tribes may enable the identification of key host proteins for further research. Since all these analyses are in silico, further experimental validation is necessary. Interaction between effectors and host targets may be confirmed through in vitro double-hybrid tests and in vivo through the

bimolecular fluorescence complementation (BiFC) assay (Strohmayr *et al.*, 2021). The host's hub proteins may be used as molecular markers to assist genetic improvement programs or for direct genetic manipulation through CRISPR-Cas9 technology for the development of phytoplasma-resistant plants (Ijaz *et al.*, 2023). Likewise, the identification of targets in insect vectors may enable us to control phytoplasmas through the modulation of gene expression in insects with gene silencing (Pacheco *et al.*, 2020; Wang *et al.*, 2023). In addition to their practical applications, these experimental evaluations also uncover the mechanisms used by phytoplasma effectors and effector roles during host–pathogen interactions. Findings here in the amino acid sequences of phytoplasma effectors open exciting lines of research to unravel effector roles and explore novel strategies for phytoplasma control.

In addition to the novel insights from the identification of protein domains, *de novo* motifs and known motifs, the analysis of amino acid compositions revealed interesting results about phytoplasma effectors. They were found to be rich in the amino acids Lys, Leu and Asn, and poor in Trp and Cys. A similar composition of amino acids was reported by Singh and Lakhanpaul (2020) in ortholog proteins of SAP54, which were rich in the amino acids asparagine (19.7%), leucine (16%) and lysine (10.9%), while cysteine and tryptophan were absent.

The proteomes of the unicellular eukaryotes *Plasmodium falciparum*, *P. berghei*, *P. chabaudi* and *P. yoelii* have been dominated by Lys-rich, Asn-rich and Ile-rich proteins (Filisetti *et al.*, 2013). The genomes of these parasites are AT-rich, as phytoplasma genomes usually are (Tan *et al.*, 2021; Liefting and Kirkpatrick, 2003; Wei *et al.*, 2008), which leads to an abundance of the trinucleotide AAT, coding for Asn (Filisetti *et al.*, 2013; Muralidharan and Goldberg, 2013). In *P. falciparum*, the causal agent of the malaria disease, the Asn-rich proteins are fibrillar (Rajapandi, 2020), and their abnormal extracellular accumulation leads to the formation of the amyloid, irreversible insoluble protein aggregates that deposit in organs and tissues and affect the heart, kidneys, liver, nerves and digestive system of the human host (Moles *et al.*, 2015). For this biochemical behavior, these proteins have been named “prion-like” (Pallarès *et al.*, 2018). Prions have been linked with diseases in eukaryotes, but a prion-like protein was first described in *Clostridium botulinum* that behaved like a prion when it was expressed in *Escherichia coli* bacteria and yeast (Choi, 2017). To

date, it is known that prion-like proteins are part of bacterial proteomes, and the largest fractions are found in the *Mycoplasma* genus (Harrison, 2019), a relative of the *Phytoplasma* genus; both belong to the bacterial class Mollicutes but the former are bacterial parasites of animal pathogens (Chen *et al.*, 2012). Prions may be ancient in the history of life (Jheeta *et al.*, 2021; Navarro *et al.*, 2015) since they have been found in phages of various groups of bacteria and archaea (Tetz and Tetz, 2017), and the spike protein of the SARS-CoV-2 virus (Tetz and Tetz, 2022); therefore, their presence in phytoplasmas would not be rare. Prionic signatures should be identified in phytoplasma proteomes and may help to unravel the function of some members in the effectoromes. The prion-like aggregated structure probably confers protection to the phytoplasma effectors against proteolytic degradation (Shahnawaz *et al.*, 2017). Richness in Asn in the proteins is typically associated with regions of low complexity, probably forming bulging domains (Filisetti *et al.*, 2013); in *P. falciparum*, these low-complexity regions in the proteins play diverse roles throughout the parasite's lifecycle, from mediating protein–protein interactions to enabling the parasite to evade the host immune system (Davies *et al.*, 2017). In the phytoplasma effectors named phyllogens, Iwabuchi *et al.* (Iwabuchi *et al.*, 2020) demonstrated the importance of asparagine and asparagine–lysine amino acids in the induction of changes in flower morphology.

In the case of Lys-rich proteins, Lys-rich secreted effectors were recently reported in the fungus *Fusarium oxysporum* f. sp. *lycopersici* and were found regulated by acetylation (Li *et al.*, 2020). Acetylation is a dynamic and reversible posttranslational modification that regulates protein functions through modifying enzymatic activity, interactions with DNA, protein stability, protein localization and the interaction with other proteins (Narita *et al.*, 2019). Most of the genes of *F. oxysporum* f. sp. *lycopersici* for Lys-rich proteins were upregulated during initial infection of the plant host (Li *et al.*, 2020), revealing that they are indeed involved in pathogenesis. In prokaryotes, lysine acetylation regulates proteins involved in transcription, translation, pathways associated with central metabolism and stress responses (Jones and O'Connor, 2011). These regulatory functions are consistent with the roles of a number of phytoplasma effectors that bind other proteins and regulate gene transcription and the metabolisms of phytohormones (Bai *et al.*, 2023; Aurin *et al.*, 2020; MacLean *et al.*, 2014); however, whether acetylation occurs in phytoplasma proteins remains to be determined.

The amino acid leucine contributes to the aliphatic index in proteins. Leu is common to both plant receptors that recognize pathogen effectors (Bauer *et al.*, 2021; Qi *et al.*, 2012) and bacterial effectors, such as the SlrP effector from *Salmonella enterica* that targets the human chaperone ERdj3 (Pillay *et al.*, 2023; Bernal-Bayard *et al.*, 2010b); the SspH1, SspH2 and Slrp effectors of *S. typhimurium* required for normal pathogenesis in animal models (Miao *et al.*, 1999); the XC1553 effector from *Xanthomonas campestris* pv. *campestris* that is recognized in vascular tissue of *Arabidopsis thaliana* (R.-Q. Xu *et al.*, 2008); the YopM effector from *Yersinia pestis*, which has a C-terminal E3 ligase domain (NEL domain) (Hu *et al.*, 2016); and the GALA5 effector from *Ralstonia syzygii* (Remigi *et al.*, 2011), among others. Leu-rich proteins are widely distributed in bacteria, playing important roles in various protein–protein interaction processes (Kibby *et al.*, 2023; Eshghi *et al.*, 2019; Loimaranta *et al.*, 2009); this suggests that this effector feature existed before the phytoplasmas arose from their bacterial ancestor (Galletto *et al.*, 2008; Singh and Lakhanpaul, 2020). However, all these bacterial effectors are secreted through the type III secretion system (Bernal-Bayard *et al.*, 2010; R.-Q. Xu *et al.*, 2008), while most of the effectors of phytoplasmas are secreted through the Sec-dependent pathway (Bai *et al.*, 2009). Leucine also plays a role in pathogenicity in other microorganisms, for example, in fungi and oomycetes, the motif RXLR at the C-terminus of some effectors often suppresses plant immunity (Liu *et al.*, 2019; Padilla-Ramos *et al.*, 2019).

Regarding the amino acid cysteine, it forms inter- or intramolecular disulfide bonds, which confer thermal stability to the protein structure under oxidizing conditions (Bhopatkar *et al.*, 2020). In fungi, many effectors are Cys-rich proteins; this characteristic confers stability in the acidic environment of the host apoplast (Singh *et al.*, 2021; Zhang *et al.*, 2021). Conversely, in phytoplasma effectors, the presence of the amino acid Cys was found to be limited or absent. Miseta and Csutora (Miseta and Csutora, 2000) analyzed the content of Cys in the proteins of phylogenetically distant organisms and found an evolutionary trend favoring the incorporation of more cysteine residues into proteins of more complex organisms, as an “order-promoting” condition (Gérard *et al.*, 2022; Miseta and Csutora, 2000). Stability in phytoplasma effectors is probably driven by other means, for example, through low-complexity regions or prionic signatures in the proteins. In the case of Trp, this amino acid is usually rare in proteins, since it is encoded by only one codon. Further research is needed to establish the functional significance of these amino acid patterns.

All these novel findings shed light on the world of effectoromics in phytoplasmas. It will be interesting to address some of the hypotheses that are raised from these computational results.

CAPÍTULO III

ARTÍCULO 2. PHYEFFECTOR, THE FIRST ALGORITHM THAT IDENTIFIES CLASSICAL AND NON-CLASSICAL EFFECTORS IN PHYTOPLASMAS

Autores:

Karla Gisel Carreón-Anguiano⁺, Sara Elena Vila-Luna⁺, Luis Sáenz-Carbonell and Blondy Canto-Canché^{*}

⁺ Estos autores contribuyeron igualmente en este trabajo

Cita: Carreón-Anguiano, K.G.; Vila-Luna, S.E.; Sáenz-Carbonell, L.; Canto-Canché, B. PhyEffector, the First Algorithm That Identifies Classical and Non-Classical Effectors in Phytoplasmas. *Biomimetics* 2023, 8,550. <https://doi.org/10.3390/biomimetics8070550>

Abstract: Phytoplasmas are the causal agents of more than 100 plant diseases in economically important crops. Eleven genomes have been fully sequenced and have allowed us to gain a better understanding of the biology and evolution of phytoplasmas. Effectors are key players in pathogenicity and virulence, and their identification and description are becoming an essential practice in the description of phytoplasma genomes. This is of particular importance because effectors are possible candidates for the development of new strategies for the control of plant diseases. To date, the prediction of effectors in phytoplasmas has been a great challenge; the reliable comparison of effectoromes has been hindered because research teams have used the combination of different programs in their predictions. This is not trivial since significant differences in the results can arise, depending on the predictive pipeline used. Here, we tested different predictive pipelines to create the PhyEffector algorithm; the average value of the F1 score for PhyEffector was 0.9761 when applied to different databases or genomes, demonstrating its robustness as a predictive tool. PhyEffector can recover both classical and non-classical phytoplasma effectors, making it an invaluable tool to accelerate effectoromics in phytoplasmas.

3.1 INTRODUCTION

Phytoplasmas are pleomorphic, mycoplasma-like bacterial pathogenic microorganisms. These organisms are between 0.2 to 0.8 μm in diameter and lack cell walls; they reside in plant phloem and are transmitted by phloem-feeding insect vectors (Bertaccini, 2007; Hogenhout *et al.*, 2008; Oshima *et al.*, 2013). Phytoplasmas cause severe damage to the agriculture and horticulture industries worldwide, with extensive yield losses in thousands of economically important crops such as vegetables, spices, medicinal plants, ornamentals, palms, fruit trees, among others (Kirdat *et al.*, 2023). Infected plants display symptoms such as phyllody (transformation of floral organs into tissues similar to leaves), virescence (green coloration in floral organs), yellowing (chlorosis), stunting, Little Leaf disease, and Witches' broom (proliferation of shoots) (Hemmati *et al.*, 2021; Bertaccini, 2022). Phytoplasmas cause these abnormal plant morphologies through the secretion of virulent proteins known as effectors, which interfere with the hormone signaling of their hosts (Sugio *et al.*, 2011a; Rashid *et al.*, 2018).

Effectors are defined as small molecules that selectively bind to a protein and regulates its biological activity (Singh *et al.*, 2019). These molecules are fundamental in the parasite-host interaction, where the parasite is metabolically dependent on its host; a struggle of forces is established between the attack of the parasite and the defense of the host. Effectors play key roles in the successful infection of the plant host by the pathogen (Padilla-Ramos *et al.*, 2019).

To date, eleven phytoplasma genomes have been sequenced (Cho *et al.*, 2020); KEGG, <https://www.genome.jp/kegg/genome/>, ranging from ~570Mb to ~950Mb, and comprised ~470 to 870 proteins (Debonneville *et al.*, 2022). The number of predicted effectors reported per genome ranges from 10 in the genome of 'Ca. Phytoplasma asteris' De Villa (Cho *et al.*, 2020) to 65 effectors in the genome of the AY-WB strain OY-M (Chung *et al.*, 2013), and 98 in 'Ca. Phytoplasma aurantifolia' (Tan *et al.*, 2021).

Phytoplasma effectors are secreted proteins, usually with no transmembrane domains. However, although this description of the effectors is relatively simple, their identification has followed different routes. For signal peptide identification, different versions of SignalP have been used, a program which has been trained to predict the presence of signal peptides (SP) in gram-negative or gram-positive bacterial proteins. (Bai *et al.*, 2009) evaluated

SignalP v3.0 on 369 mollicute protein sequences that included 46 experimentally verified secreted proteins; the program was able to identify 43 of these 46 secreted proteins. After using SignalP v3.0, and eliminating the SP (20-50 amino acids) sequence, the proteins were analyzed using the TMHMM v2.0 program to identify and eliminate proteins with transmembrane domains (TMDs). Using this approach, 56 effector candidates were found in the genome of the aster yellows phytoplasma strain witches' broom.

(Anabestani *et al.*, 2017) used a similar approach, but compared SignalP v3.0 and SignalP v4.0, selecting the latter as the predictor for their secreted proteins; they predicted 28 effectors in '*Ca. Phytoplasma aurantifolia*', nine of them with nuclear localization. (Wang *et al.*, 2018b) used the same pipeline, combining SignalP v4.0 and TMHMM v2.0 after eliminating the SP and found 28 effectors in the Jujube Witches' Broom (JWB) Phytoplasma genome. Recently, (Tan *et al.*, 2021) compared the use of SignalP v4.1 vs SignalP v5.0. With the latter they found 28 extracellular proteins, while SignalP v4.1 excluded one of these proteins, but retrieved 70 additional secreted proteins. However, they found by manual analysis that many of these proteins were false positives, warning users about accuracy of the predictions depending on the version of SignalP used. Previously, (Cho *et al.*, 2020), used SignalP v5.0 (in Gram-positive bacterial mode), then, filtered with TMHMM v2.0 to retain the secreted proteins with no TMDs; they analyzed 11 phytoplasma genomes and found differences in the results previously reported by (Music *et al.*, 2019) using SignalP v3.0. For example, these authors predicted eight effectors for AYWB strain, and 15 effectors for OY-M strain, while (Cho *et al.*, 2020) predicted 33 and 37 effectors respectively.

Most of the previous pipelines focused on predicting effectors as secreted proteins with no TMDs. However, a few known phytoplasma effectors are transmembrane proteins, such as Amp (Rashidi *et al.*, 2015) and Imp (Boonrod *et al.*, 2012). Very recently, (Debonneville *et al.*, 2022) combined the use of SignalP v5.0 with Phobius to predict the presence of signal peptide and TMDs in effectors from the '*Flavescence Dorée*' Phytoplasma. This combination identified 17 effector candidates; Phobius identified three secreted proteins with a TMD and a SP. In addition, Phobius was able to retrieve two putative secreted proteins which are unique to the genome of this phytoplasma, which suggests that Phobius must be considered in pipelines for the identification of effectors in phytoplasmas.

Interestingly, (Gao *et al.*, 2023) identified 95 non-classically secreted proteins (ncSecPs) in the '*Ca. Phytoplasma ziziphi*' genome for the first time and confirmed that 25 of them are

true secreted proteins. Non-classically secreted proteins lack SPs or translocation signals, but are still exported to the extracellular space through a Sec-independent secretion pathway. These proteins are not identified by the SignalP program but by SecretomeP. These novel results show that classic pipelines for the identification of phytoplasma effectors must be revised and innovated.

Here, we performed a comparison of different pipelines comprising either SignalP v3.0, or SignalP v4.1 or SignalP v5.0, with TMHMM v2.0 (on SP-lacking proteins), followed by sequential analysis with Phobius, SecretomeP and sequence-homology-based Blastp. The false positives retrieved by any of the steps were eliminated. The performance of the pipelines was tested on a positive and a testing datasets (comprising phytoplasma effectors) and negative dataset (phytoplasma non-effectors proteins), resulting in F1 scores from 0.56 to 1.0. The best algorithm was named PhyEffector and it was further evaluated on phytoplasma genomes (i.e., deduced proteomes), resulting in an average F1 score of 0.9761. These results showed that PhyEffector is a robust predictor of phytoplasma effectors, with very low false negatives (generally <10). PhyEffector is able to retrieve both classical effectors (containing a SP and without TMD) as well as non-classical (atypical) effectors from phytoplasmas (effectors that contain a TMD or lack SP and are secreted by a Sec-independent secretion pathway), which makes PhyEffector a powerful computational tool to speed up phytoplasma effectoromics.

3.2 MATERIALS AND METHODS

3.2.1. Creation of databases

Positive dataset: The positive dataset was constructed with 10 of 21 validated phyto- plasma effectors reported in the literature (Table 3.1): 5 of them canonical effectors (secreted and identified by any version of the SignalP program) and 5 non-canonical ones (with a TMD, or secreted by a non-classical pathway). The list was completed by searching in the UNIPROT database with the keywords: SAP01, SAP02, SAP03, until SAP80; TENGU, phyllody, phyll, antigenic membrane protein (Amp), Immunodominant membrane protein A (idpA), Immunodominant membrane protein (Imp), and Variable membrane protein A (VmpA). The complete retrieved list contained 738 protein sequences. To prevent over-representation due to large effector families, 4 members were chosen per family, except for SAP11, for

which sequence identity among them was lower than 45%. Similarly, to avoid over-fitting due to highly conserved effector sequences, members with ~50% sequence identity were selected, except for largely conserved families such as TENGU, idpA and PME. The final list of positive datasets of phytoplasma effectors comprised 64 sequences (Table S1).

A second database was constructed for testing PhyEffector, comprising the other 11 true effectors (6 canonicals and 5 non-canonicals), plus 215 of the other potential effectors retrieved from UNIPROT. These 226 potential effectors share ~50% identity with protein sequences in the positive database.

Negative dataset: Core phytoplasma proteomes were identified after manual revision of phytoplasma genomes available at KEGG (<https://www.genome.jp/kegg/genome/>; accessed on 15 July 2023). Conserved KEGG pathways were identified, such as: glycolysis/gluconeogenesis, citrate cycle (TCA cycle), pentose phosphate pathway, fructose and mannose metabolism, purine metabolism, pyrimidine metabolism, glycine, serine and threonine metabolism, valine, leucine and isoleucine degradation, etc. Among these KEGG pathways, 64 core protein sequences with 40 different functional annotations were used to construct the negative dataset (Table S2).

3.2.2 In Silico Characterization of Effectors from Phytoplasmas

To gain knowledge about the features of currently known phytoplasma effectors, the positive dataset was analyzed with SignalP v4.1, TMHMM v2.0 on mature proteins (without signal peptide), Phobius and SecretomeP v2.0.

3.2.3 Characterization of Different Pipelines to Predict Effectors in Phytoplasmas: Construction of PhyEffector Algorithm

Different versions of SignalP: v3.0 (Bendtsen *et al.*, 2004), 4.1 (Petersen *et al.*, 2011), and 5.0 (Almagro *et al.*, 2019) were tested independently in Gram positive and Gram negative modes to identify secreted proteins, and after removal of the SP, the proteins were filtered with TMHMM v2.0 (Krogh *et al.*, 2001), resulting in each case in the set #1.

Prediction by Phobius server (set #2) (Käll *et al.*, 2007), SecretomeP (set # 3) (Bendtsen *et al.*, 2005), and Blastp analysis using the dataset of phytoplasma effectors as query (set # 4) were pooled, and then redundancies were eliminated in the results of each pipeline, becoming the “total potential effectorome” for each pipeline. These proteins were annotated performing a Blastp against the GenBank non-redundant protein database (<https://www.ncbi.nlm.nih.gov/>; accessed on 10 August 2023); proteins annotated as hypothetical proteins, unknown function, predicted proteins, no hits, as well as pathogenic/virulence related functions and those with annotations related to phytoplasma effectors, were selected and comprise the final list of effectors, while proteins with metabolic essential functions were discarded (Figure 3.1).

Table S3 shows the list of functional descriptors (annotations) of known effectors. The list of annotations of essential metabolic activities was obtained from (Cho *et al.*, 2020).

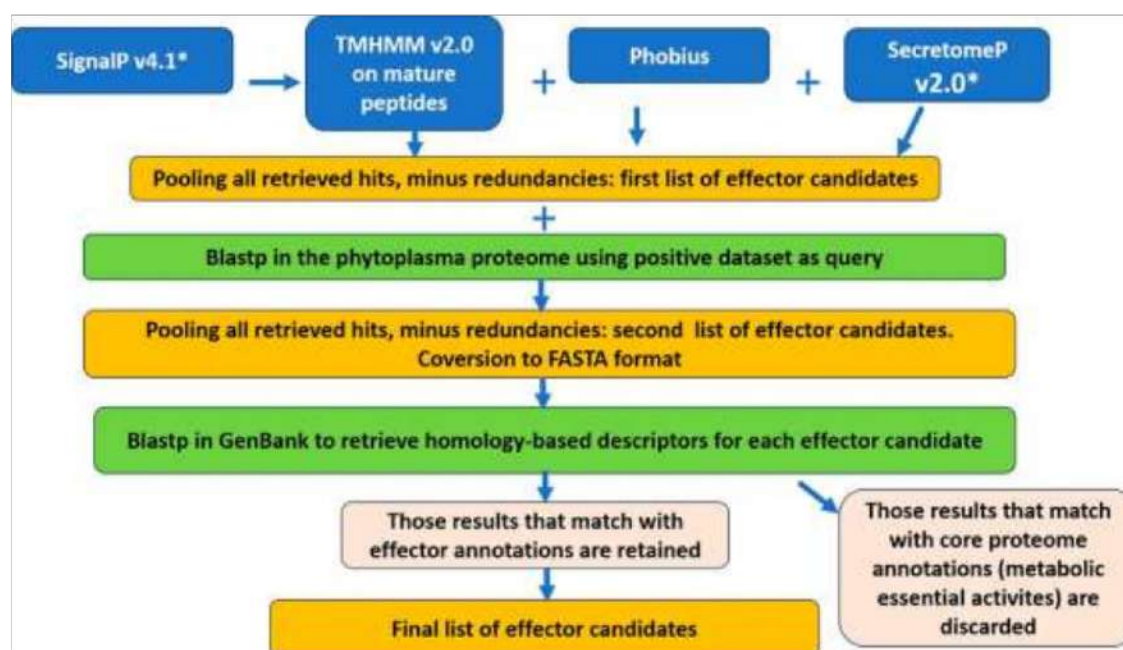


Figure 3.1 PhyEffector workflow for the prediction of effectors in phytoplasma proteomes. *, programs are run in Gram-positive mode. Mature proteins are signal peptide-lacking proteins. Blue squares, pre-installed prediction tools; green squares, Blastp analyses; pink squares, custom tools for filtering results based on descriptors; yellow squares, lists of effector candidates. All databases were constructed here, except the database of annotations of metabolic essential activities, which was based on the open access report from (Cho *et al.*, 2020).

These pipelines were run on the positive dataset (phytoplasma effectors) and the negative dataset (non-effector phytoplasma proteins). True positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) were determined to calculate sensitivity, specificity, precision and accuracy parameters. F1 scores were calculated to measure and compare performances of the pipelines.

3.2.4 Validation of PhyEffector Algorithm

PhyEffector was used to carry out the identification of effector candidates on a testing database comprising 11 true effectors (different from those 10 true effectors in the positive dataset) and the 385 protein sequences retrieved from UNIPROT, which shared ~50% identity with protein sequences in the positive dataset.

PhyEffector was validated through the identification of effectors in the genomes (i.e., deduced proteomes) of '*Ca. P. asteris*' AYWB (16SrI-A), '*Ca. P. asteris*' OY-M (16SrI-B), '*Ca. P. aurantifolia*' (16SrII), '*Ca. P. ziziphi*' Jwb-nky (16SrV-B), '*Ca. P. vitis*' of Flavescence dorée phytoplasma' (16SrV-C), '*Ca. P. luffae*' (16SrVIII), '*Ca. P. mali*' (16SrX), '*Ca. P. australiense*' PAa (16SrXII) and '*Ca. P. solani*' SA-1 (16SrXII); prediction by PhyEffector on each phytoplasma genome was compared with the results from the respective report in the literature. Discrepancies between the prediction from PhyEffector and the scientific literature were solved by determining the number of FP and FN for each prediction by following the criteria described above, i.e., identifying those FN using the list of effector annotations, and those FP with the list of annotations of essential metabolic activities.

Table 3.1 List of true phytoplasma effectors used in the present work.

Effector	Accession at GenBank/ UNIPROT	Homolog	Phytoplasma	Phenotype or Function	Observations	Dataset	Reference
Canonical, typical, or classical							
TENGU	BAH29766.1/ A0A4P6MDK8	-----	'Ca. Phytoplasma Asteris', strain Onion yellows phytoplasma OY-M. Group 16Srl.	Dwarfism, witches' broom symptoms and plant sterility. Pleiotropic effects on auxin and jasmonic acid	First reported witches' broom-inducing effector. Small protein (70-amino acid preprotein, of which 38 C-terminal amino acids are released into plant host)	Positive set	(Hoshi <i>et al.</i> , 2009; Sugawara <i>et al.</i> , 2013)
SAP05	8PFC A WP 011412316.1	-----	Aster Yellows phytoplasma strain Witches' Broom (AY-WB)	Induces witches' broom symptoms, proliferation of vegetative tissue and shoots.	Binds plant SPL and GATA transcription factors and mediates their degradation in a ubiquitin-independent manner	Positive set	(Huang <i>et al.</i> , 2021)
SAP11	GI:85057650	-----	Aster Yellows phytoplasma strain Witches' Broom (AY-WB). Crinkled leaves and siliques	CIN-TCP binding and destabilization, and impaired synthesis of jasmonic acid, and increase in leafhopper oviposition activity.	Modular organization; at least three domains are required for efficient CIN-TCP destabilization in plants	Positive set	(Sugio <i>et al.</i> , 2014)
SAP54	WP_252861407.1	-----	Aster Yellows phytoplasma strain Witches' Broom (AY-WB). Virescence	Degrading MADS-box Proteins; induces phyllody and sterile plants	-----	Positive set	(MacLean <i>et al.</i> , 2014)
PHYL1	LC388988.1, LC388989.1, LC388990.1, LC388981.1, LC388982.1, LC388983.1, LC388992.1, LC492887.1, LC388972.1, LC388985.1, LC388987.1	SAP-54	"Ca. Phytoplasma" species	witches' broom symptoms	Phyllogens (four groups: phyl-A, -B, -C, and -D)	Positive set	(Iwabuchi <i>et al.</i> , 2020)
SWP1	WP_024563292.1	SAP11-like	wheat blue dwarf phytoplasma	witches' broom symptoms	-----	Testing set	(Wang <i>et al.</i> , 2018b)
SWP11	No GenBank accession. Arbitrary authors' code WBD_0004	-----	Wheat blue dwarf phytoplasma	cell death and defence responses, including H ₂ O ₂ accumulation and callose deposition.	Up-regulation of HIN1, PR1, PR2 and PR3	Testing set	(Wang <i>et al.</i> , 2017)
SWP12	No GenBank accession. Arbitrary authors' code WBD_0238	-----	Wheat blue dwarf phytoplasma	suppress SWP11-, BAX-, and/or INF1-induced cell death	-----	Testing set	(Wang <i>et al.</i> , 2017)

SWP21	No GenBank accession. Arbitrary authors' code WBD_0274	TENGU-like	Wheat blue dwarf phytoplasma	suppress SWP11-, BAX-, and/or INF1-induced cell death	SWP21 has a distinct role in virulence compared with TENGU	Testing set	(Sugawara <i>et al.</i> , 2013; Wang <i>et al.</i> , 2017)
Zaofeng3	AYJ01078.1	SAP54-like	'Ca. Phytoplasma ziziphi' (JWB phytoplasma) (16SrV-B)	Overexpression showed phytoplasma-like symptoms	87% identity with SAP54PnWB	Testing set	(Chen <i>et al.</i> , 2022)
Zaofeng6	AYJ01297.1	SAP11-like	JWB phytoplasma	Overexpression resulted in shoot proliferation; triggered hypersensitive response and induced the expression of defense-related genes	48% identity with SAP11AYWB.	Testing set	(Chen <i>et al.</i> , 2022)
Non-canonical, atypical, or non-classical							
IdpA	ADD52250.1	-----	Poinsettia branch-inducing phytoplasma	Crucial role in plant and insect vector transmission	Immunodominant membrane protein A; transmembrane domain present	Positive set	(Kakizawa <i>et al.</i> , 2009)
Imp	CBJ17020.1	-----	'Ca. Phytoplasma mali'	Binds to plant actin; probably involved in phytoplasma motility in host plants	Immunodominant membrane protein; transmembrane domain present	Positive set	(Boonrod <i>et al.</i> , 2012)
VmpA	ULR56812.1	-----	Flavescence dorée phytoplasma	Binds the midgut of the insect vector and promotes adhesion to its epithelial cells.	Variable membrane protein A; transmembrane domain present	Testing set	(Arricau-Bouvery <i>et al.</i> , 2021)
Amp	WP071345415.1	-----	Rice orange leaf Phytoplasma	Suppresses host defenses. Interacts with actin of its vector; probably involved in vector specificity	Antigenic membrane protein; transmembrane domain present	Positive set	(Wang <i>et al.</i> , 2023)
ncSecP3	WP_161554967.1	-----	'Ca. P. ziziphi'	Suppresses hypersensitive cell death response (HR) in <i>Nicotiana bentamiana</i> , triggered by the pro-apoptotic mouse protein Bax and the <i>Phytophthora infestans</i> elicitor INF1	Non-classically secreted proteins (ncSecPs); non-secreted by Sec-pathway	Positive set	(Gao <i>et al.</i> , 2023)
ncSecP9	WP_121463838.1	-----	'Ca. P. ziziphi'	Suppresses HR in <i>Nicotiana bentamiana</i> , triggered by Bax and INF1	ncSecPs	Positive set	(Gao <i>et al.</i> , 2023)
ncSecP1 2	WP_161554974.1	-----	'Ca. P. ziziphi'	Suppresses HR in <i>Nicotiana bentamiana</i> , triggered by Bax and INF1	ncSecPs	Testing set	(Gao <i>et al.</i> , 2023)

ncSecP1 4	WP_121463915.1	-----	'Ca. P. ziziphi'	Suppresses HR in <i>Nicotiana bentamiana</i> , triggered by Bax and INF1	ncSecPs	Testing set	(Gao <i>et al.</i> , 2023)
ncSecP1 6	WP_161554978.1	-----	'Ca. P. ziziphi'	Suppresses HR in <i>Nicotiana bentamiana</i> , triggered by Bax and INF1	ncSecPs	Testing set	(Gao <i>et al.</i> , 2023)
ncSecP2 2	WP_121463976.1	-----	'Ca. P. ziziphi'	Suppresses HR in <i>Nicotiana bentamiana</i> , triggered by Bax and INF1	ncSecPs	Testing set	(Gao <i>et al.</i> , 2023)

3.3 RESULTS

3.3.1 Construction of Positive Dataset

The positive dataset was constructed with 21 validated phytoplasma effectors reported in the literature (Table 3.1), 11 of them canonical effectors (secreted, with SP and no TMD), and 10 non-canonical ones (with a TMD, or secreted by a non-classical pathway). The first attempt to complete the list of phytoplasma effectors was done through searches in GenBank using certain keywords (for example effector names), but results were highly redundant. For example, “phytoplasma effector sap11” retrieved 71 results, but 18 sequences for SAP11 effector protein from ‘Lime witches’-broom phytoplasma (IDs from QAB44970.1 to QAB44987.1) were 100% identical. High redundancy was observed as well in other results while “phytoplasma effector TENGU” retrieved no results from the GenBank (10 March 2023). Phytoplasma effectors were then searched for in the UNIPROT database using the keywords SAP01, SAP02, SAP03, until SAP80; other keywords used were TENGU, phyllody, phyll, antigenic membrane protein (Amp), Immunodominant membrane protein A (idpA), Immunodominant membrane protein (Imp) and Variable membrane protein A (VmpA).

The first list of phytoplasma effectors comprised 738 amino acid sequences. The list was carefully revised to eliminate potential false positives; 229 effector candidates annotated as “Candidatus effector” or “phytoplasma effector” were eliminated because they do not share identity with any known phytoplasma effectors and none of them have been experimentally validated. The preferred proteins were phytoplasma effectors with ~50% identity shared among each other to avoid over-representation or over-fitting of effector families; only 15% of the sequences share > 80% identity. The final list comprised 64 phytoplasma effectors (Table S1).

3.3.2 Characterization of Phytoplasma Effectors

The positive dataset was analyzed to classify the phytoplasma effectors in classical (secreted, with SP, with no TMD), and non-classical candidates (those that do not meet any of the characteristics of classical phytoplasma effectors) (Table 3.2). Most of the known phytoplasma effectors are predicted to be secreted through the canonical type II secretion system. From the other ~10%, half are predicted to be secreted by the non-classical pathway and the other half are predicted as non-secreted. Understandably, it was found that almost 92% were predicted to have no TMD, while ~8% have one TMD, but the occurrence of TMD indicates that a “no TMD” criterion leads to the underestimation of the phytoplasma effectoromes.

Table 3.2 Analyses of phytoplasma effectors features.

Characteristics	Number of Effectors	% of the Total *
SP **	58	90.6
Nc-SecP	3 ***	4.7
Non-secreted	3 ****	4.7
0 TMD *****	59	92.2
1 TMD	5	7.8

* Considering 64 phytoplasma effector sequences in the positive dataset; ** by SignalP v4.1; *** Considering the 6 sequences which were not recognized by SignalP v4.1 but recognized by SecretomeP v2.0; **** Results which were not recognized by SignalPv4.1 neither by SecretomeP v2.0; ***** by TMHMM v2.0 on mature proteins (without signal peptide).

3.3.3 Comparison of Multiple Pipelines to Identify Phytoplasma Effectors

The most common pipelines for the identification of phytoplasma effectors use SignalP v4.1 or SignalP v5.0 and TMHMM v2.0 on SP-lacking mature proteins (Bai *et al.*, 2009; Cho *et al.*, 2020; Chung *et al.*, 2013; Music *et al.*, 2019; Wang *et al.*, 2018b). Only a few reports have modified these pipelines, to included other programs such as Phobius (Debonneville *et al.*, 2022) and SecretomeP v2.0 (Gao *et al.*, 2023). Most authors have used SignalP programs in Gram-positive mode (Bai *et al.*, 2009; Cho *et al.*, 2020; Chung *et al.*, 2013; Music *et al.*, 2019; Wang *et al.*, 2018b). Recently, (Gao *et al.*, 2023) argued that phytoplasmas have a distinct membrane composition, and they are neither in the Gram-positive group nor the Gram-negative group; these authors used both SignalP program modes to identify effectors in ‘*Ca. Phytoplasma ziziphi*’. Based on these reports, different pipelines were tested in our analyses using SignalP v4.1 or SignalP v5.0 in either Gram-positive or Gram-negative modes.

Taking into consideration that ~8% of phytoplasma effectors were found to have TMDs, and ~10% were elusive to the SignalP programs (Table 3.2), the pipelines constructed here included additional programs, Phobius and/or SecretomeP. The results of these programs were pooled and redundancies were eliminated. Table 3.3 shows the performance of pipelines with programs used in Gram-positive mode, and Table 3.4 for Gram-negative mode.

Table 3.3 Comparison of SignalP v4.1 and SignalP v5.0, in Gram-positive mode, in different pipelines for the prediction of Phytoplasma effectors.

Pipeline 1 (Signalp4.1+phobius+secretomeP2.0+TMHMM2.0)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	61	0.93	0.89	0.89	0.91	0.10	0.91
Negative dataset	64	7						
Pipeline 2 (Signalp4.1+phobius+secretomeP2.0+TMHMM2.0+BLASTP+ elimination of false positive)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	64	1	1	1	1	0	1
Negative dataset	64	0						
Pipeline 3 (Signalp4.1+phobius+TMHMM2.0)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	59	0.92	0.92	0.92	0.92	0.07	0.92
Negative dataset	64	5						

Pipeline 4 (Signalp4.1+phobius+TMHMM2.0+BLASTP+ elimination of false positive)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	64	1	1	1	1	0	1
Negative dataset	64	0						
Pipeline 5 (Signalp5.0+phobius+secretomeP2.0+TMHMM2.0)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	55	0.85	0.95	0.94	0.90	0.04	0.90
Negative dataset	64	3						
Pipeline 6 (Signalp5.0+phobius+secretomeP2.0+TMHMM2.0+BLASTP+ elimination of false positive)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	64	1	1	1	1	0	1
Negative dataset	64	0						
Pipeline 7 (Signalp5.0+phobius+TMHMM2.0)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	52	0.81	0.98	0.98	0.89	0.01	0.88
Negative dataset	64	1						
Pipeline 8 (Signalp5.0+phobius+TMHMM2.0+BLASTP+ elimination of false positive)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	64	1	1	1	1	0	1
Negative dataset	64	0						

BlastP = [BlastP with the positive dataset constructed in this work (results added to positive predictions); Blast of false positives or suspected false positives against a database with annotations of known effectors (hits are included); Blast of false positives or suspected false positives against a database with annotations of metabolic essential activities (hits are discarded)]. Sen: Sensitivity; Spe: Specificity; PPV: Positive Predictive Value; ACC: Accuracy; FPR: False positive rate; F1 score: Measure of the success of binary classifier (score reaches its best value at 1, and worst score at 0).

Table 3.4 Comparison of SignalP v4.1 and SignalP v5.0, in Gram-negative mode, in different pipelines for the prediction of Phytoplasma effectors.

Pipeline 1 (Signalp4.1+phobius+secretomeP2.0+TMHMM2.0)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	60	0.93	0.93	0.93	0.93	0.06	0.93
Negative dataset	64	4						
Pipeline 2 (Signalp4.1+phobius+secretomeP2.0+TMHMM2.0+BLASTP+ elimination of false positive)								
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Positive dataset	64	64	1	1	1	1	0	1
Negative dataset	64	0						

Cont...Table 3.4

Content: Table 3.4

Pipeline 3 (Signalp4.1+phobius+TMHMM2.0)									
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score	
Positive dataset	64	56	0.87	0.96	0.96	0.92	0.03	0.0.91	
Negative dataset	64	2							
Pipeline 4 (Signalp4.1+phobius+TMHMM2.0+BLASTP+ elimination of false positive)									
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score	
Positive dataset	64	64	1	1	1	1	0	1	
Negative dataset	64	0							
Pipeline 5 (Signalp5.0+phobius+secretomeP2.0+TMHMM2.0)									
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score	
Positive dataset	64	49	0.76	0.95	0.94	0.85	0.04	0.84	
Negative dataset	64	3							
Pipeline 6 (Signalp5.0+phobius+secretomeP2.0+TMHMM2.0+BLASTP+ elimination of false positive)									
Control set	Proteins	Prediction	Sen/Rec	Spe	PPV/Prec	ACC	FPR	F1 score	
Positive dataset	64	64	1	1	1	1	0	1	
Negative dataset	64	0							
Pipeline 7 (Signalp5.0+phobius+TMHMM2.0)									
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score	
Positive dataset	64	45	0.70	0.98	0.97	0.84	0.01	0.81	
Negative dataset	64	1							
Pipeline 8 (Signalp5.0+phobius+TMHMM2.0+BLASTP+ elimination of false positive)									
Control set	Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score	
Positive dataset	64	64	1	1	1	1	0	1	
Negative dataset	64	0							

BlastP = [BlastP with the positive dataset constructed in this work (results added to positive predictions); Blast of false positives or suspected false positives against a database with annotations of known effectors (hits are included); Blast of false positives or suspected false positives against a database with annotations of metabolic essential activities (hits are discarded)]. Sen: Sensitivity; Spe: Specificity; PPV: Positive Predictive Value; ACC: Accuracy; FPR: False positive rate; F1 score: Measure of the success of binary classifier (score reaches its best value at 1, and worst score at 0).

The pipelines were constructed as follows:

1. Signalp4.1 + phobius + secretomeP2.0 + TMHMM2.0
2. Signalp4.1 + phobius + secretomeP2.0 + TMHMM2.0 + BLASTP+ elimination of false positive
3. Signalp4.1 + phobius + TMHMM2.0

4. Signalp4.1 + phobius + TMHMM2.0 + BLASTP+ elimination of false positive
5. Signalp5.0 + phobius + secretomeP2.0 + TMHMM2.0
6. Signalp5.0 + phobius + secretomeP2.0 + TMHMM2.0 + BLASTP+ elimination of false positives
7. Signalp5.0 + phobius + TMHMM2.0
8. Signalp5.0 + phobius + TMHMM2.0 + BLASTP+ elimination of false positives

For both classes of pipelines, either with SignalP v.4.1 or SignalP v5.0, the Gram- negative mode achieved F1 scores lower than their counterpart in the Gram-positive mode. These results indicate that the Gram-positive mode is more suitable for the prediction of phytoplasma effectors. The pipelines with SignalP + TMHMM + Phobius retrieved consistently fewer false positives (pipelines 3 and 7) compared to pipelines that included these programs plus SecretomeP v2.0 (pipelines 1 and 5), but also excluded more effectors in comparison with the pipelines that include SecretomeP (see pipelines 3 and 7 and compare with pipelines 1 and 5). These results support pipelines 1 and 5 to continue the analysis. Pipelines 1 and 5 differ in the version of SignalP in use. The pipeline that includes SignalP v4.1 retrieved 16 more false positives (7) than retrieved by SignalP v5.0 (3), but the pipeline with SignalP v5.0 excluded 9 effectors (false negatives) while the pipeline with SignalP v4.1 excluded 3; in other words, pipeline 1 is able to retrieve 6 effectors that pipeline 5 could not (Table 3.3). The balance of false positives and false negatives lead to an F1 score of 0.91 for pipeline 1, and an F1 score of 0.90 for pipeline 5.

To improve the performance of pipeline 1, three databases were independently linked for the effector identification: (A) the positive dataset (Table S1), which was compiled with the results of the Blastp analysis and contains the potential homologs of known effectors; these results were included as effector candidates. (B) a list of annotations/functional descriptors of known effectors (Table S3); this step provides additional supports for effector candidates (reinforces true positives). When some predictors fail to identify “true positive” effectors, the outcome is “false negatives” for that predictor. (C) a list of 40 annotations/descriptors of essential metabolic activities (the list of these annotations was obtained from KEGG) (Table S2). This step allows for the exclusion of false positives.

3.3.4 PhyEffector Pipeline

Based on the performance of the pipelines analyzed above to identify phytoplasma effectors, the PhyEffector pipeline was constructed with SignalP v4.1, TMHMM v2.0, Phobius and SecretomeP v2.0. Their results were pooled, redundancies were eliminated, and the candidates were subsequently added to the results from Blastp using the positive dataset as query. This second list of effector candidates was converted to FASTA format and submitted to Blastp against the non-redundant protein database at GenBank to obtain homology- based descriptions of all the retrieved hits. Their descriptions were then compared with the list of functional descriptions of known phytoplasma effectors (Table S3). Those hits that did not have any effector-related annotations were then further compared with the list of functional annotations of metabolic essential activities (proteins that correspond to phytoplasmas' core proteome); those which matched with descriptions in this set were discarded. Figure 3.1 shows the complete PhyEffector workflow.

3.3.5 PhyEffector Performance: Prediction of Effectors on a Testing Dataset and on Phytoplasma Genomes and Comparison with Literature

A second validation of the different pipelines was conducted on a testing dataset composed of a different set of 11 validated effectors, along with 192 proteins annotated in UNIPROT as "*Candidatus* effector" or "phytoplasma effector" and 23 homologs of the proteins comprising the positive dataset to give 226 proteins in total (Table S4). These protein sequences share ~50% identity with each other in the testing dataset (except for effector families which are largely conserved) and ~50% identity with protein sequences from the positive dataset. Those annotated as "effectors" are largely divergent from each other and share no identity with the positive dataset. A second negative dataset, comprising 226 core proteins (Table S5), different from the first negative dataset, was constructed as well. Validation of PhyEffector on these sets calculates a realistic F1 score since positive and testing datasets comprise different proteins with low identity among them. Table 3.5 shows the performance of PhyEffector on the testing dataset, as well as the comparison with the other pipelines, to re-analyze whether pipeline # 2 (denominated PhyEffector) is the best predictor for phytoplasma effectors. The F1 score of PhyEffector was 0.90 and it was higher

than in all the other pipelines, reinforcing that this pipeline is the best option for further analyses.

Table 3.5 Prediction of Phytoplasma effectors using different pipelines on a testing dataset.

Pipeline 1 (Signalp4.1+phobius+secretomeP2.0+TMHMM2.0)								
Set	Num. Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Testing set	226	181	0.80	0.84	0.83	0.82	0.15	0.81
Negative set	226	35						
Pipeline 2 (Signalp4.1+phobius+secretomeP2.0+TMHMM2.0+BLASTP+ elimination of false positive)								
Set	Num. Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Testing set	226	189	0.83	0.99	0.99	0.91	0.004	0.90
Negative set	226	1						
Pipeline 3 (Signalp4.1+phobius+TMHMM2.0)								
Set	Num. Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Testing set	226	158	0.69	0.94	0.92	0.82	0.05	0.79
Negative set	226	13						
Pipeline 4 (Signalp4.1+phobius+TMHMM2.0+BLASTP+ elimination of false positive)								
Set	Num. Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Testing set	226	172	0.76	0.99	0.98	0.87	0.008	0.86
Negative set	226	1						
Pipeline 5 (Signalp5.0+phobius+secretomeP2.0+TMHMM2.0)								
Set	Num. Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Testing set	226	127	0.56	0.87	0.81	0.71	0.12	0.66
Negative set	226	29						
Pipeline 6 (Signalp5.0+phobius+secretomeP2.0+TMHMM2.0+BLASTP+ elimination of false positive)								
Set	Num. Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Testing set	226	138	0.61	0.98	0.97	0.79	0.01	0.75
Negative set	226	1						
Pipeline 7 (Signalp5.0+phobius+TMHMM2.0)								
Set	Num. Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Testing set	226	89	0.39	0.93	0.86	0.66	0.06	0.54
Negative set	226	12						
Pipeline 8 (Signalp5.0+phobius+TMHMM2.0+BLASTP+ elimination of false positive)								
Set	Num. Proteins	Prediction	Sen	Spe	PPV	ACC	FPR	F1 score
Testing set	226	106	0.46	0.99	0.98	0.73	0.008	0.63
Negative set	226	2						

Pipelines 1–4, SignalP is in Gram-positive mode; Pipelines 5–8, SignalP is in Gram-negative mode. BlastP = [BlastP with the positive dataset constructed in this work (results added to positive predictions); Blast of false positives or suspected false positives against a database with annotations of known effectors (hits are included); Blast of false positives or suspected false positives against a

database with annotations of metabolic essential activities (hits are discarded)]. Sen: Sensitivity; Spe: Specificity; PPV: Positive Predictive Value; ACC: Accuracy; FPR: False positive rate; F1 score: Measure of the success of binary classifier (score reaches its best value at 1, and worst score at 0).

In a third analysis for the validation of PhyEffector, the algorithm was used to identify effectors in phytoplasma genomes (i.e., deduced proteomes), and the results were compared with corresponding scientific reports. In order to solve discrepancies between the prediction of PhyEffector and the literature, and to determine which prediction is stronger, false negatives (FN) were identified comparing the annotations of the predicted effectors with the descriptions of known effectors; those that coincide with “true effector” annotations were considered “true”, and those that were not recognized (i.e., lacking effector-associated annotations) were considered “FN”. False positives (FP) were identified with a similar strategy, but instead, the effector results were compared with functional annotations of essential (core) proteins; those that coincided with essential proteins were considered “false positives”.

The number of predicted effectors by PhyEffector was highly variable, from 41 in the *Flavescence dorée* phytoplasma, to 97 effector candidates in ‘*Ca. Phytoplasma asteris*’ DY2014. In all cases, PhyEffector predicted a higher number of effector candidates than the scientific reports (Table 3.6). Since the prediction by PhyEffector included the FP elimination step, the number of FP was zero in all predictions, while FN were less than 10 in general, except in ‘*Ca. Phytoplasma asteris*’ (AY-WB), where PhyEffector had 17 FN. However, the predictions from the corresponding literature had higher numbers of FN and higher numbers of FP. Based on the results obtained from the various phytoplasma genomes, the F1 score for PhyEffector ranked from 0.8957 to 1.0, with an average value of 0.9783. These results indicate that the PhyEffector is a robust algorithm to identify effector candidates in phytoplasmas.

For this evaluation, the proteins that were not retrieved by the pipelines from the “Testing set” were considered as “false negatives” for F1 score determination. However, these proteins are actually “ambiguous” since they may be false positives and the PhyEffector’ F1 score may be higher for each pipeline in Table 3.5.

Table 3.6 Effectors prediction by PhyEffector on phytoplasma genomes.

Phytoplasma	Effectors predicted by the authors	Pipeline used for effector prediction*	Reference	PhyEffector prediction	Shared candidates	Unshared candidates	False negatives	False positives	F1 score***
'Ca. Phytoplasma mali'	31	SignalP v4.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Chung <i>et al.</i> , 2013)	49	18	A=13 P= 31	A=31 P=6	A=7 P=0	0.9423
'Ca. Phytoplasma australiense'	61	SignalP v4.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Chung <i>et al.</i> , 2013)	89	43	A=18 P= 46	A=46 P=10	A=8 P=0	0.9518
'Ca. Phytoplasma asteris' (AY-WB)	58	SignalP v4.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Chung <i>et al.</i> , 2013)	73	35	A= 23 P=38	A=38 P=17	A=6 P=0	0.8957
'Ca. Phytoplasma asteris' OY-M	65	SignalP v4.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Chung <i>et al.</i> , 2013)	85	54	A= 11 P=50	A=50 P=7	A=4 P=0	0.9674
'Ca. Phytoplasma solani' strain SA-1	38	SignalP v3.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Music <i>et al.</i> , 2019)	96	26	A=12 P=83	A=83 P=4	A=8 P=0	0.9819
'Ca. Phytoplasma asteris' AYWB	33	SignalP v3.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	73	23	A= 10 P=40	A=40 P=5	A=5 P=0	0.9668
'Ca. Phytoplasma. asteris' NJAY	23	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	95	18	A= 5 P=77	A=77 P=0	A=5 P=0	1
'Ca. Phytoplasma asteris' WEID	17	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	64	11	A= 6 P=53	A=53 P=1	A=5 P=0	0.9922
'Ca. Phytoplasma asteris' OY-M	37	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	84	15	A= 22 P= 69	A=84 P=5	A=17 P=0	0.9710

'Ca. Phytoplasma asteris' OY-V	36	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	87	24	A=12 P=63	A=129 P=4	A=8 P=0	0.9870
'Ca. Phytoplasma asteris' DY2014	45	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	97	31	A= 14 P=52	A=52 P=2	A=12 P=0	0.9944
'Ca. Phytoplasma asteris' MBP-M3	13	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	64	9	A= 4 P= 51	A=56 P=0	A=4 P=0	1
'Ca. Phytoplasma asteris' De Villa	10	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	55	5	A= 5 P= 50	A=50 P=1	A=4 P=0	0.9909
'Ca. Phytoplasma asteris' LD1	14	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	60	9	A= 5 P= 46	A=56 P=1	A=4 P=0	0.9923
'Ca. Phytoplasma asteris' CYP	21	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	91	14	A= 7 P= 77	A=77 P=0	A=7 P=0	1
'Ca. Phytoplasma asteris' TW1	19	SignalP v5.0, for SP, and then TMHMM v2.0 on mature protein sequence without the SP	(Cho <i>et al.</i> , 2020)	57	14	A= 5 P= 38	A=51 P=2	A=3 P=0	0.9827
'Ca. Phytoplasma hytoplasma aurantifolia'	98	Comparison of SignalP v4.1 and SignalP v5.0; the former retrieved ~ 70 false positives.	(Tan <i>et al.</i> , 2021)	93	53	A= 45 P=40	A=40 P=8	A=37 P=0	0.9587
'Ca. Phytoplasma aurantifolia'	27**	Comparison of SignalP v4.1 and SignalP v5.0; the former retrieved ~ 70 false positives.	(Tan <i>et al.</i> , 2021)	93	20	A= 7 P=73	A=73 P=5	A=2 P=0	0.9738
'Ca. Phytoplasma vitis' (Flavescence dorée)	17	SignalP v5.0 and Phobius. Effectors with transmembrane domains (TMDs) were also identified.	(Debonneville <i>et al.</i> , 2022)	41	6	A= 11 P= 35	A=35 P=2	A=9 P=0	0.9791
'Ca. Phytoplasma	8 (Zaofeng1 to Zaofeng8).	Signal peptide by SignalP4.1 and TMDs by the TMHMM 2.0. Potential mobile units (PMUs) were identified by the	(Chen <i>et al.</i> , 2022)	87	5	A= 3 P=82	A=89 P=1	A=2 P=0	0.9942

ziziphi' (Jujube witches'-broom Phytoplasma)	presence of flanking tra5 insertion sequences and DNA replication genes (dnaG, dnaB, ssb, tmk). Secreted proteins harbored in PMUs were identified as JWB phytoplasma putative effectors.
--	--

* SignalP v3.0, v4.0, v5.0 were used on the Gram-positive bacteria mode. A = authors; P = PhyEffector. ** Authors recognized this number as the correct prediction. *** F1 score calculated using the effector prediction by PhyEffector in each genome.

3.4 DISCUSSION

Effectors are essential for virulence and are therefore susceptible targets for the control of phytoplasma-associated diseases (Galetto *et al.*, 2021; Tomkins *et al.*, 2018). It is necessary to have a robust computational tool that specifically identifies these effector candidates to accelerate phytoplasma Effectoromics.

A positive database (phytoplasma effectors) was constructed by searching with the keywords such as “Phytoplasma effectors”, “SAP effector”, etc. The first attempt was the search in the GenBank database, but it resulted in highly redundant data. Better results were obtained in the UNIPROT database, but manual revision was also necessary. Finally, the positive dataset comprised 64 non-redundant proteins. Difficulties involved in the prediction of effectors include the small number of true (validated) currently known effectors, and the number of potential false positives in the public databases. We found that many proteins annotated in UNIPROT as “phytoplasma effectors” and “*Candidatus* effectors” correspond to “hypothetical proteins” in GenBank. Although many hypothetical proteins are expected to be effectors, it cannot be ruled out that false positives are among them. Here, care was taken and we did not include proteins annotated as “effectors” (phytoplasma effectors or *Candidatus* effectors) according to UNIPROT, which have no homology to known validated effectors and have no experimental validation either, in the positive dataset. However, they were included in another

dataset to test the algorithm. Those from the testing dataset that were not retrieved by PhyEffector were counted as false negatives, although they are ambiguous because it is not possible to classify them as effectors or non-effectors at this time.

We reasoned that, if the positive dataset comprises phytoplasma effectors, which are dispensable proteins, i.e., most of them are not ubiquitous in all phytoplasmas, and have a patchy phylogenetic distribution (Dillon *et al.*, 2019; Morgado y Vicente, 2022), the negative dataset must comprise indispensable proteins from phytoplasmas. Indispensable proteins are widely distributed in phytoplasmas, and are related to essential metabolic activities (part of the core proteome). (Music *et al.*, 2019) compared the proteomes of phytoplasmas from 16SrI, 16SrXII, and 16SrXVI and found 259 orthologous gene clusters shared among them. Recently, (Cho *et al.*, 2020) compared 11 genomes of 16SrI-phytoplasmas and found 303 single-copy genes shared among all of them. This list is representative of the core proteome in phytoplasmas and the size is similar to the extended negative dataset constructed in the present study.

Usually, phytoplasma effectors are described as secreted proteins, but recent reports have uncovered the existence of phytoplasma effectors having transmembrane domains (TMDs) or being secreted by a non-classical pathway (Gao *et al.*, 2023). Analyzing the positive dataset (64 phytoplasma effector proteins) showed that these proteins are more common than previously believed: ~8% of the phytoplasma effectors have TMDs and ~5% are non-secreted. Based on these findings, different pipelines comprising different predicted programs were constructed and evaluated.

Gram-positive bacteria are surrounded by a thick peptidoglycan cell wall, while Gram-negative bacteria have a much thinner peptidoglycan cell wall with an outer membrane surrounding the cell containing lipopolysaccharides (Mai-Prochnow *et al.*, 2016). Phytoplasmas strictly are neither Gram-positive nor Gram-negative in terms of their membrane composition (Bai *et al.*, 2009), but they are closely related to the non-sterol-requiring achleoplasmas (Kerr y Gibb, 1997). To choose the best option for phytoplasmas, SignalP programs were tested both in Gram-positive and Gram-negative modes on the positive dataset; better retrieval of phytoplasma effectors was achieved in the Gram-positive mode. This is consistent with phytoplasmas' phylogeny, since they belong to the Mollicutes class—cell wall-less microorganisms derived from a *Bacillus/Clostridium*-like ancestor (Cao

et al., 2020; Francis *et al.*, 2010). In other words, phytoplasmas are derived from Gram-positive bacteria but they lack the cell wall.

The mechanisms for delivering effectors are also different between Gram-negative and Gram-positive bacteria. Gram-negative bacteria have type III, type IV and type VI secretion systems that form hollow tubes through which effectors are directly translocated from the bacterial cytosol directly into the cytosol of host cells (Meir *et al.*, 2023). In the case of Gram-positive bacteria, they predominantly use the Sec-dependent pathway for effectors delivery (Hogenhout and Loria, 2008). Therefore, the SignalP program is suitable for the identification of phytoplasma effectors, and different versions of this program have been used (Bai *et al.*, 2009; Cho *et al.*, 2020; Chung *et al.*, 2013). Here, a comparison of pipelines only differing in the version 4.1 or 5.0 of SignalP showed that specificity (proportion of negatives that were correctly identified) of SignalP v5.0 was higher than that of SignalP v4.1, but sensitivity of SignalP v5.0 (proportion of positives that are correctly identified) was lower; simply put, v5.0 excluded more true effectors. Similar results were found by (Tan *et al.*, 2021) with SignalP v5.0 identifying less false positives in '*Ca. Phytoplasma aurantifolia*' than SignalP v4.1, but also SignalP v5.0 identified a smaller number of effectors. In order to distinguish the performance of the pipelines evaluated here, the F1 scores (these measure the success of the pipelines; best value at 1 and the worst score at 0) were calculated. F1 scores were higher for pipelines that harbor SignalP v4.1 (Tables 3.3 and 3.4), supporting this version for the search for effectors in the phytoplasma proteomes. However, none of the pipelines was able to retrieve all sequences from the positive dataset. The best pipeline (pipeline 1, indicated in Table 3.3 with a F1 score of 0.91) excluded three true effectors. We reasoned that the best pipeline should be able to recover all or most of the effectors, even if a large number of false positives are initially recovered, followed by the elimination of the false positives present.

During the *in silico* characterization of the effectors, we realized that some of them were identified by homology because selection based on common characteristics (having signal peptide and not having TMD) would exclude them. For example, protein ID A0A859I9H9 (GenBank QKX95313.1) from the rapeseed phyllody phytoplasma is homologous to the effector SAP01 from the phytoplasma AY-WB, but the mature protein (after *in silico* removal of the signal peptide) has a TMD according to TMHMM. Therefore, to improve the pipeline's sensitivity, a Blastp step was included using the phytoplasma effectors' positive dataset as query. Subsequently, the functional annotation of known effectors and essential proteins

was used to distinguish effector candidates from false positives. The hits with descriptors that match with annotations of essential proteins were ruled out.

The search for effectors based on homology was also used in the prediction of fungal and oomycete effectors by WideEffHunter and demonstrated to improve its accuracy on real tests beyond positive control (Carreón-Anguiano *et al.*, 2022).

This strategy showed no significant differences in the identification of effectors on the positive dataset (Tables 3.3 and 3.4), but the combination of the different programs used in the different reports, including SignalP v4.1, TMHMM v2.0, Phobius and SecretomeP, along with the search for homologous sequences of known effectors, significantly expanded the number of effector candidates retrieved in the testing dataset (see in Table 3.5, pipeline 2 vs. pipeline 1; pipeline 4 vs. pipeline 3; pipeline 6 vs. pipeline 5; and pipeline 8 vs. pipeline 7). Later, the last step of the pipeline specifically eliminated FP, resulting in higher F1 scores in all cases of pipelines that include the steps of Blastp and elimination of FP. The pipeline # 2 was chosen as the best predictor and it was named “PhyEffector”. To challenge the pipeline, “PhyEffector” was applied on a number of deduced proteomes of phytoplasmas. PhyEffector identified double to triple the number of effector candidates identified by the authors of previous phytoplasma-related reports. The analyses showed that PhyEffector identified zero FP and a low number (<10) of FN, while both parameters were greater numbers in all previous reports. The F1 score for PhyEffector ranked from 0.8957 to 1, with an average value of 0.9783, evidencing that this is a robust predictor of phytoplasma effectors. The results from PhyEffector indicate that phytoplasma effectoromes have been underestimated. Larger phytoplasma effectoromes were predicted than previously reported, and these large sizes are congruent with the fact that many phytoplasma proteins are annotated as “hypothetical protein”; for example, 257 hypothetical proteins in ‘*Ca. Phytoplasma solani*’ SA-1, and 337 in ‘*Ca. Phytoplasma australiense*’ PAa (Music *et al.*, 2019), and 81 in Peanut Witches’-Broom Phytoplasma (Chung *et al.*, 2013). Hypothetical proteins are not part of the core proteins because core proteins are conserved and have known essential functions. Therefore, some or all of these hypothetical proteins may be part of phytoplasma effectoromes (Bai *et al.*, 2009; Gao *et al.*, 2023). PhyEffector demonstrated the ability to identify these novel effectors at the genome level, retrieving not only classical effector candidates but also the non-classical ones.

It is important to emphasize that PhyEffector does not rank effectors based on probability. Researchers may prioritize effectors according to their own interests. Conventional predictions pay attention to homologs of known effector families (Tan *et al.*, 2021), or to genomic location (in PMUs) (Chung *et al.*, 2013; Huang *et al.*, 2022) or high expression during host or insect vector infection (Anabestani *et al.*, 2017; MacLean *et al.*, 2011). Other researchers look for novel effectors, choosing candidates with non-canonical characteristics (Gao *et al.*, 2023; Zhang *et al.*, 2019). Regardless of the type of phytoplasma effectors that the researcher is interested in, the PhyEffector algorithm may be useful to them. PhyEffector may be found at <https://github.com/Gisel-Carreon/PhyEffector>. Researchers are encouraged to experimentally validate the chosen effector candidates after in silico prediction, especially the novel non-classical ones.

CAPÍTULO IV

SECUENCIACIÓN PARCIAL DEL GENOMA DE 'Ca. P. palmae', FITOPLASMA ASOCIADO AL ALC

Autores:

Sara Elena Vila-Luna⁺, Karla Gisel Carreón-Anguiano⁺, Iván Córdova-Lara, Luis Sáenz-Carbonell, Carlos Oropeza Salín y Blondy Canto-Canché*

⁺ Estos autores contribuyeron igualmente en este trabajo

Cita: Vila-Luna SE, Carreón-Anguiano KG, Córdova-Lara I, Sáenz-Carbonell L, Oropeza-Salín C, Canto-Canché B. (2025). The draft genome of 'Candidatus Phytoplasma palmae' strain LY-C2, the phytoplasma associated with coconut lethal yellowing disease, reveals insights into its biological characteristics. *World J Microbiol Biotechnol.* 41(7):242. <https://doi.org/10.1007/s11274-025-04418-3>

4.1 INTRODUCCIÓN

Los fitoplasmas son fitopatógenos que infectan al menos 700 especies vegetales alrededor del mundo, lo que provoca pérdidas significativas de producción en cultivos de importancia comercial (Jomantiene *et al.*, 2007). La palma de coco (*Cocos nucifera*) es una especie vegetal de inmenso valor económico a nivel mundial, conocida como el "Árbol de la Vida" porque cada una de sus partes puede utilizarse para la obtención de alimentos, productos de higiene, cosméticos, construcción y artesanías (Henrietta *et al.*, 2022). Desafortunadamente, la palma de coco es susceptible a la enfermedad del amarillamiento letal (ALC), lo que provoca pérdidas de producción en países donde se cultiva la palma de coco (Córdova *et al.*, 2019). Esta enfermedad está vinculada a 'Candidatus Phytoplasma palmae', miembro del grupo 16SrIV-A transmitida por *Haplaxius crudus*, que se alimenta del floema de la palma de coco (Córdova *et al.*, 2019). Existen otras cepas que afectan a más de 30 especies de palmeras, incluidas 16SrIV-B, -C, -D, -E y -F. 16SrIV-B afecta a *Acrocomia aculeata* en Honduras, 16SrIV-C se ha informado en cocoteros en África y la cepa 16SrIV-D es predominante en México y afecta a *Adonidia merrillii*, *Carludovica palmata*, *Sabal mexicana* y *Pseudophoenix sargentii*. El 16SrIV-E afecta a las palmas de coco en la República Dominicana y el 16SrIV-F afecta a *Washingtonia robusta* en EE. UU. (Oropeza *et al.*, 2020). Los fitoplasmas son patógenos que secretan proteínas conocidas

como “efectores”, que son pequeñas moléculas que suprimen las defensas de la planta o atraen insectos vectores (Sugio *et al.*, 2011a; Tomkins *et al.*, 2018). Los estudios genómicos están ayudando en la predicción de efectores de fitoplasmas; por ejemplo, el genoma de ‘*Ca. P. asteris*’ agente causal de la escoba de bruja del amarillamiento del aster (AY-WB), contiene 56 efectores (Bai *et al.*, 2009), mientras que se predijeron 30 efectores en el genoma del fitoplasma del amarillamiento de la cebolla (OY-M) (Hoshi *et al.*, 2009). Actualmente, 21 efectores de fitoplasmas han sido validados experimentalmente (Wang *et al.*, 2023; Chen *et al.*, 2022; Arricau-Bouvery *et al.*, 2021; Huang *et al.*, 2021; Wang *et al.*, 2017; MacLean *et al.*, 2014; Maejima *et al.*, 2014; Sugio *et al.*, 2014; Boonrod *et al.*, 2012; Hoshi *et al.*, 2009; Kakizawa *et al.*, 2009). Estos efectores desempeñan un papel crucial en la patogenicidad y son responsables de los síntomas característicos inducidos por los fitoplasmas, como la escoba de bruja, la filodia, el enanismo o la esterilidad (Bai *et al.*, 2009; Hoshi *et al.*, 2009).

Se han descrito varios genomas de fitoplasmas secuenciados completamente; once de ellos se seleccionaron para compararlos con ‘*Ca. P. palmae*’-ALC. De acuerdo con el gen 16S ARNr, cuatro genomas pertenecientes a fitoplasmas del grupo I: ‘*Ca. P. asteris*’ responsable del amarillamiento de la cebolla (OY-M), del grupo 16Srl-B; ‘*Ca. P. asteris*’ del amarillamiento del aster escoba de bruja (AY-WB), del grupo 16Srl-A; ‘*Ca. P. asteris*’ del fitoplasma del enanismo arbustivo del maíz (MBSP M3), del grupo 16Srl; y ‘*Ca. P. asteris*’ De Villa de *Catharanthus roseus*, del grupo 16Srl; uno pertenece al grupo II: ‘*Ca. P. aurantifolia*’ del fitoplasmas de escoba de bruja de la *Echinacea purpurea* (EpWB), del grupo 16SrlI; dos pertenecen al grupo V: ‘*Ca. P. ziziphi*’ responsable de la escoba de bruja del jujube o dátíl rojo (JWB), del grupo 16SrV-B y ‘*Ca. P. vitis*’, fitoplasma de la flavescencia dorada (FDp), del grupo 16SrV-D; uno perteneciente al grupo VIII: ‘*Ca. P. luffae*’ de la enfermedad de la escoba de bruja de la lufa (*Luffa aegyptiaca*), del grupo 16SrVIII; uno correspondiente al grupo X: ‘*Ca. P. mali*’ de proliferación del manzano (AP), grupo 16SrX; y dos pertenecen al grupo XII: ‘*Ca. P. australiense*’ del amarillamiento letal de la fresa (SLY), grupo 16SrXII; y ‘*Ca. P. australiense*’ PAa del amarillamiento de la vid australiana (AUSGY), del grupo 16SrXII-B (Böhm *et al.*, 2023; Debonneville *et al.*, 2022; Huang *et al.*, 2022; Tan *et al.*, 2021; Coetzee *et al.*, 2019; Wang *et al.*, 2018a; Orlovskis *et al.*, 2017; Andersen *et al.*, 2013; Kube *et al.*, 2008; Tran-Nguyen *et al.*, 2008; Bai *et al.*, 2006; Oshima *et al.*, 2004). Los tamaños del genoma varían desde aproximadamente 570 Kb de ‘*Ca. Phytoplasma asteris*’ del MBSP (16Srl) a aproximadamente 950 Kb de ‘*Ca. P. australiense*’ SLY (16SrXII)

(Orlovskis *et al.*, 2017; Andersen *et al.*, 2013). El genoma del fitoplasma 'Ca. P. palmae'-TPPD (cepa ACPD, fitoplasma del Declive de la palma Phoenix de Texas), del grupo 16SrIV-D, está disponible en GenBank (GCA_005774685.2), pero, hasta donde sabemos, no existen manuscritos que describan dicho genoma de fitoplasma 16SrIV. Este genoma también se incluyó para su comparación con el genoma de 'Ca. P. palmae'-ALC.

En este estudio, se presenta el genoma parcial de 'Ca. P. palmae'-ALC, que tiene un tamaño de 434 Kb. El análisis de identidad de nucleótidos promedio (ANI) reveló una identidad del 90 % con el fitoplasma del declive de la palma Phoenix de Texas cepa ACPD ('Ca. P. palmae'-TPPD) y del 75-77 % con fitoplasmas de otros grupos 16Sr. Se identificaron 54 proteínas efectoras en el fitoplasma 'Ca. P. palmae'-ALC, con 23 efectores específicos de este fitoplasma y 11 que compartían homología únicamente con efectores del fitoplasma del Declive de la palma Phoenix de Texas. También se identificaron dos unidades potencialmente móviles (PMU).

Mientras que, en muchos fitoplasmas, los efectores se localizan típicamente dentro de las regiones PMU, en el fitoplasma 'Ca. P. palmae'-ALC se encontraron principalmente fuera de las PMUs, en regiones genómicas sin homología con los genomas de otros fitoplasmas. Este reporte sobre el genoma de 'Ca. P. palmae'-ALC mejora nuestra comprensión de la organización y evolución de los fitoplasmas.

4.2 MATERIALES Y MÉTODOS

4.2.1. MATERIAL VEGETAL Y EXTRACCIÓN DE ADN

Las muestras de tallo de palmas de coco (*C. nucifera*) con síntomas de amarillamiento letal se colectaron en Telchac Puerto, Yucatán, México (coordenadas: 21°20'26"N 89°15'47"O), de acuerdo con Córdova *et al.* (2019). La extracción de ADN se realizó con el kit de aislamiento de ADN Plant/Fungi (Norgen Biotek™). La presencia de fitoplasmas se confirmó mediante la prueba de PCR en tiempo real (qPCR), usando la sonda TaqMan LYGroEL, específica para el subgrupo 16Sr-IVA (Córdova *et al.*, 2019). Los ADN extraídos se cuantificaron con un espectrofotómetro NanoDrop One (Thermo Scientific™) y la integridad del ADN se evaluó mediante un gel de agarosa al 1 %.

Se utilizó el kit *Microbiome DNA Enrichment* (NEBNext®) para enriquecer el ADN de los fitoplasmas, de acuerdo con las instrucciones del fabricante. El procedimiento consiste en cuatro etapas: 1) preunión proteína-perlas magnéticas; 2) captura del ADN-metilado del hospedero; 3) colecta del ADN-enriquecido microbiano y 4) purificación del ADN-enriquecido. Posteriormente, las muestras enriquecidas con el ADN del fitoplasma se cuantificaron en el espectrofotómetro (NanoDrop One, Thermo Scientific™) y se realizaron los análisis de validación de enriquecimiento.

a) Validación mediante qPCR: Para la validación del enriquecimiento del ADN del fitoplasma en las muestras obtenidas se usó la técnica de qPCR, empleando el ensayo de la sonda TaqMan LYGroEL (Córdova *et al.*, 2019). Como control se utilizó 100 ng del ADN de la muestra inicial, no enriquecida. Adicionalmente, se analizaron las muestras empleando los cebadores (primers) universales de control de bacterias del gen 16S ARNr que trae el kit Microbiome DNA Enrichment (NEBNext®).

Para verificar que las muestras se limpiaron del ADN-hospedero se analizaron también mediante qPCR utilizando el ensayo de la sonda TaqMan 18S, para amplificar el ADN de la planta. Como control se utilizó 100 ng del ADN de la muestra inicial, no enriquecida.

Las condiciones utilizadas para qPCR fueron las siguientes: Primer ciclo a 50°C por 2 minutos (1 vez), segundo ciclo a 95°C por 10 minutos (1 vez), tercer ciclo a 95°C por 15 segundos y 61°C por 60 segundos (35 veces).

Resultados esperados en la validación positiva (qPCR): en la fracción microbiana se debe observar poco o ningún cambio en el valor de Ct con el ensayo TaqMan LYGroEL (específicas para la enfermedad de AL, Córdova *et al.*, 2019) al comparar las muestras de antes y después del enriquecimiento. También debe observarse un cambio significativo en los Cts (aumento) entre la fracción enriquecida y no enriquecida con los cebadores 18S que amplifican regiones de la planta. Estos resultados corroborarían que no se perdió ADN del fitoplasma (ensayo TaqMan LY16S), pero sí se eliminó ADN del hospedante vegetal eucariota (ensayo TaqMan 18S).

b) Validación en gel de agarosa: Para visualizar la integridad del ADN enriquecido se preparó un gel de agarosa al 1 % con 5 µL de cada muestra (enriquecida y no enriquecida). Se espera observar una banda tenue para el ADN-enriquecido.

4.2.2. SECUENCIACIÓN Y ANOTACIÓN DEL GENOMA DE ALC

El ADN enriquecido del fitoplasma se secuenció mediante el sistema DNBSEQ™ (nanoesferas de ADN) en la plataforma de secuenciación MGISEQ-2000 de LANGEBIO-CINVESTAV en Irapuato, México.

El control de calidad de las lecturas se realizó con FastQC V0.11.15 (Brown *et al.*, 2017). Los adaptadores se eliminaron con Cutadapt V4.1 (Martin, 2011) y las lecturas se recortaron con trimmomatic v0.40 (Bolger *et al.*, 2014) con PE-phred28.

Las secuencias genómicas de la palma de coco se filtraron con base en la secuencia completa concatenada del genoma del coco (GCA_008124465.1), el genoma del cloroplasto de coco (NC_022417.1) y el genoma mitocondrial de coco (KX028885.1) utilizando la herramienta hisat2 V2.0.5.

Las lecturas resultantes (10,540,293) se ensamblaron *de novo* utilizando Spades 3.15.4 (opción -meta) (Prijbelski *et al.*, 2020), Megahit 1.2.9 (Li *et al.*, 2016) y el programa PATRIC, utilizando los genomas almacenados en la base de datos PATRIC (Parrello *et al.*, 2021). Para la estimación completa del genoma se usó BUSCO v5.8.2.cv1 (Manni *et al.*, 2021), con el conjunto de datos de linaje “mollicutes_odb10”, creado el 8 de enero de 2024. También, se utilizó el servidor web MiGa (<http://microbial-genomes.org>, consultado el 20 de enero de 2025) (Rodriguez *et al.*, 2018) para evaluar la calidad del genoma. Los plásmidos se buscaron utilizando el programa Plasmer (Zhu *et al.*, 2023) con parámetros predeterminados.

La predicción génica se realizó con el programa PRODIGAL (Hyatt *et al.*, 2010), y la anotación funcional se realizó mediante una combinación de RASTtk (Brettin *et al.*, 2015), Interproscan5 (Jones *et al.*, 2014) y BLASTP (Johnson *et al.*, 2008) contra la base de datos NCBI NR. El mapeo de ontología génica (Gene ontology, GO por sus siglas en inglés) para procesos biológicos (PB), funciones moleculares (FM) y componentes celulares (CC) se realizó con la base de datos de GO y la base de datos integrada de dominios funcionales de InterProScan v.86.0 (Jones *et al.*, 2014). La reconstrucción metabólica se realizó con la base de datos KEGG (Kanehisa y Goto, 2000).

4.2.3 ANÁLISIS COMPARATIVO CON DIFERENTES GENOMAS DE FITOPLASMAS

El genoma de '*Ca. P. palmae*'-ALC se comparó con 11 genomas completos de fitoplasma reportados previamente. Estos genomas incluyen '*Ca. P. asteris*' OY-M (Oshima *et al.*, 2004) [GenBank AP006628], '*Ca. P. asteris*' AY-WB (Bai *et al.*, 2006) [GenBank CP000061], '*Ca. P. mali*' AT (Kube *et al.*, 2008) [GenBank CU469464], '*Ca. P. australiense*' PAa (Tran-Nguyen *et al.*, 2008) [GenBank AM422018], '*Ca. P. australiense*' SLY (Andersen *et al.*, 2013) [GenBank CP002548], '*Ca. P. asteris*' MBPS M3 (Orlovskis *et al.*, 2017) [GenBank CP015149], '*Ca. P. ziziphi*' JWB (Wang *et al.*, 2018a) [GenBank CP025121], '*Ca. P. asteris*' De Villa (Coetzee *et al.*, 2019) [GenBank CP035949], '*Ca. P. aurantifolia*' EpWB (Tan *et al.*, 2021) [GenBank CP040925], '*Ca. P. luffae*' (Huang *et al.*, 2022) [GenBank CP054393] y el fitoplasma de la Flavescencia dorada (Debonneville *et al.*, 2022) [GenBank CP097583]. '*Ca. P. palmae*'-TPPD cepa ACPD (GCA_005774685.2) también se incluyó en los análisis. Estos genomas se recuperaron de GenBank.

La comparación de genomas completos se realizó con FastANI v1.3.3 (Jain *et al.*, 2018) y los resultados se visualizaron mediante el servidor en línea Proksee (Grant *et al.*, 2023). BLASTN se corrió con los siguientes parámetros: E-value= 1×10^{-4} , número de alineamientos=10000, número de núcleos (threads) =8 y el parámetro *dust* habilitado. Se buscaron Unidades Móviles Potenciales (PMU) con base en ocho genes principales: *tra5*, *dnaB*, *dnaG*, *tmk*, *hflB*, *himA*, *ssb* y *rpoD*, de acuerdo con Bai *et al.* (2006).

Se analizó la conservación de la sintenia a nivel de genoma para '*Ca. P. palmae*'-ALC (fitoplasma del amarillamiento del cocotero) y '*Ca. P. palmae*'-TPPD (cepa ACPD, fitoplasma del Declive de la palma Phoenix de Texas) utilizando Mummer v3.23 (Kurtz *et al.*, 2004) con las opciones "maxmatch -mincluster 30-b". El resultado se visualizó con gnuPlot, disponible en el servidor Galaxy (<https://usegalaxy.eu/>).

4.2.4 IDENTIFICACIÓN Y CARACTERIZACIÓN COMPUTACIONAL DE EFECTORES PUTATIVOS EN '*Ca. P. palmae*'-ALC

Las proteínas efectoras putativas se identificaron utilizando el algoritmo PhyEffector (programa computacional creado en este trabajo correspondiente al artículo 2 ya publicado y el capítulo 3 de este documento. Disponible en <https://github.com/Gisel->

[Carreon/PhyEffector](#)). Los efectores predichos de 'Ca. P. palmae'-ALC se analizaron para determinar el péptido señal utilizando SignalP v4.1 (Petersen *et al.*, 2011), el dominio transmembrana (TMD) utilizando TMHMM v2.0 (Krogh *et al.*, 2001), y Phobius (Käll *et al.*, 2007). Además, los efectores se sometieron a análisis utilizando la herramienta "Maximización de expectativas múltiples para la obtención de motivos" (MEME, Multiple expectation maximization for motif elicitation) versión 5.5.6 (<https://meme-suite.org/meme/index.html>) y se compararon con las tribus de efectores de fitoplasmas basadas en MEME reportadas en este trabajo (artículo 1 publicado, capítulo 2). Además, se realizó la búsqueda de motivos *de novo* en los efectores predichos de 'Ca. P. palmae'-ALC en modo clásico, con un límite máximo de 25 motivos. Los motivos lineales cortos (SLiMs), también conocidos como ELMs, se identificaron utilizando la base de datos Eukaryotic Linear Motif (ELM) (<http://elm.eu.org/>) con la configuración predeterminada.

4.3 RESULTADOS

4.3.1 ENSAMBLAJE DEL GENOMA

Inicialmente, se obtuvo un total de 42,982,611 lecturas, que se redujeron a 10,540,293 lecturas después de eliminar las secuencias de coco. Este proceso resultó en 50 contigs y un genoma parcial de 434,453 pb (Figura 4.1), con una cobertura promedio de 24.76X y un contenido de G/C del 21.9 %; no se detectaron plásmidos. El análisis BUSCO evaluó 151 marcadores del conjunto de datos "Mollicutes", lo que resultó en 135 genes completos y 16 genes faltantes (89.4 % y 10.6 %, respectivamente). El análisis del genoma 'Ca. P. palmae'-ALC utilizando MiGa indicó una puntuación de completado del 88.7 %, clasificada como "muy alta", y una puntuación de calidad del 84.2 %, clasificada como "excelente", lo que confirma la calidad del genoma obtenido a pesar de ser parcial. El análisis de predicción de genes identificó 412 genes y 365 proteínas, de los cuales 215 se encontraron anotadas, 2 operones de ARNr adyacentes y 25 genes de ARNt.

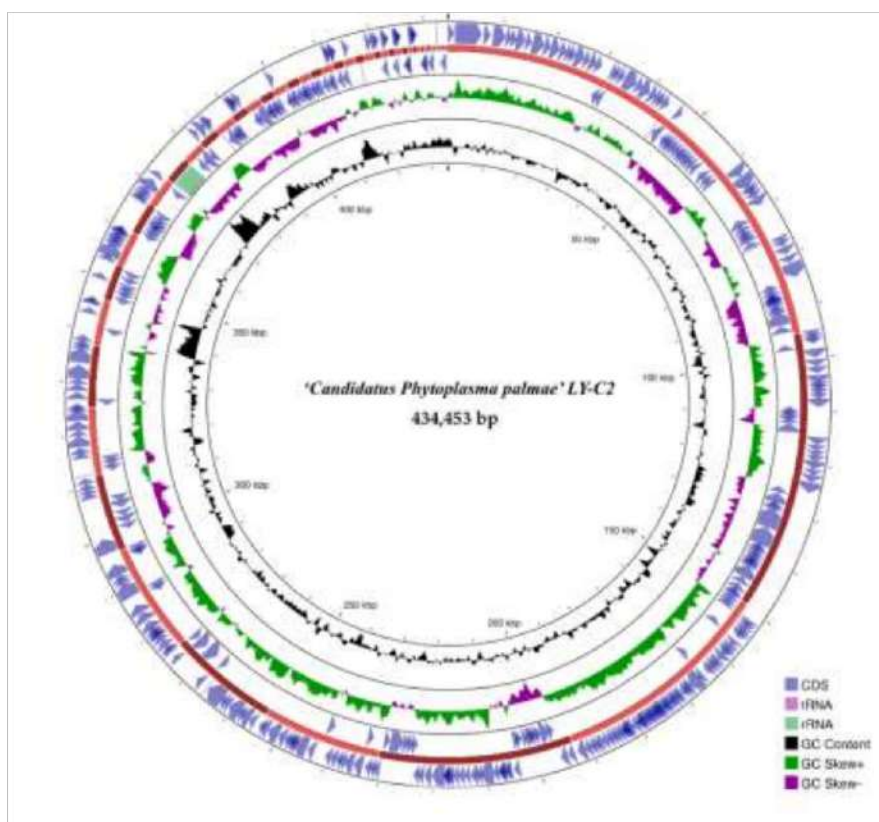


Figura 4.1 Representación circular del genoma parcial de ‘Ca. P. palmae’ fitoplasma del ALC. Se observa el genoma del fitoplasma de ALC marcados por colores las secuencias codificantes de ADN (CDS en color azul), ARN de transferencia (RNA_t en color lila), ARN ribosómico (tRNA en color azul agua marina), contenido de Guanina-Citocina (GC content en color negro), sesgo positivo GC (GCskew+ en color verde), sesgo negativo GC (GC skew- en color morado).

4.3.2 ANOTACIÓN FUNCIONAL DEL GENOMA PARCIAL DE ‘Ca. P. palmae’-ALC

Utilizando la base de datos de GO (<https://geneontology.org/>), se anotaron 342 genes y se categorizaron en tres categorías funcionales: procesos biológicos, componentes celulares y funciones moleculares. El término común fue “proceso celular” para procesos biológicos, “parte celular” para componentes celulares y “actividad catalítica” para funciones moleculares (Figura 4.2).

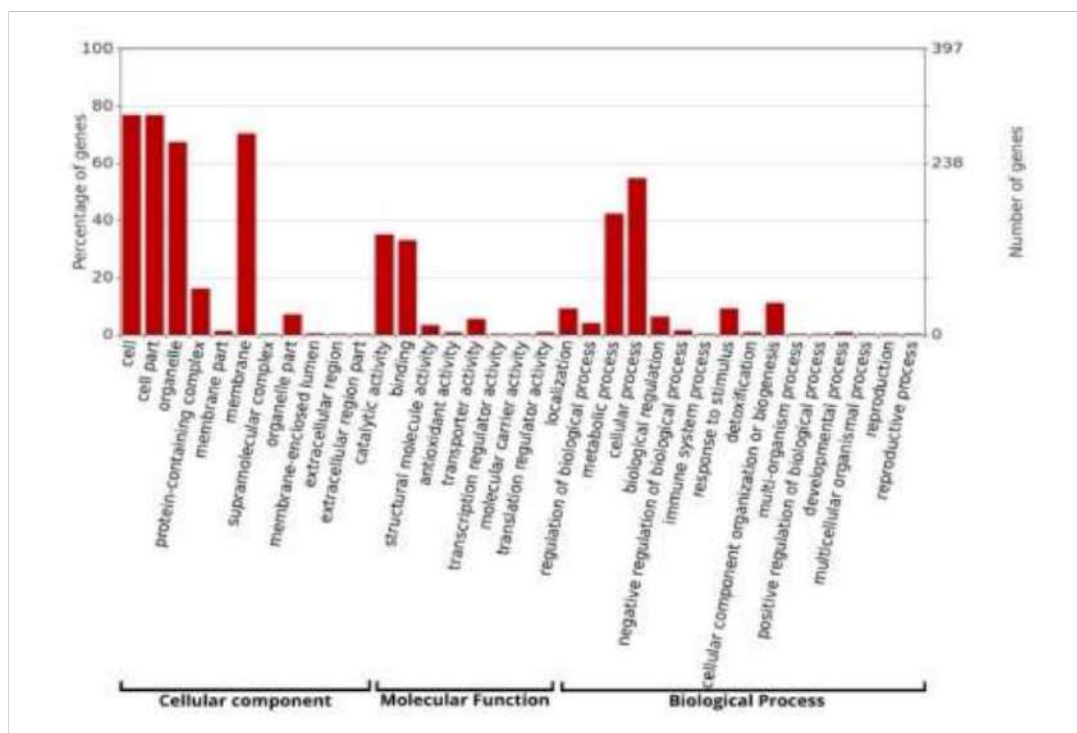


Figura 4.2 Mapa de clasificación del análisis de anotación funcional de ontología génica (GO) del genoma parcial de ‘*Ca. P. palmae*’-ALC. Se observan las distintas categorías obtenidas en el análisis de GO: procesos biológicos, componentes celulares y funciones moleculares.

La Figura 4.3 muestra la clasificación de las proteínas de ‘*Ca. P. palmae*’-ALC de acuerdo con la vía KEGG. El mayor número de genes en el metabolismo corresponde al metabolismo de carbohidratos y nucleótidos (14 genes cada uno). En el procesamiento de la información genética, el mayor número de genes se observó para traducción (68 genes), replicación y reparación (19 genes) y biogénesis de ribosomas (17 genes). En procesos celulares y señalización, los genes más frecuentes fueron en transportadores (7 genes).

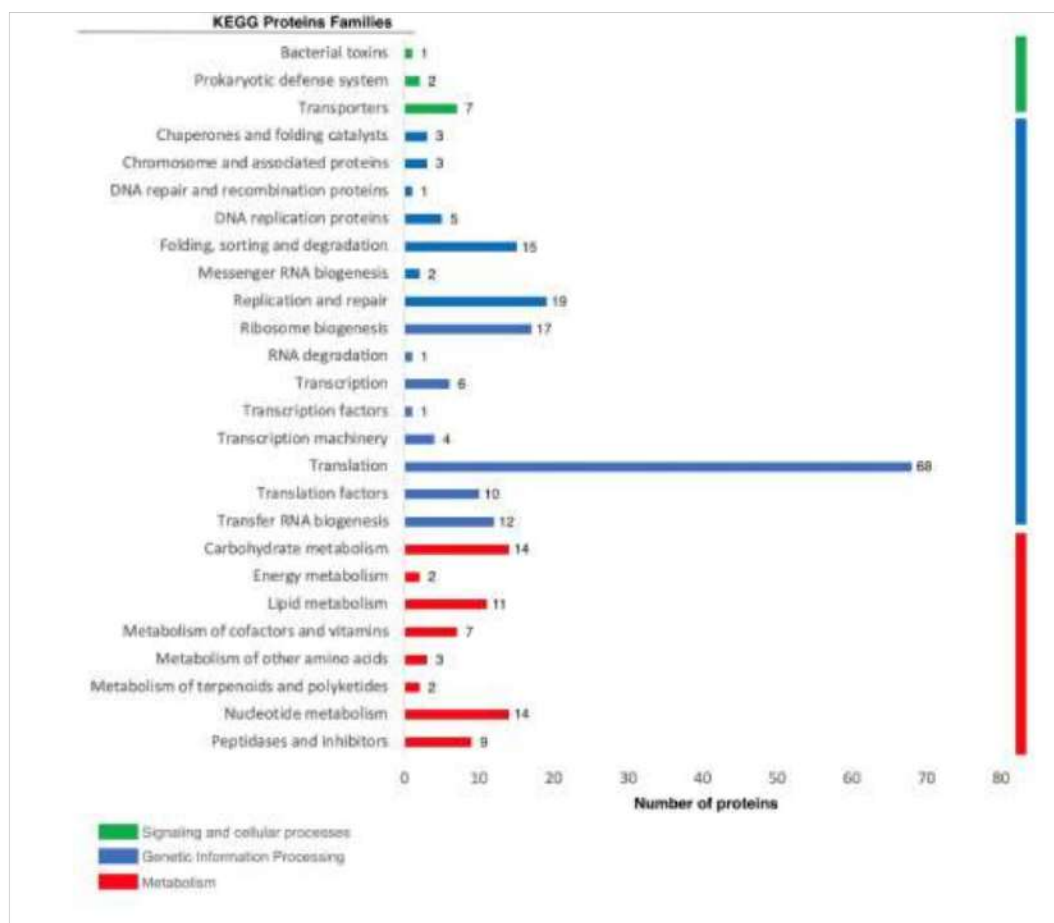


Figura 4.3 Mapa de clasificación del análisis de anotación de la Enciclopedia de Kioto de Genes y Genomas (KEGG) de 'Ca. P. palmae'-ALC. Se observa, marcados por colores las diferentes clases de familias de proteínas en: procesos celulares y señalización (color verde), procesamiento de información genética (color azul) y metabolismo (color rojo).

4.3.2.1 Metabolismo energético

En este estudio se identificaron seis genes putativos que codifican proteínas implicadas en la vía glucolítica: glucosa-6 fosfato isomerasa (*pgi*), 6-fosfofructocinasa (*pfkA*), fructosa-bisfosfato aldolasa (*fbA*), gliceraldehído 3-fosfato deshidrogenasa (*GapA*), piruvato cinasa (*pyK*) y triosa fosfato isomerasa (*tpiA*). Sin embargo, la fosfoglicerato cinasa (*pgK*), la fosfoglicerato mutasa (2,3-difosfoglicerato independiente) (*gpmI*) y la enolasa (*enO*) no estaban presentes.

La vía de acetogénesis putativa, responsable de la conversión de piruvato en acetato, se encontró completa. Esta vía incluye la subunidad alfa de la enzima piruvato deshidrogenasa (*pdhA*), subunidad beta de la deshidrogenasa (*pdhB*), dihidrolipoamida acetiltransferasa (*pdhC*), dihidrolipoamida deshidrogenasa (*pdhD*), fosfotransacetilasa (*ptA*) y acetato cinasa (*ackA*). Además, se encontraron los genes de la enzima málica (*sfcA*) y fosfato propanoil transferasa (conocida también como enzima propionil-CoA fosfatotransferasa) (*pduL*), lo que potencialmente permite a este fitoplasma convertir malato en acetato (Kube *et al.*, 2012).

Otra vía energética identificada en 'Ca. *P. palmae*'-ALC es la degradación de glicerol. Este proceso está mediado por las enzimas glicerol-3-fosfato deshidrogenasa (*GpsA*) y dihidroxiacetona cinasa. Además, la enzima CDP-diacilglicerol-glicerol-3-fosfato 3-fosfatidiltransferasa (*PgsA*) participa en la conversión de glicerol-3-fosfato en fosfatidilglicerol-fosfato.

De igual manera, se observó que el metabolismo de los glicerofosfolípidos (conversión de dihidroxiacetona fosfato en fosfatidiletanolamina) se encontró completa en 'Ca. *P. palmae*'-ALC: se identificaron la holo-proteína transportadora de acilo sintasa (*AcpS*), fosfato aciltransferasa (*PlsX*), acil fosfato glicerol-3-fosfato aciltransferasa (*PlsY*), 1-acil-sn-glicerol-3-fosfato (*PlsC*), fosfátido citidililtransferasa (*CdsA*), la fosfatidilserina transferasa (*PssA*) y la fosfatidilserina descarboxilasa (*PsD*).

4.3.2.2 Biosíntesis de proteínas

Se predijeron permeasas específicas para metionina y aminoácidos básicos (ejm. glutamina, arginina) en 'Ca. *P. palmae*'-ALC. Se identificó la N (6)-L-treonil carbamoil adenina sintasa, también conocida como el complejo treonil-carbamoil transferasa (*TsaE*) (Su *et al.*, 2022). Esta enzima es responsable de la producción de N (6)-treonil carbamoil adenosina en varios ARNt, incluidos ARNt (Ile), ARNt (Thr), ARNt (Asn), ARNt (Lys), ARNt (Ser) y ARNt (Arg). Adicionalmente, los genes putativos que codifican las enzimas 16S ARNr (guanina⁹⁶⁶-N2)-metiltransferasa (*rsmD*) y 16S ARNr de guanina⁵²⁷-N7-metiltransferasa (*rmsG*). Estas proteínas metilan el ARN ribosomal 16S en puntos específicos, la guanina en la posición 966 (N2) en el caso de *rsmD* y la guanina en la posición 527 (N7) en el caso de *rmsG*. La ARNt sintetasa para cada ARNt-aminoácido

(aminoacil-ARNt) está presente en 'Ca. P. palmae'-ALC, junto con la ARNt (guanina³⁷-N¹)-metiltransferasa (*TrmD*), la cual es esencial para mantener el marco de lectura correcto durante la traducción.

4.3.2.3 Metabolismo de purinas y pirimidinas

Es probable que 'Ca. P. palmae'-ALC sintetice derivados de timidilato mediante la enzima desoxiuridina 5'-trifosfato nucleótido hidrolasa (*Dut*), que produce desoxiuridina monofosfato (dUMP), el precursor inmediato de los nucleótidos de timidina; otros genes putativos encontrados en este fitoplasma fueron para la uridilato cinasa (*pyrH*), la timidilato cinasa (*tmk*) y la timidilato sintasa (*thyA*) (Kube *et al.*, 2012). La enzima timidilato sintasa cataliza la conversión de dUMP en desoxitimidina monofosfato (dTMP) (Kube *et al.*, 2012, p. 202; Bizarro y Schuck, 2007). El CTP (citidina trifosfato) puede generarse a partir de UTP (uridina trifosfato) mediante la enzima citidina trifosfato sintetasa (CTPS) (Goto *et al.*, 2004), ya que no se encontraron los genes *pyrG*, *udk*, *cmk*, *nrdA* ni *nrdF*, implicados en la biosíntesis de citidilatos (Kube *et al.*, 2012). Además, está presente el gen putativo que codifica la 2',3'-nucleótido cíclico 2'-fosfodiesterasa (*cpdB*). Esta enzima cataliza la conversión del nucleótido 2',3'-fosfato cíclico en nucleótido 3'-fosfato y participa en el metabolismo de las purinas y las pirimidinas (López-Villamizar *et al.*, 2016).

4.3.2.4 Estrés oxidativo

Los transportadores putativos de espermidina/putrescina y Mn/Zn, así como el gen que codifica para superóxido dismutasa de manganeso (Mn-SodA) y para fosfodiesterasa cíclica-di-AMP (GdpP), se identificaron en 'Ca. P. palmae'-ALC.

4.3.2.5 Maquinaria secretora

Fueron predichos algunos genes involucrados en el sistema de secreción (Sec) en 'Ca. P. palmae'-ALC. Se identificó la presencia de *secA* y *secY*, junto con la translocasa de la membrana interna y la chaperona insertasa para el ensamblaje de proteínas, *yidC*, los

receptores de partículas de reconocimiento de señales (*ffh* y *ftsY*), también se encontraron las chaperonas *dnaJ*, *dnaK*, *grpE* y *groEL*. Sin embargo, no se observaron en este fitoplasma los genes *secB*, *secD*, *secE* ni *secG*.

4.3.3 IDENTIFICACIÓN COMPUTACIONAL DE EFECTORES EN 'Ca. P. palmae'-ALC

Se utilizó PhyEffector (algoritmo creado en este trabajo, correspondiente al artículo 2 ya publicado, capítulo 3 de este documento) para predecir los efectores en el genoma de 'Ca. P. palmae'-ALC. Se identificaron 54 proteínas efectoras (Tabla 4.1). SignalP v4.1 (Petersen *et al.*, 2011) identificó 27 efectores secretados putativos; TMHMM v2.0 (Krogh *et al.*, 2001) predijo un dominio transmembrana (TMD) en dos efectores putativos, fig|33926.128.peg.310 y fig|33926.128.peg.246, mientras que Phobius predijo un TMD en cuatro proteínas (fig|33926.128.peg.310, fig|33926.128.peg.231, fig|33926.128.peg.246 y fig|33926.128.peg.287). Los resultados de BLASTP indicaron que veinte proteínas efectoras predichas compartían similitudes con efectores predichos de otros fitoplasmas, mientras que once efectores compartían homología exclusivamente con efectores predichos del fitoplasma del Declive de la palma Phoenix de Texas. Veintitres efectores predichos fueron específicos de 'Ca. P. palmae'-ALC, sin mostrar homología con efectores de otras especies. Se identificaron dominios funcionales en 12 efectores predichos, incluyendo los siguientes dominios: secuencia señal de proteína SVM (IPR021970), glicerol-3-fosfato deshidrogenasa dependiente de NAD (GPDH) (IPR011128), receptor de di-AMP cíclico (IPR010375), complejo de captación de zinc (IPR006127), ATPasa (IPR041569), exonucleasa-like (IPR019080), dominio central DnaJ (IPR001623), ARNasa Y (IPR003607), dedos de zinc (IPR001878), familia de proteínas no caracterizadas UPF0154 (IPR005359) y proteína de exclusión de superinfección B (IPR025982). Además, siete efectores predichos tuvieron similitud con miembros del repertorio SAP y una coincidencia para IdpA (Tabla 4.1).

Tabla 4.1 Efectores putativos predichos en 'Ca. P. palmae'-ALC

ID	Localización	PS SignalP 4.1	TMHMM 2.0	Phobius (TMD - SP)	Función	Efectores conocidos por homología	Uniprot ID (Hit)	Query cover	Porcentaje de identidad	Tribu*
fig 33926.128.peg.19	NODE_1_length_94959_ cov_11.939349_30833- 519	Sí	0	0TMD – 0SP	-	-				Tribu 1
fig 33926.128.peg.78	NODE_2_length_54417_ cov_12.079762_2362- 276	Sí	0	0TMD – YSP	-	-				-
fig 33926.128.peg.120	NODE_3_length_43610_ cov_12.819561_294-132	No	0	0TMD – YSP	-	-				-
fig 33926.128.peg.137	NODE_3_length_43610_ cov_12.819561_13643+1 152	Sí	0	0TMD – 0SP	-	-				Tribu 1
fig 33926.128.peg.190	NODE_4_length_37875_ cov_12.413776_10081- 585	Sí	0	0TMD – 0SP	-	-				Tribu 1
fig 33926.128.peg.193	NODE_4_length_37875_ cov_12.413776_12636+2 85	No	0	0TMD – 0SP	-	SAP53-like	A0A421NUE5 (Ca. P. solani SA-1)	95%	78.85%	-
fig 33926.128.peg.194	NODE_4_length_37875_ cov_12.413776_12947+3 75	Sí	0	0TMD – 0SP	Secuencia señal de la proteína SVM	SAP11-like	A0A1V0PKR7 (Lime witches'-broom phytoplasma)	98%	39.67%	Tribu 1
fig 33926.128.peg.195	NODE_4_length_37875_ cov_12.413776_13773+1 17	Sí	0	0TMD – 0SP	-	SAP53-like	A0A859IA89 (Rapeseed phyllody phytoplasma RP166)	68%	53.85%	-
fig 33926.128.peg.196	NODE_4_length_37875_ cov_12.413776_13961+1 65	No	0	0TMD – 0SP	-	SAP53-like	A0A2S8NU00 (Candidatus Phytoplasma phoenicium ChiP)	50%	66.67%	-
fig 33926.128.peg.197	NODE_4_length_37875_ cov_12.413776_14152+3 87	Sí	0	0TMD – 0SP	Secuencia señal de la proteína SVM	SAP11-like	A0A531Y0A4 (Periwinkle leaf yellowing phytoplasma)	19%	58.33%	Tribu 1
fig 33926.128.peg.208	NODE_4_length_37875_ cov_12.413776_24620+1 23	Sí	0	0TMD – 0SP	-	-				-
fig 33926.128.peg.231	NODE_5_length_23186_ cov_13.178894_15757+2 46	Sí	0	1TMD – 0SP	-	-				-
fig 33926.128.peg.246	NODE_6_length_21488_ cov_12.151169_12752- 729	No	0	0TMD – 0SP	-	ldpA	A0A3G3BKC1 (Phytoplasma sp.)	50%	25.20%	Tribu 1

fig 33926.128.peg.249	NODE_6_length_21488_cov_12.151169_19572-2118	Sí	0	0TMD – 0SP	-	Putative effector	A0A421NYW6 (Candidatus <i>Phytoplasma solani</i> SA-1)	47%	29.41%	-
fig 33926.128.peg.255	NODE_7_length_21416_cov_12.317963_2709+975	No	0	0TMD – YSP	Glicerol-3-fosfato deshidrogenasa dependiente de NAD N-terminal	-				-
fig 33926.128.peg.270	NODE_7_length_21416_cov_12.317963_20160-357	No	0	0TMD – YSP	Receptor de di-AMP cíclico	-				-
fig 33926.128.peg.271	NODE_7_length_21416_cov_12.317963_21123-765	Sí	0	0TMD – YSP	-	-				Tribu 10
fig 33926.128.peg.275	NODE_8_length_15276_cov_11.730701_7250-981	Sí	0	0TMD – 0SP	-	-				-
fig 33926.128.peg.278	NODE_8_length_15276_cov_11.730701_11001+1110	Sí	0	0TMD – 0SP	Componente A del complejo de captación de zinc periplásmico	-				Tribu 1
fig 33926.128.peg.287	NODE_9_length_13147_cov_12.203559_5688+2025	Sí	0	1TMD – 0SP	Familia ATPasa asociada con diversas actividades celulares (AAA), RecA-like_FtsH	-				Tribu 1
fig 33926.128.peg.292	NODE_10_length_12012_cov_12.105210_54+144	No	0	0TMD – 0SP	-	-				-
fig 33926.128.peg.295	NODE_10_length_12012_cov_12.105210_2991+1407	No	0	0TMD – 0SP	-	-				Tribu 1
fig 33926.128.peg.300	NODE_14_length_9020_cov_7.210597_3+84	No	0	0TMD – YSP	-	-				-
fig 33926.128.peg.301	NODE_14_length_9020_cov_7.210597_893+150	No	0	0TMD – 0SP	-	-				-

fig 33926.128.peg.302	NODE_14_length_9020_cov_7.210597_3026-123	No	0	0TMD – 0SP	-	-	-	-	-	-
fig 33926.128.peg.304	NODE_14_length_9020_cov_7.210597_7847+249	No	0	0TMD – 0SP	-	-	-	-	-	-
fig 33926.128.peg.305	NODE_14_length_9020_cov_7.210597_8162+405	No	0	0TMD – YSP	-	-	-	-	-	-
fig 33926.128.peg.306	NODE_22_length_7233_cov_11.174840_720-453	Sí	0	0TMD – 0SP	-	-	-	-	-	Tribu 1
fig 33926.128.peg.307	NODE_22_length_7233_cov_11.174840_2048-750	Sí	0	0TMD – YSP	-	SAP61-like	A0A531Y053 (Periwinkle leaf yellowing phytoplasma)	18%	30.43%	-
fig 33926.128.peg.308	NODE_22_length_7233_cov_11.174840_2869-672	Sí	0	0TMD – 0SP	-	-	-	-	-	Tribu 1
fig 33926.128.peg.309	NODE_22_length_7233_cov_11.174840_3554-324	No	0	0TMD – 0SP	-	-	-	-	-	-
fig 33926.128.peg.310	NODE_22_length_7233_cov_11.174840_3856+246	No	1	1TMD – 0SP	-	-	-	-	-	-
fig 33926.128.peg.311	NODE_22_length_7233_cov_11.174840_5156-801	Sí	0	0TMD – 0SP	-	SAP11-like	A0A859I9W5 (Rapeseed phyllody phytoplasma)	45%	29.75%	Tribu 7
fig 33926.128.peg.312	NODE_22_length_7233_cov_11.174840_6495-1029	Sí	0	0TMD – 0SP	-	-	-	-	-	Tribu 1
fig 33926.128.peg.321	NODE_25_length_6842_cov_18.486371_6252+132	No	0	0TMD – YSP	-	-	-	-	-	-
fig 33926.128.peg.334	NODE_41_length_5793_cov_11.528581_4726-594	No	0	0TMD – 0SP	PDDEXK_I ambda_exo nucleasa-like	-	-	-	-	-
fig 33926.128.peg.336	NODE_72_length_4227_cov_21.635666_1882-1116	No	0	0TMD – 0SP	Dominio central de DnaJ	-	-	-	-	-
fig 33926.128.peg.339	NODE_86_length_4030_cov_12.998491_254+429	No	0	0TMD – 0SP	-	-	-	-	-	-
fig 33926.128.peg.340	NODE_86_length_4030_cov_12.998491_1284+1569	Sí	0	0TMD – 0SP	Región N-terminal de la ARNasa Y, dominio HD/PDEas a	-	-	-	-	Tribu 7
fig 33926.128.peg.350	NODE_130_length_3229_cov_5.795841_2551+369	No	0	0TMD – 0SP	Dedos de Zinc	-	-	-	-	-

fig 33926.128.peg.351	NODE_134_length_3210 _cov_34.709033_504-132	No	0	0TMD – YSP	-	-	-
fig 33926.128.peg.363	NODE_182_length_2704 _cov_15.003775_2224-183	No	0	0TMD – OSP	-	-	-
fig 33926.128.peg.366	NODE_230_length_2398 _cov_11.486129_1905-1008	Sí	0	0TMD – OSP	-	-	Tribu 1
fig 33926.128.peg.373	NODE_267_length_2288 _cov_27.952530_1923+132	No	0	0TMD – YSP	-	-	-
fig 33926.128.peg.379	NODE_440_length_1830 _cov_17.078310_1154+135	Sí	0	0TMD – OSP	-	-	-
fig 33926.128.peg.380	NODE_769_length_1444 _cov_10.636429_3+651	Sí	0	0TMD – OSP	-	-	Tribu 1
fig 33926.128.peg.381	NODE_769_length_1444 _cov_10.636429_689+384	Sí	0	0TMD – YSP	-	-	Tribu 1
fig 33926.128.peg.383	NODE_787_length_1430 _cov_36.215273_727+414	Sí	0	0TMD – YSP	-	-	-
fig 33926.128.peg.384	NODE_787_length_1430 _cov_36.215273_1171+132	No	0	0TMD – YSP	-	-	-
fig 33926.128.peg.387	NODE_1103_length_125 8_cov_61.716542_576+165	Sí	0	0TMD – OSP	Familia de proteínas no caracteriza da (UPF0154)	-	-
fig 33926.128.peg.388	NODE_1103_length_125 8_cov_61.716542_774+129	No	0	0TMD – YSP	-	-	-
fig 33926.128.peg.389	NODE_1230_length_120 6_cov_45.813206_573-132	No	0	0TMD – YSP	-	-	-
fig 33926.128.peg.390	NODE_1230_length_120 6_cov_45.813206_1058-345	Sí	0	0TMD – OSP	Proteína B de exclusión por superinfecc ión	-	-
fig 33926.128.peg.393	NODE_2486_length_903 _cov_82.926887_426+132	No	0	0TMD – YSP	-	-	-

TMD: dominio transmembrana; SP: péptido señal. *Tribus: de acuerdo con Carreón-Anguiano *et al.* 2023 (artículo 1 publicado, capítulo 2).

En el trabajo se clasificaron los efectores de fitoplasma en 15 tribus (artículo 1 publicado, capítulo 2) basados en los motivos MEME LOGOS (>50 secuencias de aminoácidos). En el presente estudio, de los efectores predichos de '*Ca. P. palmae*'-ALC, un total de 15 efectores se clasificaron en la tribu 1, 2 efectores en la tribu 7 y 1 efector en la tribu 10 (Tabla 4.1). Interesantemente, cinco motivos nuevos fueron descubiertos en 16 de los efectores predichos de '*Ca. P. palmae*'-ALC (Figura 4.4). Específicamente, el motivo 1 se encontró en 8 efectores predichos, mientras que los 4 motivos restantes estaban presentes en 2 efectores predichos cada uno.

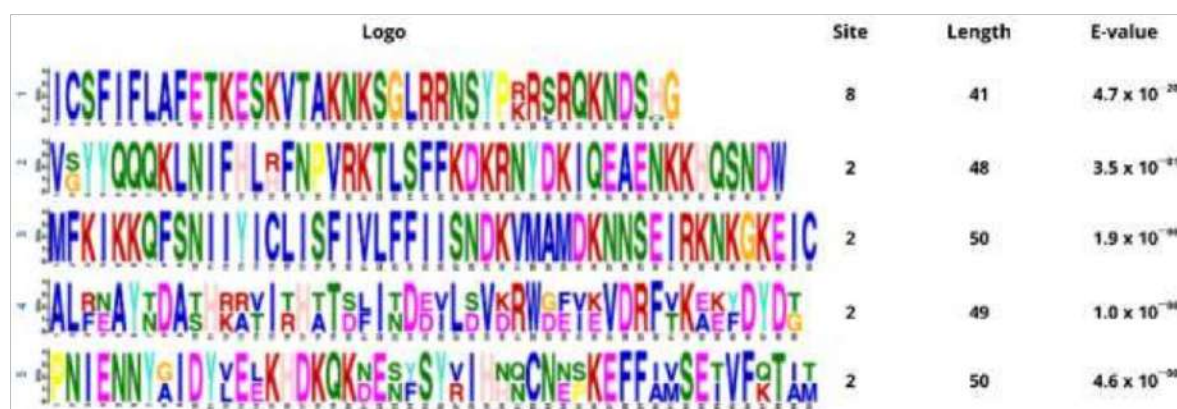


Figura 4.4 Motivos de novo (MEME) encontrados exclusivamente en los efectores predichos de '*Ca. P. palmae*'-ALC. Site: corresponde al número de efectores predichos que contienen ese motivo. Length: longitud en aminoácidos del motivo. Las posiciones con una sola letra indican una conservación del 100% de ese aminoácido en ese lugar; las posiciones con dos aminoácidos corresponden a lugares con aminoácidos alternativos en los efectores predichos.

Previamente se identificaron 42 motivos lineales eucariotas (ELMs/SLiMs) (artículo 1 publicado, capítulo 2) en efectores de fitoplasmas. Los ELMs o SLiMs son péptidos lineales cortos (3-10 aminoácidos) reconocidos como dominios de interacción que participan en

interacciones clave transitorias con proteínas, ADN o ARN, regulando diversos procesos celulares. Se denominaron a los 10 ELM más frecuentes como “Top10”. Se realizó una búsqueda de estos ELM en los efectores predichos de ‘*Ca. P. palmae*’-ALC, lo que reveló la presencia de siete motivos en veintisiete efectores predichos (Tabla 4.2). Los ELMs más comunes fueron LIG_PTB_Phospho_1, presente en 10 efectores candidatos, y LIG_PDZ_Class_2, presente en siete efectores candidatos de este fitoplasma.

Tabla 4.2 Motivos ELM encontrados en los efectores putativos de ‘*Ca. P. palmae*’-ALC

ELM	Motivo	Efectores	ID del efector	Localización
1	LIG_PDZ_Class_2	7	fig 33926.128.peg.78, fig 33926.128.peg.137, fig 33926.128.peg.270, fig 33926.128.peg.300, fig 33926.128.peg.302, fig 33926.128.peg.309, fig 33926.128.peg.336	NODE_2_length_54417_cov_12.079762_2362-276, NODE_3_length_43610_cov_12.819561_13643+1152, NODE_7_length_21416_cov_12.317963_20160-357, NODE_14_length_9020_cov_7.210597_3+84, NODE_14_length_9020_cov_7.210597_3026-123, NODE_22_length_7233_cov_11.174840_3554-324, NODE_72_length_4227_cov_21.635666_1882-1116
2	LIG_DCNL_PONY_1	2	fig 33926.128.peg.305, fig 33926.128.peg.373	NODE_14_length_9020_cov_7.210597_8162+405, NODE_267_length_2288_cov_27.952530_1923+132
3	LIG_PCNA_PIPBox_1	3	fig 33926.128.peg.249, fig 33926.128.peg.255, fig 33926.128.peg.287	NODE_6_length_21488_cov_12.151169_19572-2118, NODE_7_length_21416_cov_12.317963_2709+975, NODE_9_length_13147_cov_12.203559_5688+2025
4	LIG_GBD_Chelix_1	2	fig 33926.128.peg.78, fig 33926.128.peg.287	NODE_2_length_54417_cov_12.079762_2362-276, NODE_9_length_13147_cov_12.203559_5688+2025
5	LIG_NRP_CendR_1	---	---	---
6	TRG_ER_dilys_1	---	---	---
7	DOC_PIKK_1	2	fig 33926.128.peg.287, fig 33926.128.peg.306	NODE_9_length_13147_cov_12.203559_5688+2025, NODE_22_length_7233_cov_11.174840_720-453
8	DOC_CYCLIN_yCib1_LxF_4	1	fig 33926.128.peg.249	NODE_6_length_21488_cov_12.151169_19572-2118
9	LIG_TYR_ITSM	---	---	---
10	LIG_PTB_Phospho_1	10	fig 33926.128.peg.194, fig 33926.128.peg.197, fig 33926.128.peg.249, fig 33926.128.peg.255, fig 33926.128.peg.287, fig 33926.128.peg.295, fig 33926.128.peg.305, fig 33926.128.peg.307, fig 33926.128.peg.312, fig 33926.128.peg.387	NODE_4_length_37875_cov_12.413776_12947+375, NODE_4_length_37875_cov_12.413776_14152+387, NODE_6_length_21488_cov_12.151169_19572-2118, NODE_7_length_21416_cov_12.317963_2709+975, NODE_9_length_13147_cov_12.203559_5688+2025, NODE_10_length_12012_cov_12.105210_2991+1407, NODE_14_length_9020_cov_7.210597_8162+405, NODE_22_length_7233_cov_11.174840_2048-750, NODE_22_length_7233_cov_11.174840_6495-1029, NODE_1103_length_1258_cov_61.716542_576+165

Los genomas de fitoplasmas suelen contener regiones conocidas como Unidades Potencialmente Móviles (PMU), originadas a partir de inserciones de fagos antiguos (Jomantiene *et al.*, 2007). Estas regiones PMU, presentes en diversas especies de fitoplasmas, contienen un conjunto central de genes conservados que contribuyen a funciones como la transposición y la duplicación génica (Huang *et al.*, 2022; Bai *et al.*, 2006). Estas regiones suelen albergar genes efectores, lo que indica su papel potencial en la patogenicidad y adaptación de los fitoplasmas. Se identificaron dos regiones similares a PMU en el genoma de ‘*Ca. P. palmae*’-ALC, cada una con un gen efector. La primera región

PMU esta localizada en la región de 216,500 pb a 222,342 pb y contiene el efector putativo fig|33926.128.peg.208. La segunda región PMU se encuentra entre 235,061pb y 252,454 pb y contiene el efector putativo fig|33926.128.peg.231. La mayoría de los efectores predichos se situaron fuera de las PMU, principalmente a 300-400 kpb del genoma parcial de 'Ca. P. palmae'-ALC. Adicionalmente, algunos genes centrales de las regiones PMU se encontraban dispersos por todo el genoma en lugar de agrupados en las PMU.

4.3.4 SIMILITUDES Y DIFERENCIAS DEL GENOMA DE 'Ca. P. palmae'-ALC CON OTROS GENOMAS DE FITOPLASMAS

Los resultados de FastANI muestran que 'Ca. P. palmae'-ALC está estrechamente relacionado con el fitoplasma del Declive de la palma Phoenix de Texas (TPPD), con un 90 % de identidad entre ambos genomas. El análisis con el programa Mummer reveló aproximadamente 10 regiones genómicas con sintenia conservada entre los genomas de 'Ca. P. palmae'-ALC y 'Ca. P. palmae'-TPPD. La región sinténica mayor, corresponde a la mitad del genoma. Se encontró una identidad de nucleótidos promedio (ANI) entre el 75 % y 77 % entre el genoma de 'Ca. P. palmae'-ALC y los genomas de otros fitoplasmas (Tabla 4.3).

Tabla 4.3 Comparación por pares del genoma de 'Ca. P. palmae'-ALC (16SrIV-A) y genomas de otros fitoplasmas

Genomas	Grupo 16S ARNr	Tamaño del genoma (pb)	Identidad de nucleótidos promedio (ANI)
'Ca. P. asteris' OY-M	I-B	853,092	75.6314%
'Ca. P. asteris' AY-WB	I-A	706,569	75.7647%
'Ca. P. mali' AT	X-A	601,943	76.6453%
'Ca. P. australiense' PAa	XII-B	879,324	75.8888%
'Ca. P. australiense' SLY	XII	959,779	76.0362%
'Ca. P. asteris' MBSP M3	I	576,118	ND*
'Ca. P. ziziphi' JWB	V	750,803	77.4144%
'Ca. P. asteris' De Villa	I	600,116	75.7321%
'Ca. P. aurantifolia' EpWB	II	635,584	76.0858%
'Ca. P. luffae'	VIII	769,143	77.2628%
Flavescence dorée phytoplasma	V-D	654,223	77.5991%
'Ca. P. palmae'-TPPD cepa ACPD	IV-D	486,100	90.3423%

*El programa no arrojó resultado

La comparación de la gráfica circular del genoma parcial de 'Ca. P. palmae'-ALC con los genomas de otros fitoplasmas, identificó regiones genómicas en el genoma de 'Ca. P. palmae'-ALC carentes de homología (alineamiento vacío) con los otros genomas (Figura 4.5, flechas negras). Solo el fitoplasma del Declive de la palma Phoenix de Texas (cepa ACPD) comparte algunas de estas regiones genómicas, pero no todas. Muchos efectores predichos de 'Ca. P. palmae'-ALC se localizaron en estas regiones genómicas específicas de 'Ca. P. palmae'-ALC.

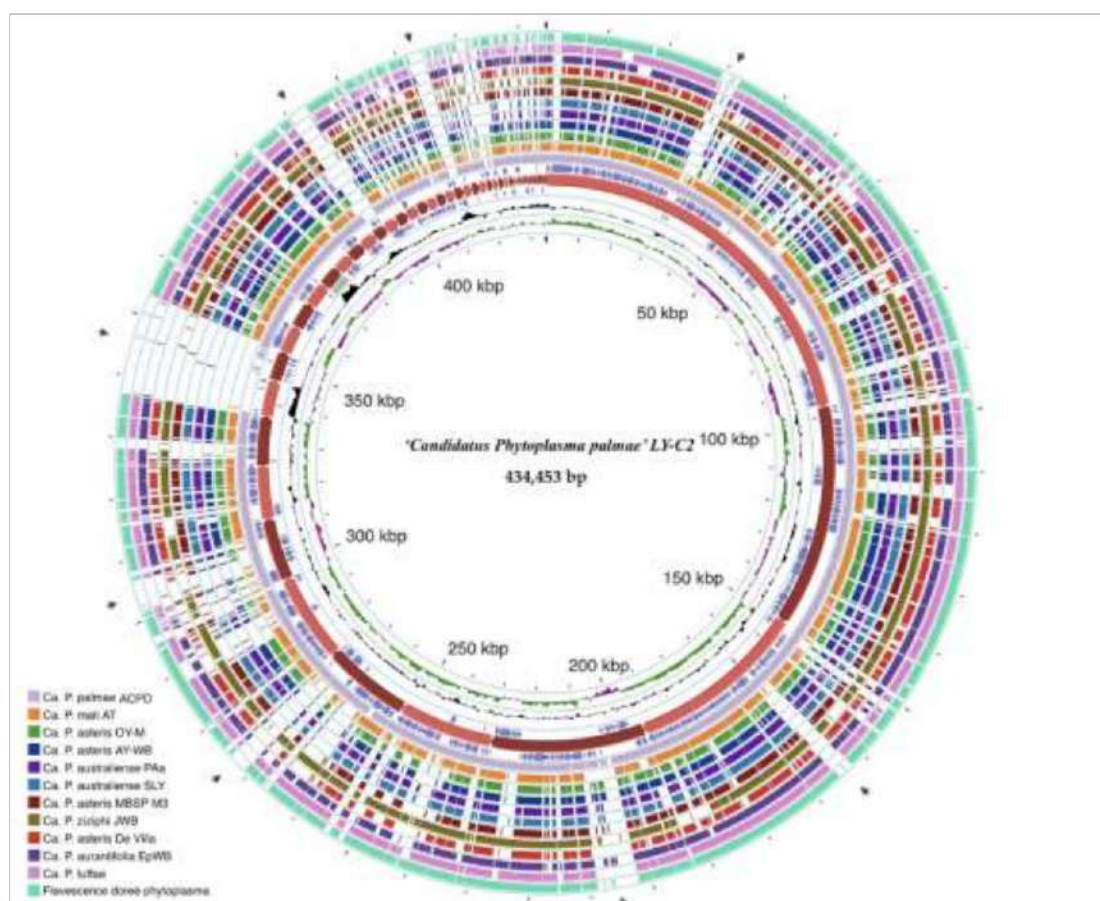


Figura 4.5 Comparación circular en BLASTN del genoma de 'Ca. P. palmae'-ALC con otros doce genomas de fitoplasmas. Marcado del centro hacia afuera: Contigs de 'Ca. P. palmae'-ALC (en color rojo marrón y salmón); el panel izquierdo indica los colores correspondientes a cada genoma de fitoplasma. Las flechas negras indican las regiones genómicas donde 'Ca. P. palmae'-ALC carece de homología con los otros genomas de fitoplasmas.

4.4 DISCUSIÓN

En general, la investigación sobre la enfermedad del amarillamiento letal se ha centrado en la caracterización de las cepas que afectan a las palmas, la identificación de posibles vectores y el desarrollo de mejores métodos de diagnóstico, principalmente mediante PCR en tiempo real (qPCR). Sin embargo, no se han reportado estudios sobre la genómica de los fitoplasmas del amarillamiento letal (Oropeza *et al.*, 2020).

El presente reporte es el primero relacionado con el genoma parcial de '*Ca. P. palmae*'-ALC y también la primera descripción de un genoma de fitoplasma perteneciente al grupo 16SrIV. El genoma parcial del fitoplasma del Declive de la palma Phoenix de Texas (cepa ACPD) está disponible en GenBank (GCA_005774685.2), pero hasta donde se conoce, este genoma no ha sido descrito.

El genoma parcial obtenido de '*Ca. P. palmae*'-ALC tiene un tamaño de 434,453 pb. El análisis BUSCO identificó 135 de 151 genes marcadores del conjunto de datos "Mollicutes" (89.4 %) en el genoma '*Ca. P. palmae*'-ALC. Este porcentaje es similar al de otros reportes (Böhm *et al.*, 2024; Debonneville *et al.*, 2022). El análisis con MiGa, para evaluar ensamblajes genómicos procariotas, calificó la completitud como "muy alta", lo que evidencia que, aunque el genoma de '*Ca. P. palmae*'-ALC es parcial, vale la pena analizarlo para su descripción.

Aunque los plásmidos están en algunos fitoplasmas (Tan *et al.*, 2021; Coetzee *et al.*, 2019; Andersen *et al.*, 2013; Tran-Nguyen *et al.*, 2008; Bai *et al.*, 2006; Oshima *et al.*, 2004), no se encontró plásmidos en '*Ca. P. palmae*'-ALC, pero no se descarta del todo hasta que el genoma sea completado. El análisis comparativo del genoma parcial de este fitoplasma con otros doce genomas de fitoplasmas reveló muchas regiones genómicas compartidas con '*Ca. P. palmae*'-ALC, pero también regiones genómicas específicas exclusivas de '*Ca. P. palmae*'-ALC; muchos efectores predichos en este fitoplasma (proteínas hipotéticas con funciones desconocidas) se localizaron en estas regiones genómicas específicas de '*Ca. P. palmae*'-ALC.

'*Ca. P. palmae*'-ALC comparte los genes *pgi*, *pfkA*, *fbA*, *gapA*, *pyK* y *pfkA* con '*Ca. P. asteris*' cepas OY-M y AY-WB, '*Ca. P. australiense*' y '*Ca. P. mali*', correspondientes a la parte superior y central de la glucólisis (Kube *et al.*, 2012). Sin embargo, los genes de la parte inferior de dicha ruta (*pgK*, *gpmI*, *enO*) no se identificaron, similar a lo reportado para '*Ca.*

P. mali' (Kube *et al.*, 2008), aunque, como ya se ha mencionado, solo se podrá confirmar cuando el genoma sea completado.

Es probable que el malato y la maltosa se utilicen como fuentes de carbono y energía en este fitoplasma. La maltosa es uno de los principales azúcares transportados en las plantas durante la descomposición del almidón, y se identificó un gen para el transportador ABC de maltosa/maltodextrina en '*Ca. P. palmae*'-ALC, lo cual, también se observa en el genoma de '*Ca. P. rubi*' (Böhm *et al.*, 2023); sin embargo, los genes para las enzimas que degradan la maltosa no se identificaron en '*Ca. P. palmae*'-ALC, similar a lo reportado en '*Ca. P. mali*' (Kube *et al.*, 2008). Por el contrario, el gen putativo para el simportador malato/citrato/H⁺ se encontró en '*Ca. P. palmae*'-ALC, junto con todos los genes putativos que codifican enzimas para convertir malato en acetato (*mleP*, *sfcA*, *pdhAB*, *pdhD*, *pduL*, *ackA*) (Kube *et al.* (2012), lo que respalda la vía de producción de energía malato-acetato en '*Ca. P. palmae*'-ALC.

Otra vía completa predicha en '*Ca. P. palmae*'-ALC es la del metabolismo de los glicerofosfolípidos, específicamente la conversión de dihidroxiacetona fosfato en fosfatidiletanolamina. Esta vía proporciona una fuente alternativa de energía. La vía metabólica de los glicerofosfolípidos también se encontró completa en los genomas de '*Ca. P. rubi*' (Böhm *et al.*, 2023) y '*Ca. P. ziziphi*' (Xue *et al.*, 2023), ambos pertenecientes al grupo 16SrV.

'*Ca. P. palmae*'-ALC probablemente produce pirimidinas a partir de dUMP, ya que se predijeron los genes *dut*, *pyrH*, *tmk* y *thyA* (Kube *et al.*, 2012), pero probablemente depende de precursores externos de nucleótidos de purina. Sin embargo, no se encontró ningún transportador de nucleósidos en '*Ca. P. palmae*'-ALC, similar a lo reportado en otros fitoplasmas (Kube *et al.*, 2012) y otros Mollicutes como *Mycoplasma genitalium* y *M. pneumoniae* (Bizarro y Schuck, 2007). Esto sugiere que un transportador no específico importa precursores de nucleótidos de purina a '*Ca. P. palmae*'-ALC.

Se identificaron permeasas específicas para metionina y aminoácidos básicos (e.g. glutamina, arginina) en '*Ca. P. palmae*'-ALC. Es probable que otros aminoácidos sean importados por transportadores no específicos. La presencia de una permeasa específica de metionina es otra característica compartida entre '*Ca. P. palmae*'-ALC, '*Ca. P. rubi*' (Böhm *et al.*, 2023) y '*Ca. P. ziziphi*' (Xue *et al.*, 2023).

Los transportadores putativos de espermidina/putrescina y Mn/Zn también fueron identificados en 'Ca. P. palmae'-ALC. Estos transportadores se han reportado en los genomas de los fitoplasmas del grupo 16SrV, el fitoplasma de la escoba de bruja del Jujube o dátil rojo (JWB) (Wang *et al.*, 2018a), 'Ca. P. rubi' (Böhm *et al.*, 2023) y 'Ca. P. ziziphi' (Xue *et al.*, 2023), y el fitoplasma del grupo 16Srl, el fitoplasmas del enanismo azul de trigo (Chen *et al.*, 2014), lo que indica diferentes orígenes y mosaicismo de los elementos adquiridos. Al igual que 'Ca. P. ziziphi' (Xue *et al.*, 2023), 'Ca. P. palmae'-ALC posee el gen que codifica para Mn-SodA, lo que sugiere que estos fitoplasmas se enfrentan al estrés oxidativo en sus hospedadores. Mn-SodA puede proporcionar capacidad redox, mientras que las poliaminas pueden prevenir la desnaturalización de las proteínas del fitoplasma. Adicionalmente, la fosfodiesterasa de di-AMP cíclico (GdpP), que modula los niveles de di-AMP cíclico en respuesta al estrés de la membrana celular (Wang *et al.*, 2017), también se identificó en 'Ca. P. palmae'-ALC.

Se identificaron varios componentes del sistema Sec en 'Ca. P. palmae'-ALC, que incluye *secA*, *secY*, *yidC*, *ffh*, *ftsY*, *dnaJ*, *dnaK*, *grpE* y *groEL*; estos genes también se encontraron en los genomas de 'Ca. P. mali' y 'Ca. P. asteris' (Kube *et al.*, 2008), pero no se encontraron los genes *secB*, *secD*, *secE* ni *secG*. La ausencia de estos genes es una característica común en muchos fitoplasmas, como 'Ca. P. tritici', 'Ca. P. asteris', 'Ca. P. australiense', 'Ca. P. mali' y 'Ca. P. ziziphi' (Xue *et al.*, 2023; Chen *et al.*, 2014; Kube *et al.*, 2008). Sin embargo, su ausencia sólo podrá confirmarse o rebatirse hasta completar el genoma.

Uno de los productos importantes secretados por los fitoplasma a través del sistema Sec son los efectores. En 'Ca. P. palmae'-ALC, se predijeron 54 efectores, 27 contienen un péptido señal y cuatro efectores un dominio transmembrana (dos de ellos adicional al péptido señal), de acuerdo con el programa Phobius. Recientemente han sido reportados efectores con TMDs en el fitoplasmas de la flavescencia dorada (Debonneville *et al.*, 2022), pero las funciones siguen siendo desconocidas. Algunos efectores predichos de 'Ca. P. palmae'-ALC comparten homología con efectores conocidos (SAP11, SAP53, SAP54 y SAP61). Sin embargo, la mayor proporción de ellos (27) se anotó como proteínas hipotéticas con funciones desconocidas, 24 no tuvieron hits con otros organismos y 11 tenían un homólogo cada uno en el fitoplasma de Declive de la palma Phoenix de Texas. La probable proteína de membrana inmunodominante, efector IdpA, también se identificó en 'Ca. P. palmae'-ALC, aunque la cobertura y la identidad fueron bajas (50 %, 25.20 %, respectivamente). Este es uno de los tres tipos de proteínas de membrana

inmunodominante (IDP), que comprende las proteínas de membrana altamente abundantes Imp (proteína de membrana inmunodominante), IdpA (proteína de membrana inmunodominante A) y Amp (proteína de membrana antigénica). Estos efectores superficiales, altamente abundantes, median la adhesión del fitoplasma a las células de insectos (Arricau-Bouvery *et al.*, 2021; Rashidi *et al.*, 2015). No se encontraron resultados para Amp o Imp en 'Ca. P. palmae'-ALC.

En 'Ca. P. palmae'-ALC, se identificaron dos regiones PMU que contienen un efector predicho en cada una. Sin embargo, la mayoría de los efectores predichos se ubicaron principalmente fuera de las regiones de PMU, asociados con regiones genómicas que no muestran homología con otros genomas de fitoplasmas. 'Ca. P. palmae'-ALC interactúa específicamente con su insecto vector y la palma de coco. Por lo tanto, la capacidad de duplicar y frecuentemente transferir efectores a través del genoma puede no ser necesaria para este fitoplasma. Esto también es consistente con la pobre relación que se encuentra entre los efectores predichos de 'Ca. P. palmae'-ALC y efectores de otros fitoplasmas.

En muchos fitoplasmas, los efectores se encuentran dentro de regiones PMUs, que pueden transferirlos horizontalmente entre fitoplasmas relacionados estrechamente o distantemente (Tokuda *et al.*, 2023; Huang *et al.*, 2022). Adicionalmente, en algunas especies, las PMUs pueden constituir hasta el 25 % del genoma (Huang *et al.*, 2022). La variabilidad en el número de PMU puede atribuirse a varios factores, lo que refleja la historia evolutiva, el nicho ecológico y la adaptación genética de cada especie de fitoplasma. Las especies de fitoplasma con amplia gamma de hospedantes y diversos nichos ecológicos pueden poseer una mayor diversidad de PMUs. Por el contrario, los fitoplasmas con hospedantes reducidos, pueden tener cambios evolutivos más lentos y requerir menos PMUs, transferidos por herencia vertical entre los linajes estrechamente relacionados, ya que se observan PMUs similares entre linajes de fitoplasmas relacionados (Tokuda *et al.*, 2023). Sin embargo, aunque la mayoría de los efectores de fitoplasmas reportados están codificados en PMUs, también hay fitoplasmas con efectores codificados principalmente fuera de las regiones de PMU (Tan *et al.*, 2021; Hoshi *et al.*, 2009), como se observó para 'Ca. P. palmae'-ALC.

Aunque los programas que identifican altos niveles de conservación no han podido encontrar familias en los efectores de fitoplasma, en este trabajo se identificaron motivos *de novo* (MEME) y ELMs. En el presente estudio, un total de 15 efectores y 2 efectores

predichos de 'Ca. P. palmae'-ALC se clasificaron en las tribus 1 y 7, respectivamente; tribus que comprenden efectores de varios fitoplasmas diferentes. Un efector predicho de 'Ca. P. palmae'-ALC fue clasificado en la Tribu 10, la cual comprende principalmente efectores de fitoplasmas pertenecientes al grupo 16SrV. Además, se identificaron secuencias de motivos lineales cortos (motivos ELM/SLiMs) en los efectores predichos de 'Ca. P. palmae'-ALC. Veintisiete efectores predichos contenían uno de los 10 motivos ELM más frecuentes (Top-10) reportados previamente en los efectores de fitoplasmas (artículo 1 publicado en el capítulo 2). Los motivos ELM más frecuentes fueron LIG_PDZ_Class_2 y LIG_PTB_Phospho_1, presentes en 7 y 10 efectores predichos, respectivamente. LIG_PDZ_Class_2 presenta el patrón [VYF)X(VIL], y los efectores con este motivo pueden dirigirse a las proteínas de membrana del hospedante que contienen el dominio PDZ y participan en señalización, similar a la estrategia infecciosa seguida por algunos virus (Caillet-Saguy *et al.*, 2021). LIG_PTB_Phospho_1 presenta los patrones [QFSNIIY], [TILNKCY] y [CLMNIIY] y se une a la proteína DOK1, una proteína adaptadora o de andamiaje que proporciona una plataforma de acoplamiento para el ensamblaje de complejos de señalización multimolecular (Kumar *et al.*, 2022). Por lo tanto, es probable que los efectores predichos de 'Ca. P. palmae'-ALC que contienen el motivo LIG_PTB_Phospho_1 también interfieran con la transducción de señales en el hospedante.

El dominio SVM, una secuencia señal de aproximadamente 30 aminoácidos, descrito como una característica única de los fitoplasmas (Jomantiene *et al.*, 2007), se encontró en dos efectores predichos de 'Ca. P. palmae'-ALC. Quizás estos motivos MEME, ELM y SVM existieron en efectores ancestros y estén presentes en los efectores modernos, aunque no muestran conservación de secuencia fuera de estos motivos proteicos. Investigaciones futuras sobre estos aminoácidos y sus funciones podrían proporcionar información valiosa sobre la biología de los fitoplasmas y sus interacciones con las plantas hospedadoras. Esta investigación podría revelar mecanismos que contribuyen a su patogenicidad y estrategias evolutivas.

Más de la mitad del genoma parcial de 'Ca. P. palmae'-ALC muestra conservación de sintenia con el genoma parcial de 'Ca. P. palmae'-TPPD. Esto es interesante porque a menudo se ha reportado una falta de conservación de sintenia a nivel genómico en comparaciones entre secuencias genómicas de fitoplasmas (Huang *et al.*, 2022), incluso entre once cepas de 'Ca. P. asteris' que comparten >98 % de ANI en todo el genoma (Cho

et al., 2020). Por lo tanto, la disponibilidad del genoma de 'Ca. P. palmae'-ALC podría ser útil para futuras investigaciones sobre el grupo de fitoplasmas 'Ca. P. palmae'.

CAPÍTULO V

CARACTERIZACIÓN DE DOS EFECTORES DE FITOPLASMA DE ALC

5.1 INTRODUCCIÓN

Los efectores son moléculas secretadas por los patógenos que alteran las funciones biológicas del hospedante, generalmente son definidas como proteínas pequeñas que intervienen en la virulencia del patógeno (Hogenhout *et al.*, 2009). Se dividen en dos grupos: efectores apoplásticos (son secretados y actúan en el espacio extracelular) y efectores citoplasmáticos (son secretados y actúan dentro de las células del hospedante), llegando a ser secretados de 30 hasta 200 efectores por fitopatógeno (Carella *et al.*, 2018). Los efectores de fitoplasmas son liberados en el citoplasma de las células cribosas interactuando con sus componentes o se dispersan del floema a otros tejidos vegetales, degradando proteínas importantes necesarias para el correcto crecimiento de la planta, lo que ocasiona distintos cambios morfológicos en las especies vegetales infectadas (Rashid *et al.*, 2018; Sugio *et al.*, 2011a; Hogenhout *et al.*, 2009), cambios que son atractivos para insectos vectores, provocando la transmisión y diseminación de estos patógenos (Sugio *et al.*, 2011a).

La identificación y caracterización de efectores permitirá comprender mejor sobre la virulencia y mecanismo de acción de estos patógenos, lo que podría ayudar a la identificación de moléculas blanco que en un futuro puedan protegerse e impedir la pérdida de importantes especies vegetales con valor comercial. La validación de efectores de fitoplasmas se ha realizado mediante la expresión de estas proteínas efectoras en plantas modelo como *Nicotiana benthamiana* y en líneas transgénicas de *Arabidopsis thaliana* (Huang *et al.*, 2021; MacLean *et al.*, 2014; Maejima *et al.*, 2014; Sugio *et al.*, 2011b; Hoshi *et al.*, 2009); siendo hasta el momento confirmados 21 efectores de fitoplasmas (Wang *et al.*, 2023; Chen *et al.*, 2022; Arricau-Bouvery *et al.*, 2021; Huang *et al.*, 2021; Wang *et al.*, 2017; MacLean *et al.*, 2014; Maejima *et al.*, 2014; Sugio *et al.*, 2014; Boonrod *et al.*, 2012; Hoshi *et al.*, 2009; Kakizawa *et al.*, 2009). Sin embargo, ningún efector del fitoplasma responsable del ALC había sido identificado. En este trabajo se caracterizaron dos efectores candidatos identificados en el genoma de 'Ca. P. Palmae', agente causal del ALC, logrando la expresión de un efector a través de muestras de dos palmas de cocotero

enfermas con este fitoplasma, mediante el diseño de cebadores específicos a partir de las secuencias de efectores candidatos previamente identificados en su genoma.

5.2 MATERIALES Y MÉTODOS

5.2.1. OBTENCIÓN DEL MATERIAL VEGETAL

Se colectó el tallo, hojas y raíces de cinco palmas de *Cocos nucifera* L. en Telchac Puerto, Mérida, Yucatán. Una palma asintomática (sana para control) y 4 palmas con síntomas asociados con ALC. Primero, se colectó la palma asintomática para evitar contaminación con las otras palmas. Las muestras se colectaron por duplicado, la primera muestra se colocó en un tubo Falcón con 5 mL del reactivo Shield™ + 5 µL de tioglicerol (preparado un día anterior de la colecta) y la segunda muestra se colocó en una bolsa, cada una debidamente etiquetada y mantenida en hielo en el momento de la colecta y transporte, posteriormente se guardaron a 4 °C para su posterior procesamiento.

5.2.2 EXTRACCIÓN DE ADN PARA LA EVALUACIÓN DE LA PRESENCIA DE FITOPLASMAS

Para detectar la presencia de fitoplasmas, se extrajo el ADN a partir de las muestras de tallo. Para ello se empleó el protocolo de extracción con amortiguador con CTAB (Bromuro de hexadeciltrimetilamonio) de acuerdo con Córdova *et al.* (2014), el cual se explica brevemente a continuación. Se utilizó un 1 mL de cada muestra de tallo (tubo Falcón con 1 g de la muestra de tallo en el reactivo Shield™) y se mezcló con 1 mL del amortiguador CTAB con 1 µL de tioglicerol (1 µL por cada mL de CTAB). Las muestras se incubaron a 65°C durante 20 minutos en el baño María, mezclando ocasionalmente. De cada extracto resultante se pasó 1 mL a un tubo Eppendorf y se añadió 1 mL (volumen 1:1) de una mezcla de fenol:cloroformo:alcohol isoamílico (25:24:1), con cuidado de tapar bien cada tubo. Se mezcló en el vortex y se centrifugó a 14,000 rpm durante 5 min a temperatura ambiente. Posteriormente se recuperó la fase acuosa, transfiriéndola a un tubo Eppendorf nuevo (~1 mL); para precipitar los ácidos nucleicos totales se agregó 0.6 volumen (600 µl) de isopropanol frío y 0.1 volumen (100 µL) de acetato de sodio (3M, pH 5.2), mezclando

suavemente por inversión y dejando incubar durante 60 min a -20 °C. Posteriormente se realizó una centrifugación a 14,000 rpm durante 10 min a 4 °C, se recuperó la pastilla o “pellet” (correspondiente al ácido nucleico) y se desechó la fase acuosa. La pastilla se lavó con 500 µL de etanol al 70 % frío y se centrifugó de nuevo a 7500 rpm durante 5 min a temperatura ambiente. Se eliminó el etanol y la pastilla se dejó secar a temperatura ambiente por unos pocos minutos. Finalmente, el ADN se resuspendió en 50 µL de agua ultrapura y se guardó en el congelador a -20°C. Se determinó la concentración de ADN mediante un espectrofotómetro (NanoDrop One, Thermo Scientific™) utilizando 2 µL de muestra.

5.2.3 CONFIRMACIÓN DE PRESENCIA DE FITOPLASMAS MEDIANTE PCR EN TIEMPO REAL (qPCR)

Se confirmó la presencia del fitoplasma en las palmas de coco con síntomas asociados al ALC y la ausencia del patógeno en la palma sana (control) mediante la técnica de PCR en tiempo real (qPCR) empleando el ensayo TaqMan LY16S reportado por Córdova *et al.* (2014). Las condiciones que se usaron fueron: 2 min a 50°C para activar la AmpErase UNG, 10 min a 95°C para activar la ADN polimerasa AmpliTaq Gold y 35 ciclos a 95°C durante 15 s y 61°C durante 1 min. Como control negativo se usó agua estéril y como control positivo el ADN de palma enferma con ALC.

5.2.4 EXTRACCIÓN DE ARN

Se seleccionaron las palmas de cocotero que dieron positivo a la presencia de fitoplasmas (Ct <30 en el ensayo con sonda TaqMan LY16S), usando la palma sana como control y se extrajo el ARN de cada una de las partes colectadas (tallos, hojas y raíces) correspondientes a esa palma.

Para extraer el ARN se utilizó el Kit de extracción Plant/Fungi Total RNA purification Kit (Norgen Biotek), siguiendo las instrucciones del fabricante. El procedimiento se describe a continuación y consta de cuatro pasos principales: *Lisado, Enlace en la columna, Lavado y Elución del ARN.*

1. Lisado: Para las hojas y raíces se pesó 100 mg de muestra y se pulverizó con nitrógeno líquido en un mortero, se colocó la muestra pulverizada en un tubo Eppendorf de 2 mL que previamente contenía 600 μ L de solución amortiguadora de lisis con tioglicerol (1 μ L por cada mL de amortiguador). Para las muestras de tallo se tomaron 600 μ L del sobrenadante (muestras de tallo con 1 g de muestra mezcladas con 5 mL de solución Shields, que se guardaron en el tubo Falcón), y se agregó al tubo con el amortiguador preparado, se mezcló en el vortex vigorosamente. Se dejó incubar 5 minutos a 55 °C en el baño María, mezclando ocasionalmente (2 o 3 veces durante la incubación). Posteriormente se colectó el sobrenadante (dejando unos segundos para que el material vegetal se asiente) y se depositó en una columna de filtrado (anillo claro) con un tubo de colecta, previamente armada, dando un spin de 14,000 rpm por 2 minutos. Posteriormente se desechó la columna de filtrado y con cuidado de no tocar la pastilla, se recuperó el sobrenadante. Al filtrado recuperado se le añadió en volúmenes iguales (1:1) etanol al 100 % (grado biología molecular) y se mezcló en vortex.

2. Enlace en la columna: Se transfirieron 600 μ L de lisado a una columna de Spin (anillo gris) previamente armada en un tubo de colecta y se centrifugó a 600 rpm por 1 minuto. Se desechó el filtrado y se agregó más sobrenadante o filtrado recuperado previamente, se repitió el procedimiento hasta que todo el filtrado recuperado del lisado se haya adherido a la columna. Al final se dio una centrifugación a 14,000 rpm por 1 min para asegurar que todo haya pasado por la columna.

3. Lavados: Para los lavados se añadió 400 μ L de solución de lavado A (Wash solution A) a la columna spin y se centrifugó a 14, 000 rpm por 1 minuto (si fue necesario se dio un spin adicional para asegurar que todo el líquido paso por la columna). Se realizaron tres lavados, en el último lavado se dio una centrifugación adicional de 14,000 rpm por 2 minuto para garantizar que no quedó líquido remanente en la columna.

4. Elución del ARN: En un tubo Eppendorf nuevo de 1.7 mL se colocó la columna spin y se añadieron en el centro de la misma 50 μ L de solución de Elución A (Elution solution A),

dejando reposar 2 minutos, posteriormente se centrifugó a 14,000 rpm por 2 minutos; si fue necesario se dio otra centrifugación de 14,000 rpm por 1 minutos.

La cantidad de ARN obtenido se cuantificó en el Nanodrop One (Thermo Scientific™) y se visualizaron 3 µL de muestra en gel de agarosa al 1 % corrido a 90 volts, 200 Amp, 30 min.

Las muestras se guardaron a -20°C para su posterior procesamiento.

Para las muestras difíciles de procesar como las raíces se utilizó el método de CTAB como segundo intento para obtener el ARN.

La obtención de ARN de las muestras de raíz es difícil; no obstante, para las muestras denominadas P1-raíz y P2-raíz se obtuvo el ARN, utilizando las muestras colectadas en bolsa (sin sol. Shields) y la técnica de CTAB. Adicionalmente para garantizar la obtención del ARN de la raíz en lugar de pesar 100 mg se pesaron 200 mg. Se cortó la raíz y se pulverizó con nitrógeno líquido, para garantizar la mayor cantidad de muestra obtenida (que no se pierda muestra por quedarse adherida en las paredes del mortero y/o pistilo), la muestra triturada se transfirió con ayuda de una espátula a la misma charola que se usó para pesar la muestra y posteriormente, se transfirió todo a un tubo con 1 ml del amortiguador CTAB con 1 µL de tioglicerol (1 µL por cada mL de CTAB), se pasó en el vortex vigorosamente y se continuó con el protocolo de CTAB.

5.2.5 DETERMINACIÓN DE LA CANTIDAD DE ARN Y VISUALIZACIÓN

La cuantificación del ARN se realizó mediante espectrofotometría en un NanoDrop One (Thermo Scientific™). Para la medición de la concentración y pureza se utilizó 2 µL de cada muestra, se midió la absorbancia a 260 nm para determinar la concentración de ácidos nucleicos, además de que proporciona dos índices de pureza con base a la relación de absorbancia A260/A280 y A260/A230 para determinar la presencia de proteínas o contaminantes en la muestra. El ARN fue analizado mediante electroforesis horizontal en gel de agarosa; los geles de agarosa se prepararon al 1 % usando 1 µL del colorante bromuro de etidio (BrEt), y se utilizó 3 µL del marcador 1Kb GeneRuler (Thermo scientific™), se cargaron 3 µL de muestra en los pocillos y se corrió a 90 volts, 200 AMP,

30 min. La visualización se realizó con luz ultravioleta (UV) en el equipo de sistema de imagen Vilber E-box.

En todas las muestras se observó ARN, pero también restos de ADN, por lo tanto, se realizó el procedimiento de eliminación del ADN.

5.2.6 LIMPIEZA DEL ARN MEDIANTE DIGESTIÓN ENZIMÁTICA DEL ADN

Para la obtención del ARN se empleó el protocolo de eliminación del ADN mediante un tratamiento con ADNasa I, libre de ARNasa (Thermo Scientific). Brevemente, se preparó una mezcla de reacción con 6 µL de agua estéril libre de DNAasa, 2 µL de amortiguador 10X con MgCl₂, 2 µL de ADNasa y 10 µL de la muestra de ácidos nucleicos (ARN con trazas de ADN), volumen total de la mezcla de reacción de 20 µL, incubando a 37 °C x 30 min en un termociclador (T100TM Thermal cycler); posteriormente se agregó 1 µL de EDTA (ácido etilendiaminotetraacético) a cada tubo para parar la reacción, incubando a 65 °C x 10 min.

Para verificar la obtención del ARN y correcta eliminación del ADN se corrió un gel de agarosa al 1 % con 5 µL de muestra a 90 volts, 200 AMP durante 35 min.

5.2.7 SÍNTESIS DEL ADN COMPLEMENTARIO (ADNc)

El ADNc se sintetizó a partir del ARN total libre de ADN, empleando un kit de transcriptasa reversa (RevertAid Reverse Transcriptase, Thermo ScientificTM), siguiendo las instrucciones del fabricante.

5.2.8 DISEÑO DE CEBADORES ESPECÍFICOS PARA EFECTORES DEL FITOPLASMA DEL ALC

Del genoma *de novo* de '*Ca. P. palmae*', secuenciado en este trabajo, se seleccionaron dos secuencias candidatas de efectores. Las secuencias de nucleótidos identificadas se seleccionaron para diseñar parejas de cebadores específicos para la amplificación de efectores del fitoplasma en palmas de cocotero enfermas con ALC. Se realizó un

alineamiento de ambas secuencias en el programa Clustal Omega < Multiple Sequence Alignment < EMBL-EBI (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) y se eligió las regiones no conservadas para el diseño manual de los cebadores. Se evaluó el contenido de GC, si existía o no complementariedad (self complementarity y self3' complementarity) con ayuda de los programas para análisis de cebadores OligoAnalyzer (<https://www.idtdna.com/pages/tools/oligoanalyzer>) y Primer-BLAST de NCBI (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Adicionalmente, se diseñaron cebadores para cada secuencia (completa) con ayuda de programas computacionales disponibles en la web, utilizando Prime3plus (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) y Primer-BLAST de NCBI (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>), eligiendo la mejor pareja de cebadores diseñados. Las secuencias de cada pareja de cebadores diseñados se enviaron a sintetizar (100 nmole DNA Oligo) en UNIPARTS, S.A. DE C.V.

5.2.9 CARACTERIZACIÓN DE DOS EFECTORES CANDIDATOS DE ALC

Los dos efectores candidatos identificados en el genoma de 'Ca. P. palmae' que se seleccionaron para el diseño de los cebadores específicos fueron analizados en búsqueda del péptido señal (PS), dominios transmembranas (DTM) y señales de localización nuclear (SLN). Para la identificación del PS se utilizaron los programas disponibles en la web SignalPv4.1, Phobius y TargetP v.2 (<https://services.healthtech.dtu.dk/services/SignalP-4.1/>, <https://phobius.sbc.su.se/>, <https://services.healthtech.dtu.dk/services/TargetP-2.0/>); para la búsqueda de DTM se utilizaron los programas TMHMM v2.0 y Phobius (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/> y <https://phobius.sbc.su.se/>); para la identificación de SLN se realizó un análisis en los programas NucPred y ELM (<https://nucpred.bioinfo.se/cgi-bin/single.cgi> y <http://elm.eu.org/search/>). Se determinó el contenido de aminoácidos y se determinó a que tribu pertenecen mediante la identificación de sus motivos *de novo* en el programa MEME (<https://meme-suite.org/meme/tools/meme>). Además, se realizó un análisis en BlastP de NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

5.2.10 AMPLIFICACIÓN DE EFECTORES DEL FITOPLASMA DE ALC MEDIANTE PCR DE PUNTO FINAL

Primero se evaluó cada pareja de cebadores con ADN de muestras de tallo de palmas de cocotero enfermas con ALC, seleccionando la pareja de cebadores donde se obtuviera amplificación para cada secuencia. Posteriormente, los cebadores seleccionados fueron evaluados mediante la técnica de PCR de punto final con ADNc de muestras de tallo, raíces y hojas de palmas de cocotero infectadas con fitoplasmas, usando una palma como control (sana). Las condiciones de amplificación que se utilizaron para cada pareja de cebadores se indica a continuación.

Para la pareja de cebadores diseñados para la secuencia 1 (cebadores 1A-Forward y 1A-Reverse) condiciones evaluadas: 94° por 2:00 min, 94° por 0:30 seg, 60° por 0:30 seg, 72° por 0:50 seg, GOTO_35X Step, 25°_∞.

Para la pareja de cebadores diseñados para la secuencia 2 (cebadores 2M-Forward y 2M-Reverse) condiciones evaluadas: 94° por 2:00 min, 94° por 0:30 seg, 50° por 0:30 seg, 72° por 0:50 seg, GOTO_35X Step, 25°_∞.

5.2.11 PURIFICACIÓN DE LAS BANDAS AMPLIFICADAS Y ENVÍO A SECUENCIACIÓN

La recuperación de las bandas amplificadas se realizó a partir de un gel de agarosa al 1.5 % donde se colocó 20µL de muestra de PCR en cada pocillo. El gel se sometió a electroforesis (90 volts, 200 AMP, 40 min), posteriormente las bandas amplificadas se visualizaron en el transiluminador UV/White Light Transilluminator Spectroline Bi-O-VISION™ y se extrajeron del gel. Una vez obtenidas, el exceso de agarosa se eliminó del ADN mediante el paquete comercial Gel DNA Recovery kit de Zymoclean™, siguiendo las instrucciones del fabricante. Las muestras se diluyeron en 15 µL de agua estéril (se procesaron por duplicado) y se cuantificaron en el espectrofotómetro Nanodrop One.

Las muestras fueron enviadas (junto con los cebadores previamente diseñados) para su secuenciación en el laboratorio Nacional de Biotecnología Agrícola, Medica y Ambiental (LANBAMA) de San Luis Potosí.

Posteriormente, para la confirmación de la amplificación de los efectores, en las palmas de cocotero con ALC, se realizaron alineamientos con los resultados obtenidos en la secuenciación de los fragmentos amplificados comparados con las secuencias originales identificadas en el genoma del fitoplasma del ALC, utilizando el programa ClustalW (<https://www.genome.jp/tools-bin/clustalw>). Para obtener la secuencia reverso complementario de cada secuencia reverse se utilizó el programa Reverse Complement (https://www.bioinformatics.org/sms/rev_comp.html) disponible en la web.

5.3 RESULTADOS

5.3.1. CONFIRMACIÓN DE FITOPLASMAS EN PALMAS DE *Cocos nucifera* L.

En las palmas con síntomas asociados con ALC se observó caída de frutos, necrosis de las inflorescencias y amarillamiento de las hojas inferiores, colectando distintas partes de cada una (hojas, raíces y tallo), además de colectar una palma sana (control) (Figura 5.1).

Se confirmó la presencia de fitoplasmas, mediante la prueba de qPCR con sonda TaqMan LY16S, en tres de las palmas colectadas (P1, P2 y P4) y la ausencia de estos patógenos en la palma sin síntomas utilizada como control (PC). De las muestras se recuperó ADN >350 ng/μL (Tabla 5.1) de acuerdo con lo indicado en el espectrofotómetro Nanodrop One (Thermo Scientific™).





Figura 5.1 Palmas de *Cocos nucifera* L. colectadas en Telchac Puerto. Palma control (PC), palma 1 (P1), palma 2 (P2), palma 3 (P3) y palma 4 (P4).

En la tabla 5.1 se observa la severidad de avance de la enfermedad de ALC en cada palma colectada, la concentración de ADN obtenida en la extracción y la confirmación de la presencia o ausencia de fitoplasmas en cada una. La palma 3 (P3) se descartó porque no se confirmó la presencia del fitoplasma.

Tabla 5.1 Confirmación (qPCR) y cuantificación de fitoplasmas en palmas de cocotero.

No. de palma colectada	Severidad de la enfermedad	Extracción de ADN del tallo (método CTAB)			qPCR sonda TaqMan LY16S (Ct)
		ADN ng/μl	A260/A280	A260/A230	
PC	---	386.4	1.50	1.61	ND
P1	Grado 4	1049.1	1.54	2.00	28.47
P2	Grado 4	380.1	1.61	1.56	27.23
P3	asociado al ALC	281.4	1.57	1.47	31.10
P4	Grado 4	355.6	1.45	1.52	22.10

PC: palma control (palma sana), P1: palma 1, P2: palma 2, P3: palma 3, P4: palma 4.
Ct Control-positivo: 25.38, Ct Control-negativo: 30.89

5.3.2 OBTENCIÓN DE ARN EN TALLO, RAÍCES Y HOJAS EN PALMAS CON ALC

En la tabla 5.2 se observa la cantidad de ARN obtenida para cada muestra procesada (tallo, hojas y raíz) de cada una de las palmas (PC, P1, P2 y P4).

Tabla 5.2 Concentración de ARN obtenida en tallo, raíces y hojas en palmas con ALC.

Muestra	Concentración de ARN		
	ng/ μ L	A260/A280	A260/A230
PC-T	46.2	2.08	1.50
P1-T	57.5	2.17	1.27
P2-T	35.7	1.86	1.12
P4-T	33.4	1.84	1.08
*PC-H	119.8	1.69	1.02
P1-H	95.2	2.02	1.69
P2-H	159.8	2.06	1.91
**P4-H	---	---	---
PC-R	136.3	1.73	1.06
*P1-R	265.9	1.49	1.19
*P2-R	98.2	1.44	1.00
P4-R	67.7	1.52	0.79

P: palma, **C:** control, **T:** tallo, **H:** Hojas, **R:** raíz.

*Se obtuvo el ARN utilizando el método de CTAB

**P4-H sin muestra (no se pudieron colectar las hojas)

Se realizó un gel de agarosa al 1 % para la visualización del ARN, observando en todas las muestras el ARN con trazas de ADN (Figura 5.2)

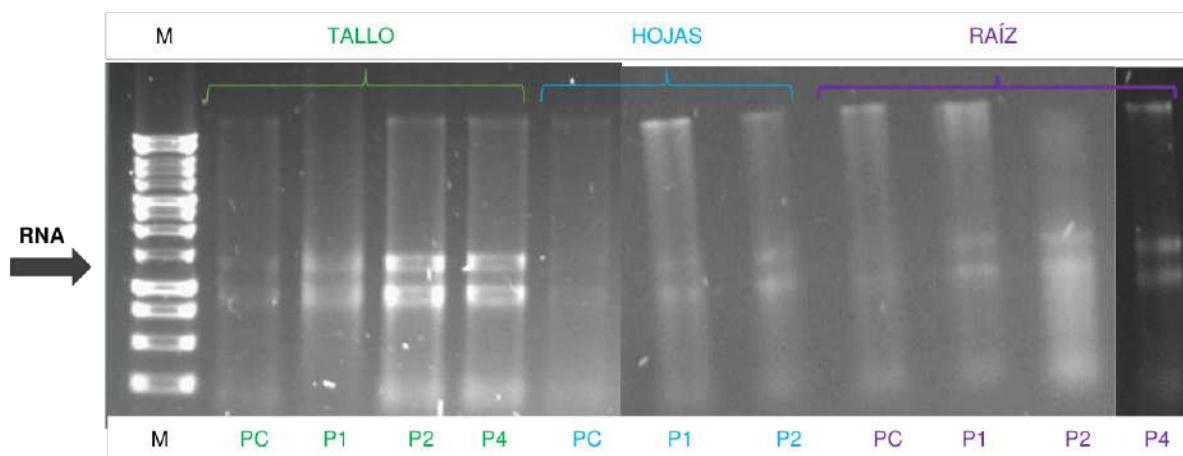


Figura 5.2 Visualización del ARN en gel de agarosa al 1 % en muestras de tallo, hojas y raíz en palmas con ALC. Se observa el ARN con trazas de ADN en todas las muestras. Marcadas en verde las muestras de tallo de la palma control (PC), palma 1 (P1), palma 2 (P2) y palma 4 (P4). Marcadas en azul las muestras de hojas de la palma control (PC), palma 1 (P1) y palma 2 (P2), (P4

no se colectaron las hojas). Marcadas en morado las muestras de raíz de la palma control (PC), palma 1 (P1), palma 2 (P2) y palma 4 (P4).

Posteriormente se realizó el procedimiento de eliminación del ADN y se corrió un nuevo gel de agarosa al 1 %. Se observó el ARN, confirmando la correcta eliminación del ADN en las muestras (Figura 5.3).

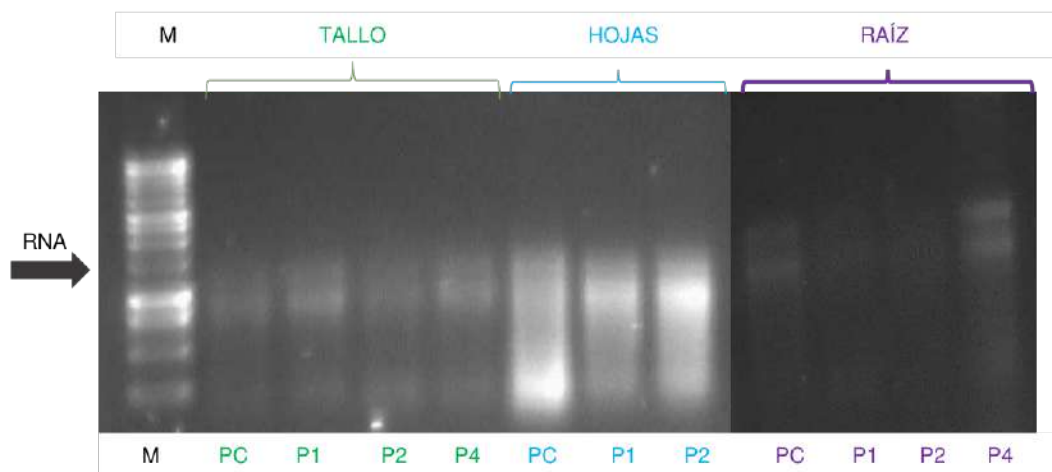


Figura 5.3 Observación del ARN (sin trazas de ADN). Gel de agarosa al 1 % con muestras de ARN de tallo, hojas y raíz en palmas con ALC (P1, P2 y P4) y palma control (PC), marcador (M).

Se continuó con la obtención del ADNc, a partir de las muestras de ARN, el cual se utilizó para realizar las pruebas para la amplificación de los efectores del fitoplasma de ALC con los cebadores (iniciadores) previamente diseñados a partir de las secuencias de efectores candidatos identificados en el genoma de '*Ca. P. palmae*'.

5.3.3 SELECCIÓN DE EFECTORES CANDIDATOS IDENTIFICADOS EN EL GENOMA DE '*Ca. P. palmae*' PARA EL DISEÑO DE CEBADORES

En el genoma *de novo* de '*Ca. P. palmae*' se identificaron 54 efectores candidatos. En la región de 12900-14600 pb del genoma se seleccionaron dos secuencias candidatas de

efectores, las cuales codifican para dos proteínas de 124 aa (375 nt) y 128 aa (387 nt). Estas dos secuencias de efectores predichos se utilizaron para diseñar los cebadores específicos, los cuales se enviaron a sintetizar y para realizar los análisis de expresión de los efectores en palmas enfermas con ALC. Las secuencias de aminoácidos y su respectiva secuencia de nucleótidos (para ambos efectores candidatos) se muestran a continuación en la tabla 5.3.

Tabla 5.3 Efectores candidatos identificados en el genoma de ‘Ca. P. palmae’

Efector candidato 1
Secuencia de nucleótidos_CDS (375 nt) >fig 33926.128.peg.194 hypothetical protein [Candidatus Phytoplasma 1 33926.128 12947..13321] ATGTTTAAAAATAAAAAAGCAATTTAGTAACATAATTTATATTGTTTAATAAGTTTTATAGTATTATTTTTAT AATTAGTAATGATAAAGTTATGGCAATGGATAAAAAATAACTCTGAAATAAGAAAAAATAAGGGAAAGAAA TTTGCCTACTACTAATGAACCTAGTGATGAAGAAGTTGATGAATTTTTTCGAAAATTAAAAAGAGCAAAA CAAAGTATTCAACAAATAAGAGAAAAACAATCCTCATTAAAAAAATCTTCTGTGCGAGAAATAGAAGCATA TATTTTATCAAGTATTAAGAGAAATCTGAAAAAGTAACAAAAAAGACAAAAAATAGATTTAAACGTTAT TC CAGAAGAAAATAA Secuencia de aminoácidos deducida (124 aa) >fig 33926.128.peg.194 hypothetical protein [Candidatus Phytoplasma 1 33926.128 12947..13321] MFKIKKQFSNIIYICLISFIVLFFIISNDKVMAMDKNNSEIRKNKGKEICTTTNEPSDEEVDEFFRK LKRAKQSIQQIRENNPHLKKSSVGEIEAYILSSIKRESEKSNKKRQKIDLNVIPEEK
Efector candidato 2
Secuencia de nucleótidos_CDS (387 nt) >fig 33926.128.peg.197 hypothetical protein [Candidatus Phytoplasma 1 33926.128 14152..14538] ATGTTTAAAAATAAAAAAGCAATTTAGTAACATAATTTATATTGTTTAATAAGTTTTATAGTATTATTTTTAT AATTAGTAATGATAAAGTTATGGCAATGGATAAAAAATAACTCTGAAATAAGAAAAAATAAGGGAAAGAAA TTTGCCTAATGAATCATCTGATGATGAAGAAGTTGATGATTTTATGGAATTATTTCAAATATTAAGCG CAAAACAAAAAATTCAACAAATAAGAGAAAAACATCCTCATTAAAAAGATCTTTCCCAAAGAGATTCTT TCAAAGAATTAGAAGCAGCAAAAATTTATTAGATCTTCGAAAAGGATCTAAAAAAGACAAAAAATAGAT T TAAACACTATTCCAGAAGAAAATAA Secuencia de aminoácidos deducida (128 aa) >fig 33926.128.peg.197 hypothetical protein [Candidatus Phytoplasma 1 33926.128 14152..14538] MFKIKKQFSNIIYICLISFIVLFFIISNDKVMAMDKNNSEIRKNKGKEICTNESSDDEEVDDFMEL FQILKSAKQKIQIREKHPHLKRSFPKEDSFKELEAAKILLDLRKGSKKRQKIDLNTIPEEK

En la tabla 5.4 se muestra la pareja de cebadores diseñados para cada secuencia de efectores, utilizando los programas disponibles en la web: Primer-BLAST y Primer3Plus, así como la pareja de cebadores diseñada de manera manual. Se solicitaron sintetizar ambas parejas de cebadores, para cada gen.

Tabla 5.4 Cebadores (iniciadores) específicos diseñados para cada secuencia de efectores candidatos identificados en el genoma de 'Ca. P. palmae'.

Pareja de cebadores diseñados para la secuencia 1 (375 nt/124 aa)		
	Diseño Manual (1M*)	Diseño con Primer-BLAST (1A*)
Primer Forward	CTACTAATGAACCTAGTG	TGCACTACTACTAATGAACCTAGTG
Primer Reverse	ATATGCTTCTATTTCTCCG	TGCTTCTATTTCTCCGACAGAAG
Inicio-F	155	148
Inicio-R	288	285
Producto	134 pb	138 pb
Pareja de cebadores diseñados para la secuencia 2 (387 nt/128 aa)		
	Diseño Manual (2M*)	Diseño con Primer3Plus (2A*)
Primer Forward	CATCTGATGATGAAGAAG	GGGAAAGAAATTTGCACTAATGA
Primer Reverse	AGATCCTTTTCGAAGATC	TTTTTGCTGCTTCTAATTCTTTG
Inicio-F	161	136
Inicio-R	339	313
Producto	179 pb	178 pb

*Los cebadores o iniciadores diseñados de modo manual para la secuencia 1 y secuencia 2 se denominaron 1M y 2M, respectivamente; para la pareja de cebadores diseñados con ayuda de programas computacionales para la secuencia 1 y secuencia 2, se denominaron 1A y 2A, respectivamente.

5.3.4 CARACTERIZACIÓN DE 2 EFECTORES CANDIDATOS SELECCIONADOS DE ALC

Las dos secuencias identificadas como efectores candidatos en el genoma del fitoplasmas del ALC y seleccionadas para el diseño de los cebadores codifican para dos proteínas de 124 aa y 128 aa con 71.3 % de identidad entre ellos. Se realizó un análisis *in silico* para caracterizar estos dos efectores candidatos y se determinó la presencia de un PS (programas SignalPv4.1 y TargetPv.2.0), ausencia de DTM (programas TMHMMv.2.0 y Phobius) e identificación de SLN para ambas secuencias (programas NucPred y ELM) (Tabla 5.5). El efector candidato 1 de acuerdo con los análisis bioinformáticos corresponde a una preproteína de 124 aa con un péptido señal (PS) de 33 aa, generando una proteína madura de 91 aa (10.1 kDa.) de longitud; para el efector candidato 2 los análisis indican una preproteína de 128 aa y posee un PS de 33 aa, generando una proteína madura de 95 aa (10.55 kDa.) de longitud (Figura 5.4). Además, se logró identificar una SLN en ambos efectores cuando se realizaron los análisis en NucPred y ELM (Figura 5.4).

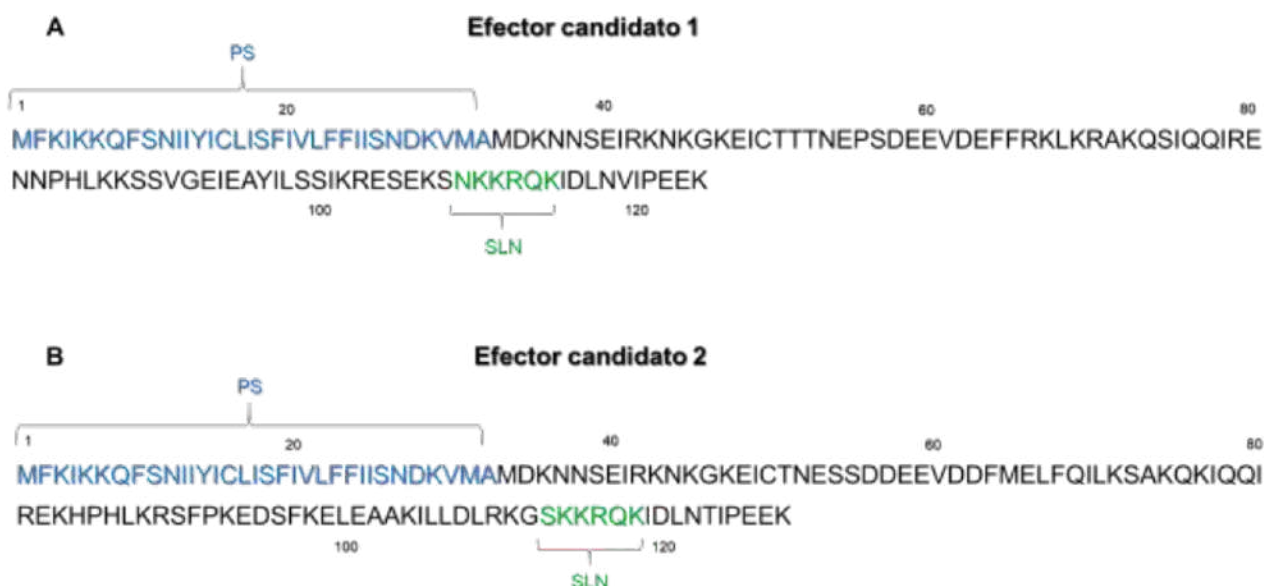


Figura 5.4 Secuencias de proteínas de los efectores candidatos del fitoplasma de ALC. (A) Secuencia proteica del efector candidato 1, marcado en azul la secuencia del péptido señal (PS) y en verde la secuencia señal de localización nuclear (SLN). (B) Secuencia proteica del efector candidato 2, marcado en azul la secuencia del péptido señal (PS) y en verde la secuencia señal de localización nuclear (SLN).

Se determinó el contenido de los aminoácidos lisina, leucina, asparagina, cisteína y triptófano, observando entre 6 hasta 21 residuos para los tres primeros y de 2 a 0 residuos para cisteína y triptófano (Tabla 5.5); el escaso o nulo contenido en estos dos últimos aminoácidos es una característica común en la mayoría de los efectores de fitoplasmas (resultados observados en este trabajo, capítulo 2). Adicionalmente, se encontró que estos dos efectores candidatos pertenecen a la tribu 1. Además, cuando se comparó con la base de datos previamente creada en este trabajo, para ambos se observó homología con proteínas semejantes a SAP11, SAP54 y PHYLL1 (Tabla 5.5). En el análisis con BlastP de NCBI se observó homología con la familia de proteínas SVM, entre los que se encuentra el fitoplasma de la palma Phoenix de Texas con arriba del 50 % de similitud.

Tabla 5.5 Análisis *in silico* de dos efectores candidatos de 'Ca. P. palmae', fitoplasma del ALC.

Efectores Candidatos de 'Ca. P. palmae'										
	Contenido de aminoácidos					PS	DTM	NLS	Tribu	Homología con la BD
	K	L	N	C	W					
Efector candidato 1	19	6	10	2	0	Sí	No	Sí	1	SAP11-like, SAP54-like, Phyl1
Efector candidato 2	21	10	7	2	0	Sí	No	Sí	1	SAP11-like, SAP54-like, Phyl1

K: lisina, L: leucina, N: Asparagina, C: cisteína y W: triptófano

5.3.5 AMPLIFICACIÓN DE EFECTORES DE FITOPLASMA DE ALC

Para la amplificación de los efectores en palmas de *C. nucifera* infectadas con fitoplasmas se realizó la prueba de PCR de punto final y se evaluó cada pareja de cebadores diseñados para cada secuencia. Se realizó una primera prueba con muestras de ADN de palmas enfermas con ALC y se observó mejores resultados con la pareja de cebadores denominados 1A (forward y reverse) para la secuencia 1 y la pareja de cebadores denominados 2M (forward y reverse) para la secuencia 2.

Se seleccionaron las parejas de cebadores 1A (forward y reverse) y 2M (forward y reverse) (ver tabla 5.4, mostrada en la sección anterior) para evaluar con el ADNc en todas las muestras de tallo, hojas y raíces de las palmas PC, P1, P2 y P4.

Se corrió un gel de agarosa al 1.5 % en el cual se logró observar amplificación en dos de las muestras con ALC. Sin embargo, solamente se consiguió amplificación con la pareja de cebadores 2M, con los cuales amplificaron una banda intensa con una de las muestras de hojas con la palma 2 (P2-H) y una banda tenue para una de las muestras de tallo con la Palma 1 (P1-T) por debajo de 250 pb (Figura 5.5). De acuerdo con los cebadores diseñados el producto esperado es de 179 pb (Tabla 5.4), lo que sugirió que sí hubo amplificación de los efectores de 'Ca. P. palmae'. En la figura 5.5 se observa las muestras en donde hubo amplificación usando las muestras de ADNc, se evaluaron dos controles positivos y dos negativos y se corrió el gel de agarosa junto con una de las muestras donde no hubo amplificación (palma 2-raíces, P2-R).

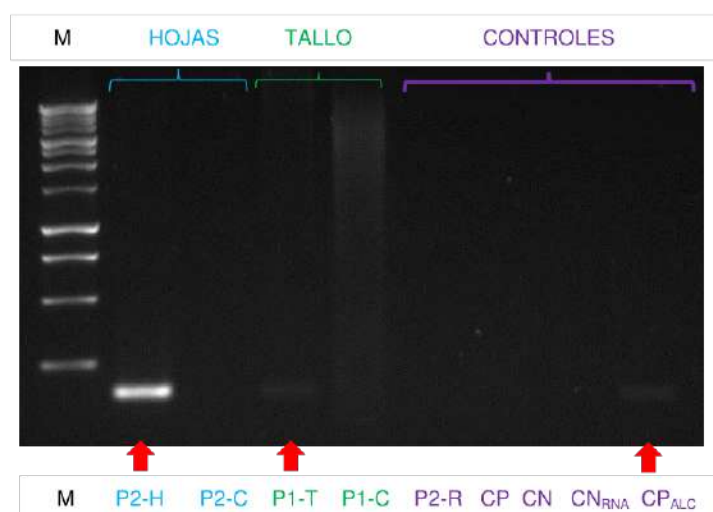


Figura 5.5 Amplificación de efectores de ‘*Ca. P. palmae*’ en palmas con ALC en gel de agarosa al 1.5 % con ADNc. Señalados con flechas rojas (carriles 2, 4 y 10) donde hubo amplificación con la pareja de cebadores 2M (diseñados para la secuencia 2) con la muestra de hojas de la palma 2 (P2-H), con la muestra de tallo de la palma 1 (P1-T) y con el control positivo (segundo control positivo evaluado con ADN de una palma infectada con ALC). En el tercer y quinto carril se observa la palma 2 control (P2-C) y palma 1 control (P1-C). En el sexto carril se observa una muestra de raíz de la palma 2 (P2-R) donde no hubo amplificación. Se evaluaron dos controles positivos y dos negativos (carriles 7-10), control positivo cDNA obtenido del control del RNA del Kit utilizado (CP), control negativo agua (CN), segundo control negativo RNA (CN_{RNA}), segundo control positivo ADN de una palma infectada con ALC (CP_{ALC}), en este último también se observó amplificación.

5.3.6 SECUENCIAS DE FRAGMENTOS AMPLIFICADOS CON LAS MUESTRAS P1-T Y P2-H

Las bandas de las dos muestras que amplificaron de la palma1 con la muestra de tallo (P1-T) y de la palma 2 con la muestra de hojas (P2-H) se recuperaron y se cuantificaron en el nanodrop. Se obtuvo una concentración de 21.7 ng/μL para la muestra P1-T y una concentración de 36.6 ng/μL para la muestra P2-H en un volumen de 26 μL para cada una; en los duplicados se obtuvo menor cantidad de muestra, teniendo una concentración de 10.8 ng/μL y 11.5 ng/μL en un vol. de 13 μL para P1-T y P2-H, respectivamente. Posteriormente, fueron enviadas para su secuenciación (en ambos sentidos) en el

laboratorio Nacional de Biotecnología Agrícola, Medica y Ambiental (LANBAMA) de San Luis Potosí, junto con los cebadores diseñados 2M-Forward y 2M-Reverse.

En la tabla 5.6 se observan las secuencias de los fragmentos P1-T y P2-H amplificados (forward, reverse y reverseo complementario).

Tabla 5.6 Secuencia de los fragmentos amplificados con la muestra de la palma 1 con tallo (P1-T) y de la palma 2 con hojas (P2-H).

Secuencias del fragmento amplificado de la palma 1 con tallo (P1-T)
Fragmento Forward_Longitud: 289 pb >P1T-2MF CGAACCATTATTGTCAAATATTAAGCGCAAACAAAAATTCAACAAATAAGAGAAAAACATCCTCATT TAAAAAGAGCAATAACCCAAAGAAGATTCTTTCAAGAATTAGAAGCACCCAAAATTTTATTAGATCTTCG AAAAGGATCTAAAGAAAGTGTTTGCTGCTTCTCATTCTTTGAAGGAATCTTCTTTCCGAAAAGATCTTTTT AAATGAGGATATTTTCTCTATTTGATGAATTTTGGCAGGCTTTTCTCTTTGAATATTTAAATCAAACC CCCCCA
Fragmento Reverse_Longitud: 154 pb >P1T-2MR GTACAGCGGTCCAATTCCTTTTGAAGATCTTCTTTGGGAAAAGGATCTTTTTAAATGAGGATGTTTTTC TCTTATTTGTTGAATTTTTGTTTTCGCTTTTAAATATTTGAAATAATTCCATAAAATCATCAACTTCTTCA TCATCAGATGA
Reverso Complementario (programa Reverse Complement) >P1T-2MR_ TCATCTGATGATGAAGAAGTTGATGATTTTATGGAATTATTTCAAATATTAAGCGCAAACAAAAAT TCAACAAATAAGAGAAAAACATCCTCATTTTAAAAGATCCTTTTCCCAAAGAAGATCTTCAAAGGAATT GGACCGCTGTAC
Secuencias del fragmento amplificado de la palma 2 con hojas (P2-H)
Fragmento Forward_Longitud: 311 pb >P2H-2MF GAGGAATATGTCAATATTAAGCGCAAACAAAAATTCAACAAATAAGGAGAAAAACATCCTCATTAA AAAGATCTTTTCCCAAAGAAGATTCTTTCAAAGAATTAGAAGCAGCAAAAATTTTATTAGATCTTCGAAAA GGATCTGAATAAAGTTTTTGCTGCTTCTAATTCTTTGAAAGAATCTTCTTTGGGAAAAGATCTTTTTAAAT GAGGATGTTTTCTCTTATTTGTTGAATTTTTGTTTTCGCTTTTAAATATTTGAAATAATTCATAAAATC ATCAACTTCTTCTCAAAGGAGAGCTA
Fragmento Reverse_Longitud: 304 pb >P2H-2MR TGGTCACTATTCTTTGAAGATCTTCTTTGGGAAAAGATCTTTTTAAATGAGGATGTTTTCTCTTATTTGTT GAATTTTTGTTTTGCGCTTTTAAATATTTGAAATAATTCATAAAATCATCAACTTCTTCATCATCAGATA AAAAGTTTTTGCTGCTTCTAATTCTTTGAAAGAATCTTCTTTGGGAAAAGATCTTTTTAAATGAGGATGTT TTTCTCTTATTTGTTGAATTTTTGTTTTCGCTTTTAAATATTTGAAATAATTCATAAAATCATCAACTTCT TCACCTCAAGGAGGGTA
Reverso Complementario (programa Reverse Complement) >P2H-2MR TACCCTCCTTGAGGTGAAGAAGTTGATGATTTTATGAATTATTTCAAATATTAAGCGCAAACAAAAA ATTCACAAATAAGAGAAAAACATCCTCATTTAAAAGATCTTTTCCCAAAGAAGATTCTTTCAAAGAATT AGAAGCAGCAAAAACTTTTATCTGATGATGAAGAAGTTGATGATTTTATGGAATTATTTCAAATATTTAA AAGCGCAAACAAAAATTCAACAAATAAGAGAAAAACATCCTCATTTAAAAGATCTTTCCCAAAGAAG ATCTTCAAAGAATAGTGACCA

5.3.7 CONFIRMACIÓN DE LA AMPLIFICACIÓN DE EFECTORES DE 'Ca. P. palmae' EN PALMAS CON ALC

Para confirmar que los fragmentos que amplificaron corresponden a efectores del fitoplasma causal del amarillamiento letal se realizó un alineamiento de las secuencias forward y reverso complementario de P1-T y P2-H con la secuencia 2 identificada como efector candidato del genoma *de novo* de 'Ca. P. palmae' utilizando el programa ClustalW (<https://www.genome.jp/tools-bin/clustalw>).

En la figura 5.6 se muestra el alineamiento realizado para P1-T con el efector candidato, observando mejor complementariedad con la secuencia reverso complementario (marcado en colores verde y amarillo).



Figura 5.6 Alineamiento con ClustalW del efector candidato de 'Ca. P. palmae' con la secuencia de la palma 1 con la muestra de tallo (P1-T). Se observa el alineamiento realizado entre la secuencia del efector candidato identificado en el genoma del fitoplasma del ALC con la secuencia reverso complementario del fragmento amplificado de P1-T, marcado en asteriscos las áreas donde hay similitud, señalado por colores verde y amarillo las regiones donde hay mayor complementariedad.

En la figura 5.7 se muestra el alineamiento realizado para P2-H con el efector candidato, se observó mejor complementariedad con la secuencia reverso complementario (marcado en colores verde y amarillo).

Los alineamientos nos permitieron comprobar que los fragmentos que amplificaron fue un efector del fitoplasma de ALC (denominado efector candidato 2, de los efectores identificados en el genoma *de novo*).

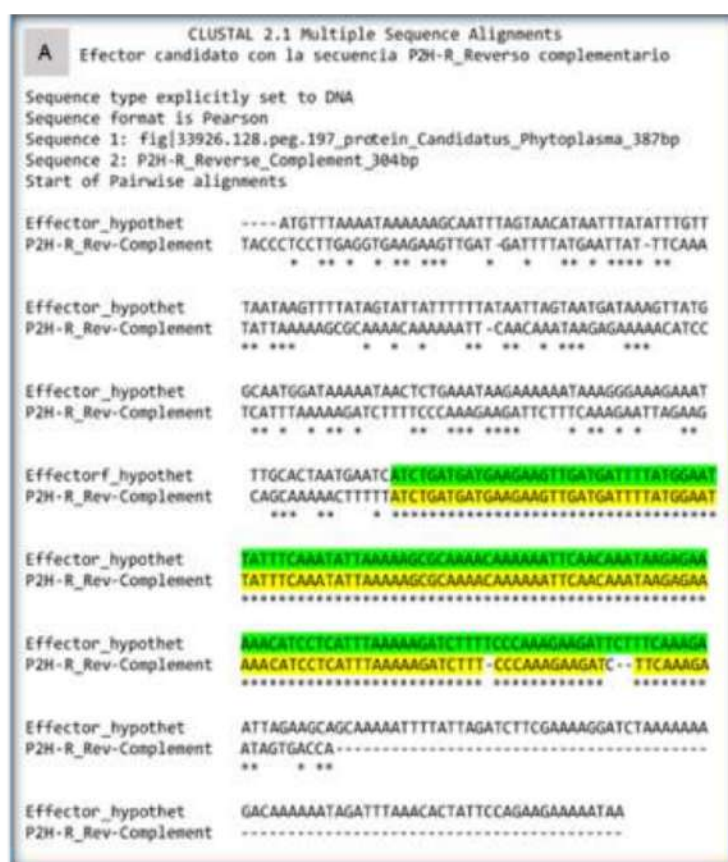


Figura 5.7 Alineamiento con ClustalW del efector candidato de 'Ca. P. palmae' con la secuencia de la palma 2 con la muestra de hojas (P2-H). Se observa el alineamiento realizado entre la secuencia del efector candidato identificado en el genoma del fitoplasma del ALC con la secuencia reverso complementario del fragmento amplificado de P2-H, marcado en asteriscos las áreas donde hay similitud, señalado por colores verde y amarillo las regiones donde hay mayor complementariedad.

5.4 DISCUSIÓN

En este trabajo, se realizó la caracterización de dos efectores fitoplasmáticos de '*Ca. P. palmae*' asociados con la enfermedad del amarillamiento letal en cocotero (Oropeza *et al.*, 2020). Estos dos efectores se obtuvieron mediante la secuenciación del genoma del fitoplasma de ALC. Se seleccionaron por su alta homología con SAP11 de la cepa de fitoplasma del amarillo del áster, escoba de bruja (fitoplasma AY-WB), uno de los pocos efectores fitoplasmáticos que parecen estar conservados en diferentes especies de fitoplasma (Chang *et al.*, 2018).

Se realizó el estudio de expresión de los dos efectores del fitoplasma de ALC, el cual se logró a través de cebadores específicos diseñados desde secuencias de efectores candidatos identificados en el genoma de '*Ca. P. palmae*'. Para la amplificación de los efectores de fitoplasmas responsables del ALC se colectaron distintas partes de palmas enfermas con AL (tallo, hojas y raíces), para obtener el ADNc y realizar la prueba de PCR de punto final con los cebadores específicos diseñados. Confirmando la amplificación de un efector de ALC (correspondiente a la secuencia del efector candidato 2), de acuerdo con la complementariedad observada al realizar los alineamientos en el programa ClustalW, en dos muestras colectadas, en una muestra de tallo con la palma 1 (P1-T) y en una muestra de hojas con la palma 2 (P2-H).

Se han identificado y caracterizado varios efectores en fitoplasmas, como SAP11 (Bai *et al.*, 2009) de '*Ca. P. asteris*', 16Srl-A, que modula la expresión de los factores de transcripción TEOSINTE BRANCHED1, CYCLOIDEA, PROLIFERATING CELL FACTORS (TCP) relacionados con CINCINNATA (CIN), lo que resulta en niveles bajos de ácido jasmónico endógeno en plantas infectadas (Sugio *et al.*, 2011a). TENGU, identificado en el genoma del fitoplasma Onion Yellow (OY, '*Ca. P. asteris*', 16Srl-B), altera la expresión de los factores de respuesta a la auxina (ARF), lo que conduce a niveles bajos de auxinas y ácido jasmónico en plantas, desarrollo floral deficiente y esterilidad (Minato *et al.*, 2014). SAP54 de '*Ca. P. asteris*' (AY-WB) degrada los factores de transcripción homeóticos del dominio MADS mediados por las proteínas de la familia RADIATION SENSITIVE23 (RAD23), alterando el desarrollo floral a tejido vegetativo y causando síntomas similares a la filodia (MacLean *et al.*, 2011). PHYL1 (homólogo de SAP54) de '*Ca. P. asteris*' (OY) degrada los factores de transcripción del dominio MADS a través de la vía de ubiquitinación, produciendo síntomas similares a la filodia (Maejima *et al.*, 2014). SAP05 de '*Ca. P. asteris*'

(AY-WB) degrada las familias de factores de transcripción SPL y GATA de manera independiente de ubiquitina, los SPL están involucrados en la transición de la fase juvenil a la adulta; y los GATA regulan el desarrollo de los órganos florales y las hojas, generando plantas estériles y promoviendo síntomas similares a los de la escoba de bruja (Huang *et al.*, 2021).

Los efectores predichos caracterizados en este estudio codificaron dos proteínas de 124 aa (375 nt) y 128 aa (387 nt) respectivamente. El efector candidato 1, según el análisis bioinformático, corresponde a una preproteína de 124 aa con un péptido señal (PS) de 33 aa, generando una proteína madura de 91 aa (10,1 kDa) de longitud; para el efector candidato 2, los análisis indican una preproteína de 128 aa y un PS de 33 aa, generando una proteína madura de 95 aa (10,55 kDa) de longitud. Además, la identificación de PS, SLN, ausencia de DTM y una longitud <400 aa en su secuencia permite clasificarlos dentro del tipo clásico o canónico, ya que cumplen los criterios reportados en la literatura. Adicionalmente, se observó homología con proteínas semejantes a SAP11, este efector ha sido bien caracterizado, conteniendo 121 aa, así como PS y motivos SLN (Bai *et al.*, 2009)

Los candidatos a efectores de este estudio pertenecen a la tribu uno, que esta ampliamente distribuida entre los fitoplasmas (Carreón-Anguiano *et al.*, 2023). Adicionalmente, el análisis *in silico* de los efectores candidatos seleccionados reveló un contenido bajo o nulo de cisteína y triptófano, características únicas encontradas previamente al analizar la base de datos con 738 efectores de fitoplasmas (resultados de este trabajo con el art. publicado en el capítulo 2). Aunque el contenido de lisina, leucina y asparagina no fue sobresaliente (<25 residuos) en estos efectores candidatos, es importante considerar que no todos los efectores de fitoplasmas cumplen con este criterio, mostrando un alto contenido (>25 residuos) de lisina el ~50 % de los efectores, en cuanto al contenido de leucina y asparagina solo el ~20 % y ~15 %, respectivamente, poseen esta característica (Carreón-Anguiano *et al.*, 2023), tal es el caso del efector Tengu, de '*Ca. P. asteris*', responsable del amarillamiento de la cebolla (OY, por sus siglas en inglés) cuyo aminoácido más enriquecido fue la leucina, teniendo 11 residuos (resultados observados en el análisis *in silico*).

En este trabajo, se analizó la expresión de efectores candidatos del fitoplasma de ALC mediante la técnica de PCR con muestras de ADNc de palmas infectadas con fitoplasmas. Sin embargo, solo se determinó la expresión del efector candidato 2. No se detectó el

efector candidato 1, posiblemente debido a la baja cantidad de ADNc sintetizado o a que no se expresó en esta etapa de progreso de ALC. Las palmas analizadas se encontraban en la etapa cuatro de progreso de ALC, lo que indica que se encuentran en la etapa intermedia de la enfermedad, donde el amarillamiento de las hojas atrae al insecto vector (Narváez *et al.*, 2018). Los ortólogos de SAP11 podrían estar activos durante esta etapa, como el efector candidato 2.

Es necesario realizar más estudios para comparar distintas partes de una misma palma enferma con ALC y poder determinar si hay diferencia en el nivel de expresión de estos efectores, así como coleccionar distintas palmas con diferente grado de avance de la enfermedad.

CAPÍTULO VI

DISCUSIÓN, CONCLUSIONES GENERALES Y PERSPECTIVAS

6.1. DISCUSIÓN

El amarillamiento letal del cocotero (ALC) es una enfermedad que causa importantes pérdidas a nivel mundial en el cultivo de la palma de coco, una enfermedad ocasionada por patógenos llamados fitoplasmas, que inducen síntomas como el amarillamiento y cambios morfológicos (Ramos-Hernández, *et al.*, 2018). Los fitoplasmas secretan proteínas de virulencia denominados efectores, los cuales degradan proteínas importantes del hospedante ocasionando los cambios morfológicos (Padilla-Ramos *et al.*, 2019; Hogenhout *et al.*, 2009). Por lo tanto, estudiar los efectores ayudaría a comprender mejor el mecanismo de interacción de estos patógenos.

En este trabajo se estudiaron los efectores de los fitoplasmas en general y luego los asociados al amarillamiento letal en particular, para ello se utilizaron cuatro estrategias; las cuales consistieron en: a) un análisis *in silico* de los efectores de fitoplasmas disponibles en las bases de datos, buscando características conservadas (firmas) en las secuencias de estos efectores; b) desarrollo de una herramienta computacional para la identificación de estas proteínas efectoras; c) secuenciación del genoma *de novo* del fitoplasmas asociado al ALC e identificación de sus efectores y d) un análisis de expresión de efectores selectos de 'Ca. P. palmae' en palma de coco enfermas con ALC.

En la primera estrategia se realizó un estudio de caracterización *in silico* con los efectores de fitoplasmas verdaderos (validados) y los predichos (candidatos) que se lograron obtener de la base de datos (eliminando los redundantes). Al realizar el análisis bioinformático se observó que los efectores de fitoplasmas se clasificaban en dos grupos; se encontró que muchas de estos efectores son proteínas pequeñas, de menos de 400 aminoácidos, con presencia de péptido señal y sin presencia de dominios transmembranales, lo que coincide con lo reportado en la literatura para efectores de fitoplasmas (Singh y Lakhanpaul, 2020; Lamilla Monje y Franco Lara, 2019; Bai *et al.*, 2009), denominados como efectores clásicos o canónicos. Sin embargo, los resultados obtenidos también demostraron que existen efectores de fitoplasmas de tipo no-canónico. El último tipo de efectores se han encontrado

en otro tipo de patógenos, como en hongos y oomicetos que se han descubierto nuevos efectores validados con tamaños mayores a 400 aminoácidos o con TMD, o sin un péptido señal (Carreón-Anguiano *et al.*, 2022). Actualmente, se ha reportado efectores con TMD en el fitoplasma de la Flavescencia dorada (Debonneville *et al.*, 2022), así como efectores secretados que carecen de un péptido señal (SP) en '*Ca. P. ziziphi*' (Gao *et al.*, 2023), evidenciando la existencia de efectores no clásicos en los fitoplasmas.

Por otro lado, se observó que a diferencia de los efectores canónicos fúngicos que son ricos en cisteína (Carreón-Anguiano *et al.*, 2020), los efectores de fitoplasma no lo son. Cuando se intentó usar el algoritmo EffHunter (Carreón-Anguiano *et al.*, 2020) se observó que éste no reconoce a los efectores verdaderos de fitoplasma SAP11, SAP54 y TENGU. Al hacer el análisis manual, se observó que los criterios de pequeño tamaño, presencia de péptido señal, ausencia de dominios transmembranales, sí coinciden entre estos efectores de fitoplasmas y los efectores canónicos fúngicos, pero la diferencia fue la composición en términos de cisteína, dado que, para los efectores de fitoplasmas no resulta una característica estructural importante. Por ejemplo, los efectores SAP11 y SAP54 no contienen cisteína (cero) y TENGU contiene sólo una, mientras que el programa EffHunter selecciona proteínas con al menos 4 cisteínas. Esta descripción no solo es un progreso de nuestro grupo en el conocimiento de los efectores de fitoplasmas, sino que estos resultados representan una contribución a la efectorómica de los fitoplasmas.

Algunas investigaciones han reportado en los genomas de los fitoplasmas la pérdida de genes de distintas rutas metabólicas, como por ejemplo de la síntesis de diferentes aminoácidos (Bai *et al.*, 2006; Oshima *et al.*, 2004). Es probable que el grado de evolución de estas proteínas esté relacionado con la ausencia o el bajo contenido de algunos aminoácidos. Aunque en los efectores de fitoplasmas no se han reportado miembros de la familia del sistema de secreción tipo 3 (SST3) que poseen algunas bacterias (Oshima *et al.*, 2004), estos patógenos poseen el sistema Sec, el cual les sirve para la secreción e integración de proteínas en el citoplasma de la célula hospedante (Oshima *et al.*, 2013). En el caso de los hongos, muchos efectores son realmente secretados y tienen que resistir en el ambiente del apoplasto (Navarrete *et al.*, 2021; Tanaka y Kahmann, 2021; Rocafort *et al.*, 2020; Li *et al.*, 2020; Wang y Wang, 2018b).

Los análisis de composición de aminoácidos en los efectores de fitoplasmas revelaron que los residuos de asparagina, lisina y leucina están evidentemente enriquecidos, contrastante

con un escaso o nulo contenido de cisteína y triptófano; además, se observó que la cantidad de aminoácidos es independiente de la longitud de la misma.

La asparagina es un aminoácido imprescindible en la *N*-glicosilación, una modificación postraducciona importante en la patogénesis bacteriana, durante la cual el azúcar se enlaza al grupo γ -amido de la asparagina. La glicosilación es un proceso involucrado en la protección contra la degradación proteolítica, la variación antigénica, adhesión, aumento de la solubilidad y ensamble de proteínas (Córdova, 2009). Otros estudios resaltan la importancia de este aminoácido en la interacción proteína-proteína, para la formación del complejo antígeno anticuerpo (Yokota *et al.*, 2010). Por otro lado, Singh y Lakhanpaul (2020), al estudiar una proteína semejante a SAP54 encontraron que la asparagina es un aminoácido funcionalmente importante dentro de la secuencia de este efector de fitoplasmas, ya que se encuentra bajo una fuerte selección purificadora (normalizadora), ocasionando que se conserve y se mantenga en mayor proporción dentro de su secuencia.

Iwabuchi *et al.* (2020), al estudiar una familia de efectores de fitoplasmas causantes de filodia, denominados filógenos, observaron que los residuos de asparagina y lisina se conservaron, además de demostrar la importancia de la asparagina y de ambos residuos, asparagina-lisina, en la inducción de la filodia. Por otro lado, la lisina es un residuo importante en las señales de localización nuclear (NLS) (Lu *et al.*, 2021), este tipo de secuencia señal ha sido reportada en efectores de fitoplasmas como SAP11 (Sugio *et al.*, 2014).

Con respecto al residuo de leucina encontrado también en mayor proporción en estas proteínas efectoras, es sabido que en algunos microorganismos patógenos como los oomicetos, poseen efectores de tipo RXLR, los cuales se caracterizan por poseer un dominio de aminoácidos de arginina-leucina-arginina en el extremo C-terminal, los cuales a menudo suprimen la resistencia vegetal (Padilla-Ramos *et al.*, 2019; Izarra y Lindqvist-Kreuze, 2016). La leucina es un aminoácido imprescindible para este tipo de receptores, lo que lleva a considerar que la leucina es también un residuo importante en los efectores fitoplasmáticos, congruente con su mayor proporción en las secuencias. Es probable que la leucina en estas proteínas esté relacionada con la virulencia.

Recientemente Singh y Lakhanpaul (2020), analizaron una proteína semejante a SAP54 denominada S54LP, causante de filodia en ajonjolí (*Sesamum indicum* L.). La estudiaron

junto con distintos ortólogos de SAP54, reportando un alto contenido de asparagina, leucina y lisina, como se encontró en el presente estudio, ampliando el análisis a los efectores conocidos y predichos. Sus resultados en cisteína y triptófano también contrastaron con los otros aminoácidos, pues no encontraron estos residuos, similar de lo observado en nuestros análisis. El presente estudio no solo ayuda a corroborar esos hallazgos, pero también refuerza que los efectores de fitoplasmas comparten características de composición de aminoácidos en su secuencia.

Lo anterior sugiere el grado de evolución de las proteínas efectoras de fitoplasmas al carecer de determinados aminoácidos (cisteína y triptófano) en su secuencia, así como la importancia funcional de otros residuos (lisina, leucina y asparagina), lo que explicaría el alto contenido de estos aminoácidos dentro de estos, que a su vez se relacionaría con la virulencia de estas proteínas efectoras.

Otra característica no reportada previamente y que parece ser estructuralmente importante en los efectores de fitoplasmas es la presencia de motivos ELMs o SLiMs. Los motivos lineales cortos (SLiM, de acuerdo con sus siglas en inglés) o motivos lineales eucariotas (ELM, de acuerdo con sus siglas en inglés) son péptidos lineales cortos, de menos de 10 aminoácidos, con un patrón de secuencia específico, que son de importancia crucial para numerosos procesos biológicos al mediar las interacciones proteína-proteína (Ruhanen *et al.*, 2014). Se ha descrito que muchos motivos en las proteínas efectoras de virulencia de los patógenos imitan motivos eucariotas, lo que permite al patógeno secuestrar, desregular y controlar los procesos intracelulares de sus hospedantes (Kumar *et al.*, 2020; Ruhanen *et al.*, 2014).

Por otro lado, los efectores que no comparten una identidad de secuencia general alta, pero comparten motivos, dominios o estructuras tridimensionales similares, pueden compartir funciones similares y se denominan “ortólogos funcionales”; razones por lo cual se buscaron los motivos ELMs/SLiMs. Los resultados en este trabajo mostraron 10 motivos ELMs/SLiMs mayoritarios involucrados en diferentes procesos celulares. Los motivos encontrados permiten conocer más sobre los mecanismos de virulencia de los efectores de fitoplasmas y comprender mejor sobre estos patógenos; como por ejemplo el motivo identificado LIG_DCNL_PONY_1 que se une a la enzima conjugadora de ubiquitina, indicando que los efectores de fitoplasma que contienen este motivo pueden regular sus objetivos mediante proteólisis. Este tipo de actividad se ha descrito en los ya caracterizados efectores SAP54

y SAP05 (Huang *et al.*, 2021; MacLean *et al.*, 2014); otro motivo encontrado es el motivo LIG_PCNA_PIPBox_1 que es un dominio de localización nuclear; se ha descrito que efectores como SAP11 y SWP1 se dirigen a las proteínas nucleares del hospedante (Bai *et al.*, 2023; Strohmayer *et al.*, 2021), entre otros motivos que fueron identificados.

Sin embargo, un factor que ha limitado la caracterización en los efectores de fitoplasmas, es la falta de conservación y cada efector ha sido estudiado individualmente, siendo escasas las familias (o tribus) basadas en homología que se han descrito (ejm. SAP11-like, SAP54-like) (Drcelic *et al.*, 2024; Singh y Lakhanpaul, 2020). Además, la homología también suele ser baja o inexistente entre los efectores de otros reinos taxonómicos (Jones *et al.*, 2021), por lo cual otra estrategia utilizada para conocer más sobre los efectores de fitoplasmas fue la búsqueda de motivos *de novo*; encontrando 50 motivos principales y permitiendo clasificar a los efectores en 15 tribus o familias; lo que ayudaría al estudio por grupos de los efectores. La clasificación de los efectores en tribus (familias de proteínas que comparten motivos) es útil ya que al compartir motivos podría estar relacionado con funciones comunes.

La segunda estrategia utilizada y una gran aportación de este trabajo fue la creación del algoritmo PhyEffector, una herramienta para la identificación de efectores de fitoplasmas. Al carecer de un programa computacional para la identificación de estas proteínas de virulencia, distintos autores recurrían a diferentes programas como SignalP y TMHMM (Anabestani *et al.*, 2017; Bai *et al.*, 2006), en busca de proteínas que cumplan con los criterios establecidos que definen a un efector, proteínas pequeñas (<400 aa), presencia de PS y ausencia de DTM (Huang *et al.*, 2021; MacLean *et al.*, 2014; Maejima *et al.*, 2014; Hoshi *et al.*, 2009; Sugio *et al.*, 2011b), descartando las proteínas que no cumplan con estos criterios, proceso que además de dificultar su identificación, limitaba la búsqueda de los efectores. En este trabajo la creación de una nueva herramienta bioinformática que combina distintos programas computacionales (SignalP v4.1, TMHMM v2.0, Phobius y SecretomeP v2.0) permite una búsqueda más completa de estas moléculas de virulencia al lograr identificar efectores clásicos/canónicos (criterios establecidos en la literatura) y efectores no-clásicos/no-canónicos (que no cumplen con los criterios establecidos).

La tercera estrategia planteada fue la identificación de efectores del fitoplasma del ALC, pero debido a la dificultad en la obtención de estas proteínas efectoras por la ausencia de cebadores específicos para el fitoplasma del amarillamiento letal del cocotero, se consideró el aislamiento y secuenciación del genoma de este patógeno, utilizando un kit especial para el enriquecimiento de ADN microbiano.

El kit Microbiome DNA Enrichment (NEBNext®) es un kit creado para separar el ADN bacteriano del ADN del hospedero, con el objetivo de obtener una mayor concentración del primero, componente minoritario en la muestra original, conduciendo de esta manera a un enriquecimiento de éste. El componente mayoritario, ADN-hospedero, se separa del ADN-microbiano mediante el uso de perlas magnéticas las cuales se enlazan a una región constante del IgG humana fusionada genéticamente con un dominio de unión al metil-CpG (MBD-Fc), que interactúa específicamente con sitios metil-CpG (ADN metilados en sitios Citosina-fosfato-Guanina), el cual es común en los ADN de eucariotas (Schmitz *et al.*, 2019). El complejo formado por las perlas magnéticas-proteína MBDFc-ADN metilado se quedan adheridos en un rack magnético, lo que permite la recuperación del ADN microbiano enriquecido (Yigit *et al.*, 2016; Feehery *et al.*, 2013). La metilación consiste en la adición de un grupo metilo en el ADN, pudiendo ocurrir en los nucleótidos de citosina o adenina; en plantas, bacterias y mamíferos es común la metilación de la citosina, mientras que la metilación de la adenina se da en bacterias (Nasrullah *et al.*, 2022; Capuano *et al.*, 2014). El kit de enriquecimiento mencionado reconoce sitios metilados en sitios de citosina, este tipo de metilación es más frecuente en eucariotas llegando a representar hasta un 14 % en plantas, mientras que en bacterias es poco frecuente, 2.3 % (Capuano *et al.*, 2014). Adicionalmente, los fitoplasmas son procariotas con un bajo contenido de citosina (23-31 %) en comparación con otros procariontes (25-80 %) (Leyva-López *et al.*, 2006). Por lo cual este método de enriquecimiento de ADN microbiano resulta útil para obtener una mayor concentración del ADN del fitoplasma.

Esta técnica ha sido utilizada recientemente para la obtención de otros genomas de fitoplasmas, como '*Ca. P. vitis*', '*Ca. P. asteris*' de *hydrangea phyllody* strain (HP) y '*Ca. P. ziziphi*' (Debonneville *et al.*, 2022; Nijo *et al.*, 2021; Cai *et al.*, 2020). El kit de ADN-enriquecido permitió secuenciar el genoma borrador del fitoplasma '*Ca. P. pini*' del grupo 16SrXXI, asociado con la enfermedad de escoba de bruja en pinos, reportando un genoma de 474,136 bases con un contenido de G+C del 22.22 % (Cai *et al.*, 2020); otro genoma borrador que se logró obtener es el de '*Ca. P. asteris*' responsable de la filodia en la especie

Hydrangea (HP, hydrangea phyllody strain), del grupo 16Srl, con un tamaño de 597,762 bp (Nijo *et al.*, 2021) y el genoma completo de 'Ca. P. ziziphi' causante de la flavesencia dorada de la vid (Flavescence dorée), perteneciente al grupo 16SrV, con la obtención de un genoma *de novo* de 654,223 bp (Debonneville *et al.*, 2022).

En este trabajo se pudo observar que el kit separó gran parte del ADN vegetal al comparar las muestras de ADN-enriquecido (Ct =21-25) con el ADN no enriquecido (Ct =15, palma sin procesar con el kit). Al utilizar la técnica de ADN-enriquecido se logró obtener el ADN del fitoplasmas asociado al ALC, el cual fue enviado a secuenciar permitiendo el exitoso ensamblaje y la obtención del genoma parcial. El genoma parcial *de novo* de 'Ca. P. palmae'-ALC posee un tamaño de 434,453 pb. con la identificación de 54 efectores candidatos. Lo que permitió la identificación de distintas rutas metabólicas como la identificación de genes importantes para la obtención de energía e involucrados en el sistema de secreción. La obtención del genoma parcial del fitoplasma del ALC contribuye a una mejor comprensión sobre la organización y evolución de los genomas de fitoplasmas.

Por último, la cuarta estrategia planteada permitió la caracterización y expresión de un efector de 'Ca. P. palmae'-ALC, al diseñar cebadores específicos desde la secuencia del efector candidato identificado en el genoma parcial de 'Ca. P. palmae'-ALC y mediante la técnica de PCR de punto final usando muestras de ADNc obtenidas de palmas enfermas con ALC, la posterior recuperación de los fragmentos amplificados, secuenciación y alineamiento con la secuencia identificada como candidata. De acuerdo con los análisis bioinformáticos se observó que estos efectores candidatos pertenecen al grupo de efectores clásicos, al cumplir con los criterios establecidos en la literatura, siendo proteínas pequeñas (<400 aa), con presencia de PS y ausencia de DTM en sus secuencias; además de mostrar homología con el efector SAP11. Se determinó la expresión del efector candidato 2 en muestras de hojas y tallos de palmas de coco infectadas con el fitoplasma de ALC.

La identificación de los efectores de fitoplasmas permitió conocer las características sobresalientes de los mismos, las cuales ayudan a una mejor comprensión en como contribuyen a la virulencia de este patógeno. La creación de PhyEffector, el primer algoritmo

para identificar efectores de fitoplasmas, es otro aporte que facilita la identificación de estas proteínas efectoras, siendo anteriormente un proceso difícil la identificación de los mismos al carecer de un programa computacional. Por otro lado, la secuenciación del genoma parcial de 'Ca. P. palmae' asociado al ALC, permite ampliar el conocimiento sobre sus genes y la identificación de sus efectores. Un primer paso será estudiar sus proteínas de virulencia y en un futuro conocer las proteínas cognado con las cuales interactúan.

La presente tesis sienta las bases para caracterizar mejor los efectores de fitoplasmas, específicamente el asociado al ALC. Los efectores son moléculas claves asociados a la virulencia de estos microorganismos. Caracterizarlos a profundidad permite comprender el modo de acción de los fitoplasmas para eventualmente mejorar su control.

6.2. CONCLUSIONES GENERALES

En este trabajo se realizaron diferentes estrategias, las cuales, al final, permitieron ampliar el conocimiento sobre los efectores de fitoplasmas, descubriendo características sobresalientes como la identificación de firmas en estas proteínas de virulencia y la existencia de dos distintas clases de efectores: clásicos y no-clásicos. Se observó diferencias en su contenido de aminoácidos comparado con otros patógenos, teniendo un escaso o nulo contenido de cisteína y triptófano, y un alto contenido de lisina, leucina y asparagina en sus secuencias, aminoácidos que pueden estar relacionados con su virulencia; por otra parte, los motivos *de novo* encontrados permitió clasificarlos en 15 tribus, ésta clasificación permitirá facilitar su estudio, ya que al carecer de conservación en su secuencia muchos de ellos tenían que ser estudiados individualmente.

La creación de un algoritmo para la búsqueda de efectores de estos patógenos es otra contribución de este trabajo, una herramienta que no se tenía, por lo que se recurría a diferentes programas que dificultaban su búsqueda y limitaban la identificación de estas proteínas efectoras en los genomas de fitoplasmas.

Otro gran aporte, es la obtención de la secuencia del genoma parcial de 'Ca. P. palmae' patógeno asociado a la enfermedad de amarillamiento letal en el cocotero (ALC), genoma que no se conocía, además de la identificación de sus efectores candidatos. Adicionalmente, se logró la expresión de un efector del fitoplasma de ALC.

Con base en todos los análisis *in silico* realizados se puede concluir que la hipótesis planteada en este trabajo no se rechaza y que el fitoplasma causante del amarillamiento letal del cocotero comparte algunos efectores conservados con otros grupos de fitoplasmas, observando similitud con efectores semejantes a SAP11 y SAP54, además de pertenecer dentro de tribus ampliamente distribuidas en otros fitoplasmas, como la tribu 1. Sin embargo, estos análisis son solamente predictivos y aún falta validar y caracterizar estos efectores por medio de experimentos funcionales en plantas modelo.

6.3. PERSPECTIVAS

Las aportaciones realizadas en este trabajo permiten ampliar el conocimiento sobre las proteínas efectoras implicadas en la virulencia de los fitoplasmas. Sin embargo, aún falta realizar más experimentos que permitan validar los efectores candidatos identificados en el genoma de 'Ca. P. palmae', como la caracterización de los mismos.

Se han realizado pocos análisis de localización subcelular (Bai *et al.*, 2009) para estas proteínas y todavía faltan análisis de presencia de motivos señal de dirección nuclear, ya que estos son secretados dentro de la célula hospedante donde se dirigen al núcleo y manipulan a la célula del hospedante (Mittelberger *et al.*, 2019; Rivas y Genin, 2011), por lo que sería interesante hacer los análisis de localización subcelular para los efectores predichos.

BIBLIOGRAFÍA

- Aguin-Pombo, D., & Bourgoïn, T. (2012). Capítulo 30. Hemípteros. En *El árbol de la vida: Sistemática y evolución de los seres vivos*. Vargas, P. y Zardoya, R. (1a y 3a reimpresión ed., pp. 303-311.). International Union of Biological Sciences.
- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4), 420-423. <https://doi.org/10.1038/s41587-019-0036-z>
- Alnomasy, S. F. (2021). Virus-receptor interactions of SARS-CoV-2 spike-receptor-binding domain and human neuropilin-1 b1 domain. *Saudi Journal of Biological Sciences*, 28(7), 3926-3928. <https://doi.org/10.1016/j.sjbs.2021.03.074>
- Alonso, J. R. (2011). *La habilidad de un fitoplasma*.
- Alto, N. M., & Orth, K. (2012). Subversion of Cell Signaling by Pathogens. *Cold Spring Harbor Perspectives in Biology*, 4(9), a006114-a006114. <https://doi.org/10.1101/cshperspect.a006114>
- Alvarez-Buylla, E. R., Benítez, M., Corvera-Poiré, A., Chaos Cador, Á., De Folter, S., Gamboa De Buen, A., Garay-Arroyo, A., García-Ponce, B., Jaimes-Miranda, F., Pérez-Ruiz, R. V., Piñeyro-Nelson, A., & Sánchez-Corrales, Y. E. (2010). Flower Development. *The Arabidopsis Book*, 8, e0127. <https://doi.org/10.1199/tab.0127>
- Álvarez-Sánchez, F. . J., Rodríguez-Gúzmañ, P., & Alarcón, A. (2019). Cap. 2. Fitoplasmas y su estudio en México. Poghosyan Melkonyan, A.; Lebsky, V.; Rojas-Martíne, R.I.; Oropeza-Salín, C. Narváez-Cab, M. e Córdova-Lara, I. En *Biodiversidad de microorganismos de México: Importancia, aplicación y conservación*. editoriales@ciencias.unam.mx.

- Anabestani, A., Izadpanah, K., Abbà, S., Galetto, L., Ghorbani, A., Palmano, S., Siampour, M., Veratti, F., & Marzachi, C. (2017). Identification of putative effector genes and their transcripts in three strains related to 'Candidatus Phytoplasma aurantifolia'. *Microbiological Research*, 199, 57-66. <https://doi.org/10.1016/j.micres.2017.03.001>
- Andersen, M. T., Liefting, L. W., Havukkala, I., & Beever, R. E. (2013). Comparison of the complete genome sequence of two closely related isolates of 'Candidatus Phytoplasma australiense' reveals genome plasticity. *BMC Genomics*, 14(1), 529. <https://doi.org/10.1186/1471-2164-14-529>
- Arricau-Bouvery, N., Duret, S., Dubrana, M.-P., Desqué, D., Eveillard, S., Brocard, L., Malembic-Maher, S., & Foissac, X. (2021). Interactions between the flavescence dorée phytoplasma and its insect vector indicate lectin-type adhesion mediated by the adhesin VmpA. *Scientific Reports*, 11(1), 11222. <https://doi.org/10.1038/s41598-021-90809-z>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25-29. <https://doi.org/10.1038/75556>
- Asif, R., Yasmin, R., Mustafa, M., Ambreen, A., Mazhar, M., Rehman, A., Umbreen, S., & Ahmad, M. (2022). Phytohormones as Plant Growth Regulators and Safe Protectors against Biotic and Abiotic Stress. En C. Hano (Ed.), *Plant Hormones—Recent Advances, New Perspectives and Applications*. IntechOpen. <https://doi.org/10.5772/intechopen.102832>
- Aurin, M.-B., Haupt, M., Görlach, M., Rümpler, F., & Theißen, G. (2020). Structural Requirements of the Phytoplasma Effector Protein SAP54 for Causing Homeotic

- Transformation of Floral Organs. *Molecular Plant-Microbe Interactions*®, 33(9), 1129-1141. <https://doi.org/10.1094/MPMI-02-20-0028-R>
- Azcón-Bieto, J., & Talón, M. (2013). Cap. 29. Fisiología vegetal de las plantas y el estrés. En *Fundamentos de Fisiología Vegetal*. (Segunda edición.). McGraw-Hill Interamericana.
- Bai, B., Zhang, G., Pei, B., Song, Q., Hao, X., Zhao, L., & Wu, Y. (2023). The function of the phytoplasma effector SWP12 depends on the properties of two key amino acids. *Journal of Biological Chemistry*, 299(4), 103052. <https://doi.org/10.1016/j.jbc.2023.103052>
- Bai, X., Correa, V. R., Toruño, T. Y., Ammar, E.-D., Kamoun, S., & Hogenhout, S. A. (2009). AY-WB Phytoplasma Secretes a Protein That Targets Plant Cell Nuclei. *Molecular Plant-Microbe Interactions*®, 22(1), 18-30. <https://doi.org/10.1094/MPMI-22-1-0018>
- Bai, X., Zhang, J., Ewing, A., Miller, S. A., Jancso Radek, A., Shevchenko, D. V., Tsukerman, K., Walunas, T., Lapidus, A., Campbell, J. W., & Hogenhout, S. A. (2006). Living with Genome Instability: The Adaptation of Phytoplasmas to Diverse Environments of Their Insect and Plant Hosts. *Journal of Bacteriology*, 188(10), 3682-3696. <https://doi.org/10.1128/JB.188.10.3682-3696.2006>
- Balakireva, A., & Zamyatnin, A. (2018). Indispensable Role of Proteases in Plant Innate Immunity. *International Journal of Molecular Sciences*, 19(2), 629. <https://doi.org/10.3390/ijms19020629>
- Bar Barroeta, A., Marquart, J. A., Bakhtiari, K., Meijer, A. B., Urbanus, R. T., & Meijers, J. C. M. (2022). Nanobodies against factor XI apple 3 domain inhibit binding of factor IX and reveal a novel binding site for high molecular weight kininogen. *Journal of Thrombosis and Haemostasis*, 20(11), 2538-2549. <https://doi.org/10.1111/jth.15815>
- Bari, R., & Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Molecular Biology*, 69(4), 473-488. <https://doi.org/10.1007/s11103-008-9435-0>

- Bauer, S., Yu, D., Lawson, A. W., Saur, I. M. L., Frantzeskakis, L., Kracher, B., Logemann, E., Chai, J., Maekawa, T., & Schulze-Lefert, P. (2021). The leucine-rich repeats in allelic barley MLA immune receptors define specificity towards sequence-unrelated powdery mildew avirulence effectors with a predicted common RNase-like fold. *PLOS Pathogens*, 17(2), e1009223. <https://doi.org/10.1371/journal.ppat.1009223>
- Behringer, C., & Schwechheimer, C. (2015). B-GATA transcription factors – insights into their structure, regulation, and role in plant development. *Frontiers in Plant Science*, 6, 90. <https://doi.org/10.3389/fpls.2015.00090>
- Bender, K. W., & Zipfel, C. (2023). Paradigms of receptor kinase signaling in plants. *Biochemical Journal*, 480(12), 835-854. <https://doi.org/10.1042/BCJ20220372>
- Bendtsen, J. D., Kiemer, L., Fausbøll, A., & Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiology*, 5(1), 58. <https://doi.org/10.1186/1471-2180-5-58>
- Bendtsen, J., Nielsen, H., Von Heijne, G., & Brunak, S. (2004). Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology*, 340(4), 783-795. <https://doi.org/10.1016/j.jmb.2004.05.028>
- Benítez-Arellano, M., & Sosa-Alcaraz, M. (2017). *Situación actual del sistema producto coco en Yucatán, desde una perspectiva de la sustentabilidad*. ECORFAN Mérida, Yucatán.
- Bernal Bayard, J. (2010a). *Análisis de la función de S1rP, un efector de los sistemas de secreción de tipo III de Salmonella entérica, en la interacción con la célula hospedadora*. [Doctoral]. Universidad de Sevilla.
- Bernal-Bayard, J., Cardenal-Muñoz, E., & Ramos-Morales, F. (2010). The Salmonella Type III Secretion Effector, Salmonella Leucine-rich Repeat Protein (SlrP), Targets the Human Chaperone ERdj3. *Journal of Biological Chemistry*, 285(21), 16360-16368. <https://doi.org/10.1074/jbc.M110.100669>

- Bernal-Bayard, J., Cardenal-Muñoz, E., & Ramos-Morales, F. (2010b). The Salmonella Type III Secretion Effector, Salmonella Leucine-rich Repeat Protein (SlrP), Targets the Human Chaperone ERdj3. *Journal of Biological Chemistry*, 285(21), 16360-16368. <https://doi.org/10.1074/jbc.M110.100669>
- Bertaccini, A. (2007). Phytoplasmas: Diversity, taxonomy, and epidemiology. *Frontiers in Bioscience*, 12(1), 673. <https://doi.org/10.2741/2092>
- Bertaccini, A. (2021). Containment of Phytoplasma-Associated Plant Diseases by Antibiotics and Other Antimicrobial Molecules. *Antibiotics*, 10(11), 1398. <https://doi.org/10.3390/antibiotics10111398>
- Bertaccini, A. (2022). Plants and Phytoplasmas: When Bacteria Modify Plants. *Plants*, 11(11), 1425. <https://doi.org/10.3390/plants11111425>
- Bertaccini, A., Oshima, K., Kube, M., & Rao, G. P. (Eds.). (2019). *Phytoplasmas: Plant Pathogenic Bacteria - III: Genomics, Host Pathogen Interactions and Diagnosis*. Springer Singapore. <https://doi.org/10.1007/978-981-13-9632-8>
- Bhopatkar, A. A., Uversky, V. N., & Rangachari, V. (2020). Disorder and cysteines in proteins: A design for orchestration of conformational see-saw and modulatory functions. En *Progress in Molecular Biology and Translational Science* (Vol. 174, pp. 331-373). Elsevier. <https://doi.org/10.1016/bs.pmbts.2020.06.001>
- Bizarro, C. V., & Schuck, D. C. (2007). Purine and pyrimidine nucleotide metabolism in Mollicutes. *Genetics and Molecular Biology*, 30(1 suppl), 190-201. <https://doi.org/10.1590/S1415-47572007000200005>
- Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., ... Finn, R. D. (2021). The InterPro protein families and domains database: 20

- years on. *Nucleic Acids Research*, 49(D1), D344-D354.
<https://doi.org/10.1093/nar/gkaa977>
- Böhm, J. W., Duckeck, D., Duduk, B., Schneider, B., & Kube, M. (2023). Genome Comparison of 'Candidatus Phytoplasma rubi' with Genomes of Other 16SrV Phytoplasmas Highlights Special Group Features. *Applied Microbiology*, 3(3), 1083-1100. <https://doi.org/10.3390/applmicrobiol3030075>
- Böhm, J. W., Huettel, B., Schneider, B., & Kube, M. (2024). The complete genome of "Candidatus Phytoplasma fraxini" AshY1 from the ash yellows group. *Microbiology Resource Announcements*, 13(7), e00318-24. <https://doi.org/10.1128/mra.00318-24>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
<https://doi.org/10.1093/bioinformatics/btu170>
- Boonrod, K., Munteanu, B., Jarausch, B., Jarausch, W., & Krczal, G. (2012). An Immunodominant Membrane Protein (Imp) of 'Candidatus Phytoplasma mali' Binds to Plant Actin. *Molecular Plant-Microbe Interactions*, 25(7), 889-895.
<https://doi.org/10.1094/MPMI-11-11-0303>
- Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R., & Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, 5(1), 8365.
<https://doi.org/10.1038/srep08365>
- Brown, J., Pirrung, M., & McCue, L. A. (2017). FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, 33(19), 3137-3139. <https://doi.org/10.1093/bioinformatics/btx373>

- Bulgari, D., Casati, P., & Faoro, F. (2011). Fluorescence in situ hybridization for phytoplasma and endophytic bacteria localization in plant tissues. *Journal of Microbiological Methods*, 87(2), 220-223. <https://doi.org/10.1016/j.mimet.2011.08.001>
- Burbano-Figueroa, Ó. (2020). Resistencia de plantas a patógenos: Una revisión sobre los conceptos de resistencia vertical y horizontal. *Revista Argentina de Microbiología*, 52(3), 245-255. <https://doi.org/10.1016/j.ram.2020.04.006>
- Cabañas-Romero, P. V., & Huerta-Saquero, A. (2014). *Nanomáquinas biológicas: Los sistemas de secreción bacterianos*. 7(13), 28-36.
- Cadiou, L., Brunisholz, F., Cesari, S., & Kroj, T. (2023). Molecular engineering of plant immune receptors for tailored crop disease resistance. *Current Opinion in Plant Biology*, 74, 102381. <https://doi.org/10.1016/j.pbi.2023.102381>
- Cai, W., Shao, J., Zhao, Y., Davis, R. E., & Costanzo, S. (2020). Draft Genome Sequence of ‘*Candidatus* Phytoplasma pini’-Related Strain MDPP: A Resource for Comparative Genomics of Gymnosperm-Infecting Phytoplasmas. *Plant Disease*, 104(4), 1009-1010. <https://doi.org/10.1094/PDIS-10-19-2127-A>
- Caillet-Saguy, C., Durbesson, F., Rezelj, V. V., Gogl, G., Tran, Q. D., Twizere, J., Vignuzzi, M., Vincentelli, R., & Wolff, N. (2021). Host PDZ-containing proteins targeted by SARS-CoV-2. *The FEBS Journal*, 288(17), 5148-5162. <https://doi.org/10.1111/febs.15881>
- Camarena-Gutiérrez, G., & De la Torre-Almaraz, R. (2008). *Fitoplasmas: Síntomas y características moleculares*. 14(2), 81-87.
- Cao, Y., Trivellone, V., & Dietrich, C. H. (2020). A timetree for phytoplasmas (Mollicutes) with new insights on patterns of evolution and diversification. *Molecular Phylogenetics and Evolution*, 149, 106826. <https://doi.org/10.1016/j.ympev.2020.106826>

- Capuano, F., Mülleder, M., Kok, R., Blom, H. J., & Ralser, M. (2014). Cytosine DNA Methylation Is Found in *Drosophila melanogaster* but Absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* , and Other Yeast Species. *Analytical Chemistry*, 86(8), 3697-3702. <https://doi.org/10.1021/ac500447w>
- Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L.-P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., Mushayahama, T., ... Elser, J. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, 49(D1), D325-D334. The Gene Ontology Consortium. <https://doi.org/10.1093/nar/gkaa1113>
- Carella, P., Evangelisti, E., & Schornack, S. (2018). Sticking to it: Phytopathogen effector molecules may converge on evolutionarily conserved host targets in green plants. *Current Opinion in Plant Biology*, 44, 175-180. <https://doi.org/10.1016/j.pbi.2018.04.019>
- Carreón-Anguiano, K. G., Islas-Flores, I., Vega-Arreguín, J., Sáenz-Carbonell, L., & Canto-Canché, B. (2020). EffHunter: A Tool for Prediction of Effector Protein Candidates in Fungal Proteomic Databases. *Biomolecules*, 10(5), 712. <https://doi.org/10.3390/biom10050712>
- Carreón-Anguiano, K. G., Todd, J. N. A., Chi-Manzanero, B. H., Couoh-Dzul, O. J., Islas-Flores, I., & Canto-Canché, B. (2022). WideEffHunter: An Algorithm to Predict Canonical and Non-Canonical Effectors in Fungi and Oomycetes. *International Journal of Molecular Sciences*, 23(21), 13567. <https://doi.org/10.3390/ijms232113567>
- Carreón-Anguiano, K. G., Vila-Luna, S. E., Sáenz-Carbonell, L., & Canto-Canché, B. (2023). Novel Insights into Phytoplasma Effectors. *Horticulturae*, 9(11), 1228. <https://doi.org/10.3390/horticulturae9111228>

- Cesari, S., Xi, Y., Declerck, N., Chalvon, V., Mammri, L., Pugnière, M., Henriquet, C., De Guillen, K., Chochois, V., Padilla, A., & Kroj, T. (2022). New recognition specificity in a plant immune receptor by molecular engineering of its integrated domain. *Nature Communications*, 13(1), 1524. <https://doi.org/10.1038/s41467-022-29196-6>
- Chang, L., Mondal, A., & Perez, A. (2022). Towards rational computational peptide design. *Frontiers in Bioinformatics*, 2, 1046493. <https://doi.org/10.3389/fbinf.2022.1046493>
- Chang, S. H., Tan, C. M., Wu, C.-T., Lin, T.-H., Jiang, S.-Y., Liu, R.-C., Tsai, M.-C., Su, L.-W., & Yang, J.-Y. (2018). Alterations of plant architecture and phase transition by the phytoplasma virulence factor SAP11. *Journal of Experimental Botany*. <https://doi.org/10.1093/jxb/ery318>
- Chen, L.-L., Chung, W.-C., Lin, C.-P., & Kuo, C.-H. (2012). Comparative Analysis of Gene Content Evolution in Phytoplasmas and Mycoplasmas. *PLoS ONE*, 7(3), e34407. <https://doi.org/10.1371/journal.pone.0034407>
- Chen, P., Chen, L., Ye, X., Tan, B., Zheng, X., Cheng, J., Wang, W., Yang, Q., Zhang, Y., Li, J., & Feng, J. (2022). Phytoplasma effector Zaofeng6 induces shoot proliferation by decreasing the expression of ZjTCP7 in *Ziziphus jujuba*. *Horticulture Research*, 9, uhab032. <https://doi.org/10.1093/hr/uhab032>
- Chen, W., Li, Y., Wang, Q., Wang, N., & Wu, Y. (2014). Comparative Genome Analysis of Wheat Blue Dwarf Phytoplasma, an Obligate Pathogen That Causes Wheat Blue Dwarf Disease in China. *PLoS ONE*, 9(5), e96436. <https://doi.org/10.1371/journal.pone.0096436>
- Chen, Y.-T., Chang, C.-C., Chen, C.-W., Chen, K.-C., & Chu, Y.-W. (2019). MADS-Box Gene Classification in Angiosperms by Clustering and Machine Learning Approaches. *Frontiers in Genetics*, 9, 707. <https://doi.org/10.3389/fgene.2018.00707>

- Cheng, H.-C., Skehan, B. M., Campellone, K. G., Leong, J. M., & Rosen, M. K. (2008). Structural mechanism of WASP activation by the enterohaemorrhagic *E. coli* effector EspFU. *Nature*, *454*(7207), 1009-1013. <https://doi.org/10.1038/nature07160>
- Cho, S.-T., Kung, H.-J., Huang, W., Hogenhout, S. A., & Kuo, C.-H. (2020). Species Boundaries and Molecular Markers for the Classification of 16S rDNA Phytoplasmas Inferred by Genome Analysis. *Frontiers in Microbiology*, *11*, 1531. <https://doi.org/10.3389/fmicb.2020.01531>
- Cho, S.-T., Lin, C.-P., & Kuo, C.-H. (2019). Genomic Characterization of the Periwinkle Leaf Yellowing (PLY) Phytoplasmas in Taiwan. *Frontiers in Microbiology*, *10*, 2194. <https://doi.org/10.3389/fmicb.2019.02194>
- Choe, J. E., & Welch, M. D. (2016). Actin-based motility of bacterial pathogens: Mechanistic diversity and its impact on virulence. *Pathogens and Disease*, *74*(8), ftw099. <https://doi.org/10.1093/femspd/ftw099>
- Choi, C. Q. (2017). Prion-like protein spotted in bacteria for the first time. *Nature*, *nature.2017.21293*. <https://doi.org/10.1038/nature.2017.21293>
- Christensen, N. R., Čalyševa, J., Fernandes, E. F. A., Lüchow, S., Clemmensen, L. S., Haugaard-Kedström, L. M., & Strømgaard, K. (2019). PDZ Domains as Drug Targets. *Advanced Therapeutics*, 1800143. <https://doi.org/10.1002/adtp.201800143>
- Chung, W.-C., Chen, L.-L., Lo, W.-S., Lin, C.-P., & Kuo, C.-H. (2013). Comparative Analysis of the Peanut Witches'-Broom Phytoplasma Genome Reveals Horizontal Transfer of Potential Mobile Units and Effectors. *PLoS ONE*, *8*(4), e62770. <https://doi.org/10.1371/journal.pone.0062770>
- Cock, J. M., Vanoosthuyse, V., & Gaude, T. (2002). Receptor kinase signalling in plants and animals: Distinct molecular systems with mechanistic similarities. *Current Opinion in Cell Biology*, *14*(2), 230-236. [https://doi.org/10.1016/S0955-0674\(02\)00305-8](https://doi.org/10.1016/S0955-0674(02)00305-8)

- Coetzee, B., Douglas-Smit, N., Maree, H. J., Burger, J. T., Krüger, K., & Pietersen, G. (2019). Draft Genome Sequence of a “*Candidatus Phytoplasma asteris*”-Related Strain (Aster Yellows, Subgroup 16SrI-B) from South Africa. *Microbiology Resource Announcements*, 8(17), e00148-19. <https://doi.org/10.1128/MRA.00148-19>
- Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). (s. f.). <https://www.biodiversidad.gob.mx/Difusion/cienciaCiudadana/urbanos/Cocosnucifera>.
- Córdova, I., Oropeza, C., Harrison, N., Ku-Rodríguez, S., Puch-Hau, C., Narváez, M., & Sáenz, L. (2019). Simultaneous detection of coconut lethal yellowing phytoplasmas (group 16SrIV) by real-time PCR assays using 16Sr- and GroEL-based TaqMan probes. *Journal of Plant Pathology*, 101(3), 609-619. <https://doi.org/10.1007/s42161-019-00249-y>
- Córdova, L. E. (2009). *Glicosilación de proteínas en bacterias patógenas*. IV(3), 105-110.
- Curto-Prieto. (2013). *Función del factor transcripcional de tipo MADS-box, AGAMOUS-LIKE 67 (AGL67), en la señalización hormonal durante la germinación de semillas de Arabidopsis thaliana*. [Memoria]. Universidad de Salamanca. Centro Hispanoluso de Investigaciones Agrarias (Ciale).
- Davey, N. E., Simonetti, L., & Ivarsson, Y. (2023). The next wave of interactomics: Mapping the SLiM-based interactions of the intrinsically disordered proteome. *Current Opinion in Structural Biology*, 80, 102593. <https://doi.org/10.1016/j.sbi.2023.102593>
- Davey, N. E., Travé, G., & Gibson, T. J. (2011). How viruses hijack cell regulation. *Trends in Biochemical Sciences*, 36(3), 159-169. <https://doi.org/10.1016/j.tibs.2010.10.002>
- Davies, H. M., Nofal, S. D., McLaughlin, E. J., & Osborne, A. R. (2017). Repetitive sequences in malaria parasite proteins. *FEMS Microbiology Reviews*, 41(6), 923-940. <https://doi.org/10.1093/femsre/fux046>

- Davis, R. E., & Sinclair, W. A. (1998). Phytoplasma Identity and Disease Etiology. *Phytopathology*[®], 88(12), 1372-1376. <https://doi.org/10.1094/PHYTO.1998.88.12.1372>
- De Paz Fernández, H. (2010). *Estudio comparativo de sistemas de secreción tipo IV implicados en transferencia conjugativa de DNA y virulencia bacteriana*. [Memoria de Doctorado]. Facultad de Medicina. Universidad de Cantabria.
- Dean, P. (2011). Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiology Reviews*, 35(6), 1100-1125. <https://doi.org/10.1111/j.1574-6976.2011.00271.x>
- Debonneville, C., Mandelli, L., Brodard, J., Groux, R., Roquis, D., & Schumpp, O. (2022). The Complete Genome of the “Flavescence Dorée” Phytoplasma Reveals Characteristics of Low Genome Plasticity. *Biology*, 11(7), 953. <https://doi.org/10.3390/biology11070953>
- Delaunay, M., & Ha-Duong, T. (2023). Computational design of cyclic peptides to inhibit protein-peptide interactions. *Biophysical Chemistry*, 296, 106987. <https://doi.org/10.1016/j.bpc.2023.106987>
- Delucchi, G., & Hurrell. (2008). Capítulo: Arecaceae. En *Flora Rioplatense* (Vol. 3, pp. 91-132). <https://www.researchgate.net/publication/272709881>.
- Dermastia, M. (2019). Plant Hormones in Phytoplasma Infected Plants. *Frontiers in Plant Science*, 10, 477. <https://doi.org/10.3389/fpls.2019.00477>
- Dickinson, M. (2010). Mobile units of DNA in phytoplasma genomes. *Molecular Microbiology*, 77(6), 1351-1353. <https://doi.org/10.1111/j.1365-2958.2010.07308.x>
- Dillon, M. M., Almeida, R. N. D., Laflamme, B., Martel, A., Weir, B. S., Desveaux, D., & Guttman, D. S. (2019). Molecular Evolution of Pseudomonas syringae Type III Secreted Effector Proteins. *Frontiers in Plant Science*, 10, 418. <https://doi.org/10.3389/fpls.2019.00418>

- Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: Towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics*, 11(8), 539-548. <https://doi.org/10.1038/nrg2812>
- Doi, Y., Teranaka, M., Yora, K., & Asuyama, H. (1967). Mycoplasma- or PLT Group-like Microorganisms Found in the Phloem Elements of Plants Infected with Mulberry Dwarf, Potato Witches' Broom, Aster Yellows, or Paulownia Witches' Broom. *Japanese Journal of Phytopathology*, 33(4), 259-266. <https://doi.org/10.3186/jjphytopath.33.259>
- Domazakis, E., Lin, X., Aguilera-Galvez, C., Wouters, D., Bijsterbosch, G., Wolters, P. J., & Vleeshouwers, V. G. A. A. (2017). Effectoromics-Based Identification of Cell Surface Receptors in Potato. En L. Shan & P. He (Eds.), *Plant Pattern Recognition Receptors* (Vol. 1578, pp. 337-353). Springer New York. https://doi.org/10.1007/978-1-4939-6859-6_29
- Douriet, N. R. (2011). *Análisis de la expresión diferencial de genes involucrados en la interacción planta de papa-fitoplasma "Mexican potato purple top"*. [Tesis de maestría]. Instituto Politécnico Nacional (IPN).
- Drcelic, M., Skiljaica, A., Polak, B., Bauer, N., & Seruga Music, M. (2024). 'Candidatus Phytoplasma solani' Predicted Effector SAP11-like Alters Morphology of Transformed Arabidopsis Plants and Interacts with AtTCP2 and AtTCP4 Plant Transcription Factors. *Pathogens*, 13(10), 893. <https://doi.org/10.3390/pathogens13100893>
- Ekchaweng, K., Evangelisti, E., Schornack, S., Tian, M., & Churngchow, N. (2017). The plant defense and pathogen counterdefense mediated by *Hevea brasiliensis* serine protease HbSPA and *Phytophthora palmivora* extracellular protease inhibitor PpEPI10. *PLOS ONE*, 12(5), e0175795. <https://doi.org/10.1371/journal.pone.0175795>

- Elkhaligy, H., Balbin, C. A., & Siltberg-Liberles, J. (2022). Comparative Analysis of Structural Features in SLiMs from Eukaryotes, Bacteria, and Viruses with Importance for Host-Pathogen Interactions. *Pathogens*, 11(5), 583. <https://doi.org/10.3390/pathogens11050583>
- Eng Sánchez, F. (2008). Jasmonatos: Compuestos de alto valor para la agricultura. Parte I. Actividad biológica y ruta biosintética del ácido jasmónico en plantas. *Sobre los Derivados de la Caña de Azúcar.*, XLII(1-3), 51-59.
- Eshghi, A., Gaultney, R. A., England, P., Brûlé, S., Miras, I., Sato, H., Coburn, J., Bellalou, J., Moriarty, T. J., Haouz, A., & Picardeau, M. (2019). An extracellular *LEPTOSPIRA INTERROGANS* leucine-rich repeat protein binds human E- and VE-cadherins. *Cellular Microbiology*, 21(2). <https://doi.org/10.1111/cmi.12949>
- Espadaler, J., Querol, E., Aviles, F. X., & Oliva, B. (2006). Identification of function-associated loop motifs and application to protein function prediction. *Bioinformatics*, 22(18), 2237-2243. <https://doi.org/10.1093/bioinformatics/btl382>
- Fanning, A. S., & Anderson, J. M. (1999). PDZ domains: Fundamental building blocks in the organization of protein complexes at the plasma membrane. *Journal of Clinical Investigation*, 103(6), 767-772. <https://doi.org/10.1172/JCI6509>
- Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A., & Pradhan, S. (2013). A Method for Selectively Enriching Microbial DNA from Contaminating Vertebrate Host DNA. *PLoS ONE*, 8(10), e76096. <https://doi.org/10.1371/journal.pone.0076096>
- Ferriol Marchena, X., Paredes, T. C., Hernández Rodríguez, L., & Luis Pantoja, M. (2016). *Los fitoplasmas. Características generales y sintomatología que producen en las plantas*. 33(1), 27-37.

- Figaj, D., Ambroziak, P., Przepiora, T., & Skorko-Glonek, J. (2019). The Role of Proteases in the Virulence of Plant Pathogenic Bacteria. *International Journal of Molecular Sciences*, 20(3), 672. <https://doi.org/10.3390/ijms20030672>
- Filisetti, D., Théobald-Dietrich, A., Mahmoudi, N., Rudinger-Thirion, J., Candolfi, E., & Frugier, M. (2013). Aminoacylation of Plasmodium falciparum tRNAAsn and Insights in the Synthesis of Asparagine Repeats. *Journal of Biological Chemistry*, 288(51), 36361-36371. <https://doi.org/10.1074/jbc.M113.522896>
- Fletcher, J., & Wayadanda, A. (2002). Fastidious Vascular-Colonizing Bacteria. *The Plant Health Instructor*. <https://doi.org/10.1094/PHI-I-2002-1218-02>
- Fouché, S., Plissonneau, C., & Croll, D. (2018). The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology*, 46, 34-42. <https://doi.org/10.1016/j.mib.2018.01.020>
- Francis, I., Holsters, M., & Vereecke, D. (2010). The Gram-positive side of plant–microbe interactions. *Environmental Microbiology*, 12(1), 1-12. <https://doi.org/10.1111/j.1462-2920.2009.01989.x>
- Galetto, L., Abbà, S., Rossi, M., Ripamonti, M., Palmano, S., Bosco, D., & Marzachi, C. (2021). Silencing of ATP synthase β reduces phytoplasma multiplication in a leafhopper vector. *Journal of Insect Physiology*, 128, 104176. <https://doi.org/10.1016/j.jinsphys.2020.104176>
- Galetto, L., Fletcher, J., Bosco, D., Turina, M., Wayadande, A., & Marzachi, C. (2008). Characterization of putative membrane protein genes of the ‘*Candidatus* Phytoplasma asteris’, chrysanthemum yellows isolate. *Canadian Journal of Microbiology*, 54(5), 341-351. <https://doi.org/10.1139/W08-010>
- Gao, X., Ren, Z., Zhao, W., & Li, W. (2023). Candidatus Phytoplasma ziziphi encodes non-classically secreted proteins that suppress hypersensitive cell death response in

- Nicotiana benthamiana. *Phytopathology Research*, 5(1), 11.
<https://doi.org/10.1186/s42483-023-00166-0>
- Garg, A., Singhal, N., & Kumar, M. (2022). Investigating the eukaryotic host-like SLiMs in microbial mimitopes and their potential as novel drug targets for treating autoimmune diseases. *Frontiers in Microbiology*, 13, 1039188.
<https://doi.org/10.3389/fmicb.2022.1039188>
- Gastaldi, V. (2020). *El papel de factores de transcripción de la familia TCP en la regulación del desarrollo y la arquitectura vegetal*. [Tesis doctoral]. Universidad Nacional del Litoral. Facultad de Bioquímica y Ciencias Biológicas.
- Gérard, C., Carrière, F., Receveur-Bréchet, V., Launay, H., & Gontero, B. (2022). A Trajectory of Discovery: Metabolic Regulation by the Conditionally Disordered Chloroplast Protein, CP12. *Biomolecules*, 12(8), 1047.
<https://doi.org/10.3390/biom12081047>
- Góngora-Canúl, C. C., Pérez-Hernández, O., Pech-Cauich, O., Escamilla-Bencomo, J. A., & Mora-Aguilera, G. (2004). *Gradientes de Diseminación del Amarillamiento Letal en Cocotero (Cocos nucífera) en Yucatán, México*. 22(3), 370-376.
- González-Grandío, E., & Cubas, P. (2016). TCP Transcription Factors: Evolution, Structure, and Biochemical Function. En *Plant Transcription Factors* (pp. 139-151). Elsevier.
<https://doi.org/10.1016/B978-0-12-800854-6.00009-9>
- Goswami, S., Samanta, D., & Duraivelan, K. (2023). Molecular mimicry of host short linear motif-mediated interactions utilised by viruses for entry. *Molecular Biology Reports*, 50(5), 4665-4673. <https://doi.org/10.1007/s11033-023-08389-2>
- Goto, M., Omi, R., Nakagawa, N., Miyahara, I., & Hirotsu, K. (2004). Crystal Structures of CTP Synthetase Reveal ATP, UTP, and Glutamine Binding Sites. *Structure*, 12(8), 1413-1423. <https://doi.org/10.1016/j.str.2004.05.013>

- Granados Sánchez, D., & López Ríos, G. F. (2002). *Manejo de la palma de coco (cocos nucifera L.) en México*. 8(1), 39-48.
- Granados Sánchez, D. & López Ríos, G.F. (2002). Manejo de la palma de coco (cocos nucifera L.) en México. *Revista Chapingo. Serie Ciencias Forestales y del Ambiente*, 8(1), 39-48.
- Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C., Graham, M., Van Domselaar, G., & Stothard, P. (2023). Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, 51(W1), W484-W492. <https://doi.org/10.1093/nar/gkad326>
- Green, E. R., & Meccas, J. (2016). Bacterial Secretion Systems: An Overview. *Microbiology Spectrum*, 4(1), 4.1.13. <https://doi.org/10.1128/microbiolspec.VMBF-0012-2015>
- Guiney, D. G., & Lesnick, M. (2005). Targeting of the actin cytoskeleton during infection by Salmonella strains. *Clinical Immunology*, 114(3), 248-255. <https://doi.org/10.1016/j.clim.2004.07.014>
- Guzmán, P. (2017). *Participación de proteínas tipo efector de Trichoderma spp. En interacción benéfica con plantas de A. thaliana*. [Tesis de posgrado]. Universidad de Guanajuato. División de ciencias naturales y exactas.
- Harrison, P. M. (2019). Evolutionary behaviour of bacterial prion-like proteins. *PLOS ONE*, 14(3), e0213030. <https://doi.org/10.1371/journal.pone.0213030>
- He, Q., McLellan, H., Boevink, P. C., & Birch, P. R. J. (2020). All Roads Lead to Susceptibility: The Many Modes of Action of Fungal and Oomycete Intracellular Effectors. *Plant Communications*, 1(4), 100050. <https://doi.org/10.1016/j.xplc.2020.100050>
- Hemmati, C., Nikooei, M., Al-Subhi, A. M., & Al-Sadi, A. M. (2021). History and Current Status of Phytoplasma Diseases in the Middle East. *Biology*, 10(3), 226. <https://doi.org/10.3390/biology10030226>

- Henrietta, H. M., Kalaiyarasi, K., & Raj, A. S. (2022). Coconut Tree (*Cocos nucifera*) Products: A Review of Global Cultivation and its Benefits. *Journal of Sustainability and Environmental Management*, 1(2), 257-264. <https://doi.org/10.3126/josem.v1i2.45377>
- Hogenhout, S. A., & Loria, R. (2008). Virulence mechanisms of Gram-positive plant pathogenic bacteria. *Current Opinion in Plant Biology*, 11(4), 449-456. <https://doi.org/10.1016/j.pbi.2008.05.007>
- Hogenhout, S. A., Oshima, K., Ammar, E., Kakizawa, S., Kingdom, H. N., & Namba, S. (2008). Phytoplasmas: Bacteria that manipulate plants and insects. *Molecular Plant Pathology*, 9(4), 403-423. <https://doi.org/10.1111/j.1364-3703.2008.00472.x>
- Hogenhout, S. A., Van Der Hoorn, R. A. L., Terauchi, R., & Kamoun, S. (2009). Emerging Concepts in Effector Biology of Plant-Associated Organisms. *Molecular Plant-Microbe Interactions*, 22(2), 115-122. <https://doi.org/10.1094/MPMI-22-2-0115>
- Hoischen, C., Dijkstra, A., Rottem, S., Reizer, J., & Saier, M. H. (1993). Presence of protein constituents of the gram-positive bacterial phosphotransferase regulatory system in *Acholeplasma laidlawii*. *Journal of Bacteriology*, 175(20), 6599-6604. <https://doi.org/10.1128/jb.175.20.6599-6604.1993>
- Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M., Komatsu, K., Kagiwada, S., Yamaji, Y., & Namba, S. (2009). A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proceedings of the National Academy of Sciences*, 106(15), 6416-6421. <https://doi.org/10.1073/pnas.0813038106>
- Hu, Y., Huang, H., Hui, X., Cheng, X., White, A. P., Zhao, Z., & Wang, Y. (2016). Distribution and Evolution of *Yersinia* Leucine-Rich Repeat Proteins. *Infection and Immunity*, 84(8), 2243-2254. <https://doi.org/10.1128/IAI.00324-16>

- Huang, Cho, S.-T., Lin, Y.-C., Tan, C.-M., Chiu, Y.-C., Yang, J.-Y., & Kuo, C.-H. (2022). Comparative Genome Analysis of 'Candidatus Phytoplasma luffae' Reveals the Influential Roles of Potential Mobile Units in Phytoplasma Evolution. *Frontiers in Microbiology*, 13, 773608. <https://doi.org/10.3389/fmicb.2022.773608>
- Huang, J., & Gogarten, J. (2006). Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends in Genetics*, 22(7), 361-366. <https://doi.org/10.1016/j.tig.2006.05.004>
- Huang, W., MacLean, A. M., Sugio, A., Maqbool, A., Busscher, M., Cho, S.-T., Kamoun, S., Kuo, C.-H., Immink, R. G. H., & Hogenhout, S. A. (2021). Parasitic modulation of host development by ubiquitin-independent protein degradation. *Cell*, 184(20), 5201-5214.e12. <https://doi.org/10.1016/j.cell.2021.08.029>
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>
- Ijaz, M., Khan, F., Zaki, H. E. M., Khan, M. M., Radwan, K. S. A., Jiang, Y., Qian, J., Ahmed, T., Shahid, M. S., Luo, J., & Li, B. (2023). Recent Trends and Advancements in CRISPR-Based Tools for Enhancing Resistance against Plant Pathogens. *Plants*, 12(9), 1911. <https://doi.org/10.3390/plants12091911>
- Irwin, N. A. T., Pittis, A. A., Richards, T. A., & Keeling, P. J. (2021). Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nature Microbiology*, 7(2), 327-336. <https://doi.org/10.1038/s41564-021-01026-3>
- Iwabuchi, N., Kitazawa, Y., Maejima, K., Koinuma, H., Miyazaki, A., Matsumoto, O., Suzuki, T., Nijo, T., Oshima, K., Namba, S., & Yamaji, Y. (2020). Functional variation in phyllogen, a phyllody-inducing phytoplasma effector family, attributable to a single amino acid polymorphism. *Molecular Plant Pathology*, 21(10), 1322-1336. <https://doi.org/10.1111/mpp.12981>

- Izarra, M., & Lindqvist-Kreuze, H. (2016). *Expresión de efectores RXLR en el linaje clonal EC-1 de Phytophthora infestans en Perú*. 23(3), 293-299.
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, 9(1), 5114. <https://doi.org/10.1038/s41467-018-07641-9>
- Jheeta, S., Chatzitheodoridis, E., Devine, K., & Block, J. (2021). The Way forward for the Origin of Life: Prions and Prion-Like Molecules First Hypothesis. *Life*, 11(9), 872. <https://doi.org/10.3390/life11090872>
- Jiang, R. H. Y., Tripathy, S., Govers, F., & Tyler, B. M. (2008). RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proceedings of the National Academy of Sciences*, 105(12), 4874-4879. <https://doi.org/10.1073/pnas.0709303105>
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, 36(Web Server), W5-W9. <https://doi.org/10.1093/nar/gkn201>
- Jomantienė, R., Zhao, Y., & Davis, R. E. (2007). Sequence-Variable Mosaics: Composites of Recurrent Transposition Characterizing the Genomes of Phylogenetically Diverse Phytoplasmas. *DNA and Cell Biology*, 26(8), 557-564. <https://doi.org/10.1089/dna.2007.0610>
- Jones, D. A. B., Moolhuijzen, P. M., & Hane, J. K. (2021). Remote homology clustering identifies lowly conserved families of effector proteins in plant-pathogenic fungi. *Microbial Genomics*, 7(9). <https://doi.org/10.1099/mgen.0.000637>
- Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323-329. <https://doi.org/10.1038/nature05286>

- Jones, J. D., & O'Connor, C. D. (2011). Protein acetylation in prokaryotes. *PROTEOMICS*, 11(15), 3012-3022. <https://doi.org/10.1002/pmic.201000812>
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236-1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Kakizawa, S., Oshima, K., Ishii, Y., Hoshi, A., Maejima, K., Jung, H.-Y., Yamaji, Y., & Namba, S. (2009). Cloning of immunodominant membrane protein genes of phytoplasmas and their *in planta* expression. *FEMS Microbiology Letters*, 293(1), 92-101. <https://doi.org/10.1111/j.1574-6968.2009.01509.x>
- Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—The Phobius web server. *Nucleic Acids Research*, 35(Web Server), W429-W432. <https://doi.org/10.1093/nar/gkm256>
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27-30. <https://doi.org/10.1093/nar/28.1.27>
- Kazmirchuk, T. D. D., Bradbury-Jost, C., Withey, T. A., Gessese, T., Azad, T., Samanfar, B., Dehne, F., & Golshani, A. (2023). Peptides of a Feather: How Computation Is Taking Peptide Therapeutics under Its Wing. *Genes*, 14(6), 1194. <https://doi.org/10.3390/genes14061194>
- Kerr, A., & Gibb, K. (s. f.). *Bacteria and phytoplasma as plant parasites*.
- Kibby, E. M., Conte, A. N., Burroughs, A. M., Nagy, T. A., Vargas, J. A., Whalen, L. A., Aravind, L., & Whiteley, A. T. (2023). Bacterial NLR-related proteins protect against phage. *Cell*, 186(11), 2410-2424.e18. <https://doi.org/10.1016/j.cell.2023.04.015>

- Kirdat, K., Tiwarekar, B., Sathe, S., & Yadav, A. (2023). From sequences to species: Charting the phytoplasma classification and taxonomy in the era of taxogenomics. *Frontiers in Microbiology*, 14, 1123783. <https://doi.org/10.3389/fmicb.2023.1123783>
- Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes¹¹Edited by F. Cohen. *Journal of Molecular Biology*, 305(3), 567-580. <https://doi.org/10.1006/jmbi.2000.4315>
- Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes¹¹Edited by F. Cohen. *Journal of Molecular Biology*, 305(3), 567-580. <https://doi.org/10.1006/jmbi.2000.4315>
- Kube, M., Mitrovic, J., Duduk, B., Rabus, R., & Seemüller, E. (2012). Current View on Phytoplasma Genomes and Encoded Metabolism. *The Scientific World Journal*, 2012, 1-25. <https://doi.org/10.1100/2012/185942>
- Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A. M., Reinhardt, R., & Seemüller, E. (2008). The linear chromosome of the plant-pathogenic mycoplasma «Candidatus Phytoplasma mali». *BMC Genomics*, 9(1), 306. <https://doi.org/10.1186/1471-2164-9-306>
- Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J. A., Bukirova, D., Čalyševa, J., Palopoli, N., Davey, N. E., Chemes, L. B., & Gibson, T. J. (2020). ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Research*, gkz1030. <https://doi.org/10.1093/nar/gkz1030>
- Kumar, M., Michael, S., Alvarado-Valverde, J., Mészáros, B., Sámano-Sánchez, H., Zeke, A., Dobson, L., Lazar, T., Örd, M., Nagpal, A., Farahi, N., Käser, M., Kraleti, R., Davey, N. E., Pancsa, R., Chemes, L. B., & Gibson, T. J. (2022). The Eukaryotic

- Linear Motif resource: 2022 release. *Nucleic Acids Research*, 50(D1), D497-D508.
<https://doi.org/10.1093/nar/gkab975>
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, 5(2), R12. <https://doi.org/10.1186/gb-2004-5-2-r12>
- Lamilla Monje, J. R., & Franco Lara, L. (2019). *Revisión: Genómica de las interacciones entre fitoplasmas, hospederos vegetales e insectos vectores*. 17(3), 143-165.
- Latorre, S. M., Reyes-Avila, C. S., Malmgren, A., Win, J., Kamoun, S., & Burbano, H. A. (2020). Differential loss of effector genes in three recently expanded pandemic clonal lineages of the rice blast fungus. *BMC Biology*, 18(1), 88.
<https://doi.org/10.1186/s12915-020-00818-z>
- Lee, I.-M., Davis, R. E., & Gundersen-Rindal, D. E. (2000). Phytoplasma: Phytopathogenic Mollicutes. *Annual Review of Microbiology*, 54(1), 221-255.
<https://doi.org/10.1146/annurev.micro.54.1.221>
- Lepka, P., Stitt, M., Moll, E., & Seemüller, E. (1999). Effect of phytoplasmal infection on concentration and translocation of carbohydrates and amino acids in periwinkle and tobacco. *Physiological and Molecular Plant Pathology*, 55(1), 59-68.
<https://doi.org/10.1006/pmpp.1999.0202>
- Leyva-López, N. E., Ochoa-Sánchez, J. C., Aviña-Padilla, K., & Martínez-Soriano, J. P. (2006). *Fitoplasmas, patógenos mínimos de plantas*. 1-19.
- Li, C., Barroeta, A. B., Wong, S. S., Kim, H. J., Pathak, M., Dreveny, I., Meijers, J. C. M., & Emsley, J. (2023). Structures of factor XI and prekallikrein bound to domain 6 of high-molecular weight kininogen reveal alternate domain 6 conformations and exosites. *Journal of Thrombosis and Haemostasis*, 21(9), 2378-2389.
<https://doi.org/10.1016/j.jtha.2023.03.042>

- Li, C., Xu, M., Cai, X., Han, Z., Si, J., & Chen, D. (2022). Jasmonate Signaling Pathway Modulates Plant Defense, Growth, and Their Trade-Offs. *International Journal of Molecular Sciences*, 23(7), 3945. <https://doi.org/10.3390/ijms23073945>
- Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., & Lam, T.-W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, 102, 3-11. <https://doi.org/10.1016/j.ymeth.2016.02.020>
- Li, D., Zhu, Z., Deng, X., Zou, J., Ma, C., Li, C., Yin, T., Liu, C., Wang, J., Chen, Q., & Xin, D. (2023). GmPBS1, a Hub Gene Interacting with Rhizobial Type-III Effectors NopT and NopP, Regulates Soybean Nodulation. *Agronomy*, 13(5), 1242. <https://doi.org/10.3390/agronomy13051242>
- Li, J., Gao, M., Gabriel, D. W., Liang, W., & Song, L. (2020). Secretome-Wide Analysis of Lysine Acetylation in *Fusarium oxysporum* f. Sp. *Lycopersici* Provides Novel Insights Into Infection-Related Proteins. *Frontiers in Microbiology*, 11, 559440. <https://doi.org/10.3389/fmicb.2020.559440>
- Liefting, L. W., & Kirkpatrick, B. C. (2003). Cosmid cloning and sample sequencing of the genome of the uncultivable mollicute, Western X-disease phytoplasma, using DNA purified by pulsed-field gel electrophoresis. *FEMS Microbiology Letters*, 221(2), 203-211. [https://doi.org/10.1016/S0378-1097\(03\)00183-6](https://doi.org/10.1016/S0378-1097(03)00183-6)
- Limones-Briones, V., & Fernandez-Barrera, M. A. (2016). El cocotero: “El árbol de la vida”. *Herbario CICY*, 8, 107-110.
- Lipsitch, M., Siller, S., & Nowak, M. A. (1996). THE EVOLUTION OF VIRULENCE IN PATHOGENS WITH VERTICAL AND HORIZONTAL TRANSMISSION. *Evolution*, 50(5), 1729-1741. <https://doi.org/10.1111/j.1558-5646.1996.tb03560.x>

- Liu, L., Xu, L., Jia, Q., Pan, R., Oelmüller, R., Zhang, W., & Wu, C. (2019). Arms race: Diverse effector proteins with conserved motifs. *Plant Signaling & Behavior*, 14(2), 1557008. <https://doi.org/10.1080/15592324.2018.1557008>
- Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., & Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology*, 66(1), 513-545. <https://doi.org/10.1146/annurev-arplant-043014-114623>
- Loimaranta, V., Hytönen, J., Pulliainen, A. T., Sharma, A., Tenovuo, J., Strömberg, N., & Finne, J. (2009). Leucine-rich Repeats of Bacterial Surface Proteins Serve as Common Pattern Recognition Motifs of Human Scavenger Receptor gp340. *Journal of Biological Chemistry*, 284(28), 18614-18623. <https://doi.org/10.1074/jbc.M900581200>
- López-Villamizar, I., Cabezas, A., Pinto, R. M., Canales, J., Ribeiro, J. M., Cameselle, J. C., & Costas, M. J. (2016). The Characterization of Escherichia coli CpdB as a Recombinant Protein Reveals that, besides Having the Expected 3'-Nucleotidase and 2',3'-Cyclic Mononucleotide Phosphodiesterase Activities, It Is Also Active as Cyclic Dinucleotide Phosphodiesterase. *PLOS ONE*, 11(6), e0157308. <https://doi.org/10.1371/journal.pone.0157308>
- Lorrain, C., Hecker, A., & Duplessis, S. (2015). Effector-Mining in the Poplar Rust Fungus Melampsora larici-populina Secretome. *Frontiers in Plant Science*, 6. <https://doi.org/10.3389/fpls.2015.01051>
- Lu, J., Wu, T., Zhang, B., Liu, S., Song, W., Qiao, J., & Ruan, H. (2021). Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Communication and Signaling*, 19(1), 60. <https://doi.org/10.1186/s12964-021-00741-y>

- Luan, S. (2002). Tyrosine phosphorylation in plant cell signaling. *Proceedings of the National Academy of Sciences*, 99(18), 11567-11569. <https://doi.org/10.1073/pnas.182417599>
- MacLean, A. M., Orlovskis, Z., Kowitwanich, K., Zdziarska, A. M., Angenent, G. C., Immink, R. G. H., & Hogenhout, S. A. (2014). Phytoplasma Effector SAP54 Hijacks Plant Reproduction by Degrading MADS-box Proteins and Promotes Insect Colonization in a RAD23-Dependent Manner. *PLoS Biology*, 12(4), e1001835. <https://doi.org/10.1371/journal.pbio.1001835>
- MacLean, A. M., Sugio, A., Makarova, O. V., Findlay, K. C., Grieve, V. M., Tóth, R., Nicolaisen, M., & Hogenhout, S. A. (2011). Phytoplasma Effector SAP54 Induces Indeterminate Leaf-Like Flower Development in Arabidopsis Plants. *Plant Physiology*, 157(2), 831-841. <https://doi.org/10.1104/pp.111.181586>
- Maejima, K., Iwai, R., Himeno, M., Komatsu, K., Kitazawa, Y., Fujita, N., Ishikawa, K., Fukuoka, M., Minato, N., Yamaji, Y., Oshima, K., & Namba, S. (2014). Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. *The Plant Journal*, 78(4), 541-554. <https://doi.org/10.1111/tpj.12495>
- Maidment, J. H., Shimizu, M., Bentham, A. R., Vera, S., Franceschetti, M., Longya, A., Stevenson, C. E., De La Concepcion, J. C., Bialas, A., Kamoun, S., Terauchi, R., & Banfield, M. J. (2023). Effector target-guided engineering of an integrated domain expands the disease resistance profile of a rice NLR immune receptor. *eLife*, 12, e81123. <https://doi.org/10.7554/eLife.81123>
- Mai-Prochnow, A., Clauson, M., Hong, J., & Murphy, A. B. (2016). Gram positive and Gram negative bacteria differ in their sensitivity to cold plasma. *Scientific Reports*, 6(1), 38610. <https://doi.org/10.1038/srep38610>

- Mak, H., & Thurston, T. L. M. (2021). Interesting Biochemistries in the Structure and Function of Bacterial Effectors. *Frontiers in Cellular and Infection Microbiology*, 11, 608860. <https://doi.org/10.3389/fcimb.2021.608860>
- Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, 38(10), 4647-4654. <https://doi.org/10.1093/molbev/msab199>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10. <https://doi.org/10.14806/ej.17.1.200>
- Martínez Pacheco, J. (2015). *Efectores de oomycetes fitopatogénicos: En la primera línea de ataque*. 19(3), 251-257.
- Martins, Y. C., Jurberg, A. D., & Daniel-Ribeiro, C. T. (2023). Visiting Molecular Mimicry Once More: Pathogenicity, Virulence, and Autoimmunity. *Microorganisms*, 11(6), 1472. <https://doi.org/10.3390/microorganisms11061472>
- Meir, A., Macé, K., Vegunta, Y., Williams, S. M., & Waksman, G. (2023). Substrate recruitment mechanism by gram-negative type III, IV, and VI bacterial injectisomes. *Trends in Microbiology*, 31(9), 916-932. <https://doi.org/10.1016/j.tim.2023.03.005>
- Mendoza González, P., & Encina García, A. E. (2018). Bases moleculares de la floración. *Ambiociencias*, 31-42. <https://doi.org/10.18002/ambioc.v0i0.5557>
- Mészáros, B., Sámano-Sánchez, H., Alvarado-Valverde, J., Čalyševa, J., Martínez-Pérez, E., Alves, R., Shields, D. C., Kumar, M., Rippmann, F., Chemes, L. B., & Gibson, T. J. (2021). Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications. *Science Signaling*, 14(665), eabd0334. <https://doi.org/10.1126/scisignal.abd0334>

- Miao, E. A., Scherer, C. A., Tsolis, R. M., Kingsley, R. A., Adams, L. G., Baumler, A. J., & Miller, S. I. (1999). Salmonella typhimurium leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. *Molecular Microbiology*, 34(4), 850-864. <https://doi.org/10.1046/j.1365-2958.1999.01651.x>
- Milo, B. (2001). *Plantas medicinales y drogas vegetales. Abril 2001. Cocotero. Cocos nucifera L. (arecácea). Unidad de Farmacología y Farmacognosia. Facultad de Farmacia. Universidad de Barcelona.* 195-196.
- Minato, N., Himeno, M., Hoshi, A., Maejima, K., Komatsu, K., Takebayashi, Y., Kasahara, H., Yusa, A., Yamaji, Y., Oshima, K., Kamiya, Y., & Namba, S. (2014). The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of the jasmonic acid and auxin pathways. *Scientific Reports*, 4(1), 7399. <https://doi.org/10.1038/srep07399>
- Miseta, A., & Csutora, P. (2000). Relationship Between the Occurrence of Cysteine in Proteins and the Complexity of Organisms. *Molecular Biology and Evolution*, 17(8), 1232-1239. <https://doi.org/10.1093/oxfordjournals.molbev.a026406>
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412-D419. <https://doi.org/10.1093/nar/gkaa913>
- Mittelberger, C., Stellmach, H., Hause, B., Kerschbamer, C., Schlink, K., Letschka, T., & Janik, K. (2019). A Novel Effector Protein of Apple Proliferation Phytoplasma Disrupts Cell Integrity of Nicotiana spp. Protoplasts. *International Journal of Molecular Sciences*, 20(18), 4613. <https://doi.org/10.3390/ijms20184613>
- Moles, E., Valle-Delgado, J. J., Urbán, P., Azcárate, I. G., Bautista, J. M., Selva, J., Egea, G., Ventura, S., & Fernández-Busquets, X. (2015). Possible roles of amyloids in

- malaria pathophysiology. *Future Science OA*, 1(2), fso.15.43.
<https://doi.org/10.4155/fso.15.43>
- Mondino, S., Schmidt, S., & Buchrieser, C. (2020). Molecular Mimicry: A Paradigm of Host-Microbe Coevolution Illustrated by *Legionella*. *mBio*, 11(5), e01201-20.
<https://doi.org/10.1128/mBio.01201-20>
- Morgado, S., & Vicente, A. C. (2022). Diversity and distribution of Type VI Secretion System gene clusters in bacterial plasmids. *Scientific Reports*, 12(1), 8249.
<https://doi.org/10.1038/s41598-022-12382-3>
- Mukhtar, M. S., McCormack, M. E., Argueso, C. T., & Pajerowska-Mukhtar, K. M. (2016). Pathogen Tactics to Manipulate Plant Cell Death. *Current Biology*, 26(13), R608-R619. <https://doi.org/10.1016/j.cub.2016.02.051>
- Muralidharan, V., & Goldberg, D. E. (2013). Asparagine Repeats in Plasmodium falciparum Proteins: Good for Nothing? *PLoS Pathogens*, 9(8), e1003488.
<https://doi.org/10.1371/journal.ppat.1003488>
- Music, M. S., Samarzija, I., Hogenhout, S. A., Haryono, M., Cho, S.-T., & Kuo, C.-H. (2019). The genome of 'Candidatus Phytoplasma solani' strain SA-1 is highly dynamic and prone to adopting foreign sequences. *Systematic and Applied Microbiology*, 42(2), 117-127. <https://doi.org/10.1016/j.syapm.2018.10.008>
- Nagpal, P., Ellis, C. M., Weber, H., Ploense, S. E., Barkawi, L. S., Guilfoyle, T. J., Hagen, G., Alonso, J. M., Cohen, J. D., Farmer, E. E., Ecker, J. R., & Reed, J. W. (2005). Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development*, 132(18), 4107-4118.
<https://doi.org/10.1242/dev.01955>
- Nair, S., & Manimekalai, R. (2021). Phytoplasma diseases of plants: Molecular diagnostics and way forward. *World Journal of Microbiology and Biotechnology*, 37(6), 102.
<https://doi.org/10.1007/s11274-021-03061-y>

- Narita, T., Weinert, B. T., & Choudhary, C. (2019). Functions and mechanisms of non-histone protein acetylation. *Nature Reviews Molecular Cell Biology*, 20(3), 156-174. <https://doi.org/10.1038/s41580-018-0081-3>
- Narváez, M., Vázquez-Euán, R., Harrison, N. A., Nic-Matos, G., Julia, J. F., Dzido, J. L., Fabre, S., Dollet, M., & Oropeza, C. (2018). Presence of 16SrIV phytoplasmas of subgroups A, D and E in planthopper *Haplaxius crudus* Van Duzee insects in Yucatán, Mexico. *3 Biotech*, 8(1), 61. <https://doi.org/10.1007/s13205-018-1094-5>
- Nasrullah, Hussain, A., Ahmed, S., Rasool, M., & Shah, A. J. (2022). DNA methylation across the tree of life, from micro to macro-organism. *Bioengineered*, 13(1), 1666-1685. <https://doi.org/10.1080/21655979.2021.2014387>
- Navarrete, F., Grujic, N., Stirnberg, A., Saado, I., Aleksza, D., Gallei, M., Adi, H., Alcântara, A., Khan, M., Bindics, J., Trujillo, M., & Djamei, A. (2021). The Pleiades are a cluster of fungal effectors that inhibit host defenses. *PLOS Pathogens*, 17(6), e1009641. <https://doi.org/10.1371/journal.ppat.1009641>
- Navarro, S., Marinelli, P., Diaz-Caballero, M., & Ventura, S. (2015). The prion-like RNA-processing protein HNRPDL forms inherently toxic amyloid-like inclusion bodies in bacteria. *Microbial Cell Factories*, 14(1), 102. <https://doi.org/10.1186/s12934-015-0284-7>
- Nguyen Ba, A. N., Pogoutse, A., Provart, N., & Moses, A. M. (2009). NLStradamus: A simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, 10(1), 202. <https://doi.org/10.1186/1471-2105-10-202>
- Nijo, T., Iwabuchi, N., Tokuda, R., Suzuki, T., Matsumoto, O., Miyazaki, A., Maejima, K., Oshima, K., Namba, S., & Yamaji, Y. (2021). Enrichment of phytoplasma genome DNA through a methyl-CpG binding domain-mediated method for efficient genome sequencing. *Journal of General Plant Pathology*, 87(3), 154-163. <https://doi.org/10.1007/s10327-021-00993-z>

- Noroy, C., Lefrançois, T., & Meyer, D. F. (2019). Searching algorithm for Type IV effector proteins (S4TE) 2.0: Improved tools for Type IV effector prediction, analysis and comparison in proteobacteria. *PLOS Computational Biology*, 15(3), e1006847. <https://doi.org/10.1371/journal.pcbi.1006847>
- Oliveira, M. J. R. A., Roriz, M., Vasconcelos, M. W., Bertaccini, A., & Carvalho, S. M. P. (2019). Conventional and novel approaches for managing “flavescence dorée” in grapevine: Knowledge gaps and future prospects. *Plant Pathology*, 68(1), 3-17. <https://doi.org/10.1111/ppa.12938>
- Omelchenko, M. V., Makarova, K. S., Wolf, Y. I., Rogozin, I. B., & Koonin, E. V. (2003). Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. Genome biology. *Genome Biology*, 4(9), R55. <https://doi.org/10.1186/gb-2003-4-9-r55>
- Örd, M., Venta, R., Möll, K., Valk, E., & Loog, M. (2019). Cyclin-Specific Docking Mechanisms Reveal the Complexity of M-CDK Function in the Cell Cycle. *Molecular Cell*, 75(1), 76-89.e3. <https://doi.org/10.1016/j.molcel.2019.04.026>
- Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAOSTAT). <https://www.fao.org/faostat/es/>. Boletín The coconut palm. (s. f.).
- Orlovskis, Z., Canale, M. C., Haryono, M., Lopes, J. R. S., Kuo, C.-H., & Hogenhout, S. A. (2017). A few sequence polymorphisms among isolates of Maize bushy stunt phytoplasma associate with organ proliferation symptoms of infected maize plants. *Annals of Botany*, mcw213. <https://doi.org/10.1093/aob/mcw213>
- Oropeza, C., Sáenz, L., Narvaez, M., Nic-Matos, G., Cordova, I., Myrie, W., Ortiz, C., & Ramos, E. (2020). *Dealing with Lethal Yellowing and Related Diseases in Coconut* (pp. 169-197). https://doi.org/10.1007/978-3-030-44988-9_9

- Oropeza, C., Zizumbo, D., Sáenz, L., Narváez, M., & Córdova, I. (2010). *Interacciones plantapatógeno-vector-ambiente en el amarillamiento letal del cocotero*. 31, 388-402. <https://doi.org/10.56866/issn.0121-2923>.
- Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H.-Y., Wei, W., Suzuki, S., Arashida, R., Nakata, D., Miyata, S., Ugaki, M., & Namba, S. (2004). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nature Genetics*, 36(1), 27-29. <https://doi.org/10.1038/ng1277>
- Oshima, K., Maejima, K., & Namba, S. (2013). Genomic and evolutionary aspects of phytoplasmas. *Frontiers in Microbiology*, 4. <https://doi.org/10.3389/fmicb.2013.00230>
- Pacheco, I. D. S., Galdeano, D. M., Maluta, N. K. P., Lopes, J. R. S., & Machado, M. A. (2020). Gene silencing of *Diaphorina citri* candidate effectors promotes changes in feeding behaviors. *Scientific Reports*, 10(1), 5992. <https://doi.org/10.1038/s41598-020-62856-5>
- Padilla-Ramos, R., Salas-Muñoz, S., Velásquez-Valle, R., & Reveles-Torres, L. R. (2019). Un nuevo enfoque molecular en el estudio de la interacción parásito-hospedero. *Revista Mexicana de Fitopatología, Mexican Journal of Phytopathology*, 37(1). <https://doi.org/10.18781/R.MEX.FIT.1808-6>
- Pallarès, I., De Groot, N. S., Iglesias, V., Sant'Anna, R., Biosca, A., Fernández-Busquets, X., & Ventura, S. (2018). Discovering Putative Prion-Like Proteins in *Plasmodium falciparum*: A Computational and Experimental Analysis. *Frontiers in Microbiology*, 9, 1737. <https://doi.org/10.3389/fmicb.2018.01737>
- Palma de coco mexicana. Secretaría de Agricultura, Ganadería y Desarrollo Rural, Pesca y Alimentación (SAGARPA). Investigación de Planeación Agrícola Nacional 2017-2030. (Primera edición). (2017).*

- Parrello, B., Butler, R., Chlenski, P., Pusch, G. D., & Overbeek, R. (2021). Supervised extraction of near-complete genomes from metagenomic samples: A new service in PATRIC. *PLOS ONE*, 16(4), e0250092. <https://doi.org/10.1371/journal.pone.0250092>
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., ... Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, 51(D1), D418-D427. <https://doi.org/10.1093/nar/gkac993>
- Pearson, J. S., Giogha, C., Mühlen, S., Nachbur, U., Pham, C. L. L., Zhang, Y., Hildebrand, J. M., Oates, C. V., Lung, T. W. F., Ingle, D., Dagley, L. F., Bankovacki, A., Petrie, E. J., Schroeder, G. N., Crepin, V. F., Frankel, G., Masters, S. L., Vince, J., Murphy, J. M., ... Hartland, E. L. (2017). EspL is a bacterial cysteine protease effector that cleaves RHIM proteins to block necroptosis and inflammation. *Nature Microbiology*, 2(4), 16258. <https://doi.org/10.1038/nmicrobiol.2016.258>
- Pecher, P., Moro, G., Canale, M. C., Capdevielle, S., Singh, A., MacLean, A., Sugio, A., Kuo, C.-H., Lopes, J. R. S., & Hogenhout, S. A. (2019). Phytoplasma SAP11 effector destabilization of TCP transcription factors differentially impact development and defence of Arabidopsis versus maize. *PLOS Pathogens*, 15(9), e1008035. <https://doi.org/10.1371/journal.ppat.1008035>
- Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785-786. <https://doi.org/10.1038/nmeth.1701>
- Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785-786. <https://doi.org/10.1038/nmeth.1701>

- Pillay, T. D., Hettiarachchi, S. U., Gan, J., Diaz-Del-Olmo, I., Yu, X.-J., Muench, J. H., Thurston, T. L. M., & Pearson, J. S. (2023). Speaking the host language: How Salmonella effector proteins manipulate the host: This article is part of the Bacterial Cell Envelopes collection. *Microbiology*, 169(6). <https://doi.org/10.1099/mic.0.001342>
- Planeación Agrícola Nacional 2017-2030. Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA). Palma de coco mexicana. (s. f.).*
- Polacco, B. J., & Babbitt, P. C. (2006). Automated discovery of 3D motifs for protein function annotation. *Bioinformatics*, 22(6), 723-730. <https://doi.org/10.1093/bioinformatics/btk038>
- Ponting, C. P. (1997). Evidence for PDZ domains in bacteria, yeast, and plants: Novel PDZ domains. *Protein Science*, 6(2), 464-468. <https://doi.org/10.1002/pro.5560060225>
- Preston, J. C., & Hileman, L. C. (2013). Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Frontiers in Plant Science*, 4. <https://doi.org/10.3389/fpls.2013.00080>
- Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics*, 70(1), e102. <https://doi.org/10.1002/cpbi.102>
- Qi, D., DeYoung, B. J., & Innes, R. W. (2012). Structure-Function Analysis of the Coiled-Coil and Leucine-Rich Repeat Domains of the RPS5 Disease Resistance Protein. *Plant Physiology*, 158(4), 1819-1832. <https://doi.org/10.1104/pp.112.194035>
- Qian, Y., Zheng, X., Wang, X., Yang, J., Zheng, X., Zeng, Q., Li, J., Zhuge, Q., & Xiong, Q. (2022). Systematic identification and functional characterization of the CFEM proteins in poplar fungus *Marssonina brunnea*. *Frontiers in Cellular and Infection Microbiology*, 12, 1045615. <https://doi.org/10.3389/fcimb.2022.1045615>

- Rajapandi, T. (2020). Chaperoning of asparagine repeat-containing proteins in *Plasmodium falciparum*. *Journal of Parasitic Diseases*, 44(4), 687-693.
<https://doi.org/10.1007/s12639-020-01251-3>
- Ramos-Hernández, E., Torres de la Cruz, M., Oropeza-Salín, C., Ortiz-García, C.F., Leshner-Gordillo, J., & Magaña-Alejandro, M.A. (2018). Manejo del agroecosistema cocotero, con énfasis en el amarillamiento letal del cocotero (*Cocos nucifera* L.). *Agroproductividad*, 11(1), 80-87.
<https://www.researchgate.net/publication/335565608>.
- Rashid, U., Bilal, S., Bhat, K. A., Shah, T. A., Wani, T. A., Bhat, F. A., Mughal, M. N., & Nazir, N. (2018). Phytoplasma Effectors and their Role in Plant-Insect Interaction. *International Journal of Current Microbiology and Applied Sciences*, 7(2), 1136-1148.
<https://doi.org/10.20546/ijcmas.2018.702.141>
- Rashidi, M., Galetto, L., Bosco, D., Bulgarelli, A., Vallino, M., Veratti, F., & Marzachi, C. (2015). Role of the major antigenic membrane protein in phytoplasma transmission by two insect vector species. *BMC Microbiology*, 15(1), 193.
<https://doi.org/10.1186/s12866-015-0522-5>
- Remigi, P., Anisimova, M., Guidot, A., Genin, S., & Peeters, N. (2011). Functional diversification of the *GALA* type III effector family contributes to *Ralstonia solanacearum* adaptation on different plant hosts. *New Phytologist*, 192(4), 976-987.
<https://doi.org/10.1111/j.1469-8137.2011.03854.x>
- Reveles-Torres, Padilla-Ramos R., Salas-Muñoz S., & Velásquez-Valle R. (2017). Diseño de transmisión de fitoplasmas a *Catharanthus roseus* como reservorio natural. *Folleto Técnico. Campo Experimental Zacatecas. CIRNOC-INIFAP*, 87, 1-57.
- Rivas, S., & Genin, S. (2011). A Plethora of Virulence Strategies Hidden Behind Nuclear Targeting of Microbial Effectors. *Frontiers in Plant Science*, 2.
<https://doi.org/10.3389/fpls.2011.00104>

- Rocafort, M., Bowen, J. K., Hassing, B., Cox, M. P., McGreal, B., De La Rosa, S., Plummer, K. M., Bradshaw, R. E., & Mesarich, C. H. (2022). The *Venturia inaequalis* effector repertoire is dominated by expanded families with predicted structural similarity, but unrelated sequence, to avirulence proteins from other plant-pathogenic fungi. *BMC Biology*, 20(1), 246. <https://doi.org/10.1186/s12915-022-01442-9>
- Rocafort, M., Fudal, I., & Mesarich, C. H. (2020). Apoplastic effector proteins of plant-associated fungi and oomycetes. *Current Opinion in Plant Biology*, 56, 9-19. <https://doi.org/10.1016/j.pbi.2020.02.004>
- Rodriguez-R, L. M., Gunturu, S., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., Cole, J. R., & Konstantinidis, K. T. (2018). The Microbial Genomes Atlas (MiGA) webserver: Taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Research*, 46(W1), W282-W288. <https://doi.org/10.1093/nar/gky467>
- Rozano, L., Jones, D. A. B., Hane, J. K., & Mancera, R. L. (2023). Template-Based Modelling of the Structure of Fungal Effector Proteins. *Molecular Biotechnology*. <https://doi.org/10.1007/s12033-023-00703-4>
- Ruan, J., Zhou, Y., Zhou, M., Yan, J., Khurshid, M., Weng, W., Cheng, J., & Zhang, K. (2019). Jasmonic Acid Signaling Pathway in Plants. *International Journal of Molecular Sciences*, 20(10), 2479. <https://doi.org/10.3390/ijms20102479>
- Ruhanen, H., Hurley, D., Ghosh, A., O'Brien, K. T., Johnston, C. R., & Shields, D. C. (2014). Potential of known and short prokaryotic protein motifs as a basis for novel peptide-based antibacterial therapeutics: A computational survey. *Frontiers in Microbiology*, 5. <https://doi.org/10.3389/fmicb.2014.00004>
- Russell, S. L. (2019). Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: A systematic review and meta-analysis. *FEMS Microbiology Letters*, 366(3). <https://doi.org/10.1093/femsle/fnz013>

- Saedler, H., Becker, A., Winter, K. U., Kirchner, C., & Theissen, G. (2001). MADS-box genes are involved in floral development and evolution. *Acta Biochimica Polonica*, 48(2), 351-358.
- Sámano-Sánchez, H., & Gibson, T. J. (2020). Mimicry of Short Linear Motifs by Bacterial Pathogens: A Drugging Opportunity. *Trends in Biochemical Sciences*, 45(6), 526-544. <https://doi.org/10.1016/j.tibs.2020.03.003>
- Sampietro, D., Sámano-Sánchez, H., Davey, N. E., Sharan, M., Mészáros, B., Gibson, T. J., & Kumar, M. (2018). *Conserved SQ and QS motifs in bacterial effectors suggest pathogen interplay with the ATM kinase family during infection* [Preprint]. Bioinformatics. <https://doi.org/10.1101/364117>
- Saunders, D. G. O., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., & Raffaele, S. (2012). Using Hierarchical Clustering of Secreted Protein Families to Classify and Rank Candidate Effectors of Rust Fungi. *PLoS ONE*, 7(1), e29847. <https://doi.org/10.1371/journal.pone.0029847>
- Sautua, F., & Carmona, M. (2021). *Cátedra de Fitopatología. Herbario virtual*.
- Schmitz, R. J., Lewis, Z. A., & Goll, M. G. (2019). DNA Methylation: Shared and Divergent Features across Eukaryotes. *Trends in Genetics*, 35(11), 818-827. <https://doi.org/10.1016/j.tig.2019.07.007>
- Scholthof, K.-B. G. (2007). The disease triangle: Pathogens, the environment and society. *Nature Reviews Microbiology*, 5(2), 152-156. <https://doi.org/10.1038/nrmicro1596>
- Schweizer, G., Münch, K., Mannhaupt, G., Schirawski, J., Kahmann, R., & Dutheil, J. Y. (2018). Positively Selected Effector Genes and Their Contribution to Virulence in the Smut Fungus *Sporisorium reilianum*. *Genome Biology and Evolution*, 10(2), 629-645. <https://doi.org/10.1093/gbe/evy023>

- Secretaría de Agricultura y Desarrollo Rural (SADER).
<https://www.gob.mx/agricultura/es/articulos/produccion-de-copra-y-coco-en-mexico>,. (2022).
- Seong, K., & Krasileva, K. V. (2021). Computational Structural Genomics Unravels Common Folds and Novel Families in the Secretome of Fungal Phytopathogen *Magnaporthe oryzae*. *Molecular Plant-Microbe Interactions*, 34(11), 1267-1280. <https://doi.org/10.1094/MPMI-03-21-0071-R>
- Shahnawaz, M., Park, K.-W., Mukherjee, A., Diaz-Espinoza, R., & Soto, C. (2017). Prion-like characteristics of the bacterial protein Microcin E492. *Scientific Reports*, 7(1), 45720. <https://doi.org/10.1038/srep45720>
- Shiu, S.-H., & Bleecker, A. B. (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proceedings of the National Academy of Sciences*, 98(19), 10763-10768. <https://doi.org/10.1073/pnas.181141598>
- Singh, A., & Lakhanpaul, S. (2020). Detection, characterization and evolutionary aspects of S54LP of SP (SAP54 Like Protein of Sesame Phyllody): A phytoplasma effector molecule associated with phyllody development in sesame (*Sesamum indicum* L.). *Physiology and Molecular Biology of Plants*, 26(3), 445-458. <https://doi.org/10.1007/s12298-020-00764-8>
- Singh, A., Verma, P., & Lakhanpaul, S. (2019). Phytoplasma effector molecules and their structural aspects: A review. *Phytopathogenic Mollicutes*, 9(2), 241. <https://doi.org/10.5958/2249-4677.2019.00121.X>
- Singh, Y., Nair, A. M., & Verma, P. K. (2021). Surviving the odds: From perception to survival of fungal phytopathogens under host-generated oxidative burst. *Plant Communications*, 2(3), 100142. <https://doi.org/10.1016/j.xplc.2021.100142>

- Solé, M., Popa, C., Mith, O., Sohn, K. H., Jones, J. D. G., Deslandes, L., & Valls, M. (2012). The *awr* Gene Family Encodes a Novel Class of *Ralstonia solanacearum* Type III Effectors Displaying Virulence and Avirulence Activities. *Molecular Plant-Microbe Interactions*, 25(7), 941-953. <https://doi.org/10.1094/MPMI-12-11-0321>
- Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology*, 19(9), 2094-2110. <https://doi.org/10.1111/mpp.12682>
- Stergiopoulos, I., Van Den Burg, H. A., Ökmen, B., Beenen, H. G., Van Lier, S., Kema, G. H. J., & De Wit, P. J. G. M. (2010). Tomato Cf resistance proteins mediate recognition of cognate homologous effectors from fungi pathogenic on dicots and monocots. *Proceedings of the National Academy of Sciences*, 107(16), 7610-7615. <https://doi.org/10.1073/pnas.1002910107>
- Stern, D. L., & Han, C. (2022). Gene Structure-Based Homology Search Identifies Highly Divergent Putative Effector Gene Family. *Genome Biology and Evolution*, 14(6), evac069. <https://doi.org/10.1093/gbe/evac069>
- Strohmayr, A., Schwarz, T., Braun, M., Krczal, G., & Boonrod, K. (2021). The Effect of the Anticipated Nuclear Localization Sequence of 'Candidatus Phytoplasma mali' SAP11-like Protein on Localization of the Protein and Destabilization of TCP Transcription Factor. *Microorganisms*, 9(8), 1756. <https://doi.org/10.3390/microorganisms9081756>
- Sugawara, K., Honma, Y., Komatsu, K., Himeno, M., Oshima, K., & Namba, S. (2013). The Alteration of Plant Morphology by Small Peptides Released from the Proteolytic Processing of the Bacterial Peptide TENGU. *Plant Physiology*, 162(4), 2005-2014. <https://doi.org/10.1104/pp.113.218586>
- Sugio, A., Kingdom, H. N., MacLean, A. M., Grieve, V. M., & Hogenhout, S. A. (2011b). Phytoplasma protein effector SAP11 enhances insect vector reproduction by

- manipulating plant development and defense hormone biosynthesis. *Proceedings of the National Academy of Sciences*, 108(48).
<https://doi.org/10.1073/pnas.1105664108>
- Sugio, A., MacLean, A. M., & Hogenhout, S. A. (2014). The small phytoplasma virulence effector SAP 11 contains distinct domains required for nuclear targeting and CIN - TCP binding and destabilization. *New Phytologist*, 202(3), 838-848.
<https://doi.org/10.1111/nph.12721>
- Sugio, A., MacLean, A. M., Kingdom, H. N., Grieve, V. M., Manimekalai, R., & Hogenhout, S. A. (2011a). Diverse Targets of Phytoplasma Effectors: From Plant Development to Defense Against Insects. *Annual Review of Phytopathology*, 49(1), 175-195.
<https://doi.org/10.1146/annurev-phyto-072910-095323>
- Taiz, L., & Zeiger, E. (2002). *Plant Physiology*. (3rd.). Sinauer Associates.
- Tan, Lin, Y.-C., Li, J.-R., Chien, Y.-Y., Wang, C.-J., Chou, L., Wang, C.-W., Chiu, Y.-C., Kuo, C.-H., & Yang, J.-Y. (2021). Accelerating Complete Phytoplasma Genome Assembly by Immunoprecipitation-Based Enrichment and MinION-Based DNA Sequencing for Comparative Analyses. *Frontiers in Microbiology*, 12, 766221.
<https://doi.org/10.3389/fmicb.2021.766221>
- Tanaka, S., & Kahmann, R. (2021). Cell wall–associated effectors of plant-colonizing fungi. *Mycologia*, 113(2), 247-260. <https://doi.org/10.1080/00275514.2020.1831293>
- Tayal, S., Bhatia, V., Mehrotra, T., & Bhatnagar, S. (2022). ImitateDB: A database for domain and motif mimicry incorporating host and pathogen protein interactions. *Amino Acids*, 54(6), 923-934. <https://doi.org/10.1007/s00726-022-03163-3>
- Tetz, G., & Tetz, V. (2017). Prion-Like Domains in Phagobiota. *Frontiers in Microbiology*, 8, 2239. <https://doi.org/10.3389/fmicb.2017.02239>

- Tetz, G., & Tetz, V. (2022). Prion-like Domains in Spike Protein of SARS-CoV-2 Differ across Its Variants and Enable Changes in Affinity to ACE2. *Microorganisms*, 10(2), 280. <https://doi.org/10.3390/microorganisms10020280>
- Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J. T., Münster, T., Winter, K. U., & Saedler, H. (2000). A short history of MADS-box genes in plants. *Plant Molecular Biology*, 42(1), 115-149.
- Theißen, G., Melzer, R., & Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: Linking plant development and evolution. *Development*, 143(18), 3259-3271. <https://doi.org/10.1242/dev.134080>
- Theißen, G., & Saedler, H. (2001). Plant biology. Floral quartets. *Nature*, 409(6819), 469-471. <https://doi.org/10.1038/35054172>
- Thomas, E., & Van Der Hoorn, R. (2018). Ten Prominent Host Proteases in Plant-Pathogen Interactions. *International Journal of Molecular Sciences*, 19(2), 639. <https://doi.org/10.3390/ijms19020639>
- Tiwari, P., & Bae, H. (2020). Horizontal Gene Transfer and Endophytes: An Implication for the Acquisition of Novel Traits. *Plants*, 9(3), 305. <https://doi.org/10.3390/plants9030305>
- Todd, J. N. A., Carreón-Anguiano, K. G., Islas-Flores, I., & Canto-Canché, B. (2022). Fungal Effectoromics: A World in Constant Evolution. *International Journal of Molecular Sciences*, 23(21), 13433. <https://doi.org/10.3390/ijms232113433>
- Tokuda, R., Iwabuchi, N., Kitazawa, Y., Nijo, T., Suzuki, M., Maejima, K., Oshima, K., Namba, S., & Yamaji, Y. (2023). Potential mobile units drive the horizontal transfer of phytoplasma effector phylogen genes. *Frontiers in Genetics*, 14, 1132432. <https://doi.org/10.3389/fgene.2023.1132432>
- Tomkins, M., Kliot, A., Marée, A. F., & Hogenhout, S. A. (2018). A multi-layered mechanistic modelling approach to understand how effector genes extend beyond phytoplasma

- to modulate plant hosts, insect vectors and the environment. *Current Opinion in Plant Biology*, 44, 39-48. <https://doi.org/10.1016/j.pbi.2018.02.002>
- Torres Pacheco, I., Guevara, R. G., González, C. A., Mejía Teniente, L., & Chapa Oliver, A. M. (2012). *Relación de la Inmunología vegetal y la producción de alimentos*. 5(1).
- Tran-Nguyen, L. T. T., Kube, M., Schneider, B., Reinhardt, R., & Gibb, K. S. (2008). Comparative Genome Analysis of “*Candidatus* Phytoplasma australiense” (Subgroup *tuf* -Australia I; *rp* -A) and “*Ca. Phytoplasma asteris*” Strains OY-M and AY-WB. *Journal of Bacteriology*, 190(11), 3979-3991. <https://doi.org/10.1128/JB.01301-07>
- Trivellone, V., & Dietrich, C. H. (2021). Evolutionary Diversification in Insect Vector–Phytoplasma–Plant Associations. *Annals of the Entomological Society of America*, 114(2), 137-150. <https://doi.org/10.1093/aesa/saaa048>
- Tully, J. G., Bove, J. M., Laigret, F., & Whitcomb, R. F. (1993). Notes: Revised Taxonomy of the Class Mollicutes: Proposed Elevation of a Monophyletic Cluster of Arthropod-Associated Mollicutes to Ordinal Rank (Entomoplasmatales ord. nov.), with Provision for Familial Rank To Separate Species with Nonhelical Morphology (Entomoplasmataceae fam. nov.) from Helical Species (Spiroplasmataceae), and Emended Descriptions of the Order Mycoplasmatales, Family Mycoplasmataceae. *International Journal of Systematic Bacteriology*, 43(2), 378-385. <https://doi.org/10.1099/00207713-43-2-378>
- Tzec-Simá, M., Félix, J. W., Granados-Alegría, M., Aparicio Ortiz, M., Juárez-Monroy, D., Mayo-Ruiz, D., Vivas-López, S., Gómez-Tah, R., Canto-Canché, B., & Berezovski, M. V. (2022). *Potential of Omics to Control Diseases and Pests in the Coconut Tree*.
- Upadhyay, S., & Shaw, B. D. (2008). The role of actin, fimbrin and endocytosis in growth of hyphae in *Aspergillus nidulans*. *Molecular Microbiology*, 68(3), 690-705. <https://doi.org/10.1111/j.1365-2958.2008.06178.x>

- Valois-Cuesta, H., Martínez-Ruiz, C., Rentería Cuesta, Y. Y., & Sol María, P. H. (2013). Diversidad, patrones de uso y conservación de palmas (Arecaceae) en bosques pluviales del Chocó, Colombia. [Palms conservation, patterns of use and diversity (Arecaceae) in rain forests from the Chocó, Colombia]. *Revista De Biología Tropical*, 61(4), 1869-1889.
- Van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., Budd, A., Gibson, T. J., & Davey, N. E. (2014). Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chemical Reviews*, 114(13), 6733-6778. <https://doi.org/10.1021/cr400585q>
- Vázquez-Euán, R., Harrison, N., Narvaez, M., & Oropeza, C. (2011). Occurrence of a 16SrIV Group Phytoplasma not Previously Associated with Palm Species in Yucatan, Mexico. *Plant Disease*, 95(3), 256-262. <https://doi.org/10.1094/PDIS-02-10-0150>
- Vleeshouwers, V. G. A. A., Rietman, H., Krennek, P., Champouret, N., Young, C., Oh, S.-K., Wang, M., Bouwmeester, K., Vosman, B., Visser, R. G. F., Jacobsen, E., Govers, F., Kamoun, S., & Van Der Vossen, E. A. G. (2008). Effector Genomics Accelerates Discovery and Functional Profiling of Potato Disease Resistance and Phytophthora Infestans Avirulence Genes. *PLoS ONE*, 3(8), e2875. <https://doi.org/10.1371/journal.pone.0002875>
- Wadie, B., Kleshchevnikov, V., Sandaltzopoulou, E., Benz, C., & Petsalaki, E. (2022). Use of viral motif mimicry improves the proteome-wide discovery of human linear motifs. *Cell Reports*, 39(5), 110764. <https://doi.org/10.1016/j.celrep.2022.110764>
- Wagner, N., Teper, D., & Pupko, T. (2022). Predicting Type III Effector Proteins Using the Effectidor Web Server. En O. Gal-Mor (Ed.), *Bacterial Virulence* (Vol. 2427, pp. 25-36). Springer US. https://doi.org/10.1007/978-1-0716-1971-1_3

- Wang, H., Shi, S., & Hua, W. (2023). Advances of herbivore-secreted elicitors and effectors in plant-insect interactions. *Frontiers in Plant Science*, 14, 1176048. <https://doi.org/10.3389/fpls.2023.1176048>
- Wang, J., Lian, N., Zhang, Y., Man, Y., Chen, L., Yang, H., Lin, J., & Jing, Y. (2022). The Cytoskeleton in Plant Immunity: Dynamics, Regulation, and Function. *International Journal of Molecular Sciences*, 23(24), 15553. <https://doi.org/10.3390/ijms232415553>
- Wang, J.-W., Czech, B., & Weigel, D. (2009). miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in *Arabidopsis thaliana*. *Cell*, 138(4), 738-749. <https://doi.org/10.1016/j.cell.2009.06.014>
- Wang, N., Li, Y., Chen, W., Yang, H. Z., Zhang, P. H., & Wu, Y. F. (2017). Identification of wheat blue dwarf phytoplasma effectors targeting plant proliferation and defence responses. *Plant Pathology*, 67(3), 603-609. <https://doi.org/10.1111/ppa.12786>
- Wang, N., Yang, H., Yin, Z., Liu, W., Sun, L., & Wu, Y. (2018b). Phytoplasma effector SWP1 induces witches' broom symptom by destabilizing the TCP transcription factor BRANCHED1: Phytoplasma effector SWP1 destabilizes BRC1. *Molecular Plant Pathology*, 19(12), 2623-2634. <https://doi.org/10.1111/mpp.12733>
- Wang, Song, L., Jiao, Q., Yang, S., Gao, R., Lu, X., & Zhou, G. (2018a). Comparative genome analysis of jujube witches'-broom Phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. *BMC Genomics*, 19(1), 689. <https://doi.org/10.1186/s12864-018-5075-1>
- Wang, X., Davlieva, M., Reyes, J., Panesso, D., Arias, C. A., & Shamoo, Y. (2017). A Novel Phosphodiesterase of the GdpP Family Modulates Cyclic di-AMP Levels in Response to Cell Membrane Stress in Daptomycin-Resistant Enterococci. *Antimicrobial Agents and Chemotherapy*, 61(3), e01422-16. <https://doi.org/10.1128/AAC.01422-16>

- Wang, X.-Y., Zhang, R.-Y., Li, J., Li, Y.-H., Shan, H.-L., Li, W.-F., & Huang, Y.-K. (2022). The Diversity, Distribution and Status of Phytoplasma Diseases in China. *Frontiers in Sustainable Food Systems*, 6, 943080. <https://doi.org/10.3389/fsufs.2022.943080>
- Wang, Y., & Wang, Y. (2018b). Trick or Treat: Microbial Pathogens Evolved Apoplastic Effectors Modulating Plant Susceptibility to Infection. *Molecular Plant-Microbe Interactions*, 31(1), 6-12. <https://doi.org/10.1094/MPMI-07-17-0177-FI>
- Wang, Z., Yang, X., Zhou, S., Zhang, X., Zhu, Y., Chen, B., Huang, X., Yang, X., Zhou, G., & Zhang, T. (2023). The Antigenic Membrane Protein (Amp) of Rice Orange Leaf Phytoplasma Suppresses Host Defenses and Is Involved in Pathogenicity. *International Journal of Molecular Sciences*, 24(5), 4494. <https://doi.org/10.3390/ijms24054494>
- Weerasekera, D., Stengel, F., Sticht, H., De Mattos Guaraldi, A. L., Burkovski, A., & Azevedo Antunes, C. (2018). The C-terminal coiled-coil domain of *Corynebacterium diphtheriae* DIP0733 is crucial for interaction with epithelial cells and pathogenicity in invertebrate animal model systems. *BMC Microbiology*, 18(1), 106. <https://doi.org/10.1186/s12866-018-1247-z>
- Wei, W., Davis, R. E., Jomantiene, R., & Zhao, Y. (2008). Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. *Proceedings of the National Academy of Sciences*, 105(33), 11827-11832. <https://doi.org/10.1073/pnas.0805237105>
- Wei, W., & Zhao, Y. (2022). Phytoplasma Taxonomy: Nomenclature, Classification, and Identification. *Biology*, 11(8), 1119. <https://doi.org/10.3390/biology11081119>
- Weintraub, P. G., & Beanland, L. (2006). INSECT VECTORS OF PHYTOPLASMAS. *Annual Review of Entomology*, 51(1), 91-111. <https://doi.org/10.1146/annurev.ento.51.110104.151039>

- Wilson, R. A., & McDowell, J. M. (2022). Recent advances in understanding of fungal and oomycete effectors. *Current Opinion in Plant Biology*, 68, 102228. <https://doi.org/10.1016/j.pbi.2022.102228>
- Wood, K. J., Nur, M., Gil, J., Fletcher, K., Lakeman, K., Gann, D., Gothberg, A., Khuu, T., Kopetzky, J., Naqvi, S., Pandya, A., Zhang, C., Maisonneuve, B., Pel, M., & Michelmore, R. (2020). Effector prediction and characterization in the oomycete pathogen *Bremia lactucae* reveal host-recognized WY domain proteins that lack the canonical RXLR motif. *PLOS Pathogens*, 16(10), e1009012. <https://doi.org/10.1371/journal.ppat.1009012>
- Xu, M., Hu, T., Zhao, J., Park, M.-Y., Earley, K. W., Wu, G., Yang, L., & Poethig, R. S. (2016). Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in *Arabidopsis thaliana*. *PLOS Genetics*, 12(8), e1006263. <https://doi.org/10.1371/journal.pgen.1006263>
- Xu, R.-Q., Blanvillain, S., Feng, J.-X., Jiang, B.-L., Li, X.-Z., Wei, H.-Y., Kroj, T., Lauber, E., Roby, D., Chen, B., He, Y.-Q., Lu, G.-T., Tang, D.-J., Vasse, J., Arlat, M., & Tang, J.-L. (2008). AvrAC_{Xcc8004}, a Type III Effector with a Leucine-Rich Repeat Domain from *Xanthomonas campestris* Pathovar *campestris* Confers Avirulence in Vascular Tissues of *Arabidopsis thaliana* Ecotype Col-0. *Journal of Bacteriology*, 190(1), 343-355. <https://doi.org/10.1128/JB.00978-07>
- Xue, C., Zhang, Y., Li, H., Liu, Z., Gao, W., Liu, M., Wang, H., Liu, P., & Zhao, J. (2023). The genome of Candidatus phytoplasma ziziphi provides insights into their biological characteristics. *BMC Plant Biology*, 23(1), 251. <https://doi.org/10.1186/s12870-023-04243-6>
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., & Wang, J. (2006). WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research*, 34(Web Server), W293-W297. <https://doi.org/10.1093/nar/gkl031>

- Yigit, E., Feehery, G. R., Langhorst, B. W., Stewart, F. J., Dimalanta, E. T., Pradhan, S., Slatko, B., Gardner, A. F., McFarland, J., Sumner, C., & Davis, T. B. (2016). A Microbiome DNA Enrichment Method for Next-Generation Sequencing Sample Preparation. *Current Protocols in Molecular Biology*, 115(1). <https://doi.org/10.1002/cpmb.12>
- Yokota, A., Tsumoto, K., Shiroishi, M., Nakanishi, T., Kondo, H., & Kumagai, I. (2010). Contribution of Asparagine Residues to the Stabilization of a Proteinaceous Antigen-Antibody Complex, HyHEL-10-Hen Egg White Lysozyme. *Journal of Biological Chemistry*, 285(10), 7686-7696. <https://doi.org/10.1074/jbc.M109.089623>
- Yong, E. (2014). Bacterial tricks for turning plants into zombies. *Nature*, nature.2014.15011. <https://doi.org/10.1038/nature.2014.15011>
- Zdrzałek, R., Stone, C., De La Concepcion, J. C., Banfield, M. J., & Bentham, A. R. (2023). Pathways to engineering plant intracellular NLR immune receptors. *Current Opinion in Plant Biology*, 74, 102380. <https://doi.org/10.1016/j.pbi.2023.102380>
- Zhang, S., Li, C., Si, J., Han, Z., & Chen, D. (2022). Action Mechanisms of Effectors in Plant-Pathogen Interaction. *International Journal of Molecular Sciences*, 23(12), 6758. <https://doi.org/10.3390/ijms23126758>
- Zhang, Wang, Liu, Fan, Zhang, Zhou, & Li. (2019). A Novel 'Candidatus Liberibacter asiaticus'-Encoded Sec-Dependent Secretory Protein Suppresses Programmed Cell Death in *Nicotiana benthamiana*. *International Journal of Molecular Sciences*, 20(22), 5802. <https://doi.org/10.3390/ijms20225802>
- Zhang, X., Zhou, Y., Ding, L., Wu, Z., Liu, R., & Meyerowitz, E. M. (2013). Transcription Repressor HANABA TARANU Controls Flower Development by Integrating the Actions of Multiple Hormones, Floral Organ Specification Genes, and GATA3 Family Genes in *Arabidopsis*. *The Plant Cell*, 25(1), 83-101. <https://doi.org/10.1105/tpc.112.107854>

- Zhang, Z.-H., Jin, J.-H., Sheng, G.-L., Xing, Y.-P., Liu, W., Zhou, X., Liu, Y.-Q., & Chen, X.-R. (2021). A Small Cysteine-Rich Phytotoxic Protein of *Phytophthora capsici* Functions as Both Plant Defense Elicitor and Virulence Factor. *Molecular Plant-Microbe Interactions*[®], 34(8), 891-903. <https://doi.org/10.1094/MPMI-01-21-0025-R>
- Zhao, J., & Song, J. (2021). NLR immune receptor RB is differentially targeted by two homologous but functionally distinct effector proteins. *Plant Communications*, 2(6), 100236. <https://doi.org/10.1016/j.xplc.2021.100236>
- Zhu, Q., Gao, S., Xiao, B., He, Z., & Hu, S. (2023). Plasmer: An Accurate and Sensitive Bacterial Plasmid Prediction Tool Based on Machine Learning of Shared k-mers and Genomic Features. *Microbiology Spectrum*, 11(3), e04645-22. <https://doi.org/10.1128/spectrum.04645-22>
- Zhu, W., Wei, W., Wu, Y., Zhou, Y., Peng, F., Zhang, S., Chen, P., & Xu, X. (2017). BcCFEM1, a CFEM Domain-Containing Protein with Putative GPI-Anchored Site, Is Involved in Pathogenicity, Conidial Production, and Stress Tolerance in *Botrytis cinerea*. *Frontiers in Microbiology*, 8, 1807. <https://doi.org/10.3389/fmicb.2017.01807>
- Zizumbo Villarreal, D., Hernández Roque, F., & Harries, H. C. (1993). Coconut varieties in Mexico. *Economic Botany*, 47(1), 65-78. <https://doi.org/10.1007/BF02862207>

ANEXO(S)