

Tesis doctoral

Búsqueda a escala genómica de marcadores moleculares de colonización
exofítica/endofítica en una bacteria ubicua asociada a plantas:
La asociación *Pantoea* spp. – alfalfa.

Lic. Sofia Agustina Erdozain Bagolin

Director: Dr. Antonio Lagares

Codirector: Dr. José Luis Lopez

2024



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Trabajo de Tesis Doctoral:

**Búsqueda a escala genómica de marcadores moleculares de
colonización exofítica/endofítica en una bacteria
ubicua asociada a plantas:
La asociación *Pantoea* spp. – alfalfa.**

Tesista: Lic. Sofia Agostina Erdozain Bagolin

Director: Dr. Antonio Lagares

Codirector: Dr. José Luis Lopez

Año:2024

El presente trabajo de Tesis para optar al grado de Doctor de la Facultad de Ciencias Exactas ha sido realizado en el Instituto de Biotecnología y Biología Molecular (IBBM) de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata – CCT – La Plata CONICET, bajo la dirección del Dr. Antonio Lagares y la codirección del Dr. José Luis Lopez.

A mis hermanas.

Mi reconocimiento

A la educación pública, gratuita y de calidad, sin la cual nada de esto sería posible.

A la Universidad Nacional de La Plata, al servicio de la ciencia y la patria.

Al Instituto de Biotecnología y Biología Molecular y al personal que lo conforma, por brindarme el espacio físico y equipamiento para desarrollar este trabajo de tesis doctoral

Al Consejo Nacional de Investigaciones Científicas y Técnicas y a la Agencia Nacional de Promoción Científica y Tecnológica, por el financiamiento que me permitió dedicarme a la investigación científica.

Al programa de cooperación entre Argentina y Alemania (ALE-ARG DAAD) por la beca otorgada para desarrollar parte de este trabajo en Philipps-Universität de Marburg, Alemania.

Mi más profundo agradecimiento

A Tony, por dirigir esta tesis. Por su guía, enseñanza y ayuda. Por confiar en mí trabajo y capacidad. Fue un gusto y orgullo trabajar con vos. Gracias por transmitir tu visión y entusiasmo por la ciencia.

A Jose, por su codirección, amistad e insistencia para que aprenda a programar y me gusten los Beatles. Al menos una de esas cosas quedó.

A la Dra. Anke Becker y todo su grupo de trabajo. Especialmente a los que hicieron mejor cada día de trabajo en Marburg: Fran, Bettina, Javier, Jutta y Andreas. Como bonus, gracias a Belti por el trabajo juntas y hacerme reír con tus ocurrencias.

A Carito, por ser la voz de la razón. Disfruté cada momento juntas, corriendo atrás tuyo para seguirte el paso y siempre con algún chiste a lo caritovacca™. Gracias por ayudarme a compaginar esta tesis.

A Abrula, por ser la voz de la locura. De vos aprendí más que cortar y peinar rulitos. Gracias por ser un espacio seguro para decir lo que se me viene a la cabeza, y por hacer mejor cada día de este recorrido. Se que aún nos queda mucho por delante juntas.

A Delfi, por las salidas, pijamas party, caminatas, charlas y quejas compartidas. Hiciste mejor cada día en Marburg, y me queda tu compañía y cariño marcado para siempre.

A Yuli y Florcita, por su guía académica y espiritual. Por las largas videollamadas, llenas de risas y chusmerío.

A Rodrigo, por crecer conmigo. Gracias por las largas llamadas, hablando sin parar o en silencio para acompañarnos.

A Emily, por abrazarme desde el día que nos conocimos y darme una cantidad de amor que iguala todo el cariño que tengo en Argentina. Gracias por compartir tu casa, tu cultura y tu vida conmigo. Ευχαριστώ

A Agos, por hacerme reír como nadie. Gracias por ser un cable a tierra y enseñarme la paciencia más difícil de todas, con una misma.

A todos mis amigos en Mainz (y en todas partes), que hicieron que Alemania siempre se sintiera un hogar: Bulat, Marcel, Mariane, Eric, Mert, Naima, Noor, Luqui, Santi, Juan, Emily y Agos. Danke.

A Abri y Juli, por reencontrarnos cada vez más juntas y con más cosas por compartir. Gracias por elegirnos como el primer día.

A los Erdozain, por hacer que todo esté atravesado por la ternura y diversión como nos enseñó la abuela.

A la familia heredera de María Brenda, por el amor y delirio que nos caracteriza.

A los Diez, por recibirme en su familia con tanto amor. Por todas las fiestas, risas y charlas compartidas.

A todo RPS chiqui, los que me adoptaron cuando empecé el doctorado, y los que me recibieron como si hubiésemos compartido años juntos. A todos los que están y a los que se fueron, gracias por enseñarme, y hacer que el laboratorio jamás se sienta como un trabajo: Nico, Eze, Hila, Lucas, Yuli, Caro, Abrula, Jose, Coti, Abru P, Rama, Agus, Kei, Charly, Martin, Delfi, Cata y Pili.

A todo RPS grande, Flor, Mariano, Gonza, Mauri y Walter. A Mauri gracias por el análisis de datos del TnSeq. A Walter gracias por hacerme reír y preguntarme por mis gatitos.

A todo el IBBM, por su calidad científica de la que aprendí año a año, y sobre todo por su inmensa calidad humana.

Finalmente,

A mis hermanas, Cami, Chia y Nare. Mis más grandes amigas, gracias por la complicidad y por hacerme sentir tan orgullosa de pertenecer a esta familia.

A Chia, por ser mí hogar. Estos cinco años (y los anteriores) no hubiesen sido posible sin esta mini familia que formamos en La Plata. Gracias por apoyarme siempre, y por tener una neurona compartida conmigo.

A mis abuelas, Elva y María, de las que aprendí la importancia de ser constante, nostálgica, graciosa y un poco irreverente. A las dos, un extracto de la agenda de la abu: "Me dejaste raíces y alas".

A Lore, registro de mí vida. Por escucharme hasta cuando no digo nada, con cariño elástico y atemporal. Deseo que, como hasta ahora, cada vez que mire atrás hayas estado ahí.

A mi mamá y mi papá, que me dieron todas las oportunidades para crecer y la sabiduría para hacer lo mejor con ellas. Gracias por enseñarme que la mejor ambición es una vida que disfrute.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS	15
ÍNDICE DE TABLAS	13
ABREVIATURAS	19
I. CAPÍTULO I	
I.1. Microorganismos asociados a plantas y su importancia en el contexto global actual	23
I.2. Microbioma de plantas	23
I.2.a Microbioma asociado a la rizosfera	24
I.2.b Microbioma asociado a la filósfera	26
I.2.c Microbioma endofítico.....	29
I.2.d Microbioma de semillas	33
I.2.e Medicago sativa y su microbioma asociado.....	35
I.3. Microorganismos multifacéticos: una mirada al género <i>Pantoea</i> y su asociación a plantas	38
I.4. Metodologías de inserción de transposones para el estudio de interacciones simbióticas y nichos de difícil acceso	42
I.4.a Técnicas de Secuenciación de Inserción de Transposones : Transposon - Sequencing (Tn-Seq) y Signature Tagged Mutagenesis (STM),.....	42
I.4.b Metodología TnSeq: Secuenciación de Transposones	45
I.5. Nuevas herramientas en el estudio de la ecología microbiana: Machine Learning para la predicción de fenotipos de interés y búsqueda <i>in silico</i> de características bacterianas responsables o asociadas	46
CONSTRUCCIÓN DE LA HIPÓTESIS DE TRABAJO	49
OBJETIVO GENERAL	50
OBJETIVOS ESPECÍFICOS	50
II. .CAPÍTULO II	
II.1. Técnicas microbiológicas.....	53
II.1.a Cepas bacterianas usadas en este trabajo	53
II.1.b Plásmidos usados en este trabajo	53
II.1.c Medios de cultivo y condiciones de crecimiento	53
II.1.d Medios de conservación.....	54
II.1.e Antibióticos	54
II.2. Procedimientos microbiológicos.....	55

II.2.a	Curvas de crecimiento	55
II.2.b	Recuento de bacterias en medio líquido	55
II.2.c	Conjugaciones biparentales	55
II.2.d	Preparación de cepas electrocompetentes. Electrotransofrmación de E.coli y <i>Pantoea</i> sp.	56
II.3.	Técnicas de biología molecular	56
II.3.a	Preparación de ADN plasmídico en pequeña escala (miniprep).....	56
II.3.b	Extracción de ADN total	56
II.3.c	Electroforesis de ADN	56
II.3.d	Tratamientos enzimáticos del ADN.....	57
II.3.e	Reacción en cadena de la polimerasa (PCR) y reacción en cadena de la polimerasa en tiempo real (RT-PCR)	57
II.3.f	Preparación de adaptadores doble cadena	60
II.3.g	Secuenciamiento genómico de cepas bacterianas	60
II.4.	Ensayos con plantas	61
II.4.a	Esterilización de semillas y germinación	61
II.4.b	Condiciones de cultivo	61
II.5.	Ensayos TnSeq (<i>transposon insertion sequencing</i>)	63
II.5.a	Desafío de la biblioteca de mutantes en condiciones de colonización de rizósfera, endosfera de plantas de alfalfa, medio rico LB y medio mínimo M9	63
II.5.b	Amplificación de las poblaciones de mutantes recuperadas de rizósfera y endosfera por crecimiento en medio rico.	64
II.5.c	Extracción de ADN genómico de las muestras y su procesamiento para la posterior identificación y cuantificación de los sitios de inserción del Tn mariner	64
II.5.d	Mezcla de los productos de amplificación etiquetados.....	65
II.5.e	Secuenciamiento de los sitios de inserción del Tn mariner y procesamiento de datos	65
II.6.	Algoritmos de <i>Machine learning</i> para la predicción del tropismo a planta.....	66
II.6.a	Construcción y curado del set de datos utilizado para el entrenamiento de modelos de <i>Machine Learning</i>	66
II.6.b	Análisis filogenético.....	66
II.6.c	Algoritmos de <i>Machine Learning</i> utilizados.....	66
II.7.	Bases de datos utilizadas.....	67
III.	CAPÍTULO III	
III.1.	Introducción	71

III.2. <i>Pantoea agglomerans</i> LPU12, la bacteria endófitas de alfalfa que estudiaremos en esta tesis.....	72
III.2.a <i>Pantoea sp.</i> LPU12 resistente a ácido nalidíxico	74
III.2.b Secuenciación genómica y posicionamiento taxonómico de <i>Pantoea sp.</i> LPU12 como <i>Pantoea agglomerans</i> LPU12.....	74
III.3. Caracterización del sistema modelo propuesto en este trabajo, la asociación <i>Medicago sativa</i> – <i>Pantoea agglomerans</i> LPU12: Grado de colonización y progreso temporal del abordaje rizosférico y endofítico.....	77
III.3.a Colonización rizosférica. Condiciones experimentales.....	77
III.3.b Colonización endofítica. Condiciones experimentales. Consideraciones numéricas.....	79
III.3.c Ingreso a la endosfera radicular. Relación entre la composición endofítica y la composición del inóculo. Estimación del número mínimo de ingresantes iniciales (“fundadores”).....	82
III.4. Consideraciones finales del capítulo.....	84
 IV. CAPÍTULO IV	
IV.1. Introducción	87
IV.2. Construcción del vector pSAM_placO ^c funcional en <i>P. agglomerans</i>	87
IV.3. Caracterización de la eficiencia transposicional en <i>P. agglomerans</i> del Tn <i>mariner</i> asociado al plásmido pSAM_placO ^c . Eficiencia transposicional en alfaproteobacterias.....	89
IV.4. Construcción de una biblioteca de mutantes transposicionales en <i>P. agglomerans</i> LPU12.....	90
IV.4.a Caracterización de la biblioteca de mutantes de <i>P. agglomerans</i> LPU12 mediante secuenciación masiva.....	91
IV.5. Consideraciones finales del capítulo.....	93
 V. CAPÍTULO V	
V.1. Introducción	97
V.2. Diseño de un ensayo TnSeq para la búsqueda de marcadores genéticos de <i>P. agglomerans</i> asociados a la colonización exo- y endofítica de plantas de alfalfa: Procedimiento general (adaptador, cebadores, etiquetado) para la identificación y contaje de mutantes.....	98
V.2.a Estructura general del ensayo. Muestras, etapas, diseño de adaptador y reacciones de PCR. Muestra para secuenciación.....	98
V.2.b Optimización de las condiciones experimentales del ensayo TnSeq. Consideraciones prácticas experimentales.....	100
V.3. Implementación del ensayo TnSeq con la biblioteca de mutantes <i>P. agglomerans</i> LPU12::Tn <i>mariner</i> : Tratamientos, resultados de las amplificaciones, confirmación de secuencias en amplicones testigo. Muestra para secuenciación de alta capacidad.....	103

V.3.a	Tratamientos y réplicas. Muestra de entrada y muestras de salida	103
V.3.b	Colección de las muestras provenientes de cultivos en medio mínimo (M9) y rico (LB) (búsqueda de grupos génicos B y C).	104
V.4.	Consideraciones finales del capítulo	110
VI. CAPÍTULO VI		
VI.1.	Introducción	113
VI.2.	Búsqueda de genes esenciales, genes de auxotrofia y genes de pérdida de parcial de aptitud (<i>fitness</i>) para crecer en medio rico.	113
VI.2.a	Genes esenciales del genoma de <i>P. agglomerans</i> LPU12.	113
VI.2.b	Genes de auxotrofia y genes cuya mutación resulta en efectos menos severos (pero aún detectables) para crecer en medio mínimo M9 respecto de medio LB.....	115
	Genes cuya mutación resulta en un mejor crecimiento en medio M9 respecto de medio LB.....	115
VI.2.c	Genes asociados a la pérdida de aptitud (<i>fitness</i>) de <i>P. agglomerans</i> LPU12 para crecer en medio rico LB.....	117
VI.4.	Búsqueda a escala genómica de marcadores genéticos asociados a la colonización de raíces de alfalfa por <i>P. agglomerans</i> LPU12.....	118
VI.4.a	En busca de los genes vinculados a la colonización de la rizósfera de alfalfa por <i>P. agglomerans</i> LPU12.	118
VI.4.b	El ingreso a la planta de <i>P. agglomerans</i> LPU12: en busca de genes asociados a la colonización endofítica de alfalfa.....	122
VI.5.	Consideraciones finales del capítulo	129
VII. CAPÍTULO VII		
VII.1.	Introducción	135
VII.1.a	Algoritmos de aprendizaje automático - Machine Learning – para la clasificación: Random forest y Logistic Regression	136
VII.2.	Construcción y curado de la base de datos utilizada para el entrenamiento de los algoritmos aprendizaje automático que permiten la clasificación categórica.....	139
VII.3.	Modelos de clasificación generados para la predicción de tropismo a plantas en el género <i>Pantoea</i> , basados en la información funcional de los genomas.....	141
VII.3.a	Reducción de variables: búsqueda de marcadores con alto grado de relación entre ellos y reducción de las matrices binarias utilizadas para entrenar los modelos	144
VII.3.b	Identificación de variables con importancia para la identificación de bacterias: búsqueda in silico de características funcionales en en los genomas relacionados a la asociación a plantas en bacterias del género <i>Pantoea</i>	147
VII.4.	Consideraciones finales del capítulo	149
VIII. CAPÍTULO VIII		

Discusión general y conclusiones.....	153
ANEXO	
VIII.1. Figuras suplementarias	163
VIII.2. Tablas suplementarias.....	164
BIBLIOGRAFÍA	
BIBLIOGRAFÍA.....	217

ÍNDICE DE FIGURAS

Figura I.1	Endófitos pasajeros, oportunistas y competentes en raíces de plantas.	30
Figura I.2	Mecanismos de promoción del crecimiento vegetal, colonización y factores que afectan a la diversidad de bacterias endofíticas en la planta huésped.	32
Figura I.3	Desarrollo vegetal de semilla a semilla y oportunidades de adquisición de microbios por las semillas en distintos estadios de este ciclo.	35
Figura I.4	Diversidad bacteriana en clones endofíticos derivados de semillas de alfalfa esterilizadas en superficie	37
Figura I.5	Diversidad y potencial de las cepas del género <i>Pantoea</i>	39
Figura I.6	Metodologías de Secuenciación de Inserción de Transposones (TIS).	44
Figura III.1	Esquema circular de los cuatro replicones de la cepa <i>Pantoea</i> LPU12.	75
Figura III.2	Árbol filogenético generado por el servidor TYGS que incluye el aislamiento LPU12 cuya capacidad de colonización de alfalfa estudiamos en esta tesis.	76
Figura III.3	Evolución temporal de la colonización rizosférica de alfalfa por parte de la cepa indicadora <i>P. agglomerans</i> LPU12 NaIR.	79
Figura III.4	Evolución temporal de la colonización endofítica de alfalfa por parte de la cepa indicadora <i>P. agglomerans</i> LPU12 NaIR durante los primeros 22 días postinoculación.	80
Figura III.5	Evolución temporal de la colonización endofítica de alfalfa por parte de la cepa indicadora <i>P. agglomerans</i> LPU12 NaIR durante las primeras 24 h postinoculación.	81
Figura III.6	Colonización endofítica de plantas de alfalfa inoculadas con diferentes proporciones de dos variantes fluorescentes de la cepa <i>P. agglomerans</i> LPU12.	83
Figura IV.1	Estrategia utilizada para la construcción del vector pSAM_placO ^c .	88
Figura IV.2	Distribución y número de relativo de inserciones del Tn <i>mariner</i> en cada uno de los replicones de la cepa <i>P. agglomerans</i> LPU12 en la biblioteca transposicional que hemos generado para los ensayos TnSeq.	92
Figura V.1	Esquema del ensayo TnSeq diseñado para la búsqueda de genes de <i>P. agglomerans</i> LPU12 vinculados a la colonización radicular de alfalfa.	98
Figure V.2	Esquema de la estrategia de corte de los extremos del Tn <i>mariner</i> , ligación de adaptador, y amplificación de las regiones genómicas flanqueantes.	99
Figura V.3	Resultado de las reacciones de qPCR empleadas en el experimento TnSeq para la amplificación y cuantificación de cada una de las regiones de unión (inserciones) Tn <i>mariner</i> - genoma de <i>P. agglomerans</i> LPU12	102
Figura V.4	Esquema de tratamientos del experimento TnSeq para la caracterización de la genética de la colonización rizosférica y endofítica de alfalfa por parte de la cepa <i>P. agglomerans</i> LPU12.	104
Figura V.5	Número de bacterias en las aguas de lavado y en el interior de las plantas luego del macerado	106
Figura V.6	Electroforesis de alta resolución de la mezcla de los productos de amplificación de las qPCR.	108

Figura V.7	Electroforesis en gel de agarosa de la mezcla de productos de amplificación de las qPCR. Secuencias de amplicones clonados	109
Figura VI.1	Distribución de funciones en el grupo de genes identificados como posiblemente esenciales en <i>P. agglomerans</i> LPU12.	115
Figura VI.2	Distribución de funciones en el grupo de genes de <i>P. agglomerans</i> LPU12 asociados a fenotipos de auxótrofia o de deficiencia de crecimiento en medio mínimo M9 respecto del crecimiento en medio rico LB	116
Figura VI.3	Gráfico tipo volcán en el que se representan los valores M y sus correspondientes valores p para los distintos genes mutados de la biblioteca al comparar el inóculo (Muestra de entrada_Inóculo_E) y la muestra de salida de la rizósfera a 3 d.p.i. (Muestra de salida_Rizósfera).	119
Figura VI.4	Localización genómica y distribución de funciones en el grupo de genes de <i>P. agglomerans</i> LPU12 que modifican la capacidad de colonizar la rizósfera de alfalfa. Distribución de funciones en el grupo de genes de <i>P. agglomerans</i> LPU12 asociados a la colonización rizosférica.	120
Figura VI.5	Densidad de sitios TA mutados en cada una de las muestras analizadas: inóculo, rizósfera, endósfera temprana (24 h) y endósfera tardía (20 d.p.i.).	123
Figura VI.6	Representación esquemática de un cuello botella y efecto fundador operando durante el acceso de los distintos genotipos al comportamiento endofítico de una raíz	125
Figura VI.7	Gráfico tipo volcán en el que se representan los valores M y los correspondientes valores p para genes seleccionados de la biblioteca al comparar el inóculo con las muestras de salida de la endósfera de alfalfa.	127
Figura VI.8	Localización genómica y distribución de funciones en el grupo de genes de <i>P. agglomerans</i> LPU12 que modifican la capacidad de colonizar la endósfera de alfalfa. Distribución de funciones en el grupo de genes de <i>P. agglomerans</i> LPU12 asociados a la colonización endofítica.	128
Figura VII.1	Entrenamiento de un árbol individual de un modelo de <i>Random Forest</i>	137
Figura VII.2	Árbol filogenético generado con la herramienta GTDB-Tk para el análisis de la distribución filogenética de 418 cepas del género <i>Pantoea</i> utilizadas en este estudio	140
Figura VII.3	Evaluación del desempeño de los modelos de <i>Machine Learning</i> generados para la clasificación del tropismo a planta en aislamientos del género <i>Pantoea</i> .	143
Figura VII.4	Variación del valor f_1 de cada modelo en función de la dimensión del set de datos con el que fueron entrenados cuando se reduce la correlación entre las variables.	146
Figura VII.5	Coeficientes de importancia asignados a las variables más relevantes para la predicción del tropismo a plantas por modelos generados con el algoritmo <i>Logistic Regression</i> .	149
Figura VII.S1	Variación del valor F_1 en modelos de <i>Random Forest</i> con diferentes números de árboles de decisión	163
Figura VII.S2	Distribución del Coeficiente de Importancia de asignado por los modelos COG-LR_0.4, KO-LR_0.4 y PFAM-LR_0.4	163

ÍNDICE DE TABLAS

Tabla I.1.	Factores bióticos y abióticos que afectan la rizodeposición.	25
Tabla I.2.	Resumen de los endófitos bacterianos de semillas conocidos aislados de diferentes especies vegetales y el orden taxonómico al que pertenecen.	34
Tabla II.1	Cepas bacterianas usadas en este trabajo y sus características.	53
Tabla II.2	Plásmidos usados en este trabajo y sus características.	53
Tabla II.3	Medios de cultivos.	54
Tabla II.4	Antibióticos utilizados en este trabajo (concentraciones finales empleadas en los medios en los que crecieron los cultivos).	55
Tabla II.5	Secuencia de oligonucleótidos utilizados en este trabajo.	59
Tabla II.6	Secuencia de oligonucleótidos firma utilizada como identificador para etiquetar cada condición del ensayo TnSeq.	60
Tabla II.7	Composición del medio mineral de irrigación de plantas.	62
Tabla II.8.	Bases de datos, herramientas bioinformáticas y estadísticas utilizadas en este trabajo.	67
Tabla III.1	Características fenotípicas del aislamiento Pan12 usado en este trabajo de tesis respecto de las del resto de los aislamientos del mismo género recuperados de semillas y plantas de alfalfa.	73
Tabla IV.1.	Frecuencia de transposición del Tn mariner en <i>P. agglomerans</i> LPU12 y en <i>S. meliloti</i> 2011 al emplear el vector pSAM_placOc o el pSAMRI.	90
Tabla IV.2.	Caracterización de la biblioteca de mutantes de <i>P. agglomerans</i> LPU12::Tn mariner.	93
Tabla VI.S1.	Genes esenciales de <i>P. agglomerans</i> LPU12 identificados en los ensayos TnSeq (Capítulo VI).	164
Tabla VI.S2.	Genes de auxotrofia y relacionados en <i>P. agglomerans</i> LPU12 identificados en los ensayos TnSeq.	183
Tabla VI.S3.	Genes cuya mutación en <i>P. agglomerans</i> LPU12 favorece el crecimiento en medio mínimo respecto del medio rico, identificado en los ensayos TnSeq.	189
Tabla VI.S4.	Genes cuya mutación en <i>P. agglomerans</i> LPU12 causa una pérdida parcial en la habilidad para crecer en medio rico LB identificados en los ensayos TnSeq.	194
Tabla VI.S5.	Genes relacionados a la colonización rizosférica de alfalfa por <i>P. agglomerans</i> LPU12 (excluidos los genes de auxotrofia) identificados en los ensayos TnSeq.	195
Tabla VI.S6.	Genes relacionados con la colonización endofítica de alfalfa por <i>P. agglomerans</i> LPU12 (excluidos los genes de auxotrofia) identificados en los ensayos TnSeq.	197
Tabla VIII.S7.	Genomas y MAGs (genomas ensamblados a partir de metagenomas) del género <i>Pantoea</i> utilizados para el entrenamiento y prueba de modelos de <i>Machine Learning</i> diseñados para clasificar aislamientos según su capacidad de asociación con plantas.	200

ABREVIATURAS

°C Grados Celsius	nM Nanomolar.
ADN Ácido Desoxirribonucleico	P/P peso / peso
ARN Ácido Ribonucleico	P/V peso / volumen
cm Centimetro.	pb Pares de Bases
dNTPs Desoxirribonucleótidos	PCR Reacción en Cadena de la Polimerasa, "Polymerase Chain Reaction"
DO Densidad Óptica	pH Potencial Hidrógeno
DO600: Densidad Óptica a 600 nm.	rpm Revoluciones por minuto
d.p.i. Días Post Inoculación	SDS Dodecil sulfato de sodio
g Gramo	Sm Estreptomicina
Gb Gigabase= 1x10 ⁹ pb	STM Signature tagged Mutagenesis
hs Horas	TA Tiamina-Adenina
Kb Kilobase= 1x10 ³ pb	Ta Temperatura de annealing o hibridación
Km Kanamicina	Tb Terabase=1x10 ¹² pb
l Litro	Tc Tetraciclina
m Metro	Tm Temperatura del melting o fusión
M Molar	Tn Transposón
Mb Megabase= 1x10 ⁶ pb	TnSeq Transposon Sequencing
mg Miligramo	TIS Transposon Insertion Sequencing
min Minuto	u.f.c. Unidades Formadoras de Colonias.
ml Mililitro	µg Microgramo
mm Milímetro	µl Microlitro
mM Milimolar	µM Micromolar
Nal Acido Nalidixico	V/V volumen /volumen
Nm Neomicina	



CAPÍTULO I

INTRODUCCIÓN



I.1. Microorganismos asociados a plantas y su importancia en el contexto global actual

Las plantas son hospedadoras de una gran cantidad y diversidad de microorganismos con los que establecen estrechas relaciones a través de un constante intercambio de información que influye en el desarrollo de la planta, de los microorganismos, y del sistema planta-microorganismos como un conjunto. Si bien la comunidad microbiana que habita los tejidos vegetales es muy diversa, existe una estructura taxonómica inherente de cada planta como se ha evidenciado en múltiples trabajos que describen el microbioma de diversas organismos modelo como *Arabidopsis thaliana*, *Medicago truncatula*, *Lotus japonicus* y *Marchantia polymorpha* (Alcaraz et al., 2018; Lundberg et al., 2012; Tkacz et al., 2015; Zgadza et al., 2016), y de cultivos como maíz, arroz, soja y cebada (Bulgarelli et al., 2015; Edwards et al., 2015; Peiffer et al., 2013; Sugiyama et al., 2014). Estudios han evidenciado la dificultad de cultivar injertos de plantas en ausencia de microorganismos (Hardoim et al., 2008), lo que indica la importancia de estos para el desarrollo de la planta huésped, y conduce a la necesidad de considerar a la planta ya no como un organismo aislado, sino como parte integral de un sistema que incluye a todos los microorganismos que coexisten con ella, concepto conocido como holobionte (Zilber-Rosenberg & Rosenberg, 2008). De este concepto se desprende también el holobioma, que comprende el genoma del huésped y el genoma del microbioma asociado, y puede considerarse como "el reflejo genómico de la red compleja de interacciones simbióticas que vinculan a un individuo de un taxón dado con su microbioma asociado" (Guerrero et al., 2013). Esto implica entender al holobionte como la unidad sujeta a procesos evolutivos que conducen a la selección y adaptación de los individuos que lo componen, así como de la información genética que cada uno aporta. En la actualidad, numerosos estudios buscan ampliar el conocimiento y la comprensión de la estructura del microbioma vegetal, los diversos factores que lo modelan y los procesos moleculares que subyacen a las interacciones microorganismos-planta, así como a las interacciones microorganismo-microorganismo en el contexto del microbioma. Se espera que estos estudios conduzcan a la identificación de microorganismos, procesos biológicos y subsistemas dentro del microbioma que impacten de manera positiva en la salud y el crecimiento de la planta. Además, se busca explorar su potencial como bioinsumos para una industria agrícola sustentable que se enfrenta a demandas crecientes de producción, en un contexto climático global cada vez más comprometido

I.2. Microbioma de plantas

Aunque la diversidad de microorganismos que habitan las plantas es considerable, se encuentra limitada en comparación con la diversidad presente en el suelo circundante, que constituye una de las principales fuentes de biodiversidad microbiana para la formación del microbioma vegetal. Dado que las plantas reclutan microorganismos del suelo, este se considera un factor determinante clave en la composición de la comunidad microbiana de la planta, especialmente en la región de las raíces. Otro factor relevante que afecta la estructura del microbioma es la filogenia de la planta, como lo evidencian diversos estudios que señalan variaciones en la composición de microorganismos asociados a plantas filogenéticamente distantes (Bulgarelli et al., 2013). La especificidad en la diversidad y composición microbiana propia de cada planta refleja relaciones concretas y específicas entre los huéspedes y las plantas, así como eventos de selección de unos microorganismos sobre otros. Estas relaciones dentro de una comunidad microbiana particular son complejas, con miembros que pueden vincularse de manera transitoria y otros que se presentan de forma persistente y ubicua en la población de plantas estudiada. A este último conjunto de microorganismos se le conoce como "microbioma núcleo", cuya definición y

constitución dependen de la escala del sistema que se analiza (por ejemplo, holobionte, población de plantas, distintas poblaciones de una especie vegetal). Por lo tanto, es esperable que la complejidad del sistema disminuya en jerarquías más altas (Vandenkoornhuysen et al., 2015). El estudio del microbioma núcleo ofrece un escenario para analizar la similitud del microbioma entre diferentes plantas en diversos niveles taxonómicos, así como el papel de miembros específicos y los efectos que puedan tener en la estructura del microbioma y su impacto en el fitness vegetal. Esto conduce a la identificación de "microorganismos fundacionales" esenciales para el sistema (Agler et al., 2016; Astudillo-García et al., 2017; Yeoh et al., 2017).

Como se describió anteriormente, la diversidad y complejidad del microbioma varían según el sistema que se analiza, y una planta individual no es una excepción. Diferentes compartimentos de la planta exhiben composiciones microbianas diversas y una clara reducción en la complejidad de la comunidad, con unos pocos colonizadores exitosos en nichos específicos (Trivedi et al., 2020). Nuevamente, este escenario plantea interrogantes sobre las fuentes de microorganismos para distintos nichos de la planta, las fuerzas que rigen el ensamblaje de estos sub-microbiomas y su impacto en el crecimiento y salud de la planta. Existe una amplia diversidad de nichos vegetales que los microorganismos pueden habitar, y en las siguientes secciones profundizaremos en los más relevantes: la rizosfera, filósfera, endosfera y las semillas.

I.2.a Microbioma asociado a la rizosfera

La rizosfera es la región próxima a las raíces que sufre una fuerte influencia por parte de la planta debido a la presencia de compuestos orgánicos liberados por la raíz de la planta que hacen de la zona un ambiente rico en nutrientes en comparación al suelo circundante, donde proliferan bacterias, hongos (incluidas micorrizas arbusculares), oomicetos, virus y arqueas (Philippot et al., 2013). Se conoce como "*efecto rizosférico*" al fenómeno que permite el establecimiento de una microbiota enriquecida debido a la influencia de la raíz. Las plantas liberan compuestos orgánicos denominados rizodepósitos que dan lugar a este efecto mediante distintos mecanismos como secreción, exudación o deposición. Los rizodepósitos están constituidos por compuestos de bajo peso molecular como azúcares, aminoácidos, ácidos orgánicos, hormonas vegetales, vitaminas, compuestos aromáticos y otros metabolitos secundarios, y compuestos de alto peso molecular como el mucílago (azúcares polimerizados), enzimas, proteínas, células vegetales de la cofia y la capa más externa de la raíz, y otros polisacáridos (Bais et al., 2006; Jones et al., 2009; Nguyen, 2009). Esta riqueza nutricional de la rizosfera conduce a una comunidad microbiana compleja cualitativa y cuantitativamente, con microorganismos saprofitos, epifitos, endófitos, patógenos y microorganismos beneficiosos (Avis et al., 2008), y el perfil de rizodepósitos en la rizosfera es un factor determinante en la diversidad taxonómica de esta comunidad. Diversos factores bióticos y abióticos propios de la planta y el suelo influyen en el proceso de rizodeposición, como se muestra en la Tabla I.1 descrita por Prashar et al. 2014. Como consecuencia, es de esperar que la microbiota asociada a la rizósfera varíe entre distintos genotipos vegetales, estadios de la planta, diferentes suelos y condiciones. Por otro lado, la versatilidad metabólica o la diversidad funcional de las poblaciones microbianas establecidas en la rizosfera se rige por la variedad de factores genéticos que portan y las interacciones con otros participantes, incluida la propia planta (Barriuso et al., 2008).

Tabla I.1. Extraída y adaptada de Prashar et al. 2014. Factores bióticos y abióticos que afectan la rizodeposición

Suelo		Planta	
<i>Biótico</i>	<i>Abiótico</i>	<i>Biótico</i>	<i>Abiótico</i>
Estructura de la comunidad microbiana	Tipo de suelo	Especie vegetal	Temperatura
Tipo de comunidad microbiana	Textura del suelo	Fotosíntesis	Agua en superficie
Actividad de la comunidad microbiana	pH	Fase de desarrollo	Humedad
Producción de fitohormonas	Impedancia	Edad de la raíz	CO2 elevado
Producción de toxinas	Salinidad	Arquitectura de la raíz	Intensidad de la luz
<i>Quorum sensing</i>	Disponibilidad de agua	Deficiencia nutricional	Pesticidas
Agentes de biocontrol	Materia orgánica	Nodulación	Irrigación
Patógenos	Potencial redox	Permeabilidad de la membrana	Ozono
Liberación de moléculas de señalización radicular	Contenido de iones metálicos	Liberación de señales microbianas	Velocidad del viento
Micorrizas	Compactación	Concentración citosólica	Fuego
	Profundidad de las raíces	Liberación de aleloquímicos	Superficie disponible
	Densidad		Latitud, Altitud
	Drenaje y aireación		Erosión

Como se ha descrito, diversos factores modelan el microbioma vegetal, resultando en comunidades específicas en cada sistema planta-microbioma. Sin embargo, muchos estudios revelan distribuciones notablemente similares de los filos microbiano, donde las diferencias entre cultivares de plantas se hacen evidentes al comparar especies y cepas microbianas. Las Proteobacterias suelen dominar las muestras, en particular las de las clases α y β . Otros grupos importantes son *Actinobacteria*, *Firmicutes*, *Bacteroidetes*, *Planctomycetes*, *Verrucomicrobia* y *Acidobacteria* (Turner et al., 2013). La diversidad compartida entre plantas tan diversas señala al suelo como la principal fuente de microorganismos, y las similitudes y diferencias pueden entenderse desde un enfoque donde la formación de la microbiota de la rizosfera es el resultado de una secuencia de eventos jerárquicos, donde el suelo actúa como la fuente inicial de microorganismos. Las propiedades fisicoquímicas del suelo, junto con los procesos biogeográficos, influyen en la estructuración de esta comunidad microbiana. En un nivel subsecuente, la ubicación donde las plantas crecen determina la biota indígena a la que las raíces están expuestas. Finalmente, la especie y el genotipo de la planta influyen en qué miembros de este reservorio pueden establecerse y prosperar en la rizósfera (Philippot et al., 2013).

Como resultado de la colonización microbiana de la rizósfera surgen relaciones asociativas, simbióticas, neutrales o parasitarias, dependiendo de factores tales como el estado nutricional y el entorno del suelo, los mecanismos de defensa de la planta, y el propio microorganismo colonizante (Parmar & Dufresne, 2011). Así el microbioma puede afectar a la planta de distintas maneras, desde el crecimiento y nutrición, hasta la susceptibilidad a enfermedades y el desarrollo de fitopatógenos. Dentro del conjunto de bacterias presentes en la rizósfera, las rizobacterias son un subconjunto que tiene la capacidad de, tras su reinsertión en la semilla o partes vegetativas de la planta, colonizar el sistema radicular en presencia de la microbiota del suelo competidora. Aquellas rizobacterias que afectan de forma negativa a la planta

se las conoce como “depletorias”, mientras que las que influyen de forma positiva se las denomina Rizobacterias “promotoras del crecimiento vegetal” (PGPR, por sus siglas en inglés *Plant Growth Promoting Rhizobacteria*) (Prashar et al., 2014).

Las PGPR promueven el crecimiento y salud de la planta mediante varios mecanismos directos e indirectos. La promoción directa del crecimiento vegetal se basa en el suministro a la planta de compuestos bacterianos favorables, o en la mejora de la absorción de nutrientes del suelo por parte de la planta mediante procesos tales como la fijación de nitrógeno atmosférico, la producción de sideróforos, la solubilización de fosfato y la síntesis de fitohormonas, entre otros (Glick, 1995). Por otro lado, la promoción indirecta del crecimiento vegetal se basa principalmente en la reducción o prevención de los efectos deletéreos de fitopatógenos, como hongos y nematodos, controlando así el desarrollo de enfermedades. Diversos mecanismos llevan a la supresión de patógenos, como la producción de cianuro, antibióticos o enzimas líticas extracelulares como quitinasas, proteasas y celulasas, la competencia por nutrientes y nichos en la rizósfera, el parasitismo y la depredación (Prashar et al., 2014). Finalmente, las PGPR pueden mejorar la tolerancia de la planta a diversos estreses medioambientales mediante la producción de fitohormonas y de la enzima ACC desaminasa que actúa degradando el ácido 1-aminociclopropano-1-carboxílico (ACC), precursor del etileno, impidiendo la acumulación de este en diversas condiciones de estrés e incrementando la tolerancia de la planta. Dependiendo de los mecanismos que subyacen las contribuciones al crecimiento y sanidad de la planta, las PGPR pueden catalogarse como bioprotectores, biofertilizantes y bioestimulantes (Martínez-Viveros et al., 2010). La efectividad de los procesos mediados por PGPR dependen fuertemente de factores como la competencia y persistencia de la bacteria en la rizósfera, su capacidad de colonizar la raíz, la síntesis de metabolitos, la especie y genotipo vegetal, y la microbiota competidora presente en la rizósfera (Nowak, 1998).

I.2.b Microbioma asociado a la filósfera

La filósfera comprende la parte aérea de la planta, dominada por las hojas, y se encuentra influenciada por diversos factores que crean de este nicho un ambiente distintivo a los otros provistos por la planta. A macro-escala, factores como las condiciones climáticas, la ubicación de las hojas y la composición de la cutícula cerosa hidrofóbica que recubre las células epidérmicas reduciendo la evaporación del agua afectan el microambiente de la filósfera, dando lugar a un entorno oligotrófico; a microescala, las nervaduras, estomas y apéndices superficiales como tricomas e hidatodos, alteran la disponibilidad de nutrientes. Por otro lado, los microorganismos epifitos que habitan la filósfera están expuestos a la atmósfera y sometidos al ciclo diurno, recibiendo la influencia directa de la luz solar e indirecta del metabolismo vegetal. Así, deben hacerle frente a la radiación ultravioleta y a la baja o fluctuante disponibilidad de agua, lo que resulta en una distribución microbiana desigual en la superficie de las hojas (Remus-Emsermann et al., 2012), donde las comunidades que se establecen no representan conjuntos aleatorios de microorganismos, sino que hay una selección que da a lugar, al menos parcialmente, a poblaciones microbianas predecibles con unos pocos filos dominantes y otros subgrupos (Delmotte et al., 2009; Whipps et al., 2008; Yang et al., 2001).

Si bien la riqueza de especies microbianas presentes en la filósfera es alta, esta compleja diversidad es reducida cuando se la compara con la comunidad de microorganismos en la rizosfera o en el suelo (Delmotte et al., 2009; Knief et al., 2012). Estudios realizados con técnicas independientes del cultivo han permitido caracterizar el microbioma de una gran variedad de cultivos agrícolas como trigo, arroz, manzana, lechuga y espinaca, y de plantas y árboles de crecimiento natural, evidenciando cuatro filos dominantes: *actinobacterias*, *bacteroidetes*, *firmicutes* y *proteobacterias* (Bulgarelli et al., 2013). Un análisis más detallado sugiere a los géneros *Pseudomonas*, *Sphingomonas*, *Methylobacterium*, *Bacillus*,

Massilia, *Arthrobacter* y *Pantoea* como parte de la microbiota de la filósfera en una amplia gama de especies vegetales (Rastogi et al., 2013).

Diversos estudios han demostrado que, en la filósfera, factores ambientales como la radiación, la contaminación y la fertilización con nitrógeno, así como factores bióticos como la edad de las hojas y la presencia de otros microorganismos, afectan la comunidad bacteriana presente. Por otro lado, investigaciones adicionales han evidenciado que factores inherentes a la planta determinan también la estructura de la comunidad bacteriana. Los posibles mecanismos subyacentes a estos efectos genotípicos podrían originarse en las propiedades de la superficie foliar, la señalización hormonal que implica al ácido jasmónico o la señalización a través de la vía del GABA (ácido γ -aminobutírico (Vorholt, 2012).

El origen de los microorganismos en la filósfera es una interrogante que numerosos estudios se esfuerzan por dilucidar. La observación de que las poblaciones bacterianas se restablecen en plantas anuales siguiendo patrones similares a los observados en años anteriores sugiere la existencia de reservorios locales de bacterias. Aunque una fracción de la microbiota de la filósfera consiste en colonizadores provenientes del aire, para las comunidades bacterianas se ha sugerido que el reclutamiento temprano de una comunidad “núcleo” a partir de las semillas, el suelo u otras plantas podría explicar la consistencia de los miembros dominantes de la comunidad en plantas anuales. Se necesitan análisis específicos para esclarecer las contribuciones relativas de la transmisión vertical a través de las semillas y la transmisión horizontal a través del suelo, el aire y/u otras plantas (Knief et al., 2010).

Una vez que los microorganismos alcanzan la filósfera, la adaptación a este nicho se vuelve esencial para su reproducción exitosa, lo que implica la capacidad de soportar estreses tanto abióticos como bióticos, como la escasez de nutrientes y la presencia de metabolitos secundarios antimicrobianos de origen vegetal o microbiano. Para lograrlo, los microorganismos exhiben una variedad de características como las que se describen a continuación:

Resistencia a estrés abiótico: La exposición al oxígeno y a la radiación solar en la filósfera puede resultar en la producción de especies reactivas de oxígeno (ROS), las cuales pueden causar daños en ácidos nucleicos, proteínas y lípidos de los microorganismos. Para protegerse de estos daños, los microorganismos despliegan una serie de mecanismos, tales como la producción de pigmentos, la activación de sistemas de reparación del ADN, y procesos enzimáticos que desintoxican las ROS mediante catalasas y superóxido dismutasas. Además, las condiciones de la filósfera también pueden provocar desecación. Para hacer frente a este desafío, las bacterias pueden formar agregados mediante la producción de exopolisacáridos (EPS) (Chang et al., 2007; Rigano et al., 2007), secretar compuestos bioactivos que ayudan a mitigar esta condición (Monier & Lindow, 2003), o segregar biosurfactantes que aumentan la humectabilidad (Schreiber et al., 2005).

Compuestos antimicrobianos, resistencia y producción: tanto las plantas como los propios microorganismos generan una amplia variedad de metabolitos secundarios con actividad antimicrobiana, que contribuyen a la estructuración de la comunidad microbiana en la planta (Giddens et al., 2003).

Movilidad y adhesión: si bien la movilidad bacteriana permite el movimiento activo hacia regiones favorables de la superficie foliar, facilitando la propagación y la localización en nichos preferenciales (Haefele & Lindow, 1987), no se ha descrito la producción de flagelina como una característica asociada a miembros comensales de la filósfera. Esto plantea interrogantes sobre una posible supresión de la expresión de la flagelina en bacterias presentes en la filósfera, lo que les permitiría evadir los sistemas de defensa de la planta que reconocen el péptido Flg22 en la flagelina, descrito como un patrón molecular

asociado a microorganismos (MAMP, por sus siglas en inglés “*microorganism-associated molecular pattern*”) (Boller & Felix, 2009). De este modo, la adhesión y replicación local podrían representar el estilo de vida predominante de estas bacterias, favoreciendo la resistencia a la eliminación por lluvias, y la producción de sustancias poliméricas extracelulares que proporcionan protección frente a la desecación.

Adaptación metabólica: a diferencia de la rizósfera, la filósfera presenta una disponibilidad reducida y heterogénea de nutrientes. Las fuentes de carbono identificadas en este nicho incluyen carbohidratos, aminoácidos, ácidos orgánicos y glicoles o azúcares alcohólicos (Fiala et al., 1990; Weibull et al., 1990). Las bacterias pueden influir en la disponibilidad de estos sustratos mediante la producción de biosurfactantes que aumentan la humectabilidad, de fitohormonas como el ácido indol-3-acético que aumentan la disponibilidad de nutrientes al estimular el desprendimiento de la pared celular y la liberación de hidratos de carbono contenidos en ella (Vorholt, 2012). Además de las fuentes de carbono mencionadas, las plantas producen compuestos volátiles que pueden ser utilizados como fuente de carbono, como se ha evidenciado para bacterias metilotróficas del género *Methylobacterium* asociadas con frecuencia a plantas (Corpe & Rheem, 1989; Knief et al., 2010). En cuanto a la disponibilidad de nitrógeno y de micro y macroelementos, estudios indican la utilización de aminoácidos como fuente de nitrógeno y el proceso de fijación biológica por parte de bacterias en la filósfera (Fürnkranz et al., 2008), así como la expresión in situ de varios sistemas de transporte de fosfato, sulfato y hierro para la adquisición de macro y microelementos (Gourion et al., 2006; Marco et al., 2005). Finalmente, estudios recientes han reportado bacterias de la filósfera con el potencial de beneficiarse de la luz mediante la expresión de bacteriorrodopsinas, y los fotótrofos anoxigénicos aerobios podrían ser capaces de capturar la luz en los centros de reacción fotosintéticos (Atamna-Ismaeel, Finkel, Glaser, Sharon, et al., 2012; Atamna-Ismaeel, Finkel, Glaser, von Mering, et al., 2012).

Regulación de los atributos de aptitud: en la filósfera, las condiciones de estrés son una norma más que una excepción, por lo que es de interés conocer con en qué grado los sistemas de regulación que funcionan bajo estrés son “sistemas alternativos”. Entre estos sistemas se encuentran el uso de factores sigma alternativos, que gobierna la respuesta general a estrés en *Pseudomonas* (Hagen et al., 2009), sistema de regulación de dos componentes como GacSR envuelto en la producción de toxinas y formación de lesiones en cepas patógenas de *Pseudomonas*, la regulación mediada por segundos mensajeros como el guanósín monofosfato cíclico dimerico (di-c-GMP) envuelto en procesos distintivos como la adhesión, y finalmente, el proceso de *quorum sensing*, mediado por diversas moléculas dependiendo del sistema biológico, que permite a las bacterias procesar información respecto a la densidad poblacional y la difusión de metabolitos, y responder de manera adecuada al ambiente en el que se encuentran (Ryan & Dow, 2011; von Bodman et al., 2003).

Los rasgos descritos anteriormente, cruciales para el éxito del crecimiento epifítico, están estrechamente relacionados con las interacciones con otros miembros de la comunidad; estos rasgos, en última instancia, determinarán la coexistencia o exclusión de otras bacterias de la filósfera (Hibbing et al., 2010). En las comunidades microbianas, la competencia por el espacio y los recursos nutritivos, así como la producción de antibióticos y la interferencia con los sistemas de señalización celular, se consideran los principales mecanismos mediante los cuales las bacterias y los hongos autóctonos se antagonizan entre sí (McCormack et al., 1994; Wilson & Lindow, 1994).

I.2.c Microbioma endofítico

La endosfera refiere a los tejidos internos de la planta, y comprende la endo-rizósfera (tejido interno de la raíz) y la endo-filósfera (tejidos internos del tallo y las hojas). La presencia de microorganismos en la endosfera se ha registrado por años, lo que condujo a la definición de “endófito” como aquellos aislamientos de órganos vegetales esterilizados en su superficie. Desde la perspectiva agronómica, a este concepto se añade la condición de no generar un daño visible en la planta hospedadora. De acuerdo con sus estrategias de vida, los microorganismos que habitan este nicho pueden ser clasificados como endófitos “obligados” o “facultativos”, donde los primeros dependen de la planta para su crecimiento, supervivencia y transmisión, y los últimos tienen un ciclo vital bifásico, alternando entre los tejidos vegetales y el ambiente (mayoritariamente el suelo) (Hardoim et al., 2008). Considerando que la mayor fuente de diversidad microbiana para las plantas es el suelo, es esperable que la gran mayoría de microorganismos endófitos tengan un estilo de vida bifásico. Muchos microorganismos endófitos se originan en la rizósfera, atraídos por los exudados radiculares y rizodépósitos, e infectan la planta ingresando por grietas en las raíces laterales y pelos radiculares, extendiéndose a los espacios intercelulares de la raíz (Chi et al., 2005; Prieto et al., 2011). Las grietas o heridas en la raíz se consideran las principales regiones de ingreso para los microorganismos, a pesar de que se han descrito otras vías de acceso a la endosfera como los estomas, hidatodos y heridas causadas por fitopatógenos microbianos o nematodos en el tejido foliar (McCully, 2001), la colonización de las semillas que hacen de banco inicial de diversidad microbiana, o mediante la colonización de la endósfera y carpósfera (Compant et al., 2010). Es por esto por lo que se infiere que, para establecerse con éxito en la endosfera, los microorganismos que provienen del suelo deben ser colonizadores competentes de la zona radicular.

Se considera que una combinación de factores ambientales y genéticos contribuyen a que una bacteria tenga la capacidad de colonizar la endosfera. Esto da lugar al concepto de “endófitos competentes” para señalar aquellas bacterias que poseen la maquinaria genética clave requerida para colonizar la endosfera y persistir en ella, en contraste con aquellas bacterias oportunistas que colonizan con éxito nichos externos de la planta y podrían convertirse en endófitos sin contar con genes claves para habitar la endosfera (Hardoim et al., 2008). En la Figura I.1 se muestra una imagen adaptada de Hardoim et al. 2008 que exhibe distintos tipos de bacterias endófitas de acuerdo a si colonizan la planta transitoriamente por una casualidad circunstancial (como la colonización de heridas), los endófitos oportunistas que pueden exhibir características asociadas a la colonización de la raíz como la quimiotaxis, pero al igual que los endófitos pasajeros se encuentran restringidos a algunos compartimientos vegetales como la corteza, y los endófitos competentes, adaptados a vivir en la planta, y capaces de expandirse y colonizar tejidos específicos.

también determinan la colonización endofítica y las estructuras de la comunidad endosférica (Hallmann & Berg, 2006 ; Van Overbeek & Van Elsas, 2008). Además, los rasgos bacterianos intrínsecos importantes para la colonización tienen un papel significativo como determinantes de la diversidad de endófitos. La capacidad de las bacterias del suelo para acercarse a las raíces de las plantas mediante la motilidad inducida por quimiotaxis, así como colonizarlas eficazmente a través de la adhesión y la formación de microcolonias, probablemente constituye uno de los factores determinantes más relevantes para su éxito en convertirse en endófitas (Bacilio-Jiménez et al., 2003).

Durante el proceso de colonización, la comunicación entre la planta y las bacterias es clave. Si bien se considera que las plantas se comunican simultáneamente con microorganismos comensales, mutualísticos, simbióticos y patogénicos mediante la exudación de compuestos en las raíces (Bais et al., 2006), se ha sugerido que puede tomar lugar una comunicación que tiene como fin atraer microorganismos específicamente para su propio beneficio ecológico y evolutivo (Sørensen & Sessitsch, 2006). Si bien el estudio de los procesos que llevan a esta hipotética selección es difícil de abordar debido a la complejidad que suponen las interacciones planta-microorganismos, muchos estudios han indicado una selección de bacterias en la rizósfera, y es de esperarse que procesos similares lleven a la selección de bacterias en la endosfera (Van Overbeek & Van Elsas, 2008). Así lo sugieren estudios sobre la asociación endofítica de *Azorhizobium caulinodans*, *Azospirillum brasilense* y *Serratia* spp con trigo y arroz, que implican compuestos específicos como flavonoides, en la comunicación química entre la planta y bacterias colonizantes (Balachandar et al., 2006; Webster et al., 1998).

Como se describió para microorganismos que colonizan otros nichos vegetales, las bacterias asociadas a la endosfera también pueden sintetizar fitohormonas como auxinas, giberelinas y citoquininas (Glick, 2012). La producción externa a la planta de estas hormonas impacta tanto en la salud de la planta como en el proceso de colonización por parte de los microorganismos. La producción de la auxina ácido indol-3-acético (AIA) se ha reportado en bacterias asociadas a plantas, influyendo en el desarrollo de la planta y en ocasiones funcionando como una molécula señal para la comunicación bacteriana (Spaepen et al., 2007). La asociación planta-microorganismo se puede ver afectada por la producción de IAA considerando que el exceso de esta auxina puede causar un desbalance en el crecimiento, e incluso que las mismas vías de síntesis pueden llevar a interacciones específicas, como se ha indicado en bacterias benéficas que producen AIA mediante la vía de indol-3-piruvato, y en bacterias patogénicas utilizando mayormente la vía de indol-3-acetamida (Persello-Cartieaux et al., 2003; Xie et al., 1996). Por otro lado, la comunicación cruzada reportada entre AIA y etileno cumple un rol esencial en la interacción entre las bacterias endófitas y las plantas. El etileno es una fitohormona que afecta el desarrollo de las plantas mediante diversos mecanismos, entre los que se encuentran la promoción de la biogénesis de raíces, la inhibición de su elongación, la promoción de la maduración de los frutos y del marchitamiento de las flores, la estimulación de la germinación de las semillas, la promoción de la abscisión de las hojas, la activación de la síntesis de otras hormonas vegetales, la inhibición de la formación de nódulos por rizobios, la inhibición de la interacción micorriza-planta y la respuesta a las presiones bióticas y abióticas (Abeles et al., 2012). La fitohormona AIA induce la transcripción de la enzima 1-aminociclopropane-1-carboxylate deaminasa (ACC deaminasa), que degrada el ácido 1-aminociclopropano-1-carboxílico, precursor del etileno, en amoníaco y α -cetobutirato, compuestos fácilmente asimilables (Glick, 2012). Hasta la fecha no se ha reportado que la ACC deaminasa se excrete al citoplasma (Glick et al., 2007), por lo que la capacidad de las bacterias en disminuir los niveles de etileno depende de su capacidad para absorber el ACC antes de que sea oxidado por la ACC oxidasa de la planta (Glick et al., 1998). En este contexto, los endófitos bacterianos con una elevada actividad de ACC deaminasa inducida localmente podrían ser excelentes promotores del crecimiento vegetal, ya que mejoran el estrés de la planta bloqueando eficazmente la producción de etileno (Cheng et al., 2007), y las plantas podrían seleccionar

bacterias productoras de ACC-deaminasa para que ingresen a los tejidos endofíticos, atenuando así el estrés de la planta causado por altos niveles de etileno (Arshad et al., 2007). Así, el equilibrio entre las hormonas vegetales, IAA y etileno podría ser de importancia fundamental para el mantenimiento de las bacterias en el interior de la planta.

En la Figura I.2 extraída de Afzal et al. 2019 se resume en un esquema lo descrito en esta sección. Las bacterias endófitas pueden beneficiar a su hospedador de forma directa, mediante la absorción de nutrientes o la modulación de fitohormonas, o de forma indirecta como lo es el biocontrol de plagas o patógenos. Sin embargo, para lograr un impacto en la planta estas bacterias deben primero colonizar los tejidos internos. Durante la invasión y el establecimiento en la planta tienen lugar eventos que ponen a prueba la competitividad de las bacterias resultado de una batería de rasgos de colonización. Este es un proceso secuencial que para la mayoría de los microorganismos endófitos comienza en la rizosfera de la planta, seguido del rizoplasma (superficie de la raíz) y la colonización del interior de la raíz. Una vez dentro, los endófitos competentes pueden pasar a las partes aéreas de las plantas (tallo y hojas). Finalmente, la diversidad de endófitos colonizadores se ve afectada por factores propios de las bacterias, las plantas y el medio ambiente, y el análisis de esta diversidad y la capacidad de encontrar nuevos microorganismos endófitos y evaluar su impacto sobre la salud vegetal está sujeto a las técnicas disponibles con las que podamos estudiarlos.

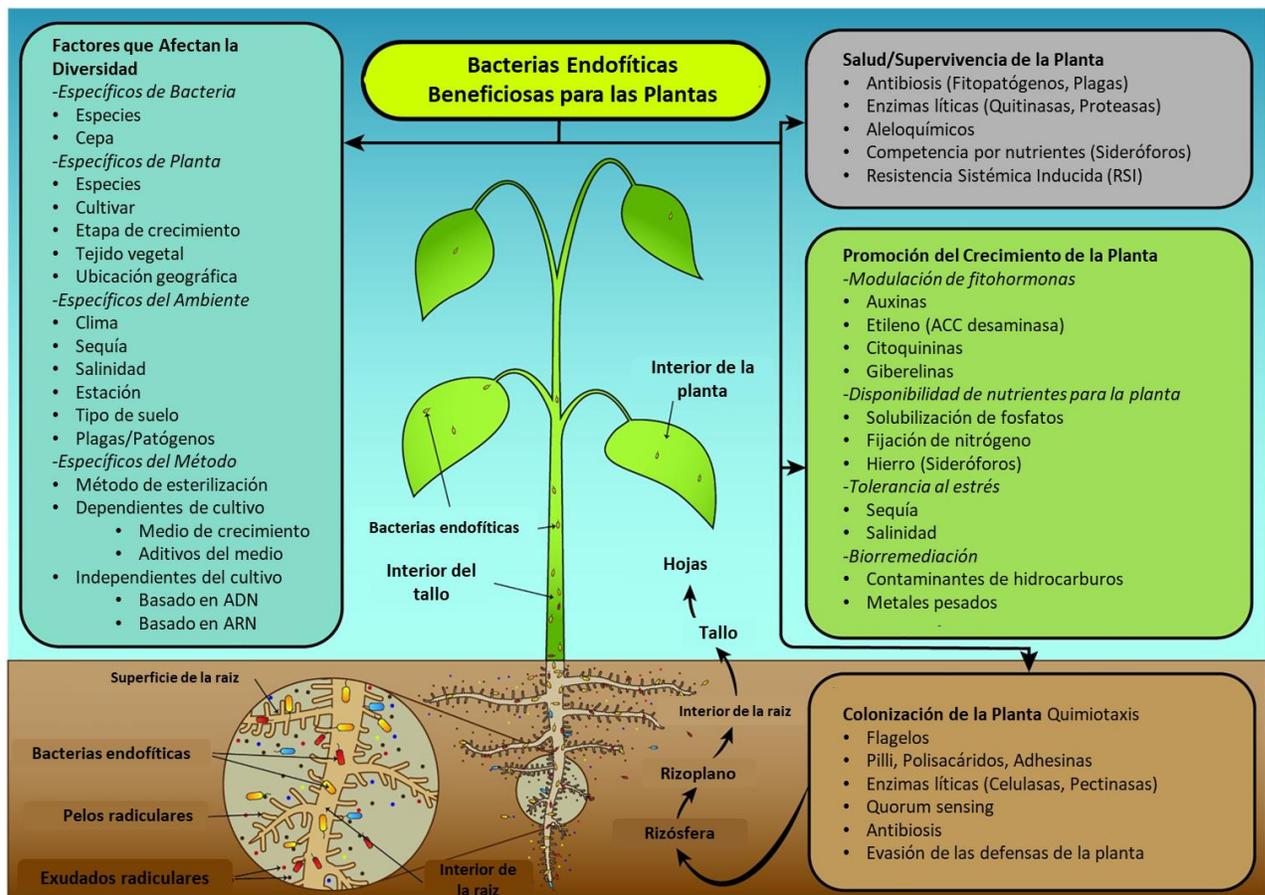


Figura I.2. Extraída y adaptada de Afzal et al. 2019. Mecanismos de promoción del crecimiento vegetal, colonización y factores que afectan a la diversidad de bacterias endofíticas en la planta huésped.

I.2.d Microbioma de semillas

La microbiota que habita las plantas proviene principalmente del ambiente que las rodea. Sin embargo, la primera fuente de microorganismos con los que toma contacto es aquella presente en las semillas que, si bien es considerablemente menor a la proporcionada por el ambiente, la capacidad de relacionarse en etapas tempranas del desarrollo vegetal y de permanecer en las semillas en dormancia, hacen a estos microorganismos de particular interés. Cuando las semillas comienzan a germinar, absorben agua y comienzan a secretar exudados que atraen bacterias que colonizan la espermósfera, rizósfera y finalmente la plántula (Nelson, 2004). Además de los microorganismos presentes en el exterior de las semillas, las bacterias presentes en el interior de estas pueden ser igualmente importantes para la evolución de la comunidad microbiana en la planta (Johnston-monje & Raizada, 2011).

Quizás una de las propiedades más destacables de este microbioma, es su potencial para ser transmitido de generación en generación. Esta transmisión vertical implica una selección por parte de la planta para convocar y heredar microorganismos mutualistas, que dependen en gran medida o por completo de la planta para su supervivencia y reproducción (Ewald, 1983; Rudgers et al., 2009). Por otro lado, las condiciones dentro de la semilla cambian durante su maduración, lo que determina que las bacterias que son capaces de habitarlas presenten un perfil característico que las distingue de aquellas presentes en raíces y tallos. La tolerancia a la presión osmótica es una característica relevante para estos endófitos debido a la acumulación de almidón y a la pérdida de agua que ocurre durante el proceso de maduración de la semilla. La capacidad de degradar el almidón mediante el uso de amilasas y de detener y retomar su crecimiento luego de un largo período de supervivencia en las semillas es otra característica asociada a estos microorganismos endófitos (Mano et al., 2006). El uso de fitato, principal fuente de almacenamiento de fósforo en las semillas, por parte de estos microorganismos ha sido también reportado (López-López et al., 2010) junto con la secreción de acetoína y la fijación de nitrógeno, lo que refleja una conservación de propiedades comunes necesarias para establecerse en la semilla y la espermósfera (Nelson, 2004).

Diversos trabajos han reportado bacterias pertenecientes a los filos *Proteobacteria*, *Actinobacteria*, *Firmicutes* y *Bacteroidetes*, como endófitos naturales de semillas de distintas especies vegetales. En la Tabla I.2 extraída de Truyens et al. 2014, se muestra la diversidad de aislamientos bacterianos pertenecientes a distintos filos descriptos como endófitos de semillas para una amplia variedad de especies vegetales, donde se observa también que las Proteobacterias, especialmente las γ -Proteobacterias, son el filo endofítico aislado dominante (Rosenblueth & Martínez-Romero, 2006). En general, los géneros bacterianos más frecuentes en semillas de especies vegetales muy diferentes son *Bacillus*, *Pseudomonas*, *Paenibacillus*, *Micrococcus*, *Staphylococcus*, *Pantoea* y *Acinetobacter*. Los taxones encontrados en el interior de las semillas son similares a los que se observan en el suelo indicando que los microorganismos allí presentes son capaces de colonizar la planta hasta establecerse en las semillas. Sin embargo, el camino desde el suelo hasta la semilla es complejo, lo que limita y reduce significativamente la diversidad en este microbioma, y señala a su vez características asociadas a la colonización de las plantas y la capacidad endofítica. Una de las cualidades descripta en diversos estudios como esencial para la colonización es la motilidad inducida por quimiotaxis que presentan las bacterias y les permite establecerse en la rizósfera (Bacilio-Jiménez et al., 2003), como lo evidencia la reducción en la capacidad de colonización de la raíz mutantes en el gen *cheA* o *pilA* de cepas de *Pseudomonas* defectuosas en la quimiotaxis impulsada por flagelos (Camacho Carvajal, 2001; de Weert et al., 2002). En el siguiente paso en su camino a la semilla, las bacterias necesitan ingresar a la raíz, lo que puede ocurrir de forma pasiva al colonizar heridas o raíces laterales emergentes, o activamente mediante la expresión de enzimas que degradan la pared celular de las células vegetales, como celulasas y pectinasas (Elbeltagy

et al., 2000; Hurek et al., 1994; James et al., 2002). Una vez dentro de la planta, el nuevo ambiente impulsa un cambio en el metabolismo y la regulación de mecanismos que les permitan la transición de un crecimiento en vida libre a una vida endofítica, como la expresión de proteínas que absorban los nutrientes sintetizados por la planta y el uso de mecanismos de detoxificación que los protegen del estrés oxidativo inducido por la planta luego de la infección (Ali et al., 2014; Sessitsch et al., 2012). Finalmente, algunos endófitos logran dispersarse de forma sistémica a otros órganos vegetales, utilizando los vasos del xilema e impulsados por la motilidad de sus flagelos y el flujo de transpiración de las plantas, mientras que otros colonizan los espacios intracelulares ricos en nutrientes utilizando enzimas líticas que degradan la pared celular, alcanzando así flores, frutos y finalmente, las semillas.

Tabla I.2. Extraída y adaptada de Truyens et al. 2014. Resumen de los endófitos bacterianos de semillas conocidos aislados de diferentes especies vegetales y el orden taxonómico al que pertenecen.

Planta	Actinobacteria	Bacteroidetes	Frimicutes	Proteobacteria			
				α	β	γ	δ
Píce de Noruega (<i>Picea abies</i>)	0	0	0	0	0	2	0
Vid (<i>Vitis vinifera</i>)	0	0	3	0	0	0	0
Eucalipto (<i>Eucalyptus</i>)	1	0	3	3	0	0	0
Arroz (<i>Oryza sativa</i>)	12	2	5	5	3	13	0
Poroto (<i>Phaseolus vulgaris</i>)	4	0	4	3	0	2	0
Tabaco (<i>Nicotiana tabacum</i>)	1	0	1	0	0	4	0
Cactus (<i>Pachycereus pringlei</i>)	0	0	2	0	0	3	0
Café (<i>Coffea arabica</i>)	3	0	1	0	1	5	0
Canola (<i>Brassica napus</i>)	1	1	2	0	2	4	0
Maíz (<i>Zea mays</i>)	18	5	14	16	17	33	1
Fresno (<i>Fraxinus</i>)	3	0	4	0	0	3	0
Soja (<i>Glycine max</i>)	1	1	1	2	0	2	1
Trigo (<i>Triticum aestivum</i> ; <i>Elymus trachycaukus</i> ; <i>Agropyron fragile</i>)	3	0	2	0	0	4	0
Remolacha azucarera (<i>Beta vulgaris</i>)	0	0	1	0	0	0	0
Arveja (<i>Pisum sativum</i>)	0	0	1	0	0	0	0
Calabaza (<i>Cucurbita pepo</i>)	0	0	1	0	0	1	0
Maní (<i>Arachis hypogaea</i>)	0	0	2	0	0	1	0
Coliflor (<i>Brassica oleracea</i>)	0	0	1	0	0	0	0
Mostaza salvaje (<i>Sinapis arvensis</i>)	0	0	1	0	0	0	0
Tomate (<i>Lycopersicon esculentum</i>)	0	0	1	0	0	0	0
Frutilla (<i>Fragaria</i>)	0	0	0	0	0	1	0
<i>Arabidopsis thaliana</i>	1	0	3	5	2	1	0
Pastos (<i>Lolium multiflorum</i> ; <i>Phleum pretense</i> ; <i>Panicumvirgatum</i> ; <i>Agrostis capillaris</i>)	1	0	3	0	0	3	0
Especies de malezas de hoja ancha	0	1	1	0	1	7	0
Especies herbáceas y leñosas	5	2	3	0	2	7	0

Aunque nuestra comprensión de las asociaciones entre semillas y microbios es más completa en algunas etapas que en otras, este marco proporciona un medio para generar nuevas hipótesis sobre el ensamblaje, la dinámica y la función del microbioma de las semillas. Es probable que las fuerzas que facilitan el ensamblaje y la estructura de los microbiomas de semillas sean bastante complejas, con microbios reclutados no solo del microbioma del suelo (Hardoim et al., 2012; Klaedtke et al., 2016), sino también de agentes de dispersión como por ejemplo polinizadores y microbiomas florales, algunos de cuyos miembros serán reclutados en última instancia en las semillas en desarrollo, se multiplicarán en la espermosfera durante la germinación de las semillas y posteriormente colonizarán las plántulas (Bangert et al., 1988; Czczuga & Godlewska, 2009; Gandolfi et al., 2013). En la Figura I.3 extraída de Nelson 2018 se ilustra el ciclo de vida de la planta y la importancia ecológica de cada etapa, donde los microbios adquiridos a lo largo del camino pueden contribuir potencialmente al microbioma que finalmente prolifera durante la germinación de las semillas y se transfiere a las plántulas y sus diversos órganos.

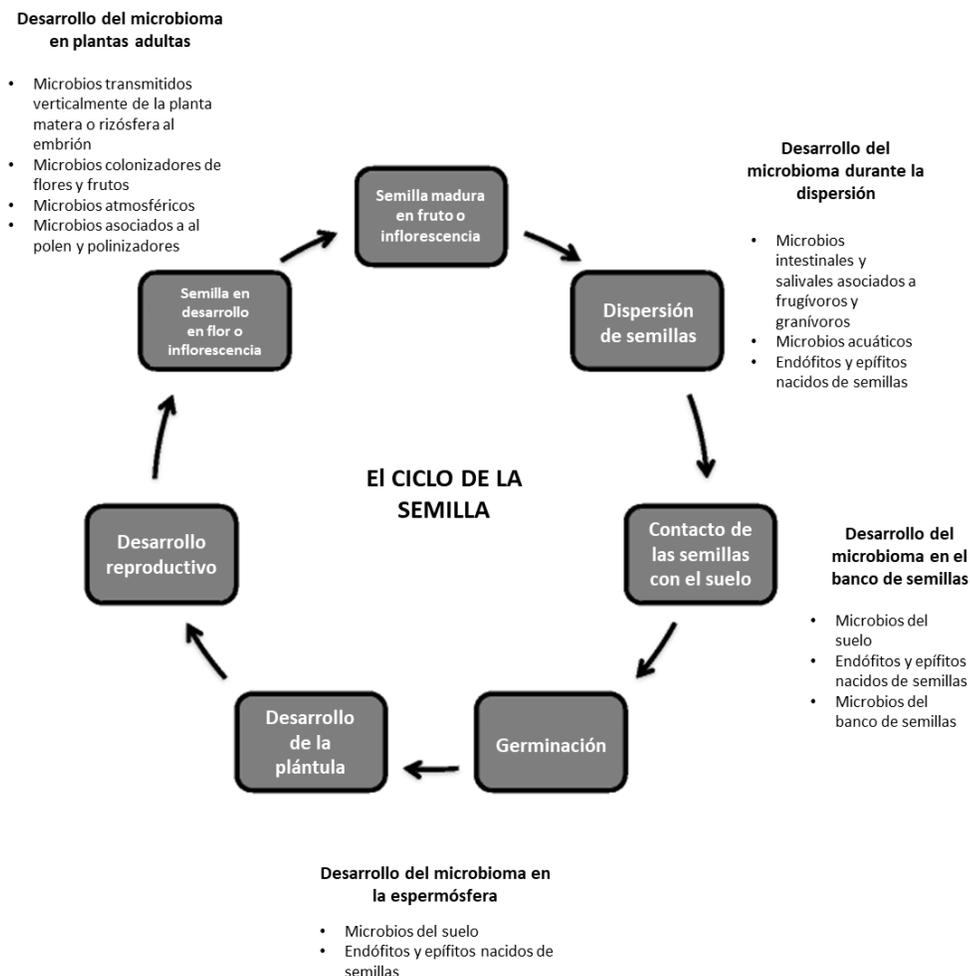


Figura I.3. Extraída y adaptada de Nelson 2018. Desarrollo vegetal de semilla a semilla y oportunidades de adquisición de microbios por las semillas en distintos estadios de este ciclo.

I.2.e Medicago sativa y su microbioma asociado

Las leguminosas son la tercera familia más grande de plantas superiores con 20.000 especies, la segunda familia más importante a nivel agronómico después de las gramíneas (Gepts et al., 2005), y

poseen también la particularidad de establecer relaciones simbióticas con rizobios. La leguminosa *Medicago sativa* posee un gran valor agronómico debido a su alto contenido proteico, alta producción de materia seca y digestibilidad, lo que la convierte en la principal fuente de forraje para alimento vacuno, con más de 32 millones de hectáreas en el mundo ocupadas por este cultivo (Michaud et al., 2015). En Argentina, cerca de 3 millones de hectáreas se utilizan para la producción de alfalfa (Basigalup, 2022). *Medicago* es una planta muy versátil, pudiendo adaptarse a un amplio rango de condiciones climáticas (James A. Duke, 1983), con excepción de períodos de inundaciones y suelos ácidos. Por otro lado, es importante para la producción de bioenergía (Sanderson & Adler, 2008) y es una de las especies vegetales más adecuadas para la restauración del suelo (A. D. Bradshaw and M. J. Chadwick, 1980). Finalmente, esta especie es utilizada como modelo para el estudio de los aspectos moleculares en la simbiosis planta-bacteria, especialmente en relación con la α -proteobacteria *Sinorhizobium* (*syn. Ensifer*) *meliloti* (Downie, 2010; Gibson et al., 2008; Oldroyd & Downie, 2008).

Si bien la asociación de *Medicago* con sus simbiontes rizobios ha sido vasta y profundamente estudiada por años, el conocimiento del microbioma asociado a la alfalfa es reducido. Pini et al. (2012) analiza la comunidad bacteriana asociada a la alfalfa, abarcando niveles taxonómicos que van desde categorías amplias, como clase y familia, hasta niveles más específicos, como especie y comunidades de *Sinorhizobium meliloti*. En dicho trabajo se estudió mediante métodos independientes de cultivo la población bacteriana asociada a suelo, nódulos, tallo y hojas de plantas de alfalfa crecidas en mesocosmos, revelando que las comunidades bacterianas del suelo y de nódulos estaban fuertemente diferenciadas de las comunidades del tallo y de las hojas, con una gran heterogeneidad en las comunidades presentes en estos dos últimos nichos. Como era de esperarse, la mayor diversidad bacteriana se encontró en el suelo, y la menor en nódulos. Las clases bacterianas *Alfaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Actinobacteria*, *Deltaproteobacteria*, *Bacili* y *Sphingobacteria* representadas tanto en las comunidades del suelo como en las del tallo y hojas, y las primeras cuatro también en los nódulos, donde la clase dominante fue *Alfaproteobacteria*. En el suelo esta clase también tuvo una alta prevalencia, junto con una gran abundancia de *Acidobacteria* y *Crenarchaeota*. *Flavobacteria* se encontró solo en nódulos, aunque no se puede descartar una presencia reducida en otros nichos. Las α -proteobacteria presentes en los nódulos correspondieron, como era de esperar, al género *Sinorhizobium/Ensifer*, y las presentes en el suelo pertenecían a las familias *Rhizobiaceae*, *Bradyrhizobiaceae*, *Methylocystaceae*, *Hypomicrobiaceae* y *Caulobacteraceae*. *Rhizobiaceae*, *Aurantimonadaceae* y *Methylobacteriaceae*, todas pertenecientes a *Rhizobiales*, además de taxones del orden *Sphingomonadales*, se encontraron en la biblioteca de Tallo + Hoja.

Como se describió en secciones anteriores, las bacterias asociadas a las plantas provienen de distintas fuentes de diversidad con las que la planta tiene contacto, y los microorganismos presentes en las semillas son, quizás, el primer contacto de la planta con una fuente de diversidad microbiana. Atento a la importancia de las bacterias presentes en las semillas, en nuestro grupo de trabajo se caracterizó el microbioma asociado a semillas de alfalfa mediante una exhaustiva búsqueda de aislamientos presentes en los exudados de cinco variedades de semillas (Super Monarca, GAPP810+, GAPP 909+, F401Y, F301Y y F111X) esterilizadas en superficie, utilizando métodos dependientes de cultivo, con el fin de caracterizar la diversidad presente, y sus principales rasgos fenotípicos. Se evaluó con el mismo fin las bacterias presentes en plántulas de alfalfa crecidas de semillas esterilizadas en superficie en condiciones asépticas, que fueron maceradas para recuperar aislamientos tanto del exterior como interior de la planta. Se obtuvo así una colección de 235 aislamientos pertenecientes 40 géneros diferentes, identificados mediante espectrometría de masas MALDI-TOF y secuenciación del gen 16s rADN, que conforman la colección de endófitos de alfalfa disponible en nuestro laboratorio (López et al., 2018). En la figura I.4, extraída de dicho trabajo, se muestran los géneros a los que pertenecen los aislamientos recuperados

tanto de los exudados de semillas como de las plántulas, indicados son una S y una P respectivamente. La diversidad recuperada pertenece a cuatro filos diferentes, donde las *Actinobacteria* fueron las más representadas (42,5%), seguidas de las *Proteobacteria* (32,5%), los *Firmicutes* (22,5%) y, en una proporción menor, los *Bacteroidetes* (2,5%). Muchos de los géneros identificados eran redundantes entre las diferentes variedades de alfalfa, apoyando así la noción de un núcleo bacteriano conservado que coloniza las semillas de alfalfa. Así, se recuperaron aislamientos de los géneros *Bacillus*, *Kocuria*, *Arthrobacter*, *Arsenicoccus*, *Microbacterium*, *Micrococcus* y *Paenibacillus* de las semillas de al menos 3 variedades diferentes. Los aislamientos pertenecientes al género *Pantoea* fueron menos frecuentes, pero resultaban los endófitos de semillas más numerosos cuando se encontraban.

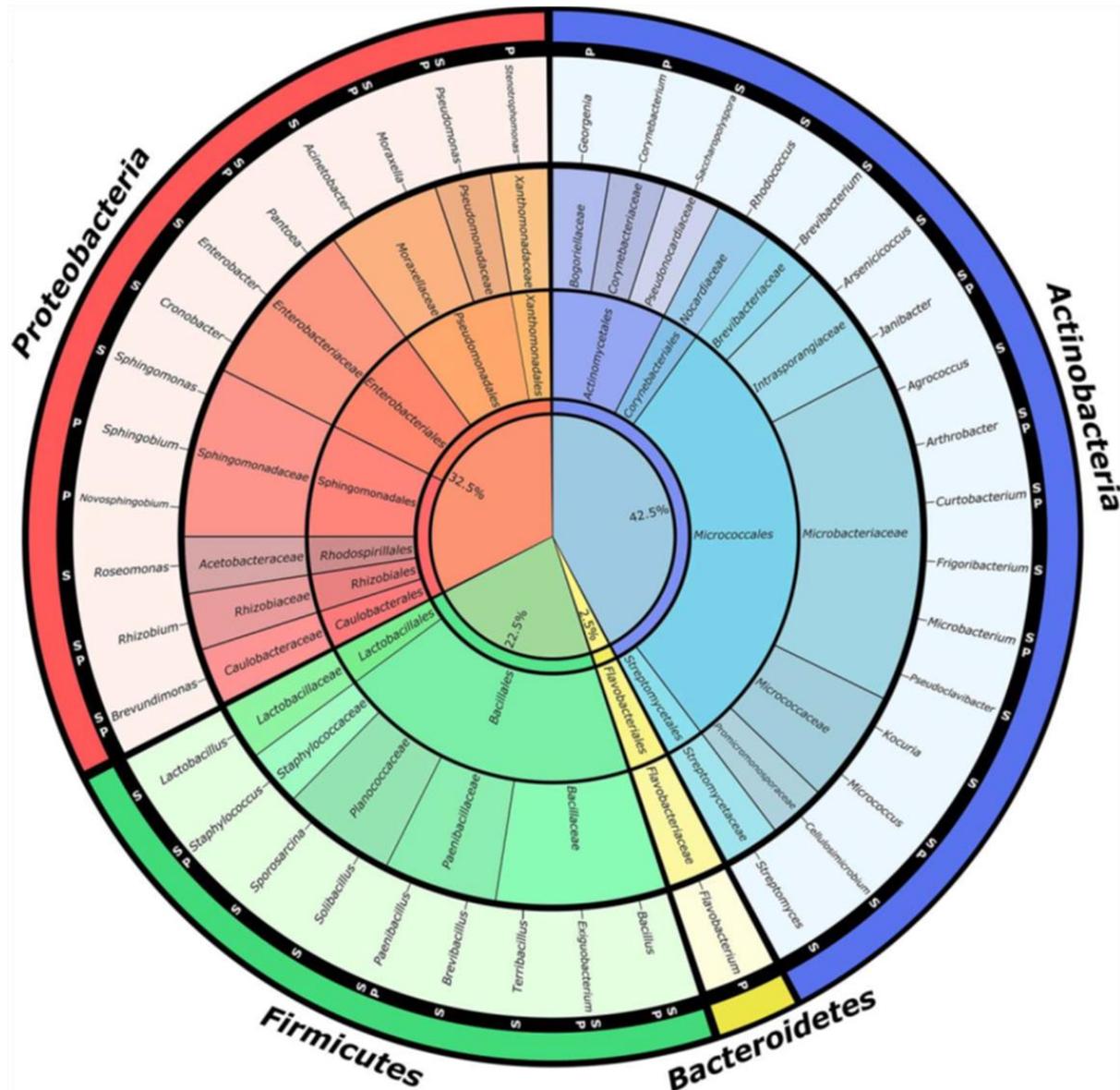


Figura I.4. Extraída de López et al. 2018. Diversidad bacteriana en clones endófitos derivados de semillas de alfalfa esterilizadas en superficie. Representación porcentual de los géneros bacterianos presentes en los endófitos de semillas. Los porcentajes corresponden a la representación de cada uno de los filos indicados, calculada a partir del

número de géneros diferentes asociados a las semillas. Los filos y el número de géneros (entre paréntesis) fueron los siguientes: *Actinobacterias* (17), *Proteobacterias* (13), *Firmicutes* (9) y *Bacteroidetes* (1).

La caracterización fenotípica de actividades relacionadas con la promoción del crecimiento vegetal en los aislamientos de esta colección reveló que las más prevalentes fueron: pectinasas, producción de sideróforos y auxinas, celulasas, amilasas, proteasas, solubilización de fosfatos, actividad quitinasa e inhibición del crecimiento fúngico. Se encontró que varias cepas de la colección mostraban más de una de estas actividades, como aislamientos del género *Bacillus*, que presentaban una notable presencia de enzimas hidrolíticas y actividades inhibitorias de hongos, y miembros del género *Pantoea*, que exhibían una destacada actividad de solubilización de fosfatos, así como producción de sideróforos y ácido indolacético.

Estudios recientes se enfocaron en evaluar la posibilidad de utilizar aislamientos de esta colección para la promoción del crecimiento vegetal y el biocontrol frente a patógenos de interés. Estos estudios demostraron la capacidad de dos aislamientos de esta colección, pertenecientes al género *Bacillus*, para inhibir el crecimiento de *Colletotrichum trifolii* y *Phytophthora megasperma*, causantes de dos de las enfermedades fúngicas más críticas en Argentina: la antracnosis y la podredumbre radicular, respectivamente (Gieco, 2007). Uno de estos aislamientos, B4, mostró además compatibilidad con el rizobio INTA5, un aislamiento de interés debido a su baja emisión de N₂O, sentando así las bases para generar un consorcio bacteriano que promueva el biocontrol y la fijación biológica de nitrógeno (Moreno et al., 2024).

La colección de endófitos de semillas de alfalfa en nuestro laboratorio se convierte así en un recurso invaluable para estudios de ciencia básica y aplicada, que permitan profundizar el conocimiento y caracterización de la asociación planta-bacteria, desde sistemas reducidos con dos participantes (aislamiento-planta) hasta sistemas complejos con múltiples integrantes (comunidades sintéticas-planta huésped), y evaluar el potencial de estos aislamientos para desarrollar nuevas estrategias en programas de mejoramiento de cultivos de alfalfa.

I.3. Microorganismos multifacéticos: una mirada al género *Pantoea* y su asociación a plantas

El género *Pantoea*, perteneciente a la familia *Enterobacteriaceae*, representa un grupo diverso de γ -*proteobacterias* gramnegativas que ha captado la atención de la comunidad científica debido a su capacidad para establecer una amplia gama de interacciones con diferentes organismos, particularmente con las plantas. Desde su primer aislamiento y descripción, *Pantoea*, filogenéticamente cercano a *Tatumella* y *Erwinia*, ha experimentado una notable expansión. Actualmente, se reconoce que el género está compuesto por aproximadamente 20 especies, fenotípicamente similares, que se agrupan en 13 grupos de hibridación, como se ha propuesto en varios estudios. Las cepas tipo de *Pantoea* que definen cada uno de estos grupos han sido aisladas de una gran diversidad de fuentes, con las plantas siendo una de las principales. Algunas especies como *P. deleyi*, *P. anthophila*, *P. allii*, *P. cyripedii*, *P. wallisii*, *P. rodasii* y *P. rwandensis* han sido aisladas exclusivamente de plantas. Por otro lado, especies como *P. conspicua*, *P. breneri*, *P. septica* y *P. eucrina* han sido reportadas únicamente en muestras clínicas. Otras especies, como *P. calida*, *P. dispersa* y *P. gaviniae*, han sido identificadas tanto en ambientes naturales como en productos procesados (Walterson & Stavrínides, 2015). La diversidad de ambientes de los que se han aislado cepas de este género, incluyendo suelos, agua y organismos vivos, subraya su versatilidad

ecológica y funcional, y la composición de especies, tanto comensales benignas como patógenos oportunistas de plantas, animales y humanos, sugiere una considerable plasticidad genética y adaptativa dentro del grupo (Figura I.5). Esta capacidad para colonizar y sobrevivir en ambientes diversos destaca la complejidad y el potencial biotecnológico de este género.

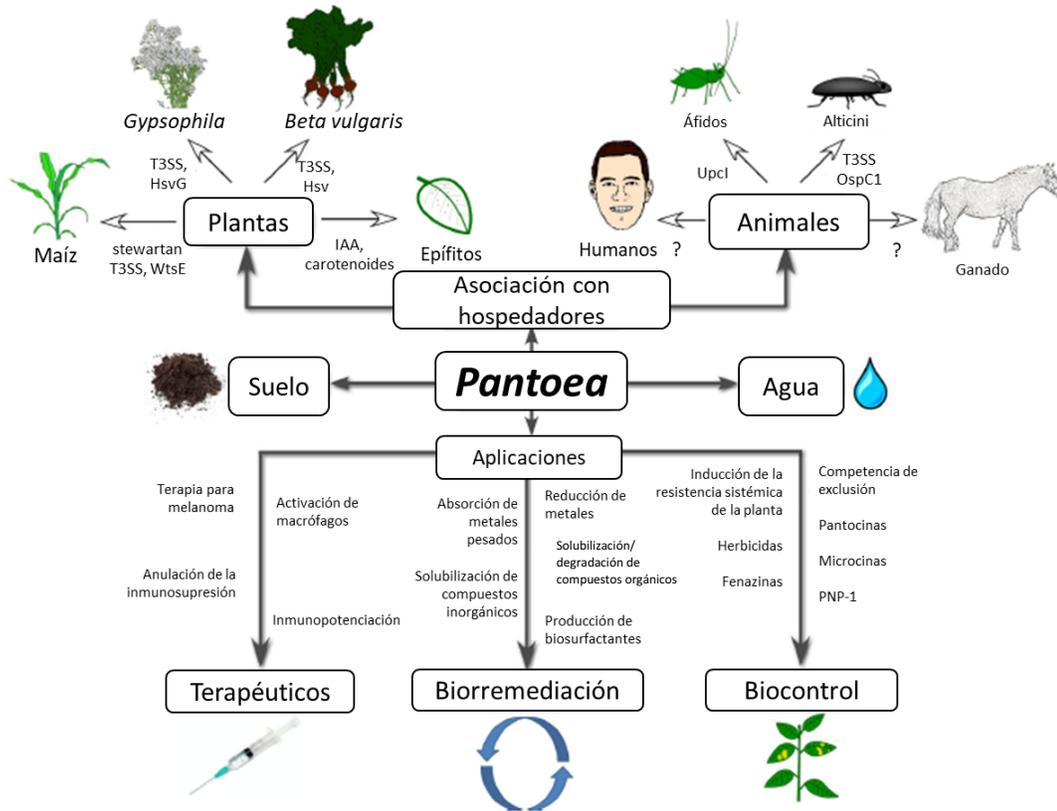


Figura I.5. Extraída y adaptada de Walterson & Stavrinides (2015). Diversidad y potencial de las cepas del género *Pantoea*. Aislamientos de *Pantoea* han sido aislados de diversos entornos de suelo y agua, así como en asociación con diferentes hospedadores. Los determinantes de la asociación con diversas plantas incluyen las proteínas T3SS y T3SEs (HsvG, HsvB y WtsE), así como carotenoides y ácido indol-3-acético (IAA). También se ha determinado la participación de varios determinantes en la asociación de *Pantoea* con diversos insectos, sin que se disponga prácticamente de información sobre los factores que facilitan la colonización de huéspedes mamíferos, incluidos los seres humanos. Algunas cepas específicas y sus productos naturales han demostrado ser prometedores para aplicaciones terapéuticas, incluyendo la terapia del cáncer y la reversión de la inmunosupresión, mientras que otros han sido explorados por sus capacidades de biorremediación, incluyendo la degradación de materiales orgánicos y la biosíntesis de surfactantes. Se siguen evaluando aislamientos de *Pantoea* para el control biológico y la gestión de plagas de plantas, como su capacidad para superar a los patógenos mediante la exclusión competitiva y la producción de antibióticos, y su capacidad para inducir la resistencia sistémica de las plantas.

La asociación de bacterias *Pantoea* con plantas ha sido reportada desde mediados del siglo XX, cuando fue aislada de tejidos vegetales y se observó su capacidad para causar enfermedades en cultivos agrícolas, así como para vivir como endófito sin causar daño aparente a la planta que la hospeda (Barash & Manulis-Sasson, 2009; Brady et al., 2008; Handelsman & Brill, 1985). A lo largo de las décadas, se han reportado diversas especies de *Pantoea* que colonizan diferentes partes de las plantas, incluyendo la

rizósfera y la endósfera, en cultivos de importancia agrícola como el maíz (*Zea mays*), el trigo (*Triticum aestivum*), y el arroz (*Oryza sativa*), así como de plantas ornamentales y silvestres. Las interacciones que establece con sus hospedadoras son diversas, y pueden dar lugar a asociaciones beneficiosas o perjudiciales dependiendo del contexto ecológico y de la especie involucrada (Walterson & Stavrínides, 2015). Por ejemplo, *Pantoea agglomerans* es conocida tanto por su capacidad para promover el crecimiento de las plantas mediante la producción de fitohormonas como por su rol como patógeno, causando necrosis en las hojas y frutos de varias especies vegetales (Cooksey, 1986; Dutkiewicz et al., 2016; Lorenzi et al., 2022; Lv et al., 2022). En la rizósfera, *Pantoea* puede interactuar con las raíces de las plantas, contribuyendo a la solubilización de fosfatos y la producción de sideróforos, lo que mejora la disponibilidad de nutrientes para la planta y, en algunos casos, proporciona resistencia a patógenos. Por otro lado, en la endósfera, algunas especies han mostrado capacidad para inducir enfermedades, destacando su rol como patógenos emergentes en la agricultura (Brady et al., 2008).

Los estudios realizados en *Pantoea* y su interacción con plantas se han enfocado principalmente en su asociación parasitaria con especies como maíz, algodón, melón y cebolla (Walterson & Stavrínides, 2015). Si bien las enfermedades que puede causar son diversas, en muchos casos se ha mostrado que la patogenicidad está facilitada por factores que incluyen sistemas de secreción tipo III, sistemas mediados por quorum sensing y exopolisacáridos (EPS) (Roper, 2011). En *P. stewartii subsp. stewartii*, el quorum sensing influye en la adhesión, motilidad y dispersión, así como en la producción de EPS, convirtiéndose en un factor clave para el desarrollo de la enfermedad de Stewartii en cultivos de maíz. En este contexto, la bacteria coloniza el xilema de la planta y produce EPS, bloqueando el flujo de agua, lo que provoca que el cultivo se marchite (Koutsoudis et al., 2006; von Bodman et al., 2003). La producción de EPS en estas bacterias ha sido caracterizada por su rol central en el desarrollo de la enfermedad y se encuentra regulada por el sistema de quorum sensing EsaI/EsaR, que gobierna la expresión del clúster de genes *cps* responsables de la biosíntesis de exopolisacáridos en un sistema dependiente de la densidad celular (Koutsoudis et al., 2006). En este sistema, EsaR actúa mediante la represión de *rcaA*, donde cumple el papel de coactivador para la transcripción del clúster de genes *cps* mediado por RcsA/RcsB (Minogue et al., 2002, 2005). También se ha descrito el rol del EPS en la protección de la bacteria frente a los sistemas de defensa de la planta (Wehland).

Para algunos aislamientos de *Pantoea*, la competencia y supervivencia epifítica es facilitada por la producción de ácido indol-3-acético (AIA) (Brandl et al., 1996; Brandl & Lindow, 1998). De manera similar, la producción de carotenoides, responsables de la pigmentación amarilla característica de estas bacterias, juega un rol central en la resistencia tanto a la luz ultravioleta (300-400 nm) como frente a especies reactivas de oxígeno (ROS). Esto ha sido evidenciado en estudios con mutantes carentes de pigmentación, que muestran una disminución general en la supervivencia en presencia de peróxido de hidrógeno (Mohammadi et al., 2012). A nivel genético, se ha documentado una considerable variabilidad en los genes asociados a la adaptación a distintos hospedadores y nichos ecológicos. Estudios recientes han identificado genes en *Pantoea* que están relacionados con la adhesión a superficies vegetales, la resistencia a condiciones adversas como la desecación y la presencia de compuestos antimicrobianos, así como la capacidad de degradar compuestos complejos presentes en los tejidos vegetales (Herrera et al., 2008; Maayer et al., 2017; Walterson & Stavrínides, 2015). Algunos de estos genes son compartidos entre cepas de *Pantoea* que ocupan nichos ecológicos similares, lo que sugiere la existencia de elementos genéticos móviles que facilitan la adaptación a entornos específicos (Maayer et al., 2017).

La plasticidad genética de este género ha facilitado su uso en estudios experimentales, permitiendo aplicar técnicas como la mutagénesis dirigida y la transposon-insertion sequencing (TIS) para identificar genes clave involucrados en interacciones específicas con plantas (Duong et al., 2018). Esta capacidad de ser manipulada genéticamente resulta especialmente valiosa en estudios fenómicos, que

buscan correlacionar genes específicos con fenotipos observados bajo distintas condiciones ambientales o en presencia de diversos hospedadores. La plasticidad genética no solo facilita la identificación de genes implicados en la interacción con otros microorganismos y hospedadores, sino que también abre la posibilidad de explorar cómo la variabilidad genética dentro del género influye en su capacidad para colonizar y afectar a diferentes plantas

El conocimiento sobre los determinantes genéticos específicos que permiten a las cepas de *Pantoea* colonizar exitosamente entornos ambientales o interactuar con hospedadores sigue siendo limitado. En el caso de *P. stewartii* subsp. *stewartii* y su asociación con el maíz dulce, los principales factores de virulencia son el sistema de secreción tipo III (T3SS, PSI-1) y el exopolisacárido stewartan, que facilitan la infección de la planta (ROPER, 2011), mientras que para su dispersión a través del escarabajo pulgoso, se involucra otro sistema de secreción tipo III, el PSI-2. En *P. agglomerans*, la patogenicidad requiere del T3SS en conjunto con efectores específicos (T3SE), que se encuentran codificados en un plásmido conocido como pPATH (Ezra et al., 2004; Nizan et al., 1997). Sin embargo, para la mayoría de las especies de *Pantoea*, los determinantes genéticos que les permiten establecerse tanto en entornos hospedadores como no hospedadores siguen siendo en gran medida desconocidos. La disponibilidad de genomas secuenciados ha facilitado la identificación y comparación de los factores genéticos que podrían contribuir a la capacidad de ciertos aislados para prosperar en distintos ambientes. Aunque aún no se ha llevado a cabo un análisis genómico comparativo o evolutivo exhaustivo a gran escala, los estudios han revelado características genómicas comunes en algunas especies, como la presencia de acilhomoserina lactonas y otros genes de detección de quórum (Smits et al. 2010; Hong et al. 2012), genes que promueven el crecimiento vegetal (Kim et al., 2012), genes de reparación del ADN (Remus-Emsermann et al., 2013), factores de patogenicidad (De Maayer et al., 2012) y sistemas de secreción tipo IV y VI (Kim et al., 2012; Medrano & Bell, 2012). Además, algunas cepas poseen genes responsables de la producción de fitotoxinas y compuestos antimicrobianos, así como enzimas líticas que les permiten causar síntomas de enfermedad en las plantas y competir por nichos ecológicos. Estos genes y su variabilidad destacan el potencial de *Pantoea* para interactuar con diversos entornos y hospederos, lo que subraya la necesidad de continuar investigando los mecanismos genéticos subyacentes que definen estas interacciones (De Maayer et al., 2011; Walterson & Stavrínides, 2015).

Además de su sorprendente versatilidad y adaptabilidad ambiental, algunos aislamientos de *Pantoea* poseen una amplia gama de capacidades biosintéticas y biodegradativas que pueden ser aprovechadas para aplicaciones en entornos agrícolas, medioambientales y clínicos, lo que incrementa su valor e interés en el campo de la biotecnología (Figura I.5). Muchas cepas son competidoras eficientes en el ambiente y producen diversos productos naturales con actividad antibiótica, tales como pantocinas, herbicolinas, microcinas y fenazinas. Un descubrimiento reciente ha sido el PNP-1 producido por *P. ananatis* BRT175, que presenta actividad inhibidora contra *E. amylovora*. Debido a la diversidad de compuestos naturales que producen, junto con su capacidad para persistir en el medio ambiente, algunas cepas se han desarrollado como productos comerciales de biocontrol, como Blight-Ban C9-1 y Bloomtime Biological, que utilizan *P. vagans* C9-1 y *P. agglomerans* E325, respectivamente (Walterson & Stavrínides, 2015). Por otro lado, algunos aislamientos de *Pantoea* recuperados del suelo muestran capacidades prometedoras como la reducción de metales, la solubilización y degradación de materiales orgánicos, la producción de biosurfactantes, y la solubilización de inorgánicos insolubles, lo que las convierte en un candidato prometedor para procesos relacionados con la biorremediación en diversos entornos (Walterson & Stavrínides, 2015).

Las características descritas para el género *Pantoea* posicionan a estas bacterias como un sistema modelo versátil y genéticamente manipulable para el estudio de las interacciones bacteria-planta, además de su potencial aplicación en tecnologías de biocontrol y biorremediación. Su capacidad para establecer

relaciones tanto beneficiosas como perjudiciales con una amplia variedad de plantas, combinada con su notable plasticidad genética, convierte a *Pantoea* en un objeto de estudio ideal para investigar los mecanismos genéticos que subyacen a estas interacciones. Estos estudios han revelado que, si bien existen sistemas generales compartidos por las distintas especies del género, también hay particularidades que permiten una asociación única entre una cepa y un hospedador determinado. La creciente disponibilidad de datos genómicos de miles de aislamientos de *Pantoea*, facilitada por las nuevas tecnologías de secuenciación, representa una valiosa fuente de información para identificar tanto determinantes genéticos comunes como específicos en cada interacción. El uso de técnicas avanzadas de screening, como la secuenciación por inserción de transposones (TIS), permite identificar genes clave en estos procesos, lo que proporcionará una comprensión más profunda de la dinámica bacteria-planta. Esta información abrirá nuevas oportunidades para desarrollar estrategias que ayuden al manejo de enfermedades vegetales y promuevan el crecimiento vegetal, haciendo de *Pantoea* una herramienta clave en la agricultura sostenible y el manejo ambiental.

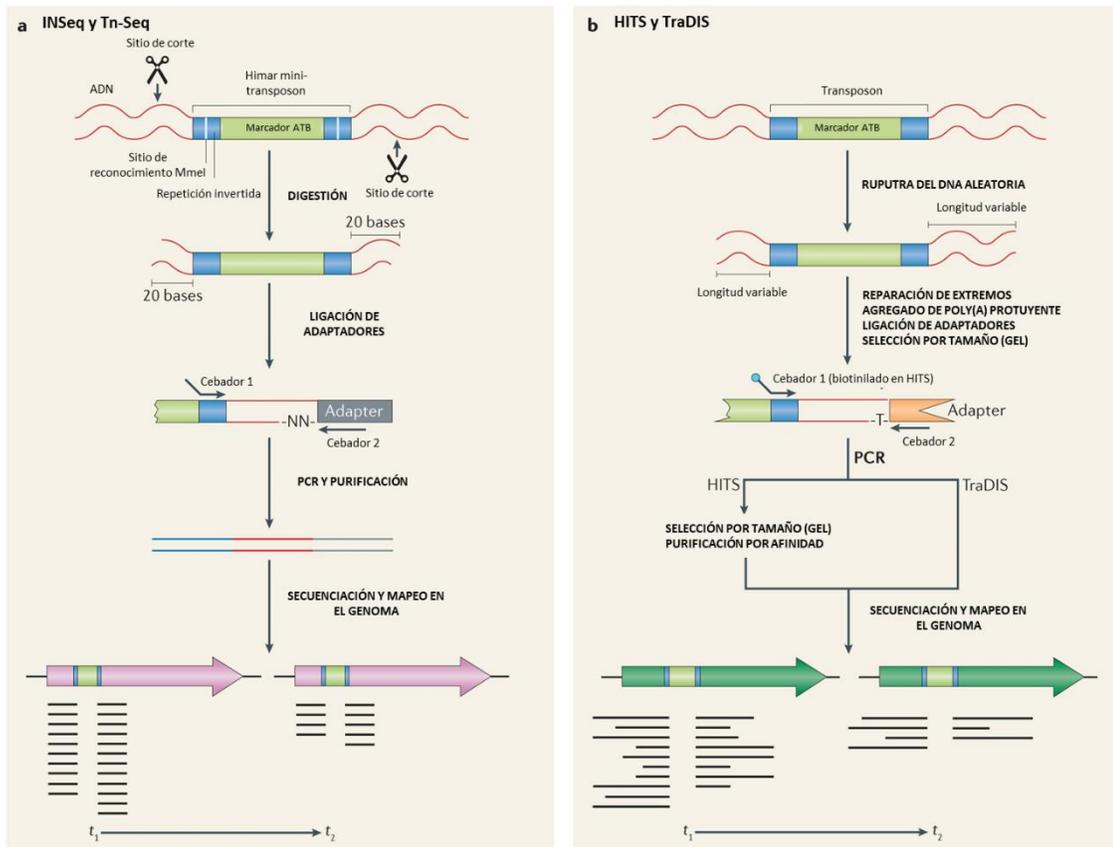
I.4. Metodologías de inserción de transposones para el estudio de interacciones simbióticas y nichos de difícil acceso

Los transposones son elementos genéticos que pueden moverse dentro de los genomas o entre ellos mediante mecanismos replicativos o de «cortar y pegar» mediados por una enzima denominada **transposasa**. Esta enzima reconoce las repeticiones invertidas en los extremos del transposón y la secuencia diana donde realiza una corte de doble hebra e inserta el transposón (Craig, 1997). Desempeñan papeles importantes, como en la evolución de la especiación y la resistencia a los antibióticos en los microorganismos (Alekhshun & Levy, 2007; Kidwell & Lisch, 2001). En la década de 1970, los investigadores propusieron el uso de transposones como herramienta para manipular los genes de bacterias, fagos y plásmidos de formas más sencillas a las que contaban en ese momento. Con la incorporación de los transposones a la lista de herramientas moleculares, se hizo mucho más fácil generar mutantes e identificar sus fenotipos mediante cruces transduccionales y ensayos de complementación. Desde entonces, la secuenciación de inserción (Kleckner et al., 1975) de transposones (TIS, por sus siglas en inglés *Transposon Insertion Sequencing*) se ha consolidado como una herramienta fundamental en la biología molecular para el análisis funcional de genomas. Esta técnica, basada en la inserción aleatoria de transposones en el ADN de un organismo, permite la identificación de genes esenciales y la elucidación de su papel en una variedad de procesos biológicos. Su versatilidad ha facilitado su aplicación en un amplio rango de estudios, incluyendo la patogenicidad bacteriana y la interacción de microorganismos con sus hospedadores. En esta sección analizaremos las principales variantes de la metodología TIS, especialmente Tn-Seq, comparando sus fortalezas y debilidades, y destacando su utilidad en la investigación de interacciones bacteria-planta.

I.4.a Técnicas de Secuenciación de Inserción de Transposones : Transposon - Sequencing (Tn-Seq) y Signature Tagged Mutagenesis (STM)

La secuenciación de inserción de transposones (TIS) representa un conjunto de técnicas poderosas para identificar funciones génicas críticas en organismos. Estas técnicas permiten la creación de bibliotecas de mutantes a través de la inserción aleatoria de transposones en el genoma, seguidas de

la identificación y cuantificación de las inserciones bajo diferentes condiciones experimentales. La potencialidad de esta herramienta molecular fue tal que en 2008 distintos grupos de trabajo desarrollaron técnicas diferentes basadas en el uso de transposones (Bentley et al., 2008; Hernandez et al., 2008; Holt et al., 2008). En la Figura I.6 se esquematiza las cuatro técnicas desarrolladas, separadas en dos grupos principales. El principio básico de los cuatro métodos implica la purificación del ADN genómico a partir de una población de mutantes, el corte del ADN (mediante enzimas específicas en las técnicas TnSeq e INSeq, o rompimiento aleatorio para las técnicas HITS y TraDIS), la unión de uno o más adaptadores a los fragmentos de ADN para facilitar la amplificación por PCR de los fragmentos que contienen secuencias de transposones y, finalmente, la secuenciación masiva en paralelo de los fragmentos amplificados para determinar la ubicación del transposón y la abundancia relativa de mutantes que contienen un transposón en este sitio. La principal diferencia entre estas técnicas es el transposón utilizado, Tnseq e INSeq emplean el transposón Himar I Mariner con una mutación en un nucleótido en las repeticiones invertidas terminales que introduce el sitio de reconocimiento de la enzima Mmel (Goodman et al., 2009a; van Opijnen et al., 2009a), mientras que las metodologías HITS y TraDIS utilizan transposones que carecen de sitios de restricción de tipo IIS dentro de su repetición invertida. Los trabajos que introdujeron estas técnicas fueron innovadores y demostraron la potencia y versatilidad de la tecnología al utilizar diferentes especies bacterianas para abordar distintos tipos de investigaciones. Trabajos con un gran impacto pueden encontrarse para cada una de estas técnicas, donde podemos destacar el uso del método de seguimiento de inserción de alto rendimiento mediante secuenciación profunda, HITS, se aplicó a una biblioteca de 75.000 mutantes de *Haemophilus influenzae* para estudiar un modelo de infección pulmonar en ratones e identificar genes de virulencia (Gawronski et al., 2009); TraDis (secuenciación de sitios de inserción dirigida por transposones) se utilizó para determinar que los genes esenciales en *Salmonella enterica* subsp. enterica serovar Typhi corresponden a aproximadamente del 8% del genoma (Langridge et al., 2009); el método de secuenciación por inserción (INSeq) se utilizó para determinar genes necesarios para la supervivencia en el colon en el simbionte humano *Bacteroides thetaiotaomicron* (Goodman et al., 2009a), y TnSeq se utilizó para identificar genes esenciales para el crecimiento en medio rico en el patógeno *S. pneumoniae* e identificar interacciones genéticas entre genes de interés y el resto del genoma (van Opijnen et al., 2009b).



Nature Reviews | Microbiology

Figura I.6. Adaptada de van Opijnen & Camilli (2013). Metodologías de Secuenciación de Inserción de Transposones (TIS). Se ilustra cuatro métodos de secuenciación de transposones, partiendo de un pool de ADN genómico de la biblioteca de inserción de transposones y terminando con la secuenciación de las uniones izquierda y derecha del transposón. El número de secuencias (lecturas) para cada unión puede diferir entre el inicio del experimento (t_1) y el final (t_2), después de haber realizado una selección en la biblioteca de mutantes en una condición e interés. En ambos ejemplos, la mutación debido a la inserción del transposón disminuye la aptitud durante el crecimiento en las condiciones probadas, indicado con menos lecturas al final del experimento respecto al principio. A) Los métodos Tn-seq (Secuenciación de Transposón) e INSeq (Secuenciación de inserción) son muy similares, pero INSeq incluye un paso de purificación en gel PAGE tras la ligación del adaptador y la PCR. B) Los métodos HITS (Seguimiento de inserción de alto rendimiento mediante secuenciación profunda) y TraDIS (secuenciación de sitios de inserción dirigida por transposones) son más similares entre sí que a Tn-seq e INSeq: tras el fraccionamiento del ADN, se reparan los extremos del ADN y se añade una cola de poli(A). Sin embargo, los métodos divergen después del paso PCR; en HITS, los productos PCR se someten a selección de tamaño (en un gel) y purificación por afinidad antes de la secuenciación, mientras que en TraDIS, los productos PCR se secuencian directamente.

La ventaja de las técnicas HITS y TraDIS reside en la posibilidad de utilizar cualquier transposón o elemento de inserción. Por otro lado, Tn-seq e INSeq tienen un protocolo de preparación de muestras más simple y generan un producto final de longitud precisa, mientras que las longitudes variables del ADN adyacente al transposón en las técnicas HITS y TraDIS (Figura I.6) puede dar lugar a potenciales sesgos en la etapa de amplificación por PCR.

I.4.b Metodología TnSeq: Secuenciación de Transposones

Esta técnica combina la mutagénesis por inserción de transposones con la secuenciación masiva de ADN, permitiendo el análisis simultáneo de miles de mutantes en una única muestra. La alta resolución de TnSeq la convierte en una herramienta ideal para identificar genes esenciales y para mapear rutas metabólicas críticas en condiciones específicas. La aplicación de una variante del TnSeq, RB-TnSeq (Random Barcoded TnSeq) en *Pseudomonas syringae*, un patógeno de plantas, permitió identificar genes esenciales para la virulencia y la capacidad de este organismo para evadir las defensas del huésped. Este estudio reveló que genes implicados en la biosíntesis de lipopolisacáridos y la resistencia a estreses oxidativos son fundamentales para la colonización efectiva de la planta (Helmann et al., 2019). TnSeq se destaca por su capacidad para analizar grandes bibliotecas de mutantes de manera simultánea, proporcionando una cobertura global del genoma con alta precisión. Su capacidad para identificar no solo genes esenciales, sino también aquellos que confieren ventajas adaptativas bajo condiciones específicas, es una de sus mayores fortalezas. En un estudio sobre *Xanthomonas campestris*, un patógeno vegetal, RB-TnSeq fue crucial para identificar genes que participan en la adaptación al ambiente apoplástico de la planta, un entorno que presenta desafíos únicos como la disponibilidad limitada de nutrientes y la presencia de compuestos antimicrobianos (Luneau et al., 2022). Este análisis permitió destacar genes involucrados en la biosíntesis de exopolisacáridos, cruciales para la formación de biopelículas protectoras y la promoción de la virulencia.

Sin embargo, TnSeq puede no ser la opción más adecuada para todos los contextos experimentales. En situaciones donde la interacción directa con un hospedador y la competencia entre diferentes mutantes es esencial para estudiar un nicho específico, como en estudios de infección *in vivo*, pueden poner en jaque la competencia de esta técnica si no se tienen las consideraciones acordes. Estas pruebas *in vivo* son muy valiosas para descubrir mecanismos de virulencia desconocidos y para estudiar el metabolismo de los patógenos o bacterias endofíticas. Una limitación para el uso de métodos de secuenciación de transposones *in vivo* es que muchas bacterias mueren o se eliminan estocásticamente durante el establecimiento de la infección. Estos **cuellos de botella** ocasionan la dificultad en la determinación de si los mutantes desaparecen debido a su falta de aptitud o por razones aleatorias no relacionadas a sus capacidades de infección. Cuando este escenario se presenta en nuestro sistema en estudio es necesario ajustar el número de mutantes diferentes presentes en el inóculo (es decir, la complejidad de la población inoculada) para que sea inferior al cuello de botella. Por ejemplo, si el modelo de infección permite que solo 10^4 bacterias colonicen el huésped independientemente del tamaño del inóculo (el tamaño del cuello de botella es igual a 10^4), se debe utilizar un inóculo que contenga menos de 10^4 mutantes únicos (lo recomendado es contar con en el inóculo entre 10-100 células por mutante) (van Opijnen & Camilli, 2013). Caracterizar entonces el cuello de botella y estimar su tamaño se torna crucial para una correcta implementación de metodologías como TnSeq. Un enfoque utilizado para esto se basa en el análisis independiente de un conjunto de inserciones neutras que no deberían desaparecer de la biblioteca. Así es posible conocer el número efectivo de células que pueden infectar el hospedador y corregirlos en los cálculos de aptitud posteriores, o rediseñar la estrategia experimental.

I.4.b.1 TnSeq y el estudio de interacciones bacteria-planta

La aplicación de técnicas TIS, y en particular de TnSeq, ha proporcionado nuevas perspectivas sobre las interacciones entre bacterias y plantas, un área de investigación crucial para el desarrollo de estrategias de control de enfermedades agrícolas. La patogénesis bacteriana en plantas es un proceso complejo que involucra múltiples factores genéticos y ambientales. La capacidad de TnSeq para identificar

genes esenciales bajo condiciones de estrés específicas ha sido clave para comprender cómo los patógenos vegetales manipulan su entorno y evaden las defensas del huésped. Trabajos realizados con esta metodología en *Pantoea stewartii* ssp. *stewartii*, agente causante de la enfermedad de Stewart del maíz, permitió determinar genes involucrados en la supervivencia en el xilema de maíz, incluidas dos proteínas de la membrana externa, la proteasa Lon y dos factores de transcripción asociados a *quorum sensing* RcsA y LrhA (Duong et al., 2018). En *Dickeya dadantii*, causante de la podredumbre blanda bacteriana en muchas especies vegetales, se identificaron cerca de 100 genes que mostraban una disminución o un aumento de la aptitud en el crecimiento en plantas de achicoria. Esta búsqueda a gran escala permitió identificar la ruta de biosíntesis del monofosfato de uridina y la ruta de biosíntesis de la purina eran esenciales para la supervivencia de *D. dadantii* en la planta (Royet et al., 2019).

Las herramientas TIS son invaluable en el estudio de la genética bacteriana y su interacción con las plantas. Cada técnica tiene sus fortalezas y debilidades, lo que las hace adecuadas para diferentes tipos de estudio. La aplicación de estas técnicas en la investigación de interacciones bacteria-planta ha avanzado significativamente nuestra comprensión de los mecanismos de patogenicidad y resistencia, lo que contribuye al desarrollo de nuevas estrategias para la protección de cultivos y la seguridad alimentaria global.

I.5. Nuevas herramientas en el estudio de la ecología microbiana: Machine Learning para la predicción de fenotipos de interés y búsqueda *in silico* de características bacterianas responsables o asociadas

Para entender la biología a nivel sistémico, es necesario examinar la estructura y la dinámica de la función celular y del organismo, en lugar de las características de partes aisladas de una célula u organismo. El desarrollo de técnicas de alto rendimiento y de la tecnología asociada para gestionar y extraer datos a gran escala ha supuesto una revolución de la biología de sistemas en la última década (Chuang et al., 2010). En este contexto, las tecnologías “ómicas”, como la genómica, transcriptómica, proteómica, metabolómica, epigenómica y metagenómica, permiten recopilar de forma rápida y paralela cantidades masivas de diferentes tipos de datos para el mismo sistema modelo. Contar con herramientas informáticas para el análisis integrador de datos a escala ómica son cruciales para hacer frente a su complejidad inherente. En particular, los algoritmos de aprendizaje automático conocidos como *Machine Learning* (ML), un subconjunto de la inteligencia artificial (IA), desempeñan un papel central en el proceso de extracción de conocimiento (Larrañaga et al., 2006; Verikas et al., 2011). Un área creciente de aplicación de estos algoritmos es en el reconocimiento supervisado de patrones en conjuntos de datos que pueden utilizarse para entrenar modelos de clasificación que permite separar muestras de diferentes clases basándose en variables, y para estimar qué variables fueron importantes para esta clasificación.

Los algoritmos clásicos de ML ya han hecho contribuciones sustanciales en campos como la bioinformática, utilizando la regresión y la clasificación para modelar diagnósticos. Los enfoques tradicionales implican la elaboración de características para mejorar la interpretabilidad, con modelos que optimizan estas características en un proceso secuencial (Jordan & Mitchell, 2015). Con la llegada del aprendizaje profundo, nuevos algoritmos son capaces de aprender de forma autónoma características complejas (Wu & Gadsden, 2023). La salud y saneamiento es una de las áreas que más ha sido impactada y beneficiada por la aplicación de estos algoritmos. En la última década, se han ideado multitud de nuevos sistemas de ML para facilitar el proceso de toma de decisiones en diversos contextos médicos, donde las técnicas clásicas de ML se han utilizado ampliamente en tareas de clasificación y regresión. Sin embargo,

estos enfoques de ML convencionales requieren una participación humana sustancial en la curación de las variables informativas a utilizar, conformadas por atributos distintivos o características cuantificables. Modelos como las máquinas de vectores soporte (Support Vector Machines SVM), Random Forest y Logistic regression han demostrado una notable eficacia en aplicaciones sanitarias, especialmente en tareas de clasificación.

El uso de algoritmos como Random Forest y la Regresión Logística ha demostrado ser particularmente útil en este contexto, debido a su capacidad para manejar la complejidad y la naturaleza multidimensional de los datos. Random Forest es un algoritmo de aprendizaje supervisado que opera creando múltiples árboles de decisión durante el entrenamiento y obteniendo la predicción final como una agregación (generalmente por voto mayoritario en clasificación, o promedio en regresión) de las predicciones de estos árboles individuales. Una de las ventajas principales de Random Forest es su capacidad para manejar grandes conjuntos de datos con muchas variables de entrada sin sobreajustarse, lo que lo hace ideal para estudios genómicos y metagenómicos donde la cantidad de atributos puede ser abrumadora (López et al., 2023). Además, Random Forest proporciona una medida de importancia de las variables, lo que permite a los investigadores identificar qué características son más relevantes para la predicción del fenotipo. Esto es especialmente valioso en ecología microbiana, donde entender qué factores microbianos influyen en los fenotipos de las plantas puede conducir a nuevos descubrimientos sobre las interacciones planta-microbio. Sin embargo, Random Forest también tiene desventajas. El modelo puede volverse computacionalmente costoso con datos extremadamente grandes, y aunque es robusto contra el sobreajuste, puede ser menos interpretable en comparación con modelos más simples, lo que puede complicar la interpretación biológica de los resultados.

En contraste a este modelo descrito se encuentra *Logistic Regression*, un algoritmo de clasificación que modela la probabilidad de un evento binario (por ejemplo, presencia o ausencia de un fenotipo) en función de una o más variables independientes. Este modelo es altamente interpretable, ya que los coeficientes obtenidos pueden ser directamente relacionados con el logaritmo de las probabilidades de los eventos, proporcionando una visión clara de cómo cada variable afecta la probabilidad del resultado. Una ventaja importante de la Regresión Logística es su simplicidad y la facilidad con la que se puede interpretar el impacto de cada variable. Esto es particularmente útil cuando se busca entender la influencia de un número limitado de factores microbianos en los fenotipos de plantas. Además, debido a su naturaleza paramétrica, requiere menos datos para generar un modelo robusto en comparación con Random Forest. Sin embargo, la Regresión Logística también presenta limitaciones. A medida que el número de variables independientes aumenta, el modelo puede volverse inestable y propenso al sobreajuste. Además, no es tan eficaz como Random Forest para manejar interacciones complejas entre variables, lo que limita su uso en estudios donde las relaciones no lineales entre los microbios y los fenotipos de plantas son comunes.

Ambos algoritmos tienen aplicaciones valiosas en la ecología microbiana y la interacción con plantas. Random Forest es más adecuado cuando se trata de grandes volúmenes de datos con muchas variables y cuando se requiere la identificación de características importantes. En contraste, la Regresión Logística es preferible cuando se necesita un modelo interpretable y cuando el número de variables es relativamente pequeño. El uso combinado de ambos enfoques puede proporcionar una estrategia poderosa para la predicción de fenotipos y la identificación de factores clave en la ecología microbiana. La aplicación de Machine Learning en estudios de ecología microbiana no solo facilita la predicción precisa de fenotipos, sino que también contribuye a una comprensión más profunda de las interacciones entre plantas y microorganismos, lo que podría tener implicaciones significativas para la agricultura sostenible y la biotecnología vegetal. En un contexto actual en el que la cantidad de datos genómicos se produce a

una velocidad mayor de la que se analiza, estas herramientas bioinformáticas se vuelven invaluable para explorar la información genética de una forma eficiente, veloz y económica.

CONSTRUCCIÓN DE LA HIPÓTESIS DE TRABAJO

Los eucariotas multicelulares, incluidas las plantas, son hoy considerados como holobiontes dinámicos, esto es, como sistemas multipartitos conformados por la propia entidad eucariota y por el conjunto de microorganismos asociados a ella (su microbioma) (Guerrero et al., 2013; Salvucci, 2014). Ese conjunto interactuante de biontes da lugar a un sistema complejo que es el que, colectivamente, evoluciona y se adapta al medio circundante. En tal contexto, y en relación a este trabajo de tesis, son cada vez más numerosos los estudios que enfocan el análisis de la interacción vegetal-microbioma, como vía formal hacia la comprensión del aporte de cada componente al comportamiento del conjunto.

Conocemos hoy que las semillas de las plantas son, ellas mismas holobiontes, portadores (tanto fuera como en su interior) de una vasta diversidad de microorganismos que, junto a los presentes en el suelo, serán los que pueblen el tejido vegetal en las primeras instancias post-germinación (Truyens et al., 2014). El interés principal de este trabajo de tesis se centró en avanzar en el conocimiento de los genes y funciones que las bacterias asociadas a plantas necesitan para colonizar raíces, el órgano en mayor contacto con el suelo y puerta principal de entrada de quienes conformarán finalmente el microbioma vegetal maduro. Con ese propósito, y dada la gran diversidad de bacterias que pueblan las plantas, decidimos estudiar un género ubico en tejidos vegetales, de modo que el conocimiento que se genere tenga impacto en un número amplio de sistemas. Para satisfacer esta condición, como sistema modelo estudiamos —con distintas herramientas— la interacción de una cepa del género *Pantoea* con raíces de plantas de alfalfa (una planta muy estudiada en su asociación con diversos microorganismos).

Las principales hipótesis subyacentes al presente trabajo de tesis fueron las siguientes:

- La aproximación experimental TnSeq, usada en el estudio de otras interacciones bacteria-planta (Cain et al., 2020), será una herramienta experimental potente para identificar genes y describir los procesos de colonización/invasión asociados a los distintos estadios (temprano y tardíos) de la interacción *Pantoea* - alfalfa.
- El uso de métodos de clasificación por aprendizaje automático (*Machine Learning*) supervisado junto a la información disponible sobre fenotipo de colonización de plantas para miembros de un género de bacterias particular (como *Pantoea* en nuestro caso) puede ser una herramienta eficiente y útil para la predicción de la capacidad de asociarse a plantas (o de no asociarse) en aislamientos nuevos aun no caracterizados frente a hospedadores vegetales.

De las hipótesis anteriores se desprendieron las siguientes **PREGUNTAS** que fueron motivo de análisis en el presente trabajo:

- ¿Cómo es la dinámica temporal de la asociación inicial (rizosférica) de *Pantoea* con una raíz?
- ¿Cómo comparan los tiempos de colonización y títulos que se alcanzan en la rizósfera con los observados en otras bacterias y plantas?
- ¿Cómo es la dinámica de la colonización endofítica? ¿Cuáles son sus tiempos? ¿Cómo comparan esos tiempos con los de la colonización rizosférica que la precede? ¿Cuál es el número de genotipos distintos que entran a la endosfera hasta que la misma se puebla por completo? ¿La magnitud de esa diversidad cambia con la edad de la planta?
- ¿Cuántos son los genes asociados a la colonización de la rizósfera y cuántos los asociados a poblar la endosfera? ¿Qué proporción del genoma representa cada uno? ¿Cuál es la predominancia de funciones en cada uno de los grupos de genes anteriores? ¿Son comparables? ¿Existe una base genética conservada para interactuar con raíces? ¿si existen evidencias de conservación, es mayor entre los genes de la interacción temprana o la tardía?
- Usando métodos de aprendizaje automático supervisado ¿puede inferirse con un grado de precisión aceptable (tasa de falsos positivos y falsos negativos) si un aislamiento desconocido de *Pantoea* tiene (o no) capacidad de asociarse a plantas?

OBJETIVO GENERAL

Caracterización fenotípica y genómica de la colonización de raíces de *Medicago sativa* (alfalfa) por una cepa de *Pantoea*.

OBJETIVOS ESPECÍFICOS

- # 1. Construcción de una biblioteca de mutantes transposicionales con un transposón tipo *mariner* en una cepa de *Pantoea* aislada de semillas de alfalfa y con capacidad de asociarse a la planta. Para ello, diseñar y construir una herramienta de mutagénesis que exprese la transposasa en *Pantoea*.
- # 2. Secuenciación genómica de la cepa de *Pantoea* para la asignación de especie a la misma, y para su uso en ensayos TnSeq.
- # 3. Experimentos de colonización rizosférica y endofítica utilizando la biblioteca de mutantes del punto 1 previo. Identificación por la metodología TnSeq de los genes de la cepa de *Pantoea* involucrados en la colonización de la rizósfera y de la endosfera de raíces de alfalfa. Caracterización funcional de los genes identificados.
- # 4. Exploración del uso de algoritmos clasificatorios de aprendizaje automático supervisado para la predicción de capacidades de asociación a plantas en bacterias del género *Pantoea*.



CAPÍTULO II

MATERIALES Y MÉTODOS



II.1. Técnicas microbiológicas

II.1.a Cepas bacterianas usadas en este trabajo

Las cepas bacterianas usadas en este trabajo se muestran en la Tabla II.1 junto con sus características más importantes.

Tabla II.1 Cepas bacterianas usadas en este trabajo y sus características

Cepas	Propiedades	Referencias
<i>Ensifer meliloti</i> 2011	Sm ^R , Nod ⁺ Fix ⁺ en alfalfa	J. Dénarié., France
<i>Escherichia coli</i> DH5α	recA, vlacU169, OdlacZDM15	Bethesda Res. Lab.
<i>Escherichia coli</i> S 17-1	E. coli 294 RP4-2-Tc:: Mu-Km: :Tn7	(Simon et al., 1983a)
<i>Escherichia coli</i> S 17-1 λ-pir	Tp ^R Sm ^R recA, thi, pro, hsdR ^M +RP4: 2-Tc:Mu: Km Tn7 λpir.	(Simon et al., 1983a)
<i>Pantoea sp.</i> LPU12 NaIR	NaI ^R , endófito de semilla de alfalfa	Este trabajo
<i>Pantoea agglomerans</i> (Pan12)	Endófito de semilla de alfalfa, cepa salvaje	(López et al., 2018)
<i>Pantoea agglomerans</i> (Pan12) pK18::eGFP	Endófito de semilla de alfalfa portadora del plásmido pK18::eGFP	Este trabajo
<i>Pantoea agglomerans</i> (Pan12) pHc60	Endófito de semilla de alfalfa portadora del plásmido pHc60	Este trabajo

II.1.b Plásmidos usados en este trabajo

Los plásmidos utilizados en este trabajo se muestran en la Tabla II.2 junto con sus características más importantes.

Tabla II.2 Plásmidos usados en este trabajo y sus características

Plásmido	Propiedades	Referencia
pSAM_RI	Amp ^R Kan ^R ; pSAM_Km con el protomotr del gen rpoD de <i>B. thetatiotamicron</i> reemplazado con la región promotra del gen rpoD de <i>R. leguminosarum</i> 3841	(Perry & Yost, 2014)
pSAM_placOc_II_1	Amp ^R Kan ^R ; pSAM_RI con el protomotr del gen rpoD de <i>R. leguminosarum</i> reemplazado con la región promotra del gen lacOc de <i>E.coli</i>	Este trabajo
pK18::eGFP	Kan ^R ; derivado de pK18 con el gen de la proteína eGFP en sitio SmaI.	Luchetti Abril, 2019
pHc60	Tc ^R , contiene gende la proteína mCherry	Cedido por la Dra. Elizaveta Krol

II.1.c Medios de cultivo y condiciones de crecimiento

Los cultivos para las cepas del género *Ensifer* y *Rhizobium* fueron realizados en medio completo Triptona-Extracto de levadura (TY)(BERINGER, 1974). Las restantes cepas bacterianas utilizadas en este trabajo fueron crecidas en medio completo Lysogeny Broth (LB) (Sambrook, 1989). El medio mínimo M9 suplementado con glucosa (Harwood, 1990) fue utilizado para evaluar el crecimiento de la biblioteca de mutantes de *Pantoea sp.* LPU12 en búsqueda de mutantes auxótrofos.

Tabla II.3 Medios de cultivos

MEDIOS COMPLEJOS		
Tritpona-Extracto de levadura levadura (TY) (Beringer, 1974)	CaCl ₂ .H ₂ O	0,7 g/L
	Triptona	5 g/L
	Extracto de levadura	3 g/L
Luria-Bertani (LB) (Sambrook, 1989)	Glucosa	1 g/L
	NaCl	5 g/L
	Triptona	10 g/L
	Extracto de levadura	5 g/L
MEDIOS MÍNIMOS		
M9 (Harwood, 1990)	M9 sales	200 mL
	MgSO ₄ 1M	2 mL
	Glucosa 20%	20 mL
	CaCl ₂ 1M	0,1 mL
M9 sales (Harwood, 1990)	Na ₂ HPO ₄ .7H ₂ O	64 g/L
	KH ₂ PO ₄	15 g/L
	NaCl	2,5 g/L
	NH ₄ Cl	5,0 g/L

Para la preparación de medios sólidos se utilizó en todos los casos 15 g/L de agar.

Los medios de cultivo fueron esterilizados en autoclave a 120 °C durante 20 minutos.

La temperatura de incubación para el crecimiento de las cepas de *E. coli* fue de 37 °C; las restantes cepas bacterianas se crecieron a 28 °C.

II.1.d Medios de conservación

La conservación a largo plazo de las cepas bacterianas fue realizada por congelamiento de alícuotas de cultivos crecidos hasta fase logarítmica tardía suplementadas con glicerol estéril hasta alcanzar una concentración final de 25 % V/V y 50 % V/V, según fueron preservadas a -80 °C o -20 °C, respectivamente.

II.1.e Antibióticos

Cuando fue requerido, los medios de cultivos fueron suplementados con antibióticos a partir de soluciones concentradas de los mismos preparadas en agua desionizada, salvo las soluciones de tetraciclina que fueron preparadas con metanol. Las soluciones preparadas con agua fueron esterilizadas por filtración utilizando membranas de nitrocelulosa de 0,2 µm de tamaño de poro. Las concentraciones finales utilizadas de cada antibiótico para cada género bacteriano se detallan en la Tabla II.4.

Tabla II.4 Antibióticos utilizados en este trabajo (concentraciones finales empleadas en los medios en los que crecieron los cultivos).

Antibiótico	<i>S. meliloti</i>	<i>E. coli</i>	<i>Pantoea sp.</i>
Kanamicina (km)	-	50 µg/ml	50 µg/ml
Ampicilina (Amp)	-	200 µg/ml	200 µg/ml
Tetraciclina (Tc)	5 µg/ml	10 µg/ml	
Estreptomicina (Sm)	400 µg/ml	100 µg/ml	-
Neomicina (Nm)	20 µg/ml	-	-
Ac. Nalidíxico (Nal)	10 µg/ml	30 µg/ml	30 µg/ml

II.2. Procedimientos microbiológicos

II.2.a Curvas de crecimiento

Las curvas de crecimiento para caracterizar el comportamiento de las cepas endófitas de alfalfa en medio rico LB y en exudados de raíz fueron realizadas en pocillos de placas *multiwell*. Para esto se partió de cultivos saturados (crecidos “*overnight*”) iniciados a una DO_{600} de 0,1 y crecidos hasta alcanzar la fase logarítmica tardía. Las lecturas de DO_{600} fueron tomadas cada 15 minutos a 28 °C empleando un lector de microplacas multifuncional (Tecan Infinite M200 PRO).

II.2.b Recuento de bacterias en medio líquido

Para determinar las unidades formadoras de colonias (UFC) en medio líquido se utilizó el método de recuento de gotas en placa de Petri (Herigstad et al., 2001). Para ello, se realizaron diluciones sucesivas en solución fisiológica estéril del cultivo cuyo número de UFC se quiso determinar. Para cada dilución fueron sembradas 8 gotas de 10 µl cada una en el medio sólido correspondiente, conteniendo o no antibiótico, según el caso. Luego la placa resultante fue incubada durante 24 o 48 hs a 28 °C, dependiendo el microorganismo de modo de obtener microcolonias aisladas, para finalmente contar las microcolonias en cada gota. En las diluciones donde se visualizó entre 1 y 20 colonias se realizó un promedio del número de colonias/gota y el mismo se refirió al número de UFC/ml de cultivo original.

II.2.c Conjugaciones biparentales

Las conjugaciones fueron realizadas utilizando la técnica de Simon et al., 1983 con modificaciones menores como se detalla a continuación. En microtubos de polipropileno de 1,5ml fueron mezclados 0,75 ml de cultivo de la cepa dadora ($DO_{600} = 0,4-0,5$) con 0,75 ml de cultivo de la cepa receptora a igual densidad óptica. La mezcla fue centrifugada a 2500 rpm durante 8 minutos y el sedimento fue resuspendido suavemente en aproximadamente 100 µl del medio de cultivo rico correspondiente. La suspensión fue colocada en una caja de Petri con medio rico sólido sin antibióticos e incubada a 28 °C durante 16 horas. Finalmente, la mezcla de bacterias levantada del medio sólido fue resuspendida en medio líquido rico y plaqueada en el mismo medio sólido suplementado con el/los antibiótico/s para la selección de transconjugantes. En todas las conjugaciones fueron realizados controles con el sólo agregado de la cepa dadora o la receptora.

II.2.d Preparación de cepas electrocompetentes. Electrotransformación de *E.coli* y *Pantoea* sp.

Para la preparación de bacterias electrocompetentes se utilizó a la técnica de Lin Tung & Chow, 1995 con modificaciones menores. Para ello, 500 ml de medio LB sin NaCl fueron inoculados con 1 ml de un cultivo de la cepa de interés crecida desde el día anterior. El cultivo inoculado fue incubado a 37 °C o 28 °C según correspondió al género bacteriano (ver sección II.1.c) a 180 rpm hasta alcanzar una $DO_{600} = 0,8$. Las células fueron centrifugadas a 4000 x g por 15 minutos a 4 °C, y luego lavadas 2 veces con glicerol 10% a 4°C. Finalmente, las células fueron resuspendidas suavemente en glicerol 10% a 4 °C y fraccionadas en alícuotas de 100 µl previo a su congelación a -80 °C.

Las transformaciones de las células fueron realizadas usando un equipo Gene Pulser (Bio-Rad) en cubetas de 0,2 cm de luz entre los electrodos bajo las condiciones recomendadas por el fabricante (25 µF; 200 µA; 2,5 kV). Luego de la transformación, a la mezcla le fue adicionado inmediatamente 1 ml de medio LB, incubándola luego 1 hora a 37 °C o 28 °C para permitir que la información contenida en el ADN introducido se expresara. Posterior a la incubación, las bacterias fueron plaqueadas en medio LB sólido selectivo e incubadas a 37 °C o 28 °C para el crecimiento de los transformantes.

II.3. Técnicas de biología molecular

II.3.a Preparación de ADN plasmídico en pequeña escala (miniprep)

Los plásmidos de *E. coli* fueron preparados siguiendo la técnica de lisis alcalina descrita por Sambrook (1989). Luego de obtener el sobrenadante de la lisis, la purificación fue realizada con sílica (Sigma 112945-52-5) según se indica a continuación. Al extracto de la lisis alcalina y la precipitación ácida le fueron agregados 10 µl de suspensión de sílica (ver preparación al pie de esta sección). Luego de mezclar, la sílica se sedimentó con un pulso de centrifuga y la misma se lavó con 500 µl de solución de lavado (ver composición al pie de esta sección). La mezcla de lavado fue centrifugada 10 seg a máxima velocidad, el sobrenadante descartado, y la sílica secada a 37 °C. Para finalizar, la sílica fue resuspendida en 30 µl de agua destilada estéril e incubada a 60 °C 5 minutos para inactivar ADNasas que pudieran estar presente. Se centrifugó a máxima velocidad por 3 minutos y se colectó el sobrenadante que contenía el ADN.

Suspensión de sílica: 15 ml de agua destilada fueron mezclados con 2 gr de sílica (Sigma 112945-52-5). La misma fue lavada 3 veces con agua destilada, sedimentándola en cada paso por centrifugación a 2000 g por 2 minutos. El producto lavado fue finalmente resuspendido en dos volúmenes de agua milliQ.

Solución de lavado: 50 mM NaCl; 10 mM Tris pH 7,5; 2,5 mM EDTA; 50% (V/V) etanol.

II.3.b Extracción de ADN total

Las extracciones de ADN de las muestras de los experimentos de TnSeq fueron realizadas utilizando el *kit* comercial NucleoSpin Microbial DNA (Macherey-Nagel) siguiendo las instrucciones del fabricante.

II.3.c Electroforesis de ADN

II.3.c.1 Separación de fragmentos de restricción y productos de amplificación por PCR mediante electroforesis en geles de agarosa

La separación de fragmentos de ADN mediante electroforesis horizontal en geles de agarosa (0,8-2 % P/V) fue realizada usando TAE (40 mM Tris, 20 mM acetato y 1 mM EDTA) en el gel y como tampón de corrida. El voltaje de corrida varió entre 5 y 12 volts/cm. Para la siembra de las muestras se utilizó una

solución de sacarosa 40% y azul de bromofenol 0,25%, preparada en una concentración 5 veces concentrada (5X).

II.3.c.2 Revelado de geles, fotografía

El revelado se realizó agregando 0,5 µg/ml de bromuro de etidio a los geles de agarosa, lo que permitió la visualización del ADN por transiluminación con luz ultravioleta de $\lambda = 260$ nm. Los geles fueron fotografiados con una cámara digital bajo iluminación ultravioleta para registrar los resultados.

II.3.c.3 Estimación de pesos moleculares de fragmentos de restricción por electroforesis en gel de agarosa

Para identificar el tamaño de los fragmentos de restricción se utilizaron como patrones de peso molecular ADN del fago lambda digerido con la endonucleasa de restricción *HindIII* (fragmentos de 23.1330, 9.416, 6.557, 4.361, 2.322, 2.0227 y 564 pb) y los marcadores de peso molecular comerciales a intervalos de 100 bp DNA ladder y GeneRuler 100 bp Plus *DNA-ladder*.

II.3.c.4 Extracción de fragmentos de ADN de geles de agarosa

Las bandas de interés en geles de agarosa fueron cortadas y colocadas en tubos de polipropileno. A las mismas se adicionaron 5 volúmenes de NaI 6 M y se calentó a 55 °C durante 5 minutos agitando esporádicamente. Luego de dejar enfriar la mezcla a temperatura ambiente, se agregó 1/10 de volumen de acetato de sodio 3 M pH 5,4. Para la purificación se agregó 10 µl de suspensión de sílica y se prosiguió como se describe para la purificación de plásmidos en la sección II.3.a.

II.3.d Tratamientos enzimáticos del ADN

Las reacciones enzimáticas del ADN fueron realizadas en las condiciones óptimas de la enzima, siguiendo las indicaciones de temperatura y tampón propuestas por los fabricantes (Invitrogen, Fermentas, Promega Corp. y New England Biolabs), y las recomendaciones descritas en Sambrook, (1989). Cuando fue necesario, luego de los tratamientos se procedió a la inactivación de las enzimas por desnaturalización térmica calentando las mezclas postdigestión a 65 °C durante 30 minutos.

II.3.e Reacción en cadena de la polimerasa (PCR) y reacción en cadena de la polimerasa en tiempo real (RT-PCR)

Los ensayos de PCR fueron realizados utilizando los cicladores térmicos Mastercycler Nexus X2 (Eppendorf) y Biometra TRIO serie (Analytic Jena). Se utilizaron las ADN polimerasas ADN *Taq* polimerasa de *PB-L Productos Bio-Lógicos* (temperatura óptima de elongación 72 °C) o de New England Biolabs (temperatura óptima de elongación 68 °C), siguiendo las instrucciones de los respectivos fabricantes. Los cebadores fueron utilizados en una concentración de 10 µM para PCR específicas, con volúmenes de 10 µl en cada mezcla de reacción. Los cebadores fueron sintetizados por GenBiotech, Sigma o Integrated DNA Technologies. Finalizada la reacción de PCR, 5 o 10 µl del producto fueron sembrados en un gel de agarosa 1,5% P/V conteniendo 0,5 µg/ml de bromuro de etidio (ver sección precedente II.3.c.1).

Los moldes de ADN para las amplificaciones por PCR fueron obtenidos siguiendo alguna de las siguientes metodologías:

- a) Para amplificaciones específicas sobre ADN genómico: células provenientes de cultivos o sólidos fueron resuspendidas en 100 µl de agua bidestilada estéril y hervidas por cinco minutos. La

suspensión luego fue centrifugada a máxima velocidad por 4 minutos y 5 μ l del sobrenadante se usaron como ADN molde. Alternativamente, se realizaron preparaciones de ADN total de acuerdo a la metodología que se describió en la sección II.3.b.

- b) Para amplificaciones de secuencias específicas de ADN contenidas en plásmidos: el ADN plasmídico se preparó según se detalla en II.3.a, y se realizó una dilución final de 1/200.

Las reacciones de PCR en tiempo real cuantitativas para el seguimiento de las muestras de ADN de los experimentos TnSeq se realizaron en el ciclador termico en tiempo real qTOWER (Analytik Jena) utilizando el kit "Taqyon™ No ROX SYBR 2X MasterMix blue dTTP" de Eurogentec. El volumen final de la mezcla de reacción fue de 20 μ l, la concentración final de oligonucleótido fue de 0,5 μ M y se utilizó 5 μ l de molde. Los oligonucleótidos utilizados en este trabajo se detallan en la Tabla II.5.

Tabla II.5 Secuencia de oligonucleótidos utilizados en este trabajo

Nombre	Secuencia	Referencia en este trabajo	Anotaciones
Oligo_promotor_BamHI	GATCGCTTACACTTTATCGTTCCGGCTCGTATGTTGTGT GGAATTATTGTTAGACAATAATTTACACAGGAAACAGC	Capítulo IV	Cadena superior del promotor sintético generado para el vector pSAM_placO ^c
Oligo_promotor_Ndel	CGAAATGTGAAATACGAAGGCCGAGCATACAACACACC TTAATAACAATCTGTTATTAAGTGTGCCTTTGTCGAT	Capítulo IV	Cadena inferior del promotor sintético generado para el vector pSAM_placO ^c
Adapter-Bottom-P	P-GCCCCGTAAACCACGTCCTTT	Capítulo V	Cadena inferior del adaptador fosforilado en la posición 5'
Adapter-Top	AAAGGACGTGGTTTACGGGGCNN	Capítulo V	Cadena superior del adaptador con dos bases aleatorias en la posición 3' que conforman las doce combinaciones posibles (AG-3', AC-3', AT-3', GA-3', GC-3', GT-3', CA-3', CG-3', CT-3', TA-3', TG-3', TC-3')
IR_Tag-n	ANNNNCGGGGACTTATCATCCAACCTG	Capítulo V	Secuencia complementaria a secuencia IR del transposon con firma única Tag-n (NNNN)* en posición 5'
Adapter-Top-PCR	AAAGGACGTGGTTTACGGGGC	Capítulo V	Cadena superior del adaptador sin bases aleatorias
TnSeq-sintético-Top	AAAGGACGTGGTTTACGGGGCGCGTCCTCGGTACCTA ACAGGTTGGATGATAAGTCCCCG	Capítulo V	Secuencia sintética del fragmento final obtenido luego del procesamiento del ADN en los ensayos TnSeq
TnSeq-sintético-Bottom	CGGGGACTTATCATCCAACCTGTTAGGTACCGAGGAC GCGCCCCGTAAACCACGTCCTTT	Capítulo V	Secuencia sintética del fragmento final obtenido luego del procesamiento del ADN en los ensayos TnSeq

* Los nucleótidos NNNN corresponden a una secuencia única utilizada como identificador para etiquetar cada condición del ensayo TnSeq realizado en el capítulo V. Las secuencias firma (Tag-n) se describen en la tabla

Tabla II.6 Secuencia de oligonucleótidos firma utilizada como identificador para etiquetar cada condición del ensayo TnSeq descrito en el capítulo V.

Condición y réplica	Nombre de la muestra	Cebador utilizado	Identificador (Tag)
Rizósfera - Réplica 1	Rizósfera - R1	IR_Tag-1	GTAT
Rizósfera - Réplica 2	Rizósfera - R2	IR_Tag-2	CAGA
Rizósfera - Réplica 3	Rizósfera - R3	IR_Tag-3	CTTT
Rizósfera - Réplica 4	Rizósfera - R4	IR_Tag-4	GTAA
Endósfera temprana 24 h p.i.- Réplica 1	Endo_E - R1	IR_Tag-5	GTAC
Endósfera temprana 24 h p.i.- Réplica 2	Endo_E - R2	IR_Tag-6	CTTC
Endósfera temprana 24 h p.i.- Réplica 3	Endo_E - R3	IR_Tag-7	ATTA
Endósfera temprana 24 h p.i.- Réplica 4	Endo_E - R4	IR_Tag-8	TCTT
Medio mínimo M9 - Réplica 1	Mínimo_M9 -R1	IR_Tag-9	AAAT
Medio mínimo M9 - Réplica 2	Mínimo_M9 -R2	IR_Tag-10	TTTT
Medio mínimo M9 - Réplica 3	Mínimo_M9 -R3	IR_Tag-11	AGAC
Medio mínimo M9 - Réplica 4	Mínimo_M9 -R4	IR_Tag-12	ATCA
Muestra de entrada - Biblioteca - Réplica 1	Inóculo - R1	IR_Tag-13	GGAA
Muestra de entrada - Biblioteca - Réplica 2	Inóculo_E - R2	IR_Tag-14	AATA
Muestra de entrada - Biblioteca - Réplica 3	Inóculo - R3	IR_Tag-15	CTCA
Muestra de entrada - Biblioteca - Réplica 4	Inóculo - R4	IR_Tag-16	AAAA
Endósfera tardía 20 d.p.i. raíz - Réplica 1	Endo-Root_L - R1	IR_Tag-17	ATTC
Endósfera tardía 20 d.p.i. raíz - Réplica 2	Endo-Root_L - R2	IR_Tag-18	TACT
Endósfera tardía 20 d.p.i. raíz - Réplica 3	Endo-Root_L - R3	IR_Tag-19	GAAA
Endósfera tardía 20 d.p.i. raíz - Réplica 4	Endo-Root_L - R4	IR_Tag-20	CAAT
Muestra de entrada enriquecida en LB - Réplica 1	Input_L - R1	IR_Tag-25	AGGT
Muestra de entrada enriquecida en LB - Réplica 2	Input_L - R2	IR_Tag-26	CACC
Muestra de entrada enriquecida en LB - Réplica 3	Input_L - R3	IR_Tag-27	GACG
Muestra de entrada enriquecida en LB - Réplica 4	Input_L - R4	IR_Tag-28	CTTG

II.3.f Preparación de adaptadores doble cadena

Los adaptadores doble cadena generados en esta tesis para diferentes ensayos se prepararon mezclando en cantidades iguales los oligonucleótidos correspondientes (indicados en cada caso y detallados en la Tabla II.5, y en una concentración 200 μ M cada uno) e hibridándolos en un ciclador térmico por descenso gradual de la temperatura: 95 °C por 2 minutos, y un descenso de 5 °C cada 1 minuto hasta alcanzar 20 °C.

II.3.g Secuenciamiento genómico de cepas bacterianas

El ADN de la cepa de interés a secuenciar se extrajo como se indica en la sección II.3.b. La muestra de ADN se envió a la compañía Plasmidsaurus (<https://www.plasmidsaurus.com>), y la secuenciación fue realizada utilizando la tecnología *Oxford Nanopore* con análisis y anotaciones personalizados. El servicio incluye la construcción de la biblioteca de lecturas largas sin amplificación, el secuenciamiento de acuerdo con el protocolo estándar de la empresa, el

ensamblaje del genoma y las anotaciones del mismo utilizando *Bakta* (<https://github.com/oschwengers/bakta>).

II.4. Ensayos con plantas

II.4.a Esterilización de semillas y germinación

En los distintos experimentos de este trabajo se utilizó *Medicago sativa* (alfalfa) de la variedad comercial GAPP 810+, de cosecha 2016, el cual fue almacenado a 5 °C hasta su utilización. Para su esterilización, primero las semillas fueron sumergidas en etanol 70% por 2 minutos, y luego tratadas con una solución de hipoclorito 7% y Tritón 0,2% durante 8 minutos, mezclando por inversión ocasionalmente. Finalmente fueron realizados 7 lavados sucesivos con agua destilada estéril.

Posterior a la esterilización, las semillas se extendieron en condiciones asépticas sobre placas de Petri con agar-agua 1,5% P/V, y se dejaron en oscuridad a 28 °C por 24 horas para su germinación. Se realizaron controles del procedimiento de esterilización sembrando en placas de Petri con LB y TY alícuotas de la última agua de lavado usada en la esterilización inspeccionando luego de incubarlas a 28 °C por 48 horas si existía crecimiento bacteriano.

II.4.b Condiciones de cultivo

Todos los ensayos se realizaron en una cámara de cultivo de plantas en las siguientes condiciones: fotoperíodo de 16/8 horas (luz/oscuridad), 22/16 °C de temperatura (día/noche).

II.4.b.1 Soporte inerte y esterilización

Como soporte inerte se utilizó vermiculita, tamaño de grano 3-8 mm, que fue lavada sucesivamente con abundante agua destilada y esterilizada por calor húmedo en autoclave (121 °C durante 20 minutos).

II.4.b.2 Solución nutritiva para crecimiento y riego de plantas

El crecimiento de las plantas se realizó en vermiculita con medio mineral Fåhraeus (Fåhraeus, 1957) libre de nitrógeno, cuya composición se detalla en la Tabla II.6. Las soluciones que componen el medio se prepararon 1.000 veces concentradas, y se esterilizaron individualmente por calor húmedo en autoclave (121 °C durante 20 minutos), con excepción de la solución de micronutrientes que fue filtrada usando filtros nitrocelulosa de 0,22 µm de diámetro de poro. Con las soluciones estériles se preparó el medio Fåhraeus empleando como base agua destilada o desionizada estéril.

Tabla II.7 Composición del medio mineral de irrigación de plantas

Solución		
Medio Mineral Fahraeus	CaCl ₂ .2H ₂ O	0,114 g/L
	MgSO ₄ .7H ₂ O	0,12 g/L
	Citrato Férrico	0,005 g/L
	KH ₂ PO ₄	0,1 g/L
	Na ₂ HPO ₄	0,15 g/ml
	Micronutrientes	1ml
	H ₂ O desionizada	1 L
Solucion de micronutrientes (1000x)	KCl	3,73 g/L
	H ₃ BO ₃	1,55 g/L
	SO ₄ Mn.H ₂ O	0,85 g/L
	CuSO ₄ .5H ₂ O	0,13 g/L
	ZnSO ₄ .7H ₂ O	0,58 g/L
	(NH ₄) ₆ Mo7O ₂₄ .4H ₂ O	0,01 g/L

II.4.b.3 Inoculación de plantas con los cultivos bacterianos de interés

Macetas de aproximadamente 330 mL conteniendo vermiculita y con agujeros en su base fueron inoculadas por inundación emulando “capacidad de campo” utilizando la solución inoculante de interés según el experimento. Para ello, la maceta con vermiculita y sus orificios de base obturados de forma manual fue inundada con la solución inoculante, permitiendo luego el drenaje del líquido excedente a través de los orificios que fueron liberados. Cada una de las macetas así inoculada fue luego apoyada sobre una bolsa de polietileno para evitar contaminaciones cruzadas durante los riegos sucesivos a lo largo del ensayo cuando correspondió. Las macetas inoculadas, y las macetas control con medio Fåhraeus y sin bacterias, fueron sembradas con semillas o con plántulas dependiendo del experimento según se indica en cada caso.

Posterior a la incubación de las plantas (control e inoculadas), a los tiempos indicados en cada ensayo se descalzaron las plantas y se removió la vermiculita adherida a raíces por movimiento manual enérgico de las plantas. A continuación, se procedió a recuperar las bacterias asociadas a los distintos nichos vegetales de interés, rizósfera o endosfera (de raíz o parte aérea) según se detalla en las siguientes secciones.

II.4.b.4 Recuperación de bacterias de la rizósfera

Para obtener las bacterias asociadas a la rizósfera, luego de eliminar la vermiculita unida a raíces (punto precedente), se cortó la parte aérea de la planta y la raíz fue sumergida en 5ml de solución fisiológica para experimentos con una sola raíz, o en 15 ml para experimentos con grupos de 15 raíces. Las suspensiones anteriores fueron agitadas vigorosamente en vortex durante 15 segundos, las plantas retiradas y el sobrenadante procesado para realizar recuentos bacterianos según se indicó en la sección II.2.b.

II.4.b.5 Recuperación de bacterias en endósfera

Para obtener las bacterias asociadas a la endósfera fueron utilizados dos métodos, atendiendo a si en el experimento se requería o no esterilizar la planta en superficie.

- a) Con esterilización superficial: En este procedimiento las plantas a las que ya se les había removido la vermiculita adherida (ver sección II.4.c. precedente) fueron sumergidas en una solución de peróxido de hidrógeno 25 % V/V preparada a partir de una solución comercial 12% Cl (Carl Roth – Art. 9062.3), resultando en una solución 10 volúmenes, durante 30 segundos. Luego, las plantas así tratadas fueron sujetas a 3 lavados con abundante agua destilada estéril, y las plantas enteras o raíz según se trate del análisis de endósfera tardía o temprana, fueron maceradas en mortero con 1 ml o 5 ml de solución fisiológica dependiendo si se analizaba sólo una planta o un conjunto de 15 plantas.
- b) Sin esterilización superficial: En este procedimiento las plantas a las que ya se les había removido la vermiculita adherida fueron sumergidas en solución fisiológica (5ml para experimentos con una sola planta, o en 15 ml para experimentos con grupos de 15). La suspensión de plantas fue agitada vigorosamente en vortex durante 15 segundos, el sobrenadante descartado, y las plantas lavadas durante 7 ciclos sucesivos con 10 mL de agua destilada estéril cada uno. Finalmente, las plantas fueron maceradas en mortero con 1 ml o 5 ml de solución fisiológica dependiendo que se trate de 1 o de 15 plantas.

En todos los casos, se realizó un plaqueo del último agua de lavado en medio rico LB sin antibiótico para evaluar la eventual presencia y cantidad de bacterias residuales (sean rizosféricas, desprendidas del rizoplaneo o derivadas del comportamiento endosférico por rotura mecánica de las plantas durante su lavado). Dichos números en el agua del último lavado nunca superaron el 10% de las bacterias recuperadas del interior de la planta.

II.5. Ensayos TnSeq (*transposon insertion sequencing*)

II.5.a Desafío de la biblioteca de mutantes en condiciones de colonización de rizósfera, endosfera de plantas de alfalfa, medio rico LB y medio mínimo M9

II.5.a.1 Esquema general, mezcla de entrada e inóculo

Los experimentos TnSeq consistieron en la evaluación de los miles de mutantes presentes en la biblioteca de mutantes de *Pantoea agglomerans* LPU12 generada para esta tesis como se describe en el Capítulo IV. Los mutantes se usaron como inóculo para las condiciones propuestas a evaluar (Figura V.4), y se recuperaron aquellos capaces de crecer en esas condiciones. La representación numérica de cada mutante en el inóculo inicial y luego de la evaluación en cada condición se obtuvieron posterior al procesamiento de cada muestra según se describe en la sección V.1, y es esta variación la que nos permitirá identificar mutantes afectados en las condiciones de interés.

Para la preparación del inóculo inicial, la biblioteca de mutantes Tn *mariner* de la cepa *Pantoea agglomerans* LPU12 conservada a -80 °C fue crecida en medio rico LB suplementado con los antibióticos kanamicina y ácido nalidíxico durante toda la noche, y de este cultivo se extrajo una alícuota de 500 µl para inocular 400 ml de medio LB. Este cultivo se creció hasta alcanzar una DO₆₀₀ de 0,3-0,5; y se utilizó para preparar los inóculos iniciales diluyendo en: medio mineral Fåhraeus a las distintas concentraciones de UFC/mL (baja con 10⁵ UFC/mL o alta con 10⁷ UFC/mL según se indica en cada ensayo para estudios de colonización rizosférica o endofítica, respectivamente), en medio Mínimo M9 y en medio rico LB nuevamente de modo de evaluar si el crecimiento en medio rico genera sesgos en la composición de mutantes de la biblioteca. El esquema experimental final propuesto junto con el detalle del procesamiento de las muestras obtenidas se describe en profundidad en el capítulo V, en las secciones V.1.a y V.3.a. La inoculación de las plantas, la extracción de bacterias de cada compartimento, y la amplificación de las poblaciones de mutantes por crecimiento de los mismos se describe en las secciones

siguientes II.5.a.2. y II.5.b. La muestra de entrada y los crecimientos finales provenientes de medio LB, medio M9 y de las muestras rizosféricas y endofíticas fueron todos procesados de igual manera para la extracción de ADN (como se detalla en la sección II.5.c).

II.5.a.2 Ensayos con plantas y su procesamiento

Las macetas fueron inoculadas con distintas concentraciones de bacterias según se trató de estudios de colonización rizosférica o endofítica (ver Figura V.4.), siguiendo el procedimiento detallado en la sección II.4.b.3, y se sembraron con 6-7 plántulas preparadas a partir de semillas esterilizadas y germinadas según el protocolo detallado en la sección II.4.a. Una vez inoculadas, las plantas fueron cultivadas según lo indicado en la sección II.4.b. por un período de 24 horas, 3 días (3 dpi) o 20 días (20 dpi) según el tipo de colonización que fue analizada (endosfera temprana, rizósfera y endosfera tardía, respectivamente). Cumplidos los tiempos anteriores las bacterias asociadas a las plantas se recuperaron como se detalla en las secciones II.4.b.4 para bacterias asociadas a la rizósfera, y como se indica en la sección II.4.b.5. con un protocolo de esterilización superficial para bacterias presentes en la endosfera.

II.5.b Amplificación de las poblaciones de mutantes recuperadas de rizósfera y endosfera por crecimiento en medio rico.

En la sección V.2.c se discute en profundidad la necesidad de enriquecer las muestras obtenidas luego de evaluar cada condición para aumentar la biomasa bacteriana que se procesará. Así, las muestras de las poblaciones de mutantes recuperadas de las distintas condiciones analizadas fueron utilizadas para inocular 50 ml de medio LB suplementado con los antibióticos Km y Nal (correspondientes a las resistencias del transposón y de la cepa de *P. agglomerans* LPU12 usada, respectivamente). **Todos los cultivos, que iniciaron en una concentración cercana a 10^2 UFC/ml fueron crecidos durante la noche hasta alcanzar una concentración de bacterias superior a 10^9 UFC/ml. Las células fueron cosechadas por centrifugación a $6000 \times g$ por 10 minutos y procesadas según se detalla en la sección siguiente para la obtención de ADN genómico de cada una de las muestras.**

II.5.c Extracción de ADN genómico de las muestras y su procesamiento para la posterior identificación y cuantificación de los sitios de inserción del Tn mariner

En el procesamiento de las muestras de ADN se detalla en la sección V.2.b donde se optimizó el protocolo utilizado, y la Figura V.1 presenta un esquema de los pasos, enzimas, adaptadores y cebadores de PCR utilizados para el procesamiento del ADN genómico de las mezclas de mutantes, para la posterior identificación de los sitios de inserción por secuenciación. De las células recuperadas según se detalló en la sección anterior se extrajo ADN genómico empleando el procedimiento detallado en la sección II.3.b. Luego, 5 μ g de ADN fueron digeridos con la enzima de restricción de clase IIS *MmeI*, en las condiciones que indica el fabricante durante toda la noche. Luego de inactivar la enzima mediante tratamiento térmico, se ligaron adaptadores doble cadena, que se prepararon previamente según se indica en la sección I.3.f. Los fragmentos de ADN con los adaptadores ligados fueron purificados utilizando el *GeneJET PCR purification kit* (ThermoFisher), y 3,12 ng de cada muestra fue utilizada como molde para las reacciones de qPCR que amplificaron los sitios de inserción y que sirvieron para etiquetar los productos de amplificación correspondientes a cada tratamiento. Los cebadores utilizados para cada muestra incluyeron el Adapter-Top-PCR y alguno de los oligonucleótidos designados IR_Tag-n (donde n corresponde a un número único, indicados en la Tabla II.6) que se detallan en la Tabla II.5. Estos últimos oligonucleótidos, que difieren en su secuencia 5'- terminal (en negrita subrayados 5'-**ANNNN**CGGGGACTTATCATCCAACCTG-3'), sirvieron para etiquetar de manera diferencial los productos

de amplificación de cada una de las muestras. La reacción de PCR cuantitativa en tiempo real se realizó como se describe en la sección II.3.e, y las condiciones del ciclado utilizado fueron las siguientes: desnaturalización inicial 95 °C 3 minutos y 40 ciclos con los siguientes pasos: desnaturalización 95 °C, 10 segundos; hibridación 55 °C, 15 segundos; elongación 60 °C, 45 segundos.

II.5.d Mezcla de los productos de amplificación etiquetados

La mezcla para el secuenciamiento fue armada con cantidades equivalentes de cada uno de los productos de amplificación. Para esto se tuvo en cuenta la intensidad máxima de fluorescencia alcanzada por cada reacción al final de la reacción de qPCR en el *plateau* (ciclo 38). Para elaborar la mezcla, en primer lugar, se tomaron 10 µl de la muestra que mostró la máxima intensidad en el *plateau* (muestra de referencia). A continuación, se agregaron sucesivamente distintas cantidades del resto de las muestras de modo de incorporar, en cada caso, la misma cantidad de unidades de fluorescencia que las presente en los 10 µl de la muestra de referencia (los volúmenes fueron siempre mayores que 10 µl dado que la muestra de referencia fue la que tenía la máxima fluorescencia en el *plateau*).

II.5.e Secuenciamiento de los sitios de inserción del Tn mariner y procesamiento de datos

Los productos de amplificación obtenidos con el procedimiento detallado en el punto precedente fueron secuenciados empleando la plataforma Illumina NextSeq del CeBiTec (Universidad de Bielefeld, Alemania) atendiendo a la necesidad de acceder a un alto número lecturas de productos cortos (64-66 pb). Con el diseño experimental propuesto se planificó analizar los productos de PCR correspondientes a 24 muestras diferentes: muestra de entrada ME y 5 tratamientos con cuatro réplicas en cada caso (indicados en la Figura V.4). Se obtuvieron un total de 80.387.446 lecturas “*paired-end*” con una calidad promedio de 33 (Phred Score), un contenido GC de 52% y una longitud promedio de 51 pares de bases. Las lecturas fueron unidas utilizando el software *NGmerge*, obteniéndose 74.730.990 lecturas unidas exitosamente (93% del total de lecturas). Posteriormente, las lecturas se procesaron con scripts desarrollados en el laboratorio por el Dr. Lozano, utilizando el software *Flexbar* para filtrar secuencias de longitud mínima de 65 nucleótidos, conforme a la estructura esperada “5’-**AAAGGACGTGGTTACGGGGC**NNNNNNNNNNNNNNNNNN**ACAGGTTGGATGATAAGTCCCGNNNN-3’**” (adaptador en azul, la secuencia del extremo del Tn en rojo junto con la etiqueta NNNN que identifica la condición y réplica a la que pertenecen, descritas en la Tabla II.6, y la secuencia genómica contenida entre ellos en verde, descrita en la sección V.II.a del Capítulo V). Tras el filtrado, se descartaron 7.349.039 lecturas por presencia de bases ‘N’ y 3.600.739 por ser secuencias cortas, resultando en un total de 10.949.778 secuencias descartadas. El análisis subsiguiente se realizó con 63.781.212 lecturas restantes (85% del total de lecturas unidas).

Posteriormente, con el mismo script, se separaron las lecturas según su etiqueta asignada (*demultiplexing*), en una serie de etapas. Primero, las lecturas fueron clasificadas de acuerdo a su etiqueta para agruparlas por condición (entrada-inóculo, salida-muestra) y número de réplica, extrayendo de cada una su correspondiente secuencia informativa de 16 pb. En total, cerca de 54.000.000 lecturas (68,3%) fueron clasificadas correctamente según réplica y condición. La siguiente etapa implicó el mapeo de las secuencias al genoma de la cepa *Pantoea agglomerans* LPU12 mediante el programa TPP (*Transit preprocessor*, DeJesus et al., 2015) con el algoritmo Bwa en modo “aln”. El genoma de la cepa *P. agglomerans* LPU12, secuenciado, ensamblado y anotado como parte de este trabajo (sección II.3.g), fue verificado como completo mediante CheckM (100%). Dado el corto tamaño de las secuencias (16/17 nucleótidos), se optó por utilizar el algoritmo “aln” sin errores o “*mismatch*” (con la opción “n 0” del software).

Con los archivos obtenidos, se llevó a cabo el análisis estadístico mediante el software *Transit* (DeJesus et al., 2015). Los datos reportados corresponden al análisis de comparación por pares de condiciones empleando la opción “Resampling”, con normalización “b-geométrica”, y los tres análisis disponibles para determinación de genes esenciales: Gumbel/Binomial, HMM y TTN fitness. A partir del recuento del número de inserciones por gen en la muestra de entrada (ME con sus réplicas) y las muestras de salida (detalladas en la Figura V.4 del Capítulo V), el software calcula los valores del parámetro M^{TnSeq} para cada uno de los genes anotados:

$$M^{\text{TnSeq}}_{(\text{gen-x})} = \log_2 \left\{ \frac{[\text{lecturas en gen-x}]^{\text{muestra de salida}}}{[\text{lecturas en gen-x}]^{\text{muestra de entrada, t = 0}}} \right\}$$

Valores de M^{TnSeq} negativos y positivos fueron asociados a genes que perturban o favorecen la colonización rizosférica/endofítica, respectivamente. Valores de $M^{\text{TnSeq}} = 0$ fueron asociados a genes sin relevancia para los fenotipos estudiados.

II.6. Algoritmos de *Machine learning* para la predicción del tropismo a planta

II.6.a Construcción y curado del set de datos utilizado para el entrenamiento de modelos de

Machine Learning

Un total de 417 genomas y genomas ensamblados de metagenomas (MAGs, Metagenome assembled Genomes) pertenecientes al género *Pantoea* se obtuvieron de la base de datos *Integrated Microbial Genomes and Microbiomes* (Chen et al., 2023) junto con sus anotaciones funcionales *Clusters of Orthologous Genes* (COG), *KEGG Orthology* (KO) y dominios PFAM). El genoma de la cepa *Pantoea agglomerans* LPU12 (López et al., 2018) utilizada en este trabajo se incluyó en el conjunto de datos usado, y sus anotaciones funcionales se realizaron usando *eggNong-mapper* (Cantalapiedra et al., 2021). La clasificación de los genomas y MAGs de acuerdo con su nicho de aislamiento se anotó manualmente en base a la información disponible en la base de datos IMG y, cuando era posible, en la publicación del genoma de la cepa bacteriana. Los genomas bacterianos utilizados para este análisis se detallan en la Tabla S.7 donde se indica el identificador propio de la base de datos IMG, el nombre asignado a la cepa, el origen de aislamiento reportado (clasificado en Ambiental, Animal, Hongo, Artrópodo o Planta), el tamaño de su genoma y el fenotipo patogénico descrito cuando estuviese disponible.

II.6.b Análisis filogenético

Se construyó un árbol filogenético con las cepas del género *Pantoea* usadas en este trabajo utilizando la herramienta GTDB-Tk (Chaumeil et al., 2020a). La visualización y anotación del árbol se realizó con la herramienta *Interactive Tree Of Life* (iTOL) (Letunic & Bork, 2021).

II.6.c Algoritmos de *Machine Learning* utilizados

Las matrices binarias (presencia/ausencia) utilizadas para entrenar los algoritmos de clasificación de *Machine Learning* se construyeron utilizando programas escritos en el lenguaje *Python* desarrollados para este estudio, en base a las tres diferentes anotaciones funcionales disponibles para cada cepa.

Los algoritmos de clasificación implementados fueron *Random Forest* (RF) y *Logistic Regression* (LR), y su desempeño se analizó mediante el análisis de la curva ROC (*Receiver operating characteristic*), usando la biblioteca *Scikit-learn* en el lenguaje de programación *Python* (Pedregosa, 2011). Los Modelos de Random Forest utilizaron hiperparámetros estándar, $max_features = 'sqrt'$, $criterion = 'entropy'$,

random_state = 123, y un número total de 600 árboles de decisión (*n_estimators=600*). Por otro lado, los modelos entrenados con el algoritmo *Logistic Regression* utilizaron como parámetro estándar *solver = liblinear*.

II.7. Bases de datos utilizadas

Las bases de datos, herramientas bioinformáticas y estadísticas utilizadas en este trabajo para obtener información de genes, genomas, y rutas metabólicas reportadas, así como también para evaluar la significancia estadística de los resultados en los casos en que fue posible, indicadas en la Tabla II.8.

Tabla II.8. Bases de datos, herramientas bioinformáticas y estadísticas utilizadas en este trabajo.

Base de datos / Herramienta	Utilidad	Fuente
IMG	Sistema de anotación y búsqueda de genomas.	https://img.jgi.doe.gov/
KEEG	Información de genes, proteínas, metabolitos y vías metabólicas.	http://www.genome.jp/kegg
EggNOG mapper	Herramienta para la anotación Funcional de grandes conjuntos de secuencias.	http://eggno-mapper.embl.de/
GraphPad	Visualización y análisis estadístico de resultados .	https://www.graphpad.com/
COG	Identificación de funciones evolutivamente conservadas .	https://www.ncbi.nlm.nih.gov/research/cog
iTOL Interactive Tree Of Life	Visualización y anotación de árboles filogenéticos.	https://itol.embl.de/
NGmerge	Software para la unión (<i>merge</i>) de lecturas "paired-end".	https://github.com/harvardinformatics/NGmerge
Flexbar	Eliminación flexible de etiquetas (barcodes) y adaptadores de secuencias genómicas.	https://github.com/seqan/flexbar
TPP (Transit preprocessor)	Parte del software Transit, permite mapear lecturas en el genoma de un organismo.	https://transit.readthedocs.io/en/v3.2.2/
Transit	Software para el análisis de datos de experimentos TnSeq.	https://transit.readthedocs.io/en/v3.2.2/
TYGS - Type (Strain) Genome Server	Identificación / clasificación de forma fiable cepas procariotas.	https://tygs.dsmz.de/
GTDB-Tk	Asignar clasificaciones taxonómicas a genomas de bacterias y arqueas.	https://ecogenomics.github.io/GTDBTk/index.html



CAPÍTULO III

CARACTERIZACIÓN DE UN NUEVO SISTEMA MODELO PARA EL ESTUDIO DE LA COLONIZACIÓN RIZOSFÉRICA Y ENDOFÍTICA DE RAÍCES: LA ASOCIACIÓN *PANTOEA AGGLOMERANS* – *M. SATIVA* (ALFALFA).



III.1. Introducción

Como hemos mencionado en la introducción de esta tesis, muchos microorganismos del suelo tienen capacidad de abordar la colonización de las plantas, primero a través del acceso a la rizósfera y luego penetrando al interior de la propia raíz. Desde hace tiempo tal acceso de la comunidad microbiana al interior de la planta ha recibido particular atención debido a la estrecha asociación que se establece entre los participantes como parte del holobionte que conforman (Zilber-Rosenberg & Rosenberg, 2008). Más allá de las interacciones de carácter patogénico (donde la planta es agredida por las bacterias que la invaden (Agrios, 2005; Agrios, 2009), para el resto de las bacterias colonizantes—muchas de ellas promotoras del crecimiento y/o la sanidad vegetal—la endosfera de las plantas sanas representa un nicho marcadamente homeostático (y por tanto protector) que contrasta con la variabilidad natural del suelo. A pesar de estas características que hacen al notable número de bacterias que habitan los tejidos de las plantas (cientos de miles/gramos) existe limitada información respecto a los genes y funciones que son requeridos por las bacterias endófitas para abordar el interior de las raíces. De todos modos, las evidencias disponibles muestran claras diferencias entre los microbiomas rizosféricos y endofíticos en distintas especies de plantas (Dong et al., 2019; Trivedi et al., 2020), de lo que puede inferirse que existen mecanismos a través de los cuales se seleccionan los microorganismos que efectivamente acceden al interior de las plantas, como resultado de restricciones (barreras) impuestas por la planta y de capacidades colonizadoras expresadas por las bacterias endófitas. Así, comparando genomas completos de nueve proteobacterias endófitas Ali et al. (2014) identificaron una lista de genes con posibles funciones relacionadas a la vida endofítica. Como parte de dicha lista se identificaron transportadores, sistemas de secreción, reguladores transcripcionales, sistemas de detoxificación, genes involucrados en degradación y modificación de polímeros vegetales, y genes cuyos productos están asociados al mantenimiento del potencial redox. Hasta el momento, sólo para algunos de esos genes pudo corroborarse experimentalmente su importancia en la colonización endofítica. Diferentes técnicas han sido aplicadas para la identificación de genes de colonización. Entre estas podemos destacar técnicas de delección/interrupción y complementación de genes (Sun et al., 2009; Zúñiga et al., 2013), PCR en tiempo real para evaluar la expresión de genes puntuales (Yousaf et al., 2011; Zhao et al., 2016), tecnología de expresión *in vivo* (IVET) (Rediers et al., 2003), sistemas reporteros de fusión al gen *gusA* (Roncato-Maccari et al., 2003), sobreexpresión de genes (Fan, 2013), introducción y expresión de genes foráneos en bacterias de interés (Cho et al., 2007), estudios del perfil transcriptómico completo basado en la secuenciación de ARN (RNA-seq) (Sheibani-Tezerji et al., 2015) y estudios fenómicos como el de esta tesis (Duong et al., 2018; Royet et al., 2019; Salas et al., 2017; Wheatley et al., 2020), entre otras. Si bien todas estas técnicas han sido aplicadas a distintas bacterias endófitas, varias de ellas han enfocado *Paraburkholderia phytofirmans* PsJN, una bacteria endófito modelo con la capacidad de colonizar de forma competente (tanto la rizosfera como la endosfera) y promover el crecimiento de una variedad de plantas hospedadoras diferentes, incluyendo *Arabidopsis thaliana*, uva, maíz, patata, pasto varilla, tomate y trigo (Sheibani-Tezerji et al., 2015; Sørensen & Sessitsch, 2006). Se ha demostrado, por ejemplo, que la cepa PsJN requiere la degradación de ácido indol-3-acético (AIA), la actividad de la enzima ACC deaminasa y del mecanismo de *quorum sensing* para colonizar plantas huésped y producir efectos beneficiosos (Sun et al., 2009; Zúñiga et al., 2013). Además, los perfiles de expresión génica *in planta* revelaron que, durante el crecimiento de la bacteria en el interior de las plantas huésped, la misma expresa una serie de rasgos particulares relacionados con la homeostasis de óxido-reducción (redox) en la célula; la producción de energía; el metabolismo general de aminoácidos, lípidos, nucleótidos y azúcares; y la regulación de la transcripción (Sheibani-Tezerji et al., 2015). Utilizar cepas bacterianas modelo, como en el caso anterior,

permitió estudiar interacciones complejas en profundidad, en sistemas que eran bien conocidos en sus características funcionales.

Con el propósito de investigar en detalle la colonización rizosférica y endofítica de una bacteria asociada de forma ubicua a plantas, buscamos un sistema que cumpla con dicha condición (presencia ubicua en plantas), que destaque por la extensión de la colonización, que sea de simple manipulación en condiciones de laboratorio, y que las evidencias que puedan colectarse sirvan para realizar estudios comparados en otras bacterias saprófitas o patógenas asociadas a plantas y a otros organismos multicelulares. Al considerar las condiciones anteriores, los candidatos principales pertenecieron al género *Pantoea*. Como se describió en las secciones I.2.e y I.3. de la introducción de esta tesis, *Pantoea* spp. está presente no sólo como endófito de tejidos vegetativos en plantas (en números cercanos a 10^6 – 10^7 u.f.c./g de tejido), sino también en semillas, con características en algunos casos promotoras del crecimiento del huésped y en otras como patógeno. Además, ciertas *Pantoea* han sido descritas como patógenas en algunos sistemas animales (Dutkiewicz et al., 2016; Walterson & Stavrínides, 2015). Esto permitirá que los resultados que presentamos en este trabajo de tesis puedan ser empleados para evaluar su grado de ubicuidad en otras *Pantoea* con distintos fenotipos asociativos (saprófitas, PGPR o patógenas) y en distintos hospedadores.

En este capítulo presentamos la cepa de *Pantoea* que estudiaremos en esta tesis, sus características genómicas, taxonomía y dinámica de colonización de *Medicago sativa* (alfalfa), la planta hospedadora que usaremos en este trabajo.

III.2. *Pantoea agglomerans* LPU12, la bacteria endófito de alfalfa que estudiaremos en esta tesis.

En la sección I.2.e de esta tesis se describió el microbioma endofítico que pudimos recuperar de semillas y plantas de alfalfa, donde los aislamientos del género *Pantoea* mostraron una presencia particularmente abundante y frecuente en las distintas variedades de *Medicago sativa* analizadas, junto con cepas pertenecientes a los géneros *Arsenicoccus*, *Arthrobacter*, *Bacillus*, *Kocuria*, *Microbacterium*, *Micrococcus* y *Paenibacillus*, entre otros (López et al., 2018). Las bacterias del género *Pantoea* asociadas a alfalfa estuvieron presente en números elevados comparadas con la mayoría de los endófitos de semilla, representando aproximadamente 12% del total de aislamientos entre los más 40 géneros identificados. Los aislamientos de *Pantoea* mostraron *in vitro* producción de auxinas, celulasas y fenotipos muy destacados de solubilización de fosfatos y producción de sideróforos. Estas dos últimas características distinguieron a varios aislamientos de *Pantoea* spp. del resto de los endófitos en un análisis de componentes principales (PCA) en el que se incluyeron distintas actividades bioquímicas vinculadas a la promoción del crecimiento vegetal (López et al., 2018). En la colección de 240 endófitos, disponible en el cepario de nuestro grupo en el Instituto de Biotecnología y Biología Molecular (IBBM), contamos con 12 cepas del género *Pantoea* aisladas de semillas y plantas de alfalfa de distintos cultivares, las cuales presentaban varias actividades promotoras del crecimiento, previamente mencionadas (ver Tabla III.1 extraída de López et al. (2018). De las distintas *Pantoea* de nuestra colección, seleccionamos la cepa LPU12 recuperada originalmente de exudados de semillas de alfalfa. Además de sus destacadas propiedades bioquímicas usualmente asociadas a la promoción del crecimiento de plantas (Tabla III.1), como mostraremos en este capítulo y en los siguientes, la cepa LPU12 tiene capacidad de colonizar eficientemente la rizósfera y la endosfera radicular de alfalfa. Atendiendo a estas características escogimos a la cepa LPU12 para la búsqueda de genes de colonización con la metodología TnSeq que presentamos en esta tesis.

Tabla III.1 Características fenotípicas del aislamiento Pan12 usado en este trabajo de tesis respecto de las del resto de los aislamientos del mismo género recuperados de semillas y plantas de alfalfa. Los distintos aislamientos fueron recuperados de distintos cultivares de *Medicago sativa*. Para cada aislamiento se indica el origen (semillas (S) o plantas (P)), cultivar del que fue aislado (C: GAPP810+, D: F301Y, G: F111X, SM: Super Monarca), y las actividades promotoras del crecimiento *in vitro* que fueron analizadas. En rojo se indica el aislamiento seleccionado para los experimentos de esta tesis. La presente Tabla ha sido extraída y adaptada de López et al. (2018).

Aislamiento de <i>Pantoea</i> spp	Cultivar	Aislado de semilla(S)/ Planta(P)	Género	Actividad enzimática					Otras actividades			Inhibición fungica
				Proteasa	Celulasa	Quitinasa	Amilasa	Pectinasa	Solubilización de fosfatos	Producción de sideróforos	Producción de AIA (auxina)	
LPU1	SM	P	<i>Pantoea</i>	-	+	-	-	-	+++	++	++	-
LPU10	G	P	<i>Pantoea</i>	-	+	-	-	-	++	+++	-	-
LPU11	SM	S	<i>Pantoea</i>	-	-	-	-	-	+++	++	+	-
LPU12	SM	S	<i>Pantoea</i>	-	+	-	-	-	+++	+	++	-
LPU2	SM	P	<i>Pantoea</i>	-	+	-	-	-	+++	++	++	-
LPU3	SM	P	<i>Pantoea</i>	-	+	-	-	-	+++	++	+++	-
LPU4	SM	P	<i>Pantoea</i>	-	+	-	-	-	+++	++	++	-
LPU5	SM	P	<i>Pantoea</i>	-	+	-	-	-	+	++	++	-
LPU6	C	P	<i>Pantoea</i>	-	-	-	-	-	+	++	+	-
LPU7	D	P	<i>Pantoea</i>	-	-	-	-	-	++	-	-	-
LPU8	D	P	<i>Pantoea</i>	-	-	-	-	-	+++	++	-	-
LPU9	G	P	<i>Pantoea</i>	-	+	-	-	-	++	+	-	-

III.2.a *Pantoea* sp. LPU12 resistente a ácido nalidíxico

Para estudiar la colonización de plantas por una bacteria particular y distinguirla de otras que colonizan el mismo huésped, es necesario marcar la cepa de interés (la cepa “indicadora”). Las semillas de alfalfa, como hemos dicho, suelen portar diversas bacterias incluyendo cepas de *Pantoea*, y por tanto resultó indispensable contar con una variante marcada de la cepa *Pantoea* sp. LPU12 que luego usamos como inoculante. Para ello, generamos un derivado resistente a ácido nalidíxico que nos permitió diferenciar nuestra cepa LPU12 indicadora de las *Pantoea* naturalmente existentes en las plantas. Cabe recordar que la resistencia a ácido nalidíxico, asociada a cambios en la enzima topoisomerasa girasa, es de localización cromosomal y por tanto no transmisible horizontalmente con facilidad. Para generar la cepa resistente se buscaron mutantes espontáneos que crecieran en placas con medio LB suplementadas con 5 µg/ml de ácido nalidíxico (Materiales y Métodos sección II.1.c). A partir de clones que crecieron en esa condición se repitió el procedimiento empleando concentraciones crecientes del antibiótico hasta obtener clones resistentes a 30 µg/ml de ácido nalidíxico. Uno de los clones con esas características, *Pantoea* sp. LPU12 Nal^R, fue criopreservado y utilizado en los diferentes experimentos que se detallan en esta tesis.

III.2.b Secuenciación genómica y posicionamiento taxonómico de *Pantoea* sp. LPU12 como *Pantoea agglomerans* LPU12

Como se describió en la sección I.4 de la Introducción, en la metodología TnSeq que se utilizará en este trabajo de tesis la identificación de las regiones de inserción del Tn *mariner* se realiza mediante el mapeo de las secuencias que bordean al transposón en el genoma de referencia. Fue por tanto esencial contar con el genoma de la cepa *Pantoea* sp. LPU12 que se empleará para la búsqueda de los genes de colonización por TnSeq. Por otra parte, conocer la secuencia genómica de la cepa LPU12 fue importante para asignar una posición taxonómica adecuada a nuestra cepa modelo LPU12. La secuenciación de la misma fue realizada empleando tecnología Oxford Nanopore, según se detalló en la sección II.3.g de Materiales y Métodos. El tamaño del genoma ensamblado fue de 4.878.567 bases (aproximadamente 4,9 Mb) distribuidas en 4 replicones (un cromosoma de 4,09 Mpb (4.093.002 pb) y 3 plásmidos: pPagLPU12_1 de 181,6 Kpb (181.617 pb); pPagLPU12_2 de 525,7 Kpb (525.669 pb) y pPagLPU12_3 de 78,3 Kpb (78.279 pb). A partir de dicha secuencia pudieron anotarse 4.605 genes: 3.863 genes en el cromosoma, 149 genes en el pPagLPU12_1; 506 genes en el pPagLPU12_2 y 87 genes en el pPagLPU12_3. En la Figura III.1 se muestran los cuatro replicones con su tamaño relativo, las regiones codificantes en cada hebra, el contenido GC para cada región génica y el desbalance del contenido G vs. C (GC skew).

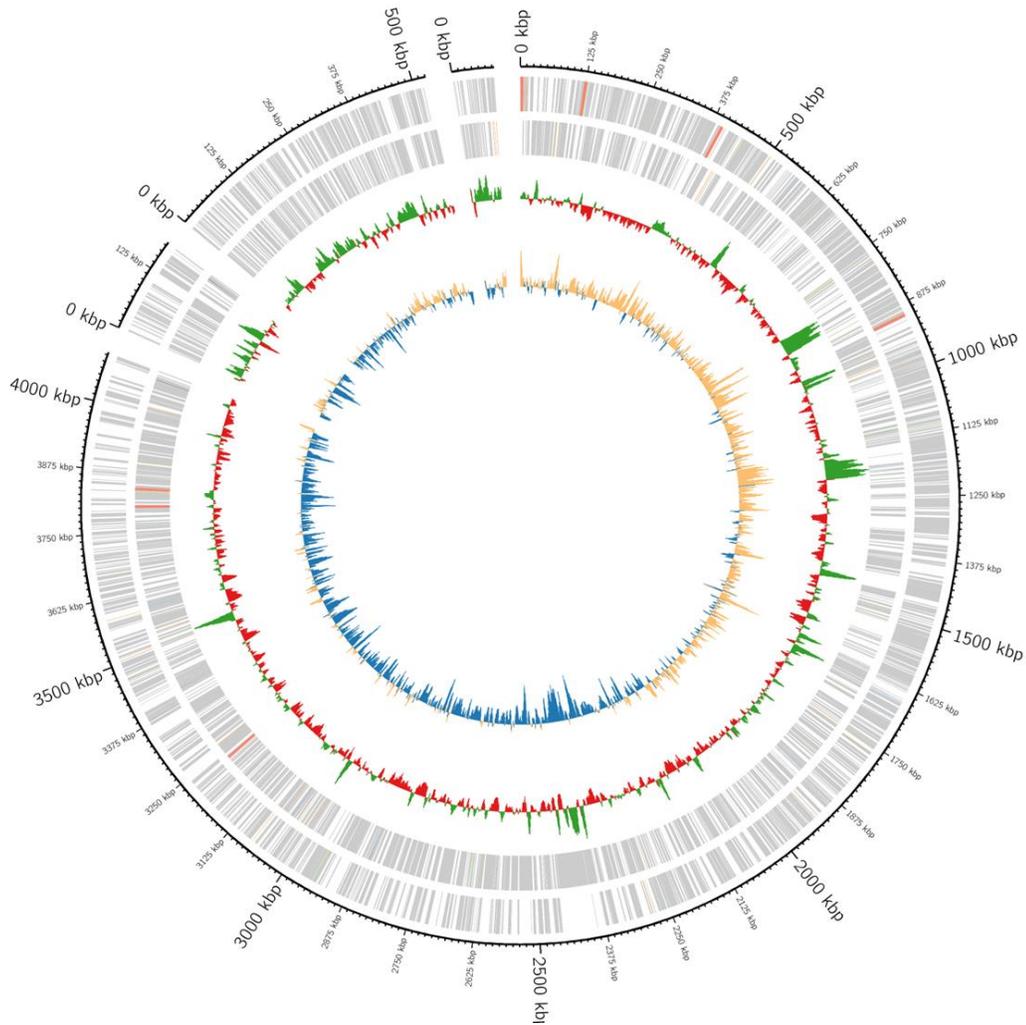


Figura III.1 Esquema circular de los cuatro replicones de la cepa LPU12. Línea exterior: cromosoma y plásmidos con escala de tamaños en kbp. Líneas grises: marcos abiertos de lectura (ORFs) en ambas hebras. Círculo en rojo y verde: contenido GC. Círculo interno en naranja y azul: GC skew.

Para analizar la posición taxonómica del aislamiento LPU12, se utilizó el servidor TYGS, *Type (Strain) Genome Server* (Meier-Kolthoff & Göker, 2019) (<https://tygs.dsmz.de/>). Brevemente, el servidor implementa las siguientes operaciones para inferir la posición taxonómica de aislamientos de interés:

- Busca y extrae de la secuencia genómica del rADN 16S del aislamiento problema.
- Realiza diferentes BLASTn de dicha secuencia contra las secuencias de rADN de todas las cepas tipo disponibles.
- Para todos aquellos pares que tienen un puntaje (*score*) elevado calcula las distancias GBDP (*Genome Blast Distance Phylogeny*).
- Ordena los resultados obtenidos e identifica los 10 genomas más cercanamente relacionados a la cepa de interés ("grupo 1" de secuencias genómicas).
- Dado que hay genomas incompletos que pueden no incluir secuencias rADN 16S, se calculan las distancias MASH entre el genoma de la cepa de interés y las distintas cepas tipo extrayéndose los 10 genomas más relacionados ("grupo 2" de secuencias genómicas). Para el cálculo de distancias MASH

se usa el algoritmo MinHash que crea vocetos pequeños y representativos de secuencias grandes para estimar su distancia de forma rápida.

- (f) Calcula las distancias GBDP entre el genoma de interés y los genomas que resultan de la unión del “grupo 1” y el “grupo 2” de secuencias.
 (g) En base a las distancias anteriores se infiere un árbol filogenético.

En la Figura III.2 se presenta el árbol filogenético obtenido con TYGS, donde la cepa LPU12 se ubica en un clado compartido con *Pantoea agglomerans* y *Curtobacterium plantarum* con los que comparte especie, porcentaje GC, valor δ , tamaño del genoma y proteínas totales. Si bien *Curtobacterium plantarum* aún es descrito con este nombre, la bibliografía y bases de datos indicada por el *NCBI Taxonomy Browser* (Schoch et al., 2020) señala esta cepa como perteneciente al género *Pantoea*, en acuerdo con la *American Type Culture Collection* (ATCC) y el instituto Leibniz DSMZ (*Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen*). Apoyados en el análisis que hemos realizado precedentemente empleando la secuencia genómica completa, de aquí en adelante designaremos al aislamiento LPU12 como *Pantoea agglomerans* LPU12.

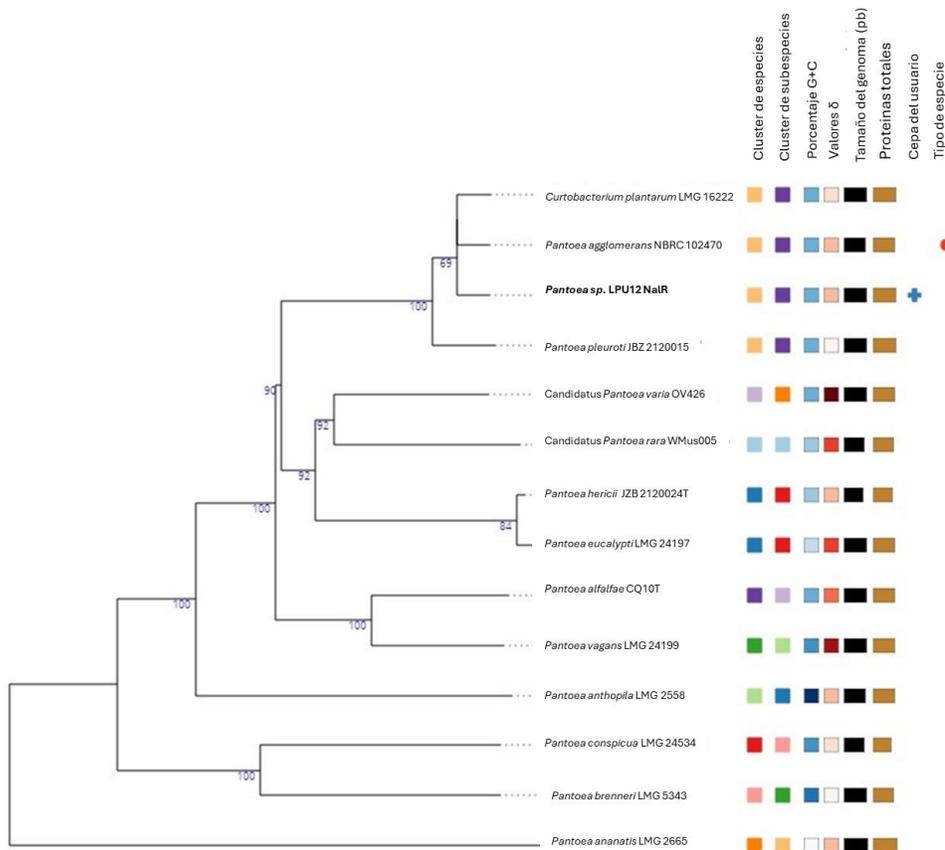


Figura III.2 Árbol filogenético generado por el servidor TYGS que incluye el aislamiento LPU12 cuya capacidad de colonización de alfalfa estudiamos en esta tesis. Las longitudes de las ramas están expresadas en términos de la fórmula d_5 de la distancia GBDP (<https://tygs.dsmz.de/>). Las etiquetas en color de las columnas de la derecha corresponden a grupos de: especies y subespecies, contenido genómico GC, valores δ , longitud total de la secuencia genómica y número de proteínas. En las dos últimas columnas se señala la cepa problema LPU12 (cruz celeste) y la cepa tipo más cercana (punto rojo, *Pantoea agglomerans* NBRC 102470).

III.3. Caracterización del sistema modelo propuesto en este trabajo, la asociación *Medicago sativa* – *Pantoea agglomerans* LPU12: Grado de colonización y progreso temporal del abordaje rizosférico y endofítico.

Como ya hemos mencionamos, la asociación establecida entre bacterias endófitas y las plantas que las hospedan es el resultado de una compleja serie de eventos a lo largo de los cuales se producen cambios en ambos participantes en respuesta a un dialogo molecular activo (Badri et al., 2009; Chiquito-Contreras et al., 2024). Así, como resultado del progreso de la interacción con la planta hospedadora, la bacteria que accede secuencialmente a los nichos rizosférico y endofítico adapta y compatibiliza su metabolismo al nuevo ambiente que coloniza. Dado que la planta ofrece un nicho con capacidad de alojamiento limitada (tanto espacialmente como en recursos), a la endosfera llegarán aquellas bacterias que presenten las mejores aptitudes para colonizar inicialmente la raíz (la denominada colonización “fundacional”). Tales bacterias fundadoras deberán luego establecerse, esto es multiplicarse y persistir.

Para abordar los ensayos que describiré en los Capítulos V y VI orientados a identificar la información genómica de LPU12 que determina sus propiedades de colonización rizosférica y endofítica de alfalfa, caracterizamos en primera instancia la dinámica de abordaje a dichos nichos. Para los ensayos de colonización se priorizó atender las siguientes consideraciones generales:

(A) En los experimentos de colonización rizosférica utilizar inóculos bajos (ver condiciones en secciones siguientes), de modo que la población “fundadora” se constituya a partir de las bacterias (genotipos) que accedan (se acerquen) y se establezcan (multipliquen, persistan) de la manera más eficiente en el nicho rizosférico.

(B) En los experimentos de colonización endofítica utilizar inóculos altos de bacterias (ver condiciones en secciones siguientes), para que en esas condiciones no haya limitaciones al acceso rizosférico. De ese modo, es de esperar en estudios TnSeq los mutantes que se muestren alterados en la colonización de la endosfera sean deficientes en el propio ingreso a la raíz o en el crecimiento dentro de la misma, descartando que las deficiencias observadas puedan deberse a efectos rizosféricos más tempranos.

III.3.a Colonización rizosférica. Condiciones experimentales.

Los microorganismos endófitos, que acceden al interior de la estructura radicular, tienen como principal vía de entrada las grietas que se generan entre la raíz primaria y las raíces laterales, así como heridas que se producen en los tejidos como resultado del crecimiento de la planta (AGARWAL, 1987; Sørensen & Sessitsch, 2006; Sprent & de Faria, 1988). Por cuestiones espaciales, el acceso a la endosfera radicular requiere siempre como etapa inicial el acceso a la rizósfera, motivo por el cual decidimos abordar la búsqueda de los determinantes genéticos que se asocian específicamente a la competencia por el acceso a cada uno de los nichos anteriores, rizosférico y endofítico.

En la sección siguiente se presentará el diseño de los ensayos con los que caracterizamos la dinámica de la colonización rizosférica.

III.3.a.1 Cinética de colonización rizosférica: Título bacteriano del inóculo inicial, periodo de incubación y poblaciones de bacterias asociadas a la rizosfera.

Con el propósito de caracterizar la dinámica de la colonización rizosférica, empleando inóculos bajos realizamos inicialmente un seguimiento temporal del número de bacterias que acceden a la

rizósfera durante los primeros 11 días que siguen a la inoculación. Como inóculo para el análisis temporal de la colonización utilizamos 10^5 u.f.c./ml de *P. agglomerans* LPU12 NaI^R. Por un lado, tal concentración compara con el que número en que se encuentran algunas bacterias del suelo. Por otra parte, atendiendo a los experimentos con bibliotecas de mutantes como los que presentaré en los ensayos TnSeq (ver Capítulo V) aseguramos la presencia —en promedio— de cerca de 20 representantes para cada uno de los 4-5 mil genes que posee *P. agglomerans* LPU12. El uso de inóculos más bajos comprometería la adecuada representación de mutantes en cada uno de los genes de *P. agglomerans*. Como describimos en Materiales y Métodos (II.4.b.3), plantas de 1 días postgerminación se colocaron en macetas con vermiculita y se inocularon por inundación y posterior escurrimiento para emular la retención de líquido a “capacidad de campo”. El sistema de inoculación se eligió para conseguir la distribución homogénea del inóculo en toda la maceta. Por otra parte, la vermiculita ofreció un medio estructurado y poroso a modo de “suelo” para favorecer la estabilización de gradientes químicos derivados de la exudación de la raíz. En la misma línea de razonamiento, para favorecer la quimiotaxis bacteriana hacia la planta, el inóculo de 10^5 u.f.c./ml de *P. agglomerans* LPU12 NaI^R fue preparado en medio mineral de Fåhraeus sin nitrógeno.

A los distintos tiempos de incubación y progreso del ensayo, procedimos a cuantificar el acceso de la cepa LPU12 a la rizósfera. Dado que el entorno radicular está constituido por poblaciones de bacterias asociadas en distinto grado con la planta (ej. pobladores de la rizósfera, bacterias unidas al rizoplano con distinta firmeza, colonizadores endofíticos) (Rovira, 1965, Albareda, et al., 2006, Rodríguez-Navarro, et al., 2007), la metodología utilizada para obtener las bacterias asociadas a plantas define qué población particular es recuperada. En el trabajo previo realizados por Salas et al. (2017) en nuestro grupo de investigación, se evaluaron distintas estrategias para recuperar bacterias rizosféricas de plantas de alfalfa. A partir de esa experiencia decidimos utilizar agitación mecánica (vortex) en solución salina NaCl 0,9% durante 20 segundos para recuperar la población de bacterias “laxamente asociada” a la raíz, que debería en principio corresponder a la población rizosférica y posiblemente a también a la población laxamente unida al rizoplano (los detalles metodológicos de este procedimiento se encuentran descritos en la sección II.4.c. I de Materiales y Métodos). Según indicamos, la presencia de la cepa LPU12 en la suspensión obtenida fue cuantificada por recuento de bacterias en placas de LB suplementadas con ácido nalidíxico. Los resultados del ensayo se muestran en la Figura III.3. Según se aprecia, existe un progreso de acceso a la rizósfera cercano a la linealidad durante los primeros 2 días hasta llegar a cerca de 10^6 u.f.c. rizosféricas/planta, momento a partir del cual alcanza un *plateau* (saturación) con un leve decrecimiento. En ensayos empleando inóculos más bajos de 10^3 u.f.c./ml ó 10^4 u.f.c./ml se alcanzaron títulos finales comparables (no se muestran).

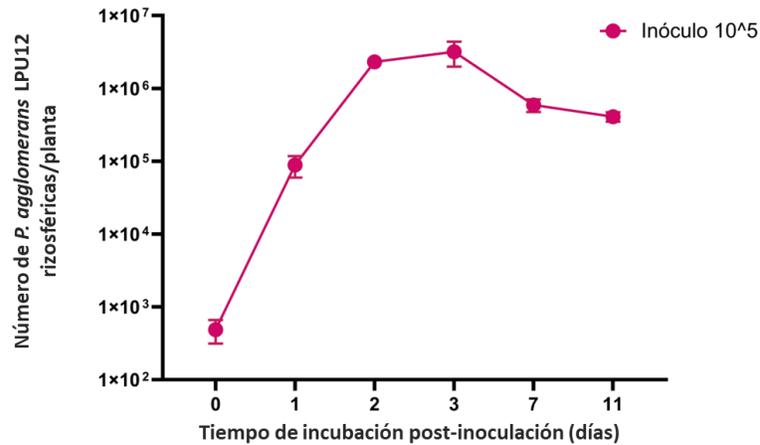


Figura III.3. Evolución temporal de la colonización rizosférica de alfalfa por parte de la cepa indicadora *P. agglomerans* LPU12 NaI^R. Plantas de alfalfa fueron inoculadas con aprox. 1×10^5 u.f.c./ml por inundación y escurrimiento (ver texto y Materiales y Métodos sección II.4.b.3). A los tiempos indicados en la figura, plantas individuales fueron descalzadas de las macetas, las bacterias rizosféricas asociadas a ellas recuperadas por lavado con solución salina NaCl 0,9% (vortex 20 segundos) y la presencia de la cepa LPU12 NaI^R fue cuantificada por recuento en placa con ácido nalidíxico. Para cada una de los tiempos indicados se procesaron al menos 4 plantas independientes. Las barras de la figura corresponden al error estándar. Ensayos similares usando inóculos de 10^3 u.f.c./ml ó 1×10^4 u.f.c./ml resultaron en densidades rizosféricas finales comparables (no se muestran).

Una dinámica cuali- y cuantitativamente similar fue observada para la colonización de raíces de alfalfa por parte de *S. meliloti* (Salas et al., 2017). La zona ascendente de la curva está posiblemente definida por un abordaje activo quimiotáctico a la rizósfera como ha sido descrito en otros sistemas (Salas et al., 2017) acompañado de crecimiento bacteriano como producto de la disponibilidad de nutrientes provenientes de la raíz. **Los resultados obtenidos indican que el recuento de bacterias asociadas a plántulas de alfalfa alcanzan el *plateau* a los 3 d.p.i., empleando inóculos cercanos a 10^5 u.f.c./ml, constituye una condición robusta para caracterizar la colonización rizosférica por parte de la cepa LPU12.** Según ya indicamos, el inóculo utilizado en estos ensayos (10^5 u.f.c./ml) debería ser compatible con la inoculación de bibliotecas transposicionales en experimentos TnSeq, de modo de asegurar la adecuada representación de mutantes en los diferentes genes de *P. agglomerans* LPU12.

III.3.b Colonización endofítica. Condiciones experimentales. Consideraciones numéricas.

Las bacterias endofitas pueden considerarse como una clase de rizobacterias especializadas que han desarrollado la capacidad de invadir las raíces de las plantas tras establecer en ellas una población rizosférica (Compant et al., 2010) o filosférica (McCully, 2001). La colonización endofítica implica un conjunto de factores ambientales y genéticos que permiten a una bacteria entrar en la endosfera de la planta (Compant et al., 2010), por lo que es de esperar que las bacterias endofitas presenten una serie de rasgos que las distinguen de aquellas que sólo se establecen en la rizosfera. Una dificultad inherente a la búsqueda de genes “puros” de colonización endofítica es que toda deficiencia en el acceso a la rizósfera será esperable que también impacte (indirectamente) en la ocupación de la endosfera radicular. Por tanto, dado que el acceso a la rizósfera es una etapa previa y necesaria para ingresar a la raíz, para cuantificar la colonización endofítica buscamos condiciones experimentales en las que no haya —o se

reduzcan tanto como sea posible— las restricciones a ocupar la rizósfera. Para ello usamos, como ya indicamos, inóculos altos del orden de 10^7 u.f.c./ml. Bajo esas condiciones experimentales, el comportamiento de mutantes para los que se observen falencias para ocupar la endosfera es esperable que pueda disociarse de eventuales defectos tempranos que acontecen en etapas anteriores, en la rizósfera o el rizoplano. El procedimiento de siembra de plantas en macetas y de inoculación de las mismas fue similar al utilizado en los ensayos de colonización rizosférica que describí en la sección precedente. La recuperación de bacterias endófitas por macerado de la raíz requirió previamente la eliminación de las bacterias asociadas externamente a la raíz (las rizosféricas y las adheridas laxamente al rizoplano) por lavado exhaustivo con agua estéril como se describe en la sección II.4.c.2 de Materiales y Métodos.

En la sección siguiente presentaré el diseño de los ensayos con los que caracterizamos la dinámica de la colonización endofítica.

III.3.b.1 Cinética de colonización endosférica: Título bacteriano del inóculo inicial, periodo de incubación y poblaciones de bacterias asociadas a la endósfera.

A partir de las consideraciones precedentes, plantas sembradas en macetas con vermiculita e inoculadas con solución de *Fåhraeus* conteniendo 10^7 u.f.c./ml de la cepa LPU12 fueron incubadas en cámara de plantas; descalzadas a los días 1, 2, 3, 4, 7, 15 y 22 postinoculación; lavadas exhaustivamente y procesadas por maceración para cuantificar por recuento en placa el número de endófitos empleando medio LB con ácido nalidíxico. En la Figura III.4A se presentan los resultados obtenidos. Se observa que ya al día 1 se alcanza un *plateau* en el nicho endofítico, con una densidad de la cepa LPU12 que osciló entre 10^5 y 10^6 u.f.c./planta. Por otra parte, en la Figura III.4B se observa que las bacterias remanentes en el último agua de lavado representan sólo 10% de la cantidad de bacterias del compartimento endofítico.

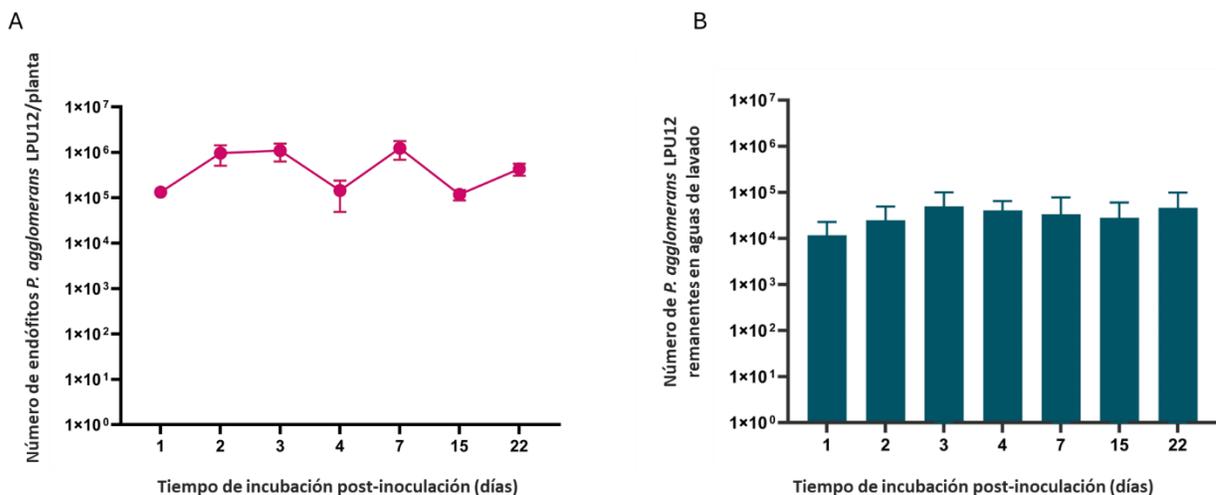


Figura III.4. Evolución temporal de la colonización endofítica de alfalfa por parte de la cepa indicadora *P. agglomerans* LPU12 NaI^R durante los primeros 22 días postinoculación. Plantas de alfalfa fueron inoculadas con aprox. 1×10^7 u.f.c./ml por inundación y escurrimiento (ver texto y Materiales y Métodos sección II.4.b.3). A los tiempos indicados en la figura, plantas individuales fueron descalzadas de las macetas, lavadas exhaustivamente y las bacterias endófitas recuperadas por maceración (Materiales y Métodos sección II.4.c.2). La presencia de la cepa LPU12 NaI^R fue cuantificada por recuento en placa con ácido nalidíxico. Para cada una de las condiciones se procesaron al menos 4 plantas independientes. Las barras de la figura corresponden al desvío estándar. **Panel A:**

Número de *P. agglomerans* LPU12 endófitas / planta a los días indicados. **Panel B:** Número de *P. agglomerans* LPU12 remanentes en el último agua de lavado a los tiempos indicados.

Dado que, por lo visto en el experimento anterior, el ingreso al interior de la planta toma lugar en una ventana temporal previa a las 24 horas postinoculación donde parecen establecerse los primeros colonizadores (“fundadores”), decidimos evaluar la cinética de colonización a tiempos más cortos previos a la saturación del número de endófitos. Para esto se repitieron las mismas condiciones experimentales que en el ensayo presentado en la Figura III.4, pero se recuperaron las bacterias del interior de la planta a 0 h, 1 h, 2 h, 3 h, 4 h, 10 h y 24 h postinoculación. En la Figura III.5 se presentan los resultados. En el Panel A se observa que los endófitos comienzan a aumentar entre las 4 y las 10 h post inoculación, con un ascenso abrupto a partir de las 10 h hasta alcanzar los valores ya observados a las 24 h en el ensayo de la sección anterior. En el Panel B, a las 24 h las bacterias remanentes luego de los lavados representan menos que 5% de las presentes en el interior de las raíces.

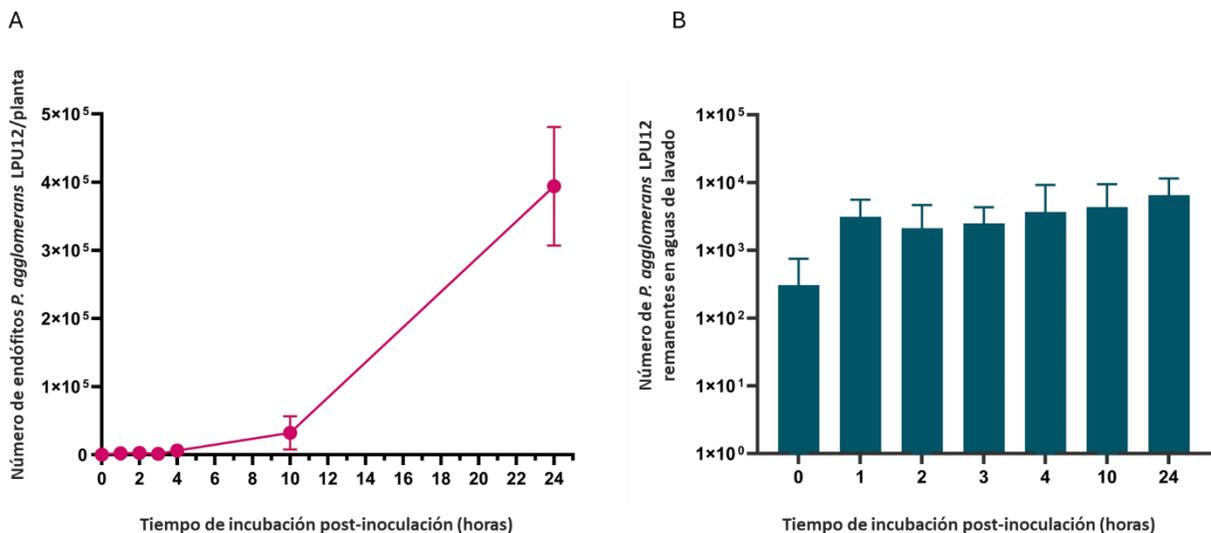


Figura III.5. Evolución temporal de la colonización endofítica de alfalfa por parte de la cepa indicadora *P. agglomerans* LPU12 NaI^R durante las primeras 24 h postinoculación. Plantas de alfalfa fueron inoculadas con aprox. 1×10^7 u.f.c./ml por inundación y escurrimiento (ver texto y Materiales y Métodos sección II.4.b.3). A los tiempos indicados en la figura, plantas individuales fueron descalzadas de las macetas, lavadas exhaustivamente y las bacterias endófitas recuperadas por maceración (Materiales y Métodos sección II.4.c.2). La presencia de la cepa LPU12 NaI^R fue cuantificada por recuento en placa con ácido nalidíxico. Para cada una de las condiciones se procesaron al menos 4 plantas independientes. Las barras de la figura corresponden al desvío estándar. **Panel A:** Número de *P. agglomerans* LPU12 endófitas / panta a las horas indicadas. **Panel B:** Número de *P. agglomerans* LPU12 remanentes en el último agua de lavado a los tiempos indicados.

Los experimentos que hemos presentado en esta sección nos permitieron definir condiciones experimentales para la recuperación exitosa de bacterias endófitas de raíz. En experimentos de TnSeq será valioso contar con la posibilidad de analizar tanto la población de endófitos tempranos (24 h postinoculación) como la que corresponde a endófitos tardíos (20 días postinoculación) que podrán ser luego comparadas. Las bacterias que hemos observado en las aguas del último lavado — en bajos números respecto de los endófitos — pueden corresponder ya sea a una fracción remanente del inóculo inicial, o a bacterias endófitas que se liberan a partir de pequeños daños mecánicos generados en la raíz durante

el mismo lavado. Si bien dichos números fueron de un orden inferior a la de los endófitos que se recuperan luego de la maceración de las raíces, en el Capítulo V —donde mostraremos el diseño de un protocolo TnSeq— presentaremos una variante de los ensayos aquí descriptos para eliminar la presencia de dichas bacterias en las aguas de lavado.

III.3.c Ingreso a la endosfera radicular. Relación entre la composición endofítica y la composición del inóculo. Estimación del número mínimo de ingresantes iniciales (“fundadores”)

La búsqueda de condiciones experimentales para el estudio de la colonización endofítica que hemos presentado en este capítulo tuvo como propósito principal establecer las condiciones experimentales para su uso en experimentos TnSeq, como los que presentaremos en detalle en los Capítulos V y VI siguientes. Según detallamos en la Introducción de esta tesis, la búsqueda por TnSeq de genes relacionados a un fenotipo particular se basa en el uso de una biblioteca de mutantes que permite, por secuenciación, contar cómo cambia la representación de cada uno de los mutantes luego de ser expuestos a una condición de interés (desafío). Aquellos mutantes que luego del desafío disminuyan su representación respecto del cambio que se observe en la cepa salvaje, serán portadores de alteraciones que afectaron su natural sobrevivencia/crecimiento en la condición que se ensayó. Por el contrario, mutantes que incrementen su representación luego del ensayo de desafío respecto de la cepa salvaje serán portadores de alteraciones que mejoran su performance en la condición ensayada. Es importante señalar que en los experimentos TnSeq tanto el inóculo como la muestra de salida deben contener un número de bacterias lo suficientemente alto como para representar bien a todos mutantes de una biblioteca (lo suficientemente diversa) preparada para cubrir toda posición funcionalmente importante del genoma. **Así, si el nicho bajo análisis restringe muy fuertemente el ingreso de las bacterias al mismo, no podrá contarse con un número adecuado de individuos en la muestra de salida de modo de conocer (con significación estadística) cómo ha cambiado la distribución de abundancia de cada mutante respecto a la que tenían los mismos en el inóculo inicial.** De existir una restricción como esa para acceder al nicho de interés (esto es, que efectivamente existe un “cuello de botella”), la misma puede representar una limitación crítica al uso exitoso de técnicas como TnSeq. En nuestro caso particular, si el número de bacterias que ingresan a la endosfera fuese extremadamente bajo, unas muy pocas unidades por planta, el tamizado (*screening*) de un número de endófitos que permita estudiar con peso estadístico a todos los genes de *P. agglomerans* (esto es, contar con un número adecuado de mutantes en cada uno de los genes) podría requerir un número de plantas desmesuradamente grande para un ensayo en condiciones de laboratorio. Resulta importante entonces evaluar si existen “cuellos de botella”, y en caso afirmativo cuantificarlos experimentalmente de modo de conocer las limitaciones del sistema y explorar posibles modificaciones para que su uso resulte informativo con peso estadístico.

Con el propósito entonces de explorar, con una primera herramienta, la estrategia de colonización de la endosfera por *P. agglomerans* LPU12, diseñamos un experimento en el que inoculamos dos variantes de la misma cepa en distintas proporciones, donde cada una de ellas era portadora de un fluoróforo diferente, eGFP y mCherry (relaciones 1:1 hasta 1:10.000 de las variantes mGFP:mCherry). Cabe señalar que, previo al experimento con plantas, verificamos que las dos variantes de *P. agglomerans* LPU12 tengan propiedades de crecimiento comparables, lo que fue efectivamente confirmado con tiempos de generación de 43 min y 46 min en medio LB a 28°C para la cepa GFP y mCherry, respectivamente (no se muestra). Si los colonizadores mCherry y eGFP ingresan en relación a su abundancia en el inóculo, esa abundancia debe coincidir con su abundancia en la población endofítica. **El objetivo del experimento fue**

ensayar distintos inóculos, y evaluar en qué rango de proporciones de las dos variantes de *Pantoea* se cumple la afirmación anterior. Podrá ocurrir que al bajar mucho la proporción de una de las variantes, y si ingresan muy pocos colonizadores iniciales (“fundadores”), sólo se encuentre como endófito la variante mayoritaria del inóculo (ingresó sólo la variante más abundante). En la Figura III.6 se presentan los resultados obtenidos al utilizar inóculos en los que la relación *mCherry*:eGFP varió en el rango 1:1 a 10.000:1.

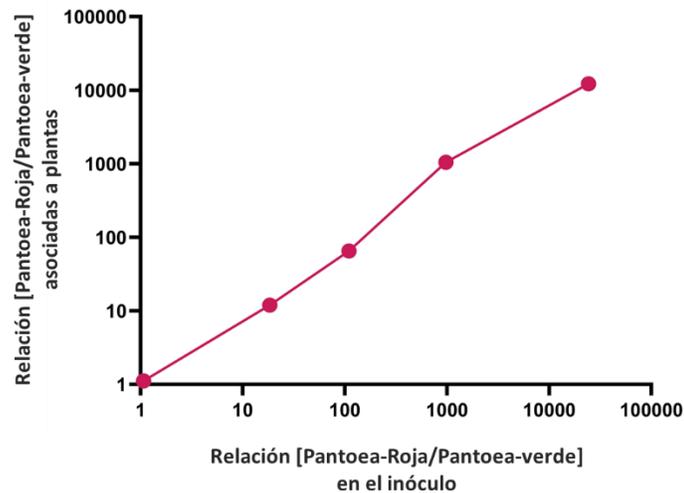


Figura III.6. Colonización endofítica de plantas de alfalfa inoculadas con diferentes proporciones de dos variantes fluorescentes de la cepa *P. agglomerans* LPU12. Para la realización del experimento de la figura se generaron dos variantes fluorescentes de *P. agglomerans* LPU12. Una de las cepas expresó eGFP a partir del plásmido pK18::eGFP que porta resistencia a kanamicina (Km^R) (Luchetti Abril, 2019), mientras la otra cepa expresó la proteína *mCherry* (Shaner et al., 2004) a partir del plásmido pHc60 que porta resistencia a tetraciclina (Tc^R) (cedido por la Dra. Lisa Kroll) (Tabla II.2 detallada en Materiales y Métodos). Las marcas anteriores permitieron distinguir ambas bacterias una de otra. Plantas de alfalfa fueron inoculadas con relaciones *mCherry*:eGFP que abarcaron el rango aproximado $10^7:10^7$ (relación 1:1) a $10^7:10^3$ (relación 10.000:1) (eje X en la figura). A las 24 h postinoculación las plantas fueron descalzadas, lavadas, los endófitos recuperados por maceración en mortero (Materiales y Métodos II.4.b.5) y la relación de las dos variantes fue obtenida luego del recuento en placas con kanamicina o tetraciclina de modo de obtener la relación *mCherry*:eGFP en el compartimento endofítico.

Los resultados muestran que el ingreso de los distintos genotipos a la endosfera radicular de alfalfa ocurre de manera proporcional a la abundancia que los mismos tienen en el inóculo, y que tal observación es válida en un amplio rango de relaciones entre las dos variantes utilizadas. Además, haber observado que la inoculación con *mCherry*:eGFP en relación 10.000:1 resulta en bacterias endofíticas que dan cuenta de esa misma relación, sugiere que el número absoluto de fundadores podría ser al menos 10^4 (como mínimo han ingresado una bacteria *mCherry* y 10.000 eGFP). Tal estimación del número de fundadores, sin embargo, debe ser considerado con precaución. En los ensayos de la sección precedente hemos observado que —aunque en mucho menor número que las endofíticas— aún existen bacterias Nal^R en las últimas aguas de lavado (ver Figuras III.4B y III.5B) que, si bien pueden ser endofíticas liberadas por daño radicular, también pueden provenir de restos de inóculo (remanente), de la rizósfera o del rizoplaneo. Desconocemos como dichas bacterias pueden afectar(modificar) la relación *mCherry*:eGFP que medimos en los macerados del experimento de la Figura III.6, especialmente cuando la cantidad de eGFP en el inóculo fue extremadamente baja. En el Capítulo VI, en el que presentaremos los resultados de los

ensayos TnSeq, estimaremos de una manera diferente el número probable de fundadores que ingresan a las raíces de cada planta de alfalfa.

III.4. Consideraciones finales del capítulo

En este capítulo hemos caracterizado la dinámica y magnitud de la colonización rizosférica y endofítica de alfalfa por parte de una cepa de *Pantoea*, que adoptamos como sistema modelo de estudio. Para la elección del género *Pantoea* y la cepa LPU12 en particular, atendimos a la presencia numérica de la cepa en plantas de alfalfa colonizadas por la microbiota endófito de semillas esterilizadas en superficie en relación a otros endófitos recuperados de las plántulas, y a las propiedades bioquímicas de la misma *in vitro* relacionadas a posibles efectos de promoción de crecimiento de plantas (López et al., 2018). Elegida la cepa, abordamos el análisis de su posición taxonómica por secuenciación genómica, pudiendo asignar la misma a la especie *P. agglomerans*. Con dicha cepa, estudiamos la dinámica de la colonización de la raíz (desarrollo en el tiempo), definiendo inóculos y tiempos de muestreo para evaluar específicamente la colonización ya sea rizosférica o endofítica. Para el caso de esta última, menos caracterizada en la literatura, pudimos además establecer que el acceso a la endosfera es linealmente dependiente de la abundancia de los genotipos que existen en el entorno radicular, y que dicha dependencia se sostiene aun para genotipos muy poco abundantes. Por último, el número mínimo de *Pantoea* que ingresan inicialmente a la endosfera pudo ser estimado (en una primera aproximación) en cerca de una decena de miles de bacterias fundadoras. Las condiciones e información que hemos colectado con los ensayos que presentamos en este capítulo sentaron las bases experimentales para el diseño de los experimentos TnSeq que describiré en los Capítulos V y VI siguientes.



CAPÍTULO IV

CONSTRUCCIÓN DE UNA BIBLIOTECA TRANSPOSICIONAL DE *P. AGGLOMERANS* LPU12 CON UN TN MARINER PARA LA BÚSQUEDA POR TNSEQ DE DETERMINANTES GENÉTICOS RELEVANTES PARA LA COLONIZACIÓN RIZOSFÉRICA Y ENDOFÍTICA DE ALFALFA



IV.1. Introducción

En el estudio de las interacciones microorganismo-hospedador se han desarrollado diversas metodologías para identificar las moléculas que participan en el proceso de comunicación y de asociación, hacia la comprensión de los mecanismos y procesos biológicos asociados a ellas. Como se mencionó en la sección I.4 de esta tesis, la secuenciación de los sitios de inserción de transposones (tecnología genéricamente designada como *Transposon Insertion Sequencing*, TIS) es una herramienta poderosa para detectar genes que influyen en la expresión de fenotipos asociativos particulares (sean simbióticos, patogénicos) (e.g., STM (Hensel et al., 1995), TraDis (Langridge et al., 2009), TnSeq (Kleckner et al., 1977)). Debido al alto rendimiento de la secuenciación de los sitios de inserción de transposones y a la sensibilidad de esta técnica (incluso para detectar mutantes con pequeñas diferencias de aptitud frente al fenotipo que se estudia), este enfoque ha resultado uno de los métodos de elección para examinar interacciones de bacterias con animales y también —aunque menos frecuentemente— con plantas (Cain et al., 2020). En el caso de plantas se han podido identificar genes implicados en la colonización y supervivencia en el nicho vegetal, así como en el establecimiento de relaciones simbióticas o perjudiciales para la especie hospedadora (Duong et al., 2018; Royet et al., 2019; Salas et al., 2017; Wheatley et al., 2020). La flexibilidad de la técnica también permite la evaluación de múltiples condiciones en un mismo ensayo, proporcionando una herramienta muy potente para el estudio de la vida asociativa. En la sección I.4 de la Introducción General de esta tesis se describieron las distintas técnicas que emplean bibliotecas de mutantes y tecnologías de secuenciación masiva para identificar genes asociados a fenotipos de interés. La técnica TnSeq, en particular, presenta la ventaja de utilizar un Tn *mariner* que —por insertarse en posiciones del genoma con dinucleótidos TA— suelen distribuirse de modo relativamente uniforme a lo largo del genoma. La metodología TnSeq también destaca por la facilidad con que pueden construirse las biblioteca de mutantes, así como y la relativamente sencilla manipulación experimental para la identificación de las regiones de inserción (detallado en las secciones I.4.b y V.1.a). Si bien la construcción de la biblioteca de mutantes por transposición con un Tn *mariner* es en principio una tarea sencilla, para ello debe asegurarse la adecuada expresión de la transposasa en la bacteria de interés para asegurar una eficiente transposición. Lamentablemente, al inicio de esta tesis no contamos con un vector portador de una transposasa de esas características que se expresara de manera adecuada en *P. agglomerans*.

En este capítulo, describiré en primer lugar la estrategia que hemos usado para adaptar el vector pSAM_RI utilizado en la mutagénesis de rizobios (Perry & Yost, 2014), para su uso como herramienta de transposicional con un Tn *mariner* en *P. agglomerans*. Mostraré a continuación como la herramienta que hemos generado puede ser empleada para la mutagénesis de otras proteobacterias de interés. Finalmente describiré las características de la biblioteca que he construido en *P. agglomerans* LPU12 que usaré en los ensayos del Capítulo VI de esta tesis.

IV.2. Construcción del vector pSAM_placO^c funcional en *P. agglomerans*

El vector pSAM fue inicialmente diseñado para estudiar el establecimiento de *Bacteroides thetaiotaomicron*, un simbiote intestinal humano, en ratones gnotobióticos de tipo salvaje e inmunodeficientes en presencia o ausencia de otros comensales intestinales humanos (Goodman et al., 2009b). Para desarrollar una herramienta similar de transposición en rizobios, distintos elementos del vector pSAM fueron utilizados por Perry & Yost (2014) para dar lugar al vector pSAM_RI, que presenta los siguientes elementos principales: **(a)** repeticiones invertidas (IRs) del Tn *mariner* que flanquean a un gen de resistencia a antibióticos que se detalla a continuación; **(b)** entre las 2 IRs anteriores, como parte del Tn, se encuentra el gen *nptII* que confiere resistencia a kanamicina/neomicina; **(c)** sitios de reconocimiento de la enzima *MmeI*, de tipo IIs, localizados a 4 pb del extremo del Tn de modo que la digestión del mismo con dicha enzima libera el Tn flanqueado por 16 pb correspondientes al ADN

genómico flanqueante; **(d)** por fuera del Tn el gen de la enzima la enzima transposasa *mariner himar1C9* (Lampe et al., 1999) bajo control del promotor *rpoD* de la cepa *Rhizobium leguminosarum* bv. *viciae* 3841, el origen de replicación del plásmido R6K, y el origen de transferencia conjugativa del plásmido RP4 (Figura VI.1.A). Lamentablemente, aunque dispusimos de los vectores pSAM y pSAM_RI, no pudimos lograr eventos de transposición con ninguno de ellos en *P. agglomerans* LPU12 sospechando que el inconveniente derivaba de la expresión inadecuada de la transposasa en una gamaproteobacteria como *Pantoea*. Ya previamente se había conseguido la transposición de un Tn *mariner* en *E. coli* (Lampe et al., 1999) e incluso en *Pantoea* (Duong et al., 2018) mediante la expresión adecuada de la transposasa. Existen reportes en los que expresando la transposasa específicamente a partir del promotor *p-lacZ* de *Escherichia coli* consiguieron transposición de un Tn *mariner* en gamaproteobacterias (Royet et al., 2019). Atendiendo a estas últimas observaciones, reemplazamos en el vector pSAMRI el promotor *rpoD* de *R. leguminosarum* por la secuencia promotora-operadora *p-lacZ-lacO^c* de *E. coli*.

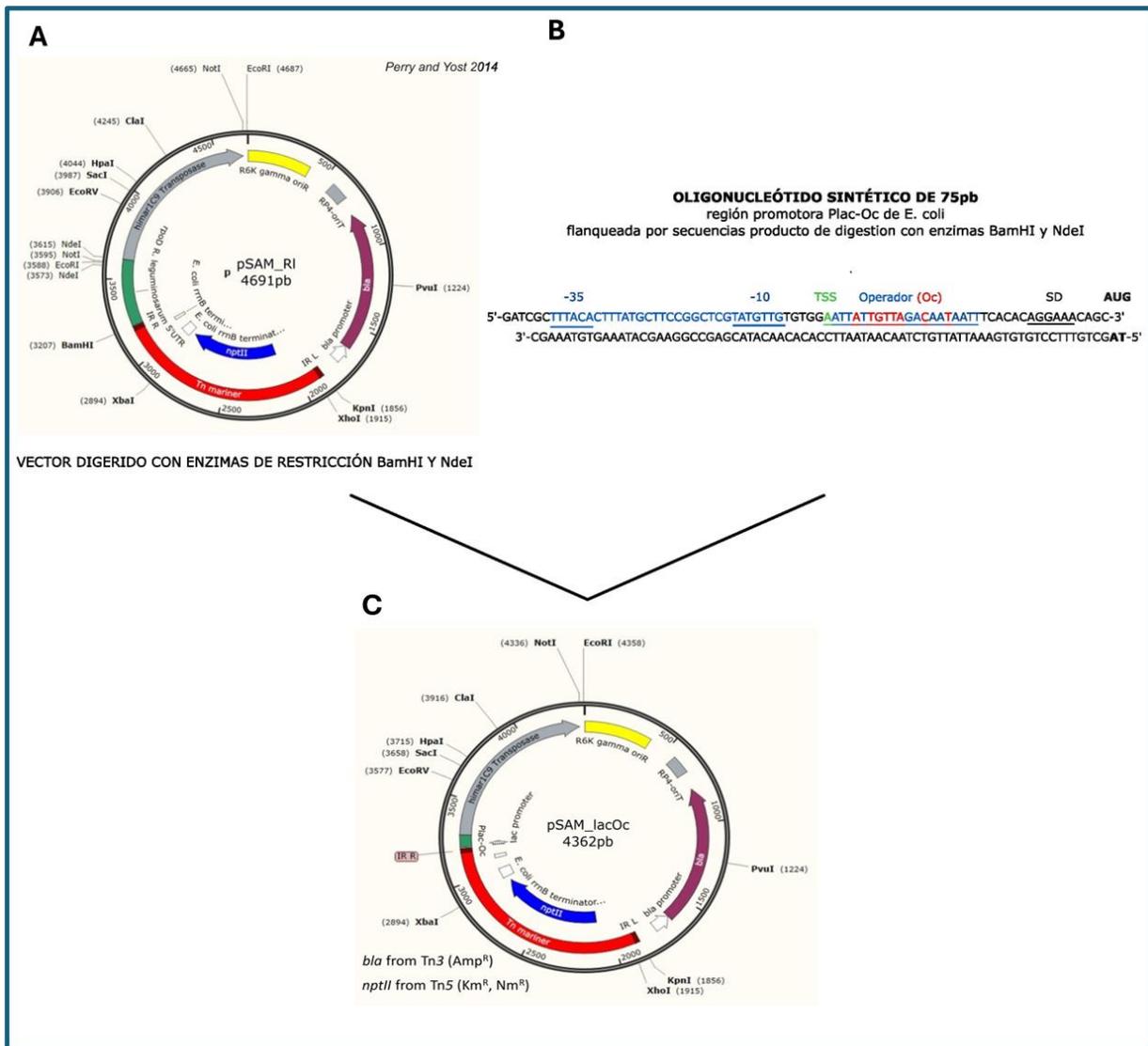


Figura IV.1. Estrategia utilizada para la construcción del vector pSAM_placOc. Panel A. Mapa del vector pSAMRI utilizado para generar mutantes transposicionales en *R. leguminosarum* y en otras alfaproteobacterias. **Panel B.** Esquema del oligonucleótido sintético que contiene la región promotora del operón *lacZ* y la zona del operador *lacO^c*

que se utilizó para controlar la transcripción de la transposasa en el vector pSAM_placO^c. Se indica en azul la región promotora, y en rojo la secuencia del operador. **Panel C.** Mapa del vector pSAM_placO^c construido en este trabajo a partir del plásmido pSAM_RI digerido con las enzimas *Bam*HI y *Nde*I al que se ligó el oligonucleótido sintético del panel B.

Utilizamos la versión *lacO^c* del operador (buscando un p-*lacZ* de expresión constitutiva) de modo de construir un vector genérico para gamaproteobacterias evitando/disminuyendo así eventuales efectos de represión asociados a la presencia de homólogos de *LacI* que puedan existir en la bacteria a ser mutagenizada (*P. agglomerans* LPU12 cuenta con ocho reguladores transcripcionales pertenecientes a la familia *LacI*). La estrategia empleada para la construcción del vector pSAM_placO^c se esquematiza en la Figura IV.1. Por un lado, se los utilizaron los oligonucleótidos sintéticos complementarios de 75 pb que se detallan en Materiales y Métodos Tabla II.5 (oligo_promotor_ *Bam*HI y oligo_promotor_ *Nde*I), y que se esquematizan en la Figura IV.1. Dichos oligonucleótidos contienen la región p-*lacZ-lacO^c* de *E. coli*, generando además en los extremos secuencias de nucleótidos protruyentes (monohebra) complementarias a las secuencias que se generan por digestión *Bam*HI y *Nde*I. El oligonucleótido de doble cadena se generó como se describe en la sección II.5.b de esta tesis para la obtención de adaptadores de doble cadena. Por otra parte, el vector pSAMRI fue digerido con *Bam*HI y *Nde*I para liberar el promotor *rpoD* de *R. leguminosarum*. Ambos fragmentos de ADN, vector y oligonucleótido de doble cadena fueron ligados, y el producto generado transformado en células de *E. coli* DH5 α . El mapa del vector pSAM_placO^c resultante con a sus características más relevantes se esquematiza en la Figura IV.1.

IV.3. Caracterización de la eficiencia transposicional en *P. agglomerans* y *S. meliloti* del Tn *mariner* asociado al plásmido pSAM_placO^c. Eficiencia transposicional en alfaproteobacterias.

El uso de una herramienta transposicional en una bacteria en particular depende, entre otras consideraciones, de la eficiencia con que el elemento móvil transpone en la especie donde se lo quiere utilizar. La construcción de una biblioteca de mutantes de buena calidad debe aspirar a generar alteraciones genómicas (por interrupción) en tantas regiones genómicas como sea posible (genes, zonas intergénicas), lo que permite luego explorar si alguna de esas interrupciones impacta en la normal expresión de algún fenotipo (esto es, en el comportamiento de la bacteria en alguna condición ambiental particular de interés). El Tn *mariner*, además de portar sitios *Mmel* en sus extremos haciéndolo muy útil para estudios TnSeq, tiene capacidad de insertarse adyacente a las secuencias TA del genoma, en principio sin otra condición que restrinja su localización. En esta sección presentamos la evaluación de la frecuencia de transposición del Tn *mariner* en *P. agglomerans* (una gamaproteobacteria) cuando se utiliza el plásmido pSAM_placO^c cuya construcción presenté en la sección precedente. Dado que la transposasa la hemos puesto bajo control del promotor del gen *lacZ*, capaz de expresarse en *S. meliloti* (una alfaproteobacteria), analizamos también cómo es la frecuencia de transposición en esta especie. Para estos ensayos se utilizó la cepa dadora *E. coli* S17-1 λ pir (Simon et al., 1983a) portadora del vector pSAM_placO^c, y se realizaron conjugaciones con las cepas receptoras *P. agglomerans* LPU12-Nal^R (este trabajo) y *S. meliloti* 2011 (Sm^R)(J. Dénarié, Francia), como se detalla en la sección II.2.c. de Materiales y Métodos. Los transconjugantes se seleccionaron en medio rico LB suplementado con kanamicina (Km) y ácido nalidíxico (Nal) en el caso de *P. agglomerans* LPU12, y en medio rico TY suplementado con neomicina (Nm) y estreptomomicina (Sm) para *S. meliloti* 2011. Las frecuencias de transposición se expresaron como la proporción de transconjugantes/número de bacterias receptoras (Tabla IV.1). En la misma Tabla IV.1 se

presentan resultados de frecuencias de trasposición para las mismas bacterias al emplear el plásmido pSAM_RI que, como puede apreciarse, no genera eventos de trasposición en *P. agglomerans* LPU12. Las frecuencias de trasposición que se muestran en la Tabla IV.1 son puras, ya que se corrigieron teniendo en cuenta los eventos de integración del vector suicida (< que 1%) calculados a partir de la proporción de clones positivos para una reacción de PCR dirigida al origen de replicación R6K de los plásmidos pSAM_placO^c y pSAM_RI. Puede observarse que la frecuencia de trasposición en *Pantoea* resultó menor a la observada en *S. meliloti* (cerca de 10⁴ veces más baja).

Tabla IV.1. Frecuencia de trasposición del Tn *mariner* en *P. agglomerans* LPU12 y en *S. meliloti* 2011 al emplear el vector pSAM_placO^c o el pSAMRI. Los valores de las frecuencias se expresan como el número de transconjugantes / célula receptora, corregido por los eventos de integración. El porcentaje de integración en ambas cepas fueron calculados a partir de la proporción de clones positivos para una reacción de PCR dirigida al origen de replicación R6K de los plásmidos, respecto a los transconjugantes totales (Km^R-Nal^R en *P. agglomerans* LPU12 y Nm^R-Sm^R en *S. meliloti* 2011). Los errores en las frecuencias de trasposición corresponden a la desviación estándar del valor medio.

Construcción	Cepa donadora	Cepa receptora	Frecuencia de trasposición	Integración
pSAM_RI	<i>E. coli</i> S17-λpir	<i>P. agglomerans</i> LPU12 Nal ^R	No se detectó trasposición	-
pSAM_RI	<i>E. coli</i> S17-λpir	<i>S. meliloti</i> 2011	6,46 ± 1,95 x 10 ⁻⁵	-
pSAM_placO ^c	<i>E. coli</i> S17-λpir	<i>P. agglomerans</i> LPU12 Nal ^R	6,75 ± 1,51 x 10 ⁻⁷	<1 %
pSAM_placO ^c	<i>E. coli</i> S17-λpir	<i>S. meliloti</i> 2011	6,46 ± 1,95 x 10 ⁻³	<1%

IV.4. Construcción de una biblioteca de mutantes transposicionales en *P. agglomerans* LPU12

Evaluada la eficiencia del vector pSAM_placO^c para obtener mutantes en la cepa *P. agglomerans* LPU12, nos enfocamos en la construcción de una biblioteca de mutantes que pueda ser utilizada en los experimentos TnSeq. La idea subyacente a la construcción de esa biblioteca fue la obtención del mayor número de mutantes independientes posible de modo que estén bien representados los diferentes genes y regiones genómicas intergénicas cuya alteración modifique el comportamiento fenotípico de colonización que queremos estudiar. Como antecedente, Duong et al. (2018) reportaron el uso de una biblioteca de *Pantoea stewartii* generados con un Tn *mariner* y constituida por aproximadamente 40.000 mutantes, que en un estudio TnSeq a partir de muestras de xilema de maíz resultó exitoso en la búsqueda de genes de patogenicidad.

Para construir nuestra biblioteca, tuvimos en cuenta la frecuencia de trasposición detallada en la Tabla IV.1 de la sección anterior, y realizamos dos conjugaciones independientes con la cepa receptora *P. agglomerans* LPU12 Nal^R y la cepa dadora *E. coli* S17 λpir portadora del vector pSAM_placO^c. Para cada conjugación se crecieron 50 ml de cultivo de cada cultivo dador y de la cepa receptora (todos a saturación), se centrifugaron y lavaron las células para eliminar restos de los antibióticos, y finalmente las bacterias de cada cultivo se resuspendieron separadamente en 25ml de medio LB fresco. Finalmente, se mezclaron 5 ml de células receptoras y 5 ml células de cada cultivo dador por separado para realizar conjugaciones independientes (ver detalles en la sección II.2.c de Materiales y Métodos). Luego de una incubación de 14 horas a 28 °C, las bacterias de cada conjugación fueron resuspendidas y los transconjugantes seleccionados en placas con medio LB suplementado con kanamicina y ácido nalidíxico. De esta forma se generaron dos bibliotecas independientes de mutantes de *P. agglomerans* LPU12 Nal^R (cada una con cerca de 78.000 transconjugantes) que fueron mezcladas para obtener la biblioteca final. En la sección siguiente

mostraremos el número de inserciones (mutantes) independientes (esto es con inserciones diferentes) que hemos conseguido generar en nuestra biblioteca transposicional. La mezcla de mutantes fue finalmente sometida a un ciclo de amplificación por crecimiento (alrededor de 5 generaciones) en medio LB líquido con kanamicina y ácido nalidíxico. Además de amplificar la biblioteca, es esperable que se haya diluido la presencia de mutantes alterados en genes letales y también la presencia de aquellos mutantes cuyas alteraciones puedan haber disminuido su crecimiento en medio nutricionalmente rico (mutantes con pérdida de *fitness* que no deseamos tener en el ensayo TnSeq de colonización de plantas). En el Capítulo VI, como parte del experimento TnSeq, presentaremos un control en el que mostraremos que mutantes de esta última clase que cambien su representación por el solo hecho de crecer en medio rico están prácticamente ausentes en nuestra biblioteca amplificada (prácticamente no encontramos mutantes con esas características). Alícuotas de la biblioteca transposicional que generamos fueron preservadas en medio LB con glicerol 20% V/V a -80°C hasta su uso.

IV.4.a Caracterización de la biblioteca de mutantes de *P. agglomerans* LPU12 mediante secuenciación masiva

En el Capítulo VI mostraremos que como parte de los ensayos TnSeq es necesario disponer del tipo y número de cada uno de los mutantes de la biblioteca que se emplea como inóculo en los ensayos con plantas. Esa información correspondiente al inóculo es luego comparada con la que corresponde a la población de mutantes que se extraen de las plantas al final del ensayo. Pueden conocerse así los genes que son relevantes para la asociación, identificando aquellos cuya representación cambia al comparar el inóculo inicial y muestra de salida. **Caracterizar la biblioteca implica, entonces, conocer en detalle los mutantes que conforman la misma, así como el número (abundancia) de cada uno de ellos.** El procedimiento experimental para la obtención de dicha información (preparación del DNA de la mezcla de mutantes, digestión, ligación de adaptadores, amplificación por PCR de las zonas de unión Tn-genoma, secuenciación de las mismas, recuento y mapeo informático de cada una de ellas en el genoma de *P. agglomerans* LPU12) se detalla en el capítulo siguiente. Según comentamos en la Introducción de esta tesis (sección I.4) el Tn *mariner* se inserta en secuencias TA del genoma. Según se muestra en la Tabla IV.2, en el genoma de *P. agglomerans* LPU 12 pueden identificarse 173.478 secuencias TA, de las cuales 135.299 (77,99%) se encuentran dentro de regiones codificantes. En la Figura IV.2 se presenta, para cada uno de los replicones de la cepa *P. agglomerans* LPU12 (cromosoma y sus tres plásmidos), cómo es la distribución de inserciones en los sitios TA de las respectivas secuencias, y cuál es la representación relativa de cada inserción estimada a través del número de lecturas que se tiene para cada una de ellas. Como puede apreciarse, en todos los replicones existe una distribución homogénea de inserciones todo a lo largo de los mismos, sin regiones apreciables donde no haya ocurrido transposición. Se observa, además, que en el replicón cromosomal hay una mayor abundancia de inserciones a medida que nos aproximamos al origen de replicación (inicio y final del eje X en el gráfico), debido seguramente al mayor número de copias físicas en esa zona. Considerando el número de lecturas en el centro del replicón respecto de las observadas en las cercanías del ori, posiblemente exista un promedio de 2 horquillas de replicación activas/cada cromosoma. Si el replicón cromosomal con cerca de 5 Mpb sólo requiere 2 horquillas para cumplimentar su duplicación en un ciclo de división (40-50 min a 28°C), los plásmidos pPagLPU1, _2 y _3 —todos de menos de 0,5 Mpb— deben replicarse posiblemente en una fracción del tiempo de división celular persistiendo la mayor parte del tiempo con un mismo número de copias a lo largo de toda su secuencia, hecho este que debe necesariamente traducirse en un número homogéneo de lecturas a lo largo de toda su secuencia como se observa en los tres paneles plasmídicos de la Figura IV.2.

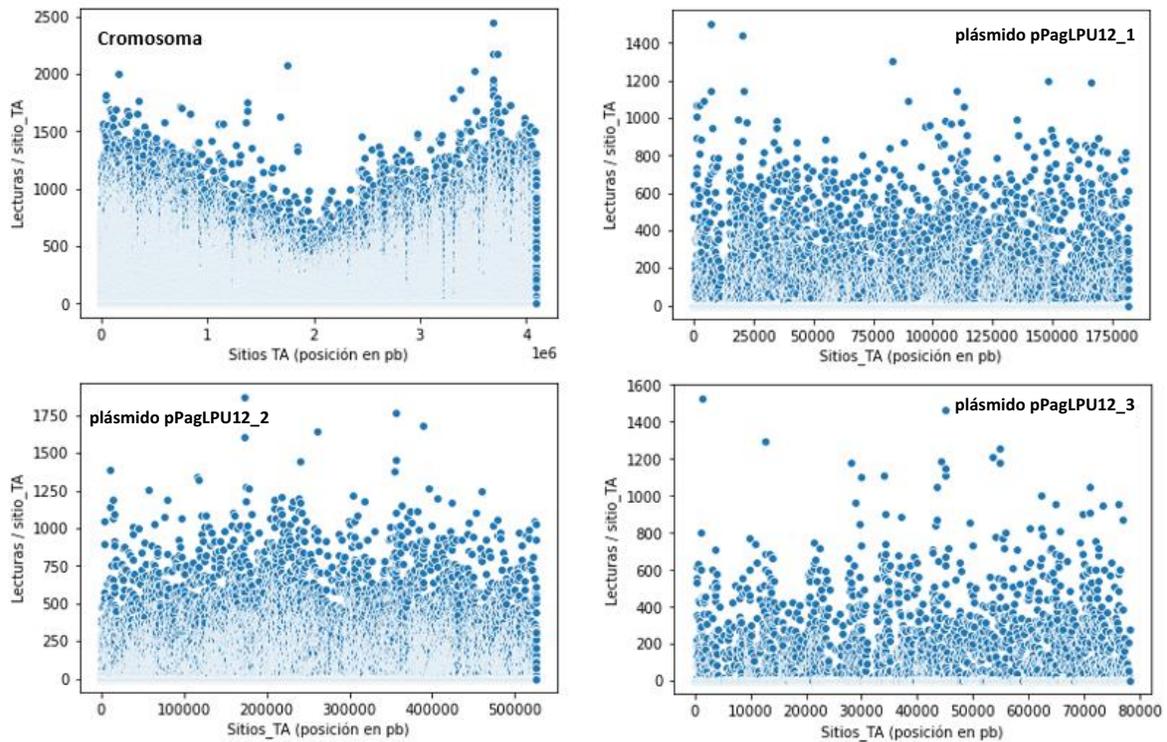


Figura IV.2. Distribución y número de relativo de inserciones del Tn *mariner* en cada uno de los replicones de la cepa *P. agglomerans* LPU12 en la biblioteca transposicional que hemos generado para los ensayos TnSeq que presentaremos en el Capítulo VI. Para cada uno de los replicones se indica en el eje X la posición de cada sitio TA en pares de bases. En el eje Y se indica el número de lecturas correspondiente a cada sitio (normalizado por replicón).

En la Tabla IV.2 se aprecia que, del total de sitios TA, en nuestra biblioteca hemos podido mutar 51.883 de los cuales 39.917 se encuentran dentro de genes, indicando por tanto que cerca de 30 % de los sitios TA disponibles pudieron ser mutados. Tal porcentaje indica la densidad de sitios que fueron efectivamente mutados respecto al total posible. Además, de la secuenciación de la biblioteca y del análisis de cada una de las inserciones surge que, del total de 4.598 genes, el número de genes que se identificaron mutados por secuenciación fue 4.076, el 88,51% del total. El 11,5% que no hemos observado entre los genes mutados se condice con las observaciones de otros autores donde cerca de 10% del total del genoma está asociado a funciones esenciales (Christen et al., 2011). Teniendo entonces en cuenta el número de inserciones dentro de genes (39.917) y el número de genes que hemos podido mutar (4.076), nuestra biblioteca debería incluir en promedio cerca de 9-10 inserciones independientes (mutantes)/gen.

Tabla IV.2. Caracterización de la biblioteca de mutantes de *P. agglomerans* LPU12::Tn *mariner*. En la tabla se indica el número total de sitios TA (totales y dentro de genes), el número de sitios TA efectivamente mutados (totales y dentro de genes). Con dichos valores se calculó la densidad porcentual de mutaciones efectivamente alcanzadas (sitios mutados/sitios totales mutables).

Nro. De sitios TA con inserción (de 173478 sitios TA totales; 135299 sitios TA en genes)					
<i>Replica 1</i>	<i>Replica 2</i>	<i>Replica 3</i>	<i>Replica 4</i>	Promedio	
52.316	51.887	55.243	48.086	51.883	Total
40.153	39.964	42.558	36.995	39.917	En genes
Densidad (%) de inserciones en sitios TA en cada réplica de la biblioteca (Inóculo)					
<i>Densidad (%) de inserciones en todos los sitios</i>					
30,16	29,91	31,84	27,72	Promedio	29,90
				DS	0,85
				%	2.83
<i>Densidad (%) de inserciones en genes</i>					
29,68	29,54	31,45	27,34	Promedio	29,50
				DS	0,84
				%	2.85

IV.5. Consideraciones finales del capítulo

En este capítulo presentamos la construcción de un vector, pSAM_placO^c, que pudo ser empleado como herramienta para la mutagénesis con un Tn *mariner* de *P. agglomerans* LPU12. El vector pSAM_placO^c que generamos, derivado del vector pSAMRI, resultó además eficiente para generar mutantes en *S. meliloti*, una alfaproteobacteria.

Con el vector pSAM_placO^c, construimos una biblioteca transposicional en *P. agglomerans* LPU12 que incluyó más de 50 mil inserciones independientes, distribuidas de forma homogénea a lo largo de los distintos replicones, con un estimado de 9-10 inserciones independientes (en promedio) /gen. La biblioteca que hemos generado la emplearemos en los ensayos TnSeq que describiré en el Capítulo VI de esta tesis para estudiar la genética de la colonización de plantas por parte de *P. agglomerans* LPU12.



CAPÍTULO V

DISEÑO E IMPLEMENTACIÓN DE UN ENSAYO TNSEQ PARA LA BÚSQUEDA E IDENTIFICACIÓN DE GENES DE *P. AGGLOMERANS* LPU12 ASOCIADOS A LA COLONIZACIÓN RIZOSFÉRICA Y ENDOFÍTICA DE RAÍCES DE *MEDICAGO SATIVA* (ALFALFA).



V.1. Introducción

La posibilidad de acceder al estudio de funciones celulares a través de la caracterización de mutantes, esto es alterando la información genética, ha sido una herramienta empleada desde hace décadas y hasta nuestros días, comenzando con los estudios en que se analizaron las primeras rutas metabólicas (Beadle & Tatum, 1941; Falk, 2010; Lacroute, 1975). Asociadas inicialmente al uso de microarreglos (hace unos 15 años) y más recientemente a las técnicas de secuenciación de alta capacidad, se han desarrollado nuevas metodologías de análisis de mutantes que, en bacterias, permiten en un sólo ensayo evaluar las consecuencias fenotípicas (de cada una) de decenas a cientos de miles de mutaciones (para revisión ver (Cain et al., 2020)). Estas metodologías que evalúan fenotipos a escala ómica, utilizan bibliotecas de mutantes transposicionales, y son designadas hoy de modo genérico como aproximaciones TIS (por *Transposon Insertion Sequencing*) (van Opijnen & Camilli, 2013). Así, las aproximaciones TIS permiten, en bacterias, la identificación de aquellos cambios genéticos que afectan (de forma negativa o positiva) la aptitud (*fitness*) para enfrentar una dada condición (desafío) ambiental.

En relación a nuestros intereses particulares, las técnicas TIS pueden utilizarse para investigar el conjunto de genes necesarios para la colonización de plantas, y el posterior crecimiento/persistencia de las bacterias en ese entorno (Cole et al., 2017). El empleo de bibliotecas de transposición a saturación (esto es, donde la densidad de mutaciones de la biblioteca incluya mutantes en todas aquellas regiones —génicas o extragénicas— cuya alteración se asocie a cambios en el fenotipo de interés) es una herramienta muy potente para el estudio a escala genómica de fenotipos complejos. Las aproximaciones basadas en TIS se pueden utilizar en bacterias de importancia agrícola, tanto en estudios *in vitro* como *in vivo*; para identificar genes, funciones y rutas metabólicas implicadas en la tolerancia a estreses y en sus distintas interacciones con plantas (Fabian et al., 2020). Es importante resaltar que las distintas variantes de experimentos basados en TIS se fundamentan en el recuento (por secuenciación) de cada forma mutante en el inóculo inicial, y luego, en la muestra de salida del ensayo. Así, para un fenotipo dado que se analice, mutantes con deterioro en su comportamiento respecto de la bacteria salvaje tendrán menor representación (recuento) en la muestra de salida respecto de la representación que tenían en el inóculo. Cuando, por el contrario, se trate de mutantes que redundan en mejoras fenotípicas, sus recuentos a la salida serán mayores que los recuentos de esos mismos mutantes en el inóculo. Mutaciones neutras no resultarán en cambios de representación al comprar sus recuentos al ingreso y egreso del ensayo. Resulta entonces importante señalar que, tanto en el inóculo como en la mezcla de salida debe garantizarse que la cantidad de bacterias sea suficiente para que cada una de las variantes mutantes se encuentren en número suficiente para asegurar su adecuado procesamiento y contaje para inferir (con peso estadístico) si como consecuencia del desafío han existido, o no, cambios entre las representaciones inicial y final.

En este capítulo, describiremos el diseño, puesta a punto y uso de un ensayo TIS aplicado a la identificación de los determinantes genéticos que usa una bacteria endófito ubicua de plantas, como *P. agglomerans*, para asociarse a las raíces de alfafa. En el diseño del ensayo TIS pusimos especial cuidado para que el mismo sea técnicamente sencillo y compatible con las condiciones de colonización rizosférica y colonización endofítica que describimos en el Capítulo III de esta tesis (concentración de los inóculos, tiempos de los ensayos, recuperación de bacterias). Los experimentos de colonización fueron realizados, según mostraremos, con la biblioteca de mutantes transposicionales cuyas características detallamos en el Capítulo IV precedente.

V.2. Diseño de un ensayo TnSeq para la búsqueda de marcadores genéticos de *P. agglomerans* asociados a la colonización exo- y endofítica de plantas de alfalfa: Procedimiento general (adaptador, cebadores, etiquetado) para la identificación y conteo de mutantes.

V.2.a Estructura general del ensayo. Muestras, etapas, diseño de adaptador y reacciones de PCR. Muestra para secuenciación.

En la Figura V.1 se presenta, de modo esquemático, el diseño y estructura del ensayo TnSeq que implementamos para estudiar la colonización rizosférica y endofítica de raíces de *M. sativa* (alfalfa) por parte de *P. agglomerans* LPU12.

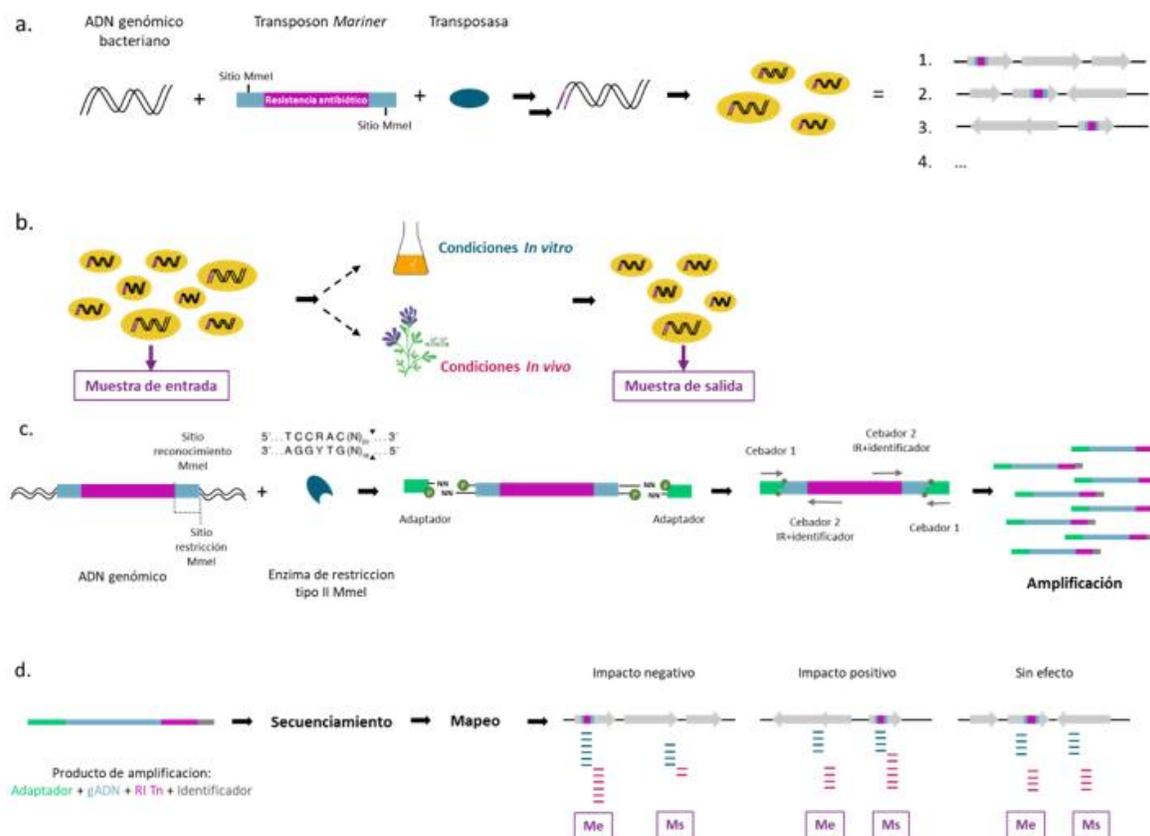


Figura V.1. Esquema del ensayo TnSeq diseñado para la búsqueda de genes de *P. agglomerans* LPU12 vinculados a la colonización radicular de alfalfa. **Panel A:** Construcción de la biblioteca de mutantes mediante la inserción aleatoria de un Tn *mariner* en regiones TA del genoma de *P. agglomerans* LPU12 (Capítulo IV precedente). **Panel B:** Evaluación (desafío) de la biblioteca en las condiciones de interés *in vitro* e *in vivo* (ej. crecimiento en medio mínimo, crecimiento en medio rico, colonización de plantas). Colección de las muestras de entrada (inóculo) y de salida para su procesamiento. **Panel C:** Preparación de ADN genómico de las muestras de entrada y salida, digestión con la enzima *MmeI*, ligación de adaptador, amplificación por PCR de las zonas de unión Tn-genoma (para detalle de los sitios de digestión en los extremos del Tn *mariner*, secuencia del adaptador y cebadores empleados ver Figura V.2).

Panel D: Flujo de trabajo para la secuenciación de los amplicones obtenidos, mapeo y conteo (abundancia) de cada inserción. Ejemplos para distintas inserciones, mostrando variaciones en la representación entre la muestra de entrada (Me) y la muestra de salida (Ms) [esto es impacto negativo ó impacto positivo de la mutación], o con ausencia de cambios [mutaciones neutras, sin impacto en el fenotipo que se analiza].

Brevemente, una vez generada la biblioteca de mutantes como indicamos en el Capítulo IV (Figura V.1., Panel A), la misma fue inoculada sobre plantas de alfalfa (muestra de entrada) para estudiar la colonización de la rizósfera y de la endosfera (Panel B), recuperando al final de los ensayos las muestras de salida. De las muestras de entrada y salida se extrajo DNA genómico que fue digerido con la enzima de restricción tipo IIs *MmeI*. El Tn *mariner* posee un sitio de reconocimiento para dicha enzima cerca de cada uno de sus extremos. El reconocimiento de esos sitios genera cortes a 18 y 20 pb de distancia (a veces a 19 y 21 pb) generando extremos 5'-P y 3', respectivamente, en dirección al genoma. Esos cortes dejan por fuera del Tn 16 (a veces 17) pb de secuencia genómica (Panel C) con dos bases protruyentes (NN) en el extremo 3'. La identidad de dichas bases depende, por tanto, de la secuencia genómica dónde se haya insertado el Tn. Los productos de las digestiones genómicas anteriores fueron ligados a un adaptador de doble cadena que contiene uno de sus extremos como, y el otro con 2 bases protruyentes 3' de secuencia aleatoria (esto es, con las 12 terminaciones posibles: AG-3', AC-3', AT-3', GA-3', GC-3', GT-3', CA-3', CG-3', CT-3', TA-3', TG-3', TC-3'). Existen así 12 versiones del adaptador que podrán ligarse a los distintos extremos 3' protruyentes que se generan con la digestión *MmeI* del ADN genómico. En la Figura V.2 se presentan en detalle la secuencia del adaptador (en azul, ver variantes -NN-3'), la secuencia del extremo del Tn (en rojo), y la secuencia genómica contenida entre ellos (en verde). Cabe señalar que el extremo 5' anidado del adaptador se utilizó fosforilado para conseguir su ligación al extremo 3' del ADN genómico digerido.



Figure V.2. Esquema de la estrategia de corte de los extremos del Tn *mariner*, ligación de adaptador, y amplificación de las regiones genómicas flanqueantes. Se muestran en rojo las secuencias IR-L e IR-R del Tn *mariner* (subrayada la secuencia de reconocimiento de la enzima *MmeI*) y en azul los oligonucleótidos que forman el adaptador (NN corresponde a dinucleótidos de secuencia al azar generadas para hibridar con las distintas secuencias que resultan de las digestiones genómicas). En naranja se muestran los cebadores que se utilizaron en las reacciones de PCR: Adapter-Top-PCR e IR_Tag-n (ver texto). La secuencia en verde y en letras minúsculas corresponden a un ejemplo arbitrario de secuencia genómica flanqueante del Tn *mariner*, y constituye la porción informativa del amplicón que luego de la secuenciación se utiliza para el mapeo de la inserción. Dado que la enzima *MmeI* puede cortar a 20 o a 21 pb de su sitio de reconocimiento, la secuencia genómica puede ser de 16 o de 17 pb (en el ejemplo es de 16 pb), dando lugar a productos de amplificación que serán de 65 o 66 pares de bases, respectivamente. N1 a N4 corresponden a secuencias destinadas a etiquetar cada uno de los amplicones, para luego de la secuenciación reconocer a que tratamiento y réplica corresponde cada uno (ej. inóculo-replica 2, rizósfera-réplica 1, compartimento endófito-replica 4, etc.).

Las regiones de unión Tn-genoma fueron luego amplificadas por PCR empleando dos cebadores (en marrón), uno de ellos que hibridó con secuencias del adaptador (cebador Adapter-Top-PCR: 5'-

AAAGGACGTGGTTTACGGGGC-3') mientras el otro lo hizo con una secuencia cercana al extremo del Tn *mariner* (cebador IR_Tag-n: **5'-AN₁N₂N₃N₄CGGGGACTTATCATCCAACCTG-3'**). Este último cebador, posee en su extremo 5' una secuencia **5'-AN₁N₂N₃N₄** (Tag-1, Tag-2, etc.) con la que se pueden etiquetar los amplicones de modo de asociar cada uno de ellos con una condición experimental y réplica específica (ej. crecimiento en medio mínimo réplica 4, colonización rizosférica réplica 2, colonización endofítica replica 1, etc.).

Los productos de amplificación etiquetados fueron mezclados en condiciones aproximadamente equimolares y secuenciados por *paired end* empleando una plataforma *Illumina* (en nuestro caso NextSeq2000 del servicio provisto por el CeBiTec, Universidad de Bielefeld, Alemania). Para la secuenciación se atendió a la necesidad de acceder a un alto número lecturas de productos cortos, de 65-66 pb. El conjunto de secuencias fue procesado empleando *scripts* en *Python* y el programa *Transit* (DeJesus et al., 2015) de acceso libre para análisis de datos TnSeq. Brevemente: **(a)** Las lecturas fueron inicialmente clasificadas de acuerdo a su etiqueta (*barcode*) de modo de agruparlas por condición (entrada-inóculo o muestra de salida) y número de réplica, **(b)** De cada una de las lecturas se extrajo su correspondiente secuencia informativa de 16/17 pb (*trimming*), **(c)** Las lecturas así procesadas fueron mapeadas sobre cada una de los genes de *P. agglomerans* LPU12, y **(d)** A partir del recuento del número de inserciones/gen en la muestra de entrada y las muestras de salida se calcularon los valores del parámetro M^{TnSeq} (*fold change*) siguiente para cada uno de los genes anotados:

$$M^{\text{TnSeq}}_{(\text{gen-x})} = \log_2 \left\{ \frac{[\text{lecturas en gen-x}]^{\text{muestra de salida}}}{[\text{lecturas en gen-x}]^{\text{muestra de entrada, t=0}}} \right\}$$

Valores de M^{TnSeq} negativos y positivos se asociaron a genes que perturban o favorecen la colonización rizosférica/endofítica, respectivamente. Valores de $M^{\text{TnSeq}} = 0$ se asociaron a genes sin relevancia para los fenotipos estudiados.

V.2.b Optimización de las condiciones experimentales del ensayo TnSeq. Consideraciones prácticas experimentales.

En las secciones siguientes presentaré el proceso de optimización de cada una de las etapas que se describen en los Paneles B y C de la Figura V.1. Para la puesta a punto de los protocolos de trabajo usamos un mutante Tn *mariner* escogido al azar de nuestra biblioteca transposicional. Describiré las consideraciones prácticas más relevantes que tuvimos en cuenta para la preparación de ADN genómico, optimización de las reacciones de digestión, ligación y qPCR previo a la secuenciación.

V.2.b.1 Amplificación de muestras y cantidades de ADN genómico a procesar, ligación y concentraciones de adaptador.

Como ya señalamos, y se muestra en la Figura V.1 panel D, la identificación y cuantificación de las regiones genómicas interrumpidas por el Tn *mariner* se realizó mediante el procesamiento del ADN extraído de las bacterias recuperadas de cada tratamiento y su posterior secuenciación. El procesamiento de cantidades adecuadas de ADN genómico extraído de cada muestra es esencial para garantizar la detección de aquellos mutantes que están en menor representación. En trabajos reportados previamente donde se utiliza la técnica TnSeq, el material genético de partida fue del orden de 4 µg a 50 µg (Arnold et al., 2017; Duong et al., 2018; Royet et al., 2019). Como se describió en el Capítulo III, en nuestro caso la cantidad de bacterias recuperadas tanto de la rizósfera como de la endósfera es usualmente cercana a 10^5 - 10^6 u.f.c./planta, una cantidad insuficiente para preparar ADN genómico en cantidades cercanas a las decenas de microgramos como las que se necesitan procesar. En nuestros ensayos de colonización de

raíces (sea rizosférica o endosférica) empleamos 15 plantas en cada una de las cuatro réplicas/tratamiento. Si bien aumentar la cantidad de plantas por réplica, por ejemplo 10 veces, hubiera permitido aproximarse a 10^7 u.f.c./condición, el material genómico que hubiera podido extraerse de ellas hubiese siendo aún insuficiente para implementar de manera adecuada las etapas de digestión, ligación del adaptador y amplificación de las regiones Tn *Mariner*-genoma. Para resolver esta limitación, se optó por enriquecer cada una de las muestras extraídas de plantas (rizósfera y endosfera) por crecimiento en 50 mL de medio líquido LB suplementado con los antibióticos kanamicina y ácido nalidíxico hasta alcanzar una OD^{600} cercana 1. A partir de muestras de esos cultivos pudimos obtener varias decenas de microgramos de ADN genómico (kit comercial NucleoSpin Microbial DNA, descrito en Materiales y Métodos, sección II.3.b). Dicho procedimiento de enriquecimiento ha sido reportado por Cole et al. (2017) sin alterar significativamente la representación relativa de cada mutante en la mezcla original. De haber ocurrido, el enriquecimiento por crecimiento hubiese interferido con la identificación de mutantes de interés con la herramienta TnSeq. Habiendo preparado suficiente cantidad de ADN genómico de cada una de las muestras, realizamos para cada una de ellas la digestión con *MmeI*, y luego la ligación con el adaptador generado por hibridación de los oligonucleótidos Adapter-Top y Adapter-Bottom-P cuyas secuencias se describen en la Tabla II.5 de Materiales y Métodos. Las reacciones de ligación realizadas en 100 μ l incluyeron, además de ligasa y su tampón correspondiente, 4 μ g de ADN digerido con *MmeI* y una concentración final 4 μ M de adaptadores. De ese modo, la relación molar entre el número de extremos *MmeI* en el ADN digerido (calculado en base al número estimado de sitios en el genoma) respecto al número de adaptadores presentes fue siempre cercana a 1:100 para favorecer el encuentro de los extremos digeridos con un adaptador. Luego de una noche de incubación a 16°C, se removió el exceso de adaptadores (columnas kit comercial *PCR GeneJET purification* - ThermoFisher Scientific) y por medida de absorbancia a 260 nm se estimó la cantidad de ADN presente en la mezcla. Con dicho ADN como molde procedimos a realizar reacciones de qPCR según el protocolo que detallo a continuación.

V.2.b.2 Condiciones de la qPCR. Cantidades de molde.

Estimación de la mínima cantidad de ADN molde de la ligación para asegurar un número de blancos de amplificación que permitan la detección de todos los mutantes presentes en la biblioteca.

Para poner a punto la cantidad de molde en las reacciones de qPCR que usamos para amplificar las regiones genómicas adyacentes a las inserciones al Tn *mariner*, realizamos en primer lugar distintas mezclas de reacción en las que variamos la cantidad de molde de la mezcla de ligación incorporada a cada tubo. La cantidad de molde/tubo varió en el rango: 6,25 ng; 3,13 ng; 1,56 ng; 0,78; 0,39 ng y 0,20 ng (molde en diluciones seriadas 1:2). En las reacciones de qPCR se emplearon los cebadores Adapter-Top-PCR e IR_Tag-1 (Figura V.2 y Materiales y Métodos Tabla II.5) y el reactivo comercial “Takyon™ No ROX SYBR 2X MasterMix blue dTTP” de Eurogentec según las condiciones que se detallan en Materiales y Métodos sección II.3.e. Como puede verse en la Figura V.3, Panel D, al menos en el rango 6,25 ng/tubo a 0,098 ng/tubo, la qPCR mostró —como era esperable— una relación lineal entre los valores de Ct (*threshold cycle*, “umbral de ciclo”) y el logaritmo de la cantidad de molde utilizada. Luego, con el propósito de establecer a qué número de copias corresponde cada una de las cantidades de ADN anteriores, realizamos reacciones de qPCR utilizando un molde sintético con la secuencia que se muestra en la Figura V.2 (5′–AAAGGACGTGGTTACGGGGCGcgtcctcggtacctaACAGGTTGGATGATAAGTCCCCG-3′), incorporando el siguientes número de copias/tubo: 10^9 , 10^8 , 10^7 , 10^6 , 10^5 , 10^4 , 10^3 , 10^2 y 10. Los resultados obtenidos se muestran en la Figura V.3, Paneles A y C. Nuevamente se observa una relación lineal entre

los valores de Ct y el logaritmo de la cantidad de molde sintético utilizada (molde en diluciones seriadas 1:10). Con estos resultados, usando los datos del Panel C para conocer a qué números de copias corresponde cada uno de los Ct del Panel D, en el Panel B se muestra (en recuadros) la relación entre los ng de molde y el número de copias al que corresponde cada uno de ellos (ej. 6,23 ng incluyen $5,81 \times 10^5$ copias; 3,13 ng incluyen $2,91 \times 10^5$ copias; 1,56 ng incluyen $1,45 \times 10^5$ copias; etc.). La estimación por cálculo del número de copias Tn-genoma (inserciones)/ng de ADN molde teniendo en cuenta la A_{260} nm y el tamaño del genoma de *P. agglomerans* LPU12, resultó en valores 35% más elevados que los obtenidos por qPCR, mostrando un sesgo (aunque no muy grande) a sobreestimar el número de copias presente. Ese sesgo de la sobreestimación anterior por cálculo deriva seguramente de al menos dos causas: (a) por un lado que no todo el ADN genómico ha ligado el adaptador, y (b) a una muy probable sobre-estimación de la cantidad de ADN genómico al medir la A_{260} nm debido a una incompleta remoción de los adaptadores.

La importancia de contar con los resultados de la qPCR —vinculando masa (ng) de ADN con número de copias— se relaciona con la necesidad de conocer cuál es la cantidad mínima de molde (en masa) que debemos emplear en nuestras reacciones de amplificación de modo de contar con un número adecuado de representantes de cada una de las inserciones que existen en nuestra biblioteca de mutantes. Según presentamos en el Capítulo IV, Tabla IV.2, la biblioteca transposicional incluye 51.883 inserciones independientes. Si empleamos, por ejemplo, 3,13 ng de molde equivalente a $2,91 \times 10^5$ copias de molde, tendríamos más de 5 copias/inserción. Y si consideramos que de todos los genes que hemos podido mutar contamos con cerca de 9 inserciones/gen, si en las qPCR usamos 3,13 ng tendríamos una oferta de molde (en promedio) cercana a las 45 copias de molde/gen.

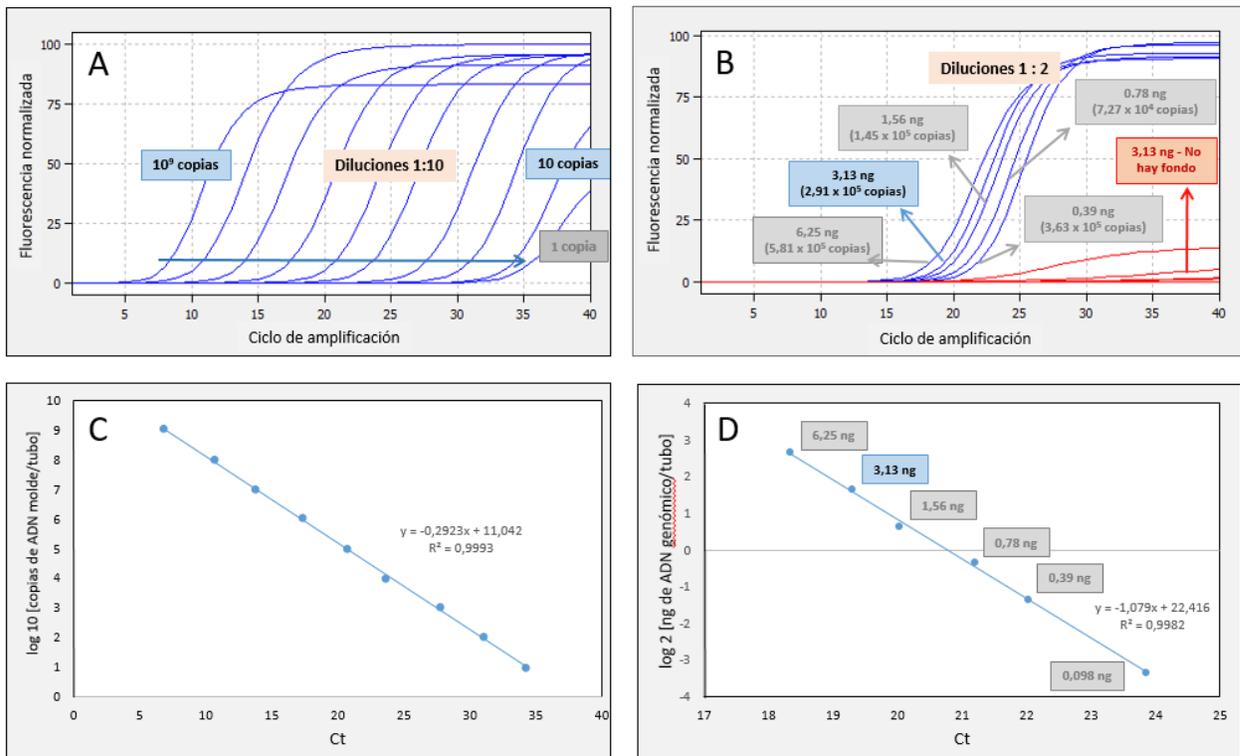


Figura V.3. Resultado de las reacciones de qPCR empleadas en el experimento TnSeq para la amplificación y cuantificación de cada una de las regiones de unión (inserciones) Tn *mariner* - genoma de *P. agglomerans* LPU12.

Panel A: Amplificación de diluciones seriadas (1:10) del oligonucleótido sintético 5'-AAAGGACGTGGTTTACGGGGC GcgtcctcggtacctaACAGGTTGGATGATAAGTCCCCG-3' que emula una secuencia arbitraria [adaptador-Tn-genoma-adaptador] como la que se esquematiza en la Figura V.2. **Panel B:** Amplificación de diluciones seriadas (1:2) de ADN genómico extraído de un mutante Tn *mariner*, digerido con *MmeI* y ligado al adaptador. **Panel C:** Gráfico de la dependencia lineal entre el Ct (*threshold cycle*, “umbral de ciclo”) y el logaritmo de las cantidades de molde (oligonucleótido sintético) que se muestran en el Panel A. **Panel D:** Gráfico de la dependencia lineal entre el Ct y el logaritmo de las cantidades de molde (ADN genómico digerido con *MmeI* y ligado al adaptador) que se muestran en el Panel B.

En esta sección hemos presentado el desarrollo de una qPCR que nos permitió amplificar de manera cuantitativa y con alta sensibilidad (hasta al menos 10 copias de molde, Figura V.3, paneles A y C) las regiones de inserción del Tn *mariner*. Además, al contar con medidas de fluorescencia que reflejan la cantidad de producto generado las mismas podrán usarse para mezclar en cantidades equimolares los distintos productos de amplificación etiquetados previo a su secuenciación como mostraremos en la sección V.3.b.3.

V.3. Implementación del ensayo TnSeq con la biblioteca de mutantes *P. agglomerans* LPU12::Tn *mariner*: Tratamientos, resultados de las amplificaciones, confirmación de secuencias en amplicones testigo. Muestra para secuenciación de alta capacidad.

V.3.a Tratamientos y réplicas. Muestra de entrada y muestras de salida.

Como se indicó en los objetivos del presente trabajo de tesis, enfocamos —a escala genómica— la búsqueda de genes (y funciones) que en bacterias estén vinculados a la colonización externa (rizosférica) e interna (endofítica) de raíces. Empleando el sistema modelo que ya hemos caracterizado en el capítulo III, en la Figura V.4 se presentó el diseño general del ensayo TnSeq con el que abordamos la búsqueda de los siguientes grupos génicos: **(A) genes esenciales** (cuya mutación resulte en un fenotipo letal); **(B) genes de auxotrofia** (por análisis del crecimiento de la biblioteca en medio mínimo M9); **(C) genes cuya mutación resulte en modificaciones de la aptitud (*fitness*) para crecer en medio sin limitaciones nutricionales severas**. Para ello las bacterias serán crecidas hasta saturación en medio rico LB, identificando luego aquellos genotipos mutantes cuyo comportamiento, por defectivo, se aleje del observado para la cepa salvaje. Este análisis será importante para conocer si efectivamente ocurren sesgos (negativos y eventualmente también positivos) en algunos mutantes durante la etapa de enriquecimiento en LB. De existir tales sesgos, deberán ser tenidos en cuenta al analizar los resultados de los ensayos de colonización de plantas. La idea será excluir de la lista de genes asociados a colonización de plantas a todos aquellos que correspondan a auxótrofos (punto B previo) o a mutantes con cambios pleiotrópicos que generan pérdida de aptitud (*fitness*) (punto C previo) para crecer en medio rico; **(D) genes asociados a la colonización de la rizósfera**; y **(E) genes asociados a la colonización endofítica de la raíz**. Dado que la colonización endofítica implica eventos tempranos vinculados al propio ingreso a la raíz, y luego funciones más tardías posiblemente diferentes, decidimos analizar la colonización endofítica a las 24 h y a los 20 días post inoculación (d.p.i.).

Para los estudios que aquí hemos listado usaremos la biblioteca transposicional que generamos y describí en el Capítulo III. En trabajos previos para estudiar la patogenicidad en plantas de *Pantoea*

stewartii con la técnica TnSeq, se utilizó con éxito una biblioteca de características similares a la que usaremos en este trabajo de tesis (Duong et al., 2018).

V.3.b Colección de las muestras provenientes de cultivos en medio mínimo (M9) y rico (LB) (búsqueda de grupos génicos B y C).

Para la realización de ambos cultivos, en medio mínimo M9 y en medio rico LB, se utilizó un inóculo bajo, del orden de 10^7 u.f.c./ml (DO= 0,01) de células lavadas proveniente de la biblioteca transposicional. La biblioteca usada como inóculo constituyó la **MUESTRA DE ENTRADA_Inóculo_E** (de *early*, en la Figura V.4). Los cultivos en M9 y en LB se dejaron crecer hasta saturación en las condiciones descritas en Materiales y Métodos, sección II.1.c., alcanzando cerca de 3×10^9 u.f.c./ml. Por tanto, debieron tener lugar cerca de 8 ciclos de división celular de modo de lavar auxótrofos en el medio mínimo M9, y también observar si en medio LB existen cambios de representación importantes por mutaciones que afecten en grado apreciable la aptitud (*fitness*) para crecer en medio rico. Alícuotas de los cultivos en M9 y en LB saturados fueron centrifugadas, y las células criopreservadas hasta su descongelamiento para preparar ADN genómico y procesarlo para el recuento de inserciones (digestión con *MmeI*, ligación de adaptador y qPCR según sección previa V.2). Para cada uno de los cultivos se procesaron 4 réplicas que constituyeron las **MUESTRAS DE SALIDA_MÍNIMO_M9** y **MUESTRAS DE SALIDA_INÓCULO_L** (de *late*).

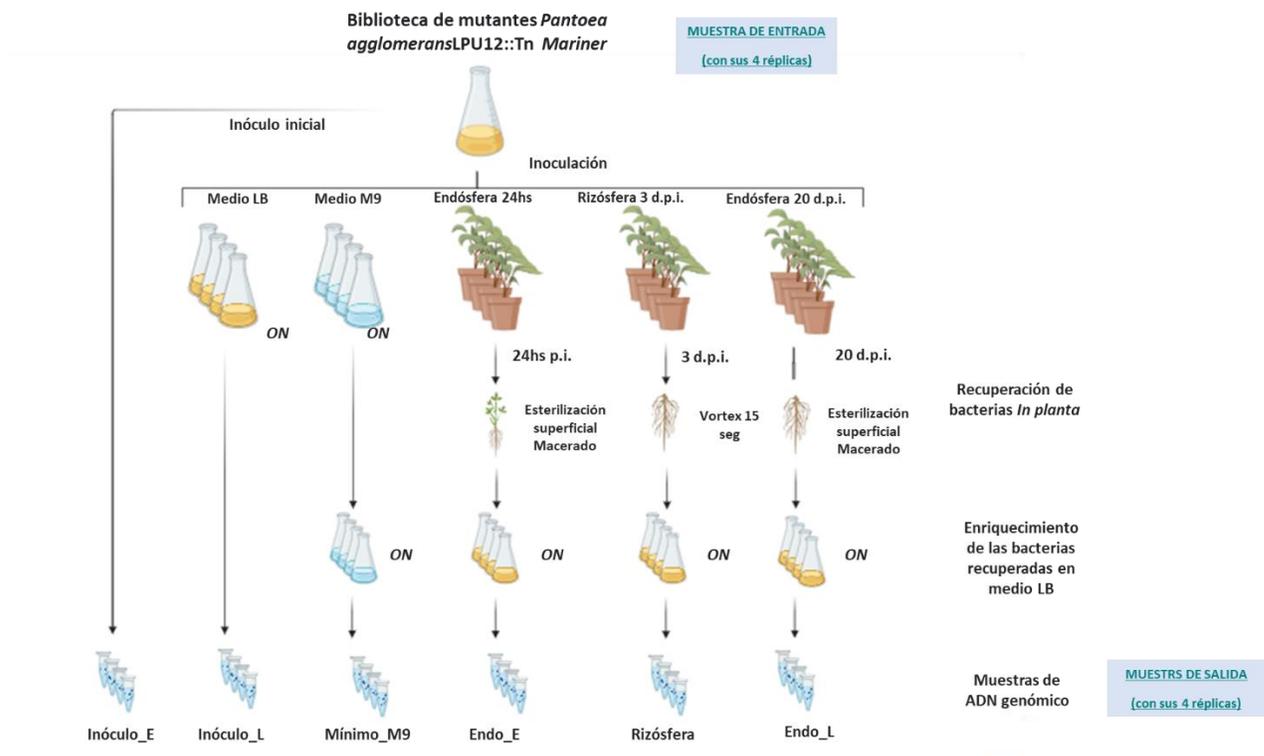


Figura V.4. Esquema de tratamientos del experimento TnSeq para la caracterización de la genética de la colonización rizosférica y endofítica de alfalfa por parte de la cepa *P. agglomerans* LPU12. La biblioteca transposicional de *P. agglomerans* LPU12 que describimos en el Capítulo IV fue utilizada como inóculo en los distintos tratamientos que se indican en la figura. Para los ensayos de colonización rizosférica y colonización endofítica se utilizaron inóculos cercanos a 10^5 y 10^7 u.f.c./ml, respectivamente (ver texto). Para cada una de las

condiciones ensayadas se realizaron 4 réplicas. Los ensayos de colonización, rizosférica o endofítica, incluyeron 15 plantas por réplica.

V.3.b.1 Colección de la muestra rizosférica (búsqueda del grupo génico D).

La inoculación con cerca de 10^5 u.f.c./ml de plantas en vermiculita y cosecha de bacterias rizosféricas por agitación 15 seg con vortex fue realizada como ya hemos descrito en el Capítulo III. Las bacterias así liberadas de la raíz, para ser amplificadas, constituyeron el inóculo de un cultivo en LB que fue crecido a saturación como se esquematiza en la Figura V.4. Como el resto de las muestras anteriores las células centrifugadas se criopreservaron hasta la extracción de ADN, procesamiento y qPCR para dar lugar a la **MUESTRA DE SALIDA_RIZÓSFERA**.

V.3.b.2 Colección de la muestra endofítica (búsqueda del grupo génico E). Procesamiento de las plantas para eliminar la población rizosférica y del rizoplano.

Para los experimentos de colonización endofítica utilizamos un inóculo cercano a 10^7 u.f.c./ml que se aplicó sobre plantas en macetas con vermiculita. Una vez inoculadas, las plantas fueron incubadas durante 24 h como he detallado en el Capítulo III para el análisis de la dinámica temporal de colonización de la endosfera. Con el diseño experimental que he utilizado, cabe recordar que las aguas de lavado siempre contenían algunas bacterias remanentes (probablemente de la rizósfera, o el rizoplano) aunque siempre en baja proporción (menos que 10%) respecto de la población endofita. Por tanto, dado que para poder obtener ADN de las bacterias endofitas recuperadas por macerado del tejido vegetal requerimos un paso de enriquecimiento por crecimiento en medio LB (ver Figura V.4), fue necesario implementar una condición de lavado que elimine toda bacteria viable exterior a la raíz que pueda ser arrastradas al cultivo en LB. Buscamos por tanto un protocolo que permita la esterilización superficial de la planta, de modo que durante el macerado sólo se recuperen bacterias endofitas. Para ello, ensayamos el uso del peróxido de hidrógeno (H_2O_2) como agente de esterilización superficial. Se evaluaron dos concentraciones del agente de esterilización (4 ó 10 volúmenes) por un período de exposición de 30 segundos. Como control se repitió el protocolo de lavado y extracción de bacterias endofitas usado en el Capítulo III sin esterilización superficial (Materiales y Métodos, sección II.4.c.2). En la Figura V.5 se presentan los resultados que muestran que la solución **25% (10 volúmenes)** H_2O_2 aplicada durante 30 seg fue suficiente para eliminar las bacterias presentes en el exterior de la planta sin afectar la población presente en la endosfera.

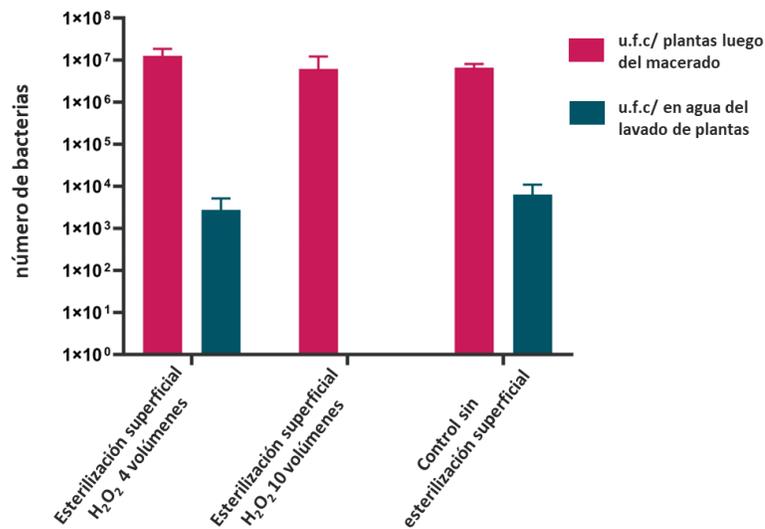


Figura V.5. Número de bacterias en las aguas de lavado y en el interior de las plantas luego del macerado. En el ensayo que se presenta en la figura, 15 plantas de alfalfa, en macetas con vermiculita, fueron inoculadas por inundación con 1×10^7 u.f.c./ml (Materiales y Métodos sección II.4.b.3). Para eliminar las bacterias de la superficie de las plantas se evaluó el uso durante 30 seg de dos concentraciones de peróxido de hidrógeno; 4 ó 10 volúmenes. Como control sin esterilización, en los lavados se utilizó sólo agua siguiendo el protocolo usado en el capítulo III, descrito en la sección II.4.c.2. Luego de los lavados indicados, 24hs postinoculación, las bacterias endófitas se recuperaron por maceración de los tejidos vegetales y contadas en medio LB con ácido nalidíxico. Para cada una de las condiciones se realizaron 4 réplicas. Las barras de la figura corresponden al error estándar.

Ajustada la esterilización externa de las raíces, para el experimento TnSeq estudiamos la colonización endofítica de plantas de alfalfa a las 24 h y a los 20 d.p.i. usando un inóculo de 10^7 u.f.c./ml (4 réplicas/condición, 15 plantas/réplica). Luego de la esterilización superficial con 25% H₂O₂ durante 30 seg, las plantas fueron maceradas y la población endofítica amplificada por crecimiento en medio LB como se hizo con la muestra rizosférica. Las células resultantes del cultivo se criopreservaron hasta su procesamiento, dando lugar a las **MUESTRAS DE SALIDA_ENDO_E Y ENDO_L**.

V.3.b.3 Reacciones de qPCR sobre las muestras de entrada y de salida. Mezcla para la secuenciación de alta capacidad.

Como se desprende de la Figura V.4, el diseño del experimento que realizamos consistió en 6 tratamientos diferentes con cuatro réplicas cada uno: **MUESTRAS DE SALIDA_INÓCULO_E**, **MUESTRAS DE SALIDA_INÓCULO_L**, **MUESTRA DE SALIDA_MÍNIMO_M9**, **MUESTRA DE SALIDA_RIZÓSFERA** y **MUESTRAS DE SALIDA_ENDO_E Y ENDO_L**. Debimos procesar por tanto 24 muestras que fueron digeridas con *MmeI*, ligadas al adaptador como ya indicamos, y amplificadas para obtener los productos de 65-66 pb que contienen los 16-17 pares de bases informativas que en el genoma se encontraban bordeando las inserciones del Tn *mariner*. Consistente con los análisis que realizamos en la sección V.2.b.2, para cada una de las qPCRs utilizamos como molde 3,125 ng proveniente de su correspondiente ligación, y los cebadores que detallamos en la Figura V.2 y en Materiales y Métodos sección II.3.e (Tabla II.5 y Tabla II.6) para etiquetar de manera diferencial cada tratamiento y replica (empleamos 24 etiquetas distintas asociadas a la colección de cebadores IR_Tag-n). En la qPCR las reacciones de amplificación tuvieron

valores de Ct comprendidos entre 16,78 y 20,48; ambos comparables al valor 16,58 que correspondió a la amplificación de la misma cantidad de molde en la Figura V.3 panel B. Con respecto a los blancos sin molde, comenzó a detectarse amplificación inespecífica muy tardíamente, con valores de Ct comprendidos entre 33 y 38. Si bien algunos de los blancos en el ciclo 38 mostraron una fluorescencia apreciable, lo regular es que cuando las reacciones de PCR específicas comienzan en ciclos muy tempranos respecto de lo observado para los blancos (como en nuestro caso), la amplificación inespecífica en los tubos donde hay molde debe ser (por competición) muy baja. Pudimos por tanto realizar las 24 reacciones de amplificación asociadas al experimento presentado en la Figura V.4. Los productos de amplificación alcanzados al final de la misma (ciclo 38) fueron mezclados, usando cantidades equivalentes normalizadas a partir de la fluorescencia relativa alcanzada en cada muestra. Para ello, se tomó como referencia el uso de 10 μL de la muestra de mayor fluorescencia, la que fue mezclada con volúmenes (siempre mayores) del resto de las muestras. El volumen final de la mezcla resultó en 346,2 μL , con una concentración de ADN doble hebra cercana a 4,82ng/ μL .

V.3.b.4 Clonado y secuenciación de amplicones para la verificación y confirmación de la estructura esperada [adaptador-genoma-Tn].

Realizada la mezcla de los productos de amplificación de las diferentes qPCR, antes de la secuenciación de la misma procedimos a analizar el perfil de tamaños de las especies moleculares que estaban presente. En la Figura V.6, paneles A y B, se muestran los resultados de una electroforesis de alta resolución en equipo Bioanalyzer (*High Sensitivity DNA Assay*, Agilent). Se observan 5 componentes principales cuyas características de tamaño y concentración se presentan en el panel C de la figura. Llamó la atención: **(a)** la aparente ausencia del producto esperado de 65-66 pb (ver Figura V.2.), y **(b)** la presencia en el pico 4 de una banda electroforética ancha, de movilidad correspondiente a un tamaño de 103 pb, y mayoritaria en la mezcla en proporción molar cercana a 55% (fila señalada en celeste, última columna de la tabla). Cuando la misma mezcla de los productos de la qPCR fue corrida en un gel convencional de agarosa 2,5 % (p/v) observamos nuevamente, en consonancia con el resultado del Bioanalyzer, un componente mayoritario con movilidad correspondiente a productos de cerca de 100 pb (Figura V.7, panel A, Muestra 2 calle 3). Con el propósito de indagar en las causas de las anomalías observadas, realizamos una reacción de PCR con los mismos cebadores Adapter-Top-PCR e IR_Tag-1 pero en la que, en lugar de emplear como molde el ADN de las ligaciones, utilizamos un oligonucleótido sintético de doble cadena con la secuencia que se muestra en la Figura V.2. El producto de amplificación obtenido mostró, ahora sí, la movilidad electroforética esperada correspondiente a un amplicón de 65-66 pb (Figura V.7, panel A, Muestra 1 calle 2).

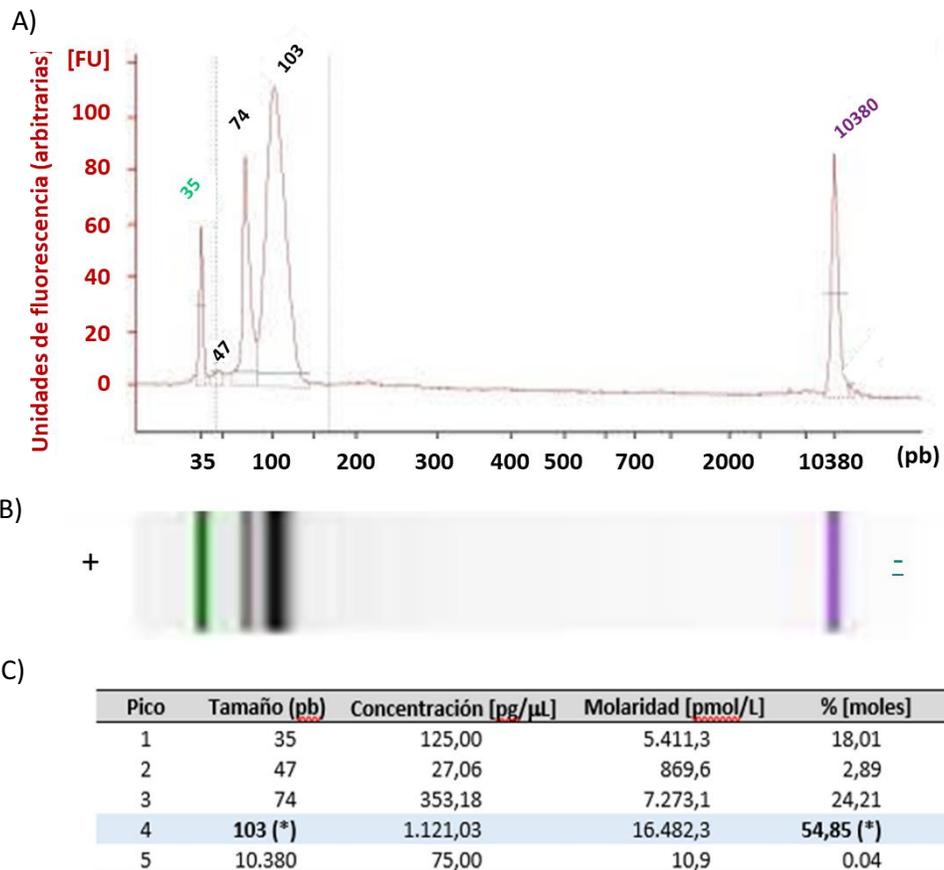


Figura V.6. Electroforesis de alta resolución de la mezcla de los productos de amplificación de las qPCR. Una alícuota de la mezcla de los productos de amplificación de la qPCR fue analizada por electroforesis de alta resolución en un Bioanalyzer (*High Sensitivity DNA Assay*, Agilent). **Paneles A y B:** electroferograma y el perfil electroforético. **Panel C:** Tabla donde se muestran las características de los 5 productos cuantitativamente más relevantes que se observan en la corrida. Los picos 2 y 5 corresponde a patrones de bajo y alto peso molecular agregado a la muestra. **Valores con asterisco (*):** Los valores para el pico 4 indicados con (*), esto es tamaño de 103 pb y abundancia porcentual en moles 54,85%; surgen de la electroforesis y están sobre y subestimados, respectivamente. El clonado y secuenciación de amplicones mostró que los productos más abundantes fueron de 66 pb, sugiriendo que los mismos deben corresponder al pico 4 y que por tanto su tamaño no es de 103 pb como se deduce de la electroforesis (ver texto para posibles razones para esta discrepancia). Considerando para los amplicones del pico 4 un tamaño de 66 pb la abundancia porcentual de los mismos en la muestra sembrada debe ser cercana a 65%, y cercana a 75% en la mezcla de qPCR sin los estándares que se agregaron para la corrida electroforética.

Por tanto, los productos con mayor movilidad, de tamaño aparente cercano a los 100 pb, aparecían en la qPCR donde los moldes eran de secuencia muy diversa en lugar de un único oligonucleótido de secuencia única. Es entonces posible que al acumularse productos de amplificación cuyo núcleo central de 16 pb es diferente (porque los adaptadores se ligaron a secuencias genómicas distintas que bordean el Tn), existan hibridaciones cruzadas generando estructuras de doble hebra con una región central de hebra simple donde no hay complementariedad o donde existe complementariedad parcial (burbuja interna) (Figura V.7 Panel c). La formación de estructuras híbridas de ese tipo se verá favorecida en los últimos ciclos de la PCR cuando la concentración de cebadores es baja o se ha agotado, y no existen por tanto reacciones de extensión que generen estructuras de doble cadena de complementariedad perfecta.

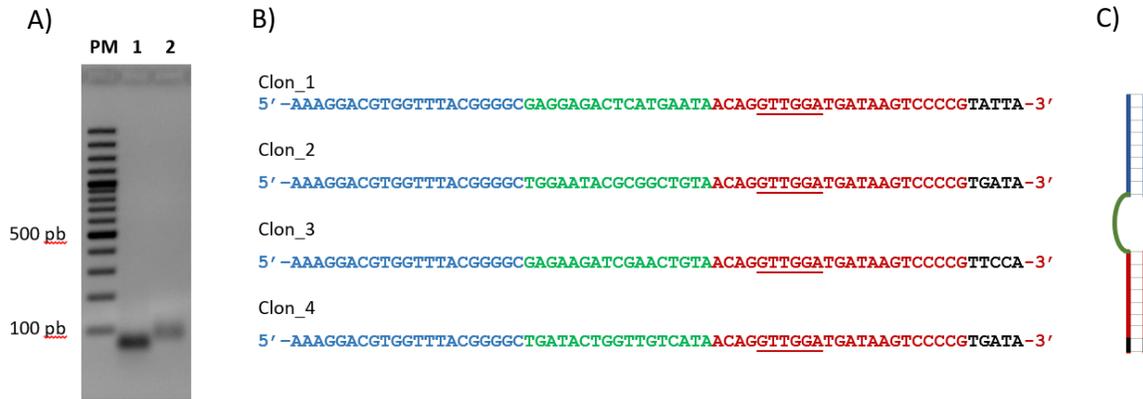


Figura V.7. Electroforesis en gel de agarosa de la mezcla de productos de amplificación de las qPCR. Secuencias de amplicones clonados. Panel A. Gel de agarosa 2,5 % (p/v). Calle 1: patrón de peso molecular (PM). Calle 2: producto de PCR resultante de la amplificación de un oligonucleótido de doble cadena ensamblado por hibridación de los oligonucleótidos complementarios TnSeq-reference-1 y TnSeq-reference-2 (Materiales y Métodos, Tabla II.5) con los cebadores Adapter-Top-PCR e IR_Tag-1 dando lugar a un producto de amplificación de 65 pb (1). Calle 3: alícuota de la mezcla de los productos de amplificación de las diferentes qPCR (2). **Panel B.** Secuencias de algunos de los productos de amplificación de la mezcla de productos de qPCR que fueron clonados en el pGEM-T (ver texto). En **rojo** se muestra la secuencia terminal del Tn *mariner* (sitio de reconocimiento *MmeI* subrayado), en **verde** se muestra la secuencia genómica adyacente al sitio de inserción del transposón y en **azul** la secuencia del adaptador ligado al ADN genómico digerido con la enzima *MmeI*. En **negro** se muestran las secuencias de las etiquetas que corresponden a cada uno de los productos amplificados (tratamiento/réplica) (ver leyenda a la Figura V.2). **Panel C.** Representación esquemática de los productos de amplificación de la qPCR en la que hibridan cadenas con secuencias “nucleo” genómicas que provienen de inserciones del Tn *mariner* diferentes (ver texto). El código de colores es el mismo que en Panel B. Verde intenso y verde claro en la “burbuja” refieren a secuencias nucleotídicas no idénticas.

La presencia de amplicones de doble hebra con una burbuja central es razonable que cambie el volumen molecular generando conformaciones más variadas (banda más ancha en el Bioanalyzer) y posiblemente más abiertas retardando por tanto la movilidad electroforética. Una situación comparable habíamos ya observado previamente en amplicones similares generados en experimentos de STM (*Signature Tagged Mutagenesis*) (Salas et al., 2017; Salas Maria Eugenia, 2015). Con el propósito de verificar si la hipótesis era correcta, procedimos al clonado (aleatorio) y secuenciación de algunos amplicones de la qPCR. En la Figura V.7, panel B se muestran las secuencias de 4 de esos amplicones. Los resultados muestran que:

- Los amplicones incluyen 17 pb centrales de secuencia genómica (recuérdese que el corte de la enzima *MmeI* puede ocurrir a 20 o a 21 pb del sitio de reconocimiento, que en el caso de las inserciones del Tn *mariner* darán lugar a 16 o 17 pb de secuencia genómica y a productos de 65 o 66 pb, respectivamente). En nuestro caso los 4 amplicones secuenciados resultaron tener un lago de 66 pb (sin contabilizar el agregado de A en los extremos 3' propio de la actividad transferasa terminal de la Taq polimerasa). Los amplicones de este largo parecieron ser los mayoritarios, y por tanto corresponder al pico 4. Atendiendo entonces a este tamaño, en la mezcla de la qPCR la proporción molar de los amplicones de 65-66 pb no será 55% como se indica en la tabla de la Figura V.6 panel C (línea celeste) sino cercana a 65% en la muestra sembrada, y cercana a 75 % en la qPCR sin los estándares agregados de 35 pb y 10.380 pb.
- En los amplicones secuenciados se pudo verificar en uno de sus extremos la presencia de diferentes “firmas” (3 firmas diferentes en 4 amplicones: TATT, TGAT y TTCC).

- Como era esperable para el Tn *mariner*, en las 4 inserciones se observa una secuencia TA (en verde) inmediatamente antes del comienzo de la secuencia IS (en rojo).

Habiendo verificado que la mezcla de la qPCR contenía los productos de amplificación esperados, enviamos la misma al servicio de secuenciación del CeBiTec en la Universidad de Bielefeld. En el Capítulo siguiente presentamos los resultados de la secuenciación y el análisis de datos del experimento TnSeq para cada uno de los tratamientos que presentamos en la Figura V.1.

V.4. Consideraciones finales del capítulo

En este capítulo presenté el diseño general y ejecución de un experimento TnSeq para abordar en *P. agglomerans* la búsqueda de determinantes genéticos asociados a su crecimiento en distintas condiciones y a su capacidad de colonizar raíces. El ensayo incluyó 6 tratamientos diferentes con cuatro réplicas cada uno de ellos. En el diseño experimental hemos tenido especialmente en cuenta que los inóculos utilizados representen de manera adecuada a la diversidad de mutantes presentes en la biblioteca transposicional que empleamos en el ensayo. Por otra parte, a la hora de amplificar las inserciones por qPCR para su secuenciación posterior, verificamos nuevamente la correcta representación de los mutantes de la biblioteca (la cuantificación de inserciones mostró que la etapa de digestión del ADN genómico junto a la de ligación de adaptador fueron ambas muy eficientes, con un número de inserciones amplificables cercano al esperado por cálculo). Finalmente, en los productos amplificados pudimos verificar, por clonación y secuenciación de Sanger, que los mismos tenían la constitución esperada [adaptador-secuencia genómica-Tn] y el largo predicho (en nuestros ejemplos 66 pb), además de representar en fracción molar el componente mayoritario de las reacciones de qPCR (cercano a 75%). En el capítulo siguiente presentaré los resultados de la secuenciación y el análisis de datos que nos permitió reconocer genes esenciales, de auxotrofia, y genes relevantes para hacer la primera aproximación a la raíz y para establecerse luego en el interior de la misma.



CAPÍTULO VI

IDENTIFICACIÓN DE MARCADORES GENÉTICOS DE *P. AGGLOMERANS*
LPU12 INVOLUCRADOS EN LA COLONIZACIÓN RIZOSFÉRICA Y
ENDOFÍTICA DE PLANTAS DE ALFALFA



VI.1. Introducción

En los capítulos precedentes de esta tesis presenté el sistema modelo *P. agglomerans* LPU12 – alfalfa (Capítulo III) con el que implementamos un ensayo TnSeq para estudiar la colonización rizosférica y el abordaje endofítico de la raíz (Capítulo V). Dicho ensayo TnSeq fue realizado empleando la colección de mutantes de una biblioteca transposicional que incluyó decenas de miles de mutantes generados por inserción de un Tn *mariner* (Capítulo IV).

Como hemos detallado en el capítulo previo, el producto emergente del experimento TnSeq es una mezcla de amplicones de PCR que contienen, para cada una de las muestras procesadas, los distintos sitios de inserción del Tn y cada uno de ellos en su representación numérica. Nuestro estudio incluyó ensayos en los que se analizó cómo la representación de cada mutante se ve afectada por el crecimiento en medio mínimo, en medio rico, o por la colonización de la planta (los tratamientos pueden verse esquematizados en la Figura V.3 del capítulo previo). Frente a cualquiera de los desafíos anteriores, la performance de mutantes en un gen particular de la biblioteca respecto de la bacteria salvaje, se expresó a través de su parámetro M^{TnSeq} (ver definición en Capítulo V, sección V.2.1); esto es el logaritmo en base 2 de la razón entre la representación de los mutantes a la salida del ensayo, respecto de la representación de los mismos en el inóculo. El conteo, mapeo de cada una de las inserciones y el cálculo de los valores M^{TnSeq} para los distintos genes, en los diferentes ensayos, se realizó con la herramienta *Transit* de acceso libre (Materiales y Métodos, Capítulo II, sección II.5.e). Para la identificación de genes esenciales, cuyos mutantes estuvieron ausentes, o en bajísima representación, dentro de la biblioteca; se emplearon tres algoritmos diferentes dedicados a ello también del paquete *Transit* (ver sección siguiente VI.2.a.). Presentaré a continuación los resultados obtenidos en el experimento TnSeq.

VI.2. Búsqueda de genes esenciales, genes de auxotrofia y genes de pérdida de parcial de aptitud (*fitness*) para crecer en medio rico.

VI.2.a Genes esenciales del genoma de *P. agglomerans* LPU12.

Conocer los genes esenciales de una especie bacteriana es una instancia importante en la caracterización genética de la misma, dado que pone de manifiesto cuál es (para ese taxón) la información no-prescindible, en un contexto genómico dado, para sostener la viabilidad en cualquier medio. Tal información permite identificar restricciones a la manipulación con fines biotecnológicos y ayuda a definir blancos de drogas que interfieran (en patógenos) o refuercen (en bacterias benéficas) procesos específicos. Desde un punto de vista básico, la identificación de la información esencial en distintos organismos contribuye a los proyectos de construcción de vida sintética (A. Ghomi et al., 2024).

En nuestro estudio, la identificación del conjunto de genes esenciales de *P. agglomerans* LPU12 resultó imprescindible para, ante la ausencia de ciertos mutantes en las muestras rizosféricas o endofíticas, diferenciar los efectos de las mutaciones letales de aquellas que generan deficiencias específicas para colonizar el ambiente planta. Para la identificación de genes posiblemente esenciales, utilizamos la información de la secuenciación de la **MUESTRA DE ENTRADA_Inóculo_E** (de *early*, en la Figura V.4, capítulo anterior). Los datos de mapeo y conteo de inserciones en esa muestra fueron analizados con las siguientes funciones que posee el programa *Transit* (DeJesus et al., 2015) para la búsqueda de genes esenciales: **(a)** Búsqueda “**Gumbel**”: A partir de las inserciones en sitios TA (propias del Tn *mariner*) se calcula la probabilidad de secuencias largas sin inserciones; **(b)** Búsqueda “**HMM (Hidden Markov Model)**”: En este caso se usan modelos de Markov ocultos para evaluar las variaciones en las frecuencias de

inserciones a lo largo del genoma. Este método incorpora diferencias locales en los conteos de lecturas y se usa para identificar genes esenciales mediante el análisis de "gaps" (secuencias largas sin inserciones); y (c) "TTN (***Transposon TnSeq Network***)": Esta aproximación hace uso de modelos de redes de transición para evaluar la distribución y frecuencia de inserciones de transposones en el genoma bacteriano. Las redes de transición ayudan a identificar patrones y relaciones entre diferentes regiones del genoma, lo que permite identificar a los posibles genes esenciales.

La búsqueda de candidatos que en nuestra biblioteca transposicional dieran resultados positivos con los 3 métodos anteriores arrojó una lista de 461 genes, mientras la exigencia de positividad para sólo 2 de los métodos resultó en una lista de 513 genes, en su mayoría (97%) de localización cromosomal como era esperable. La proporción de genes esenciales que hemos observado en el cromosoma de *P. agglomerans* LPU12 representa cerca de 13,28 % del total de genes anotados para ese replicón (3.863 genes), un valor consistente con el valor 12,87 % reportado previamente para los genes esenciales de localización cromosomal en la especie cercana *P. stewartii* (Duong et al., 2018). El número absoluto de genes esenciales, además, resultó consistente con los valores reportados en otras especies bacterianas (Luo et al., 2021). Más allá de esta consideración general, recientemente se ha publicado un estudio en enterobacterias donde se muestra que la esencialidad no es una característica inmutable, sino que depende crucialmente del contexto genómico (A. Ghomi et al., 2024). Así, en enterobacterias se ha observado que aproximadamente un tercio de los genes esenciales en cualquier cepa particular no son esenciales en otra cepa relacionada. En ese mismo trabajo se hace referencia a los efectos que pueden tener los sesgos en la inserción de los transposones en la identificación de genes esenciales, circunstancia que destaca la necesidad de analizar cuidadosamente los resultados de experimentos como los que aquí hemos presentado.

En relación a nuestro ensayo, como era esperable, de los 513 genes posiblemente esenciales 380 no tuvieron inserciones, y los 133 restantes tuvieron inserciones con un número de lecturas que se encontró dentro del 2,5% de los mutantes con lecturas menos abundantes dentro de la biblioteca (algo razonable si se tiene en cuenta que mutaciones en genes esenciales truncan la capacidad de crecer, y por tanto disminuyen su representación con los ciclos de cultivo). En la Tabla VI.S1 suplementaria se muestran los genes identificados como potencialmente esenciales con sus correspondientes códigos COG (*Clusters of Orthologous Genes*) y KO (*KEGG Orthology*). Las bases de datos COG y KEGG organizan genes y proteínas en grupos de ortología, a partir de los cuales pueden realizarse inferencias funcionales. En la Figura VI.1 se muestra de manera sintética cómo es la distribución funcional asociada a estos genes.

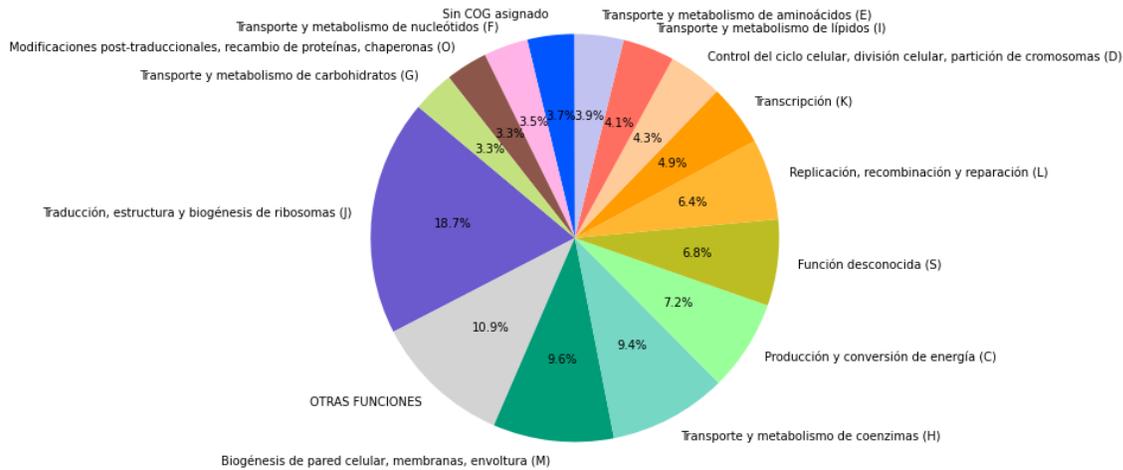


Figura VI.1. Distribución de funciones en el grupo de genes identificados como posiblemente esenciales en *P. agglomerans* LPU12. La asignación funcional se realizó a partir del análisis de los códigos COG de cada uno de los genes (ver Tabla VI.S1). Los diferentes grupos COG (A – Z, leyendas y letras en el gráfico) se representan en distinto color con su abundancia porcentual correspondiente. Todos aquellos grupos cuya representación fue menor que 3% se agruparon bajo la denominación OTRAS FUNCIONES.

Entre los genes esenciales destacó la fuerte predominancia de genes asociados a la maquinaria de traducción, y como era esperable por genes asociados a actividades muy diversas como se aprecia en la variedad de funciones que componen el gráfico (incluida la zona gris que incluye todos aquellos COGs cuya representación es menor que 3%, y que por simplicidad se incluyeron en esa categoría). Cabe también mencionar que entre los mutantes en genes esenciales emergieron inserciones en 3 ncRNAs (RNAs pequeños no codificantes).

VI.2.b Genes de auxotrofia y genes cuya mutación resulta en efectos menos severos (pero aún detectables) para crecer en medio mínimo M9 respecto de medio LB.

Genes cuya mutación resulta en un mejor crecimiento en medio M9 respecto de medio LB.

La auxotrofia refiere a la incapacidad de un organismo para sintetizar un compuesto esencial para su crecimiento. En la biblioteca de mutantes Tn mariner de *P. agglomerans* LPU12 generada para este trabajo es de esperar que muchas inserciones del transposón ocurran en genes de auxotrofia, desarrollando un requerimiento nutricional específico como resultado de la mutación. Dado que no conocemos con precisión la identidad de los nutrientes que están disponibles en la rizósfera y la endosfera, resultó importante conocer los mutantes auxótrofos dado que podría ocurrir que su crecimiento esté afectado en el ambiente planta. El reconocimiento de estos mutantes es importante para poder identificar qué deficiencias en el abordaje de la planta derivan de incapacidades biosintéticas de las bacterias (como en los auxótrofos), y cuáles de deficiencias de las bacterias propias del abordaje de la raíz (por defectos estructurales, en enzimas particulares, en la movilidad, en señalización, otros) o para usar los recursos allí presentes (defectos en genes catabólicos).

Con el propósito de identificar genes de *P. agglomerans* LPU12 cuya mutación resulte en fenotipos auxótrofos, analizamos comparativamente los mapeos y recuento de inserciones de la **MUESTRA DE SALIDA_MÍNIMO_M9** respecto de la **MUESTRA DE ENTRADA_Inóculo_E** (crecida en LB). Los genes de auxotrofia deberán tener recuento extremadamente bajo en medio M9 respecto de su recuento en medio LB. Con ese criterio de selección podrán identificarse también mutantes no-auxótrofos, pero cuya mutación de lugar a fenotipos que resulten más alterados en su crecimiento en M9 que en LB. Con un análisis como el precedente hemos detectado 162 genes no esenciales con valores de $M \leq -0.6$ (valor- $p \leq 0.01$), llegando en algunos casos a valores de M cercanos a $-4,8$. Si bien es esperable que los mutantes auxótrofos correspondan a aquellos que mostraron los valores de M más negativos, resulta *a priori* difícil establecer una línea de corte que separe esos mutantes de aquellos que, sin ser auxótrofos (porque aun crecen en M9), crecen de manera más defectiva en M9 que en LB. En la Tabla VI.S2 suplementaria se muestran los genes identificados sea como auxótrofos o con crecimiento más defectivo en M9 que en LB, todos con sus correspondientes códigos COG y KO. En la Figura VI.2 se muestra de manera sintética cómo es la distribución funcional asociada a esos genes.

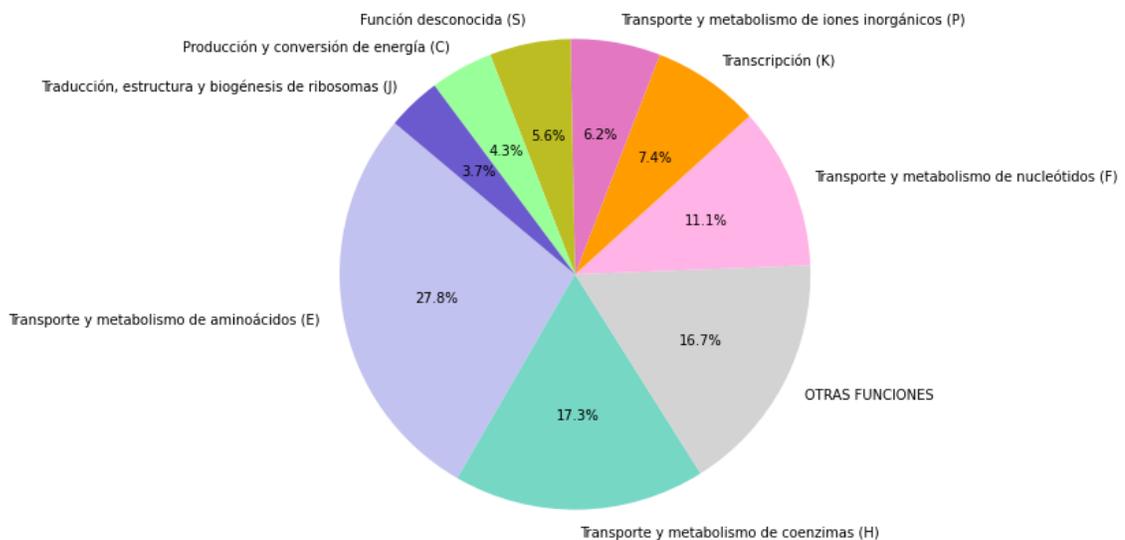


Figura VI.2. Distribución de funciones en el grupo de genes de *P. agglomerans* LPU12 asociados a fenotipos de auxotrofia o de deficiencia de crecimiento en medio mínimo M9 respecto del crecimiento en medio rico LB. La asignación funcional se realizó a partir del análisis de los códigos COG de cada uno de los genes (ver Tabla VI.S2). Los diferentes grupos COG (A – Z, leyendas y letras en el gráfico) se representan en distinto color con su abundancia porcentual correspondiente. Todos aquellos grupos cuya representación fue menor que 3% se agruparon bajo la denominación OTRAS FUNCIONES.

Entre estos genes se observa predominancia de genes metabólicos; asociados principalmente al transporte y metabolismo de aminoácidos (27%), seguido por genes del metabolismo de coenzimas (17%), metabolismo de nucleótidos (11%), seguidos por otras funciones celulares con representación menor que 10%.

Resultó especialmente llamativo que también detectamos 158 genes no esenciales asociados a mutaciones que resultan en un mejor crecimiento en M9 que en LB, con valores positivos de M comprendidos entre 0,6 y 2,53. El número de estos genes fue comparable al de los genes de auxotrofia. En la Tabla VI.S3 se presenta la lista de estos genes con sus respectivos códigos COG y KO. Entre estos genes resultó muy abundante la presencia de productos de función aún desconocida (15%). No debe descartarse que la vida en ambientes oligotróficos con limitaciones de nutrientes como el suelo, pueda

haber favorecido la selección de sistemas eficientes para captar y utilizar nutrientes esenciales, los que pueden desregularse (y operar deficientemente) en condiciones de abundancia de nutrientes, como en LB, afectando el crecimiento. Mutaciones que afecten el uso de recursos presentes en el LB pueden habilitar el uso de rutas más eficientes en condiciones naturales donde los recursos son escasos. De ser así es razonable, como se deduce de la Tabla VI.S3, que entre las categorías con funciones identificadas destaquen, además del metabolismo de pared y membrana (11%), las de metabolismo de carbohidratos (10%) y producción y conversión de energía (7%). Esta posibilidad debe ser analizada teniendo especialmente en cuenta que *P. agglomerans*, además de colonizar plantas, es un habitante regular de suelo y agua que son frecuentemente ambientes oligotróficos. Será interesante investigar si este tipo de genes cuya mutación favorece el crecimiento en medio mínimo respecto de rico se observa en otras bacterias que habiten nichos con mayor disponibilidad de nutrientes. Existe también la posibilidad adicional que metabolitos generados en medio rico generen algún tipo de interferencia que no se hace presente cuando las bacterias crecen en medio mínimo. Cabe finalmente mencionar que entre los genes 158 genes anteriores se incluyó 1 mutante en un tRNA y 2 mutantes en ncRNAs.

La lista de genes presentados en las Tablas VI.S2 y VI.S3 suplementarias fueron tenidas en cuenta en la sección VI.3 que sigue a la hora de catalogar a los mutantes afectados en la colonización radicular, sea como mutantes afectados de manera pleiotrópica, incluyendo la auxotrofia; o como mutantes afectados de manera específica en su fenotipo de abordaje a la planta.

VI.2.c Genes asociados a la pérdida de aptitud (*fitness*) de *P. agglomerans* LPU12 para crecer en medio rico LB

Según indicamos en el Capítulo V precedente, luego de la colección de bacterias de raíces (tanto rizosféricas como endofíticas) fue necesario amplificar su número a través de un ciclo de crecimiento en medio líquido LB previo al procesamiento de su ADN y posterior secuenciación. Si dicho crecimiento genera cambios significativos en la abundancia relativa de mutantes, tales cambios podrían ser erróneamente atribuidos al comportamiento de los mutantes durante la colonización de la planta; cuando en realidad esos cambios correspondieron a (o estuvieron afectados por) sesgos generados durante la etapa de enriquecimiento.

Para conocer entonces si existen cambios de abundancia de mutantes durante la etapa de crecimiento en LB, comparamos la composición de mutantes en el inóculo inicial que se usó en las plantas (**MUESTRA DE ENTRADA_Inóculo_E**) con la composición de mutantes luego del crecimiento en LB durante aproximadamente 9 generaciones (**MUESTRAS DE SALIDA_INÓCULO_L**) (ver control de crecimiento en LB en la Figura V.4 del capítulo precedente). Los resultados mostraron que si se excluyen del análisis los auxótrofos y mutantes relacionados descriptos en la sección anterior, sólo mutaciones en 16 genes mostraron cambios de abundancia significativos con valores $M \leq -0,6$ y $p \leq 0.01$ (Tabla Suplementaria VI.S4). No observamos genes con valores $M \geq 0,6$ y $p \leq 0.01$. Como en el caso de los mutantes de la sección precedente, los identificados en esta sección fueron tenidos en cuenta durante el análisis de resultados de los experimentos de colonización de raíces.

VI.4. Búsqueda a escala genómica de marcadores genéticos asociados a la colonización de raíces de alfalfa por *P. agglomerans* LPU12

En las secciones que siguen mostraré los resultados que surgen de los experimentos de colonización rizosférica y endofítica.

VI.4.a En busca de los genes vinculados a la colonización de la rizósfera de alfalfa por *P. agglomerans* LPU12.

La colonización rizosférica representa la instancia inicial de interacción entre las bacterias del suelo y las raíces de las plantas. Las bacterias perciben dicho nicho, se acercan al mismo, y se multiplican haciendo uso de los recursos disponibles provenientes de la planta, fuentes de N y especialmente de C proveniente de la fotosíntesis. Dado que la rizósfera es un nicho con espacio y recursos limitados, las bacterias del suelo compiten por el acceso a la misma. En nuestro laboratorio hemos demostrado que la eficiencia para acceder a la rizósfera es en rizobios un factor importante durante la selección de los genotipos que más tardíamente ocuparán el interior de los nódulos radiculares (Salas et al., 2017). Sin embargo, se desconoce cómo es el impacto del acceso a la rizósfera sobre la competencia por el espacio planta en otras formas de asociación menos específicas, donde el acceso a la raíz (espacios intercelulares, tejidos de conducción) está habilitado a una gran diversidad de géneros y especies. Por la relevancia tanto de las plantas como de las bacterias asociadas a ellas, resultó para nosotros de especial interés conocer cómo es la dinámica (Capítulo III) y la genética (este capítulo) del acceso rizosférico en una bacteria endofita ubicua y con marcada capacidad de colonizar los tejidos internos de la raíz como *P. agglomerans*. Conocidos los genes importantes para la colonización de la rizósfera, el uso de mutantes en esos genes habilitará una herramienta potente y versátil con la que podrá investigarse cómo las interacciones más tempranas impactan en la competencia de las bacterias por la ocupación más tardía del nicho planta. Podrá además realizarse una búsqueda de ortólogos, hacia el reconocimiento de eventuales estrategias comunes de colonización empleadas por distintos géneros asociados a plantas.

Para la identificación de genes asociados a la colonización de la rizósfera analizamos de manera comparativa los cambios en la proporción de mutantes entre la **MUESTRA DE ENTRADA_Inóculo_E** y la **MUESTRA DE SALIDA_RIZÓSFERA**. El primer análisis de los datos mostró que la densidad de sitios TA mutados respecto de la totalidad de sitios TA del genoma en la muestra rizosférica de salida fue $D^{\text{rizósfera}} = 0,299 \pm 0,013$ (cerca de un tercio del total pudieron ser mutados); un valor coincidente con el de la densidad de mutaciones observada en el inóculo ($D^{\text{inóculo}} = 0,299 \pm 0,008$). Tal resultado indica que los mutantes que no acceden a la rizósfera no representan una proporción importante de todos los mutantes inoculados. En la Figura VI.3, panel A, se presenta una gráfica de tipo "volcán" en la que representamos los valores de M^{TnSeq} (de ahora en más M, por simplicidad) (eje y) en función del valor- p (eje x) para cada uno de los genes mutados que están presente en la biblioteca. La distribución muestra que la gran mayoría de los genes analizados tienen valores de M cercanos a 0 ($-0,6 < M < 0,6$) y en su mayoría sin diferencias significativas ($p > 0,01$) (puntos grises). Tal distribución resultó comparable a la que hemos observado en nuestro estudio previo de la colonización rizosférica de alfalfa por *S. meliloti* empleando la técnica STM (Salas et al., 2017). Nuevamente, un resultado como el anterior es esperable si los genes asociados a la colonización rizosférica son pocos en relación al total. En la Figura VI.3, panel B (derecho) se muestra ampliada (*zoom*) sólo la región que incluye a los mutantes que poseen valores de $M \leq -0,6$ (en **rojo**, afectados negativamente en la colonización) ó $M \geq 0,6$ (en **verde**, afectados positivamente en la colonización) en ambos casos con $p \leq 0,01$.

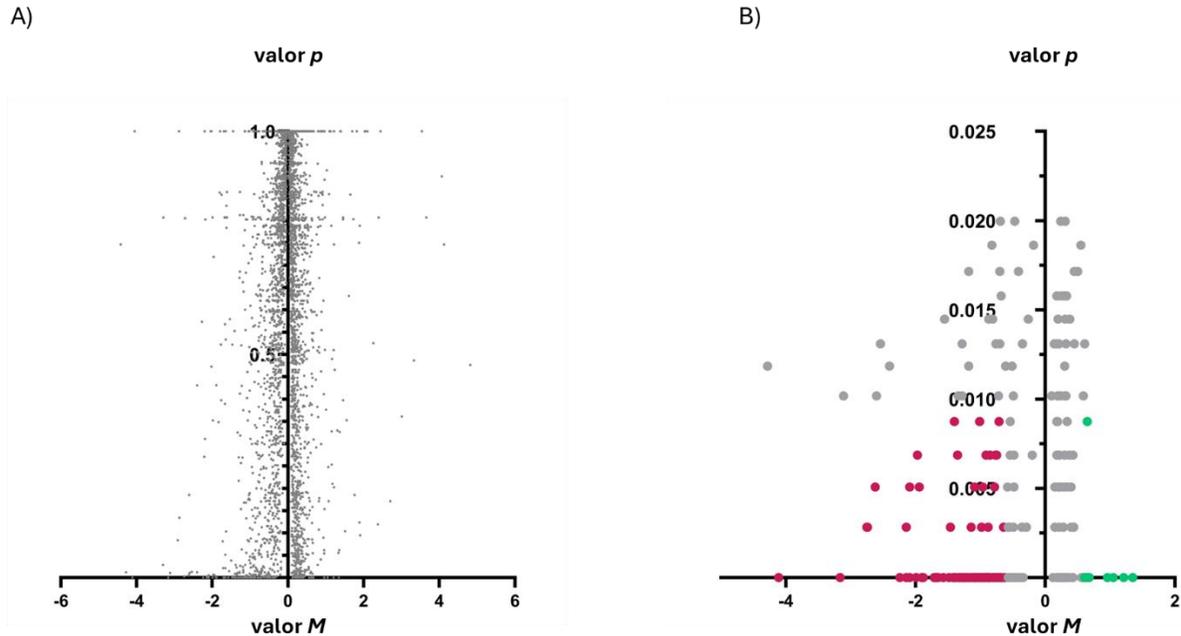


Figura VI.3. Gráfico tipo volcán en el que se representan los valores M y sus correspondientes valores p para los distintos genes mutados de la biblioteca al comparar el inóculo (**MUESTRA DE ENTRADA_Inóculo_E**) y la muestra de salida de la rizósfera a 3 d.p.i. (**MUESTRA DE SALIDA_RIZÓSFERA**). **Panel A:** gráfico incluyendo todos los genes mutados de la biblioteca. **Panel B:** gráfico en el que sólo se muestran los genes mutados con valores $p \leq 0,02$. Código de colores: **rojo** para mutantes con $M \leq -0,6$ y **verde** para mutantes con $M \geq 0,6$; en ambos casos con $p \leq 0,01$. No se colorearon en rojo los mutantes en genes esenciales, mutantes auxótrofos o relacionados, o aquellos cuya alteración en colonizar la rizósfera pudo haber sido debida a la etapa de enriquecimiento. Todo el resto de los mutantes en color gris.

En total se identificaron 108 genes que afectaron la colonización rizosférica (2,3% de los 4.605 marcos de lectura identificados en el genoma). Sin embargo, 62 de esos genes correspondieron a mutantes con efectos de auxotrofia (sección previa VI.2.b) o efectos de pérdida de aptitud (*fitness*) en su crecimiento en medio mínimo (sección previa VI.2.c). Los 46 genes restantes (1% del genoma) resultan en aquellos que afectaron específicamente la colonización rizosférica. La mayoría de este último grupo de genes (38) resultaron en efectos negativos sobre la colonización respecto de la cepa salvaje ($M \leq -0,6$); y sólo mutantes en los 8 genes restantes identificados manifestando mejores fenotipos de colonización ($M \geq 0,6$) (Figura VI.2, panel A).

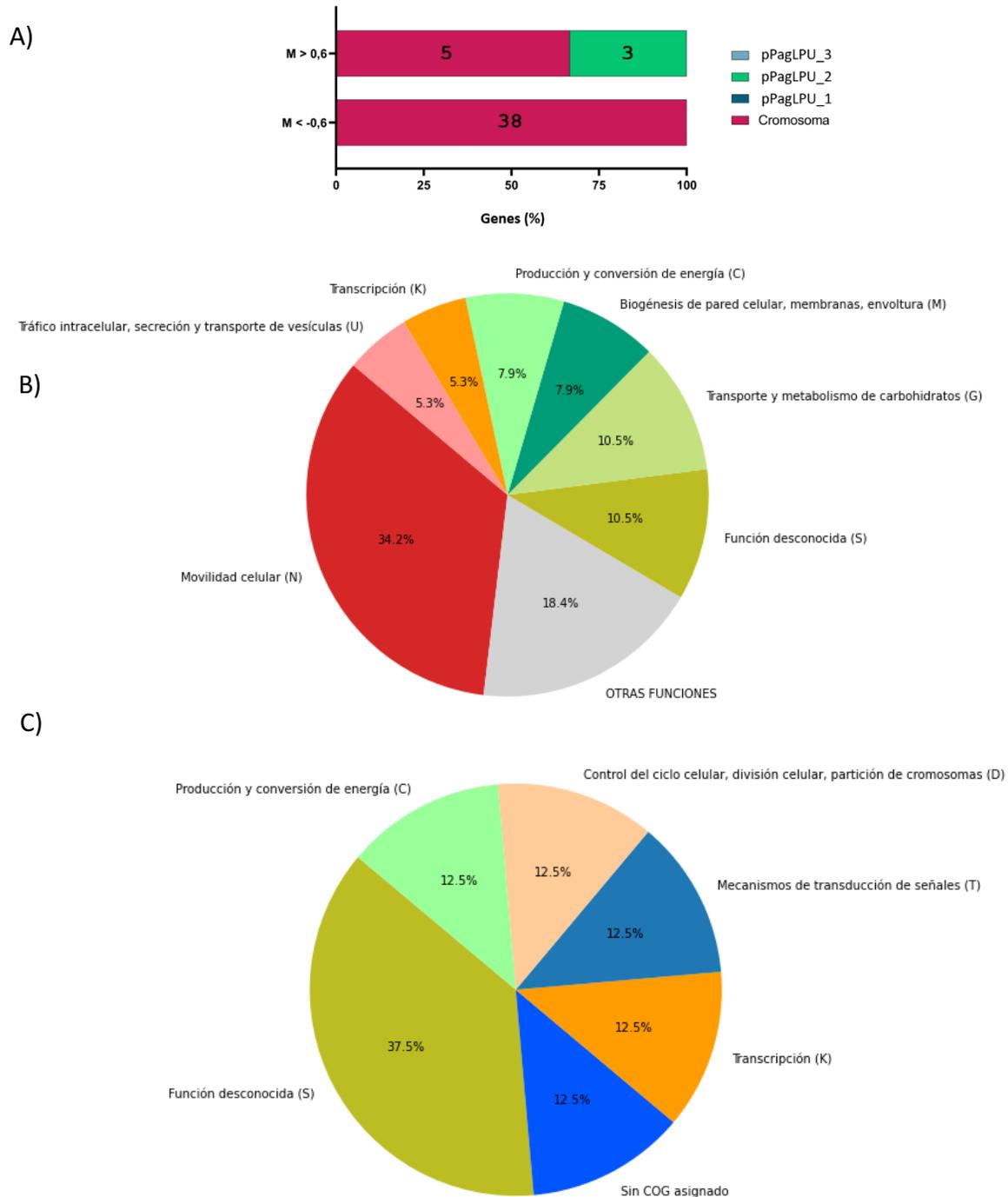


Figura VI.4. Localización genómica y distribución de funciones en el grupo de genes de *P. agglomerans* LPU12 que modifican la capacidad de colonizar la rizósfera de alfalfa. Panel A. Distribución de mutaciones según el replicón en el que se encuentran (cromosoma o plásmido) y según su valor M. **Panel B.** Asignación funcional realizada a partir del análisis de los códigos COG de cada uno de los genes vinculados a la colonización de la rizósfera cuyos mutantes resultan en valores de $M \leq -0,6$ (ver Tabla VI.S5) y valor- $p \leq 0,01$. **Panel C.** Asignación funcional realizada a partir del análisis de los códigos COG de cada uno de los genes relacionados a la colonización de la rizósfera cuyos mutantes resultan en valores de $M \geq 0,6$ y valor- $p \leq 0,01$. Tanto para el Panel B como C, los diferentes grupos COG (A – Z, leyendas y letras en el gráfico) se representan en distinto color con su abundancia porcentual correspondiente.

Todos aquellos grupos cuya representación fue menor que 3% se agruparon bajo la denominación OTRAS FUNCIONES.

En la Tabla VI.S5 suplementaria se muestra la lista de los 46 mutantes anteriores con sus correspondientes códigos COG, KO y valores M. En la Figura VI.4, paneles B y C, se muestra de manera sintética cómo es la distribución funcional asociada a los genes de cada grupo de mutantes.

En los resultados destacó que la mayoría de los mutantes con cambios en su capacidad de colonizar la rizósfera tuvieron en su mayoría valores de M negativos (38 vs. 8 mutantes, 80%) (Figura VI.4, panel A). Esta observación sugiere que a lo largo de la coevolución de *P. agglomerans* con raíces es posible que haya existido una selección purificadora en la que —siempre que fue posible— se ha ido perdiendo aquella información que perturbaba la vida asociativa temprana. De los 8 genes cuyos mutantes resultaron en mejores colonizadores, es interesante que 3 de ellos corresponden a mutantes en genes del plásmido pPaggLPU12_2, no-cromosomales y por tanto de arriba relativamente reciente al genoma de *P. agglomerans*. La conservación de dicho replicón plasmídico dependerá seguramente de cuán relevante sea la presencia de pPaggLPU12_2 para la persistencia de *P. agglomerans* en ambientes del suelo diferentes al de la vida rizosférica. Es asimismo posible que los 5 genes restantes de localización cromosomal, $M \geq 0,6$ y $p \leq 0,01$, correspondan a funciones celulares posiblemente importantes para hacer frente a circunstancias ambientales frecuentes, y que —a pesar de ser deletéreos para la colonización de la rizósfera— no ha operado sobre ellos una selección negativa.

La localización cromosomal de todos los genes que son importantes para la colonización de la rizósfera (M negativos, Figura VI.4, panel B) se condice con el carácter ancestral de las etapas más conservadas de las asociaciones entre bacterias y raíces, una habilidad cuyo origen debe remontarse, al menos, a cerca de 450 millones de años atrás cuando las plantas acuáticas comenzaron a poblar la superficie terrestre (Beerling, 2007; Delaux et al., 2015). Como se muestra en la Figura VI.4, panel B, el grupo mayoritario de los marcadores rizosféricos con valores de $M \leq -0,6$ correspondió a genes relacionados con la movilidad y la quimiotaxis bacteriana (34%). Como era de esperar, los genes mutados en *P. agglomerans* LPU12 asociados a funciones de movilidad y quimiotaxis presentan homólogos entre los genes que en *Sinorhizobium meliloti* 2011 afectan la colonización de la rizósfera de alfalfa (Salas et al., 2017; Salas, 2015). Genes relacionados con la movilidad bacteriana y la quimiotaxis han sido reportados como relevantes para la colonización de la rizósfera en interacciones diversas, que incluyen a *Pseudomonas simiae* en asociación con *Arabidopsis thaliana* (Cole et al., 2017) y en *Rhizobium leguminosarum* en asociación con *Pisum sativum* (Wheatley et al., 2020). En términos más genéricos, la motilidad flagelar ha sido recientemente reportada como un rasgo fenotípico fundamental para la adaptación bacteriana al suelo, con relación a estrategias asociadas principalmente a la disponibilidad de recursos (Ramoneda et al., 2024). En la complejidad del ambiente rizosférico, los microorganismos que logran detectar y responder de manera eficiente a los exudados de las plantas obtienen una ventaja competitiva significativa. Los resultados que presentamos muestran que la movilidad y quimiotaxis son mecanismos que operan en *P. agglomerans* para facilitar su acceso a la rizósfera. Dada la ubicuidad de esta bacteria para colonizar plantas podrá investigarse si el perfil de atractantes reconocidos por *P. agglomerans* difiere (¿más genérico?) del correspondiente a otras bacterias que a lo largo de la evolución desarrollaron sesgos más específicos de interacción con plantas como es el caso, por ejemplo, de rizobios y las leguminosas.

A los genes de movilidad le siguen en representación (Figura VI.4, panel B) genes asociados al metabolismo de carbohidratos (10%) y producción y conversión de energía (8%), lo que es consistente con que algunos/varios de esos genes posiblemente se relacionen con el uso de recursos durante la permanencia de las bacterias en la rizósfera. Es además particularmente interesante la alta proporción de genes cuyos COGs son aun de función desconocida (10%), que podrán ser blanco prioritario de análisis en busca de nuevos aspectos de la ecología bacteriana en la rizósfera.

Entre los grupos funcionales COG que tienen una representación menor que 3% y que fueron todos agrupados bajo la categoría OTRAS FUNCIONES en el gráfico de la Figura VI.4, panel B, se encuentra el grupo de genes relacionados con transducción de señales que está representado por el gen *typA* también designado como *bipA*. Dicho gen codifica una GTPasa asociada a la biogénesis de ribosomas, con homólogos en diferentes bacterias incluida *S. meliloti* 1021. En *S. meliloti*, esta proteína regula la adaptación a condiciones de estrés, como bajas temperaturas, bajo pH y presencia de detergentes como el dodecilsulfato de sodio (SDS). En *S. meliloti* se ha reportado que *typA* es relevante para el establecimiento de la simbiosis fijadora de nitrógeno con algunas líneas de *Medicago truncatula*, aunque no resulta esencial para dicha función en asociación con *Medicago sativa* (Kiss et al., 2004). Se sugirió que los mutantes de *typA* en *S. meliloti* pueden estar afectados en algún componente de la superficie celular crucial para una interacción adecuada con ciertas plantas huésped. En relación a γ -proteobacterias, más relacionadas con *P. agglomerans*, se ha reportado que en *Pseudomonas aeruginosa* el producto traducido TypA está involucrado en la virulencia, la resistencia a antimicrobianos y la formación de biopelículas (Neidig et al., 2013); y que en *E. coli* enteropatógenas BipA aumenta la virulencia y reduce la motilidad mediada por flagelos (Grant et al., 2003). Estas evidencias sugieren que posiblemente la relevancia de *typA* en la colonización de la rizósfera esté asociada a cambios en las características de la superficie bacteriana.

VI.4.b El ingreso a la planta de *P. agglomerans* LPU12: en busca de genes asociados a la colonización endofítica de alfalfa.

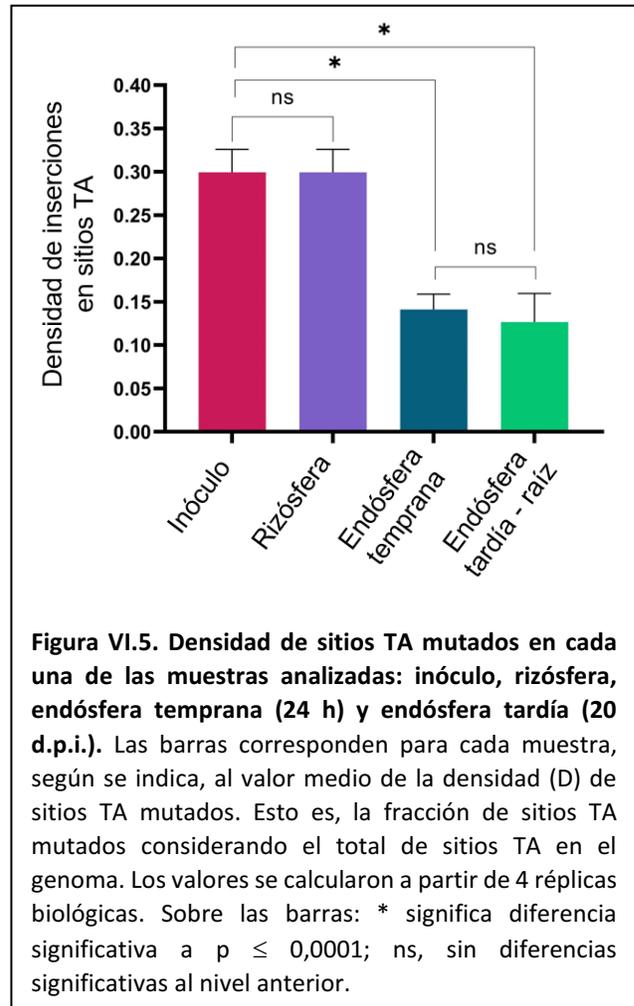
VI.4.b.1 Identificación y caracterización de un cuello de botella que se genera durante el ingreso de *P. agglomerans* LPU12 a la endósfera radicular.

Para la búsqueda de marcadores asociados a la vida endofítica analizamos inicialmente las características de los mutantes que se recuperaron a las 24 h y a los 20 d.p.i. a partir de la maceración de raíces (**MUESTRAS DE SALIDA_ENDO_E** y **ENDO_L**, respectivamente) según se esquematizó en la Figura V.4 del capítulo previo. El primer análisis de los datos mostró claramente que el valor de la densidad de sitios TA mutados respecto de la totalidad de sitios TA del genoma en la muestra endofítica de 24 h $D^{\text{endofítica-24 h}} = 0,141 \pm 0,009$ fue significativamente menor (la mitad) que el valor correspondiente a la densidad en el inóculo que fue $D^{\text{inóculo}} = 0,299 \pm 0,013$ (**MUESTRA DE ENTRADA_Inóculo_E**) (ver en la Figura VI.5 la comparación de las densidades de sitios TA mutados en las diferentes muestras). **Las densidades anteriores reflejan una diversidad de mutantes significativamente menor dentro de la planta que en el inóculo, y ponen en evidencia la existencia de un muy marcado cuello de botella (*bottleneck*) con efecto fundador durante el ingreso de las bacterias al compartimento endofítico.** Los cuellos de botella y efectos fundadores son factores importantes en la deriva génica de poblaciones. Por tanto, dada la relevancia ecológica de ambos, y con el propósito de estimar cuál es el número de bacterias que posiblemente acceden a la endósfera, simulamos el proceso de entrada presuponiendo que las frecuencias de los eventos de entrada a la raíz se producen respetando las abundancias relativas de los genotipos mutantes existentes en el inóculo. La secuenciación de la **MUESTRA DE ENTRADA_Inóculo_E** nos

permitió conocer las cantidades relativas de cada una de las inserciones TA presentes en el inóculo inicial (esto es la frecuencia asociada a cada inserción, es decir, a cada tipo de mutante). A partir de esa información, simulamos el ingreso progresivo de los distintos mutantes respetando sus frecuencias relativas de abundancia, calculando al mismo tiempo cómo era el aumento de la densidad de sitios TA mutados ($D^{\text{endofítica-simulada}}$) a medida que se permitía un ingreso cada vez más grande de mutantes. Así, cuando el valor de la $D^{\text{endofítica-simulada}}$ igualó al valor experimental de la $D^{\text{endofítica-24 h}}$ (esto es 0,141, ver Figura VI.5) sumamos el número de mutantes que virtualmente habíamos ingresado a las plantas, donde tendrán prioridad de ingreso los mutantes más abundantes. A partir de la simulación anterior pudimos inferir cuáles debieron ser los mutantes ingresantes (los afectados en qué genes) y el número de endófitos “fundadores” que resultó en aproximadamente 2.700 u.f.c. de *P. agglomerans* LPU12/planta. Una de las presunciones de nuestro modelo es que, por las características de homogeneidad del inóculo, hemos asumido que las relaciones de frecuencias entre los genotipos que ingresan es comparable antes y después del cuello de botella, situación que no siempre ocurre en cuellos de botella generados por cambios ambientales drásticos

(catástrofes) sobre poblaciones con bajos números de individuos (la representación relativa de los sobrevivientes no representa su frecuencia relativa antes del evento). En la Figura VI.6 se esquematiza de modo simplificado un cuello de botella con efecto fundador que genera una disminución de la densidad D comparable a la observada en nuestro ensayo. En la representación de la figura se muestra un inóculo con sólo 6 genotipos en distinto color (en lugar de los cerca de 50 mil genotipos mutantes de nuestra biblioteca). Cada uno de los genotipos tiene una representación distinta (cantidades y frecuencias en la tabla superior izquierda). De la muestra de salida sólo se recuperaron 3 genotipos (la mitad) que son los que consiguen ingresar a la raíz (cantidades y frecuencias en la tabla superior derecha). En el ejemplo, la densidad de clases de mutantes en la endósfera es la mitad que la densidad existente en el inóculo. Una situación análoga a la que ocurrió en nuestro ensayo.

Como se muestra también en la Figura VI.5, en la endósfera tardía, a los 20 d.p.i, el valor de la densidad $D^{\text{endofítica-20 d.p.i.}} = 0,127 \pm 0,017$ fue comparable a la densidad correspondiente a la endósfera de las 24 h postinoculación ($D^{\text{endofítica-24 h}} = 0,141$). El resultado indica que, a pesar de la presencia continua del inóculo exterior, la diversidad endofítica inicial no aumentó. Según hemos visto en el Capítulo III, ya a las 24 h postinoculación en el compartimento endofítico se alcanzan aproximadamente las 10^5 u.f.c./planta, que se sostendrán al menos hasta 20 d.p.i. (Capítulo III, Figura III.4). Es posible entonces que el ingreso de los fundadores (en un número bajo, cercano a las 2.700 bacterias como hemos descrito en este análisis) y su muy rápido crecimiento posterior, sature el nicho planta “congelando” la diversidad a la generada por esos primeros ingresantes. Los resultados que presentamos muestran que el ensayo



TnSeq, en las condiciones empleadas (especialmente la cantidad de plantas utilizadas) fue una herramienta potente para caracterizar el proceso inicial de ingreso a la endósfera, y el progreso en la diversidad interna de las bacterias que la colonizan.

Los cuellos de botella son factores importantes que afectan las tasas de adaptación de biopoblaciones en evolución (Izutsu et al., 2024). Por un lado, derivado de la pérdida de individuos, son una fuente inicial de disminución de diversidad. Sin embargo, los cuellos de botella proveen al mismo tiempo un espacio de multiplicación reservado sólo a los pocos genotipos que ganan representatividad como consecuencia del mismo. En bacterias, existen experimentos ininterrumpidos de evolución *in vitro* (LTEE, *long term evolution experiments*) que datan de más de 35 años y 75.000 generaciones (Lenski et al., 1991; Tenaillon et al., 2016; Wisser et al., 2013), cuyos resultados han sido empleados como base para nuevos experimentos recientes donde se ha analizado cuál es la dimensión ideal de los cuellos de botella (emulados en cultivos *in vitro* a través de diluciones de distinta magnitud) para que los mismos favorezcan la fijación de mutaciones benéficas (Izutsu et al., 2024). En este trabajo, las ganancias de aptitud (*fitness*) comenzaron antes y fueron mayores cuando los cuellos de botella generaron disminuciones poblacionales afectadas por un factor de 100 o 1.000, en consistencia con los modelos que predicen la aparición frecuente de mutaciones benéficas distintas y con interferencia clonal. Lamentablemente, no conocemos cuál es la dinámica mutacional de *P. agglomerans* en sus ciclos de asociación con la planta. En nuestros ensayos el número de bacterias en el entorno rizosférico fue de $10^5 - 10^6$ y luego en la endosfera 10^3 (relación 1:100 a 1:1.000), condiciones diferentes a las que operan en ambientes naturales donde es difícil generalizar cuántas de las bacterias del suelo ingresan a la raíz (cuál es el factor de dilución). Según hemos reportado previamente (López et al., 2018), del interior de semillas de alfalfa hemos podido recuperar 40 géneros diferentes de bacterias que con seguridad contribuyen a conformar la comunidad rizosférica temprana. Será importante explorar cuál es la representación de *P. agglomerans* en las rizósferas y endósferas de semillas de alfalfa germinadas a campo, de modo de conocer cómo las restricciones que aquí hemos detectado operan para esta especie bajo la diversidad de todo el microbioma en el escenario natural. Dado que el espacio endofítico es limitado, para definir la ocupación del mismo son seguramente importantes la capacidad y velocidad con que cada especie rizosférica ingresa a la planta, y la velocidad con que cada una se duplica en el nuevo nicho para conformar el microbioma endofítico final.

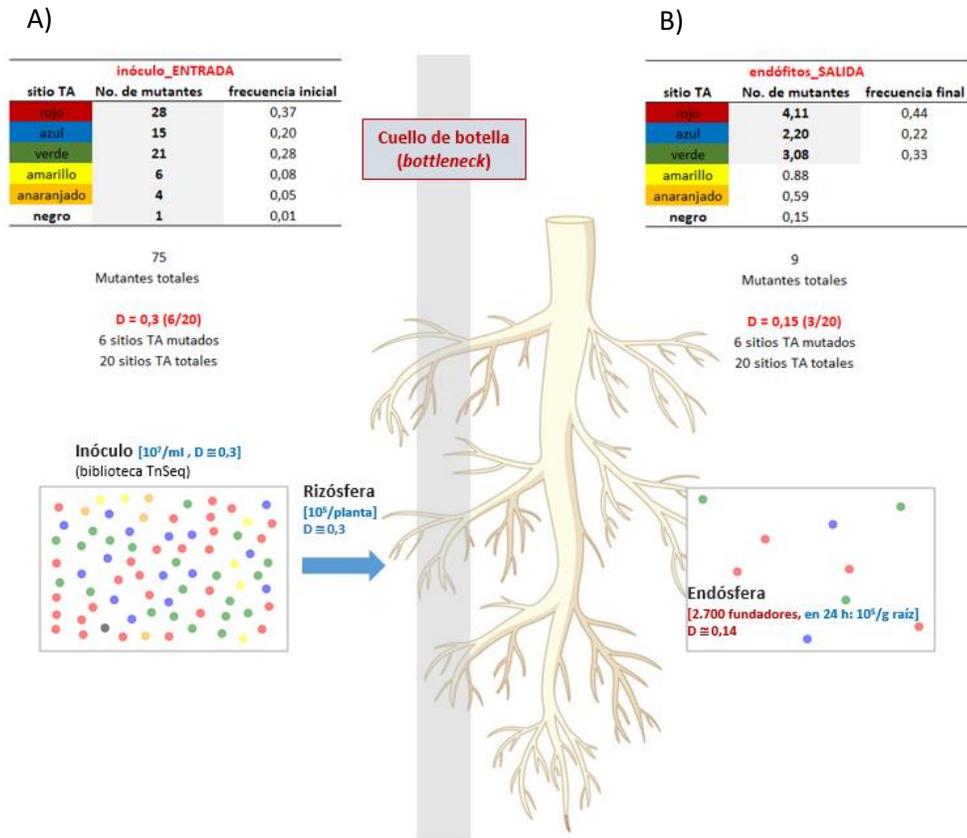


Figura VI.6. Representación esquemática de un cuello botella y efecto fundador operando durante el acceso de los distintos genotipos al comportamiento endofítico de una raíz. Panel A. En la zona izquierda del gráfico se representa la situación inicial en el inóculo, en este caso conformado por 6 genotipos distintos, esto es, mutantes Tn con inserciones en 6 sitios TA diferentes. Como se aprecia en el dibujo los mutantes tienen una representación numérica diferente, siendo los “rojos” y los “negros” los más y menos abundantes, respectivamente. En la tabla superior izquierda se indica el número de cada mutante en el inóculo, y su abundancia relativa (frecuencia). En el ejemplo se supone un total de 20 sitios TA eventualmente mutables, lo que resulta en una densidad de tipos de sitios mutados $D^{\text{inóculo}} = 0,3$. Panel B. En la zona derecha del gráfico se representa de forma esquemática el escenario temprano en el compartimento endofítico, en este caso conformado por sólo 3 genotipos mutantes diferentes como consecuencia de la restricción al ingreso impuesta por el cuello de botella. Las restricciones operaron sobre todos los genotipos, sin impedir que ingresen representantes de los genotipos más abundantes (“rojo” > “verde” > “azul”; en proporción 4 : 2 : 3), y en proporción comparable a su representación en el inóculo inicial (“rojo” > “verde” > “azul”; en proporción 28 : 15 : 21). En la tabla superior derecha se indica el número de cada uno de los 3 mutantes endofíticos con sus abundancias correspondientes (frecuencias), que resultan mayores que las que tenían en el inóculo debido a la desaparición de los 3 genotipos menos abundantes. Las sombras grises en las tablas de izquierda y derecha indican los genotipos mutantes que están presente al inicio y al final en cada compartimento, inóculo y endósfera. Como consecuencia del ingreso de sólo una parte de los genotipos mutantes, en el compartimento endofítico resulta una densidad de tipos de sitios mutados $D^{\text{compartimento endofítico}} = 0,15$ menor que la existente en el inóculo ($D=0,3$). En la parte inferior del gráfico, en celeste, se presentan los valores reales de inóculo, densidades D y fundadores correspondientes al experimento TnSeq cuyos resultados se presentan en el texto.

VI.4.b.2 Identificación de genes asociados a la colonización de la endósfera de alfalfa.

La puesta en evidencia del cuello de botella y efecto fundador que se describió en la sección precedente, debió ser tenida en cuenta a la hora de buscar e identificar mutaciones en genes asociadas a la capacidad de las bacterias de ingresar a la planta. De modo análogo al estudio de la rizósfera, procedimos en primer lugar a comparar la representación de mutantes en la **MUESTRA DE ENTRADA_Inóculo_E** respecto de la representación de mutantes en las **MUESTRAS DE SALIDA_ENDO_E y _ENDO_L**. Sin embargo, dado que como consecuencia del cuello de botella ingresó un número bajo de bacterias/panta, se generó una fuerte restricción al número de mutantes/gen que pueden ser evaluados con peso estadístico. Por tal razón, haciendo uso de nuestra predicción de fundadores y su abundancia, primeramente ordenamos los diferentes mutantes en orden decreciente de representación (de los mutantes predichos como más representados a los menos representados en la endósfera). Luego, calculada la mediana, sólo consideramos los valores de M que correspondían a mutantes del 50% más representado. Tal restricción, que atendió sólo a los 2 cuartiles mejor representados, redujo el número de genes evaluados a cerca de 600 (exactamente 591 genes) del total del genoma de más de 4 mil genes (4.605 genes). En segundo lugar, dado que aquellos genes que influyan en la colonización temprana a las 24 h deberían también modificar la colonización tardía a los 20 d.p.i. (dado que la primera es condición esencial para la segunda), seleccionamos sólo los 72 genes que en ambas condiciones (temprana y tardía) mostraron: (a) valores de M con signos coincidentes, (b) valores $M \leq -1$ ó $M \geq 1$, y (c) valor- $p \leq 0,01$. De esos genes, se excluyeron 7 mutantes que correspondieron a mutaciones que generaron fenotipos de auxotrofia y relacionados, que ya describimos en la sección VI.2.b previa. Como producto de las restricciones anteriores el conjunto de genes que asociamos a la colonización endofítica se redujo a 65. En la Figura VI.7 se presenta una gráfica de tipo "volcán" en la que se muestran los genes con $M \leq -1$ (en rojo) y $M \geq 1$ (en verde) en ambos casos con valor $p \leq 0,01$ en la endósfera a las 24 h (panel A) y en la endósfera a los 20 d.p.i. (panel B). En el panel C de la figura se presentan los 65 mutantes comunes a las dos condiciones, seleccionados con los criterios detallados en el párrafo anterior. En la Tabla VI.S6 suplementaria se lista cada uno de los 65 mutantes seleccionados con sus correspondientes códigos COG y KO y valores de M (a 24 h y 20 d.p.i.). En esa lista, en contraste con lo observado en los genes asociados a la colonización de la rizósfera, son mayoría los genes cuya alteración resulta en una mejor colonización de la endósfera (46) respecto a los que la perturban (19). La presencia de tantos genes cuya mutación mejora el ingreso a la endósfera sugiere que en el ambiente endofítico, mucho más homeostático y definido que el del suelo, *P.agglomerans* puede prescindir muchos de los genes presentes en el genoma que seguramente la bacteria usa en su vida no-asociativa. Además, a partir de la inspección de los valores de M (sean negativos o positivos) se observa que los efectos de las mutaciones se hacen más marcados con el paso del tiempo, a los 20 d.p.i. más que a las 24 h post inoculación. Esto es, la pérdida de funciones importantes (M negativos) o la ganancia de ventajas (M positivos) para la colonización de la endósfera muestran un efecto más marcado en el estadio más tardío que analizamos. En la Figura VI.8 se muestra de manera sintética cómo es la distribución funcional asociada a los genes de ambos grupos de mutantes.

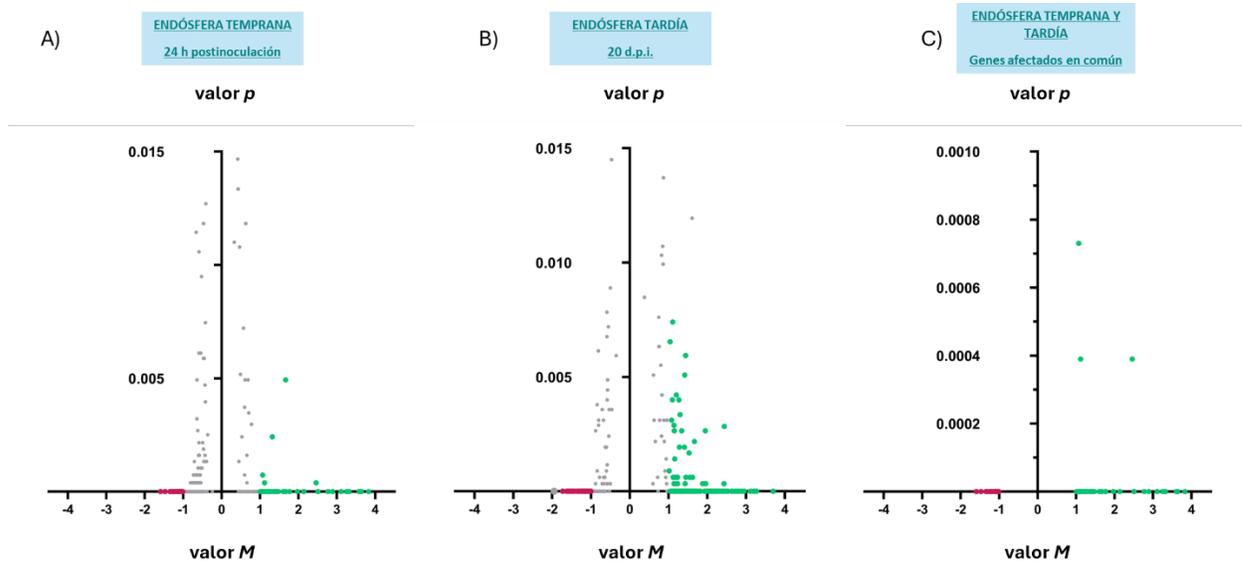


Figura VI.7. Gráfico tipo volcán en el que se representan los valores M y los correspondientes valores p para genes seleccionados de la biblioteca al comparar el inóculo con las muestras de salida de la endósfera de alfalfa. **Panel A: colonización temprana.** Comparación de la **MUESTRA DE ENTRADA_Inóculo_E** con **MUESTRAS DE SALIDA_ENDO_E** (24 h postinoculación). Como se indica en el texto, en el gráfico sólo se incluyen los mutantes en genes que correspondieron al 50% predicho como más representado entre los fundadores, y que además mostraron valores $M \leq -1$ ó $M \geq 1$; y valor $p \leq 0,01$. **Panel B: colonización tardía.** IDEM al panel A pero resultante de la comparación de la **MUESTRA DE ENTRADA_Inóculo_E** con **MUESTRAS DE SALIDA_ENDO_L** (20 d.p.i.). **Panel C: marcadores coincidentes entre ambas condiciones.** Se muestran los 65 mutantes cuyos genes corresponden a aquellos que en ambas condiciones (paneles A y B) tuvieron signo de M coincidente, $M \leq -1$ (rojo) ó $M \geq 1$ (verde) y no formaron parte de la lista de auxótrofos y relacionados (sección VI.2.b). Para los mutantes en los 65 genes que se representan en este panel se utilizaron los valores M y p del ensayo de colonización endofítica a 24 h postinoculación. En la Tabla VI.S6 suplementaria se muestran los 65 genes seleccionados con sus respectivos códigos KO y COG. De los tres paneles se excluyeron los mutantes que correspondieron a mutaciones que generaron fenotipos de auxotrofia y relacionados (ver texto).

Dentro del conjunto de los 19 mutantes con $M \leq -1$ tuvieron mayor presencia (20%) los relacionados con alteraciones de la membrana y la envoltura de la bacteria, una evidencia razonable si la misma se asocia al proceso de infección donde el contacto físico del microorganismo con el hospedador es muy estrecho. Le siguieron tres grupos igualmente representados (10% cada uno) relacionados con la transducción de señales, el tráfico intracelular/secreción y el metabolismo de carbohidratos; todas funciones compatibles con la presencia de procesos de comunicación bacteria-planta. Por su parte, en el conjunto mayoritario de los genes con $M \geq 1$ resultaron mayoritarios los de función desconocida (26%), seguidos nuevamente por los vinculados a membrana y pared celular (11%) y metabolismo de carbohidratos (6,5%); y luego por un conjunto heterogéneo de funciones diversas (24%, OTRAS FUNCIONES).

La comparación de ambos grupos resalta la importancia de la superficie celular y la energía en el proceso endofítico. Entre los productos génicos de membrana importantes para la colonización endofítica se encuentra, por ejemplo, la proteína RcsC (de *regulator of capsule synthesis*) (*locus_tag* GONCAC_06185, gen *rscC*). Dicha proteína en enterobacterias es parte de un sistema de dos componentes asociado a la regulación de respuestas a estreses de envoltura (ESR, *envelope stress responses*) y a la regulación de la virulencia (CHEN et al., 2001; Meng et al., 2021; Stout & Gottesman, 1990). La cápsula bacteriana es una estructura importante para la interacción con el entorno y la evasión del sistema inmunológico de la planta, por lo que mutantes *rscC* es razonable que estén afectados en su capacidad

colonizadora. El impacto negativo de alteraciones en la envoltura celular se puso también en evidencia en mutantes con alteraciones en la síntesis del péptidoglicano (*locus_tag* GONCAC_04745, gen *mltF*).

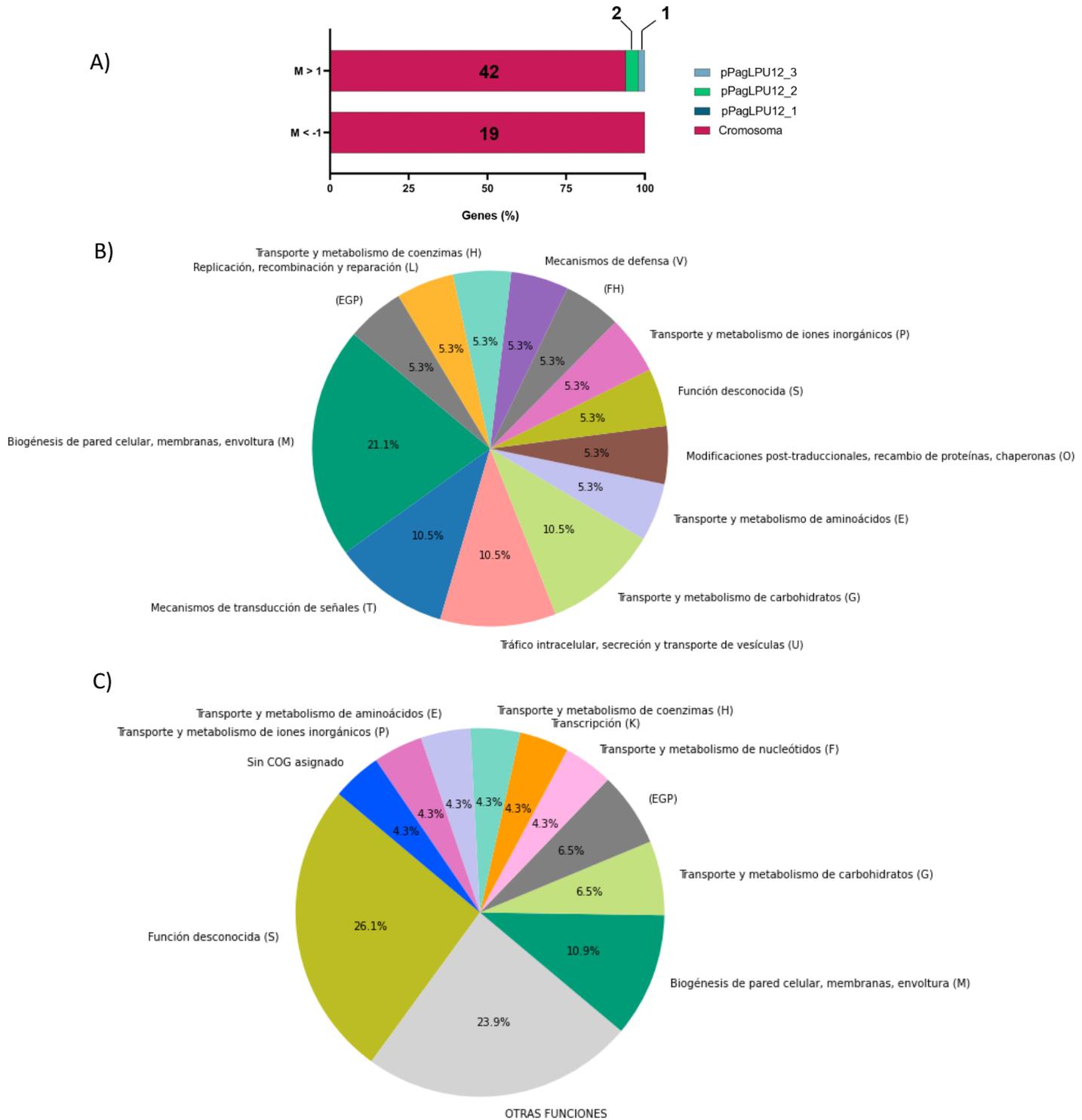


Figura VI.8. Localización genómica y distribución de funciones en el grupo de genes de *P. agglomerans* LPU12 que modifican la capacidad de colonizar la endósfera de alfalfa. Panel A. Distribución de mutaciones según el replicón en el que se encuentran (cromosoma o plásmido) y según su valor de M. **Panel B.** Asignación funcional realizada a

partir del análisis de los códigos COG de cada uno de los genes vinculados a la colonización de la rizósfera cuyos mutantes resultan en valores de $M \leq -1$ (ver Tabla VI.S6). **Panel C.** Asignación funcional realizada a partir del análisis de los códigos COG de cada uno de los genes relacionados a la colonización de la rizósfera cuyos mutantes resultan en valores de $M \geq 1$. Tanto para el Panel **B** como **C** los diferentes grupos COG (A – Z, leyendas y letras en el gráfico) se representan en distinto color con su abundancia porcentual correspondiente. Todos aquellos grupos cuya representación fue menor que 3% se agruparon bajo la denominación OTRAS FUNCIONES.

Efectos opuestos a los anteriores ($M \geq 1$, favoreciendo el fenotipo de colonización) en mutantes de superficie se observaron en los afectados en la síntesis del lípido A del lipopolisacárido (LPS), componente principal de la membrana externa. Mutantes en los genes *lpxP* (*locus_tag* GONCAC_00185) o *eptB* (*locus_tag* GONCAC_18770) impactaron positivamente en la capacidad de colonizar el compartimento endofítico. En *Xanthomonas campestris* la pérdida del antígeno-O del LPS (la porción sacarídica más externa del LPS, con formación de lipooligosacáridos, LOS) había sido reportada como una modificación que dispara mecanismos de defensa en *Arabidopsis* (Cavaco et al., 2021; Proietti et al., 2014). En nuestro caso, mutaciones que podrían modificar el lípido A, la porción más interna del LPS, resulta en una colonización más marcada. El contraste entre ambas evidencias muestra lo complejo de predecir el efecto que cambios superficiales como los anteriores podrán tener sobre la calidad de la asociación, perturbándola (como en mutantes *rscC* o *mltF*) o promoviendo (mutantes *lpxP* o *eptB*). El fenotipo resultante dependerá seguramente de cómo las modificaciones en la envoltura de cada mutante impactan en la física de la interacción (características de hidrofobicidad/polaridad) y también en el intercambio de metabolitos/señales entre ambos participantes.

En relación con el análisis que presentamos debe destacarse que, si bien hemos identificado 65 genes potencialmente asociados a la vida endofítica, debido a las restricciones del cuello de botella sólo hemos incluido en el análisis a mutantes en 591 genes de los 4.605 que anotamos en el genoma. Como ya indicamos, por las restricciones al número de mutantes que acceden a la endósfera definimos criterios de inclusión sólo para los mutantes predichos como los fundadores más abundantes. Hará falta diseñar un ensayo que alcance una mayor densidad D de sitios TA mutados en la endósfera de modo de tener bien representados a mutantes en todos los genes (los 4605 en lugar de sólo 591). Tal condición podrá ser alcanzada sólo si se aumenta significativamente el número de plantas (en un factor de al menos 10), como se ha realizado en otros estudios donde se inoculan 100 plantas por réplica para investigar la colonización del apoplasto de hojas de *Phaseolus vulgaris* (Helmann et al., 2019), de forma tal que aumente la cantidad de bacterias endofíticas que se colectan en la muestra de salida (si se tienen cerca de 2.700 fundadores iniciales/planta, con 150 plantas se podrán colectar cerca de 400.000 endófitos, que permitirán contar, en promedio, con cerca de 100 endófitos para cada uno de los cerca de 4.100 genes no esenciales de *P. agglomerans*). Un experimento como ese, factible y en el que en lugar de 591 genes podrán explorarse cerca de 4.100 (los no esenciales), es posible que aumente a cerca de 450 (en un factor cercano a 7) la cantidad de genes que resulten asociados a la vida endofítica representando muy posiblemente cerca de 10% de todo el genoma. Cabe señalar, sin embargo, que el experimento con representación completa no tiene la configuración adecuada para identificar y caracterizar cuantitativamente los cuellos de botella, ya que para ello se requieren condiciones de densidad D bajas en el compartimento endofítico respecto de las del inóculo, de modo de simular un ingreso (limitado) de mutantes que conlleve a la misma densidad D observada experimentalmente según presentamos en la sección VI.3.b.1 de este capítulo.

VI.5. Consideraciones finales del capítulo

El ensayo TnSeq cuyo análisis presenté en este capítulo permitió identificar en *P. agglomerans* LPU12: **(a)** los genes que son esenciales de esta cepa, **(b)** sus genes de auxotrofia y relacionados, **(d)** los

genes cuya mutación (en contraste con los de auxotrofia) favorecen el crecimiento de la bacteria en medio mínimo respecto del crecimiento en medio rico, **(e)** los genes cuya mutación causa pérdida parcial de aptitud (*fitness*) durante el crecimiento en medio rico, **(f)** los genes relacionados a la colonización de la rizósfera de alfalfa, y **(g)** los genes relacionados a la colonización de la endósfera de la misma planta. En cada una de las categorías anteriores pudimos caracterizar perfiles funcionales generales, y en varios casos identificar genes asociados a componentes particulares como los de movilidad, quimiotaxis, sistemas de dos componentes, estructuras superficiales, y rutas metabólicas específicas. La búsqueda de homólogos a esos genes permitirá explorar la eventual conservación de componentes y mecanismos empleados por las bacterias en su proceso de asociación con plantas.

Los resultados mostraron que en *P. agglomerans* LPU12 existen más de 500 genes esenciales y cerca de 162 genes de auxotrofia (además de un número comparable de genes con efecto contrario, cuya mutación favorece el crecimiento en medio mínimo respecto del crecimiento en medio rico). En relación a la colonización de la planta identificamos 112 genes cuya mutación afecta el acceso a (/crecimiento en) la rizósfera, de los cuales 46 (1% del total del genoma) no se asociaron a mutaciones de auxotrofia y relacionadas. La mayoría de las mutaciones en los genes anteriores afectaron específicamente la colonización rizosférica, siendo además minoría los genes cuya pérdida mejoran la representación de la bacteria en la rizósfera. Exceptuando a los genes de auxotrofia, el bajísimo número observado de genes que perturben la colonización de la rizósfera (sólo 8, de los cuales 3 fueron plasmídicos) puso en evidencia una selección génica purificadora negativa de dichos genes. En relación a la colonización de la endósfera la imagen fue diferente. Identificamos 72 genes, de los cuales la mayoría, 65 (1,4% del genoma), no se relacionaron con genes de auxotrofia. De estos 65 genes que cambiaron específicamente la colonización de la endósfera, la mutación de menos de un tercio, 19 (0,4 % del genoma), generó efectos de menor colonización mientras que mutantes en los 46 restantes (1% del genoma) fueron mejores colonizadores. Esta evidencia contrasta con lo observado para los genes rizosféricos, y sugiere fuertemente que muchos genes —que aún son parte del genoma— son seguramente importantes para distintas instancias de la vida no-asociativa, pero no durante la condición mucho más homeostática de la vida endofítica (piénsese en la reducción genómica que ha operado en bacterias intracelulares obligadas (Otero-Bravo & Sabree, 2021)). Cabe recordar que en el experimento en que estudiamos la colonización endofítica usamos una alta concentración de *P. agglomerans* LPU12 en el inóculo (cerca de 10^7 u.f.c./ml), de modo de no limitar el acceso de las bacterias a la rizósfera y concentrar el ensayo en la detección de aquellos mutantes que estaban alterados en la colonización de la endósfera. En ensayos con inóculos bajos, como los que usamos en los estudios rizosféricos (cerca de 10^5 u.f.c./ml), es muy posible que los genes que afectan la colonización de la rizósfera también sean importantes para la colonización de la endósfera (algo que no evaluamos) por constituir esta última una etapa previa y obligada de la llegada a la endósfera.

Según discutimos en detalle, pudimos describir cuantitativamente una fuerte restricción al ingreso de las bacterias a la raíz de alfalfa. Los resultados que observamos mostraron que las primeras bacterias infectantes crecen rápidamente dentro de la planta, saturan el nicho endofítico en menos de 24 h, y restringen el aumento de la diversidad interna. Tal dinámica aparece como un mecanismo básico para garantizar un espacio de propagación reservado a los genotipos fundadores. Será interesante analizar cómo progresa la diversidad dentro de la raíz en el tiempo, cuando se coinoculan dos bacterias endofíticas con diferentes tasas de multiplicación en la endósfera. Avanzar en cómo los cuellos de botella operan en bacterias con diferentes propiedades de infección y crecimiento, en mezclas, será de especial valor a la hora de diseñar mejores consorcios sintéticos (*SynCom*) con bacterias de probada actividad de promoción del crecimiento de plantas (PGP). En nuestro laboratorio hemos podido identificar que dentro de las propias semillas de alfalfa, y persistiendo por períodos muy largos, existe una comunidad muy diversa de bacterias que incluye a más de 40 géneros diferentes (López et al., 2018). A partir de las evidencias que

mostramos en este capítulo esperamos avanzar en estudios extendidos donde, con nuevos criterios de diseño experimental, podamos abordar el estudio de cómo es la dinámica relativa de esa comunidad para colonizar la raíz de la planta joven, y las consecuencias de esos eventos en la definición del microbioma radicular tardío.

Estudios como los de TnSeq que mostré en este capítulo se basan en el reconocimiento previo de bacterias que, como *P. agglomerans* LPU12, se asocien con plantas. Pero no todas las *Pantoeas* tienen tropismo por las plantas. Por tanto, en instancias previas a estudios como los de este capítulo, es valioso disponer de un método de tamizado que, de manera práctica y eficiente, discrimine que cepas particulares de un dado género son potenciales habitantes de las plantas. Tomando *Pantoea* como género modelo, en el capítulo siguiente presentaré, con una aproximación basada en aprendizaje de máquina supervisado (*Machine Learning*), el diseño y aplicación de técnicas clasificatorias de bosques aleatorios (*Random Forest*) para la predicción de fenotipos bacterianos como colonizadores o no-colonizadores de plantas.



CAPÍTULO VII

CUANDO LA MESADA NO ESTÁ DISPONIBLE:
EXPLORANDO EL USO DE MACHINE LEARNING PARA LA PREDICCIÓN DEL
TROPISMO A PLANTAS EN *GAMMAPROTEOBACTERIAS* MODELO



VII.1. Introducción

En el trabajo desarrollado hasta aquí en esta tesis utilizamos un enfoque experimental que busca asociar genes con un fenotipo en estudio, lo que nos permitió focalizar y dirigir una búsqueda, en principio extensiva, a la identificación de marcadores bacterianos involucrados en la asociación con plantas de alfalfa en el genoma de *P. agglomerans* LPU12. Nuevos interrogantes y vías de estudio se abren a partir de los resultados obtenidos respecto a cuáles son aquellas características genéticas que le permiten a una bacteria abrirse camino desde el suelo circundante hasta colonizar distintos compartimentos vegetales. Conociendo la capacidad de esta técnica usada, y de otras metodologías TIS reportadas anteriormente, para asociar un fenotipo de interés con genes bacterianos relacionados, nos propusimos abordar la posibilidad de predecir, a partir de la información genómica de un microorganismo, su capacidad de asociación con plantas ya sea como simbiote o patógeno. El desarrollo de métodos de predicción basados en el aprendizaje supervisado de algoritmos ha demostrado ser eficaz en diversas áreas, permitiendo generar predicciones precisas a partir de datos genómicos. Nos planteamos si la información contenida en los genomas bacterianos es lo suficientemente rica y consistente como para entrenar algoritmos capaces de predecir un fenotipo tan complejo y variable como la asociación entre bacterias y plantas, ya sea en relaciones mutualistas o patogénicas. El desafío radica en que la capacidad de una bacteria para interactuar con plantas involucra múltiples factores genéticos y ambientales, los cuales varían incluso al analizar distintos compartimentos de una misma planta. A priori, se desconoce cuáles genes desempeñan roles esenciales en esta asociación. No obstante, se ha reportado el uso de métodos de clasificación para predecir la relación de diversas bacterias con la rizósfera y los tejidos radiculares, en función de sus características genéticas, y han permitido también la identificación de marcadores genéticos que parecen ser ubicuos en las bacterias asociadas a plantas (Levy et al., 2018; López et al., 2023). Considerando el éxito de estos enfoques, existe un gran potencial para aplicar modelos predictivos que identifiquen patrones genéticos vinculados a la capacidad de colonización vegetal, aportando una nueva perspectiva en la investigación sobre interacciones planta-bacteria.

Como se discutió en el Capítulo VI al analizar los marcadores genéticos identificados mediante ensayos TnSeq, los genomas bacterianos muestran evidencias de adaptación evolutiva al entorno que habitan, reflejado en un bajo número de marcadores cuya eliminación mejora el fenotipo de colonización (Figura VI.3.a). Estos resultados sugieren que las bacterias asociadas a plantas poseen un contexto genómico caracterizado por la presencia de genes que favorecen su capacidad de interactuar con el entorno vegetal. En el caso de bacterias filogenéticamente cercanas, los atributos genómicos comunes en cepas que colonizan plantas pueden facilitar su clasificación en términos de tropismo hacia plantas. De manera complementaria, las diferencias genómicas entre cepas filogenéticamente relacionadas que colonizan ambientes distintos deben ser lo suficientemente consistentes como para permitir una clasificación precisa. Esta premisa subraya que la información genómica debe ser lo suficientemente robusta y específica para permitir la diferenciación entre cepas con una alta relación filogenética, pero que expresan fenotipos claramente diferentes.

Nos propusimos entonces evaluar la capacidad de algoritmos de predicción basados en aprendizaje supervisado para identificar el tropismo hacia plantas en bacterias. Dado que la alta similitud filogenética entre cepas presenta una dicotomía —por un lado, proporcionar la información común suficiente para identificar patrones y, por otro, requerir diferencias claras para distinguir fenotipos—, consideramos crucial analizar la capacidad de estos algoritmos en cepas pertenecientes a un género bacteriano. Como se describió en la sección 1.5 de la Introducción, las bacterias del género *Pantoea* son de particular interés para este tipo de análisis debido a su amplio rango de huéspedes y la diversidad de interacciones que establecen, que van desde relaciones simbióticas hasta patogénicas. Esta diversidad

convierte al género *Pantoea* en un modelo ideal para explorar la capacidad predictiva de los algoritmos supervisados. En este capítulo, se demostrará cómo los algoritmos *Random Forest* y *Logistic Regression*, entrenados con información funcional derivada de genomas disponibles en bases de datos públicas, son capaces de predecir el tropismo hacia plantas en bacterias del género *Pantoea*.

VII.1.a Algoritmos de aprendizaje automático - Machine Learning – para la clasificación:

Random forest y Logistic Regression

Los algoritmos supervisados de clasificación buscan prever la etiqueta que caracteriza a un objeto. El modelo que realiza la clasificación es entrenado con un conjunto de datos, y en base a lo aprendido se evalúa su capacidad de clasificación con un conjunto de datos de prueba antes de utilizarse para realizar predicciones con nuevos datos no vistos. Se han desarrollado diversos algoritmos que pueden entrenarse para generar modelos de clasificación. Cada algoritmo tiene un marco teórico diferente que rige el fundamento sobre el cual se basa la toma de decisiones que realiza para clasificar una variable. Difieren también en su capacidad de manejar diferentes volúmenes de datos, el tiempo de cómputo que requieren y el poder de predicción que tienen. La elección de uno u otro algoritmo dependerá de los datos con los que se cuenta para entrenar el modelo, el tipo clasificación que se quiere realizar, y la relación que exista entre la información con la que son entrenados y la variable objetivo a clasificar.

El algoritmo *Random Forest* (RF) (Breiman, 2001) se ha hecho muy popular para el reconocimiento de patrones en datos a escala ómica, principalmente porque proporciona dos aspectos que son muy importantes para la minería de datos: alta precisión de predicción e información sobre la importancia de las variables para la clasificación. El rendimiento en la predicción de RF es comparable con otros algoritmos de clasificación (Verikas et al., 2011) como las Máquinas de Vectores Soporte - *Support Vector Machines*, SVMs - (Cortes & Vapnik, 1995), las Redes Neuronales Artificiales - *Artificial neural networks* - (McCulloch & Pitts, 1943; Rosenblatt, 1958), los Clasificadores Bayesianos - *Bayesian classifiers* - (Friedman et al., 1997) y la Regresión Logística – *Logistic Regression* - (Kleinbaum et al., 1982; Nick & Campbell, 2007). Los modelos predictivos de RF se entrenan con un conjunto de árboles de decisión individuales basados en las muestras en estudio, el cuidadoso etiquetado de cada muestra y variables que caracterizan a cada una de ellas. En la Figura VII.1, extraída de Touw et al. (2013), se esquematiza el funcionamiento de *Random Forest* para la clasificación de muestras. Un modelo RF se construye creando múltiples árboles de decisión, cada uno entrenado con un conjunto de datos diferente generado mediante bootstrapping. Este proceso implica seleccionar muestras aleatorias del conjunto original con reemplazo, formando un nuevo conjunto que contiene aproximadamente dos tercios de las muestras. Las muestras no seleccionadas se denominan OOB (Out of Bag) y se usan para evaluar el rendimiento del modelo. Cada árbol crece dividiendo recursivamente las muestras del conjunto bootstrap. En cada nodo, se selecciona un subconjunto aleatorio de variables y se elige la que produce la mayor disminución de impureza para dividir las muestras en dos "nodos hijos". Las métricas más comunes para medir la impureza son el índice *Gini*, que refleja la probabilidad de una clasificación incorrecta, y la entropía, que mide el "desorden" o incertidumbre en los datos. Usar una u otra métrica suele tener un impacto mínimo en el rendimiento del modelo. El proceso de división continúa hasta que se alcanzan los nodos terminales o "hojas", que son puros (contienen solo muestras de una clase). Cada árbol, una vez entrenado, emite un voto no ponderado para clasificar nuevas muestras, y la clase final se decide por votación mayoritaria entre todos los árboles. Un árbol individual es un clasificador débil, ya que se entrena en un subconjunto de datos y variables, pero al combinar muchos árboles con baja correlación, *Random Forest* se convierte en un

clasificador fuerte. Esta baja correlación se logra gracias a la aleatoriedad en la selección de datos y variables, lo que también ayuda a prevenir el sobreentrenamiento.

Además, este algoritmo puede estimar también la importancia de las variables para la clasificación (Breiman, 2001), lo que permite interpretar la relevancia de las variables para el conjunto de datos en estudio, y la valoración de la importancia para cada variable puede utilizarse para identificar marcadores de interés (Fusaro et al., 2009) o como filtro para eliminar variables no informativas (Díaz-Uriarte & Alvarez de Andrés, 2006). Todas estas características hacen de este un potente método que permite la extracción de conocimiento relevante adicional de datos ómicos, convirtiéndolo en uno de los algoritmos más ampliamente utilizado en estudios de ciencias de la vida (Touw et al., 2013).

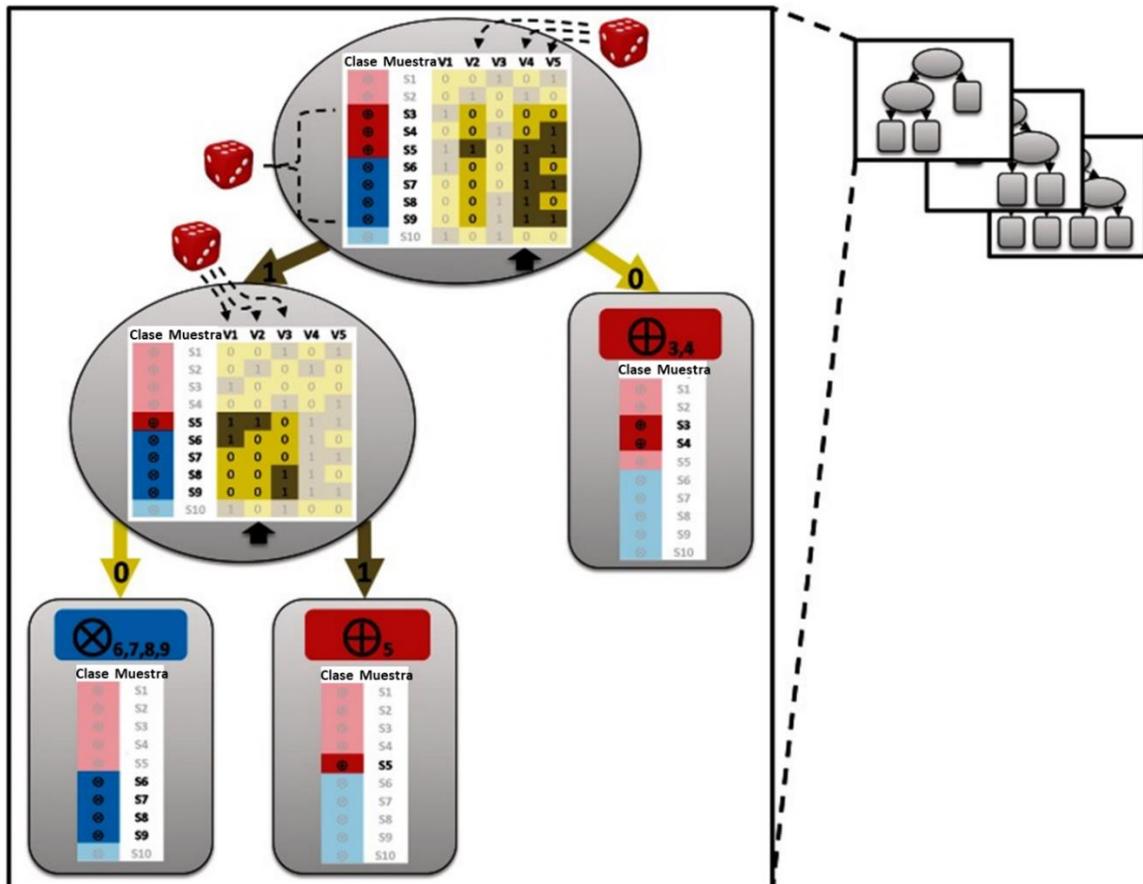


Figura VII.1: Extraída y adaptada de Touw et al. (2013). Entrenamiento de un árbol individual de un modelo de *Random Forest*. El árbol se construye a partir de una matriz de datos (que se muestra dentro de las elipses). Esta matriz consta de muestras (S1 a S10) pertenecientes a dos clases (cruces azules o signos "+" rojos) y medidas de cada muestra para distintas variables (V1-V5). Los dados simbolizan la selección aleatoria. Líneas punteadas: muestras y variables seleccionadas aleatoriamente. Para cada árbol, se crea un conjunto *bootstrap* recolectando muestras del conjunto de datos al azar y con reemplazo, hasta que contenga tantas muestras como hay en el conjunto de datos. En este ejemplo, el conjunto *bootstrap* contiene siete muestras únicas (muestras S3-S9; las muestras no seleccionadas S1, S2 y S10 están descoloridas). Para cada nodo (indicado como elipses) se seleccionan aleatoriamente algunas variables (aquí tres; las otras dos variables no seleccionadas se muestran descoloridas; en RF por defecto el número de variables seleccionadas es igual a la raíz cuadrada del número total de variables) y se evalúa su capacidad para dividir los datos. La variable que produce la mayor disminución de impurezas se elige para definir la regla de separación de las muestras. En el caso del nodo superior, es V4 y en el del segundo nodo de la

izquierda es V2 (indicado con las flechas negras). Este proceso se repite hasta que los nodos son puros y contienen solo muestras de una misma clase (indicados con recuadros de bordes redondeados). Cada árbol del modelo da como resultado un valor de clasificación, que es recopilado junto con el resultado provisto por los restantes árboles, generando un consenso que define la clasificación final de la muestra.

Si bien el método *Random Forest* es adecuado para la construcción de un modelo que prediga el tropismo hacia plantas, decidimos evaluar también el potencial de otro algoritmo de clasificación basado en un enfoque diferente. Un buen desempeño de distintos algoritmos entrenados con los mismos datos indicaría que la información contenida en los genomas es robusta para predecir fenotipos de interés, independientemente del algoritmo empleado. En este sentido, optamos por utilizar un Modelo Lineal Generalizado (GLM, por sus siglas en inglés, *Generalized Linear Models*), el cual se basa en una "función de enlace" que permite modelar relaciones no lineales dentro de un marco lineal, ofreciendo así mayor flexibilidad al analizar relaciones entre variables. Entre los algoritmos incluidos en los GLM, la regresión logística - *Logistic Regression* (LR) - es particularmente útil para realizar clasificaciones binarias, ya que modela la probabilidad de que una observación pertenezca a una de dos clases (Nick & Campbell, 2007). Este método es una extensión de la regresión lineal que utiliza una función sigmoide, la cual genera un valor entre 0 y 1 que puede interpretarse como la probabilidad de que una muestra pertenezca a una clase en particular. Si dicha probabilidad supera un umbral predeterminado (normalmente 0.5), la observación se clasifica como perteneciente a la clase positiva (clase "1" en el sistema binario que emplea); de lo contrario, se asigna a la clase negativa (clase "0").

Logistic Regression es un método donde los coeficientes que caracterizan el modelo reflejan el impacto que cada variable independiente tiene sobre la probabilidad de que ocurra el evento de interés (por ejemplo, pertenecer a una clase particular) (Stoltzfus, 2011). En el contexto del modelo, el término "solver" se refiere al algoritmo que se utiliza para la optimización de sus parámetros, donde se busca reducir el error medido por una función de costo (log-loss), que compara las probabilidades predichas con las verdaderas. Dado que la regresión logística es un algoritmo de clasificación sencillo y robusto para problemas binarios como el que planteamos aquí, consideramos que evaluar su rendimiento para la clasificación de bacterias asociadas o no a plantas nos permitirá estimar si existe una relación lineal lo suficientemente informativa entre la información genética y este complejo fenotipo.

Estos modelos de aprendizaje supervisado, que han mostrado ser eficaces en otras áreas de predicción genética, se aplicarán para identificar patrones en los datos genómicos que estén relacionados con la capacidad de las bacterias para establecer asociaciones con plantas. Al demostrar la viabilidad de estos enfoques, no solo buscamos avanzar en la comprensión de los factores genéticos que determinan las interacciones planta-microorganismo, sino también establecer una metodología aplicable a otros géneros bacterianos. Además, la evaluación de la importancia de cada variable en la predicción complementará el análisis experimental realizado en esta tesis, permitiéndonos identificar genes y funciones clave vinculados a la colonización rizosférica y endofítica que resulten generales en la asociación de cepas de *Pantoea* en su asociación con especies vegetales. Esto permitirá mejorar la predicción de la asociación entre bacterias del género *Pantoea* y plantas, arrojando luz sobre los marcadores genéticos implicados en estos procesos.

VII.2. Construcción y curado de la base de datos utilizada para el entrenamiento de los algoritmos aprendizaje automático que permiten la clasificación categórica

En la actualidad las tecnologías de secuenciación masiva cada vez más potentes y accesibles permiten secuenciar miles de genomas a una velocidad mayor a aquella con la que se analizan, lo que resulta en la disponibilidad de una gran cantidad de información genética esperando ser estudiada. Decidimos entonces utilizar los genomas disponibles en la base de datos IMG (Chen et al., 2023) para el estudio propuesto, y se descargaron 417 genomas pertenecientes al género *Pantoea* junto con sus anotaciones funcionales COG (*Clusters of Orthologous Genes*), KO (*KEGG Orthology database*) y PFAM (*Protein Family database*) que proporcionan información sobre la función molecular y los procesos biológicos asociados a un gen, y los dominios proteicos. A este conjunto de genomas se le añadió el genoma de la cepa *Pantoea agglomerans* LPU12 caracterizada y utilizada en los experimentos Tn-Seq de esta tesis, conformando así la colección final de 418 cepas. Cada cepa fue cuidadosamente analizada para determinar su origen de aislamiento según la información disponible en bases de datos y publicaciones asociadas, y fue etiquetada según se haya reportado su asociación o no a plantas (Tabla Suplementaria VII.S1). Esta información recolectada, junto con las anotaciones funcionales obtenidas, se utilizaron para entrenar a los algoritmos de predicción.

Antes de explorar los algoritmos de clasificación, se construyó un árbol filogenético con los genomas seleccionados utilizando GTDB-Tk (Chaumeil et al., 2020b) para evaluar la similitud entre las cepas dentro del género *Pantoea*. Esta es una herramienta bioinformática diseñada para la clasificación y asignación taxonómica de genomas microbianos. La clasificación de un genoma se basa en una combinación de su ubicación en el árbol de referencia GTDB (gtdb.ecogenomic.org), su divergencia evolutiva relativa (RED, relative evolutionary divergence) (Parks et al., 2018) y su valor de identidad de nucleótidos ANI con respecto a los genomas de referencia. RED se utiliza para resolver los casos en los que las asignaciones de rango son ambiguas. Las asignaciones de especies se establecen utilizando el ANI, calculado con FastANI (Jain et al., 2018). Específicamente, un genoma ubicado dentro de un género se asigna a la especie del genoma de referencia más cercano con una fracción de alineación >65% y un ANI mayor a 95%, según lo definido por la GTDB. De lo contrario, el genoma de consulta se clasifica como una nueva especie dentro del género. La Figura VII.2 muestra el árbol resultante, donde cada cepa está etiquetada con su origen de aislamiento (plantas, animales, hongos u otros ambientes diversos - cepas ambientales-), su fenotipo patogénico o no patogénico, y el tamaño de su genoma. Este análisis filogenético evidencia la gran diversidad del género *Pantoea*, con cepas que muestran alta similitud genética, pero están asociadas a una amplia variedad de hospedadores de diferentes reinos. Las relaciones que estas cepas establecen con sus hospedadores pueden variar desde interacciones beneficiosas hasta patogénicas. Sin embargo, también se observa una tendencia interesante: los aislamientos asociados a plantas tienden a agruparse en clados comunes o cercanos, lo que sugiere la posible presencia de características genéticas compartidas entre las cepas que habitan en plantas. Este hallazgo apoya nuestra hipótesis de que los atributos genómicos pueden ser utilizados para entrenar algoritmos de *Machine Learning* capaces de predecir el tropismo hacia plantas en bacterias pertenecientes a este género. La proximidad filogenética de las cepas asociadas a plantas indica que podrían compartir factores genéticos clave relacionados con la colonización o asociación con este tipo de hospedadores. Así, este análisis filogenético fortalece la base sobre la que consideramos factible el uso

de herramientas de clasificación supervisada, que pueden identificar estos patrones genéticos y ayudar en la predicción del comportamiento ecológico de otras cepas del género

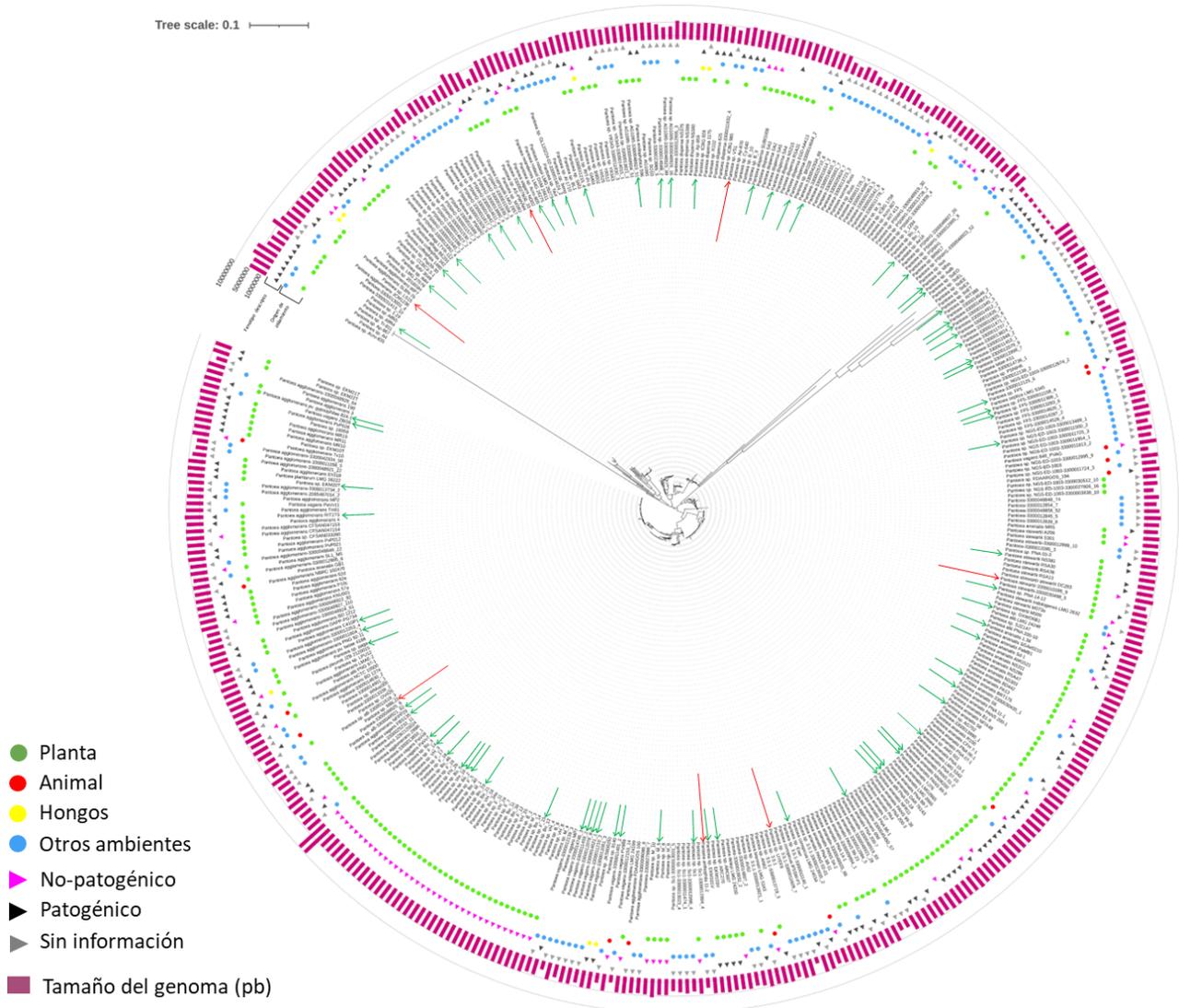


Figura VII.2. Árbol filogenético generado con la herramienta GTDB-Tk para el análisis de la distribución filogenética de 418 cepas del género *Pantoea* utilizadas en este estudio. Para cada cepa empleada en este estudio se muestra el origen de donde fueron aisladas, el fenotipo patogénico o no patogénico descrito para aquellas que han sido caracterizados en estudios reportados y el tamaño de los genomas en pares de bases. Las flechas indican los aislamientos incluidos en el conjunto de datos de prueba, donde el color de cada flecha refleja la precisión en la clasificación del tropismo por parte del modelo que mostró el mejor desempeño (COG-RF) (ver sección siguiente VII.3): verde para clasificaciones correctas y rojo para incorrectas. La metadata asociada a cada genoma se encuentra detallada en la Tabla Suplementaria VII.S1. Se observa que las bacterias asociadas a plantas se ubican en clados comunes o filogenéticamente cercanos.

Para explorar la eficacia de los algoritmos de clasificación en la predicción del tropismo vegetal de cepas de *Pantoea*, se definieron como variables (denominadas *features* en el lenguaje del algoritmo) las anotaciones funcionales derivadas de las bases de datos KO (KEGG Orthology), COG (Clusters of

Orthologous Groups) y PFAM (Protein Families). KEGG Orthology es parte del sistema de anotación de la base de datos KEGG y está diseñado para agrupar genes de diferentes organismos que realizan funciones similares en rutas metabólicas y procesos biológicos (Kanehisa et al., 2016). Permite así asociar rutas específicas en KEGG que representan redes bioquímicas de metabolismo y señalización con los genes asignados a un KO. *Clusters of Orthologous Groups* es una base de datos que agrupa proteínas en grupos de ortólogos, por lo que está especialmente enfocada en la evolución y la conservación funcional de proteínas entre organismos procariotas y algunos eucariotas. COG agrupa genes en base a la homología evolutiva, identificando proteínas ortólogas en diferentes organismos y asigna cada grupo de proteínas ortólogas a una categoría funcional específica (Galperin et al., 2021). Por otro lado, Pfam es una base de datos que clasifica las proteínas en familias basadas en sus dominios (estructuras funcionales y evolutivamente conservadas), y utiliza la información de los dominios para realizar la anotación funcional (Mistry et al., 2021). Optamos por utilizar estas anotaciones para entrenar los modelos de clasificación debido a que representan información funcional accesible y ampliamente disponible, sin necesidad de procesamiento adicional. Si estas anotaciones resultan suficientes para sustentar la predicción del tropismo hacia plantas, su accesibilidad las convierte en herramientas aún más valiosas y prácticas para estudios futuros.

El entrenamiento de los modelos se realizó utilizando matrices binarias basadas en la presencia o ausencia de estas anotaciones funcionales. Para cada cepa, se asignó un valor de 1 (presencia) o 0 (ausencia) a cada COG, KO o dominio PFAM según estuviera presente o no en su genoma. Estas matrices de variables se complementaron con un vector de etiquetas que indicaba para cada cepa si estaba asociada a plantas (valor 1) o si había sido reportada en asociación con otros hospedadores o ambientes (valor 0). Este conjunto de datos binarios permitió el entrenamiento supervisado de los algoritmos de clasificación seleccionados, como *Random Forest* y *Logistic Regression*, en la tarea de predecir la capacidad de una cepa para establecer asociaciones con plantas. Como hemos discutido más temprano en la introducción de este capítulo, la elección de estos métodos de clasificación responde a su capacidad comprobada para manejar grandes cantidades de datos binarios y detectar patrones complejos que pueden estar presentes en las características genómicas de las cepas.

VII.3. Modelos de clasificación generados para la predicción de tropismo a plantas en el género *Pantoea*, basados en la información funcional de los genomas

Elegimos dos enfoques de aprendizaje supervisado ampliamente utilizados para abordar la tarea de predecir si una cepa de *Pantoea* puede habitar plantas basándonos en sus características genómicas: ***Logistic Regression (LR)*** y ***Random Forest (RF)***. Como se describió en la sección I.6 de la introducción de esta tesis, ambos algoritmos presentan ventajas complementarias. *Logistic Regression* es un método estándar para la clasificación binaria, útil cuando los datos no son demasiado voluminosos. Este algoritmo no solo se centra en predecir un resultado, sino también en explorar la relación entre las variables individuales (*features*) utilizadas para entrenar el modelo y el resultado predicho. En nuestro caso, LR nos ayudará a evaluar si la relación entre las características funcionales de los genomas con el tropismo vegetal puede ser interpretada o aproximada como lineal, sugiriendo que la presencia o ausencia de determinadas funciones o procesos biológicos determinan la asociación de la bacteria a los nichos vegetales. *Random Forest*, por otro lado, es una técnica más potente en cuanto a la capacidad de predicción, especialmente cuando el conjunto de datos es grande y las relaciones entre las características

y el resultado pueden ser más complejas y no necesariamente lineales. Este método, basado en la creación de múltiples árboles de decisión, permite manejar la gran cantidad de características presentes en las anotaciones KO, COG y PFAM de nuestros datos sin requerir su reducción. Dado que desconocemos a priori si la relación entre las características genómicas y el resultado (capacidad de colonización de plantas) es lineal o no, decidimos explorar ambos enfoques. El uso de *Random Forest* se justifica por su capacidad para modelar relaciones complejas, mientras que *Logistic Regression* nos permite examinar cómo influyen individualmente las características genómicas en el resultado, lo cual es fundamental para entender los mecanismos biológicos subyacentes. Para los análisis y la implementación de ambos algoritmos, utilizamos *Scikit-learn*, una biblioteca de código abierto ampliamente reconocida por su eficiencia y versatilidad en el ámbito del aprendizaje automático (*Machine Learning*) en el ecosistema de Python. *Scikit-learn* es considerada una herramienta de referencia en la comunidad científica y proporciona las funcionalidades necesarias para realizar tanto predicciones precisas como interpretaciones detalladas de los modelos.

Para la utilización de *Random Forest*, primero establecimos los mejores hiperparámetros que caracterizan a cada modelo. Entre ellos se encuentra el número de árboles que lo compone (*n_estimator*), la medida de impureza (*criterion*), el número de variables que conforman el subconjunto aleatorio en cada paso de división de un nodo (*max_features*), y el parámetro semilla que permite la reproducibilidad de los resultados en los modelos controlando la aleatoriedad introducida por el *bootstrapping* y *max_features (random_state)*, cuyo valor en sí mismo no tiene importancia más allá de ser consistente en todas las ejecuciones para obtener resultados comparables. *Random Forest* puede ser utilizado sin ajustar los hiperparámetros al modelo que se pretende generar, pero una selección adecuada de los mismos con frecuencia optimiza su rendimiento. El principal parámetro a determinar es el número de árboles de decisión que componen el modelo, por lo que para determinar esto evaluamos el rendimiento de los modelos al aumentar el número de árboles que lo componen, manteniendo los restantes hiperparámetros según los valores estándar de *Random Forest* (detallados en la sección II.6.c de Materiales y Métodos). Para evaluar la performance de los modelos (tanto en RF como en LR más adelante) utilizamos el valor f1, medida que combina la precisión y exactitud del modelo en una sola métrica, dada por la ecuación $f1 = \frac{2 \times \text{precisión} \times \text{exactitud}}{\text{precisión} + \text{exactitud}}$. Evaluamos entonces la variación del valor f1 al aumentar el número de árboles, y definir así cuantos conformarán el modelo final (Figura Suplementaria VII.S1). El valor de f1 no aumenta significativamente al aumentar el número de árboles por encima de 200, pero se observó un máximo en 600 árboles, por lo que decidimos utilizar este número de estimadores para los modelos de *Random Forest*.

Para entrenar los diferentes modelos utilizando los sets de datos generados según las tres anotaciones funcionales, y con cada uno de los algoritmos elegidos, se dividió cada conjunto de datos en un set de entrenamiento (compuesto por el 80% de los datos) y un set de testeo (compuesto por el 20% de los datos). Así, el modelo aprende de la información provista y luego su capacidad de predicción es puesta a prueba con el set de testeo, compuesto con un conjunto de datos que no vio anteriormente. Se entrenaron finalmente seis modelos basados en diferentes características y algoritmos de clasificación: **KO_RF**, **KO_LR**, **COG_RF**, **COG_LR**, **PFAM_RF**, y **PFAM_LR**. Estos modelos, que combinan las características funcionales extraídas de las bases de datos **KO**, **COG** y **PFAM** con los algoritmos *Random Forest* (RF) y *Logistic Regression* (LR), demostraron una notable capacidad para predecir el tropismo vegetal en cepas del género *Pantoea*. Todos los modelos alcanzaron altos valores de f1, cercanos a 0.9, lo que indica una fuerte capacidad predictiva del tropismo a la planta. Esto se puede observar en la Figura VII.3.A, donde los resultados de predicción son consistentes entre las diferentes combinaciones de características y algoritmos.

Para un mejor análisis del rendimiento de los modelos, utilizamos las curvas ROC (*Receiver Operating Characteristic*) junto con el valor AUC (*Area Under the Curve*) para evaluar la relación entre sensibilidad (capacidad de detectar verdaderos positivos) y especificidad (capacidad de minimizar falsos positivos). La curva ROC es un gráfico que muestra la relación entre la tasa de verdaderos positivos (TPR) o sensibilidad en el eje Y, que indica la proporción de casos positivos correctamente identificados por el modelo, y la tasa de falsos positivos (FPR) en el eje X, que es la proporción de casos negativos que fueron clasificados incorrectamente como positivos. La curva ROC se construye al variar el umbral de decisión que el modelo usa para clasificar las predicciones como positivas o negativas, lo que permite evaluar el rendimiento del modelo para distintos niveles de sensibilidad y especificidad. A medida que varía el umbral, se recalculan las tasas de verdaderos positivos (TPR) y falsos positivos (FPR). El valor AUC es el área bajo la curva ROC, y proporciona una medida agregada del rendimiento del clasificador en todos los umbrales posibles, que varía entre 0 y 1, siendo 1 un clasificador perfecto y 0.5 un clasificador que realiza predicciones aleatorias. Los resultados del AUC-ROC, que se observan en la Figura VII.3.b, muestran que los modelos generados tienen un excelente equilibrio entre sensibilidad y especificidad, con un bajo número de falsos positivos. Incluso con un umbral de probabilidad bajo (por ejemplo, 0.6), los modelos son capaces de detectar un alto número de verdaderos positivos, lo que sugiere que la predicción del tropismo a planta es robusta. Para todos los modelos, los valores de AUC se aproximaron a 1, lo cual es un indicador clave de un rendimiento de clasificación fuerte y confiable. Esto significa que los modelos no solo son precisos, sino que también logran distinguir de manera efectiva entre cepas asociadas y no asociadas a plantas, con mínimos errores en la clasificación. En conjunto, los resultados confirman que las características funcionales de los genomas son suficientes para entrenar algoritmos de predicción capaces de identificar la capacidad de una bacteria de asociarse con plantas.

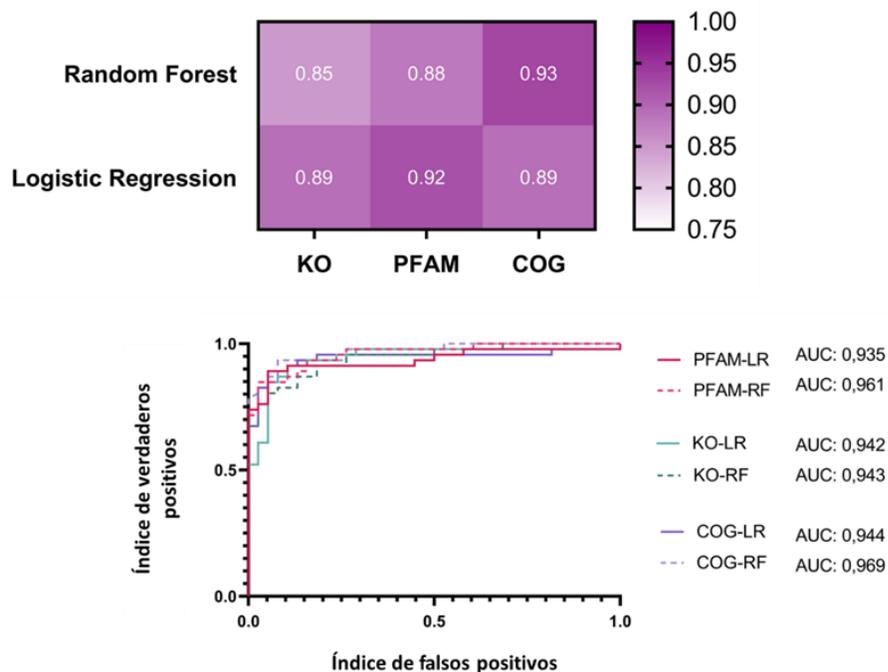


Figura VII.3. Evaluación del desempeño de los modelos de *Machine Learning* generados para la clasificación del tropismo a planta en aislamientos del género *Pantoea*. A) valores f1 para cada uno de los modelos entrenados y evaluados para la predicción del tropismo a plantas; mejores modelos presentan valores f1 cercanos a 1. B) Curvas ROC y valores AUC para cada modelo evaluado que indican el índice de verdaderos positivos respecto del índice de

falsos positivos que detecta cada modelo. Valores AUC cercanos a 1 indican un rendimiento de clasificación fuerte y confiable.

En la Figura VII.2 descrita anteriormente, se muestra el árbol filogenético construido con las cepas de *Pantoea* empleadas para el entrenamiento y testeo de los modelos de clasificación. Se indica con una flecha los aislamientos incluidos en el conjunto de datos de prueba, donde el color de la flecha refleja la precisión del modelo en la clasificación del tropismo de cada aislamiento: verde para clasificaciones correctas y rojo para incorrectas, de acuerdo con el modelo de mejor desempeño (COG-RF). Observamos que los errores del modelo ocurren en regiones del árbol donde el fenotipo de asociación, determinado según el origen del aislamiento (ver Tabla Suplementaria VII.S1), es diverso. Esto sugiere que, aunque existen agrupaciones claras de fenotipos similares en clados comunes, ciertas áreas presentan una complejidad que el análisis filogenético por sí solo no puede resolver para clasificar el tropismo a plantas en una cepa. Esto enfatiza el valor de contar con modelos predictivos robustos para inferir asociaciones a plantas en bacterias relacionadas. Además, el modelo COG-RF demuestra precisión en zonas cercanas en el árbol que exhiben fenotipos diversos, subrayando su capacidad predictiva y la utilidad de generar herramientas clasificatorias capaces de predecir asociaciones a plantas en aislamientos de un género.

Tanto *Random Forest* como *Logistic Regression* han demostrado ser algoritmos de clasificación eficientes con un alto poder predictivo. Uno de los aspectos más interesantes de estos algoritmos es su capacidad para proporcionar información sobre la importancia de las variables utilizadas para el entrenamiento ("*feature importance*"), lo que en nuestro estudio nos permite identificar qué genes o funciones biológicas están más fuertemente asociados con la predicción del tropismo vegetal en *Pantoea*. Un desafío que surge al analizar esta información es que las características genómicas utilizadas en el entrenamiento de los algoritmos pueden no ser independientes entre sí. Dado que en un genoma las regiones codificantes pueden encontrarse duplicadas o en operones, muchas de ellas podrían estar relacionadas, lo que puede resultar en información redundante, influyendo en el rendimiento del modelo que no refleja con precisión la relevancia de variables específicas. Para abordar este problema, decidimos reducir el conjunto de características eliminando aquellas que muestran un alto grado de correlación entre sí. Este proceso de reducción en la dimensión del conjunto de datos tiene como objetivo descartar variables que están estrechamente relacionadas, reduciendo la posible duplicación de información y evaluando la independencia de las características realmente importantes para la predicción. Buscando conservar la alta capacidad de predicción con un conjunto más conciso de características, podremos analizar con mayor claridad qué genes o funciones biológicas específicas son más importantes para la asociación de *Pantoea* con plantas. Esto proporcionará una visión más detallada y confiable de los determinantes genéticos de esta relación.

En la sección siguiente realizaremos un análisis de correlación entre las variables para identificar aquellas con alta redundancia, y las eliminaremos de los datos de entrenamiento. Luego, entrenaremos nuevamente los modelos y evaluaremos su rendimiento para comprobar si la capacidad predictiva se mantiene sólida con este conjunto reducido de variables. De esta manera, podremos obtener conclusiones más precisas sobre los atributos genéticos clave que impulsan la asociación planta-bacteria en *Pantoea*.

VII.3.a Reducción de variables: búsqueda de marcadores con alto grado de relación entre ellos y reducción de las matrices binarias utilizadas para entrenar los modelos

Al abordar el problema de la redundancia en las variables durante el entrenamiento de los modelos, la eliminación de aquellas altamente correlacionadas es clave para evitar que información repetitiva influya de manera desproporcionada en el rendimiento del modelo. Cuando dos o más variables

están altamente correlacionadas, tienden a proporcionar información similar, por lo que incluir ambas no añade valor al modelo, sino que lo hace más complejo y difícil de interpretar. Al eliminar variables redundantes, se simplifica el modelo sin perder información relevante, mejorando su eficiencia. Por otro lado, en modelos como *Logistic Regression* una alta correlación entre variables puede provocar que los coeficientes que describen el modelo sean inestables y difíciles de interpretar, aumentando la varianza entre ellos, lo que hace que las estimaciones del modelo sean más sensibles a pequeños cambios en los datos, y disminuya a su vez la capacidad del modelo para identificar qué variables son realmente importantes. En este contexto, el coeficiente de correlación de Pearson se convierte en una herramienta esencial para identificar aquellas variables que aportan información similar.

Para esto, primero calculamos una matriz de correlación entre todas las variables de los conjuntos de datos utilizados (COG, KO, PFAM) utilizando la librería para funciones estadísticas *scipy.stats* de Python. Esta matriz nos proporciona el coeficiente de correlación de Pearson junto con su valor p asociado para cada par de variables, lo que nos permite evaluar la existencia de una relación lineal entre ellas. El valor del coeficiente oscila entre -1 y 1, indicando la dirección y magnitud de la asociación. En este análisis decidimos descartar sólo aquellas variables que presenten un valor de correlación positiva que supere un umbral determinado y un valor $p < 0,05$. La reducción de los datos se realizó de manera progresiva, estableciendo umbrales más restrictivos para el coeficiente de Pearson, con el fin de observar el impacto en la precisión predictiva de los modelos. De este modo, analizamos si la eliminación de determinadas variables correlacionadas afecta significativamente el rendimiento de los modelos o si, por el contrario, la capacidad predictiva se mantiene con un conjunto de datos más reducido, lo que nos permitiría evaluar con mayor precisión la relevancia de atributos génicos específicos en la asociación de las cepas de *Pantoea* con plantas. La Figura VII.4 muestra cómo la precisión de los modelos de predicción varía en función de la cantidad de variables en el conjunto de datos de entrenamiento, al aplicar umbrales de correlación más estrictos. Observamos que, aunque hay una ligera disminución en la precisión de la predicción, los modelos siguen siendo robustos, manteniendo valores de precisión cercanos a 0.8 incluso cuando el coeficiente de correlación se restringe a 0.4. Este umbral reduce el número de variables a aproximadamente un 10% del tamaño original del conjunto de datos.

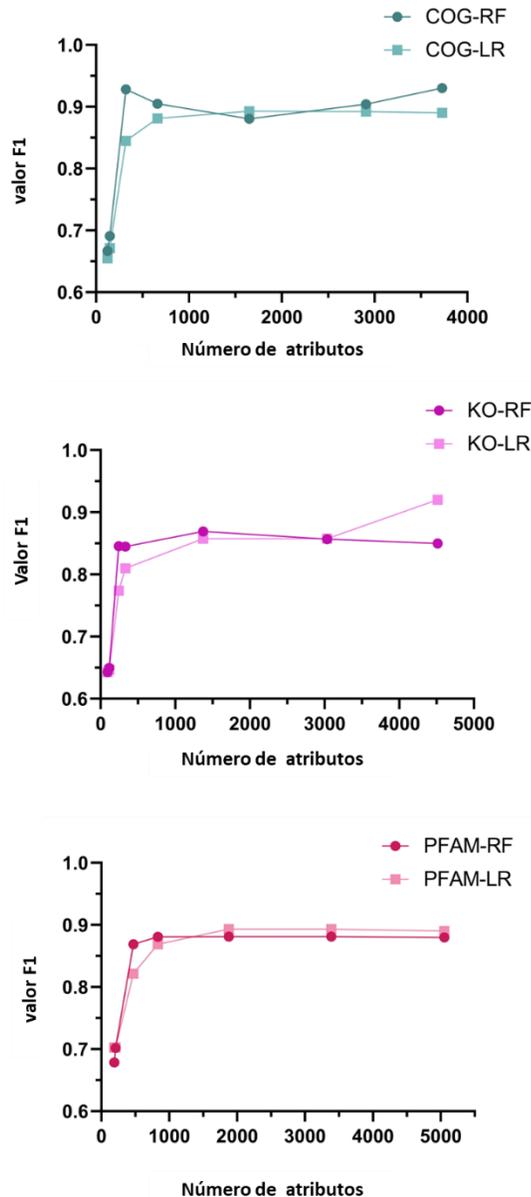


Figura VII.4. Variación del valor f1 de cada modelo en función de la dimensión del set de datos con el que fueron entrenados cuando se reduce la correlación entre las variables. Cada set de datos fue reducido eliminando las variables que presentaban un alto grado de correlación indicado por el valor del coeficiente de *Pearson*. Se generaron 6 conjuntos de datos para cada categoría funcional de variables utilizadas (COG, KO y PFAM), cuyo número de variables disminuye al eliminar aquellas variables que superaban un valor umbral de correlación asignado para cada conjunto reducido, y un valor $p < 0.05$. Los valores umbral definidos para cada conjunto de datos reducido fueron - 0.1, 0.2, 0.4, 0.5, 0.7, 0.9 y 1, donde 1 corresponde al máximo de correlación permitido entre variables, con el total de datos del conjunto de datos original.

Estos resultados sugieren la existencia de un conjunto reducido, pero altamente representativo de características funcionales en los genomas de cepas de *Pantoea* asociadas a plantas. Estas variables son suficientes para que los algoritmos de aprendizaje supervisado realicen predicciones precisas sobre el tropismo vegetal en aislamientos de este género. La identificación de estas variables de forma sencilla

y accesible, como se ha demostrado en este estudio, tiene un gran valor para investigaciones que, como la nuestra, adoptan un enfoque global, analizando múltiples cepas con diferentes relaciones con sus huéspedes.

VII.3.b Identificación de variables con importancia para la identificación de bacterias: búsqueda in silico de características funcionales en los genomas relacionados a la asociación a plantas en bacterias del género *Pantoea*

Una ventaja importante de los algoritmos de clasificación que evaluamos es la capacidad de determinar las variables utilizadas en las predicciones más relevantes en el proceso de toma de decisiones. *Random Forest* y *Logistic Regression* utilizan enfoques distintos para calcular la importancia de los variables en un modelo, dado que están basados en diferentes principios de aprendizaje. *Random Forest* realiza este cálculo basado en cómo cada atributo afecta la toma de decisiones dentro de los árboles. En cada nodo de un árbol se selecciona una característica para dividir los datos con base en alguna métrica de impureza (*Gini* o *Entropía*, descritas en la sección VII.1.a). Así, la importancia de un atributo se basa en cuánta reducción de impureza provoca una característica al ser usada en los nodos de decisión. Por otro lado, en *Logistic Regression* cada característica utilizada en el modelo tiene un coeficiente asociado que mide su influencia en la predicción, que es estimado a través de la maximización de la verosimilitud (probabilidad de que un conjunto de parámetros del modelo explique o "produzca" los datos observados). Si el coeficiente es positivo, la característica aumenta la probabilidad de que ocurra el evento positivo, por lo contrario, si el coeficiente es negativo, la característica disminuye la probabilidad del evento positivo.

Se determinaron los valores de importancia de las variables utilizando la función *feature_importances* de la biblioteca *Scikit-learn*, basada en la reducción de impurezas medida según la entropía (acorde a los parámetros con los que se entrenó el modelo, detallados en la sección II.6.c de Materiales y Métodos) para *Random Forest*, y la función *coef_* para los modelos generados con *Logistic Regression*. Este análisis se realizó en los seis modelos que emplearon datos COG, KO y PFAM: COG-RF, KO-RF, PFAM-RF, COG-LR, KO-LR, y PFAM-LR. Además, se analizó de manera similar la relevancia de los atributos cuando se redujeron las variables con altos coeficientes de correlación, tal como se describió anteriormente, con el objetivo de acotar la búsqueda de marcadores relevantes en la clasificación.

Dado que el algoritmo de regresión lineal permite asociar cada variable de entrada a un valor de salida (donde los valores de importancia positivos se asocian con la categoría de "Planta" y los valores negativos con "No-Planta"), nos enfocamos en los resultados de los valores de importancia de las variables asignados en los modelos generados con este algoritmo. Luego de analizar los resultados para todos los modelos generados, decidimos profundizar el análisis en los modelos de COG, KO y PFAM entrenados con LR cuando se reduce el número de variables cerca del 90% (admitiendo un grado de correlación baja entre variables, con un valor de coeficiente de Pearson menor a 0.4), donde aún se sostiene un gran poder predictivo (valor f_1 cercano a 0.9, Figura VII.4). La reducción de variables permite crear modelos más simples, rápidos y precisos que se generalizan mejor a nuevos datos, ahorran recursos y son más fáciles de interpretar y mantener. Con los valores de importancia obtenidos para cada uno de estos modelos (COG-LR_0.4, KO-LR_0.4, PFAM-LR_0.4), evaluamos la distribución de estos valores en un histograma donde se observa que la mayoría de los valores se concentran alrededor de la media, con algunas características que presentan valores de importancia más significativos (Figura Suplementaria VII.S2). Para identificar las variables relevantes, buscamos aquellas con los valores más altos y positivos, que corresponden a las variables con mayor impacto en la clasificación de bacterias asociadas a plantas en las

predicciones de los modelos, ubicadas en la distribución en valores mayores al umbral definido (media más la desviación estándar). En la Figura VII.5 se muestran las variables más significativas para cada modelo analizado. Al analizar estos mismos resultados para los restantes modelos generados (conjuntos de datos mayores y utilizando el algoritmo *Random Forest*), se observó una clara consistencia en las variables con mayor importancia para la capacidad de predicción (aquellas conservadas en los modelos entrenados con *Random forest* se indican subrayadas en la figura, y las variables que se conservan al utilizar conjuntos de datos de entrenamiento con un mayor número de variables se señalan en fucsia). Este análisis permitió identificar un conjunto de categorías funcionales presentes en los genomas de *Pantoea spp.* que permiten su correcta clasificación como cepas asociadas a plantas. Entre las funciones más relevantes para la clasificación las más representadas se encontraron relacionadas con la biogénesis de membranas celular, sistemas de defensa, genes asociados al mobiloma y funciones desconocidas. Sin embargo, al analizar los genes involucrados en la asociación de aislamientos de *Pantoea* con especies vegetales reportados en estudios experimentales, como el realizado en esta tesis que estudia la colonización de plantas de alfalfa por parte de *Pantoea agglomerans* LPU12 y el reportado por Duong et al. (2018) en el estudio del crecimiento de la cepa *Pantoea stewartii* DC283 en plantas de maíz, no se encontraron genes afectados en tales experimentos cuya anotación funcional coincidiera con aquellas esenciales para la predicción del tropismo a planta por los modelos generados en este trabajo. Esto sugiere que las características genómicas detectadas por los algoritmos para clasificar la asociación de cepas de *Pantoea* con plantas no están necesariamente relacionadas con interacciones específicas huésped-hospedador, como las evaluadas experimentalmente en esta tesis y en otros estudios. Es posible que estas funciones estén vinculadas a diversificaciones evolutivas en los genomas de cepas asociadas a plantas que comparten características genómicas heredadas de ancestros comunes, lo que se conoce como “señal filogenética” (Blomberg et al., 2003). La señal filogenética asociada a un fenotipo es débil cuando el mismo se encuentra influenciado mayormente por adaptaciones al medio ambiente, de lo contrario una fuerte señal ocurre cuando los rasgos analizados están determinados por herencia evolutiva. Analizar en qué grado la presencia de las características genéticas que le permiten a cepas de *Pantoea* habitar plantas se debe a la historia evolutiva que comparten permitiría entender cómo interactúan la herencia común y las presiones ambientales en la evolución de las especies.

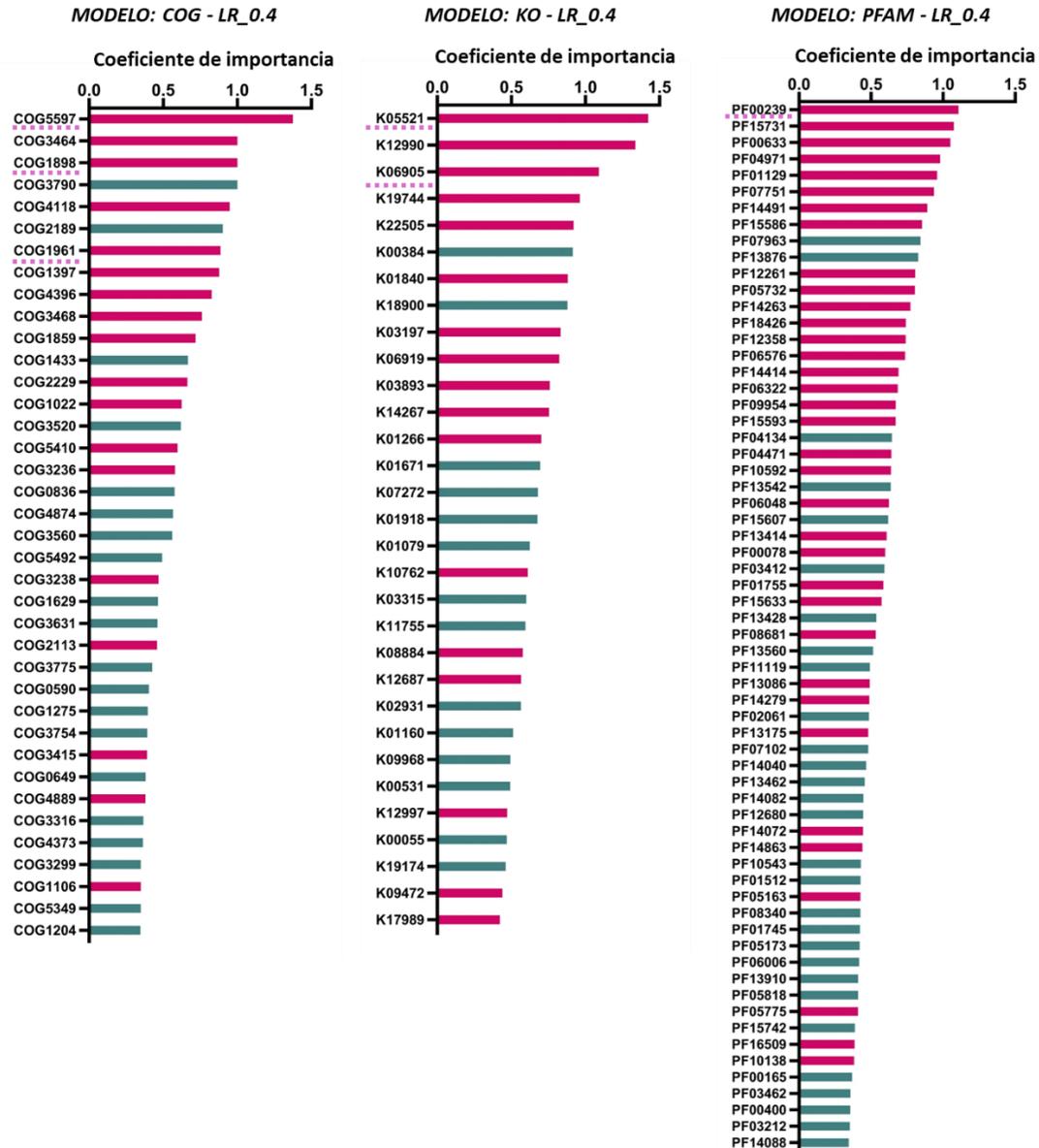


Figura V.II.5. Coeficientes de importancia asignados a las variables más relevantes para la predicción del tropismo a plantas por modelos generados con el algoritmo *Logistic Regression*. Los coeficientes corresponden a los asignados por los modelos entrenados con un set de datos reducido donde se eliminaron las variables que presentaban un coeficiente de correlación de Pearson mayor a 0.4 con un valor p asociado menor a 0.05 (modelos: COG-LR_0.4; KO-LR_0.4; PFAM-LR_0.4). En color fucsia se indican las variables que resultaron relevantes también cuando se utiliza el set de datos completo (permitiendo mayores valores de correlación hasta alcanzar el total de datos original, coeficiente de Pearson permitido=1), y se subrayaron aquellas variables de mayor importancia para la predicción por parte de los modelos entrenados con *Random Forest*.

VII.4. Consideraciones finales del capítulo

El enfoque computacional presentado en esta capítulo, basado en métodos de clasificación y regresión de *Machine Learning*, demostró ser altamente efectivo para generar modelos predictivos sobre la capacidad de *Pantoea* spp. para colonizar plantas. En la mayoría de los casos, los modelos alcanzaron

precisiones superiores al 90%, lo que sugiere la existencia de una estructura genómica característica de las bacterias capaces de asociarse con plantas.

A partir de estos modelos, fue posible identificar características funcionales específicas (como COG, PFAM y KO) cuya presencia se correlaciona positivamente con la capacidad de *Pantoea* spp. para colonizar especies vegetales. Sin embargo, no se logró establecer una relación clara entre estas características funcionales identificadas por los modelos y los genes reportados en estudios experimentales que estén directamente involucrados en asociaciones de cepas de *Pantoea* con plantas (reportados en esta tesis en el Capítulo VI, y en Duong et al. (2018)). Esto puede deberse a que las cualidades genómicas detectadas por los algoritmos de clasificación podrían estar relacionadas con una herencia colateral. Es decir, los genes que actúan directamente en la colonización de las plantas podrían haber sido heredados por bacterias de un linaje simbiote o patógeno vegetal, o adquiridos a través de la transferencia horizontal de genes. De esta manera, las características genómicas comunes en aislamientos de este género relevantes para la predicción de la asociación con plantas podrían no reflejar funciones requeridas para establecer interacciones específicas huésped-hospedador, sino ser resultado de la herencia evolutiva de las mismas.

Si bien la interpretación detallada de los atributos génicos utilizados por los modelos de clasificación presenta una complejidad significativa, lo que requiere un análisis más profundo, los resultados de este capítulo muestran que es posible desarrollar modelos clasificatorios robustos y altamente predictivos del tropismo hacia plantas en bacterias del género *Pantoea* utilizando un conjunto limitado de características genómicas (aproximadamente el 10% del genoma bacteriano). Esto proporciona una herramienta poderosa y eficiente que puede emplearse en futuros análisis, facilitando así la identificación de cepas con capacidades específicas de asociación a plantas.



CAPÍTULO VIII

DISCUSIÓN GENERAL Y CONCLUSIONES



Discusión general y conclusiones.

El suelo es un ambiente complejo donde, pese a la baja disponibilidad de recursos, hay una gran diversidad de microorganismos adaptados como comunidad a la vida en el propio suelo y en asociación con las plantas que lo habitan. La relación entre microorganismos y plantas terrestres es un vínculo que se ha desarrollado a lo largo de los últimos 450 millones de años (Beerling, 2007; Delaux et al., 2015), a través del cual microorganismos y plantas expresan capacidades funcionales diversas, y necesidades y dependencias mutuas. En un marco de gran diversidad y abundancia microbiana, dado el carácter oligotrófico del suelo, las plantas representan un nicho nutricionalmente rico por cuya ocupación compiten los microorganismos circundantes.

En este trabajo de tesis hemos abordado la caracterización de la colonización de alfalfa por parte de bacterias del género *Pantoea*, ubicuas en distintas especies de plantas (Truyens et al., 2014; Walterson & Stavrínides, 2015). Usamos en particular la cepa *P. agglomerans* LPU12 recuperada de semillas de *Medicago sativa*. Con esta cepa caracterizamos la dinámica del proceso de asociación rizosférica y endofítica (Capítulo III), diseñamos y construimos herramientas de TnSeq (Capítulo IV) con las que identificamos determinantes genéticos bacterianos relevantes para la asociación de las bacterias con raíces (Capítulo V & VI), y finalmente desarrollamos aproximaciones informáticas de aprendizaje automático supervisado que, con baja tasa de fallo, nos permitieron predecir el tropismo para colonizar plantas (o la ausencia del mismo) de diferentes especies y aislamientos del género *Pantoea* (Capítulo VII).

En el capítulo III de esta tesis presentamos la caracterización general del sistema *P. agglomerans* LPU12 – *M. sativa* como modelo de trabajo. La presencia de *Pantoea* en muy diversas especies de plantas (tanto en tejidos vegetativos como incluso en semillas), sus altos números por unidad de biomasa vegetal y la existencia de especies tanto benéficas como patogénicas para plantas y animales; hacen de *Pantoea* un género de particular interés para investigar mecanismos de invasión en estudios de fisiología molecular comparada (interacciones benéficas/patogénicas, en planta/animales). La cepa utilizada como modelo en este trabajo, inicialmente designada *Pantoea* sp. LPU12 (López et al., 2018), fue luego tipificada como perteneciente a la especie *P. agglomerans* por secuenciación genómica completa y análisis taxonómico de la misma (Capítulo III). Dicha cepa fue elegida como modelo de análisis de la colección de bacterias endófitas aisladas de semillas y de plantas de alfalfa disponible en nuestro laboratorio (López et al., 2018). La cepa *P. agglomerans* LPU12 nunca evidenció signos de patogenicidad sobre plantas de alfalfa. Cabe destacar, sin embargo, que miembros de esta especie han sido reportados en asociaciones tanto benéficas como patogénicas con plantas (Dutkiewicz et al., 2016), e incluso en reportes referidos a aislamientos patógenos oportunistas de humanos (Guevarra et al., 2021). La elección del aislamiento *P. agglomerans* LPU12 fue realizada teniendo en consideración el potencial uso de los resultados obtenidos en estudios genómicos extendidos a otras *Pantoeas*. Es importante caracterizar los determinantes genéticos propios de los distintos tipos de interacciones, patogénicas o no-patogénicas, así como aquellos que puedan ser comunes a ambas. Con esa información podrán pensarse nuevas estrategias prácticas orientadas ya sea a promover el desarrollo de interacciones positivas, como a contrarrestar el avance de asociaciones patogénicas.

Según mostramos en el Capítulo III, inicialmente caracterizamos la cinética de colonización rizosférica de *P. agglomerans* LPU12 empleando inóculos bajos (del orden de 10^5 u.f.c./ml), de modo que se preserven eventuales respuestas específicas de *quorum* generadas por moléculas de las mismas bacterias. Por otra parte, en experimentos independientes empleando inóculos elevados (del orden de 10^7 u.f.c./ml), que no limiten el acceso a la rizósfera, caracterizamos la cinética de la colonización endofítica que en esas condiciones mostró ser muy eficiente: en sólo 24 h se alcanzan cerca de 10^7 u.f.c./g de raíz que se mantienen al menos hasta cerca de las 3 semanas post-inoculación. La presencia de la cepa

P. agglomerans LPU12 en el interior de la planta (o bajo estructuras protectoras de la raíz) fue corroborada en ensayos de colonización endosférica empleando una cepa marcada, y procesando plantas esterilizadas en superficie con peróxido de hidrógeno según se indicó en el Capítulo V. Para conocer mejor la dinámica de la colonización, antes de disponer de los resultados TnSeq realizamos un ensayo en el que inoculamos sobre alfalfa dos bacterias isogénicas con marca fluorescente diferente (Capítulo III) y en distintas proporciones. A través del mismo pudimos demostrar que la cantidad de cada bacteria fluorescente dentro de la planta, a las 24 h postinoculación, ocurre en relación lineal con la abundancia de cada una de ellas en el inóculo. Ese análisis junto al experimento TnSeq permitió estimar en cerca de $2-3 \times 10^3$ u.f.c. el número de colonizadores independiente/planta, los denominados “fundadores” (ver más adelante en esta discusión general la referencia a cambios en la densidad D de sitios TA mutados en el inóculo y en la muestra endosférica, y el modelado realizado para la estimación de fundadores). Son las bacterias fundadoras las que luego por multiplicación colonizarán todo el nicho radicular endofítico. En el Capítulo VI describimos en detalle las características del cuello de botella asociado al ingreso de las bacterias rizosféricas al compartimento endofítico. Hicimos referencia también a la importancia de analizar la posible relevancia de ese cuello de botella como posible mecanismo de selección de genotipos a lo largo de los sucesivos ciclos de reducción de diversidad y expansión clonal que se dan al ingresar a la endósfera. Todos los ensayos fueron realizados utilizando vermiculita estéril como soporte, de modo de emular efectos estructurales a modo de suelo. Dicha condición ya había probado ser útil en estudios previos de colonización rizosférica realizados en nuestro laboratorio (Salas Maria Eugenia, 2015). Los ensayos que hemos realizado con rizobios (Salas et al., 2017) y los presentados en el Capítulo III de esta tesis con *Pantoea*, sientan las bases metodológicas para explorar capacidades de colonización en distintas especies bacterianas asociadas a alfalfa disponibles en nuestro laboratorio (hemos criopreservado representantes de cerca de 40 géneros bacterianos recuperados de semillas de alfalfa) (López et al., 2018). Será importante expandir los estudios microbiológicos de colonización a nuevos miembros de la comunidad de bacterias asociadas a alfalfa, de modo de explorar relaciones entre las cinéticas de crecimiento *in vitro* (de las que ya disponemos) y las de colonización de la rizósfera y la endósfera. Análisis como el precedente, acompañados con estudios de inoculación consorcial y secuenciación de rADN 16S (en curso, no se presentan en esta tesis), permitirán obtener bajo condiciones controladas (manejando la heterogeneidad del inóculo y las cantidades relativas de sus miembros) una imagen acabada de cuáles son los géneros y especies que pueblan más eficientemente la rizósfera y/o el compartimento endofítico de la planta.

Con el propósito de caracterizar los determinantes genéticos asociados a la colonización del nicho rizosférico y endofítico de alfalfa por parte de *P. agglomerans* LPU12, en el Capítulo IV de esta tesis presentamos el diseño y puesta a punto de un formato experimental TnSeq que permitió la detección de mutantes con fenotipos de colonización alterados (deficientes o exacerbados). Dicho diseño incluyó la construcción de un vector adecuado para poder realizar transposiciones con un Tn *mariner* en *Pantoea*. Para ello construimos una variante del vector pSAM_RI (Perry & Yost, 2014) en el que reemplazamos el promotor *rpoD* de *R. leguminosarum* en la región promotora de la transposasa por el promotor y operador *plac-lacO^c* de *E. coli* que resultó funcional en *P. agglomerans* LPU12. Con dicha herramienta construimos una biblioteca portadora de cerca de cincuenta mil inserciones independientes, que incluyó —en promedio— cerca de 10 inserciones independientes/gen. Las características de esta biblioteca comparan a las reportadas en la caracterización por TnSeq de una cepa de *P. stewartii* patógena de plantas (Duong et al., 2018). El sistema TnSeq que diseñamos (Capítulo IV) utilizó cebadores diseñados *in house* para el etiquetado de las diferentes condiciones y réplicas. Los resultados de la secuenciación mostraron una muy alta proporción de productos de amplificación con la estructura esperada, y por tanto procesables para su mapeo y análisis posterior. Por otra parte, el control que realizamos para evaluar eventuales

desviaciones en la abundancia de mutantes luego de un ciclo de crecimiento en medio rico LB, prácticamente no mostro cambios en las proporciones iniciales y finales de los distintos mutantes (sólo fueron identificados mutantes en 16 genes que mostraban cambios significativos, ver Tabla VI.S4 suplementaria). Dicho control validó la amplificación que realizamos de las muestras rizosféricas y endofíticas por crecimiento de las mismas en medio LB, de modo de obtener una biomasa bacteriana y cantidad posterior de ADN que resultara procesable—digestión, unión de adaptadores y amplificación—para su posterior secuenciación. Como describimos en el Capítulo VI el protocolo TnSeq que utilizamos mostró capacidad de poner en evidencia diferencias significativas entre las proporciones inicial y final de mutantes de interés aun cuando la abundancia de los mismos en el inóculo fue extremadamente baja, en algunos casos del orden de 55 u.f.c./ml. Aproximaciones experimentales del tipo TIS (STM, TnSeq) se han utilizado en diferentes especies bacterianas para detectar genes esenciales y marcadores asociados a condiciones de interés *in vitro* e *in vivo*, mayoritariamente en bacterias patógenas de humanos (Cain et al., 2020). Los trabajos reportados que emplearon estas técnicas para el estudio de asociaciones de bacterias con plantas son más escasos, y se han aplicado principalmente al estudio del comportamiento de bacterias simbióticas como rizobios (Salas et al., 2017; Wheatley et al., 2020) y en la búsqueda de marcadores involucrados en la patogenicidad de cepas (Duong et al., 2018; Helmann et al., 2019; Royet et al., 2019).

Como hemos presentado en el Capítulo VI, si se prescinde de los genes de auxotrofia la colonización rizosférica resultó demandar para su normal desarrollo 46 genes, que representan cerca de 1% del genoma de *P. agglomerans* LPU12, de los cuales 16 (35%) corresponden a genes vinculados a la movilidad y la quimiotaxis de la bacteria. Luego de estos, le siguen en representación los genes relacionados a metabolismo y transporte de carbohidratos, y los vinculados a producción y conversión de energía; en analogía con lo previamente reportados en otros microorganismos modelo como rizobios (Salas et al., 2017; Wheatley et al., 2020). Fue además particularmente interesante la alta proporción de genes cuyos COGs son aun de función desconocida (10%), que podrán ser blanco prioritario de análisis en busca de caracterizar aspectos aún desconocidos de la ecología bacteriana en la rizósfera. Destacó asimismo que la gran mayoría de los genes anteriores (38, más de 80%) al ser mutados resultan en fenotipos con colonización rizosférica deficiente. Como ya indicamos, tal resultado sugiere fuertemente que, a lo largo de la evolución, el genoma ha sufrido (cuando fue posible) una selección purificadora negativa de aquellas funciones que de alguna manera perturban el vínculo con la rizósfera, circunstancia que remarca la importancia de este nicho para una bacteria ubicua del suelo como *P. agglomerans*. En apoyo de esta última consideración, los genes relevantes para el acceso rizosférico son en su totalidad de localización cromosomal, el replicón de contenido génico altamente conservado entre los distintos miembros de la especie. Una observación similar fue reportada por Salas et al. (2017) cuando analizó la colonización rizosférica de alfalfa por parte de *S. meliloti*. También puede destacarse que de los 8 genes cuya mutación se asocia a mejoras en la colonización, 3 son de localización plasmídica —en el plásmido pPagLPU12_2— y por tanto posiblemente de adquisición reciente y de existencia transitoria. La permanencia del plásmido pPagLPU12_2, deletéreo para el acceso a la rizósfera, posiblemente se deba a la relevancia del mismo en la expresión de otro/s fenotipo/s importante/s en ambientes para nosotros desconocidos. La densidad de genes cromosomales deletéreos para la colonización rizosférica mostró ser extremadamente baja, sólo 5 en 3.863 (comparada con 3 genes en 506 totales en el plásmido pPagLPU12_2). Nuevamente, es muy probable que esos 5 genes cromosomales, a pesar de su efecto negativo para la colonización de la planta, tengan efectos positivos en condiciones que son posiblemente de ocurrencia relativamente frecuente para *P. agglomerans*. Uno de estos genes está relacionado con la tolerancia a telurito, factor cuya importancia para la sobrevivencia de *P. agglomerans* en ambientes naturales será por tanto interesante indagar.

En el análisis que realizamos de la colonización endofítica, una etapa que espacial y temporalmente le sigue al abordaje rizosférico, empleamos una concentración de bacterias del orden de 10^7 u.f.c./ml en el inóculo, de modo que no se limite el acceso rizosférico (inóculo “saturante”). La idea subyacente fue que las diferencias que se observen en el compartimento endófito se limiten sólo a aquellos genes que son relevantes para el ingreso inicial a la raíz y/o para el crecimiento posterior de los mismos en la endósfera. Bajo esta configuración de ensayo, evaluamos inicialmente la diversidad de mutantes presentes en el inóculo y la presente en la muestra endofítica luego de la colonización (evaluado en ambas muestras a través de la densidad D de sitios TA mutados). Según mostramos en el Capítulo VI, mientras el inóculo portaba mutantes con inserciones en cerca de 30% del total de los sitios TA potencialmente mutables por el Tn *mariner*; en la muestra endofítica la densidad D de inserciones fue cercana a la mitad del valor anterior (14%). Tal disminución en la densidad de inserciones puso en evidencia, como ya indicamos, un fuerte cuello de botella que limitó el ingreso de la población rizosférica al interior de la planta. Es interesante destacar, además, que dicha disminución en la densidad de inserciones en la endosfera se sostuvo hasta 20 días después de la inoculación a pesar de que el inóculo inicial permaneció en la macetas, proveyendo evidencia en favor de que los fundadores que colonizan inicialmente la raíz posiblemente limitan en cada planta el acceso de nuevos genotipos desde la mayor diversidad rizosférica (cabe recordar que la diversidad rizosférica estimada a partir de la densidad D de sitios TA mutados fue comparable a la observada en el inóculo inicial. El nicho rizosférico resultó ser tan diverso como el inóculo que se utilizó). Así como las leguminosas infectadas restringen la infección por parte de nuevos rizobios cuando la planta ha sido suficientemente infectada, los fundadores endófitos de *P. agglomerans* parecen limitar la diversidad tardía a aquella que ellos mismos alcanzaron en las etapas tempranas de la infección. Desconocemos si esa limitación ocurre por simple ocupación del espacio y recursos disponibles en el nicho endosférico, o si existen señales específicas que restringen nuevas entradas. A la luz de estos resultados, resulta entonces crítica la estimación del número de fundadores, dado que los mismos parecen ser los que modelan la composición del microbioma endofítico en la planta adulta. Según describimos en el Capítulo VI, con el propósito de estimar el número de fundadores modelamos el proceso de ingreso asignando a cada mutante una probabilidad de acceso a la raíz que se correspondió con su abundancia en el inóculo inicial (tal supuesto se apoyó en los resultados del ensayo de coinoculación de cepas isogénicas con distinta fluorescencia que ya hemos comentado en los párrafos precedentes). La predicción numérica de ingreso de cada uno de los genotipos en base a su abundancia en el inóculo, de modo que se alcance una diversidad igual a la observada en la muestra experimental endofítica, permitió estimar el número de fundadores que ingresaron a la planta. Dicho número resultó cercano a las 3 mil u.f.c./planta, que luego de 24h resultan en los cerca de $10^5 - 10^6$ u.f.c./planta observados en la endosfera (Capítulo III). Puede inferirse entonces que en dicho período de tiempo es esperable que los fundadores realicen aproximadamente 5 ciclos de división. La cantidad de fundadores resultó un estimado numérico de mucho valor a la hora de entender el proceso de colonización de raíces. Los cuellos de botella son instancias en las que, en principio, se genera una pérdida de diversidad. Sin embargo, el análisis de su contribución a la evolución de una comunidad bacteriana debe incluir la consideración de diferentes aspectos. Si el ingreso al nicho endosférico fuera accesible a un número muy elevado de genotipos, el resultado conduciría a que a lo largo de los ciclos de infección y liberación de las bacterias de las raíces la población resultaría siempre enriquecida especialmente en variantes genéticas seleccionadas por su capacidad de desempeño en el nicho planta (cómo lo infectan, como se multiplican en él). Si por el contrario la planta genera un espacio de crecimiento sólo para un número reducido de genotipos fundadores, que no siempre serán los que mejor infectan ni los más numerosos (el suelo puede ser un medio espacial y microbiológicamente muy heterogéneo), esos genotipos podrán propagarse con poca competencia dentro de la endósfera. Con una estrategia como la

que se desprende de los resultados obtenidos, se preservan espacios de crecimiento para los distintos genotipos dentro del nicho planta, de modo de preservar una diversidad poblacional que no se ajuste sólo a la capacidad infectiva. Pensando en términos de los distintos miembros (especies) de la comunidad rizosférica, restringir el acceso múltiple es esperable que preserve la posibilidad de propagación *in planta* de aquellos componentes del microbioma que son de crecimiento más lento, dando lugar a que cada planta sea capaz de albergar y enriquecer una diversidad moderada, pero no idéntica entre planta y planta. Sostener la pluralidad de especies bacterianas presente en las plantas será por otra parte esencial para preservar la diversidad endófitica del interior de las semillas (López et al., 2018), vinculado en este caso a la heredabilidad de las mismas. Cómo el cuello de botella hacia la vida endofítica contribuye a delinear la diversidad del microbioma, es un aspecto que deberá ser estudiado y modelado. En nuestro laboratorio ya hemos caracterizado las tasas de crecimiento de una colección de decenas de géneros asociados a alfalfa, y podremos iniciar experimentos con comunidades sintéticas muy simples para evaluar cómo la presencia de las plantas modela el progreso y abundancia de cada una de las especies presentes. Hasta aquí sólo hemos mostrado evidencias respecto de la extensión (número) y diversidad a lo largo del tiempo con que una única bacteria endófitica ocupa (es admitida en) el nicho planta. Nuevos ensayos deberán ser hechos con inóculos múltiples, de complejidad creciente y ajustados a los taxones dominantes en el suelo, de modo de explorar cómo es la evolución de la diversidad composicional de esas comunidades sintéticas en presencia y ausencia de la planta. En este escenario, resulta fundamental considerar las heterogeneidades composicionales que pueden presentarse tanto en el suelo como en el nicho de la planta (en la endósfera), así como la propagación de genotipos específicos en microlocalizaciones (como una única endósfera).

Las consideraciones que aquí hemos realizado referidas al cuello de botella que se genera durante la etapa de ingreso de *P. agglomerans* a la endosfera radicular (fundadores del orden de unos pocos miles que emergen de un inóculo que posee decenas de miles de variantes) impuso una restricción importante a la identificación de genes relevantes para el acceso y establecimiento de la vida endofítica. La razón de tal limitación consistió en que con sólo 15 plantas y cerca de 3 mil fundadores/planta, accedimos en cada réplica a un total de 40-50 mil fundadores totales que en principio representan—en promedio—cerca de 10 mutantes por gen en la muestra de salida teniendo en cuenta el número de genes de *P. agglomerans*. Atendiendo a que no todos los genes están representados por un número igual de mutantes, y que por tanto muchos genes mutados tuvieron una representación muy baja dentro de las plantas, para identificar genes asociados a la colonización endofítica sólo analizamos aquellos fundadores cuya representación predicha estuvo entre el 50% de los mutantes de mayor incidencia (tal predicción se realizó, como indicamos previamente, haciendo uso de las abundancias relativas de cada mutante en el inóculo, y la presencia esperada de los mismos en la endosfera). Tal restricción limitó el análisis a aproximadamente 590 genes, algo más de 10% del genoma de *P. agglomerans* LPU12. Del conjunto anterior, sin considerar los efectos de auxotrofia, seleccionamos un total de 65 genes que correspondieron a los que —con significancia estadística— mostraron efectos coincidentes (sea positivo o negativo) tanto en la muestra endofítica temprana (de las 24 h postinoculación) como en la muestra endofítica tardía cosechada a los 20 días postinoculación. Así, exigimos a cada mutante la presencia de efectos del mismo tipo sobre la colonización endofítica en muestras que correspondieron a plantas independientes, cosechadas a tiempos diferentes (24 h y 20 días post inoculación). De los 65 genes mencionados, sólo la mutación de 19 de ellos (29%) resultó en un efecto negativo sobre la colonización endofítica, mientras los mutantes en los 46 genes restantes (71%) mostraron efecto positivo sobre su presencia en la endósfera. De modo análogo a lo que ocurrió con la colonización rizosférica, entre los 65 mutantes los 3 únicos identificados en plásmidos (2 en el plásmido pPagLPU12_2 y 1 en el plásmido pPagLPU12_3) resultaron en efectos positivos en la colonización endofítica. En relación con el propio diseño experimental del ensayo, destaca

que entre los 65 genes identificados no se observó ningún gen reconocido como importantes para el acceso rizosférico. Tal ausencia, esperable, puso de manifiesto la eficacia de la inoculación con altas concentraciones de bacterias como las que usamos (del orden de 10^7 u.f.c./ml según hemos dicho), para eliminar deficiencias en la llegada de las bacterias mutantes a la rizósfera que luego puedan resultar en la detección de genes de efecto temprano y no directamente relacionados con el propio establecimiento en el interior de la raíz. Puede considerarse que los marcadores identificados en este ensayo son propios del establecimiento/crecimiento dentro de la raíz. Entre ellos, la proporción de mutantes con efectos positivos en la colonización respecto de los mutantes con efectos negativos (relación mayor que 2:1) contrasta notablemente con lo observado para los genes de importancia rizosférica, donde la presencia de mutaciones con efectos positivos resultó muy minoritaria respecto a las mutaciones que perturban el acceso a la rizósfera (genes de efectos positivos respecto de genes de efectos negativos en relación menor que 1:4,7). La existencia de numerosos genes cuya pérdida de función mejora la colonización interna de la raíz deriva muy posiblemente de la necesidad de mantener esos genes durante la vida en ambientes diferentes al del interior de las plantas. El nicho endofítico es con seguridad mucho más homeostático que el entorno suelo, para el cual se requieren muchas más funciones para hacer frente a las múltiples variaciones de origen biótico y abiótico. Es un hecho reconocido que la vida endosimbiótica, con espacios físicos y temporales alejados de entornos variables, conduce a la reducción genómica (pérdida de información) derivada de prescindencia de funciones (eliminación de genes) que sólo se requieren para la vida no asociativa (Otero-Bravo & Sabree, 2021). Cabe destacar además que, habiendo identificado un total de 65 genes de un conjunto de 590 genes analizados, es posible que a la expansión del análisis a la totalidad del genoma resulte en la identificación de más de 400 genes relacionados con la vida endofítica, representando por tanto cerca de 10% del genoma. El abordaje a dichos genes requerirá el procesamiento de varios centenares de plantas, de modo de contar con un número adecuado de endófitos que soslayen las limitaciones que impone el cuello de botella al que ya hemos hecho referencia.

En concordancia con las consideraciones anteriores, los grupos funcionales que se encontraron entre los genes vinculados a la colonización endofítica resultaron diversos, tanto en los mutantes con efectos desfavorables como en los que ejercieron efectos positivos sobre la misma. Entre los mutantes que resultan en menos colonización de la endósfera, como ya hemos mencionado, se incluyen mayoritariamente mutantes en genes relacionados con el metabolismo de la envoltura celular (20%) y luego mutantes en genes asociados a la transducción de señales (10%). El primer grupo de mutantes se encuentra seguramente relacionado con el estrecho contacto físico bacteria-planta que caracteriza a la vida endofítica. El segundo grupo destaca posiblemente que el ingreso y persistencia en la endósfera es un evento finamente regulado. A modo de ejemplo puede mencionarse el gen *rscC* que codifica para una histidin-kinasa que forma parte de un sistema de dos componentes junto a *rscD*. RcsC sensa señales ambientales, y en respuesta a ellas fosforila al regulador de respuesta RcsD. Este último gen se encuentra involucrado en la producción de la cápsula del patógeno *P. stewartii* donde ha sido descrito como componente indispensable para el éxito de la proliferación bacteriana en planta (Duong et al., 2018). Tal evidencia muestra que existen determinantes de interacción comunes a asociaciones no patogénicas y patogénicas, un aspecto que podrá investigarse con estudios de homología y en la medida que se cuente con especies de *Pantoea* cuyos fenotipos en planta hayan sido bien caracterizados.

Con una aproximación diferente a la experimental utilizada en el estudio para identificar genes de colonización en *P. agglomerans* LPU12, exploramos la posibilidad de utilizar herramientas de aprendizaje automático (*Machine Learning*), en su variante de aprendizaje supervisado, para intentar identificar—a partir de datos genómicos—especies y aislamientos de *Pantoea* que tengan tropismo positivo (asociativo) hacia plantas (Capítulo VII). Implementamos un procedimiento para reconocer la capacidad de asociarse a plantas de una dada variante genómica del género *Pantoea*, sin necesidad de ensayos de interacción en

cámara de plantas o invernáculo. Así, con algoritmos de clasificación basados en la implementación de “bosques aleatorios”- *Random Forest*- fuimos capaces de reconocer como colonizadoras o no-colonizadoras de plantas, con más de 80% de aciertos y a partir de la sola información genómica, aislamientos de *Pantoea* escogidos de una muestra aleatorizada. Tal posibilidad constituye una herramienta potente para: (a) concentrar el estudio del género *Pantoea* sobre aquellas especies cuyas variantes genómicas predicen capacidades asociativas con plantas, y (b) hacer uso de las predicciones anteriores para realizar estudios genómicos comparativos.

La extensa colonización radicular de las plantas es un evento explorado cada vez más intensamente en el contexto de las potencialidades de biocontrol y biofertilización de muchas de las bacterias que conviven en (y con) las raíces. Los resultados de este trabajo de tesis han permitido acceder al reconocimiento de aislamientos de *Pantoea* con capacidad de colonizar plantas, y en particular a una colección de genes vinculados a la colonización radicular exofítica y endofítica por *P. agglomerans*. El carácter diverso, tanto benéfico como patogénico, de las asociaciones entre *Pantoea* spp. con plantas y diversos hospedadores animales presenta a nuestro trabajo como un insumo de base para estudios genómicos comparativos. Avanzar en el conocimiento de las características que promueven o perturban cada tipo de asociación, permitirán su uso en el diseño de estrategias prácticas orientadas tanto a la promoción del desarrollo de interacciones positivas entre plantas y *Pantoea*, así como a la interferencia de asociaciones de carácter patogénico.



ANEXO



VIII.1. Figuras suplementarias

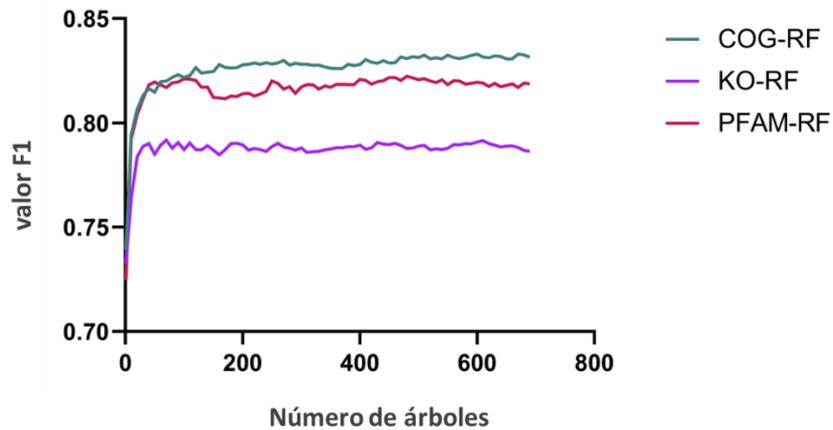


Figura Suplementaria VII.S1. Variación del valor F1 en modelos de Random Forest con diferentes números de árboles de decisión. Se generaron modelos de clasificación para el tropismo a plantas en aislamientos de *Pantoea* utilizando *Random Forest*, con hiperparámetros: *max_features = 'sqrt'*, *criterion = 'entropy'*, *random_state = 123*, y variando el número de árboles (*n_estimators*). El valor F1 se evaluó para cada modelo y se presenta en función del número de árboles utilizados.

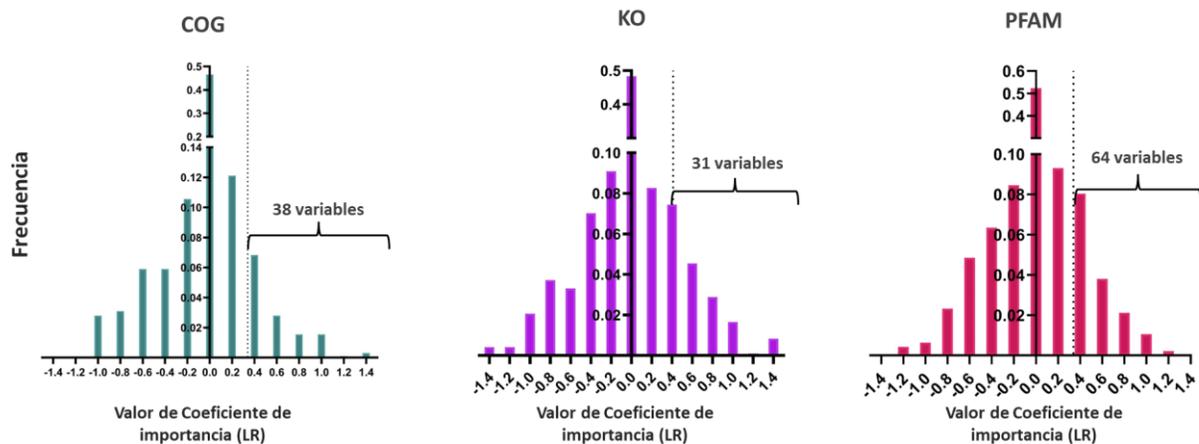


Figura Suplementaria VII.S2. Distribución del Coeficiente de Importancia asignado por los modelos COG-LR_0.4, KO-LR_0.4 y PFAM-LR_0.4. Los modelos, entrenados con conjuntos de datos reducidos al 10% del total de variables originales (admitiendo un coeficiente de correlación de Pearson entre variables de hasta 0.4) y el algoritmo *Logistic Regression*, asignan un valor de importancia a cada variable para la clasificación del tropismo a plantas en cepas de *Pantoea*. Los valores positivos corresponden a variables que correlacionan, según el modelo, con la clasificación positiva (Planta). La mayoría de las variables tienen un valor de importancia cercano a cero. Las variables más significativas fueron aquellas con valores mayores a la media más una desviación estándar, indicadas con líneas punteadas en cada gráfico.

VIII.2. Tablas suplementarias

Tabla VI.S1. Genes esenciales de *P. agglomerans* LPU12 identificados en los ensayos TnSeq (Capítulo VI).

Replicón	locus_tag	Gen (nombre)	COG	categoría COG	KO	Definición asociada al KO	Criterio*
cromosoma	GONCAC_00050	<i>dsbA</i>	COG1651	O	K03673	dsbA; protein dithiol oxidoreductase (disulfide-forming) [EC:1.8.4.15]	E3
cromosoma	GONCAC_00065	<i>yihA</i>	COG0218	D	K03978	engB; GTP-binding protein	E3
cromosoma	GONCAC_00090	<i>glnA</i>	COG0174	E	K01915	glnA, GLUL; glutamine synthetase [EC:6.3.1.2]	E3
cromosoma	GONCAC_00140	<i>spoT</i>	COG0317	KT	K01139	spoT; GTP diphosphokinase / guanosine-3',5'-bis(diphosphate) 3'-diphosphatase [EC:2.7.6.5 3.1.7.2]	E2
cromosoma	GONCAC_00145	<i>rpoZ</i>	COG1758	K	K03060	rpoZ; DNA-directed RNA polymerase subunit omega [EC:2.7.7.6]	E3
cromosoma	GONCAC_00150	<i>gmk</i>	COG0194	F	K00942	gmk, GUK1; guanylate kinase [EC:2.7.4.8]	E3
cromosoma	GONCAC_00230	<i>dut</i>	COG0756	F	K01520	dut, DUT; dUTP diphosphatase [EC:3.6.1.23]	E3
cromosoma	GONCAC_00235	<i>coaBC</i>	COG0452	H	K13038	coaBC, dfp; phosphopantothenoylecysteine decarboxylase / phosphopantothenate---cysteine ligase [EC:4.1.1.36 6.3.2.5]	E3
cromosoma	GONCAC_00265	-	COG0463	M	K12984	waaE, kdtX; (heptosyl)LPS beta-1,4-glucosyltransferase [EC:2.4.1.-]	E3
cromosoma	GONCAC_00270	<i>waaA</i>	COG1519	M	K02527	kdtA, waaA; 3-deoxy-D-manno-octulosonic-acid transferase [EC:2.4.99.12 2.4.99.13 2.4.99.14 2.4.99.15]	E3
cromosoma	GONCAC_00275	-	COG0438	M	K02844	waaG, rfaG; UDP-glucose:(heptosyl)LPS alpha-1,3-glucosyltransferase [EC:2.4.1.-]	E3
cromosoma	GONCAC_00280	<i>rfaQ</i>	COG0859	M	K02849	waaQ, rfaQ; lipopolysaccharide heptosyltransferase III [EC:2.4.99.25]	E3
cromosoma	GONCAC_00300	<i>rfaC</i>	COG0859	M	K02841	waaC, rfaC; lipopolysaccharide heptosyltransferase I [EC:2.4.99.23]	E3
cromosoma	GONCAC_00305	<i>rfaF</i>	COG0859	M	K02843	waaF, rfaF; lipopolysaccharide heptosyltransferase II [EC:2.4.99.24]	E3
cromosoma	GONCAC_00310	<i>rfaD</i>	COG0451	GM	K03274	gmhD, rfaD; ADP-L-glycero-D-manno-heptose 6-epimerase [EC:5.1.3.20]	E3
cromosoma	GONCAC_00365	<i>cysE</i>	COG1045	E	K00640	cysE; serine O-acetyltransferase [EC:2.3.1.30]	E3
cromosoma	GONCAC_00380	<i>cpxR</i>	COG0745	K	K07662	cpxR; two-component system, OmpR family, response regulator CpxR	E3
cromosoma	GONCAC_00420	<i>tpiA</i>	COG0149	G	K01803	TPI, tpiA; triosephosphate isomerase (TIM) [EC:5.3.1.1]	E3
cromosoma	GONCAC_00490	<i>priA</i>	COG1198	L	K04066	priA; primosomal protein N' (replication factor Y) (superfamily II helicase) [EC:5.6.2.4]	E3
cromosoma	GONCAC_00825	-	COG0400	S	K06999	K06999; phospholipase/carboxylesterase	E2
cromosoma	GONCAC_01040	<i>hemB</i>	COG0113	H	K01698	hemB, ALAD; porphobilinogen synthase [EC:4.2.1.24]	E3
cromosoma	GONCAC_01155	<i>ftsY</i>	COG0552	U	K03110	ftsY; fused signal recognition particle receptor	E3

cromosoma	GONCAC_01170	<i>rpoH</i>	COG0568	K	K03089	rpoH; RNA polymerase sigma-32 factor	E3
cromosoma	GONCAC_01285	<i>asd</i>	COG0136	E	K00133	asd; aspartate-semialdehyde dehydrogenase [EC:1.2.1.11]	E3
cromosoma	GONCAC_01360	<i>glpD</i>	COG0578	C	K00111	glpA, glpD; glycerol-3-phosphate dehydrogenase [EC:1.1.5.3]	E2
cromosoma	GONCAC_01465	-	28IJR	S			E3
cromosoma	GONCAC_01510	<i>aroK</i>	COG0703	F	K00891	aroK, aroL; shikimate kinase [EC:2.7.1.71]	E3
cromosoma	GONCAC_01515	<i>aroB</i>	COG0337	E	K01735	aroB; 3-dehydroquinate synthase [EC:4.2.3.4]	E3
cromosoma	GONCAC_01520	<i>damX</i>	COG3266	D	K03112	damX; DamX protein	E3
cromosoma	GONCAC_01525	<i>dam</i>	COG0338	H	K06223	dam; DNA adenine methylase [EC:2.1.1.72]	E3
cromosoma	GONCAC_01535	<i>gph</i>	COG0546	G	K01091	gph; phosphoglycolate phosphatase [EC:3.1.3.18]	E2
cromosoma	GONCAC_01540	<i>trpS</i>	COG0180	J	K01867	WARS, trpS; tryptophanyl-tRNA synthetase [EC:6.1.1.2]	E2
cromosoma	GONCAC_01600	<i>crp</i>	COG0664	K	K10914	crp; CRP/FNR family transcriptional regulator, cyclic AMP receptor protein	E3
cromosoma	GONCAC_01695	<i>tusD</i>	COG1553	J	K07235	tusD, dsrE; tRNA 2-thiouridine synthesizing protein D [EC:2.8.1.-]	E3
cromosoma	GONCAC_01700	<i>tusC</i>	COG2923	P	K07236	tusC, dsrF; tRNA 2-thiouridine synthesizing protein C	E3
cromosoma	GONCAC_01705	<i>tusB</i>	COG2168	P	K07237	tusB, dsrH; tRNA 2-thiouridine synthesizing protein B	E3
cromosoma	GONCAC_01710	-	COG0048	J	K02950	RP-S12, MRPS12, rpsL; small subunit ribosomal protein S12	E3
cromosoma	GONCAC_01715	<i>rpsG</i>	COG0049	J	K02992	RP-S7, MRPS7, rpsG; small subunit ribosomal protein S7	E3
cromosoma	GONCAC_01720	<i>fusA</i>	COG0480	J	K02355	fusA, GFM, EFG; elongation factor G	E3
cromosoma	GONCAC_01750	<i>rplC</i>	COG0087	J	K02906	RP-L3, MRPL3, rplC; large subunit ribosomal protein L3	E3
cromosoma	GONCAC_01755	<i>rplD</i>	COG0088	J	K02926	RP-L4, MRPL4, rplD; large subunit ribosomal protein L4	E3
cromosoma	GONCAC_01760	<i>rplW</i>	COG0089	J	K02892	RP-L23, MRPL23, rplW; large subunit ribosomal protein L23	E3
cromosoma	GONCAC_01770	<i>rpsS</i>	COG0185	J	K02965	RP-S19, RSM19, rpsS; small subunit ribosomal protein S19	E3
cromosoma	GONCAC_01775	<i>rplV</i>	COG0091	J	K02890	RP-L22, MRPL22, rplV; large subunit ribosomal protein L22	E3
cromosoma	GONCAC_01780	<i>rpsC</i>	COG0092	J	K02982	RP-S3, rpsC; small subunit ribosomal protein S3	E3
cromosoma	GONCAC_01785	<i>rplP</i>	COG0197	J	K02878	RP-L16, MRPL16, rplP; large subunit ribosomal protein L16	E3
cromosoma	GONCAC_01795	<i>rpsQ</i>	COG0186	J	K02961	RP-S17, MRPS17, rpsQ; small subunit ribosomal protein S17	E3
cromosoma	GONCAC_01805	<i>rplN</i>	COG0093	J	K02874	RP-L14, MRPL14, rplN; large subunit ribosomal protein L14	E3
cromosoma	GONCAC_01810	<i>rplX</i>	COG0198	J	K02895	RP-L24, MRPL24, rplX; large subunit ribosomal protein L24	E3
cromosoma	GONCAC_01815	<i>rplE</i>	COG0094	J	K02931	RP-L5, MRPL5, rplE; large subunit ribosomal protein L5	E3
cromosoma	GONCAC_01820	<i>rpsN</i>	COG0199	J	K02954	RP-S14, MRPS14, rpsN; small subunit ribosomal protein S14	E3

cromosoma	GONCAC_01825	<i>rpsH</i>	COG0096	J	K02994	RP-S8, MRPS8, rpsH; small subunit ribosomal protein S8	E3
cromosoma	GONCAC_01830	<i>rplF</i>	COG0097	J	K02933	RP-L6, MRPL6, rplF; large subunit ribosomal protein L6	E3
cromosoma	GONCAC_01835	<i>rplR</i>	COG0256	J	K02881	RP-L18, MRPL18, rplR; large subunit ribosomal protein L18	E3
cromosoma	GONCAC_01840	<i>rpsE</i>	COG0098	J	K02988	RP-S5, MRPS5, rpsE; small subunit ribosomal protein S5	E3
cromosoma	GONCAC_01845	<i>rpmD</i>	COG1841	J	K02907	RP-L30, MRPL30, rpmD; large subunit ribosomal protein L30	E3
cromosoma	GONCAC_01850	<i>rplO</i>	COG0200	J	K02876	RP-L15, MRPL15, rplO; large subunit ribosomal protein L15	E3
cromosoma	GONCAC_01855	<i>secY</i>	COG0201	U	K03076	secY; preprotein translocase subunit SecY	E3
cromosoma	GONCAC_01860	<i>rpsM</i>	COG0099	J	K02952	RP-S13, rpsM; small subunit ribosomal protein S13	E3
cromosoma	GONCAC_01865	<i>rpsK</i>	COG0100	J	K02948	RP-S11, MRPS11, rpsK; small subunit ribosomal protein S11	E3
cromosoma	GONCAC_01870	<i>rpsD</i>	COG0522	J	K02986	RP-S4, NAM9, rpsD; small subunit ribosomal protein S4	E3
cromosoma	GONCAC_01875	<i>rpoA</i>	COG0202	K	K03040	rpoA; DNA-directed RNA polymerase subunit alpha [EC:2.7.7.6]	E3
cromosoma	GONCAC_01880	<i>rplQ</i>	COG0203	J	K02879	RP-L17, MRPL17, rplQ; large subunit ribosomal protein L17	E3
cromosoma	GONCAC_01905	<i>trkA</i>	COG0569	P	K03499	trkA, ktrA, ktrC; trk/ktr system potassium uptake protein	E3
cromosoma	GONCAC_01915	<i>fmt</i>	COG0223	J	K00604	MTFMT, fmt; methionyl-tRNA formyltransferase [EC:2.1.2.9]	E3
cromosoma	GONCAC_01920	<i>def</i>	COG0242	J	K01462	PDF, def; peptide deformylase [EC:3.5.1.88]	E3
cromosoma	GONCAC_01940	<i>tsaC</i>	COG0009	J	K07566	tsaC, rimN, SUA5, YRDC; L-threonylcarbamoyladenylate synthase [EC:2.7.7.87]	E2
cromosoma	GONCAC_02045	<i>accC</i>	COG0439	I	K01961	accC; acetyl-CoA carboxylase, biotin carboxylase subunit [EC:6.4.1.2 6.3.4.14]	E3
cromosoma	GONCAC_02050	<i>accB</i>	COG0511	I	K02160	accB, bccP; acetyl-CoA carboxylase biotin carboxyl carrier protein	E3
cromosoma	GONCAC_02055	<i>aroQ</i>	COG0757	E	K03786	aroQ, qutE; 3-dehydroquinate dehydratase II [EC:4.2.1.10]	E3
cromosoma	GONCAC_02075	<i>mreB</i>	COG1077	D	K03569	mreB; rod shape-determining protein MreB and related proteins	E3
cromosoma	GONCAC_02175	<i>asd</i>	COG0688	I	K01613	psd, PISD; phosphatidylserine decarboxylase [EC:4.1.1.65]	E3
cromosoma	GONCAC_02180	<i>rsgA</i>	COG1162	S	K06949	rsgA, engC; ribosome biogenesis GTPase / thiamine phosphate phosphatase [EC:3.6.1.- 3.1.3.100]	E3
cromosoma	GONCAC_02185	<i>orn</i>	COG1949	L	K13288	orn, REX2, REXO2; oligoribonuclease [EC:3.1.-.-]	E3
cromosoma	GONCAC_02215	<i>tsaE</i>	COG0802	S	K06925	tsaE; tRNA threonylcarbamoyladenine biosynthesis protein TsaE	E3
cromosoma	GONCAC_02220	<i>amiB</i>	COG0860	M	K01448	amiABC; N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28]	E3
cromosoma	GONCAC_02260	<i>purA</i>	COG0104	F	K01939	purA, ADSS; adenylosuccinate synthase [EC:6.3.4.4]	E3
cromosoma	GONCAC_02305	<i>rpsF</i>	COG0360	J	K02990	RP-S6, MRPS6, rpsF; small subunit ribosomal protein S6	E3
cromosoma	GONCAC_02370	<i>ppa</i>	COG0221	C	K01507	ppa; inorganic pyrophosphatase [EC:3.6.1.1]	E3

cromosoma	GONCAC_02500	<i>RNaseP_bact_a</i>		ncRNA			E2
cromosoma	GONCAC_02705	<i>rpoD</i>	COG0568	K	K03086	rpoD; RNA polymerase primary sigma factor	E3
cromosoma	GONCAC_02710	-		-			E3
cromosoma	GONCAC_02715	<i>dnaG</i>	COG0358	L	K02316	dnaG; DNA primase [EC:2.7.7.101]	E3
cromosoma	GONCAC_02720	<i>rpsU</i>	COG0828	J	K02970	RP-S21, MRPS21, rpsU; small subunit ribosomal protein S21	E3
cromosoma	GONCAC_02725	<i>tsaD</i>	COG0533	O	K25706	tsaD; tRNA N6-adenosine threonylcarbamoyltransferase [EC:2.3.1.234]	E3
cromosoma	GONCAC_02745	-	COG0617	H	K00974	cca; tRNA nucleotidyltransferase (CCA-adding enzyme) [EC:2.7.7.72 3.1.3.- 3.1.4.-]	E3
cromosoma	GONCAC_02765	<i>hldE</i>	COG0615	H	K03272	gmhC, hldE, waaE, rfaE; D-beta-D-heptose 7-phosphate kinase/ D-beta-D-heptose 1-phosphate adenosyltransferase [EC:2.7.1.167 2.7.7.70]	E3
cromosoma	GONCAC_02770	-	COG2960	S	K09806	ubiK; ubiquinone biosynthesis accessory factor UbiK	E2
cromosoma	GONCAC_02775	<i>ribB</i>	COG0108	H	K02858	ribB, RIB3; 3,4-dihydroxy 2-butanone 4-phosphate synthase [EC:4.1.99.12]	E3
cromosoma	GONCAC_02820	-	COG0187	L	K02622	parE; topoisomerase IV subunit B [EC:5.6.2.2]	E3
cromosoma	GONCAC_02845	<i>parC</i>	COG0188	L	K02621	parC; topoisomerase IV subunit A [EC:5.6.2.2]	E3
cromosoma	GONCAC_02850	-	COG0204	I	K00655	plsC; 1-acyl-sn-glycerol-3-phosphate acyltransferase [EC:2.3.1.51]	E3
cromosoma	GONCAC_03260	<i>ruvX</i>	COG0816	L	K07447	ruvX; putative pre-16S rRNA nuclease [EC:3.1.-.-]	E3
cromosoma	GONCAC_03265	-	COG1678	K	K07735	algH; putative transcriptional regulator	E3
cromosoma	GONCAC_03310	<i>metK</i>	COG0192	H	K00789	metK, MAT; S-adenosylmethionine synthetase [EC:2.5.1.6]	E3
cromosoma	GONCAC_03345	<i>pgk</i>	COG0126	G	K00927	PGK, pgk; phosphoglycerate kinase [EC:2.7.2.3]	E3
cromosoma	GONCAC_03350	<i>fbaA</i>	COG0191	G	K01624	FBA, fbaA; fructose-bisphosphate aldolase, class II [EC:4.1.2.13]	E3
cromosoma	GONCAC_03375	<i>rpiA</i>	COG0120	G	K01807	rpiA; ribose 5-phosphate isomerase A [EC:5.3.1.6]	E3
cromosoma	GONCAC_03405	<i>pepP</i>	COG0006	E	K01262	pepP; Xaa-Pro aminopeptidase [EC:3.4.11.9]	E3
cromosoma	GONCAC_03410	<i>ubiH</i>	COG0654	CH	K03185	ubiH; 2-octaprenyl-6-methoxyphenol hydroxylase [EC:1.14.13.-]	E3
cromosoma	GONCAC_03415	<i>ubil</i>	COG0654	CH	K18800	ubil; 2-polyprenylphenol 6-hydroxylase [EC:1.14.13.240]	E3
cromosoma	GONCAC_03455	<i>ygfZ</i>	COG0354	S	K06980	ygfZ; tRNA-modifying protein YgfZ	E3
cromosoma	GONCAC_03475	<i>xerD</i>	COG4974	D	K04763	xerD; integrase/recombinase XerD	E3
cromosoma	GONCAC_03490	<i>prfB</i>	COG1186	J	K02836	prfB; peptide chain release factor 2	E3
cromosoma	GONCAC_03495	<i>lysS</i>	COG1190	J	K04567	KARS, lysS; lysyl-tRNA synthetase, class II [EC:6.1.1.6]	E3
cromosoma	GONCAC_03585	<i>lgt</i>	COG0682	M	K13292	lgt, umpA; phosphatidylglycerol---prolipoprotein diacylglycerol transferase [EC:2.5.1.145]	E3

cromosoma	GONCAC_03590	<i>thyA</i>	COG0207	F	K00560	thyA, TYMS; thymidylate synthase [EC:2.1.1.45]	E3
cromosoma	GONCAC_03615	<i>recC</i>	COG1330	L	K03583	recC; exodeoxyribonuclease V gamma subunit [EC:3.1.11.5]	E3
cromosoma	GONCAC_03625	<i>recB</i>	COG1074	L	K03582	recB; exodeoxyribonuclease V beta subunit [EC:3.1.11.5]	E3
cromosoma	GONCAC_03800	<i>pyrG</i>	COG0504	F	K01937	pyrG, CTPS; CTP synthase [EC:6.3.4.2]	E3
cromosoma	GONCAC_03805	<i>eno</i>	COG0148	G	K01689	ENO1_2_3, eno; enolase 1/2/3 [EC:4.2.1.11]	E3
cromosoma	GONCAC_03870	<i>ftsB</i>	COG2919	D	K05589	ftsB; cell division protein FtsB	E3
cromosoma	GONCAC_03875	<i>ispD</i>	COG1211	I	K00991	ispD; 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase [EC:2.7.7.60]	E3
cromosoma	GONCAC_03900	<i>nlpD</i>	COG1388	DM	K06194	nlpD; lipoprotein NlpD	E3
cromosoma	GONCAC_03925	-	COG2916	L	K03746	hns; DNA-binding protein H-NS	E3
cromosoma	GONCAC_04085	-		-	K03555	mutS; DNA mismatch repair protein MutS	E3
cromosoma	GONCAC_04145	<i>recA</i>	COG0468	L	K03553	recA; recombination protein RecA	E3
cromosoma	GONCAC_04155	<i>alaS</i>	COG0013	J	K01872	AARS, alaS; alanyl-tRNA synthetase [EC:6.1.1.7]	E3
cromosoma	GONCAC_04160	<i>csrA</i>	COG1551	T	K03563	csrA; carbon storage regulator	E3
cromosoma	GONCAC_04220	<i>ffh</i>	COG0541	U	K03106	SRP54, ffh; signal recognition particle subunit SRP54 [EC:3.6.5.4]	E3
cromosoma	GONCAC_04225	<i>rpsP</i>	COG0228	J	K02959	RP-S16, MRPS16, rpsP; small subunit ribosomal protein S16	E3
cromosoma	GONCAC_04230	<i>rimM</i>	COG0806	J	K02860	rimM; 16S rRNA processing protein RimM	E3
cromosoma	GONCAC_04235	<i>trmD</i>	COG0336	J	K00554	trmD; tRNA (guanine37-N1)-methyltransferase [EC:2.1.1.228]	E3
cromosoma	GONCAC_04255	<i>tyrA</i>	COG0287	E	K14187	tyrA; chorismate mutase / prephenate dehydrogenase [EC:5.4.99.5 1.3.1.12]	E2
cromosoma	GONCAC_04270	<i>bamD</i>	COG4105	M	K05807	bamD; outer membrane protein assembly factor BamD	E3
cromosoma	GONCAC_04335	<i>pssA</i>	COG1502	I	K00998	pssA; CDP-diacylglycerol---serine O-phosphatidyltransferase [EC:2.7.8.8]	E3
cromosoma	GONCAC_04610	<i>pasT</i>	COG2867	I			E3
cromosoma	GONCAC_04615	-	COG2914	S	K09801	K09801; uncharacterized protein	E3
cromosoma	GONCAC_04630	<i>nadK</i>	COG0061	H	K00858	ppnK, NADK; NAD+ kinase [EC:2.7.1.23]	E3
cromosoma	GONCAC_04635	<i>grpE</i>	COG0576	O	K03687	GRPE; molecular chaperone GrpE	E3
cromosoma	GONCAC_04665	<i>rpoE</i>	COG1595	K	K03088	rpoE; RNA polymerase sigma-70 factor, ECF subfamily	E3
cromosoma	GONCAC_04685	<i>rseC</i>	COG3086	T	K03803	rseC; sigma-E factor negative regulatory protein RseC	E2
cromosoma	GONCAC_04690	<i>lepA</i>	COG0481	M	K03596	lepA; GTP-binding protein LepA	E3
cromosoma	GONCAC_04695	<i>lepB</i>	COG0681	U	K03100	lepB; signal peptidase I [EC:3.4.21.89]	E3
cromosoma	GONCAC_04700	<i>rnc</i>	COG0571	J	K03685	rnc, DROSHA, RNT1; ribonuclease III [EC:3.1.26.3]	E3

cromosoma	GONCAC_04705	<i>era</i>	COG1159	S	K03595	era, ERAL1; GTPase	E3
cromosoma	GONCAC_04715	<i>pdxJ</i>	COG0854	H	K03474	pdxJ; pyridoxine 5-phosphate synthase [EC:2.6.99.2]	E3
cromosoma	GONCAC_04720	<i>acpS</i>	COG0736	I	K00997	acpS; holo-[acyl-carrier protein] synthase [EC:2.7.8.7]	E3
cromosoma	GONCAC_04790	<i>glyA</i>	COG0112	E	K00600	glyA, SHMT; glycine hydroxymethyltransferase [EC:2.1.2.1]	E3
cromosoma	GONCAC_04820	<i>suhB</i>	COG0483	G	K01092	E3.1.3.25, IMPA, suhB; myo-inositol-1(or 4)-monophosphatase [EC:3.1.3.25]	E3
cromosoma	GONCAC_04835	<i>iscS</i>	COG1104	H	K04487	iscS, NFS1; cysteine desulfurase [EC:2.8.1.7]	E3
cromosoma	GONCAC_04885	<i>ispG</i>	COG0821	I	K03526	gcpE, ispG; (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase [EC:1.17.7.1 1.17.7.3]	E3
cromosoma	GONCAC_04890	<i>hisS</i>	COG0124	J	K01892	HARS, hisS; histidyl-tRNA synthetase [EC:6.1.1.21]	E3
cromosoma	GONCAC_04900	<i>bamB</i>	COG1520	M	K17713	bamB; outer membrane protein assembly factor BamB	E3
cromosoma	GONCAC_04905	<i>der</i>	COG1160	S	K03977	engA, der; GTPase	E3
cromosoma	GONCAC_04930	<i>xseA</i>	COG1570	L	K03601	xseA; exodeoxyribonuclease VII large subunit [EC:3.1.11.6]	E2
cromosoma	GONCAC_04935	<i>guaB</i>	COG0516	F	K00088	IMPDH, guaB; IMP dehydrogenase [EC:1.1.1.205]	E2
cromosoma	GONCAC_04940	<i>guaA</i>	COG0518	F	K01951	guaA, GMPS; GMP synthase (glutamine-hydrolysing) [EC:6.3.5.2]	E3
cromosoma	GONCAC_05130	<i>dapE</i>	COG0624	E	K01439	dapE; succinyl-diaminopimelate desuccinylase [EC:3.5.1.18]	E3
cromosoma	GONCAC_05135	-	COG1393	P			E3
cromosoma	GONCAC_05195	<i>hemF</i>	COG0408	H	K00228	CPOX, hemF; coproporphyrinogen III oxidase [EC:1.3.3.3]	E3
cromosoma	GONCAC_05255	<i>crr</i>	COG2190	G	K02777	crr; sugar PTS system EIIA component [EC:2.7.1.-]	E2
cromosoma	GONCAC_05285	<i>zipA</i>	COG3115	D	K03528	zipA; cell division protein ZipA	E3
cromosoma	GONCAC_05290	<i>ligA</i>	COG0272	L	K01972	E6.5.1.2, ligA, ligB; DNA ligase (NAD+) [EC:6.5.1.2]	E3
cromosoma	GONCAC_05335	-	COG0008	J	K01885	EARS, gltX; glutamyl-tRNA synthetase [EC:6.1.1.17]	E3
cromosoma	GONCAC_05345	-	COG3415	L			E2
cromosoma	GONCAC_05595	-		-			E3
cromosoma	GONCAC_05770	-		-			E2
cromosoma	GONCAC_05880	<i>prmB</i>	COG2890	J	K07320	prmB; ribosomal protein L3 glutamine methyltransferase [EC:2.1.1.298]	E3
cromosoma	GONCAC_05885	<i>aroC</i>	COG0082	E	K01736	aroC; chorismate synthase [EC:4.2.3.5]	E3
cromosoma	GONCAC_05915	<i>fabB</i>	COG0304	IQ	K00647	fabB; 3-oxoacyl-[acyl-carrier-protein] synthase I [EC:2.3.1.41]	E3
cromosoma	GONCAC_05965	<i>accD</i>	COG0777	I	K01963	accD; acetyl-CoA carboxylase carboxyl transferase subunit beta [EC:6.4.1.2 2.1.3.15]	E3
cromosoma	GONCAC_05970	<i>folC</i>	COG0285	H	K11754	folC; dihydrofolate synthase / folylpolyglutamate synthase [EC:6.3.2.12 6.3.2.17]	E3

cromosoma	GONCAC_05990	<i>ubiX</i>	COG0163	H	K03186	ubiX, bsdB, PAD1; flavin prenyltransferase [EC:2.5.1.129]	E3
cromosoma	GONCAC_06070	<i>nuoA</i>	COG0838	C	K00330	nuoA; NADH-quinone oxidoreductase subunit A [EC:7.1.1.2]	E3
cromosoma	GONCAC_06075	<i>nuoB</i>	COG0377	C	K00331	nuoB; NADH-quinone oxidoreductase subunit B [EC:7.1.1.2]	E3
cromosoma	GONCAC_06080	<i>nuoC</i>	COG0649	C	K13378	nuoCD; NADH-quinone oxidoreductase subunit C/D [EC:7.1.1.2]	E3
cromosoma	GONCAC_06085	<i>nuoE</i>	COG1905	C	K00334	nuoE; NADH-quinone oxidoreductase subunit E [EC:7.1.1.2]	E3
cromosoma	GONCAC_06095	<i>nuoG</i>	COG1034	C	K00336	nuoG; NADH-quinone oxidoreductase subunit G [EC:7.1.1.2]	E3
cromosoma	GONCAC_06100	<i>nuoH</i>	COG1005	C	K00337	nuoH; NADH-quinone oxidoreductase subunit H [EC:7.1.1.2]	E3
cromosoma	GONCAC_06105	<i>nuoI</i>	COG1143	C	K00338	nuoI; NADH-quinone oxidoreductase subunit I [EC:7.1.1.2]	E3
cromosoma	GONCAC_06110	<i>nuoJ</i>	COG0839	C	K00339	nuoJ; NADH-quinone oxidoreductase subunit J [EC:7.1.1.2]	E3
cromosoma	GONCAC_06115	<i>nuoK</i>	COG0713	C	K00340	nuoK; NADH-quinone oxidoreductase subunit K [EC:7.1.1.2]	E3
cromosoma	GONCAC_06120	<i>nuoL</i>	COG1009	CP	K00341	nuoL; NADH-quinone oxidoreductase subunit L [EC:7.1.1.2]	E3
cromosoma	GONCAC_06125	<i>nuoM</i>	COG1008	C	K00342	nuoM; NADH-quinone oxidoreductase subunit M [EC:7.1.1.2]	E3
cromosoma	GONCAC_06130	<i>nuoN</i>	COG1007	C	K00343	nuoN; NADH-quinone oxidoreductase subunit N [EC:7.1.1.2]	E3
cromosoma	GONCAC_06175	<i>ubiG</i>	COG2227	H	K00568	ubiG; 2-polyprenyl-6-hydroxyphenyl methylase / 3-demethylubiquinone-9 3-methyltransferase [EC:2.1.1.222 2.1.1.64]	E3
cromosoma	GONCAC_06180	<i>gyrA</i>	COG0188	L	K02469	gyrA; DNA gyrase subunit A [EC:5.6.2.2]	E3
cromosoma	GONCAC_06245	<i>apbE</i>	COG1477	H	K03734	apbE; FAD:protein FMN transferase [EC:2.7.1.180]	E3
cromosoma	GONCAC_06280	<i>yejM</i>	COG3083	S	K07014	K07014; uncharacterized protein	E2
cromosoma	GONCAC_06285	<i>yejL</i>	COG3082	S	K09904	K09904; uncharacterized protein	E3
cromosoma	GONCAC_06295	<i>rplY</i>	COG1825	J	K02897	RP-L25, rplY; large subunit ribosomal protein L25	E3
cromosoma	GONCAC_06505	<i>metG</i>	COG0073	J	K01874	MARS, metG; methionyl-tRNA synthetase [EC:6.1.1.10]	E3
cromosoma	GONCAC_06685	-	2ABFX	S			E2
cromosoma	GONCAC_06760	-		-			E3
cromosoma	GONCAC_06775	-	COG0451	GM	K22252	tld; GDP-6-deoxy-D-talose 4-dehydrogenase [EC:1.1.1.135]	E2
cromosoma	GONCAC_06780	-	COG0438	M			E2
cromosoma	GONCAC_06790	<i>gndA</i>	COG0362	G	K00033	PGD, gnd, gntZ; 6-phosphogluconate dehydrogenase [EC:1.1.1.44 1.1.1.343]	E3
cromosoma	GONCAC_07510	-	COG1974	K			E3
cromosoma	GONCAC_07775	-	2C08M	S			E3
cromosoma	GONCAC_07810	-	COG0018	J			E2

cromosoma	GONCAC_08060	<i>argS</i>	COG0173	J	K01887	RARS, argS; arginyl-tRNA synthetase [EC:6.1.1.19]	E3
cromosoma	GONCAC_08125	<i>aspS</i>		-	K01876	DARS2, aspS; aspartyl-tRNA synthetase [EC:6.1.1.12]	E3
cromosoma	GONCAC_08130	<i>nudB</i>	COG1051	F	K08310	nudB, ntpA; dihydroneopterin triphosphate diphosphatase [EC:3.6.1.67]	E2
cromosoma	GONCAC_08140	<i>ruvC</i>	COG0817	L	K01159	ruvC; crossover junction endodeoxyribonuclease RuvC [EC:3.1.21.10]	E3
cromosoma	GONCAC_08145	<i>ruvA</i>	COG0632	L	K03550	ruvA; holliday junction DNA helicase RuvA	E3
cromosoma	GONCAC_08150	<i>ruvB</i>	COG2255	L	K03551	ruvB; holliday junction DNA helicase RuvB [EC:5.6.2.4]	E3
cromosoma	GONCAC_08175	<i>lpxM</i>	COG1560	M	K02560	lpxM, msbB; lauroyl-Kdo2-lipid IVA myristoyltransferase [EC:2.3.1.243]	E3
cromosoma	GONCAC_08360	-	COG3109	T	K03607	proQ; ProP effector	E3
cromosoma	GONCAC_08365	<i>prc</i>	COG0793	M	K03797	E3.4.21.102, prc, ctpA; carboxyl-terminal processing protease [EC:3.4.21.102]	E2
cromosoma	GONCAC_08520	<i>minE</i>	COG0851	D	K03608	minE; cell division topological specificity factor	E2
cromosoma	GONCAC_08550	<i>dsbB</i>	COG1495	O	K03611	dsbB; protein dithiol:quinone oxidoreductase [EC:1.8.5.9]	E3
cromosoma	GONCAC_08570	<i>dadX</i>	COG0787	M	K01775	alr; alanine racemase [EC:5.1.1.1]	E3
cromosoma	GONCAC_08620	<i>gapA</i>	COG0057	G	K00134	GAPDH, gapA; glyceraldehyde 3-phosphate dehydrogenase (phosphorylating) [EC:1.2.1.12]	E3
cromosoma	GONCAC_08665	<i>xthA</i>	COG0708	L	K01142	E3.1.11.2, xthA; exodeoxyribonuclease III [EC:3.1.11.2]	E3
cromosoma	GONCAC_08695	<i>rssB</i>	COG0745	KT	K02485	rssB, hnr; two-component system, response regulator	E3
cromosoma	GONCAC_08700	<i>galU</i>	COG1210	M	K00963	UGP2, galU, galF; UTP--glucose-1-phosphate uridylyltransferase [EC:2.7.7.9]	E3
cromosoma	GONCAC_08705	<i>ugd</i>	COG1004	M	K00012	UGDH, ugd; UDPglucose 6-dehydrogenase [EC:1.1.1.22]	E3
cromosoma	GONCAC_08710	-	COG0451	M	K08679	GAE, cap1J; UDP-glucuronate 4-epimerase [EC:5.1.3.6]	E3
cromosoma	GONCAC_08810	<i>ispZ</i>	COG2917	D	K06190	ispZ; intracellular septation protein	E2
cromosoma	GONCAC_08905	-	COG0550	L	K03168	topA; DNA topoisomerase I [EC:5.6.2.1]	E3
cromosoma	GONCAC_08920	<i>ribA</i>	COG0807	H	K01497	ribA, RIB1; GTP cyclohydrolase II [EC:3.5.4.25]	E3
cromosoma	GONCAC_08930	-	COG3771	S	K08992	lapA; lipopolysaccharide assembly protein A	E3
cromosoma	GONCAC_08935	<i>lapB</i>	COG2956	G	K19804	lapB; lipopolysaccharide assembly protein B	E3
cromosoma	GONCAC_08945	<i>yciH</i>	COG0023	J	K03113	EIF1, SUI1; translation initiation factor 1	E2
cromosoma	GONCAC_09020	<i>fabI</i>	COG0623	I	K00208	fabI; enoyl-[acyl-carrier protein] reductase I [EC:1.3.1.9 1.3.1.10]	E3
cromosoma	GONCAC_09025	<i>sapF</i>	COG1123	P	K19230	sapF; cationic peptide transport system ATP-binding protein	E3
cromosoma	GONCAC_09030	<i>sapD</i>	COG1123	P	K19229	sapD; cationic peptide transport system ATP-binding protein	E3
cromosoma	GONCAC_09035	<i>sapC</i>	COG4171	V	K19228	sapC; cationic peptide transport system permease protein	E3

cromosoma	GONCAC_09040	-	COG0582	L			E3
cromosoma	GONCAC_09065	<i>sapB</i>	COG4168	P	K19227	sapB; cationic peptide transport system permease protein	E3
cromosoma	GONCAC_09070	<i>sapA</i>	COG4166	E	K19226	sapA; cationic peptide transport system substrate-binding protein	E3
cromosoma	GONCAC_09510	-	2CSJ8	S			E2
cromosoma	GONCAC_09675	-	COG2771	K			E2
cromosoma	GONCAC_10300	<i>iscR</i>	COG1959	K	K19587	oqxR; Rrf2 family transcriptional regulator, repressor of oqxAB	E3
cromosoma	GONCAC_10420	<i>rsxA</i>	COG4657	C	K03617	rnfA; H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit A	E3
cromosoma	GONCAC_10435	<i>rsxD</i>	COG4658	C	K03614	rnfD; H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit D [EC:7.1.1.11 7.2.1.2]	E3
cromosoma	GONCAC_10490	<i>tyrS</i>	COG0162	J	K01866	YARS, tyrS; tyrosyl-tRNA synthetase [EC:6.1.1.1]	E3
cromosoma	GONCAC_10745	<i>grxD</i>	COG0278	C	K07390	grxD, GLRX5; monothiol glutaredoxin	E2
cromosoma	GONCAC_10785	<i>ribE</i>	COG0307	H	K00793	ribE, RIB5; riboflavin synthase [EC:2.5.1.9]	E3
cromosoma	GONCAC_10825	<i>sufE</i>	COG2166	S	K02426	sufE; cysteine desulfuration protein SufE	E3
cromosoma	GONCAC_10830	-	COG0520	E	K11717	sufS; cysteine desulfurase / selenocysteine lyase [EC:2.8.1.7 4.4.1.16]	E3
cromosoma	GONCAC_10835	<i>sufD</i>	COG0719	O	K09015	sufD; Fe-S cluster assembly protein SufD	E3
cromosoma	GONCAC_10840	<i>sufC</i>	COG0396	O	K09013	sufC; Fe-S cluster assembly ATP-binding protein	E3
cromosoma	GONCAC_10845	<i>sufB</i>	COG0719	O	K09014	sufB; Fe-S cluster assembly protein SufB	E3
cromosoma	GONCAC_10850	<i>sufA</i>	COG0316	S	K05997	sufA; Fe-S cluster assembly protein SufA	E3
cromosoma	GONCAC_10925	-	COG0776	K	K04764	ihfA, himA; integration host factor subunit alpha	E3
cromosoma	GONCAC_10930	<i>pheT</i>	COG0072	J	K01890	FARSB, pheT; phenylalanyl-tRNA synthetase beta chain [EC:6.1.1.20]	E3
cromosoma	GONCAC_10935	<i>pheS</i>	COG0016	J	K01889	FARSA, pheS; phenylalanyl-tRNA synthetase alpha chain [EC:6.1.1.20]	E3
cromosoma	GONCAC_10945	<i>rplT</i>	COG0292	J	K02887	RP-L20, MRPL20, rplT; large subunit ribosomal protein L20	E3
cromosoma	GONCAC_10950	-	COG0291	J	K02916	RP-L35, MRPL35, rpml; large subunit ribosomal protein L35	E3
cromosoma	GONCAC_10955	<i>infC</i>	COG0290	J	K02520	infC, MTIF3; translation initiation factor IF-3	E3
cromosoma	GONCAC_10960	<i>thrS</i>	COG0441	J	K01868	TARS, thrS; threonyl-tRNA synthetase [EC:6.1.1.3]	E3
cromosoma	GONCAC_11025	<i>nadE</i>	COG0171	H	K01916	nadE; NAD+ synthase [EC:6.3.1.5]	E3
cromosoma	GONCAC_11230	<i>kdsA</i>	COG2877	M	K01627	kdsA; 2-dehydro-3-deoxyphosphooctonate aldolase (KDO 8-P synthase) [EC:2.5.1.55]	E3
cromosoma	GONCAC_11245	<i>prmC</i>	COG2890	J	K02493	hemK, prmC, HEMK; release factor glutamine methyltransferase [EC:2.1.1.297]	E3
cromosoma	GONCAC_11250	<i>prfA</i>	COG0216	J	K02835	prfA, MTRF1, MRF1; peptide chain release factor 1	E3

cromosoma	GONCAC_11255	<i>hemA</i>	COG0373	H	K02492	hemA; glutamyl-tRNA reductase [EC:1.2.1.70]	E3
cromosoma	GONCAC_11260	<i>lolB</i>	COG3017	M	K02494	lolB; outer membrane lipoprotein LolB	E3
cromosoma	GONCAC_11265	<i>ispE</i>	COG1947	I	K00919	ispE; 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase [EC:2.7.1.148]	E3
cromosoma	GONCAC_11270	<i>prs</i>	COG0462	F	K00948	PRPS, prsA; ribose-phosphate pyrophosphokinase [EC:2.7.6.1]	E3
cromosoma	GONCAC_11300	-	2E11Q	-			E3
cromosoma	GONCAC_11380	-		-			E2
cromosoma	GONCAC_11450	-	2EI78	-			E3
cromosoma	GONCAC_11470	-		-			E3
cromosoma	GONCAC_11485	-		-			E3
cromosoma	GONCAC_11600	-	2DB2V	-			E3
cromosoma	GONCAC_11875	<i>icd</i>	COG0538	C	K00031	IDH1, IDH2, icd; isocitrate dehydrogenase [EC:1.1.1.42]	E3
cromosoma	GONCAC_11890	<i>mnmA</i>	COG0482	J	K00566	mnmA, trmU; tRNA-uridine 2-sulfurtransferase [EC:2.8.1.13]	E3
cromosoma	GONCAC_11895	<i>hflD</i>	COG2915	S	K07153	hflD; high frequency lysogenization protein	E3
cromosoma	GONCAC_11900	<i>purB</i>	COG0015	F	K01756	purB, ADSL; adenylosuccinate lyase [EC:4.3.2.2]	E3
cromosoma	GONCAC_11930	<i>lolE</i>	COG4591	M	K09808	lolE; lipoprotein-releasing system permease protein	E3
cromosoma	GONCAC_11935	<i>lolD</i>	COG1136	V	K09810	lolD; lipoprotein-releasing system ATP-binding protein [EC:7.6.2.-]	E3
cromosoma	GONCAC_11940	<i>lolC</i>	COG4591	M	K09808	lolE; lipoprotein-releasing system permease protein	E3
cromosoma	GONCAC_11985	<i>cdil</i>	2DS9R	-			E3
cromosoma	GONCAC_11995	-	2BP8K	-			E3
cromosoma	GONCAC_12005	-	2AUNT	-			E3
cromosoma	GONCAC_12075	<i>holB</i>	COG0470	L	K02341	holB; DNA polymerase III subunit delta' [EC:2.7.7.7]	E3
cromosoma	GONCAC_12080	<i>tmk</i>	COG0125	F	K00943	tmk, DTYMK; dTMP kinase [EC:2.7.4.9]	E3
cromosoma	GONCAC_12085	<i>yceG</i>	COG1559	S	K07082	mltG; peptidoglycan lytic transglycosylase G [EC:4.2.2.29]	E3
cromosoma	GONCAC_12100	<i>acpP</i>	COG0236	IQ	K02078	acpP, acpM; acyl carrier protein	E2
cromosoma	GONCAC_12105	<i>fabG</i>	COG1028	IQ	K00059	fabG, OAR1; 3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]	E3
cromosoma	GONCAC_12110	-	COG0331	I	K00645	fabD, MCAT, MCT1; [acyl-carrier-protein] S-malonyltransferase [EC:2.3.1.39]	E3
cromosoma	GONCAC_12115	<i>fabH</i>	COG0332	I	K00648	fabH; 3-oxoacyl-[acyl-carrier-protein] synthase III [EC:2.3.1.180]	E3
cromosoma	GONCAC_12120	<i>plsX</i>	COG0416	I	K03621	plsX; phosphate acyltransferase [EC:2.3.1.274]	E3
cromosoma	GONCAC_12125	<i>rpmF</i>	COG0333	J	K02911	RP-L32, MRPL32, rpmF; large subunit ribosomal protein L32	E3

cromosoma	GONCAC_12130	<i>yceD</i>	COG1399	S	K07040	yceD, ylbN; DUF177 domain-containing protein	E3
cromosoma	GONCAC_12145	-	COG1530	J	K08300	rne; ribonuclease E [EC:3.1.26.12]	E3
cromosoma	GONCAC_12630	<i>tusE</i>	COG2920	P	K11179	tusE, dsrC; tRNA 2-thiouridine synthesizing protein E [EC:2.8.1.-]	E3
cromosoma	GONCAC_12690	<i>ompA</i>	COG2885	M	K03286	TC.OOP; OmpA-OmpF porin, OOP family	E3
cromosoma	GONCAC_12705	<i>fabA</i>	COG0764	I	K01716	fabA; 3-hydroxyacyl-[acyl-carrier protein] dehydratase / trans-2-decenoyl-[acyl-carrier protein] isomerase [EC:4.2.1.59 5.3.3.14]	E3
cromosoma	GONCAC_12800	<i>asnS</i>	COG0017	J	K01893	NARS, asnS; asparaginyl-tRNA synthetase [EC:6.1.1.22]	E3
cromosoma	GONCAC_12830	<i>mukB</i>	COG3096	D	K03632	mukB; chromosome partition protein MukB	E2
cromosoma	GONCAC_12835	<i>mukE</i>	COG3095	D	K03804	mukE; chromosome partition protein MukE	E3
cromosoma	GONCAC_12840	<i>mukF</i>	COG3006	D	K03633	mukF; chromosome partition protein MukF	E3
cromosoma	GONCAC_12875	<i>lpxK</i>	COG1663	M	K00912	lpxK; tetraacyldisaccharide 4'-kinase [EC:2.7.1.130]	E3
cromosoma	GONCAC_12880	<i>msbA</i>	COG1132	V	K11085	msbA; ATP-binding cassette, subfamily B, bacterial MsbA [EC:7.5.2.6]	E3
cromosoma	GONCAC_12895	<i>rpsA</i>	COG0539	J	K02945	RP-S1, rpsA; small subunit ribosomal protein S1	E3
cromosoma	GONCAC_12900	<i>cmk</i>	COG0283	F	K00945	cmk; CMP/dCMP kinase [EC:2.7.4.25]	E3
cromosoma	GONCAC_12905	<i>aroA</i>	COG0128	E	K00800	aroA; 3-phosphoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19]	E3
cromosoma	GONCAC_12940	<i>serS</i>	COG0172	J	K01875	SARS, serS; seryl-tRNA synthetase [EC:6.1.1.11]	E3
cromosoma	GONCAC_12950	<i>lolA</i>	COG2834	M	K03634	lolA; outer membrane lipoprotein carrier protein	E3
cromosoma	GONCAC_12955	-	COG1674	D	K03466	ftsK, spoIIIE; DNA segregation ATPase FtsK/SpoIIIE, S-DNA-T family	E3
cromosoma	GONCAC_12970	<i>cydD</i>	COG4988	V	K16013	cydD; ATP-binding cassette, subfamily C, bacterial CydD	E2
cromosoma	GONCAC_12985	<i>infA</i>	COG0361	J	K02518	infA; translation initiation factor IF-1	E3
cromosoma	GONCAC_13160	<i>cynT</i>	COG0288	P	K01673	cynT, can; carbonic anhydrase [EC:4.2.1.1]	E3
cromosoma	GONCAC_13705	<i>gpmA</i>	COG0588	G	K01834	PGAM, gpmA; 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase [EC:5.4.2.11]	E3
cromosoma	GONCAC_13780	<i>tolB</i>	COG0823	U	K03641	tolB; TolB protein	E2
cromosoma	GONCAC_13790	<i>tolR</i>	COG0848	U	K03560	tolR; biopolymer transport protein TolR	E2
cromosoma	GONCAC_13795	<i>tolQ</i>	COG0811	U	K03562	tolQ; biopolymer transport protein TolQ	E2
cromosoma	GONCAC_13825	<i>sucD</i>	COG0074	C	K01902	sucD; succinyl-CoA synthetase alpha subunit [EC:6.2.1.5]	E3
cromosoma	GONCAC_13830	<i>sucC</i>	COG0045	C	K01903	sucC; succinyl-CoA synthetase beta subunit [EC:6.2.1.5]	E3
cromosoma	GONCAC_13835	<i>odhB</i>	COG0508	C	K00658	DLST, sucB; 2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase) [EC:2.3.1.61]	E3

cromosoma	GONCAC_13840	<i>sucA</i>	COG0567	C	K00164	OGDH, sucA; 2-oxoglutarate dehydrogenase E1 component [EC:1.2.4.2]	E3
cromosoma	GONCAC_13855	<i>sdhD</i>	COG2142	C	K00242	sdhD, frdD; succinate dehydrogenase membrane anchor subunit	E3
cromosoma	GONCAC_13860	<i>sdhC</i>	COG2009	C	K00241	sdhC, frdC; succinate dehydrogenase cytochrome b subunit	E3
cromosoma	GONCAC_13935	<i>seqA</i>	COG3057	L	K03645	seqA; negative modulator of initiation of replication	E3
cromosoma	GONCAC_13965	<i>glnS</i>	COG0008	J	K01886	QARS, glnS; glutaminyl-tRNA synthetase [EC:6.1.1.18]	E3
cromosoma	GONCAC_13980	<i>nagA</i>	COG1820	G	K01443	nagA, AMDHD2; N-acetylglucosamine-6-phosphate deacetylase [EC:3.5.1.25]	E3
cromosoma	GONCAC_14035	<i>ubiF</i>	COG0654	CH	K03184	ubiF; 3-demethoxyubiquinol 3-hydroxylase [EC:1.14.99.60]	E3
cromosoma	GONCAC_14045	<i>phoH</i>	COG1702	T	K06217	phoH, phoL; phosphate starvation-inducible protein PhoH and related proteins	E3
cromosoma	GONCAC_14050	<i>ybeY</i>	COG0319	J	K07042	ybeY, yqfG; probable rRNA maturation factor	E3
cromosoma	GONCAC_14055	<i>corC</i>	COG4535	P	K06189	corC, tlyC; hemolysin (HlyC) family protein	E3
cromosoma	GONCAC_14060	<i>Int</i>	COG0815	M	K03820	Int; apolipoprotein N-acyltransferase [EC:2.3.1.269]	E3
cromosoma	GONCAC_14095	<i>leuS</i>	COG0495	J	K01869	LARS, leuS; leucyl-tRNA synthetase [EC:6.1.1.4]	E3
cromosoma	GONCAC_14110	<i>nadD</i>	COG1057	H	K00969	nadD; nicotinate-nucleotide adenyltransferase [EC:2.7.7.18]	E2
cromosoma	GONCAC_14115	<i>rsfS</i>	COG0799	J	K09710	ybeB; ribosome-associated protein	E3
cromosoma	GONCAC_14120	<i>rlmH</i>	COG1576	J	K00783	rlmH; 23S rRNA (pseudouridine1915-N3)-methyltransferase [EC:2.1.1.177]	E3
cromosoma	GONCAC_14125	<i>mrda</i>	COG0768	M	K05515	mrda; penicillin-binding protein 2 [EC:3.4.16.4]	E3
cromosoma	GONCAC_14130	<i>mrdb</i>	COG0772	M	K05837	rodA, mrdb; rod shape determining protein RodA	E3
cromosoma	GONCAC_14145	<i>ybeD</i>	COG2921	S	K09158	K09158; uncharacterized protein	E3
cromosoma	GONCAC_14155	<i>lipA</i>	COG0320	H	K03644	lipA, LIAS, LIP1, LIP5; lipoyl synthase [EC:2.8.1.8]	E3
cromosoma	GONCAC_14420	<i>folD</i>	COG0190	H	K01491	folD; methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase [EC:1.5.1.5 3.5.4.9]	E3
cromosoma	GONCAC_14435	<i>cysS</i>	COG0215	J	K01883	CARS, cysS; cysteinyl-tRNA synthetase [EC:6.1.1.16]	E3
cromosoma	GONCAC_14440	<i>ppiB</i>	COG0652	O	K03768	PPIB, ppiB; peptidyl-prolyl cis-trans isomerase B (cyclophilin B) [EC:5.2.1.8]	E3
cromosoma	GONCAC_14445	<i>lpxH</i>	COG2908	S	K03269	lpxH; UDP-2,3-diacetylglucosamine hydrolase [EC:3.6.1.54]	E3
cromosoma	GONCAC_14565	<i>adk</i>	COG0563	F	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	E3
cromosoma	GONCAC_14595	<i>dnaX</i>	COG2812	H	K02343	dnaX; DNA polymerase III subunit gamma/tau [EC:2.7.7.7]	E3
cromosoma	GONCAC_14765	-	COG0466	O	K01338	lon; ATP-dependent Lon protease [EC:3.4.21.53]	E3
cromosoma	GONCAC_14775	<i>clpX</i>	COG1219	O	K03544	clpX, CLPX; ATP-dependent Clp protease ATP-binding subunit ClpX	E3
cromosoma	GONCAC_14780	<i>clpP</i>	COG0740	O	K01358	clpP, CLPP; ATP-dependent Clp protease, protease subunit [EC:3.4.21.92]	E3

cromosoma	GONCAC_14860	<i>ispA</i>	COG0142	H	K00795	ispA; farnesyl diphosphate synthase [EC:2.5.1.1 2.5.1.10]	E3
cromosoma	GONCAC_14865	<i>dxs</i>	COG1154	H	K01662	dxs; 1-deoxy-D-xylulose-5-phosphate synthase [EC:2.2.1.7]	E3
cromosoma	GONCAC_14880	<i>thiL</i>	COG0611	H	K00946	thiL; thiamine-monophosphate kinase [EC:2.7.4.16]	E3
cromosoma	GONCAC_14885	<i>nusB</i>	COG0781	K	K03625	nusB; transcription antitermination protein NusB	E3
cromosoma	GONCAC_14890	<i>ribE</i>	COG0054	H	K00794	ribH, RIB4; 6,7-dimethyl-8-ribityllumazine synthase [EC:2.5.1.78]	E3
cromosoma	GONCAC_14895	<i>ribD</i>	COG0117	H	K11752	ribD; diamino-hydroxyphosphoribosylaminopyrimidine deaminase / 5-amino-6-(5-phosphoribosylamino)uracil reductase [EC:3.5.4.26 1.1.1.193]	E3
cromosoma	GONCAC_14900	<i>nrdR</i>	COG1327	K	K07738	nrdR; transcriptional repressor NrdR	E3
cromosoma	GONCAC_14915	<i>secF</i>	COG0341	U	K03074	secF; preprotein translocase subunit SecF	E3
cromosoma	GONCAC_14920	<i>secD</i>	COG0342	U	K03072	secD; preprotein translocase subunit SecD	E3
cromosoma	GONCAC_14925	<i>tgt</i>	COG1862	U	K03210	yajC; preprotein translocase subunit YajC	E3
cromosoma	GONCAC_15085	<i>proC</i>	COG0345	E	K00286	proC; pyrroline-5-carboxylate reductase [EC:1.5.1.2]	E3
cromosoma	GONCAC_15250	<i>proA</i>	COG0014	E	K00147	proA; glutamate-5-semialdehyde dehydrogenase [EC:1.2.1.41]	E3
cromosoma	GONCAC_15255	<i>proB</i>	COG0263	E	K00931	proB; glutamate 5-kinase [EC:2.7.2.11]	E3
cromosoma	GONCAC_15315	<i>lpcA</i>	COG0279	G	K03271	gmhA, lpcA; D-sedoheptulose 7-phosphate isomerase [EC:5.3.1.28]	E3
cromosoma	GONCAC_15460	<i>dnaQ</i>	COG0847	L	K02342	dnaQ; DNA polymerase III subunit epsilon [EC:2.7.7.7]	E3
cromosoma	GONCAC_15465	<i>rnhA</i>	COG0328	L	K03469	rnhA, RNASEH1; ribonuclease HI [EC:3.1.26.4]	E3
cromosoma	GONCAC_15525	<i>gmhB</i>	COG0241	E	K03273	gmhB; D-glycero-D-manno-heptose 1,7-bisphosphate phosphatase [EC:3.1.3.82 3.1.3.83]	E3
cromosoma	GONCAC_15555	<i>proS</i>	COG0442	J	K01881	PARS, proS; prolyl-tRNA synthetase [EC:6.1.1.15]	E3
cromosoma	GONCAC_15580	<i>tilS</i>	COG0037	D	K04075	tilS, mesJ; tRNA(Ile)-lysine synthase [EC:6.3.4.19]	E3
cromosoma	GONCAC_15595	<i>accA</i>	COG0825	I	K01962	accA; acetyl-CoA carboxylase carboxyl transferase subunit alpha [EC:6.4.1.2 2.1.3.15]	E3
cromosoma	GONCAC_15600	<i>dnaE</i>	COG0587	L	K02337	dnaE; DNA polymerase III subunit alpha [EC:2.7.7.7]	E3
cromosoma	GONCAC_15605	<i>rnhB</i>	COG0164	L	K03470	rnhB; ribonuclease HII [EC:3.1.26.4]	E3
cromosoma	GONCAC_15610	<i>lpxB</i>	COG0763	M	K00748	lpxB; lipid-A-disaccharide synthase [EC:2.4.1.182]	E3
cromosoma	GONCAC_15615	<i>lpxA</i>	COG1043	M	K00677	lpxA; UDP-N-acetylglucosamine acyltransferase [EC:2.3.1.129]	E3
cromosoma	GONCAC_15620	<i>fabZ</i>	COG0764	I	K02372	fabZ; 3-hydroxyacyl-[acyl-carrier-protein] dehydratase [EC:4.2.1.59]	E3
cromosoma	GONCAC_15625	<i>lpxD</i>	COG1044	M	K02536	lpxD; UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase [EC:2.3.1.191]	E3
cromosoma	GONCAC_15630	<i>skp</i>	COG2825	M	K06142	hlpA, ompH; outer membrane protein	E3

cromosoma	GONCAC_15635	<i>bamA</i>	COG4775	M	K07277	SAM50, TOB55, bamA; outer membrane protein insertion porin family	E3
cromosoma	GONCAC_15640	<i>rseP</i>	COG0750	M	K11749	rseP; regulator of sigma E protease [EC:3.4.24.-]	E3
cromosoma	GONCAC_15645	<i>cdsA</i>	COG0575	I	K00981	E2.7.7.41, CDS1, CDS2, cdsA; phosphatidate cytidyltransferase [EC:2.7.7.41]	E3
cromosoma	GONCAC_15650	<i>ispU</i>	COG0020	H	K00806	uppS; undecaprenyl diphosphate synthase [EC:2.5.1.31]	E3
cromosoma	GONCAC_15655	<i>ispC</i>	COG0743	I	K00099	dxr; 1-deoxy-D-xylulose-5-phosphate reductoisomerase [EC:1.1.1.267]	E3
cromosoma	GONCAC_15660	<i>frr</i>	COG0233	J	K02838	frr, MRRF, RRF; ribosome recycling factor	E3
cromosoma	GONCAC_15665	<i>pyrH</i>	COG0528	F	K09903	pyrH; uridylate kinase [EC:2.7.4.22]	E3
cromosoma	GONCAC_15670	<i>tsf</i>	COG0264	J	K02357	tsf, TSFM; elongation factor Ts	E3
cromosoma	GONCAC_15675	<i>rpsB</i>	COG0052	J	K02967	RP-S2, MRPS2, rpsB; small subunit ribosomal protein S2	E3
cromosoma	GONCAC_15680	<i>t44</i>		ncRNA			E3
cromosoma	GONCAC_15695	<i>dapD</i>	COG2171	E	K00674	dapD; 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase [EC:2.3.1.117]	E2
cromosoma	GONCAC_15720	<i>mtnN</i>	COG0775	E	K01243	mtnN, mtn, pfs; adenosylhomocysteine nucleosidase [EC:3.2.2.9]	E3
cromosoma	GONCAC_15735	<i>erpA</i>	COG0316	C	K15724	erpA; iron-sulfur cluster insertion protein	E3
cromosoma	GONCAC_15740	<i>hemL</i>	COG0001	H	K01845	hemL; glutamate-1-semialdehyde 2,1-aminomutase [EC:5.4.3.8]	E3
cromosoma	GONCAC_15795	<i>pcnB</i>	COG0617	H	K00970	pcnB; poly(A) polymerase [EC:2.7.7.19]	E3
cromosoma	GONCAC_15800	<i>folK</i>	COG0801	H	K00950	folK; 2-amino-4-hydroxy-6-hydroxymethylidihydropteridine diphosphokinase [EC:2.7.6.3]	E3
cromosoma	GONCAC_15915	<i>acnB</i>	COG1049	C	K01682	acnB; aconitate hydratase 2 / 2-methylisocitrate dehydratase [EC:4.2.1.3 4.2.1.99]	E3
cromosoma	GONCAC_15940	<i>lpdA</i>	COG1249	C	K00382	DLD, lpd, pdhD; dihydrolipoyl dehydrogenase [EC:1.8.1.4]	E3
cromosoma	GONCAC_16010	-	COG0237	H	K00859	coaE; dephospho-CoA kinase [EC:2.7.1.24]	E2
cromosoma	GONCAC_16030	<i>secA</i>	COG0653	U	K03070	secA; preprotein translocase subunit SecA [EC:7.4.2.8]	E3
cromosoma	GONCAC_16035	<i>secM</i>	2ATX7	S	K13301	secM; secretion monitor	E3
cromosoma	GONCAC_16045	<i>ftsZ</i>	COG0206	D	K03531	ftsZ; cell division protein FtsZ	E3
cromosoma	GONCAC_16050	<i>ftsA</i>	COG0849	D	K03590	ftsA; cell division protein FtsA	E3
cromosoma	GONCAC_16055	<i>ftsQ</i>	COG1589	D	K03589	ftsQ; cell division protein FtsQ	E3
cromosoma	GONCAC_16060	<i>ddl</i>	COG1181	F	K01921	ddl; D-alanine-D-alanine ligase [EC:6.3.2.4]	E3
cromosoma	GONCAC_16065	<i>murC</i>	COG0773	M	K01924	murC; UDP-N-acetylmuramate--alanine ligase [EC:6.3.2.8]	E3
cromosoma	GONCAC_16070	<i>murG</i>	COG0707	M	K02563	murG; UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase [EC:2.4.1.227]	E3

cromosoma	GONCAC_16075	<i>ftsW</i>	COG0772	D	K03588	ftsW, spoVE; cell division protein FtsW	E3
cromosoma	GONCAC_16080	<i>murD</i>	COG0771	M	K01925	murD; UDP-N-acetylmuramoylalanine--D-glutamate ligase [EC:6.3.2.9]	E3
cromosoma	GONCAC_16085	<i>mraY</i>	COG0472	M	K01000	mraY; phospho-N-acetylmuramoyl-pentapeptide-transferase [EC:2.7.8.13]	E3
cromosoma	GONCAC_16090	<i>murF</i>	COG0770	M	K01929	murF; UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase [EC:6.3.2.10]	E3
cromosoma	GONCAC_16095	<i>murE</i>	COG0769	M	K01928	murE; UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase [EC:6.3.2.13]	E3
cromosoma	GONCAC_16100	<i>ftsI</i>	COG0768	M	K03587	ftsI; cell division protein FtsI (penicillin-binding protein 3) [EC:3.4.16.4]	E3
cromosoma	GONCAC_16105	<i>ftsL</i>	COG3116	D	K03586	ftsL; cell division protein FtsL	E3
cromosoma	GONCAC_16110	<i>rsmH</i>	COG0275	J	K03438	mraW, rsmH; 16S rRNA (cytosine1402-N4)-methyltransferase [EC:2.1.1.199]	E2
cromosoma	GONCAC_16215	<i>lptD</i>	COG1452	M	K04744	lptD, imp, ostA; LPS-assembly protein	E3
cromosoma	GONCAC_16220	<i>surA</i>	COG0760	M	K03771	surA; peptidyl-prolyl cis-trans isomerase SurA [EC:5.2.1.8]	E3
cromosoma	GONCAC_16225	<i>pdxA</i>	COG1995	H	K00097	pdxA; 4-hydroxythreonine-4-phosphate dehydrogenase [EC:1.1.1.262]	E3
cromosoma	GONCAC_16230	<i>rsmA</i>	COG0030	J	K02528	ksgA; 16S rRNA (adenine1518-N6/adenine1519-N6)-dimethyltransferase [EC:2.1.1.182]	E3
cromosoma	GONCAC_16250	<i>folA</i>	COG0262	H	K00287	DHFR, folA; dihydrofolate reductase [EC:1.5.1.3]	E3
cromosoma	GONCAC_16270	<i>dapB</i>	COG0289	E	K00215	dapB; 4-hydroxy-tetrahydronicotinate reductase [EC:1.17.1.8]	E3
cromosoma	GONCAC_16275	<i>ispH</i>	COG0761	IM	K03527	ispH, lytB; 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase [EC:1.17.7.4]	E3
cromosoma	GONCAC_16280	<i>fkpB</i>	COG1047	G	K03774	slpA; FKBP-type peptidyl-prolyl cis-trans isomerase SlpA [EC:5.2.1.8]	E3
cromosoma	GONCAC_16285	<i>lspA</i>	COG0597	M	K03101	lspA; signal peptidase II [EC:3.4.23.36]	E3
cromosoma	GONCAC_16290	<i>ileS</i>	COG0060	J	K01870	IARS, ileS; isoleucyl-tRNA synthetase [EC:6.1.1.5]	E3
cromosoma	GONCAC_16295	<i>ribF</i>	COG0196	H	K11753	ribF; riboflavin kinase / FMN adenylyltransferase [EC:2.7.1.26 2.7.7.2]	E3
cromosoma	GONCAC_16300	<i>rpsT</i>	COG0268	J	K02968	RP-S20, rpsT; small subunit ribosomal protein S20	E3
cromosoma	GONCAC_16310	<i>nhaA</i>	COG3004	P	K03313	nhaA; Na+:H+ antiporter, NhaA family	E3
cromosoma	GONCAC_16315	<i>dnaJ</i>	COG0484	O	K03686	dnaJ; molecular chaperone DnaJ	E3
cromosoma	GONCAC_16325	<i>dnaK</i>	COG0443	O	K04043	dnaK, HSPA9; molecular chaperone DnaK	E3
cromosoma	GONCAC_16380	<i>arcA</i>	COG0745	K	K07773	arcA; two-component system, OmpR family, aerobic respiration control protein ArcA	E3
cromosoma	GONCAC_16840	<i>lptG</i>	COG0795	S	K11720	lptG; lipopolysaccharide export system permease protein	E3
cromosoma	GONCAC_16845	<i>lptF</i>	COG0795	S	K07091	lptF; lipopolysaccharide export system permease protein	E3
cromosoma	GONCAC_16860	<i>holC</i>	COG2927	L	K02339	holC; DNA polymerase III subunit chi [EC:2.7.7.7]	E3

cromosoma	GONCAC_16865	<i>valS</i>	COG0525	J	K01873	VARs, valS; valyl-tRNA synthetase [EC:6.1.1.9]	E3
cromosoma	GONCAC_17040	-	COG0513	F	K05592	deaD, cshA; ATP-dependent RNA helicase DeaD [EC:5.6.2.7]	E3
cromosoma	GONCAC_17065	<i>rpsO</i>	COG0184	J	K02956	RP-S15, MRPS15, rpsO; small subunit ribosomal protein S15	E3
cromosoma	GONCAC_17075	<i>rbfA</i>	COG0858	J	K02834	rbfA; ribosome-binding factor A	E3
cromosoma	GONCAC_17080	<i>infB</i>	COG0532	J	K02519	infB, MTIF2; translation initiation factor IF-2	E3
cromosoma	GONCAC_17085	<i>nusA</i>	COG0195	K	K02600	nusA; transcription termination/antitermination protein NusA	E3
cromosoma	GONCAC_17090	<i>rimP</i>	COG0779	S	K09748	rimP; ribosome maturation factor RimP	E3
cromosoma	GONCAC_17105	<i>secG</i>	COG1314	U	K03075	secG; preprotein translocase subunit SecG	E3
cromosoma	GONCAC_17110	<i>glmM</i>	COG1109	G	K03431	glmM; phosphoglucosamine mutase [EC:5.4.2.10]	E3
cromosoma	GONCAC_17115	<i>folP</i>	COG0294	H	K00796	folP; dihydropteroate synthase [EC:2.5.1.15]	E3
cromosoma	GONCAC_17120	<i>ftsH</i>	COG0465	O	K03798	ftsH, hflB; cell division protease FtsH [EC:3.4.24.-]	E3
cromosoma	GONCAC_17125	<i>rlmE</i>	COG0293	J	K02427	rlmE, rrmJ, ftsJ; 23S rRNA (uridine2552-2'-O)-methyltransferase [EC:2.1.1.166]	E3
cromosoma	GONCAC_17215	<i>cgtA</i>	COG0536	S	K03979	obgE, cgtA, MTG2; GTPase [EC:3.6.5.-]	E3
cromosoma	GONCAC_17220	<i>rpmA</i>	COG0211	J	K02899	RP-L27, MRPL27, rpmA; large subunit ribosomal protein L27	E3
cromosoma	GONCAC_17225	<i>rplU</i>	COG0261	J	K02888	RP-L21, MRPL21, rplU; large subunit ribosomal protein L21	E3
cromosoma	GONCAC_17230	<i>ispB</i>	COG0142	H	K02523	ispB; octaprenyl-diphosphate synthase [EC:2.5.1.90]	E3
cromosoma	GONCAC_17240	<i>murA</i>	COG0766	M	K00790	murA; UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC:2.5.1.7]	E3
cromosoma	GONCAC_17245	<i>ibaG</i>	COG5007	K			E3
cromosoma	GONCAC_17280	<i>kdsD</i>	COG0517	M	K06041	kdsD, kpsF; arabinose-5-phosphate isomerase [EC:5.3.1.13]	E3
cromosoma	GONCAC_17290	<i>lptC</i>	COG3117	S	K11719	lptC; lipopolysaccharide export system protein LptC	E3
cromosoma	GONCAC_17295	<i>lptA</i>	COG1934	S	K09774	lptA; lipopolysaccharide export system protein LptA	E3
cromosoma	GONCAC_17300	<i>lptB</i>	COG1137	S	K06861	lptB; lipopolysaccharide export system ATP-binding protein [EC:7.5.2.5]	E3
cromosoma	GONCAC_17335	-	COG0642	T	K07648	arcB; two-component system, OmpR family, aerobic respiration control sensor histidine kinase ArcB [EC:2.7.13.3]	E2
cromosoma	GONCAC_17360	<i>rpsI</i>	COG0103	J	K02996	RP-S9, MRPS9, rpsI; small subunit ribosomal protein S9	E3
cromosoma	GONCAC_17365	<i>rplM</i>	COG0102	J	K02871	RP-L13, MRPL13, rplM; large subunit ribosomal protein L13	E3
cromosoma	GONCAC_17385	<i>degS</i>	COG0265	O	K04691	hhoB, degS; serine protease DegS [EC:3.4.21.-]	E3
cromosoma	GONCAC_17390	<i>epmA</i>	COG2269	J	K04568	epmA, poxA; elongation factor P--(R)-beta-lysine ligase [EC:6.3.1.-]	E3
cromosoma	GONCAC_17410	<i>efp</i>	COG0231	J	K02356	efp; elongation factor P	E3

cromosoma	GONCAC_17415	<i>epmB</i>	COG1509	C	K19810	epmB; L-lysine 2,3-aminomutase [EC:5.4.3.-]	E3
cromosoma	GONCAC_17430	<i>groL</i>	COG0459	O	K04077	groEL, HSPD1; chaperonin GroEL [EC:5.6.1.7]	E3
cromosoma	GONCAC_17435	<i>groES</i>	COG0234	O	K04078	groES, HSPE1; chaperonin GroES	E3
cromosoma	GONCAC_17555	<i>nagC</i>	COG1522	GK			E3
cromosoma	GONCAC_17560	-	COG2971	G	K18676	gspK; glucosamine kinase [EC:2.7.1.8]	E3
cromosoma	GONCAC_17660	-	COG2944	K	K07726	K07726; putative transcriptional regulator	E2
cromosoma	GONCAC_17695	-	COG3620	K			E2
cromosoma	GONCAC_17890	<i>dnaB</i>	COG0305	L	K02314	dnaB; replicative DNA helicase [EC:5.6.2.3]	E3
cromosoma	GONCAC_17940	<i>plsB</i>	COG2937	I	K00631	plsB; glycerol-3-phosphate O-acyltransferase [EC:2.3.1.15]	E3
cromosoma	GONCAC_17945	<i>ubiA</i>	COG0382	H	K03179	ubiA; 4-hydroxybenzoate polyprenyltransferase [EC:2.5.1.39]	E2
cromosoma	GONCAC_17950	<i>ubiC</i>	COG3161	H	K03181	ubiC; chorismate lyase [EC:4.1.3.40]	E2
cromosoma	GONCAC_18120	<i>hemE</i>	COG0407	H	K01599	hemE, UROD; uroporphyrinogen decarboxylase [EC:4.1.1.37]	E3
cromosoma	GONCAC_18155	<i>rpoC</i>	COG0086	K	K03046	rpoC; DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6]	E3
cromosoma	GONCAC_18160	<i>rpoB</i>	COG0085	K	K03043	rpoB; DNA-directed RNA polymerase subunit beta [EC:2.7.7.6]	E3
cromosoma	GONCAC_18170	<i>rplL</i>	COG0222	J	K02935	RP-L7, MRPL12, rplL; large subunit ribosomal protein L7/L12	E3
cromosoma	GONCAC_18175	<i>rplJ</i>	COG0244	J	K02864	RP-L10, MRPL10, rplJ; large subunit ribosomal protein L10	E3
cromosoma	GONCAC_18180	<i>rplA</i>	COG0081	J	K02863	RP-L1, MRPL1, rplA; large subunit ribosomal protein L1	E3
cromosoma	GONCAC_18185	<i>rplK</i>	COG0080	J	K02867	RP-L11, MRPL11, rplK; large subunit ribosomal protein L11	E3
cromosoma	GONCAC_18190	<i>nusG</i>	COG0250	K	K02601	nusG; transcription termination/antitermination protein NusG	E3
cromosoma	GONCAC_18195	<i>secE</i>	COG0690	U	K03073	secE; preprotein translocase subunit SecE	E3
cromosoma	GONCAC_18225	<i>coaA</i>	COG1072	H	K00867	coaA; type I pantothenate kinase [EC:2.7.1.33]	E3
cromosoma	GONCAC_18230	<i>birA</i>	COG0340	HK	K03524	birA; BirA family transcriptional regulator, biotin operon repressor / biotin---[acetyl-CoA-carboxylase] ligase [EC:6.3.4.15]	E3
cromosoma	GONCAC_18235	<i>murB</i>	COG0812	M	K00075	murB; UDP-N-acetylmuramate dehydrogenase [EC:1.3.1.98]	E3
cromosoma	GONCAC_18305	<i>ubiD</i>	COG0043	H	K03182	ubiD; 4-hydroxy-3-polyprenylbenzoate decarboxylase [EC:4.1.1.98]	E3
cromosoma	GONCAC_18340	<i>ubiB</i>	COG0661	H	K03688	ubiB, aarF; ubiquinone biosynthesis protein	E3
cromosoma	GONCAC_18345	-	COG3165	S	K03690	ubiJ; ubiquinone biosynthesis accessory factor UbiJ	E3
cromosoma	GONCAC_18350	<i>ubiE</i>	COG0500	H	K03183	ubiE; demethylmenaquinone methyltransferase / 2-methoxy-6-polyprenyl-1,4-benzoquinol methylase [EC:2.1.1.163 2.1.1.201]	E3
cromosoma	GONCAC_18470	<i>xerC</i>	COG4973	D	K03733	xerC; integrase/recombinase XerC	E3

cromosoma	GONCAC_18475	-	COG3159	S	K09921	K09921; uncharacterized protein	E3
cromosoma	GONCAC_18485	-	COG5567	N			E3
cromosoma	GONCAC_18490	<i>cyaA</i>	COG3072	F	K05851	<i>cyaA</i> ; adenylate cyclase, class 1 [EC:4.6.1.1]	E3
cromosoma	GONCAC_18495	<i>hemC</i>	COG0181	H	K01749	<i>hemC</i> , HMBS; hydroxymethylbilane synthase [EC:2.5.1.61]	E3
cromosoma	GONCAC_18500	<i>hemD</i>	COG1587	H	K01719	<i>hemD</i> , UROS; uroporphyrinogen-III synthase [EC:4.2.1.75]	E3
cromosoma	GONCAC_18605	<i>rho</i>	COG1158	K	K03628	<i>rho</i> ; transcription termination factor Rho	E3
cromosoma	GONCAC_18625	<i>rep</i>	COG0210	L	K03656	<i>rep</i> ; ATP-dependent DNA helicase Rep [EC:5.6.2.4]	E3
cromosoma	GONCAC_18830	<i>sodA</i>	COG0605	C	K04564	SOD2; superoxide dismutase, Fe-Mn family [EC:1.15.1.1]	E3
cromosoma	GONCAC_18885	-	2EKCH	-			E2
cromosoma	GONCAC_18900	<i>glyS</i>	COG0751	J	K01879	<i>glyS</i> ; glycyl-tRNA synthetase beta chain [EC:6.1.1.14]	E3
cromosoma	GONCAC_18905	<i>glyQ</i>	COG0752	J	K01878	<i>glyQ</i> ; glycyl-tRNA synthetase alpha chain [EC:6.1.1.14]	E3
cromosoma	GONCAC_18940	<i>mnmE</i>	COG0486	J	K03650	<i>mnmE</i> , <i>trmE</i> , MSS1; tRNA modification GTPase [EC:3.6.-.-]	E3
cromosoma	GONCAC_18945	<i>yidC</i>	COG0706	U	K03217	<i>yidC</i> , <i>spoIIIJ</i> , OXA1, <i>ccfA</i> ; YidC/Oxa1 family membrane protein insertase	E2
cromosoma	GONCAC_18955	<i>rnpA</i>	COG0594	J	K03536	<i>rnpA</i> ; ribonuclease P protein component [EC:3.1.26.5]	E2
cromosoma	GONCAC_18960	<i>dnaA</i>	COG0593	L	K02313	<i>dnaA</i> ; chromosomal replication initiator protein	E3
cromosoma	GONCAC_18970	<i>dnaN</i>	COG0592	L	K02338	<i>dnaN</i> ; DNA polymerase III subunit beta [EC:2.7.7.7]	E3
cromosoma	GONCAC_18980	<i>gyrB</i>	COG0187	L	K02470	<i>gyrB</i> ; DNA gyrase subunit B [EC:5.6.2.2]	E3
cromosoma	GONCAC_19105	<i>dnaC</i>	COG1484	L	K02315	<i>dnaC</i> ; DNA replication protein DnaC	E3
cromosoma	GONCAC_19175	<i>glmS</i>	COG0449	M	K00820	<i>glmS</i> , GFPT; glutamine---fructose-6-phosphate transaminase (isomerizing) [EC:2.6.1.16]	E3
cromosoma	GONCAC_19180	<i>glmU</i>	COG1207	M	K04042	<i>glmU</i> ; bifunctional UDP-N-acetylglucosamine pyrophosphorylase / glucosamine-1-phosphate N-acetyltransferase [EC:2.7.7.23 2.3.1.157]	E3
cromosoma	GONCAC_19185	<i>atpC</i>	COG0355	C	K02114	ATPF1E, <i>atpC</i> ; F-type H ⁺ -transporting ATPase subunit epsilon	E3
cromosoma	GONCAC_19190	<i>atpD</i>	COG0055	C	K02112	ATPF1B, <i>atpD</i> ; F-type H ⁺ /Na ⁺ -transporting ATPase subunit beta [EC:7.1.2.2 7.2.2.1]	E3
cromosoma	GONCAC_19195	<i>atpG</i>	COG0224	C	K02115	ATPF1G, <i>atpG</i> ; F-type H ⁺ -transporting ATPase subunit gamma	E3
cromosoma	GONCAC_19200	<i>atpA</i>	COG0056	C	K02111	ATPF1A, <i>atpA</i> ; F-type H ⁺ /Na ⁺ -transporting ATPase subunit alpha [EC:7.1.2.2 7.2.2.1]	E3
cromosoma	GONCAC_19205	<i>atpH</i>	COG0712	C	K02113	ATPF1D, <i>atpH</i> ; F-type H ⁺ -transporting ATPase subunit delta	E3
cromosoma	GONCAC_19210	<i>atpF</i>	COG0711	C	K02109	ATPF0B, <i>atpF</i> ; F-type H ⁺ -transporting ATPase subunit b	E3
cromosoma	GONCAC_19220	<i>atpB</i>	COG0356	C	K02108	ATPF0A, <i>atpB</i> ; F-type H ⁺ -transporting ATPase subunit a	E3

cromosoma	GONCAC_19225	<i>atpI</i>	COG3312	C	K02116	atpI; ATP synthase protein I	E3
cromosoma	GONCAC_19235	<i>mnmG</i>	COG0445	D	K03495	gidA, mnmG, MTO1; tRNA uridine 5-carboxymethylaminomethyl modification enzyme	E3
cromosoma	GONCAC_19240	<i>mioC</i>	COG0716	C	K06205	mioC; MioC protein	E3
pPagLPU12_1	GONCAC_19370	-	COG1475	K	K03497	parB, spo0J; ParB family transcriptional regulator, chromosome partitioning protein	E3
pPagLPU12_1	GONCAC_19375	-	COG1192	D	K03496	parA, soj; chromosome partitioning protein	E3
pPagLPU12_1	GONCAC_19380	-	28ICS	S			E3
pPagLPU12_1	GONCAC_19830	-	2ATCS	-			E2
pPagLPU12_2	GONCAC_21015	-	2EQZ7	-			E3
pPagLPU12_2	GONCAC_21190	-	COG0702	GM			E2
pPagLPU12_2	GONCAC_21380	<i>parA</i>	COG1192	D	K03496	parA, soj; chromosome partitioning protein	E3
pPagLPU12_2	GONCAC_21385	-	COG1475	K	K03497	parB, spo0J; ParB family transcriptional regulator, chromosome partitioning protein	E3
pPagLPU12_2	GONCAC_21495	-	COG1942	S			E3
pPagLPU12_2	GONCAC_21590	-	2AUUB	-			E2
pPagLPU12_2	GONCAC_21825	-	COG0600	P	K15599	thiX; putative hydroxymethylpyrimidine transport system permease protein	E2
pPagLPU12_3	GONCAC_22715	-	COG3121	NU			E3
pPagLPU12_3	GONCAC_22760	<i>ecoRIIM</i>	COG0270	H	K00558	DNMT1, dcm; DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37]	E3
pPagLPU12_3	GONCAC_22780	-	28ICS	S			E3
pPagLPU12_3	GONCAC_22990			ncRNA	ncRNA		E2

*El software *Transit* (DeJesus et al., 2015) identifica genes esenciales utilizando tres métodos: “Gumbel”, “HMM (Hidden Markov Model)” y “TTN (Transposon TnSeq Network)”. Se indica el criterio con el que fueron detectados los genes esenciales según el número de métodos que los identificaron, etiquetados como (E2) para aquellos genes identificados por dos de los métodos y (E3) para aquellos identificados por los tres métodos. Para más detalle ver Capítulo VI, Sección VI.2.a.

Tabla VI.S2. Genes de auxotrofia y relacionados en *P. agglomerans* LPU12 identificados en los ensayos TnSeq (Capítulo VI).

Replicón	locus_tag	Gen (nombre)	COG	categoría COG	KO	Definición asociada al KO	M
cromosoma	GONCAC_06540	<i>thiM</i>	COG2145	H	K00878	<i>thiM</i> ; hydroxyethylthiazole kinase [EC:2.7.1.50]	-4,81
cromosoma	GONCAC_06820	<i>hisA</i>	COG0106	E	K01814	<i>hisA</i> ; phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase [EC:5.3.1.16]	-4,63
cromosoma	GONCAC_08835	<i>trpA</i>	COG0159	E	K01695	<i>trpA</i> ; tryptophan synthase alpha chain [EC:4.2.1.20]	-4,43
cromosoma	GONCAC_18060	<i>metA</i>	COG1897	E	K00651	<i>metA</i> ; homoserine O-succinyltransferase/O-acetyltransferase [EC:2.3.1.46 2.3.1.31]	-4,3
cromosoma	GONCAC_00355	<i>secB</i>	COG1952	U	K03071	<i>secB</i> ; preprotein translocase subunit SecB	-4,29
pPagLPU12_2	GONCAC_20640	-	COG2022	H	K03149	<i>thiG</i> ; thiazole synthase [EC:2.8.1.10]	-4,15
cromosoma	GONCAC_00350	<i>grxC</i>	COG0695	O	K03676	<i>grxC</i> , GLRX, GLRX2; glutaredoxin 3	-4,09
cromosoma	GONCAC_16500	<i>serB</i>	COG0560	E	K01079	<i>serB</i> , PSPH; phosphoserine phosphatase [EC:3.1.3.3]	-4,03
cromosoma	GONCAC_03860	<i>cysC</i>	COG0529	F	K00860	<i>cysC</i> ; adenylylsulfate kinase [EC:2.7.1.25]	-3,91
cromosoma	GONCAC_16265	<i>carA</i>	COG0505	F	K01956	<i>carA</i> , CPA1; carbamoyl-phosphate synthase small subunit [EC:6.3.5.5]	-3,9
cromosoma	GONCAC_16360	<i>thrB</i>	COG0083	E	K00872	<i>thrB</i> ; homoserine kinase [EC:2.7.1.39]	-3,84
cromosoma	GONCAC_03385	-	COG0212	H	K01934	MTHFS; 5-formyltetrahydrofolate cyclo-ligase [EC:6.3.3.2]	-3,83
cromosoma	GONCAC_05105	<i>purC</i>	COG0152	F	K01923	<i>purC</i> ; phosphoribosylaminoimidazole-succinocarboxamide synthase [EC:6.3.2.6]	-3,83
cromosoma	GONCAC_05945	<i>pdxB</i>	COG0111	H	K03473	<i>pdxB</i> ; erythronate-4-phosphate dehydrogenase [EC:1.1.1.290]	-3,74
cromosoma	GONCAC_05985	<i>purF</i>	COG0034	F	K00764	<i>purF</i> , PPAT; amidophosphoribosyltransferase [EC:2.4.2.14]	-3,72
cromosoma	GONCAC_16355	<i>thrC</i>	COG0498	E	K01733	<i>thrC</i> ; threonine synthase [EC:4.2.3.1]	-3,72
cromosoma	GONCAC_08940	<i>pyrF</i>	COG0284	F	K01591	<i>pyrF</i> ; orotidine-5'-phosphate decarboxylase [EC:4.1.1.23]	-3,7
cromosoma	GONCAC_06825	<i>hisH</i>	COG0118	E	K02501	<i>hisH</i> ; imidazole glycerol-phosphate synthase subunit HisH [EC:4.3.2.10]	-3,69
cromosoma	GONCAC_00345	<i>pspE</i>	COG0607	P			-3,61
cromosoma	GONCAC_03825	<i>cysI</i>	COG0155	P	K00381	<i>cysI</i> ; sulfite reductase (NADPH) hemoprotein beta-component [EC:1.8.1.2]	-3,53
cromosoma	GONCAC_04980	<i>ppk1</i>	COG0855	H	K00937	<i>ppk1</i> ; polyphosphate kinase [EC:2.7.4.1]	-3,53
cromosoma	GONCAC_13865	<i>gltA</i>	COG0372	C	K01647	CS, <i>gltA</i> ; citrate synthase [EC:2.3.3.1]	-3,52
cromosoma	GONCAC_00535	-	COG0548	E	K00930	<i>argB</i> ; acetylglutamate kinase [EC:2.7.2.8]	-3,49
cromosoma	GONCAC_16150	<i>leuC</i>	COG0065	H	K01703	<i>leuC</i> , IPMI-L; 3-isopropylmalate/(R)-2-methylmalate dehydratase large subunit [EC:4.2.1.33 4.2.1.35]	-3,49
cromosoma	GONCAC_16900	<i>pyrB</i>	COG0540	F	K00609	<i>pyrB</i> , PYR2; aspartate carbamoyltransferase catalytic subunit [EC:2.1.3.2]	-3,47
cromosoma	GONCAC_14455	<i>purK</i>	COG0026	F	K01589	<i>purK</i> ; 5-(carboxyamino)imidazole ribonucleotide synthase [EC:6.3.4.18]	-3,45
cromosoma	GONCAC_05005	<i>purM</i>	COG0150	F	K01933	<i>purM</i> ; phosphoribosylformylglycinamide cyclo-ligase [EC:6.3.3.1]	-3,44
cromosoma	GONCAC_08840	<i>trpB</i>	COG0133	E	K01696	<i>trpB</i> ; tryptophan synthase beta chain [EC:4.2.1.20]	-3,42
cromosoma	GONCAC_16260	<i>carB</i>	COG0458	F	K01955	<i>carB</i> , CPA2; carbamoyl-phosphate synthase large subunit [EC:6.3.5.5]	-3,4

cromosoma	GONCAC_00655	<i>ilvE</i>	COG0115	E	K00826	E2.6.1.42, <i>ilvE</i> ; branched-chain amino acid aminotransferase [EC:2.6.1.42]	-3,38
cromosoma	GONCAC_16155	<i>leuD</i>	COG0066	E	K01704	<i>leuD</i> , IPMI-S; 3-isopropylmalate/(R)-2-methylmalate dehydratase small subunit [EC:4.2.1.33 4.2.1.35]	-3,37
cromosoma	GONCAC_12910	<i>serC</i>	COG1932	E	K00831	<i>serC</i> , PSAT1; phosphoserine aminotransferase [EC:2.6.1.52]	-3,36
cromosoma	GONCAC_18095	<i>purD</i>	COG0151	F	K01945	<i>purD</i> ; phosphoribosylamine---glycine ligase [EC:6.3.4.13]	-3,36
cromosoma	GONCAC_08845	<i>trpCF</i>	COG0134	E	K13498	<i>trpCF</i> ; indole-3-glycerol phosphate synthase / phosphoribosylanthranilate isomerase [EC:4.1.1.48 5.3.1.24]	-3,34
cromosoma	GONCAC_12250	<i>pyrC</i>	COG0418	F	K01465	URA4, <i>pyrC</i> ; dihydroorotase [EC:3.5.2.3]	-3,34
cromosoma	GONCAC_04750	<i>purL</i>	COG0046	F	K01952	PFAS, <i>purL</i> ; phosphoribosylformylglycinamide synthase [EC:6.3.5.3]	-3,32
cromosoma	GONCAC_08860	<i>trpE</i>	COG0147	E	K01657	<i>trpE</i> ; anthranilate synthase component I [EC:4.1.3.27]	-3,32
cromosoma	GONCAC_03380	<i>serA</i>	COG0111	EH	K00058	<i>serA</i> , PHGDH; D-3-phosphoglycerate dehydrogenase / 2-oxoglutarate reductase [EC:1.1.1.95 1.1.1.399]	-3,3
cromosoma	GONCAC_00525	<i>argE</i>	COG0624	E	K01438	<i>argE</i> ; acetylornithine deacetylase [EC:3.5.1.16]	-3,29
cromosoma	GONCAC_00545	<i>argH</i>	COG0165	E	K01755	<i>argH</i> , ASL; argininosuccinate lyase [EC:4.3.2.1]	-3,29
cromosoma	GONCAC_06840	<i>hisD</i>	COG0141	E	K00013	<i>hisD</i> ; histidinol dehydrogenase [EC:1.1.1.23]	-3,28
cromosoma	GONCAC_18380	<i>metR</i>	COG0583	K	K03576	<i>metR</i> ; LysR family transcriptional regulator, regulator for <i>metE</i> and <i>metH</i>	-3,28
cromosoma	GONCAC_08850	<i>trpD</i>	COG0547	E	K00766	<i>trpD</i> ; anthranilate phosphoribosyltransferase [EC:2.4.2.18]	-3,27
cromosoma	GONCAC_16140	<i>leuA</i>	COG0119	E	K01649	<i>leuA</i> , IMS; 2-isopropylmalate synthase [EC:2.3.3.13]	-3,25
cromosoma	GONCAC_17340	<i>gltB3</i>	COG0067	E	K00265	<i>gltB</i> ; glutamate synthase (NADPH) large chain [EC:1.4.1.13]	-3,24
cromosoma	GONCAC_13995	-	COG0367	E	K01953	<i>asnB</i> , ASNS; asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]	-3,22
cromosoma	GONCAC_14840	<i>panE</i>	COG1893	H	K00077	<i>panE</i> , <i>apbA</i> ; 2-dehydropantoate 2-reductase [EC:1.1.1.169]	-3,21
cromosoma	GONCAC_15805	<i>panB</i>	COG0413	H	K00606	<i>panB</i> ; 3-methyl-2-oxobutanoate hydroxymethyltransferase [EC:2.1.2.11]	-3,19
pPagLPU12_2	GONCAC_20635	-	COG0476	H	K21147	<i>moeZR</i> , <i>moeBR</i> ; sulfur-carrier protein adenylyltransferase/sulfurtransferase [EC:2.7.7.80 2.7.7.- 2.8.1.11 2.8.1.-]	-3,18
cromosoma	GONCAC_03820	<i>cysJ</i>	COG0369	P	K00380	<i>cysJ</i> ; sulfite reductase (NADPH) flavoprotein alpha-component [EC:1.8.1.2]	-3,14
cromosoma	GONCAC_08855	<i>pabA</i>	COG0512	EH	K13497	<i>trpGD</i> ; anthranilate synthase/phosphoribosyltransferase [EC:4.1.3.27 2.4.2.18]	-3,14
cromosoma	GONCAC_16895	<i>argF</i>	COG0078	E	K00611	OTC, <i>argF</i> , <i>argI</i> ; ornithine carbamoyltransferase [EC:2.1.3.3]	-3,13
cromosoma	GONCAC_02335	<i>cysQ</i>	COG1218	P	K01082	<i>cysQ</i> , MET22, BPNT1; 3'(2'), 5'-bisphosphate nucleotidase [EC:3.1.3.7]	-3,12
cromosoma	GONCAC_18140	<i>thiE</i>	COG0352	H	K00788	<i>thiE</i> ; thiamine-phosphate pyrophosphorylase [EC:2.5.1.3]	-3,12
cromosoma	GONCAC_12750	<i>pyrD</i>	COG0167	F	K00254	DHODH, <i>pyrD</i> ; dihydroorotate dehydrogenase [EC:1.3.5.2]	-3,11
cromosoma	GONCAC_06845	<i>hisG</i>	COG0040	E	K00765	<i>hisG</i> ; ATP phosphoribosyltransferase [EC:2.4.2.17]	-3,1
cromosoma	GONCAC_08680	<i>purU</i>	COG0788	F	K01433	<i>purU</i> ; formyltetrahydrofolate deformylase [EC:3.5.1.10]	-3,1
cromosoma	GONCAC_16365	<i>thrA</i>	COG0460	E	K12524	<i>thrA</i> ; bifunctional aspartokinase / homoserine dehydrogenase 1 [EC:2.7.2.4 1.1.1.3]	-3,1
cromosoma	GONCAC_02760	<i>glnE</i>	COG1391	H	K00982	<i>glnE</i> ; [glutamine synthetase] adenylyltransferase / [glutamine synthetase]-adenylyl-L-tyrosine phosphorylase [EC:2.7.7.42 2.7.7.89]	-3,07

cromosoma	GONCAC_06400	-	COG3339	S			-3,06
cromosoma	GONCAC_00220	<i>pyrE</i>	COG0461	F	K00762	pyrE; orotate phosphoribosyltransferase [EC:2.4.2.10]	-3,04
cromosoma	GONCAC_18610	<i>trxA</i>	COG3118	O	K03671	TXN, <i>trxA</i> ; thioredoxin	-3,04
cromosoma	GONCAC_00505	-	COG0626	E	K01739	metB; cystathionine gamma-synthase [EC:2.5.1.48]	-3,02
cromosoma	GONCAC_00215	<i>rph</i>	COG0689	J	K00989	rph; ribonuclease PH [EC:2.7.7.56]	-2,99
cromosoma	GONCAC_17345	-	COG0369	P	K00380	cysJ; sulfite reductase (NADPH) flavoprotein alpha-component [EC:1.8.1.2]	-2,97
cromosoma	GONCAC_05275	<i>cysK</i>	COG0031	E	K01738	cysK; cysteine synthase [EC:2.5.1.47]	-2,96
cromosoma	GONCAC_18090	<i>purH</i>	COG0138	F	K00602	purH; phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase [EC:2.1.2.3 3.5.4.10]	-2,92
cromosoma	GONCAC_12960	<i>lrp</i>	COG1522	K	K03719	lrp; Lrp/AsnC family transcriptional regulator, leucine-responsive regulatory protein	-2,89
cromosoma	GONCAC_00530	<i>argC</i>	COG0002	E	K00145	argC; N-acetyl-gamma-glutamyl-phosphate reductase [EC:1.2.1.38]	-2,86
cromosoma	GONCAC_00540	<i>argG</i>	COG0137	E	K01940	argG, ASS1; argininosuccinate synthase [EC:6.3.4.5]	-2,85
cromosoma	GONCAC_00085	<i>glnL</i>	COG3852	T	K07708	glnL, ntrB; two-component system, NtrC family, nitrogen regulation sensor histidine kinase GlnL [EC:2.7.13.3]	-2,84
cromosoma	GONCAC_03830	<i>cysH</i>	COG0175	EH	K00390	cysH; phosphoadenosine phosphosulfate reductase [EC:1.8.4.8 1.8.4.10]	-2,84
pPagLPU12_2	GONCAC_20650	<i>thiO</i>	COG0665	E	K03153	thiO; glycine oxidase [EC:1.4.3.19]	-2,84
cromosoma	GONCAC_00520	<i>ppc</i>	COG2352	H	K01595	ppc; phosphoenolpyruvate carboxylase [EC:4.1.1.31]	-2,83
cromosoma	GONCAC_16395	<i>gpmB</i>	COG0406	G	K15634	gpmB; 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase [EC:5.4.2.11]	-2,83
cromosoma	GONCAC_03635	<i>argA</i>	COG0548	E	K14682	argAB; amino-acid N-acetyltransferase [EC:2.3.1.1]	-2,8
cromosoma	GONCAC_06810	<i>hisIE</i>	COG0139	E	K11755	hisIE; phosphoribosyl-AMP cyclohydrolase / phosphoribosyl-ATP pyrophosphohydrolase [EC:3.5.4.19 3.6.1.31]	-2,75
cromosoma	GONCAC_06835	<i>hisC</i>	COG0079	E	K00817	hisC; histidinol-phosphate aminotransferase [EC:2.6.1.9]	-2,75
cromosoma	GONCAC_01590	<i>astC</i>	COG4992	E	K00821	argD; acetylornithine/N-succinyldiaminopimelate aminotransferase [EC:2.6.1.11 2.6.1.17]	-2,7
cromosoma	GONCAC_18135	<i>thiC</i>	COG0422	H	K03147	thiC; phosphomethylpyrimidine synthase [EC:4.1.99.17]	-2,7
cromosoma	GONCAC_03790	<i>relA</i>	COG0317	KT	K00951	relA; GTP pyrophosphokinase [EC:2.7.6.5]	-2,64
cromosoma	GONCAC_00660	<i>ilvD</i>	COG0129	EG	K01687	ilvD; dihydroxy-acid dehydratase [EC:4.2.1.9]	-2,63
cromosoma	GONCAC_03510	-	COG0583	K			-2,6
cromosoma	GONCAC_04250	<i>aroG1</i>	COG0722	E	K01626	E2.5.1.54, aroF, aroG, aroH; 3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54]	-2,59
cromosoma	GONCAC_15815	<i>panD</i>	COG0853	H	K01579	panD; aspartate 1-decarboxylase [EC:4.1.1.11]	-2,56
cromosoma	GONCAC_15950	<i>aceE</i>	COG2609	C	K00163	aceE; pyruvate dehydrogenase E1 component [EC:1.2.4.1]	-2,54
cromosoma	GONCAC_03845	<i>cysG</i>	COG0007	H	K02302	cysG; uroporphyrin-III C-methyltransferase / precorrin-2 dehydrogenase / sirohydrochlorin ferrochelataze [EC:2.1.1.107 1.3.1.76 4.99.1.4]	-2,53
cromosoma	GONCAC_03850	<i>cysD</i>	COG0175	H	K00957	cysD; sulfate adenyltransferase subunit 2 [EC:2.7.7.4]	-2,49
cromosoma	GONCAC_17305	<i>rpoN</i>	COG1508	K	K03092	rpoN; RNA polymerase sigma-54 factor	-2,48
cromosoma	GONCAC_00665	<i>ilvA</i>	COG1171	E	K01754	E4.3.1.19, ilvA, tdcB; threonine dehydratase [EC:4.3.1.19]	-2,47

cromosoma	GONCAC_02895	<i>exbB</i>	COG0811	U	K03561	exbB; biopolymer transport protein ExbB	-2,43
cromosoma	GONCAC_00080	<i>glnG</i>	COG2204	T	K07712	glnG, ntrC; two-component system, NtrC family, nitrogen regulation response regulator GlnG	-2,41
cromosoma	GONCAC_03855	<i>cysN</i>	COG2895	H	K00956	cysN; sulfate adenylyltransferase subunit 1 [EC:2.7.7.4]	-2,41
cromosoma	GONCAC_16145	<i>leuB</i>	COG0473	CE	K00052	leuB, IMDH; 3-isopropylmalate dehydrogenase [EC:1.1.1.85]	-2,4
cromosoma	GONCAC_00670	<i>ilvY</i>	COG0583	K	K02521	ilvY; LysR family transcriptional regulator, positive regulator for ilvC	-2,36
cromosoma	GONCAC_13740	<i>nadA</i>	COG0379	H	K03517	nadA; quinolinate synthase [EC:2.5.1.72]	-2,28
cromosoma	GONCAC_00675	<i>ilvC</i>	COG0059	EH	K00053	ilvC; ketol-acid reductoisomerase [EC:1.1.1.86]	-2,27
cromosoma	GONCAC_06830	<i>hisB</i>	COG0131	E	K01089	hisB; imidazoleglycerol-phosphate dehydratase / histidinol-phosphatase [EC:4.2.1.19 3.1.3.15]	-2,26
cromosoma	GONCAC_12020	<i>ndh</i>	COG1252	C	K03885	ndh; NADH:quinone reductase (non-electrogenic) [EC:1.6.5.9]	-2,26
pPagLPU12_2	GONCAC_20710	<i>metE</i>	COG0620	E	K00549	metE; 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase [EC:2.1.1.14]	-2,23
cromosoma	GONCAC_05980	<i>cvpA</i>	COG1286	S	K03558	cvpA; membrane protein required for colicin V production	-2,21
cromosoma	GONCAC_03515	<i>lysA</i>	COG0019	E	K01586	lysA; diaminopimelate decarboxylase [EC:4.1.1.20]	-2,2
cromosoma	GONCAC_17055	<i>pnp</i>	COG1185	J	K00962	pnp, PNPT1; polyribonucleotide nucleotidyltransferase [EC:2.7.7.8]	-2,13
cromosoma	GONCAC_06815	<i>hisF</i>	COG0107	E	K02500	hisF; imidazole glycerol-phosphate synthase subunit HisF [EC:4.3.2.10]	-2,09
cromosoma	GONCAC_15945	<i>aceF</i>	COG0508	C	K00627	DLAT, aceF, pdhC; pyruvate dehydrogenase E2 component (dihydrolipoyllysine-residue acetyltransferase) [EC:2.3.1.12]	-2,08
cromosoma	GONCAC_02875	<i>metC</i>	COG0626	E	K01760	metC; cysteine-S-conjugate beta-lyase [EC:4.4.1.13]	-2,02
cromosoma	GONCAC_04260	<i>pheA</i>	COG0077	E	K14170	pheA; chorismate mutase / prephenate dehydratase [EC:5.4.99.5 4.2.1.51]	-2,02
cromosoma	GONCAC_17175	<i>fepD</i>	COG0609	P	K23186	fepD, fagA, cchC, desH; iron-siderophore transport system permease protein	-1,97
cromosoma	GONCAC_17185	<i>fepB</i>	COG4592	P	K23185	fepB; ferric enterobactin transport system substrate-binding protein	-1,95
cromosoma	GONCAC_00375	<i>cpxA</i>	COG0642	T	K07640	cpxA; two-component system, OmpR family, sensor histidine kinase CpxA [EC:2.7.13.3]	-1,94
cromosoma	GONCAC_06405	<i>yieE</i>	COG0583	K			-1,93
cromosoma	GONCAC_10860	-	COG0247	C	K26063	ydiJ; (R)-2-hydroxyglutarate dehydrogenase [EC:1.1.-.-]	-1,88
cromosoma	GONCAC_10500	-	COG3895	S			-1,87
cromosoma	GONCAC_01175	<i>panM</i>	COG0454	K			-1,86
cromosoma	GONCAC_00650	<i>ilvM</i>	COG3978	S	K11258	ilvM; acetolactate synthase II small subunit [EC:2.2.1.6]	-1,85
cromosoma	GONCAC_00725	<i>mqo</i>	COG0579	C	K00116	mqo; malate dehydrogenase (quinone) [EC:1.1.5.4]	-1,81
cromosoma	GONCAC_15985	<i>nadC</i>	COG0157	H	K00767	nadC, QPRT; nicotinate-nucleotide pyrophosphorylase (carboxylating) [EC:2.4.2.19]	-1,72
cromosoma	GONCAC_17165	<i>fecE</i>	COG1120	HP	K23188	fepC, fagC, cchE, desF; iron-siderophore transport system ATP-binding protein [EC:7.2.2.17 7.2.2.-]	-1,71
cromosoma	GONCAC_04075	-	COG0697	EG			-1,67
cromosoma	GONCAC_16135	-		-			-1,66
cromosoma	GONCAC_17170	<i>fepG</i>	COG4779	P	K23187	fepG, fagB, cchD, desG; iron-siderophore transport system permease protein	-1,66

cromosoma	GONCAC_16370	-		-			-1,61
cromosoma	GONCAC_07855	<i>fliz</i>	2DBMG	N	K02425	fliz; regulator of sigma S factor Fliz	-1,54
pPagLPU12_2	GONCAC_20715	-	28H5S	S			-1,54
cromosoma	GONCAC_00645	<i>ilvG</i>	COG0028	H	K01652	E2.2.1.6L, ilvB, ilvG, ilvI; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]	-1,51
cromosoma	GONCAC_03620	<i>ptrA</i>	COG1025	O	K01407	ptrA; protease III [EC:3.4.24.55]	-1,48
cromosoma	GONCAC_00130	-	COG1200	L	K03655	recG; ATP-dependent DNA helicase RecG [EC:5.6.2.4]	-1,46
cromosoma	GONCAC_09290	<i>uspE</i>	COG0589	T	K14055	uspE; universal stress protein E	-1,44
cromosoma	GONCAC_18510	<i>hemY</i>	COG3071	H	K02498	hemY; HemY protein	-1,34
cromosoma	GONCAC_15955	<i>pdhR</i>	COG2186	K	K05799	pdhR; GntR family transcriptional regulator, transcriptional repressor for pyruvate dehydrogenase complex	-1,33
cromosoma	GONCAC_02755	-	COG3025	S	K18446	ygiF; triphosphatase [EC:3.6.1.25]	-1,28
cromosoma	GONCAC_06065	<i>lrhA</i>	COG0583	K			-1,28
cromosoma	GONCAC_00510	<i>metL</i>	COG0460	E	K12525	metL; bifunctional aspartokinase / homoserine dehydrogenase 2 [EC:2.7.2.4 1.1.1.3]	-1,26
cromosoma	GONCAC_13635	<i>bioB</i>	COG0502	H	K01012	bioB; biotin synthase [EC:2.8.1.6]	-1,24
cromosoma	GONCAC_18110	<i>yjaG</i>	COG3068	S	K09891	K09891; uncharacterized protein	-1,24
cromosoma	GONCAC_01955	-	COG0663	S			-1,21
cromosoma	GONCAC_16130	<i>ilvI</i>	COG0028	H	K01652	E2.2.1.6L, ilvB, ilvG, ilvI; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]	-1,2
cromosoma	GONCAC_16470	<i>deoR</i>	COG1349	K	K11534	deoR; DeoR family transcriptional regulator, deoxyribose operon repressor	-1,17
cromosoma	GONCAC_04660	<i>nadB</i>	COG0029	H	K00278	nadB; L-aspartate oxidase [EC:1.4.3.16]	-1,15
cromosoma	GONCAC_04080	-	COG0151	F			-1,14
cromosoma	GONCAC_03340	<i>epd</i>	COG0057	G	K03472	epd; D-erythrose 4-phosphate dehydrogenase [EC:1.2.1.72]	-1,13
cromosoma	GONCAC_04215	<i>ypjD</i>	COG4137	S			-1,09
cromosoma	GONCAC_18505	<i>hemX</i>	COG2959	H	K02496	hemX; uroporphyrin-III C-methyltransferase [EC:2.1.1.107]	-1,09
cromosoma	GONCAC_17145	-	COG1629	P	K19611	fepA, pfeA, iron, pirA; ferric enterobactin receptor	-0,96
cromosoma	GONCAC_17150	<i>fes</i>	COG2382	P	K07214	fes; iron(III)-enterobactin esterase [EC:3.1.1.108]	-0,96
cromosoma	GONCAC_04840	<i>pepB</i>	COG0260	E	K07751	pepB; PepB aminopeptidase [EC:3.4.11.23]	-0,94
cromosoma	GONCAC_16125	<i>ilvN</i>	COG0440	E	K01653	E2.2.1.6S, ilvH, ilvN; acetolactate synthase I/III small subunit [EC:2.2.1.6]	-0,93
cromosoma	GONCAC_04275	<i>rluD</i>	COG0564	J	K06180	rluD; 23S rRNA pseudouridine1911/1915/1917 synthase [EC:5.4.99.23]	-0,91
cromosoma	GONCAC_18555	-	COG0554	C	K12582	wecF, rfft; dTDP-N-acetylglucosamine:lipid II N-acetylglucosaminyltransferase [EC:2.4.1.325]	-0,91
cromosoma	GONCAC_02270	<i>rnr</i>	COG0557	J	K12573	rnr, vacB; ribonuclease R [EC:3.1.13.1]	-0,83
cromosoma	GONCAC_01380	<i>glpR</i>	COG1349	K	K02444	glpR; DeoR family transcriptional regulator, glycerol-3-phosphate regulon repressor	-0,78
cromosoma	GONCAC_02265	<i>nsrR</i>	COG1959	K	K13771	nsrR; Rrf2 family transcriptional regulator, nitric oxide-sensitive transcriptional repressor	-0,75

cromosoma	GONCAC_16490	<i>nadR</i>	COG1056	H	K06211	nadR; HTH-type transcriptional regulator, transcriptional repressor of NAD biosynthesis genes [EC:2.7.7.1 2.7.1.22]	-0,74
cromosoma	GONCAC_03050	<i>deoC</i>	COG0274	F	K01619	deoC, DERA; deoxyribose-phosphate aldolase [EC:4.1.2.4]	-0,73
cromosoma	GONCAC_18575	<i>rfbA</i>	COG1209	M	K00973	rfbA, rmlA, rffH; glucose-1-phosphate thymidyltransferase [EC:2.7.7.24]	-0,73
cromosoma	GONCAC_03675	<i>csdA</i>	COG0520	E	K01766	csdA; cysteine sulfinatase desulfinate [EC:4.4.1.-]	-0,68
cromosoma	GONCAC_02035	<i>panF</i>	COG4145	H	K14392	panF; sodium/pantothenate symporter	-0,65
cromosoma	GONCAC_17905	<i>dusA</i>	COG0042	J	K05539	dusA; tRNA-dihydrouridine synthase A [EC:1.-.-.-]	-0,65
cromosoma	GONCAC_18545	<i>rffM</i>	COG1922	M	K02852	wecG, rffM; UDP-N-acetyl-D-mannosaminouronate: lipid I N-acetyl-D-mannosaminouronosyltransferase [EC:2.4.1.180]	-0,65
cromosoma	GONCAC_03785	<i>rlmD</i>	COG2265	H	K03215	rumA; 23S rRNA (uracil1939-C5)-methyltransferase [EC:2.1.1.190]	-0,62
cromosoma	GONCAC_16525	<i>rsmC</i>	COG2813	J	K00564	rsmC; 16S rRNA (guanine1207-N2)-methyltransferase [EC:2.1.1.172]	-0,61
cromosoma	GONCAC_18320	<i>tatD</i>	COG0084	L	K03424	tatD; TatD DNase family protein [EC:3.1.21.-]	-0,61

Tabla VI.S3. Genes cuya mutación en *P. agglomerans* LPU12 favorece el crecimiento en medio mínimo respecto del medio rico, identificado en los ensayos TnSeq (Capítulo VI).

Replicón	locus_tag	Gen (nombre)	COG	categoría COG	KO	Definición asociada al KO	M
cromosoma	GONCAC_08560	<i>spoVR</i>	COG2719	S	K06415	spoVR; stage V sporulation protein R	0,6
cromosoma	GONCAC_09105	<i>pspF</i>	COG2204	K	K03974	pspF; psp operon transcriptional activator	0,6
cromosoma	GONCAC_09595	-	COG1840	P	K02012	afuA, fbpA; iron(III) transport system substrate-binding protein	0,6
cromosoma	GONCAC_09895	-	COG0477	EGP	K03288	citA, tcuC; MFS transporter, MHS family, citrate/tricarballylate:H ⁺ symporter	0,6
cromosoma	GONCAC_10100	-	COG3500	S	K06905	D; Bacteriophage probable baseplate hub protein	0,6
cromosoma	GONCAC_10140	-	COG4385	S			0,6
cromosoma	GONCAC_11070	-	COG3203	M	K09476	ompF; outer membrane pore protein F	0,6
cromosoma	GONCAC_03300	-	COG3307	M	K02847	waaL, rfaL; O-antigen ligase [EC:2.4.99.26]	0,61
cromosoma	GONCAC_10625	<i>glgX</i>	COG1523	G	K01214	ISA, treX; isoamylase [EC:3.2.1.68]	0,61
cromosoma	GONCAC_06305	<i>rsuA</i>	COG1187	J	K06183	rsuA; 16S rRNA pseudouridine516 synthase [EC:5.4.99.19]	0,62
cromosoma	GONCAC_12730	-	COG0488	S	K15738	uup; ABC transport system ATP-binding/permease protein	0,62
cromosoma	GONCAC_07830	<i>fliD</i>	COG1345	N	K02407	fliD; flagellar hook-associated protein 2	0,63
cromosoma	GONCAC_09000	-	COG0840	NT	K05876	trg; methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	0,63
cromosoma	GONCAC_11365	-	COG1397	O			0,63
cromosoma	GONCAC_09100	-	COG0477	EGP			0,64
cromosoma	GONCAC_10065	<i>oqxA</i>	COG0845	M	K19586	oqxA; membrane fusion protein, multidrug efflux system	0,64
cromosoma	GONCAC_11915	<i>roxA</i>	COG2850	S	K18850	ycfD; 50S ribosomal protein L16 3-hydroxylase [EC:1.14.11.47]	0,64
cromosoma	GONCAC_10605	-	COG0668	M	K16053	ybdG, mscM; miniconductance mechanosensitive channel	0,65
cromosoma	GONCAC_10730	<i>fadH</i>	COG1902	C	K10680	nemA; N-ethylmaleimide reductase [EC:1.-.-.-]	0,65
cromosoma	GONCAC_11315	<i>lysR</i>	COG0583	K			0,65
cromosoma	GONCAC_12925	<i>pf1B</i>	COG1882	C	K00656	E2.3.1.54, pf1D; formate C-acetyltransferase [EC:2.3.1.54]	0,65
cromosoma	GONCAC_09995	<i>leuA</i>	COG0119	E	K01649	leuA, IMS; 2-isopropylmalate synthase [EC:2.3.3.13]	0,66
cromosoma	GONCAC_10235	-	28HIF	S			0,66
cromosoma	GONCAC_10340	-	COG1168	E	K14155	patB, malY; cysteine-S-conjugate beta-lyase [EC:4.4.1.13]	0,66
cromosoma	GONCAC_11355	-	COG3367	S	K26272	dgcN; D-glutamate N-acetyltransferase [EC:2.3.1.312]	0,66
cromosoma	GONCAC_09255	<i>rstB</i>	COG0642	T	K07639	rstB; two-component system, OmpR family, sensor histidine kinase RstB [EC:2.7.13.3]	0,67
cromosoma	GONCAC_10335	<i>malX</i>	COG1263	G	K02791	malX; maltose/glucose PTS system EIICB component [EC:2.7.1.199 2.7.1.208]	0,67
cromosoma	GONCAC_07485	<i>norM</i>	COG0534	V	K26937	dinF, mepA, vmrA; MATE family, multidrug efflux pump	0,68
cromosoma	GONCAC_07720	<i>fliF</i>	COG1766	N	K02409	fliF; flagellar M-ring protein FliF	0,68
cromosoma	GONCAC_09825	-	COG4948	M			0,68

cromosoma	GONCAC_11095	-	COG4932	M			0,68
cromosoma	GONCAC_11920	<i>pepT</i>	COG2195	E	K01258	pepT; tripeptide aminopeptidase [EC:3.4.11.4]	0,68
cromosoma	GONCAC_02095	<i>rng</i>	COG1530	J	K08301	rng, cafA; ribonuclease G [EC:3.1.26.-]	0,69
cromosoma	GONCAC_07715	<i>fliG</i>	COG1536	N	K02410	fliG; flagellar motor switch protein FlIG	0,69
cromosoma	GONCAC_08980	<i>cstA</i>	COG1966	T	K06200	cstA; carbon starvation protein	0,7
cromosoma	GONCAC_09660	-	COG3188	NU	K07347	fimD, fimC, mrkC, htrE, cssD; outer membrane usher protein	0,7
cromosoma	GONCAC_12720	<i>pqiB</i>	COG1463	Q	K06192	pqiB; paraquat-inducible protein B	0,71
cromosoma	GONCAC_13255	<i>dps</i>	COG0783	P	K04047	dps; starvation-inducible DNA-binding protein	0,71
cromosoma	GONCAC_08350	<i>yebS</i>	COG2995	C	K03808	pqiA; paraquat-inducible protein A	0,72
cromosoma	GONCAC_16120	<i>cra</i>	COG1609	K	K03435	fruR1, fruR; LacI family transcriptional regulator, fructose operon transcriptional repressor	0,72
pPagLPU12_2	GONCAC_21395	-	COG3959	G	K00615	E2.2.1.1, tktA, tktB; transketolase [EC:2.2.1.1]	0,72
cromosoma	GONCAC_11925	<i>cobB</i>	COG0846	K	K12410	cobB, srtN, npdA; NAD-dependent protein deacetylase/lipoamidase [EC:2.3.1.286 2.3.1.313]	0,73
cromosoma	GONCAC_02855	<i>ftsP</i>	COG2132	D	K04753	sufI; suppressor of ftsI	0,74
cromosoma	GONCAC_07705	<i>fliI</i>	COG1157	NU	K02412	fliI; flagellum-specific ATP synthase [EC:7.4.2.8]	0,74
cromosoma	GONCAC_08565	-	COG0665	C	K00285	dadA; D-amino-acid dehydrogenase [EC:1.4.5.1]	0,74
cromosoma	GONCAC_08995	<i>malX</i>	COG1263	G	K02791	malX; maltose/glucose PTS system EIICB component [EC:2.7.1.199 2.7.1.208]	0,74
cromosoma	GONCAC_09320	-	COG2272	I	K03929	pnbA; para-nitrobenzyl esterase [EC:3.1.1.-]	0,74
cromosoma	GONCAC_11910	<i>phoQ</i>	COG0642	T	K07637	phoQ; two-component system, OmpR family, sensor histidine kinase PhoQ [EC:2.7.13.3]	0,74
cromosoma	GONCAC_08320	-	COG2807	P			0,75
cromosoma	GONCAC_08490	-	COG3338	P	K01674	cah; carbonic anhydrase [EC:4.2.1.1]	0,75
cromosoma	GONCAC_09935	<i>ulaA</i>	COG3037	S	K03475	ulaA, sgaT; ascorbate PTS system EIIC component	0,76
cromosoma	GONCAC_10325	-	COG5339	S			0,76
cromosoma	GONCAC_08575	-	COG0477	EGP	K07552	bcr, tcaB; MFS transporter, DHA1 family, multidrug resistance protein	0,77
cromosoma	GONCAC_02495	<i>rsml</i>	COG0313	H	K07056	rsml; 16S rRNA (cytidine1402-2'-O)-methyltransferase [EC:2.1.1.198]	0,78
cromosoma	GONCAC_10070	<i>oqxB</i>	COG0841	V	K19585	oqxB; multidrug efflux pump	0,78
cromosoma	GONCAC_00295	-	COG0859	M			0,79
cromosoma	GONCAC_08725	<i>adhE</i>	COG1012	C	K04072	adhE; acetaldehyde dehydrogenase / alcohol dehydrogenase [EC:1.2.1.10 1.1.1.1]	0,79
cromosoma	GONCAC_09545	<i>dcp</i>	COG0339	E	K01284	dcp; peptidyl-dipeptidase Dcp [EC:3.4.15.5]	0,79
cromosoma	GONCAC_08000	-	COG0840	NT	K05874	tsr; methyl-accepting chemotaxis protein I, serine sensor receptor	0,8
cromosoma	GONCAC_08440	-	COG4943	T	K21090	adrB, pdeD; c-di-GMP phosphodiesterase [EC:3.1.4.52]	0,8
cromosoma	GONCAC_09650	-	COG3121	NU			0,8
cromosoma	GONCAC_14815	<i>cyoC</i>	COG1845	C	K02299	cyoC; cytochrome o ubiquinol oxidase subunit III	0,8
cromosoma	GONCAC_03205	<i>mltC</i>	COG0741	M	K08306	mltC; peptidoglycan lytic transglycosylase C [EC:4.2.2.29]	0,81

cromosoma	GONCAC_09755	-	COG4952	G	K01805	xylA; xylose isomerase [EC:5.3.1.5]	0,82
cromosoma	GONCAC_09580	-	COG3387	G			0,83
cromosoma	GONCAC_03295	-	COG0438	M	K02840	waaB, rfaB; UDP-D-galactose:(glucosyl)LPS alpha-1,6-D-galactosyltransferase [EC:2.4.1.-]	0,84
cromosoma	GONCAC_08015	<i>cheR</i>	COG1352	H	K00575	cheR; chemotaxis protein methyltransferase CheR [EC:2.1.1.80]	0,84
cromosoma	GONCAC_09365	-	COG0477	EGP			0,84
cromosoma	GONCAC_09555	-	COG0438	M			0,84
cromosoma	GONCAC_09680	-	COG0657	I			0,84
cromosoma	GONCAC_11475	-	COG3209	M			0,84
cromosoma	GONCAC_12315	<i>mdoG</i>	COG3131	P	K03670	mdoG; periplasmic glucans biosynthesis protein	0,84
cromosoma	GONCAC_07685	<i>fliM</i>	COG1868	N	K02416	fliM; flagellar motor switch protein FliM	0,85
cromosoma	GONCAC_09700	-	COG1902	C			0,85
cromosoma	GONCAC_10245	-	COG0338	L	K06223	dam; DNA adenine methylase [EC:2.1.1.72]	0,85
cromosoma	GONCAC_10475	<i>dtpA</i>	COG3104	U	K03305	TC.POT; proton-dependent oligopeptide transporter, POT family	0,86
cromosoma	GONCAC_11590	-	COG3755	S			0,86
cromosoma	GONCAC_09635	-	COG0006	E			0,87
cromosoma	GONCAC_08915	<i>acnA</i>	COG1048	C	K27802	acnA; aconitate hydratase A / 2-methylisocitrate dehidratase [EC:4.2.1.3 4.2.1.99]	0,88
cromosoma	GONCAC_05975	<i>dedD</i>	COG3147	D	K03749	dedD; DedD protein	0,89
cromosoma	GONCAC_06290	<i>yejK</i>	COG3081	S	K06899	ndpA; nucleoid-associated protein	0,89
cromosoma	GONCAC_08040	<i>flhA</i>	COG1298	N	K02400	flhA; flagellar biosynthesis protein FlhA	0,89
cromosoma	GONCAC_09410	<i>rhaA</i>	COG4806	G	K01813	rhaA; L-rhamnose isomerase [EC:5.3.1.14]	0,89
cromosoma	GONCAC_09655	<i>stbD</i>	COG3539	NU			0,9
cromosoma	GONCAC_09270	-	COG0531	E	K03758	arcD, lysI, lysP; arginine:ornithine antiporter / lysine permease	0,91
cromosoma	GONCAC_13905	-	COG0327	S			0,91
cromosoma	GONCAC_09535	<i>xdhB</i>	COG4631	C	K13482	xdhB; xanthine dehidrogenase large subunit [EC:1.17.1.4]	0,93
cromosoma	GONCAC_09960	<i>pqqE</i>	COG0535	S	K06139	pqqE; PqqA peptide cyclase [EC:1.21.98.4]	0,94
cromosoma	GONCAC_08965	<i>caiD</i>	COG1024	I	K13816	rpfF; DSF synthase	0,95
cromosoma	GONCAC_10970	-	2ED4F	S			0,95
cromosoma	GONCAC_08095	-	COG0371	C	K00005	gldA; glycerol dehidrogenase [EC:1.1.1.6]	0,96
cromosoma	GONCAC_09235	<i>hrpA</i>	COG1643	L	K03578	hrpA; ATP-dependent RNA helicase HrpA [EC:5.6.2.6]	0,96
cromosoma	GONCAC_12065	<i>ptsG</i>	COG1263	G	K02779	ptsG; glucose PTS system EIICB or EIICBA component [EC:2.7.1.199]	0,96
cromosoma	GONCAC_09970	<i>pqqC</i>	COG5424	H	K06137	pqqC; pyrroloquinoline-quinone synthase [EC:1.3.3.11]	0,97
cromosoma	GONCAC_09740	-	COG1172	G	K10440	rbsC; ribose transport system permease protein	0,98
cromosoma	GONCAC_07690	<i>fliL</i>	COG1580	N	K02415	fliL; flagellar protein FliL	0,99
cromosoma	GONCAC_07730	-		-			0,99
cromosoma	GONCAC_03245	<i>ycf19</i>	COG0762	S	K02221	yggT; YggT family protein	1

cromosoma	GONCAC_08010	-	COG0840	NT	K05874	tsr; methyl-accepting chemotaxis protein I, serine sensor receptor	1,01
cromosoma	GONCAC_09520	<i>ydfG</i>	COG4221	S	K16066	ydfG; 3-hydroxy acid dehydrogenase / malonic semialdehyde reductase [EC:1.1.1.381 1.1.1.-]	1,03
cromosoma	GONCAC_11020	-	COG0477	EGP	K07552	bcr, tcaB; MFS transporter, DHA1 family, multidrug resistance protein	1,03
cromosoma	GONCAC_17275	<i>eCM27</i>	COG0530	P	K07301	yrbG; cation:H ⁺ antiporter	1,05
cromosoma	GONCAC_09760	-	COG4993	G	K00117	gcd; quinoprotein glucose dehydrogenase [EC:1.1.5.2]	1,08
cromosoma	GONCAC_12310	<i>mdoH</i>	COG2943	M	K03669	mdoH; membrane glycosyltransferase [EC:2.4.1.-]	1,08
cromosoma	GONCAC_07670	<i>fliP</i>	COG1338	N	K02419	fliP; flagellar biosynthesis protein FliP	1,09
cromosoma	GONCAC_08805	<i>yciA</i>	COG1607	I	K10806	yciA; acyl-CoA thioesterase YciA [EC:3.1.2.-]	1,1
cromosoma	GONCAC_09015	-	COG4950	S			1,1
cromosoma	GONCAC_13765	<i>Lys_trna</i>		tRNA			1,11
cromosoma	GONCAC_09565	<i>ampC</i>	COG1680	V			1,12
cromosoma	GONCAC_10145	-	COG3948	S			1,12
cromosoma	GONCAC_04880	<i>rodZ</i>	COG1426	S	K15539	rodZ; cytoskeleton protein RodZ	1,14
cromosoma	GONCAC_09525	<i>guaD</i>	COG0402	F	K01487	guaD, GDA; guanine deaminase [EC:3.5.4.3]	1,14
cromosoma	GONCAC_01435	<i>ompR</i>	COG0745	K	K07659	ompR; two-component system, OmpR family, phosphate regulon response regulator OmpR	1,15
cromosoma	GONCAC_01440	<i>envZ</i>	COG0642	T	K07638	envZ; two-component system, OmpR family, osmolarity sensor histidine kinase EnvZ [EC:2.7.13.3]	1,15
cromosoma	GONCAC_08655	<i>topB</i>	COG0550	L	K03169	topB; DNA topoisomerase III [EC:5.6.2.1]	1,15
cromosoma	GONCAC_09835	-	COG0601	P	K02033	ABC.PE.P; peptide/nickel transport system permease protein	1,17
cromosoma	GONCAC_12700	-	COG1067	O	K04770	lonH; Lon-like ATP-dependent protease [EC:3.4.21.-]	1,17
cromosoma	GONCAC_08410	<i>rlmA</i>	COG0500	Q	K00563	rlmA1; 23S rRNA (guanine745-N1)-methyltransferase [EC:2.1.1.187]	1,18
cromosoma	GONCAC_09170	<i>zntB</i>	COG0598	P	K16074	zntB; zinc transporter	1,18
cromosoma	GONCAC_09955	<i>pqqF</i>	COG1025	O			1,18
cromosoma	GONCAC_02025	<i>dusB</i>	COG0042	J	K05540	dusB; tRNA-dihydrouridine synthase B [EC:1.-.-.-]	1,19
cromosoma	GONCAC_09490	-	COG1732	M	K05845	opuC; osmoprotectant transport system substrate-binding protein	1,19
cromosoma	GONCAC_09600	-	COG1178	P	K02011	afuB, fbpB; iron(III) transport system permease protein	1,19
cromosoma	GONCAC_09640	-	COG1289	S			1,2
pPagLPU12_1	GONCAC_19715	<i>arnD</i>	COG0726	G	K13014	arnD; undecaprenyl phosphate-alpha-L-ara4FN deformylase [EC:3.5.1.-]	1,2
cromosoma	GONCAC_09625	-	COG3550	S	K07154	hipA; serine/threonine-protein kinase HipA [EC:2.7.11.1]	1,21
cromosoma	GONCAC_10460	-	COG2271	G			1,21
cromosoma	GONCAC_18600	<i>wecA</i>	COG0472	M	K02851	wecA, tagO, rfe; UDP-GlcNAc:undecaprenyl-phosphate/ decaprenyl-phosphate GlcNAc-1-phosphate transferase [EC:2.7.8.33 2.7.8.35]	1,21
cromosoma	GONCAC_06030	-	COG0280	C	K13788	pta; phosphate acetyltransferase [EC:2.3.1.8]	1,24

cromosoma	GONCAC_08755	<i>oppD</i>	COG0444	P	K15583	oppD; oligopeptide transport system ATP-binding protein	1,25
cromosoma	GONCAC_02240	<i>hflX</i>	COG2262	S	K03665	hflX; GTPase	1,26
cromosoma	GONCAC_06750	-	COG1134	GM	K09691	wzt, rfbB; homopolymeric O-antigen transport system ATP-binding protein [EC:7.5.2.14]	1,26
cromosoma	GONCAC_11950	<i>mfd</i>	COG1197	L	K03723	mfd; transcription-repair coupling factor (superfamily II helicase) [EC:5.6.2.4]	1,3
cromosoma	GONCAC_04755	<i>GlmZ_SraJ</i>		ncRNA			1,34
cromosoma	GONCAC_04760	<i>GlmY_tke1</i>		ncRNA			1,34
cromosoma	GONCAC_10885	<i>ydiA</i>	COG1806	H	K09773	ppsR; [pyruvate, water dikinase]-phosphate phosphotransferase / [pyruvate, water dikinase] kinase [EC:2.7.4.28 2.7.11.33]	1,35
cromosoma	GONCAC_17875	<i>tyrB</i>	COG1448	E	K00832	tyrB; aromatic-amino-acid transaminase [EC:2.6.1.57]	1,41
cromosoma	GONCAC_04875	<i>pilW</i>	COG3063	NU	K02656	pilF; type IV pilus assembly protein PilF	1,42
cromosoma	GONCAC_06755	-	COG4627	S			1,46
cromosoma	GONCAC_09905	-	COG0583	K			1,49
cromosoma	GONCAC_09690	-	COG0583	K	K18900	bpeT; LysR family transcriptional regulator, regulator for bpeEF and oprC	1,51
cromosoma	GONCAC_06785	-	COG0438	M			1,54
cromosoma	GONCAC_06645	<i>udk</i>	COG0572	F	K00876	udk, UCK; uridine kinase [EC:2.7.1.48]	1,55
cromosoma	GONCAC_11800	<i>ldcA</i>	COG1619	V	K01297	ldcA; muramoyltetrapeptide carboxypeptidase [EC:3.4.17.13]	1,56
cromosoma	GONCAC_06770	<i>gmd</i>	COG1089	M	K01711	gmd, GMDS; GDPmannose 4,6-dehydratase [EC:4.2.1.47]	1,57
cromosoma	GONCAC_11210	-	COG0515	KLT			1,57
cromosoma	GONCAC_09975	<i>pqqB</i>	COG1235	S	K06136	pqqB; pyrroloquinoline quinone biosynthesis protein B	1,59
cromosoma	GONCAC_06740	-	COG1109	G	K01840	manB; phosphomannomutase [EC:5.4.2.8]	1,6
cromosoma	GONCAC_06765	-	COG0297	M			1,61
cromosoma	GONCAC_06735	<i>cpsB</i>	COG0662	GM	K00971	manC, cpsB; mannose-1-phosphate guanylyltransferase [EC:2.7.7.13]	1,63
cromosoma	GONCAC_09360	-	COG0624	E			1,73
cromosoma	GONCAC_08955	<i>araD</i>	COG0235	G	K03077	araD, ulaF, sgaE, sgbE; L-ribulose-5-phosphate 4-epimerase [EC:5.1.3.4]	1,75
cromosoma	GONCAC_06160	<i>nrdB</i>	COG0208	F	K00526	E1.17.4.1B, nrdB, nrdF; ribonucleoside-diphosphate reductase beta chain [EC:1.17.4.1]	1,97
cromosoma	GONCAC_08195	<i>zwf</i>	COG0364	G	K00036	G6PD, zwf; glucose-6-phosphate 1-dehydrogenase [EC:1.1.1.49 1.1.1.363]	2,09
cromosoma	GONCAC_08985	<i>rnb</i>	COG4776	J	K01147	rnb; exoribonuclease II [EC:3.1.13.1]	2,12
cromosoma	GONCAC_10320	<i>manA</i>	COG1482	G	K01809	manA, MPI; mannose-6-phosphate isomerase [EC:5.3.1.8]	2,22
cromosoma	GONCAC_06780	-	COG0438	M			2,46
cromosoma	GONCAC_06165	<i>nrdA</i>	COG0209	F	K00525	E1.17.4.1A, nrdA, nrdE; ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1]	2,53

Tabla VI.S4. Genes cuya mutación en *P. agglomerans* LPU12 causa una pérdida parcial en la habilidad para crecer en medio rico LB identificados en los ensayos TnSeq (Capítulo VI).

Replicón	locus_tag	Gen (nombre)	COG	categoría COG	KO	Definición asociada al KO	M
cromosoma	GONCAC_10330	-	COG1609	K	K16136	mall; LacI family transcriptional regulator, maltose regulon regulatory protein	-2,13
cromosoma	GONCAC_07455	-		-			-1,56
cromosoma	GONCAC_00560	<i>sthA</i>	COG1249	C	K00322	sthA, udhA; NAD(P) transhydrogenase [EC:1.6.1.1]	-1,2
cromosoma	GONCAC_09165	-	COG4166	E	K15580	oppA, mppA; oligopeptide transport system substrate-binding protein	-1,04
cromosoma	GONCAC_09775	-	COG1457	U	K03457	TC.NCS1; nucleobase:cation symporter-1, NCS1 family	-1,04
cromosoma	GONCAC_08740	<i>oppA</i>	COG4166	E	K15580	oppA, mppA; oligopeptide transport system substrate-binding protein	-1
cromosoma	GONCAC_08425	<i>manY</i>	COG3715	G	K02795	manY; mannose PTS system EIIC component	-0,94
cromosoma	GONCAC_11480	-	COG3209	M			-0,83
cromosoma	GONCAC_12160	<i>flgJ</i>	COG1705	N	K02395	flgJ; peptidoglycan hydrolase FlgJ	-0,81
cromosoma	GONCAC_08285	-	COG1835	I			-0,78
cromosoma	GONCAC_10380	<i>araA</i>	COG2160	G	K01804	araA; L-arabinose isomerase [EC:5.3.1.4]	-0,77
cromosoma	GONCAC_11145	-	COG0243	C	K00372	nasC, nasA; assimilatory nitrate reductase catalytic subunit [EC:1.7.99.-]	-0,74
cromosoma	GONCAC_09440	-	COG0477	EGP	K08159	sotB; MFS transporter, DHA1 family, L-arabinose/isopropyl-beta-D-thiogalactopyranoside export protein	-0,72
cromosoma	GONCAC_07420	<i>xyIG</i>	COG1129	P	K10441	rbsA; ribose transport system ATP-binding protein [EC:7.5.2.7]	-0,68
cromosoma	GONCAC_10120	-	COG5283	-			-0,68
cromosoma	GONCAC_11525	-	2F5B2	-			-0,66

Aunque no se incluyen en la tabla por su carácter pleiotrópico:

- Los mutantes en genes con locus_tag GONCAC_00345, GONCAC_00350, GONCAC_04075, GONCAC_09290, GONCAC_15950 (asociados a genes de auxotrofia) modificaron su representación de la biblioteca al ser crecida en LB, todos ellos con M negativos.
- Los mutantes en genes con locus tag GONCAC_09660 y GONCAC_09895 (asociados a un crecimiento preferencial en mínimo M9 respecto de medio LB rico) modificaron su representación de la biblioteca al ser crecida en LB, ambos con M positivos.

Tabla VI.S5. Genes relacionados a la colonización rizosférica de alfalfa por *P. agglomerans* LPU12 (excluidos los genes de auxotrofia) identificados en los ensayos TnSeq (Capítulo VI).

Replicón	locus_tag	Gen (nombre)	COG	categoría COG	KO	Definición asociada al KO	M
cromosoma	GONCAC_00360	<i>gpsA</i>	COG0240	I	K00057	gpsA; glycerol-3-phosphate dehydrogenase (NAD(P)+) [EC:1.1.1.94]	-2,62
cromosoma	GONCAC_02165	<i>mdh</i>	COG0039	C	K00024	mdh; malate dehydrogenase [EC:1.1.1.37]	-1,88
cromosoma	GONCAC_07980	<i>motA</i>	COG1291	N	K02556	motA; chemotaxis protein MotA	-1,68
cromosoma	GONCAC_07850	<i>fliA</i>	COG1191	K	K02405	fliA, whiG; RNA polymerase sigma factor FliA	-1,48
cromosoma	GONCAC_12150	<i>flgL</i>	COG1344	N	K02397	flgL; flagellar hook-associated protein 3 FlgL	-1,42
cromosoma	GONCAC_07835	-	COG1344	N	K02406	fliC, hag; flagellin	-1,4
cromosoma	GONCAC_08615	-	COG0676	G	K01792	E5.1.3.15; glucose-6-phosphate 1-epimerase [EC:5.1.3.15]	-1,4
cromosoma	GONCAC_12165	<i>flgI</i>	COG1706	N	K02394	flgI; flagellar P-ring protein FlgI	-1,33
cromosoma	GONCAC_12180	<i>flgF</i>	COG4787	N	K02391	flgF; flagellar basal-body rod protein FlgF	-1,26
cromosoma	GONCAC_12155	<i>flgK</i>	COG1256	N	K02396	flgK; flagellar hook-associated protein 1	-1,25
cromosoma	GONCAC_12175	<i>flgG</i>	COG4786	N	K02392	flgG; flagellar basal-body rod protein FlgG	-1,24
cromosoma	GONCAC_12170	<i>flgH</i>	COG2063	N	K02393	flgH; flagellar L-ring protein FlgH	-1,09
cromosoma	GONCAC_12185	<i>flgE</i>	COG1749	N	K02390	flgE; flagellar hook protein FlgE	-1,07
cromosoma	GONCAC_14805	<i>cyoA</i>	COG1622	C	K02297	cyoA; cytochrome o ubiquinol oxidase subunit II [EC:7.1.1.3]	-1,07
cromosoma	GONCAC_02740	<i>bacA</i>	COG1968	V	K06153	bacA; undecaprenyl-diphosphatase [EC:3.6.1.27]	-1,03
cromosoma	GONCAC_00095	<i>typA</i>	COG1217	T	K06207	typA, bipA; GTP-binding protein	-1,01
cromosoma	GONCAC_07830	<i>fliD</i>	COG1345	N	K02407	fliD; flagellar hook-associated protein 2	-1,01
cromosoma	GONCAC_03335	<i>tkt</i>	COG0021	G	K00615	E2.2.1.1, tktA, tktB; transketolase [EC:2.2.1.1]	-0,98
cromosoma	GONCAC_18335	<i>tatA</i>	COG1826	U	K03116	tatA; sec-independent protein translocase protein TatA	-0,97
cromosoma	GONCAC_07990	<i>cheA</i>	COG0643	NT	K03407	cheA; two-component system, chemotaxis family, sensor kinase CheA [EC:2.7.13.3]	-0,96
cromosoma	GONCAC_12190	<i>flgD</i>	COG1843	N	K02389	flgD; flagellar basal-body rod modification protein FlgD	-0,96
cromosoma	GONCAC_12805	-	COG3203	M	K09476	ompF; outer membrane pore protein F	-0,95
cromosoma	GONCAC_07720	<i>fliF</i>	COG1766	N	K02409	fliF; flagellar M-ring protein FliF	-0,93
cromosoma	GONCAC_07975	<i>flhC</i>	2DBG4	K	K02402	flhC; flagellar transcriptional activator FlhC	-0,93
cromosoma	GONCAC_07660	<i>fliR</i>	COG1684	N	K02421	fliR; flagellar biosynthesis protein FliR	-0,91
cromosoma	GONCAC_04190	-	COG1238	S			-0,87
cromosoma	GONCAC_06240	-	2AUWP	-			-0,87
cromosoma	GONCAC_13520	-	COG1376	S	K19235	ybiS; L,D-transpeptidase YbiS	-0,85
cromosoma	GONCAC_00290	-	COG1442	M			-0,8
cromosoma	GONCAC_12315	<i>mdoG</i>	COG3131	P	K03670	mdoG; periplasmic glucans biosynthesis protein	-0,78
cromosoma	GONCAC_05240	<i>cysM</i>	COG0031	E	K12339	cysM; S-sulfo-L-cysteine synthase (O-acetyl-L-serine-dependent) [EC:2.5.1.144]	-0,75
cromosoma	GONCAC_18325	<i>tatC</i>	COG0805	U	K03118	tatC; sec-independent protein translocase protein TatC	-0,75
cromosoma	GONCAC_03300	-	COG3307	M	K02847	waaL, rfaL; O-antigen ligase [EC:2.4.99.26]	-0,71

cromosoma	GONCAC_12825	<i>ldtD</i>	COG2989	S	K21470	ycbB; L,D-transpeptidase YcbB	-0,71
cromosoma	GONCAC_13700	<i>galM</i>	COG2017	G	K01785	galM, GALM; aldose 1-epimerase [EC:5.1.3.3]	-0,68
cromosoma	GONCAC_13930	<i>pgm</i>	COG0033	G	K01835	pgm; phosphoglucomutase [EC:5.4.2.2]	-0,68
cromosoma	GONCAC_10105	-	COG3497	S	K06907	Fl; Bacteriophage tail sheath protein	-0,64
cromosoma	GONCAC_14810	-	COG0843	C	K02298	cyoB; cytochrome o ubiquinol oxidase subunit I [EC:7.1.1.3]	-0,63
cromosoma	GONCAC_11630	<i>tssC</i>	COG3517	S	K11900	impC; type VI secretion system protein ImpC	0,61
cromosoma	GONCAC_03985	-	COG1196	D			0,63
cromosoma	GONCAC_04015	<i>hxsD</i>	2E65T	-			0,65
cromosoma	GONCAC_02455	-	2DBAP	S			0,66
cromosoma	GONCAC_04970	<i>tehB</i>	COG4106	S	K16868	tehB; tellurite methyltransferase [EC:2.1.1.265]	0,68
pPagLPU12_2	GONCAC_20185	<i>dcuR</i>	COG4565	K	K07703	dcuR; two-component system, CitB family, response regulator DcuR	1,05
pPagLPU12_2	GONCAC_20180	-	COG3290	T	K07701	dcuS; two-component system, CitB family, sensor histidine kinase DcuS [EC:2.7.13.3]	1,21
pPagLPU12_2	GONCAC_20175	<i>mgo</i>	COG0579	C	K00116	mgo; malate dehydrogenase (quinone) [EC:1.1.5.4]	1,35

Aunque no se incluye en la tabla por su carácter pleiotrópico, el mutante en el gen con locus tag GONCAC_13255 (asociados a un crecimiento preferencial en mínimo M9 respecto de medio LB rico) también afectó la colonización de la rizósfera ($M = 0,96$).

Tabla VI.S6. Genes relacionados con la colonización endofítica de alfalfa por *P. agglomerans* LPU12 (excluidos los genes de auxotrofia) identificados en los ensayos TnSeq (Capítulo VI).

Replicón	locus_tag	Gen (nombre)	COG	Categoría COG	KO	Definición asociada al KO	M_temprano	M_tardío
cromosoma	GONCAC_06185	<i>rscC</i>	COG0642	T	K07677	rscC; two-component system, NarL family, capsular synthesis sensor histidine kinase RcsC [EC:2.7.13.3]	-1,59	-1,59
cromosoma	GONCAC_04745	<i>mltF</i>	COG4623	M	K18691	mltF; peptidoglycan lytic transglycosylase F [EC:4.2.2.29]	-1,47	-1,08
cromosoma	GONCAC_14700	<i>amtB</i>	COG0004	P	K03320	amt, AMT, MEP; ammonium transporter, Amt family	-1,34	-1,62
cromosoma	GONCAC_15325	<i>mtnK</i>	COG4857	H	K00899	mtnK; 5-methylthioribose kinase [EC:2.7.1.100]	-1,31	-1,34
cromosoma	GONCAC_05825	<i>nrfE</i>	COG1138	O	K02198	ccmF; cytochrome c-type biogenesis protein CcmF	-1,24	-1,73
cromosoma	GONCAC_15025	-	COG0531	E	K16238	eat, eutP; ethanolamine permease	-1,24	-1,31
cromosoma	GONCAC_12440	-	COG0477	EGP			-1,2	-1,21
cromosoma	GONCAC_06380	<i>fruA</i>	COG1299	G	K02770	fruA; fructose PTS system EIIBC or EIIC component [EC:2.7.1.202]	-1,19	-1,48
cromosoma	GONCAC_06205	-	COG2373	M			-1,17	-1,49
cromosoma	GONCAC_12010	-	COG3210	U	K15125	fhaB; filamentous hemagglutinin	-1,11	-1,27
cromosoma	GONCAC_04915	-	COG4993	G	K05358	quiA; quinate dehydrogenase (quinone) [EC:1.1.5.8]	-1,1	-1,18
cromosoma	GONCAC_13495	-	COG3264	M	K05802	mscK, kefA, aefA; potassium-dependent mechanosensitive channel	-1,1	-1,95
cromosoma	GONCAC_12675	-	COG1289	S			-1,09	-1,52
cromosoma	GONCAC_05145	<i>acrD</i>	COG0841	V	K18324	acrD; multidrug efflux pump	-1,06	-1,13
cromosoma	GONCAC_03910	<i>mutS</i>	COG0249	L	K03555	mutS; DNA mismatch repair protein MutS	-1,04	-1,16
cromosoma	GONCAC_13725	-	COG0642	T	K07679	evgS, bvgS; two-component system, NarL family, sensor histidine kinase EvgS [EC:2.7.13.3]	-1,04	-1,59
cromosoma	GONCAC_15220	-	COG5434	M			-1,03	-1,09
cromosoma	GONCAC_04050	-	COG0477	U	K02100	araE; MFS transporter, SP family, arabinose:H+ symporter	-1,02	-1,08
cromosoma	GONCAC_15450	<i>ncs1</i>	COG1953	FH	K03457	TC.NCS1; nucleobase:cation symporter-1, NCS1 family	-1,01	-1,17
cromosoma	GONCAC_00330	-	COG0438	M			1,01	1,18
cromosoma	GONCAC_17550	-	COG3842	P	K10111	malK, mtlK, thuK; multiple sugar transport system ATP-binding protein [EC:7.5.2.-]	1,02	2,71
cromosoma	GONCAC_18920	<i>ynieH</i>	COG0637	S			1,02	2,41
cromosoma	GONCAC_01050	-	COG4943	T	K26148	K26148; c-di-GMP phosphodiesterase [EC:3.1.4.52]	1,05	1,46
cromosoma	GONCAC_18730	<i>bcsQ</i>	COG1192	D			1,05	1,53
cromosoma	GONCAC_00285	-	COG1216	S	K19354	waaH; heptose III glucuronosyltransferase [EC:2.4.1.-]	1,07	1,38

cromosoma	GONCAC_17530	<i>ugpE</i>	COG0395	P	K10119	msmG; raffinose/stachyose/ melibiose transport system permease protein	1,07	2,52
cromosoma	GONCAC_19030	-	COG2232	S			1,07	1,96
cromosoma	GONCAC_18785	-	COG0317	KT			1,1	2,27
cromosoma	GONCAC_01275	-	2AV50	-			1,12	2,08
cromosoma	GONCAC_19090	-	COG5571	N			1,12	1,88
cromosoma	GONCAC_19250	<i>ovoA</i>	COG1262	H			1,12	1,54
cromosoma	GONCAC_18835	-	COG0160	E			1,15	2,19
cromosoma	GONCAC_19095	-	2,00E+47	-			1,15	2,4
cromosoma	GONCAC_18910	-	COG3274	S			1,18	1,54
cromosoma	GONCAC_19315	-	COG0477	EGP			1,18	1,82
cromosoma	GONCAC_17540	-	COG1653	G	K10117	msmE; raffinose/stachyose/ melibiose transport system substrate-binding protein	1,21	2,46
cromosoma	GONCAC_00440	<i>emrD</i>	COG0477	EGP	K08154	emrD; MFS transporter, DHA1 family, 2-module integral membrane pump EmrD	1,23	1,32
cromosoma	GONCAC_19260	<i>ravA</i>	COG0714	S	K03924	moxR; MoxR-like ATPase [EC:3.6.3.-]	1,23	2,01
cromosoma	GONCAC_01685	<i>fkpA</i>	COG0545	M	K03772	fkpA; FKBP-type peptidyl-prolyl cis-trans isomerase FkpA [EC:5.2.1.8]	1,24	1,94
cromosoma	GONCAC_19080	<i>ompA</i>	COG2885	M			1,25	2,04
cromosoma	GONCAC_00155	<i>ligB</i>	COG0272	L	K01972	E6.5.1.2, ligA, ligB; DNA ligase (NAD+) [EC:6.5.1.2]	1,27	2,15
cromosoma	GONCAC_18770	<i>eptB</i>	COG2194	S	K12975	eptB; KDO II ethanolaminophosphotransferase [EC:2.7.8.42]	1,28	1,93
cromosoma	GONCAC_00185	<i>lpxP</i>	COG1560	M	K12974	lpxP; KDO2-lipid IV(A) palmitoleyltransferase [EC:2.3.1.242]	1,36	2,53
cromosoma	GONCAC_01045	<i>proP</i>	COG0477	EGP	K03762	proP; MFS transporter, MHS family, proline/betaine transporter	1,42	2,08
cromosoma	GONCAC_17590	<i>gcd</i>	COG4993	G	K00117	gcd; quinoprotein glucose dehydrogenase [EC:1.1.5.2]	1,45	2,1
cromosoma	GONCAC_00160	-	COG2860	S			1,47	2,07
cromosoma	GONCAC_02950	-	2CEAX1	S			1,61	2,63
cromosoma	GONCAC_00210	<i>nupC</i>	COG1972	F	K03317	TC.CNT; concentrative nucleoside transporter, CNT family	1,67	2,25
cromosoma	GONCAC_00335	-	COG2861	S	K09798	K09798; uncharacterized protein	1,77	2,22
cromosoma	GONCAC_18855	<i>fdxH</i>	COG0437	C	K00124	fdoH, fdsB; formate dehydrogenase iron-sulfur subunit	1,97	1,67
cromosoma	GONCAC_04545	-	2E13C	S			2,14	2,16
cromosoma	GONCAC_00820	-	COG1082	G			2,46	2,43
cromosoma	GONCAC_18825	-	COG2990	S	K09824	virK; uncharacterized protein	2,51	2,88
cromosoma	GONCAC_15275	<i>gpt</i>	COG2236	F	K00769	gpt; xanthine phosphoribosyltransferase [EC:2.4.2.22]	2,77	2,81

cromosoma	GONCAC_03275	<i>rsmE</i>	COG1385	J	K09761	rsmE; 16S rRNA (uracil1498-N3)-methyltransferase [EC:2.1.1.193]	2,89	2,95
cromosoma	GONCAC_17745	-	COG2026	DJ			3,11	2,38
cromosoma	GONCAC_00635	<i>yifE</i>	COG3085	S	K09897	K09897; uncharacterized protein	3,26	3,11
cromosoma	GONCAC_00035	<i>mobA</i>	COG0746	H	K03752	mobA; molybdenum cofactor guanylyltransferase [EC:2.7.7.77]	3,3	1,97
cromosoma	GONCAC_00555	<i>oxyR</i>	COG0583	K	K04761	oxyR; LysR family transcriptional regulator, hydrogen peroxide-inducible genes activator	3,32	2,32
cromosoma	GONCAC_06520	<i>ahpC</i>	COG0450	O	K24119	ahpC; NADH-dependent peroxiredoxin subunit C [EC:1.11.1.26]	3,61	3,27
cromosoma	GONCAC_17750	<i>gltP</i>	COG1301	U	K11102	gltP, gltT; proton glutamate symport protein	3,63	2,77
Cromosoma	GONCAC_13245	<i>ompX</i>	COG3637	M	K11934	ompX; outer membrane protein X	3,83	3,2
pPagLPU12_2	GONCAC_20140	<i>argD</i>	COG0160	E	K00836	ectB, dat; diaminobutyrate-2-oxoglutarate transaminase [EC:2.6.1.76]	1,63	2,44
pPagLPU12_2	GONCAC_22130	-	COG1309	K	K16137	nemR; TetR/AcrR family transcriptional regulator, transcriptional repressor for nem operon	3,64	3,7
pPagLPU12_3	GONCAC_22995	-	COG1835	I	K16568	exoZ; exopolysaccharide production protein ExoZ	1,42	1,26

Aunque no se incluyen en la tabla por su carácter pleiotrópico, los mutantes en genes con locus_tag GONCAC_05985, GONCAC_16260, GONCAC_04980, GONCAC_03820, GONCAC_04750, GONCAC_03855 y GONCAC_03825 (asociados a genes de auxotrofia) también afectó la colonización de la endósfera, todos ellos con M negativos.

Tabla VIII.S7. Genomas y MAGs (genomas ensamblados a partir de metagenomas) del género *Pantoea*, obtenidos de la base de datos IMG (Chen et al., 2023), utilizados para el entrenamiento y prueba de modelos de *Machine Learning* diseñados para clasificar aislamientos según su capacidad de asociación con plantas (Capítulo VII).

Identificador IMG	Nombre del aislamiento	NCBI Bioproject Identificador	Tamaño del genoma (Mpb)	Tipo de hospedador	Patogenicidad
2511231025	<i>Pantoea</i> sp. YR343	PRJNA83107	5,32	Planta	No patógeno
2511231035	<i>Pantoea</i> sp. GM01	PRJNA83063	5,34	Planta	No patógeno
2511231160	<i>Pantoea ananatis</i> PA13	PRJNA74285	4,87	Planta	Patógeno
2511231178	<i>Pantoea ananatis</i> LMG 5342	PRJEA78277	4,91	Animal	Patógeno
2513237354	<i>Pantoea stewartii stewartii</i> DC283	PRJNA39561	5,23	Planta	Patógeno
2519899784	<i>Pantoea</i> sp. Sc1	PRJNA86999	4,45	Planta	Patógeno
2541047590	<i>Pantoea agglomerans</i> 299R	PRJNA181162	4,58	Planta	No patógeno
2551306469	<i>Pantoea</i> sp. A4	PRJNA173196	5,33	Planta	No patógeno
2558309123	<i>Pantoea agglomerans</i> Tx10	PRJNA202241	4,86	Animal	Patógeno
2568526489	<i>Pantoea ananatis</i> Sd-1	PRJNA232443	4,92	Planta	No patógeno
2576861011	<i>Pantoea ananatis</i> LMG 2665	PRJNA240762	4,98	Planta	Patógeno
2576861661	<i>Pantoea ananatis</i> BRT175	PRJNA202242	4,85	Planta	No patógeno
2576861666	<i>Pantoea agglomerans</i> Eh318	PRJNA222832	5,04	Planta	No patógeno
2576861786	<i>Pantoea agglomerans</i> RIT273	PRJNA239282	5,37	Planta	No patógeno
2582580715	<i>Pantoea agglomerans</i> KNU001	PRJNA243037	4,94	Planta	Patógeno
2582581300	<i>Pantoea</i> sp. 9140	PRJNA247914	4,84	Hongo	No patógeno
2585427963	<i>Pantoea ananatis</i> LMG2665	PRJNA245803	4,94	Planta	no patógeno
2585427967	<i>Pantoea ananatis</i> BD442	PRJNA245793	4,80	Planta	Patógeno
2585427976	<i>Pantoea ananatis</i> PA4	PRJNA245800	5,16	Planta	Patógeno
2597490175	<i>Pantoea agglomerans</i> 4	PRJNA255804	4,81	Planta	No patógeno
2599185299	<i>Pantoea ananatis</i> NFR11	PRJNA263085	4,85	Planta	No patógeno
2600255084	<i>Pantoea stewartii</i> indologenes LMG 2632	PRJNA252992	4,68	Planta	Patógeno
2602041550	<i>Pantoea</i> sp. AS-PWVM4	PRJNA202228	4,99	Planta	No patógeno
2602041581	<i>Pantoea dispersa</i> EGD-AAK13	PRJNA211932	4,76	Ambiental	Desconocido

2602041648	<i>Pantoea ananatis</i> DAR 76143	PRJDB1421	5,25	Planta	Patógeno
2608642108	<i>Pantoea agglomerans</i> NFPP29	PRJNA263080	4,10	Planta	No patógeno
2609460089	<i>Pantoea</i> sp. IMH	PRJNA239657	4,09	Ambiental	Desconocido
2609460211	<i>Pantoea agglomerans</i> DAPP-PG734	PRJNA251954	5,37	Planta	No patógeno
2609460322	<i>Pantoea agglomerans</i> MP2	PRJNA254768	4,73	Artrópodo	No patógeno
2616644925	<i>Pantoea</i> sp. 3.5.1	PRJNA246264	4,96	Ambiental	Desconocido
2617271107	<i>Pantoea agglomerans</i> 190	PRJNA248051	5,00	Ambiental	Desconocido
2617271108	<i>Pantoea</i> sp. FF5	PRJEB4277	4,55	Animal	No patógeno
2619619082	<i>Pantoea agglomerans</i> SL1_M5	PRJNA49341	4,92	Planta	Patógeno
2627853687	<i>Pantoea</i> sp. MBLJ3	PRJNA264308	4,82	Ambiental	Desconocido
2627853780	<i>Pantoea agglomerans</i> LMAE-2	PRJNA270082	4,98	Ambiental	Desconocido
2627853912	<i>Pantoea</i> sp. SM3	PRJNA277270	5,33	Ambiental	Desconocido
2630968876	<i>Pantoea</i> sp. PSNIH1	PRJNA242843	4,41	Ambiental	Patógeno
2630968889	<i>Pantoea</i> sp. PSNIH2	PRJNA242850	4,97	Ambiental	Patógeno
2630968904	<i>Pantoea ananatis</i> PaMB1	PRJNA259615	4,76	Planta	Patógeno
2630968991	<i>Pantoea ananatis</i> MR5	PRJNA280811	4,15	Planta	No patógeno
2636415588	<i>Pantoea</i> sp. BL1	PRJNA277252	5,33	Ambiental	Desconocido
2636415703	<i>Pantoea anthophila</i> 11-2	PRJNA274021	4,60	Ambiental	Desconocido
2636415742	<i>Pantoea dispersa</i> SUBG008	PRJNA286066	4,44	Planta	Patógeno
2636415937	<i>Pantoea stewartii</i> M009	PRJNA263996	4,82	Ambiental	Patógeno
2639763161	<i>Pantoea ananatis</i> CFH 7-1	PRJNA287251	4,60	Planta	Patógeno
2645727681	<i>Pantoea agglomerans</i> P10c	PRJNA292535	4,78	Planta	No patógeno
2645727744	<i>Pantoea ananatis</i> GB1	PRJNA275554	4,76	Planta	No patógeno
2648501511	<i>Pantoea</i> sp. Russ	-	3,94	Ambiental	Desconocido
2648501645	<i>Pantoea vagans</i> ZBG6	PRJNA287868	4,72	Ambiental	Desconocido
2648501680	<i>Pantoea stewartii</i> RSA13	PRJNA278249	4,77	Planta	No patógeno
2648501693	<i>Pantoea ananatis</i> B1-9	PRJEA68179	5,07	Planta	Desconocido
2648501765	<i>Pantoea stewartii</i> RSA36	PRJNA278253	4,77	Planta	No patógeno
2651869643	<i>Pantoea vagans</i> ND02	PRJNA261094	4,31	Ambiental	Desconocido

2651869657	<i>Pantoea</i> sp. RIT-PI-b	PRJNA289424	5,27	Planta	No patógeno
2651869808	<i>Pantoea stewartii</i> NS381	PRJNA278247	4,69	Planta	No patógeno
2654587503	<i>Pantoea stewartii</i> M073a	PRJNA265925	4,81	Ambiental	Desconocido
2654587623	<i>Pantoea stewartii</i> RSA30	PRJNA278250	4,74	Planta	No patógeno
2663763475	<i>Pantoea</i> sp. NGS-ED-1003	PRJNA255947	4,81	Animal	Desconocido
2667527843	<i>Pantoea</i> sp. CFSAN033090	PRJNA283309	4,94	Planta	Desconocido
2675903554	<i>Pantoea dispersa</i> SA4	PRJNA278433	4,91	Planta	No patógeno
2675903586	<i>Pantoea vagans</i> PaVv11	PRJEB8255	4,83	Planta	Desconocido
2681813054	<i>Pantoea ananatis</i> NS311	PRJNA278205	4,70	Planta	No patógeno
2681813055	<i>Pantoea dispersa</i> RSA31	PRJNA278252	4,85	Planta	No patógeno
2681813360	<i>Pantoea ananatis</i> S7	PRJEB7512	4,49	Planta	No patógeno
2684622535	<i>Pantoea ananatis</i> AMG521	PRJNA301127	4,86	Planta	No patógeno
2684622997	<i>Pantoea ananatis</i> NFIX48	PRJNA263067	4,62	Planta	No patógeno
2687453352	<i>Pantoea ananatis</i> R100	PRJNA310041	4,86	Planta	no patógeno
2690315618	<i>Pantoea</i> sp. YR512	PRJNA335059	5,28	Planta	No patógeno
2690315619	<i>Pantoea</i> sp. YR525	PRJNA335060	5,28	Planta	No patógeno
2690315639	<i>Pantoea</i> sp. OV426	PRJNA329711	4,75	Planta	No patógeno
2690315785	<i>Pantoea agglomerans</i> 3	PRJNA315986	4,81	Planta	No patógeno
2706794650	<i>Pantoea ananatis</i> S6	PRJEB7511	4,34	Planta	No patógeno
2706794651	<i>Pantoea ananatis</i> S8	PRJEB7513	4,48	Planta	No patógeno
2713896807	<i>Pantoea stewartii</i> A206	PRJNA292338	4,65	Planta	Patógeno
2713896810	<i>Pantoea stewartii</i> S301	PRJNA292332	4,48	Planta	Patógeno
2716884865	<i>Pantoea agglomerans</i> C410P1	PRJNA335437	5,12	Planta	No patógeno
2731957702	<i>Pantoea agglomerans</i> NBRC 102470	PRJDB388	4,65	Animal	Desconocido
2731957876	<i>Pantoea ananatis</i> NS296	PRJNA278207	4,73	Planta	No patógeno
2731957877	<i>Pantoea ananatis</i> NS303	PRJNA278206	4,73	Planta	No patógeno
2731957878	<i>Pantoea ananatis</i> RSA47	PRJNA278208	4,74	Planta	No patógeno
2731957898	<i>Pantoea dispersa</i> NS215	PRJNA278251	4,84	Planta	No patógeno
2731957899	<i>Pantoea dispersa</i> NS375	PRJNA278245	4,88	Planta	No patógeno

2731957900	<i>Pantoea dispersa</i> NS380	PRJNA278246	4,77	Planta	No patógeno
2731957901	<i>Pantoea dispersa</i> NS389	PRJNA278248	4,81	Planta	No patógeno
2731957902	<i>Pantoea dispersa</i> SA2	PRJNA278241	4,91	Planta	No patógeno
2731957903	<i>Pantoea dispersa</i> SA3	PRJNA278432	4,90	Planta	No patógeno
2731957904	<i>Pantoea dispersa</i> SA5	PRJNA278434	4,88	Planta	No patógeno
2751185886	<i>Pantoea eucrina</i> Russ	PRJNA327384	3,94	Ambiental	Desconocido
2751185887	<i>Pantoea alhagi</i> LTYR-11Z	PRJNA374633	4,32	Planta	No patógeno
2757320372	<i>Pantoea</i> sp. JKS000250	PRJNA367742	4,87	Hongo	No patógeno
2765236068	<i>Pantoea</i> sp. PNA 03-3	PRJNA439974	4,93	Planta	No patógeno
2765236069	<i>Pantoea ananatis</i> PNA 86-1	PRJNA439961	4,82	Planta	No patógeno
2765236086	<i>Pantoea ananatis</i> PNA 14-1	PRJNA439968	5,05	Planta	No patógeno
2765236466	<i>Pantoea</i> sp. PNA 14-12	PRJNA439975	4,67	Planta	No patógeno
2765236468	<i>Pantoea ananatis</i> PANS 200-1	PRJNA439969	4,82	Planta	No patógeno
2767802301	<i>Pantoea allii</i> PNA 02-18	PRJNA439973	4,86	Planta	No patógeno
2767802443	<i>Pantoea ananatis</i> PNA 07-1	PRJNA439963	4,90	Planta	No patógeno
2767802444	<i>Pantoea allii</i> PNG 97-1	PRJNA439970	4,87	Planta	No patógeno
2767802445	<i>Pantoea ananatis</i> PANS 02-01	PRJNA439964	4,99	Planta	No patógeno
2767802446	<i>Pantoea agglomerans</i> PNG 92-11	PRJNA439971	5,06	Planta	No patógeno
2767802447	<i>Pantoea allii</i> PNA 200-10	PRJNA439972	5,00	Planta	No patógeno
2767802457	<i>Pantoea ananatis</i> PNA 200-7	PRJNA439962	4,95	Planta	No patógeno
2767802458	<i>Pantoea ananatis</i> PNA 11-1	PRJNA439966	4,89	Planta	No patógeno
2767802459	<i>Pantoea ananatis</i> PNA 98-11	PRJNA439967	5,04	Planta	No patógeno
2767802468	<i>Pantoea</i> sp. AG702	PRJNA463049	4,75	Planta	No patógeno
2767802607	<i>Pantoea ananatis</i> PNA 07-10	PRJNA439965	4,90	Planta	No patógeno
2775506726	<i>Pantoea</i> sp. AG1095	PRJNA456366	4,91	Planta	No patógeno
2775507025	<i>Pantoea</i> sp. OXWO6B1	PRJNA312098	5,68	Planta	No patógeno
2775507026	<i>Pantoea</i> sp. FDAARGOS_194	PRJNA231221	4,88	Animal	No patógeno
2791355154	<i>Pantoea</i> sp. ICBG 828	PRJNA429668	5,02	Hongo	Desconocido
2791355156	<i>Pantoea conspicua</i> LMG 24534	PRJNA252986	4,30	Animal	Patógeno

2791355157	<i>Pantoea conspicua</i> IF5SW-P1	PRJNA319366	5,03	Ambiental	No patógeno
2791355159	<i>Pantoea latae</i> AS1	PRJNA376797	4,95	Planta	No patógeno
2791355160	<i>Pantoea coffeiphila</i> 342	PRJNA412994	5,85	Planta	Patógeno
2791355163	<i>Pantoea</i> sp. ICBG 985	PRJNA429667	4,93	Artrópodo	No patógeno
2791355164	<i>Pantoea</i> sp. ICBG 1758	PRJNA429666	4,17	Artrópodo	No patógeno
2791355165	<i>Pantoea cypripedii</i> LMG 2657	PRJNA252995	6,55	Planta	Patógeno
2791355166	<i>Pantoea deleyi</i> LMG 24200	PRJNA252987	4,58	Planta	Patógeno
2791355168	<i>Pantoea breunneri</i> LMG 5343	PRJNA252984	4,91	Animal	Patógeno
2795385499	<i>Pantoea septica</i> LMG 5345	PRJNA252991	4,30	Animal	Patógeno
2799112224	<i>Pantoea</i> sp. BK028	PRJNA519388	4,83	Planta	No patógeno
2802429580	<i>Pantoea rodasii</i> LMG 26273	PRJNA252990	5,74	Planta	Patógeno
2802429621	<i>Pantoea rwandensis</i> LMG 26275	PRJNA252996	5,77	Planta	Patógeno
2808606414	<i>Pantoea</i> sp. SJZ147	PRJNA536260	4,92	Planta	No patógeno
2811995295	<i>Pantoea vagans</i> 848_PVAG	PRJNA267549	4,94	Animal	Patógeno
2811995300	<i>Pantoea vagans</i> MP7	PRJNA254769	4,60	Artrópodo	No patógeno
2814123165	<i>Pantoea allii</i> LMG 24248	PRJNA252983	5,00	Planta	Patógeno
2816332477	<i>Pantoea vagans</i> PV989	PRJNA445847	4,84	Animal	Patógeno
2821047172	<i>Pantoea</i> sp. AN62	PRJNA547112	4,97	Planta	No patógeno
2821051985	<i>Pantoea agglomerans</i> 57a	PRJNA547114	4,80	Planta	No patógeno
2821056566	<i>Pantoea agglomerans</i> 62d	PRJNA547115	4,74	Planta	No patógeno
2821061094	<i>Pantoea agglomerans</i> 62e	PRJNA547116	4,74	Planta	No patógeno
2821065626	<i>Pantoea agglomerans</i> MR19	PRJNA547117	5,01	Planta	No patógeno
2821070448	<i>Pantoea agglomerans</i> MR11	PRJNA547118	5,01	Planta	No patógeno
2821075268	<i>Pantoea agglomerans</i> MR10	PRJNA547119	5,01	Planta	No patógeno
2821717141	<i>Pantoea agglomerans</i> 69d	PRJNA547113	5,02	Planta	No patógeno
2824495342	<i>Pantoea</i> sp. Alfr5	PRJNA547171	4,47	Planta	No patógeno
2824583583	<i>Pantoea agglomerans</i> WP7	PRJNA547130	4,79	Planta	No patógeno
2824860516	<i>Pantoea</i> sp. Alfr4	PRJNA547170	4,47	Planta	No patógeno
2844050306	<i>Pantoea vagans</i> LMG 24199	PRJNA505269	4,79	Planta	Patógeno

2844054866	<i>Pantoea</i> sp. SO10	PRJNA540191	5,10	Ambiental	Desconocido
2847782834	<i>Pantoea agglomerans</i> TH81	PRJNA486198	4,98	Ambiental	Desconocido
2847787641	<i>Pantoea agglomerans</i> CFSAN047153	PRJNA243331	5,16	Planta	Desconocido
2847792636	<i>Pantoea ananatis</i> SGAir0210	PRJNA388547	4,81	Ambiental	Desconocido
2847797336	<i>Pantoea ananatis</i> NN08200	PRJNA514184	5,18	Planta	No patógeno
2847802446	<i>Pantoea agglomerans</i> FDAARGOS 160	PRJNA231221	4,81	Animal	Desconocido
2847807205	<i>Pantoea agglomerans</i> CFSAN047154	PRJNA243331	5,16	Planta	Patógeno
2847812201	<i>Pantoea vagans</i> FBS135	PRJNA383128	4,95	Planta	No patógeno
2847817077	<i>Pantoea ananatis</i> YJ76	PRJNA382605	5,15	Planta	No patógeno
2871727361	<i>Pantoea</i> sp. 1.19	PRJNA356250	3,79	Planta	No patógeno
2871731064	<i>Pantoea plantarum</i> LMG 16222	PRJNA498371	4,98	Planta	No patógeno
2871735898	<i>Pantoea ananatis</i> 1.38	PRJNA393047	4,87	Planta	No patógeno
2871740659	<i>Pantoea agglomerans</i> pv. betae 4188	PRJNA320975	5,00	Planta	No patógeno
2871745582	<i>Pantoea</i> sp. PSNIH6	PRJNA430813	5,35	Ambiental	Desconocido
2871750963	<i>Pantoea</i> sp. AV62	PRJNA386770	4,97	Planta	No patógeno
2871755987	<i>Pantoea agglomerans</i> pv. gypsophilae 824-1	PRJNA320975	4,99	Planta	No patógeno
2871760914	<i>Pantoea ananatis</i> PANS 01-2	PRJNA394886	4,92	Artrópodo	No patógeno
2871765707	<i>Pantoea rodasii</i> DSM 26611	PRJNA420399	5,75	Plant	No patógeno
2871771314	<i>Pantoea</i> sp. Ae16	PRJNA338688	4,00	Artrópodo	No patógeno
2874381966	<i>Pantoea</i> sp. RIT388	PRJNA327264	5,01	Planta	No patógeno
2874386776	<i>Pantoea ananatis</i> PANS 99-23	PRJNA394879	4,94	Planta	No patógeno
2874391575	<i>Pantoea ananatis</i> PNA 06-1	PRJNA394894	4,98	Planta	No patógeno
2874396476	<i>Pantoea</i> sp. AMG 501	PRJNA389525	5,10	Planta	No patógeno
2874401498	<i>Pantoea</i> sp. 3_1284	PRJNA479572	4,07	Ambiental	Patógeno
2874405593	<i>Pantoea ananatis</i> PNA 99-7	PRJNA394888	4,97	Planta	No patógeno
2874410500	<i>Pantoea ananatis</i> PNA 15-1	PRJNA394895	5,00	Planta	No patógeno
2874415362	<i>Pantoea ananatis</i> PANS 04-2	PRJNA394887	4,92	Artrópodo	No patógeno
2874420152	<i>Pantoea ananatis</i> PANS 99-36	PRJNA394885	4,86	Planta	No patógeno

2874424882	<i>Pantoea</i> sp. BRM17	PRJNA414266	3,27	Ambiental	No patógeno
2876561215	<i>Pantoea vagans</i> PaVv9	PRJEB8259	9,75	Planta	Patógeno
2876571425	<i>Pantoea vagans</i> PaVv7	PRJEB8258	9,75	Planta	Patógeno
2876581635	<i>Pantoea ananatis</i> PANS 99-3	PRJNA394878	5,11	Planta	No patógeno
2876586586	<i>Pantoea agglomerans</i> BD 1212	PRJNA481880	4,88	Planta	No patógeno
2876591303	<i>Pantoea agglomerans</i> BD 1274	PRJNA481880	4,97	Planta	Patógeno
2876596178	<i>Pantoea ananatis</i> PNA 200-3	PRJNA394893	5,01	Planta	No patógeno
2876601092	<i>Pantoea endophytica</i> 596	PRJNA422728	5,11	Planta	No patógeno
2876605999	<i>Pantoea agglomerans</i> NCTC 10500	PRJEB6403	4,66	Animal	Desconocido
2881175132	<i>Pantoea</i> sp. LMG 27579	PRJNA244966	4,06	Hongo	Patógeno
2881609920	<i>Pantoea</i> sp. ARC607	PRJNA399065	4,41	Planta	No patógeno
2888264944	<i>Pantoea</i> sp. CCBC3-3-1	PRJNA506501	5,16	Planta	No patógeno
2888270003	<i>Pantoea</i> sp. MSR2	PRJNA416796	5,72	Planta	No patógeno
2888275480	<i>Pantoea</i> sp. 201603H	PRJNA496586	4,15	Hongo	Patógeno
2891657896	<i>Pantoea</i> sp. ARC270	PRJNA399064	4,40	Planta	No patógeno
2903071575	<i>Pantoea</i> sp. Bo_3	PRJNA563888	4,92	Planta	Patógeno
2903076451	<i>Pantoea</i> sp. M_8	PRJNA563888	4,65	Planta	Patógeno
2903080973	<i>Pantoea</i> sp. F_14	PRJNA563888	4,87	Planta	Patógeno
2903085801	<i>Pantoea</i> sp. VH_4	PRJNA563888	4,93	Planta	Patógeno
2903090699	<i>Pantoea</i> sp. VH_3	PRJNA563888	4,90	Planta	Patógeno
2903095552	<i>Pantoea</i> sp. EKM101V	PRJNA605565	4,59	Planta	No patógeno
2903100013	<i>Pantoea</i> sp. F_17	PRJNA563888	4,87	Planta	Patógeno
2903104845	<i>Pantoea</i> sp. F_18	PRJNA563888	4,87	Planta	Patógeno
2903109678	<i>Pantoea</i> sp. Acro-835	PRJNA564151	5,42	Artrópodo	No patógeno
2903114778	<i>Pantoea</i> sp. SoET	PRJNA413893	0,84	Artrópodo	No patógeno
2903115531	<i>Pantoea</i> sp. Al-1710	PRJNA564151	5,86	Artrópodo	No patógeno
2903121377	<i>Pantoea</i> sp. Taur	PRJNA413893	5,46	Artrópodo	No patógeno
2903126686	<i>Pantoea</i> sp. Bo_46	PRJNA563888	4,92	Planta	Patógeno
2904652709	<i>Pantoea</i> sp. Eser	PRJNA413893	3,93	Artrópodo	No patógeno

2904658405	<i>Pantoea</i> sp. Nvir	PRJNA413893	1,43	Artrópodo	No patógeno
2905874158	<i>Pantoea</i> sp. F_16	PRJNA563888	4,88	Planta	Patógeno
2905879010	<i>Pantoea</i> sp. Tr-811	PRJNA564151	6,66	Artrópodo	No patógeno
2905885049	<i>Pantoea</i> sp. FN_2b	PRJNA563888	4,92	Planta	Patógeno
2905889911	<i>Pantoea</i> sp. Bo_5	PRJNA563888	4,92	Planta	Patógeno
2905894795	<i>Pantoea</i> sp. VH_16	PRJNA563888	4,90	Planta	Patógeno
2905899716	<i>Pantoea</i> sp. Ap-959	PRJNA564151	5,03	Hongo	No patógeno
2905904552	<i>Pantoea</i> sp. Bo_2	PRJNA563888	4,92	Planta	Patógeno
2905909436	<i>Pantoea</i> sp. SoEX	PRJNA413893	0,83	Artrópodo	No patógeno
2905910181	<i>Pantoea</i> sp. F_5	PRJNA563888	4,87	Planta	Patógeno
2905915013	<i>Pantoea</i> sp. Bo_40	PRJNA563888	4,92	Planta	Patógeno
2905931062	<i>Pantoea</i> sp. Acro-807	PRJNA564151	3,98	Hongo	No patógeno
2905934905	<i>Pantoea</i> sp. M_10	PRJNA563888	4,65	Planta	Patógeno
2905939430	<i>Pantoea</i> sp. EKM10T	PRJNA605402	4,71	Planta	No patógeno
2905943963	<i>Pantoea</i> sp. SoEL	PRJNA413893	0,84	Artrópodo	No patógeno
2905947761	<i>Pantoea</i> sp. B_8	PRJNA563888	4,94	Planta	Patógeno
2905952654	<i>Pantoea</i> sp. M_6	PRJNA563888	4,65	Planta	Patógeno
2905957205	<i>Pantoea</i> sp. M_1	PRJNA563888	4,87	Planta	Patógeno
2905962070	<i>Pantoea</i> sp. VH_8	PRJNA563888	4,92	Planta	Patógeno
2905966976	<i>Pantoea</i> sp. Bo_19	PRJNA563888	4,92	Planta	Patógeno
2908669403	<i>Pantoea coffeiphila</i> 1480	PRJNA695719	5,74	Planta	No patógeno
2908725932	<i>Pantoea</i> sp. Mhis	PRJNA413893	1,03	Artrópodo	No patógeno
2918543456	<i>Pantoea agglomerans</i> JCM1236	PRJNA709560	4,94	Ambiental	Desconocido
2919812101	<i>Pantoea</i> alhagi BE291	PRJNA709311	4,56	Planta	No patógeno
2921922511	<i>Pantoea</i> sp. Psp39-30	PRJNA658485	4,98	Planta	No patógeno
2921928104	<i>Pantoea</i> sp. M_9	PRJNA563888	3,91	Planta	Patógeno
2921931888	<i>Pantoea</i> sp. Aalb	PRJNA413893	1,14	Artrópodo	No patógeno
2921932851	<i>Pantoea</i> sp. Ap-967	PRJNA564151	6,67	Artrópodo	No patógeno
2921938947	<i>Pantoea</i> sp. Bo_21	PRJNA563888	4,92	Planta	Patógeno

2921943832	<i>Pantoea</i> sp. VH_18	PRJNA563888	4,87	Planta	Patógeno
2921948699	<i>Pantoea</i> sp. Bo_7	PRJNA563888	3,87	Planta	Patógeno
2921952440	<i>Pantoea</i> sp. EKM103V	PRJNA605565	4,69	Planta	No patógeno
2921957013	<i>Pantoea</i> sp. Bo_34a	PRJNA563888	4,92	Planta	Patógeno
2921961884	<i>Pantoea</i> sp. F_11	PRJNA563888	4,87	Planta	Patógeno
2923634449	<i>Pantoea agglomerans</i> SLBN-76	PRJNA708487	4,75	Planta	No patógeno
2924459726	<i>Pantoea</i> sp. paga	PRJNA555164	5,08	Planta	No patógeno
2924464632	<i>Pantoea</i> sp. VH_24	PRJNA563888	4,87	Planta	Patógeno
2924469488	<i>Pantoea</i> sp. F_7	PRJNA563888	4,87	Planta	Patógeno
2924474320	<i>Pantoea</i> sp. BAV 3049	PRJNA576872	5,31	Planta	No patógeno
2924479673	<i>Pantoea</i> sp. Bo_11	PRJNA563888	4,90	Planta	Patógeno
2924484564	<i>Pantoea</i> sp. EKM21T	PRJNA605422	4,87	Planta	No patógeno
2924489260	<i>Pantoea</i> sp. 18059	PRJEB34793	4,73	Planta	No patógeno
2924493850	<i>Pantoea</i> sp. Acro-805	PRJNA564151	5,90	Planta	No patógeno
2924499653	<i>Pantoea dispersa</i> 625	PRJNA232278	4,74	Planta	No patógeno
2924504186	<i>Pantoea</i> sp. M_5	PRJNA563888	5,40	Planta	Patógeno
2924510081	<i>Pantoea</i> sp. Seng	PRJNA413893	5,05	Planta	No patógeno
2924515017	<i>Pantoea</i> sp. B9002	PRJNA607442	5,57	Planta	Patógeno
2924520353	<i>Pantoea</i> sp. Bo_39	PRJNA563888	4,92	Planta	Patógeno
2927986592	<i>Pantoea agglomerans</i> PVP026	-	5,17	Planta	No patógeno
2929943546	<i>Pantoea</i> sp. UYEF8	PRJNA1115399	4,94	Ambiental	Desconocido
2935373448	<i>Pantoea agglomerans</i> PVP012	PRJNA1115477	5,14	Planta	No patógeno
2935378324	<i>Pantoea agglomerans</i> PVP021	-	5,14	Planta	No patógeno
2937529543	<i>Pantoea</i> sp. F_12	PRJNA563888	4,87	Planta	Patógeno
2937534371	<i>Pantoea</i> sp. Ap-870	PRJNA564151	4,98	Artrópodo	No patógeno
2937539155	<i>Pantoea</i> sp. SoEE	PRJNA413893	0,85	Artrópodo	No patógeno
2937539931	<i>Pantoea</i> sp. LS15	PRJNA615287	4,64	Planta	No patógeno
2937544590	<i>Pantoea</i> sp. RIT 413	PRJNA449741	3,84	Planta	Desconocido
2937548250	<i>Pantoea</i> sp. R102	PRJNA515589	4,55	Artrópodo	No patógeno

2937552856	<i>Pantoea</i> sp. Bo_10	PRJNA563888	3,87	Planta	Patógeno
2937556594	<i>Pantoea</i> sp. B_9	PRJNA563888	4,89	Planta	Patógeno
2937561330	<i>Pantoea</i> sp. VH_25	PRJNA563888	4,91	Planta	Patógeno
2937566189	<i>Pantoea</i> sp. EKM22T	PRJNA605422	4,87	Planta	No patógeno
2937570898	<i>Pantoea</i> sp. B_2	PRJNA563888	4,92	Planta	Patógeno
2937575775	<i>Pantoea</i> sp. M_4	PRJNA563888	4,91	Planta	Patógeno
2937580670	<i>Pantoea</i> sp. VS1	PRJNA387371	5,41	Planta	No patógeno
2937640648	<i>Pantoea hericii</i> JZB2120024	PRJNA515154	4,01	Planta	Patógeno
2937644537	<i>Pantoea</i> sp. SoEO	PRJNA413893	0,84	Artrópodo	No patógeno
2937645278	<i>Pantoea</i> sp. Cy-640	PRJNA564151	5,06	Artrópodo	No patógeno
2937650227	<i>Pantoea</i> sp. M_3	PRJNA563888	4,91	Planta	Patógeno
2937655109	<i>Pantoea</i> sp. Bo_14	PRJNA563888	4,90	Planta	Patógeno
2937659990	<i>Pantoea</i> sp. F_15	PRJNA563888	4,86	Planta	Patógeno
2937664816	<i>Pantoea</i> sp. Bo_13	PRJNA563888	4,92	Planta	Patógeno
2939602548	<i>Pantoea dispersa</i> 1175	-	4,95	Planta	No patógeno
2957890251	<i>Pantoea</i> sp. B_10	PRJNA563888	4,99	Planta	Patógeno
2957895252	<i>Pantoea</i> sp. WMus005	PRJNA643572	4,49	Animal	No patógeno
2957899578	<i>Pantoea</i> sp. EKM20T	PRJNA605422	4,88	Planta	No patógeno
2957904330	<i>Pantoea pleuroti</i> JZB 2120015	PRJNA513927	4,88	Hongo	Patógeno
2065487014_2	<i>Pantoea agglomerans</i> -2065487014_2	-	4,30	Artrópodo	Desconocido
3300002953_5	<i>Pantoea</i> -3300002953_5	PRJNA365939	2,08	Ambiental	Desconocido
3300003838_10	<i>Pantoea</i> sp. NGS-ED-1003-3300003838_10	PRJNA340595	2,82	Planta	Desconocido
3300006086_8	<i>Pantoea agglomerans</i> -3300006086_8	PRJNA367205	4,59	Ambiental	Desconocido
3300009994_1	<i>Pantoea ananatis</i> -3300009994_1	PRJNA406013	4,70	Planta	Desconocido
3300010395_9	<i>Pantoea stewartii</i> -3300010395_9	PRJNA340605	3,54	Planta	Desconocido
3300011140_1	<i>Pantoea</i> sp. 3.5.1-3300011140_1	PRJNA271013	4,80	Ambiental	Desconocido
3300011158_4	<i>Pantoea</i> sp. FF5-3300011158_4	PRJNA271013	3,58	Ambiental	Desconocido
3300011158_5	<i>Pantoea agglomerans</i> -3300011158_5	PRJNA271013	3,30	Ambiental	Desconocido
3300011210_2	<i>Pantoea vagans</i> -3300011210_2	PRJNA271013	4,70	Ambiental	Desconocido

3300011215_4	<i>Pantoea</i> -3300011215_4	PRJNA271013	4,37	Ambiental	Desconocido
3300011388_1	<i>Pantoea</i> sp. FF5-3300011388_1	PRJNA271014	4,21	Ambiental	Desconocido
3300011401_2	<i>Pantoea vagans</i> -3300011401_2	PRJNA271015	2,82	Ambiental	Desconocido
3300011453_1	<i>Pantoea</i> -3300011453_1	PRJNA271016	4,48	Ambiental	Desconocido
3300011459_2	<i>Pantoea vagans</i> -3300011459_2	PRJNA271017	4,92	Ambiental	Desconocido
3300011471_1	<i>Pantoea</i> -3300011471_1	PRJNA271018	4,64	Ambiental	Desconocido
3300011474_1	<i>Pantoea</i> sp. Sc1-3300011474_1	PRJNA271019	4,21	Ambiental	Desconocido
3300011500_2	<i>Pantoea</i> sp. NGS-ED-1003-3300011500_2	PRJNA271020	4,54	Ambiental	Desconocido
3300011504_3	<i>Pantoea</i> sp. PSNIH2-3300011504_3	PRJNA271021	4,08	Ambiental	Desconocido
3300011530_4	<i>Pantoea</i> sp. PSNIH2-3300011530_4	PRJNA271022	3,01	Ambiental	Desconocido
3300011540_2	<i>Pantoea</i> sp. PSNIH2-3300011540_2	PRJNA271023	4,04	Ambiental	Desconocido
3300011577_2	<i>Pantoea</i> sp. PSNIH2-3300011577_2	PRJNA271024	3,36	Ambiental	Desconocido
3300011596_1	<i>Pantoea</i> sp. PSNIH2-3300011596_1	PRJNA271025	4,08	Ambiental	Desconocido
3300011604_1	<i>Pantoea agglomerans</i> -3300011604_1	PRJNA271026	4,23	Ambiental	Desconocido
3300011609_7	<i>Pantoea</i> sp. 3.5.1-3300011609_7	PRJNA271027	1,79	Ambiental	Desconocido
3300011614_2	<i>Pantoea</i> -3300011614_2	PRJNA271028	3,82	Ambiental	Desconocido
3300011617_2	<i>Pantoea</i> sp. PSNIH2-3300011617_2	PRJNA271029	4,18	Ambiental	Desconocido
3300011618_3	<i>Pantoea</i> sp. aB-3300011618_3	PRJNA271030	3,47	Ambiental	Desconocido
3300011626_2	<i>Pantoea</i> -3300011626_2	PRJNA271031	4,42	Ambiental	Desconocido
3300011699_3	<i>Pantoea</i> sp. PSNIH2-3300011699_3	PRJNA271032	3,60	Ambiental	Desconocido
3300011707_1	<i>Pantoea</i> -3300011707_1	PRJNA271033	4,35	Ambiental	Desconocido
3300011724_3	<i>Pantoea</i> sp. NGS-ED-1003-3300011724_3	PRJNA271034	4,17	Ambiental	Desconocido
3300011725_2	<i>Pantoea</i> -3300011725_2	PRJNA271035	3,83	Ambiental	Desconocido
3300011725_3	<i>Pantoea</i> sp. NGS-ED-1003-3300011725_3	PRJNA271036	3,44	Ambiental	Desconocido
3300011778_4	<i>Pantoea</i> -3300011778_4	PRJNA271037	3,36	Ambiental	Desconocido
3300011813_2	<i>Pantoea</i> sp. NGS-ED-1003-3300011813_2	PRJNA271038	3,71	Ambiental	Desconocido
3300011825_6	<i>Pantoea</i> -3300011825_6	PRJNA271039	2,93	Ambiental	Desconocido
3300011830_3	<i>Pantoea</i> sp. YR343-3300011830_3	PRJNA271040	4,79	Ambiental	Desconocido
3300011832_4	<i>Pantoea dispersa</i> -3300011832_4	PRJNA271041	3,89	Ambiental	Desconocido

3300011929_5	<i>Pantoea</i> sp. PSNIH2-3300011929_5	PRJNA271042	2,59	Ambiental	Desconocido
3300011939_3	<i>Pantoea</i> sp. PSNIH2-3300011939_3	PRJNA271043	3,73	Ambiental	Desconocido
3300011939_4	<i>Pantoea</i> sp. PSNIH1-3300011939_4	PRJNA271044	3,69	Ambiental	Desconocido
3300011945_3	<i>Pantoea</i> -3300011945_3	PRJNA271045	3,70	Ambiental	Desconocido
3300011948_2	<i>Pantoea</i> -3300011948_2	PRJNA271046	4,49	Ambiental	Desconocido
3300011954_1	<i>Pantoea</i> sp. NGS-ED-1003-3300011954_1	PRJNA271047	4,29	Ambiental	Desconocido
3300012061_3	<i>Pantoea vagans</i> -3300012061_3	PRJNA405282	3,91	Artrópodo	Desconocido
3300012078_8	<i>Pantoea vagans</i> -3300012078_8	PRJNA405276	3,76	Artrópodo	Desconocido
3300012079_9	<i>Pantoea</i> -3300012079_9	PRJNA406097	2,33	Artrópodo	Desconocido
3300012095_3	<i>Pantoea</i> -3300012095_3	PRJNA405232	4,16	Artrópodo	Desconocido
3300012125_6	<i>Pantoea</i> -3300012125_6	PRJNA405283	3,25	Artrópodo	Desconocido
3300012139_2	<i>Pantoea</i> -3300012139_2	PRJNA405284	4,37	Artrópodo	Desconocido
3300012139_3	<i>Pantoea vagans</i> -3300012139_3	PRJNA405238	4,34	Artrópodo	Desconocido
3300012153_14	<i>Pantoea vagans</i> -3300012153_14	PRJNA405257	2,14	Artrópodo	Desconocido
3300012253_4	<i>Pantoea agglomerans</i> -3300012253_4	PRJNA271013	2,60	Ambiental	Desconocido
3300012674_2	<i>Pantoea</i> sp. NGS-ED-1003-3300012674_2	PRJNA405239	3,55	Artrópodo	Desconocido
3300012839_8	<i>Pantoea</i> -3300012839_8	PRJNA405153	4,39	Artrópodo	Desconocido
3300012845_5	<i>Pantoea</i> -3300012845_5	PRJNA405150	4,85	Artrópodo	Desconocido
3300012854_7	<i>Pantoea</i> -3300012854_7	PRJNA405147	3,58	Artrópodo	Desconocido
3300012994_4	<i>Pantoea</i> sp. Sc1-3300012994_4	PRJNA406130	2,85	Artrópodo	Desconocido
3300012994_5	<i>Pantoea ananatis</i> -3300012994_5	PRJNA406130	2,52	Artrópodo	Desconocido
3300012995_3	<i>Pantoea</i> -3300012995_3	PRJNA406134	5,38	Artrópodo	Desconocido
3300012995_6	<i>Pantoea agglomerans</i> -3300012995_6	PRJNA406134	4,39	Artrópodo	Desconocido
3300012995_8	<i>Pantoea</i> sp. PSNIH1-3300012995_8	PRJNA406134	2,80	Artrópodo	Desconocido
3300012995_9	<i>Pantoea</i> sp. NGS-ED-1003-3300012995_9	PRJNA406134	2,27	Artrópodo	Desconocido
3300012996_4	<i>Pantoea</i> sp. Sc1-3300012996_4	PRJNA406131	3,70	Artrópodo	Desconocido
3300012998_7	<i>Pantoea</i> -3300012998_7	PRJNA406135	3,34	Artrópodo	Desconocido
3300012999_10	<i>Pantoea stewartii</i> -3300012999_10	PRJNA406136	4,60	Artrópodo	Desconocido
3300013001_4	<i>Pantoea</i> -3300013001_4	PRJNA406138	5,41	Artrópodo	Desconocido

3300013001_5	<i>Pantoea</i> sp. Sc1-3300013001_5	PRJNA406138	3,48	Artrópodo	Desconocido
3300013023_4	<i>Pantoea</i> sp. Sc1-3300013023_4	PRJNA444373	3,82	Artrópodo	Desconocido
3300013283_6	<i>Pantoea</i> sp. FF5-3300013283_6	PRJNA271013	3,55	Ambiental	Desconocido
3300013338_2	<i>Pantoea</i> -3300013338_2	PRJNA271013	3,16	Ambiental	Desconocido
3300013352_4	<i>Pantoea</i> sp. PSNIH2-3300013352_4	PRJNA271013	3,49	Ambiental	Desconocido
3300013489_1	<i>Pantoea</i> sp. NGS-ED-1003-3300013489_1	PRJNA271013	4,55	Ambiental	Desconocido
3300013719_3	<i>Pantoea</i> sp. 3.5.1-3300013719_3	PRJNA271013	4,21	Ambiental	Desconocido
3300013723_2	<i>Pantoea</i> -3300013723_2	PRJNA271013	3,92	Ambiental	Desconocido
3300013728_2	<i>Pantoea</i> sp. PSNIH1-3300013728_2	PRJNA271013	3,91	Ambiental	Desconocido
3300013734_2	<i>Pantoea agglomerans</i> -3300013734_2	PRJNA271013	4,61	Ambiental	Desconocido
3300013821_1	<i>Pantoea</i> sp.-3300013821_1	PRJNA271013	4,75	Ambiental	Desconocido
3300013824_1	<i>Pantoea</i> -3300013824_1	PRJNA271013	4,36	Ambiental	Desconocido
3300013850_1	<i>Pantoea</i> -3300013850_1	PRJNA271013	4,56	Ambiental	Desconocido
3300013921_1	<i>Pantoea</i> sp. 3.5.1-3300013921_1	PRJNA271013	4,66	Ambiental	Desconocido
3300013932_2	<i>Pantoea</i> -3300013932_2	PRJNA271013	3,97	Ambiental	Desconocido
3300013935_2	<i>Pantoea</i> sp. 3.5.1-3300013935_2	PRJNA271013	4,71	Ambiental	Desconocido
3300013990_1	<i>Pantoea</i> -3300013990_1	PRJNA271013	5,05	Ambiental	Desconocido
3300013990_4	<i>Pantoea</i> sp. PSNIH2-3300013990_4	PRJNA271013	2,87	Ambiental	Desconocido
3300014015_1	<i>Pantoea</i> sp. YR343-3300014015_1	PRJNA271013	4,46	Ambiental	Desconocido
3300014048_4	<i>Pantoea</i> -3300014048_4	PRJNA271013	3,63	Ambiental	Desconocido
3300014287_2	<i>Pantoea</i> sp. FF5-3300014287_2	PRJNA271013	4,47	Ambiental	Desconocido
3300014526_4	<i>Pantoea</i> sp. FF5-3300014526_4	PRJNA271013	3,89	Ambiental	Desconocido
3300014588_2	<i>Pantoea</i> sp.-3300014588_2	PRJNA271013	4,45	Ambiental	Desconocido
3300014610_3	<i>Pantoea</i> -3300014610_3	PRJNA271013	3,69	Ambiental	Desconocido
3300014611_4	<i>Pantoea</i> sp. PSNIH2-3300014611_4	PRJNA271013	4,16	Ambiental	Desconocido
3300014621_1	<i>Pantoea vagans</i> -3300014621_1	PRJNA271013	4,73	Ambiental	Desconocido
3300014624_2	<i>Pantoea</i> sp. GL120224-02-3300014624_2	PRJNA271013	4,17	Ambiental	Desconocido
3300014626_1	<i>Pantoea</i> sp. FF5-3300014626_1	PRJNA271013	4,30	Ambiental	Desconocido
3300014627_2	<i>Pantoea</i> sp. PSNIH2-3300014627_2	PRJNA271013	4,04	Ambiental	Desconocido

3300014630_2	<i>Pantoea</i> -3300014630_2	PRJNA271013	4,40	Ambiental	Desconocido
3300014642_2	<i>Pantoea</i> sp. PSNIH2-3300014642_2	PRJNA271013	4,15	Ambiental	Desconocido
3300014644_2	<i>Pantoea dispersa</i> -3300014644_2	PRJNA271013	4,78	Ambiental	Desconocido
3300014646_2	<i>Pantoea</i> -3300014646_2	PRJNA271014	4,72	Ambiental	Desconocido
3300014673_1	<i>Pantoea</i> -3300014673_1	PRJNA271015	4,63	Ambiental	Desconocido
3300014736_1	<i>Pantoea</i> -3300014736_1	PRJNA271016	4,10	Ambiental	Desconocido
3300014746_3	<i>Pantoea vagans</i> -3300014746_3	PRJNA271017	2,51	Ambiental	Desconocido
3300014897_2	<i>Pantoea</i> -3300014897_2	PRJNA271018	4,33	Ambiental	Desconocido
3300014901_1	<i>Pantoea</i> -3300014901_1	PRJNA271019	4,71	Ambiental	Desconocido
3300014912_2	<i>Pantoea</i> -3300014912_2	PRJNA271020	4,79	Ambiental	Desconocido
3300023065_5	<i>Pantoea</i> -3300023065_5	PRJNA271021	4,97	Ambiental	Desconocido
3300025233_11	<i>Pantoea</i> -3300025233_11	PRJNA329923	4,78	Planta	Desconocido
3300025542_6	<i>Pantoea</i> -3300025542_6	PRJNA365937	1,89	Ambiental	Desconocido
3300025624_7	<i>Pantoea</i> -3300025624_7	PRJNA365939	2,14	Ambiental	Desconocido
3300025651_7	<i>Pantoea</i> -3300025651_7	PRJNA365938	2,59	Ambiental	Desconocido
3300025710_6	<i>Pantoea</i> -3300025710_6	PRJNA365936	3,00	Ambiental	Desconocido
3300027606_16	<i>Pantoea</i> sp. NGS-ED-1003-3300027606_16	PRJNA340595	4,06	Planta	Desconocido
3300027606_33	<i>Pantoea</i> -3300027606_33	PRJNA340595	1,96	Planta	Desconocido
3300027898_7	<i>Pantoea</i> -3300027898_7	PRJNA367205	4,42	Ambiental	Desconocido
3300030435_1	<i>Pantoea ananatis</i> -3300030435_1	PRJNA178312	4,72	Ambiental	Desconocido
3300030498_5	<i>Pantoea stewartii</i> -3300030498_5	PRJNA340605	4,31	Planta	Desconocido
3300030512_10	<i>Pantoea</i> sp. NGS-ED-1003-3300030512_10	PRJNA340595	3,81	Planta	Desconocido
3300031413_10	<i>Pantoea</i> -3300031413_10	PRJNA518638	6,90	Planta	Desconocido
3300042934_50	<i>Pantoea agglomerans</i> -3300042934_50	PRJNA708766	3,24	Ambiental	Desconocido
3300042934_69	<i>Pantoea</i> -3300042934_69	PRJNA708766	3,51	Ambiental	Desconocido
3300045450_57	<i>Pantoea ananatis</i> -3300045450_57	-	4,54	Planta	Desconocido
3300048848_74	<i>Pantoea</i> -3300048848_74	-	4,31	Ambiental	Desconocido
3300048849_22	<i>Pantoea agglomerans</i> -3300048849_22	-	4,73	Ambiental	Desconocido
3300048856_52	<i>Pantoea</i> -3300048856_52	-	4,45	Ambiental	Desconocido

3300048904_51	<i>Pantoea</i> sp. AG1095-3300048904_51	-	4,78	Planta	Desconocido
3300048919_30	<i>Pantoea</i> sp. PSNIH1-3300048919_30	-	3,25	Planta	Desconocido
3300048919_65	<i>Pantoea ananatis</i> -3300048919_65	-	4,44	Planta	Desconocido
3300048920_88	<i>Pantoea ananatis</i> -3300048920_88	-	4,01	Planta	Desconocido
3300048921_22	<i>Pantoea agglomerans</i> -3300048921_22	-	4,01	Planta	Desconocido
3300048921_28	<i>Pantoea ananatis</i> -3300048921_28	-	2,63	Planta	Desconocido
3300048921_62	<i>Pantoea</i> sp. aB-3300048921_62	-	3,87	Planta	Desconocido
3300048922_82	<i>Pantoea agglomerans</i> -3300048922_82	-	3,80	Planta	Desconocido
3300048923_51	<i>Pantoea</i> sp. AG1095-3300048923_51	-	4,75	Planta	Desconocido
3300048923_52	<i>Pantoea</i> sp. PSNIH1-3300048923_52	-	3,77	Planta	Desconocido
3300048924_61	<i>Pantoea agglomerans</i> -3300048924_61	-	4,52	Planta	Desconocido
3300048925_89	<i>Pantoea</i> sp. AG1095-3300048925_89	-	3,51	Planta	Desconocido
3300048925_9	<i>Pantoea</i> -3300048925_9	-	4,07	Planta	Desconocido
3300048926_16	<i>Pantoea</i> sp. AG1095-3300048926_16	-	3,70	Planta	Desconocido
3300048927_110	<i>Pantoea agglomerans</i> -3300048927_110	-	3,81	Planta	Desconocido
3300048927_20	<i>Pantoea</i> sp. PSNIH1-3300048927_20	-	3,35	Planta	Desconocido
3300048928_64	<i>Pantoea agglomerans</i> -3300048928_64	-	3,22	Planta	Desconocido
-	<i>Pantoea agglomerans</i> LPU12	-	0,00	Planta	No patógeno



BIBLIOGRAFÍA



BIBLIOGRAFÍA

- A. D. Bradshaw and M. J. Chadwick. (1980). *The Restoration of Land: The Ecology and Reclamation of Derelict and Degraded Land*. University of California Press.
- A. Ghomi, F., Jung, J. J., Langridge, G. C., Cain, A. K., Boinett, C. J., Abd El Ghany, M., Pickard, D. J., Kingsley, R. A., Thomson, N. R., Parkhill, J., Gardner, P. P., & Barquist, L. (2024). High-throughput transposon mutagenesis in the family Enterobacteriaceae reveals core essential genes and rapid turnover of essentiality. *MBio*, 15(10). <https://doi.org/10.1128/mbio.01798-24>
- Abeles, F. B., Morgan, P. W., & Saltveit Jr, M. E. (2012). *Ethylene in plant biology*. Academic press.
- Afzal, I., Shinwari, Z. K., Sikandar, S., & Shahzad, S. (2019). Plant beneficial endophytic bacteria: Mechanisms, diversity, host range and genetic determinants. *Microbiological Research*, 221, 36–49. <https://doi.org/10.1016/j.micres.2019.02.001>
- AGARWAL, S. (1987). Tetrazolium reducing microorganisms inside the root of Brassica species. *Curr. Sci.*, 56, 187–188.
- Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S.-T., Weigel, D., & Kemen, E. M. (2016). Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLOS Biology*, 14(1), e1002352. <https://doi.org/10.1371/journal.pbio.1002352>
- AGRIOS, G. N. (2005). PLANT DISEASES CAUSED BY PROKARYOTES: BACTERIA AND MOLLICUTES. In *Plant Pathology* (pp. 615–703). Elsevier. <https://doi.org/10.1016/B978-0-08-047378-9.50018-X>
- Agrios, G. N. (2009). Plant Pathogens and Disease: General Introduction. In *Encyclopedia of Microbiology* (pp. 613–646). Elsevier. <https://doi.org/10.1016/B978-012373944-5.00344-8>
- Alcaraz, L. D., Peimbert, M., Barajas, H. R., Dorantes-Acosta, A. E., Bowman, J. L., & Arteaga-Vázquez, M. A. (2018). Marchantia liverworts as a proxy to plants' basal microbiomes. *Scientific Reports*, 8(1), 12712. <https://doi.org/10.1038/s41598-018-31168-0>
- Alekshun, M. N., & Levy, S. B. (2007). Molecular Mechanisms of Antibacterial Multidrug Resistance. *Cell*, 128(6), 1037–1050. <https://doi.org/10.1016/j.cell.2007.03.004>
- Ali, S., Duan, J., Charles, T. C., & Glick, B. R. (2014). A bioinformatics approach to the determination of genes involved in endophytic behavior in Burkholderia spp. *Journal of Theoretical Biology*, 343, 193–198. <https://doi.org/10.1016/j.jtbi.2013.10.007>
- Arnold, M. F. F., Shabab, M., Penterman, J., Boehme, K. L., Griffiths, J. S., & Walker, G. C. (2017). Genome-Wide Sensitivity Analysis of the Microsymbiont *Sinorhizobium meliloti* to Symbiotically Important, Defensin-Like Host Peptides. *MBio*, 8(4). <https://doi.org/10.1128/mBio.01060-17>

- Arshad, M., Saleem, M., & Hussain, S. (2007). Perspectives of bacterial ACC deaminase in phytoremediation. *Trends in Biotechnology*, 25(8), 356–362. <https://doi.org/10.1016/j.tibtech.2007.05.005>
- Astudillo-García, C., Bell, J. J., Webster, N. S., Glasl, B., Jompa, J., Montoya, J. M., & Taylor, M. W. (2017). Evaluating the core microbiota in complex communities: A systematic investigation. *Environmental Microbiology*, 19(4), 1450–1462. <https://doi.org/10.1111/1462-2920.13647>
- Atamna-Ismaeel, N., Finkel, O., Glaser, F., von Mering, C., Vorholt, J. A., Koblížek, M., Belkin, S., & Bèjà, O. (2012). Bacterial anoxygenic photosynthesis on plant leaf surfaces. *Environmental Microbiology Reports*, 4(2), 209–216. <https://doi.org/10.1111/j.1758-2229.2011.00323.x>
- Atamna-Ismaeel, N., Finkel, O. M., Glaser, F., Sharon, I., Schneider, R., Post, A. F., Spudich, J. L., von Mering, C., Vorholt, J. A., Iluz, D., Bèjà, O., & Belkin, S. (2012). Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environmental Microbiology*, 14(1), 140–146. <https://doi.org/10.1111/j.1462-2920.2011.02554.x>
- Avis, T. J., Gravel, V., Antoun, H., & Tweddell, R. J. (2008). Multifaceted beneficial effects of rhizosphere microorganisms on plant health and productivity. *Soil Biology and Biochemistry*, 40(7), 1733–1740. <https://doi.org/10.1016/j.soilbio.2008.02.013>
- Bacilio-Jiménez, M., Aguilar-Flores, S., Ventura-Zapata, E., Pérez-Campos, E., Bouquelet, S., & Zenteno, E. (2003). Chemical characterization of root exudates from rice (*Oryza sativa*) and their effects on the chemotactic response of endophytic bacteria. *Plant and Soil*, 249(2), 271–277. <https://doi.org/10.1023/A:1022888900465>
- Badri, D. V., Weir, T. L., van der Lelie, D., & Vivanco, J. M. (2009). Rhizosphere chemical dialogues: plant–microbe interactions. *Current Opinion in Biotechnology*, 20(6), 642–650. <https://doi.org/10.1016/j.copbio.2009.09.014>
- Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S., & Vivanco, J. M. (2006). THE ROLE OF ROOT EXUDATES IN RHIZOSPHERE INTERACTIONS WITH PLANTS AND OTHER ORGANISMS. *Annual Review of Plant Biology*, 57(1), 233–266. <https://doi.org/10.1146/annurev.arplant.57.032905.105159>
- Balachandar, D., Sandhiya, G. S., Sugitha, T. C. K., & Kumar, K. (2006). Flavonoids and Growth Hormones Influence Endophytic Colonization and in Planta Nitrogen Fixation by a Diazotrophic *Serratia* sp. in Rice. *World Journal of Microbiology and Biotechnology*, 22(7), 707–712. <https://doi.org/10.1007/s11274-005-9094-0>
- Bangert, R. L., Cho, B. R., Widders, P. R., Stauber, E. H., & Ward, A. C. S. (1988). A Survey of Aerobic Bacteria and Fungi in the Feces of Healthy Psittacine Birds. *Avian Diseases*, 32(1), 46. <https://doi.org/10.2307/1590947>
- Barash, I., & Manulis-Sasson, S. (2009). Recent Evolution of Bacterial Pathogens: The Gall-Forming *Pantoea agglomerans* Case. *Annual Review of Phytopathology*, 47(1), 133–152. <https://doi.org/10.1146/annurev-phyto-080508-081803>
- Barriuso, J., Ramos Solano, B., Lucas, J. A., Lobo, A. P., García-Villaraco, A., & Gutiérrez Mañero, F. J. (2008). Ecology, Genetic Diversity and Screening Strategies of Plant Growth Promoting

- Rhizobacteria (PGPR). In *Plant-Bacteria Interactions* (pp. 1–17). Wiley. <https://doi.org/10.1002/9783527621989.ch1>
- Basigalup, D. H. (2022). *Investigación, producción e industrialización de la alfalfa en Argentina*. INTA Ediciones.
- Beadle, G. W., & Tatum, E. L. (1941). Genetic Control of Biochemical Reactions in *Neurospora*. *Proceedings of the National Academy of Sciences of the United States of America*, 27(11), 499–506. <https://doi.org/10.1073/pnas.27.11.499>
- Beerling, D. (2007). *The Emerald Planet*. Oxford University Press. <https://doi.org/10.1093/oso/9780192806024.001.0001>
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., ... Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53–59. <https://doi.org/10.1038/nature07517>
- BERINGER, J. E. (1974). R Factor Transfer in *Rhizobium leguminosarum*. *Microbiology*, 84(1), 188–198. <https://doi.org/10.1099/00221287-84-1-188>
- Blomberg, S. P., Garland, T., & Ives, A. R. (2003). TESTING FOR PHYLOGENETIC SIGNAL IN COMPARATIVE DATA: BEHAVIORAL TRAITS ARE MORE LABILE. *Evolution*, 57(4), 717–745. <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>
- Boller, T., & Felix, G. (2009). A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annual Review of Plant Biology*, 60(1), 379–406. <https://doi.org/10.1146/annurev.arplant.57.032905.105346>
- Brady, C., Cleenwerck, I., Venter, S., Vancanneyt, M., Swings, J., & Coutinho, T. (2008). Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Systematic and Applied Microbiology*, 31(6–8), 447–460. <https://doi.org/10.1016/j.syapm.2008.09.004>
- Brandi, M., Clark, E. M., & Lindow, S. E. (1996). Characterization of the indole-3-acetic acid (IAA) biosynthetic pathway in an epiphytic strain of *Erwinia herbicola* and IAA production in vitro. *Canadian Journal of Microbiology*, 42(6), 586–592. <https://doi.org/10.1139/m96-079>
- Brandl, M., & Lindow, S. (1998). Contribution of indole-3-acetic acid production to the epiphytic fitness of *erwinia herbicola*. *Applied and Environmental Microbiology*, 64(9), 3256–3263. <https://doi.org/10.1128/AEM.64.9.3256-3263.1998>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bulgarelli, D., Garrido-Oter, R., Münch, P. C., Weiman, A., Dröge, J., Pan, Y., McHardy, A. C., & Schulze-Lefert, P. (2015). Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley. *Cell Host & Microbe*, 17(3), 392–403. <https://doi.org/10.1016/j.chom.2015.01.011>

- Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology*, *64*(1), 807–838. <https://doi.org/10.1146/annurev-arplant-050312-120106>
- Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., & van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nature Reviews Genetics*, *21*(9), 526–540. <https://doi.org/10.1038/s41576-020-0244-x>
- Camacho Carvajal, M. M. (2001). *Molecular characterization of the roles of type 4 pili, NDH-I and PyrR in rhizosphere colonization of Pseudomonas fluorescens WCS365*. University Leiden, Leiden, The Netherlands.
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Cavaco, A. R., Matos, A. R., & Figueiredo, A. (2021). Speaking the language of lipids: the cross-talk between plants and pathogens in defence and disease. *Cellular and Molecular Life Sciences*, *78*(9), 4399–4415. <https://doi.org/10.1007/s00018-021-03791-0>
- Chang, W.-S., van de Mortel, M., Nielsen, L., Nino de Guzman, G., Li, X., & Halverson, L. J. (2007). Alginate Production by *Pseudomonas putida* Creates a Hydrated Microenvironment and Contributes to Biofilm Architecture and Stress Tolerance under Water-Limiting Conditions. *Journal of Bacteriology*, *189*(22), 8290–8299. <https://doi.org/10.1128/JB.00727-07>
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020a). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020b). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
- Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S. J., Webb, C., Wu, D., Varghese, N. J., Reddy, T. B. K., Mukherjee, S., Ovchinnikova, G., Nolan, M., Seshadri, R., Roux, S., Visel, A., Woyke, T., ... Ivanova, N. N. (2023). The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Research*, *51*(D1), D723–D732. <https://doi.org/10.1093/nar/gkac976>
- CHEN, M. H., TAKEDA, S., YAMADA, H., ISHII, Y., YAMASHINO, T., & MIZUNO, T. (2001). Characterization of the RcsC→YojN→RcsB Phosphorelay Signaling Pathway Involved in Capsular Synthesis in *Escherichia coli*. *Bioscience, Biotechnology, and Biochemistry*, *65*(10), 2364–2367. <https://doi.org/10.1271/bbb.65.2364>
- Cheng, Z., Park, E., & Glick, B. R. (2007). 1-Aminocyclopropane-1-carboxylate deaminase from *Pseudomonas putida* UW4 facilitates the growth of canola in the presence of salt. *Canadian Journal of Microbiology*, *53*(7), 912–918. <https://doi.org/10.1139/W07-050>

- Chi, F., Shen, S.-H., Cheng, H.-P., Jing, Y.-X., Yanni, Y. G., & Dazzo, F. B. (2005). Ascending Migration of Endophytic Rhizobia, from Roots to Leaves, inside Rice Plants and Assessment of Benefits to Rice Growth Physiology. *Applied and Environmental Microbiology*, *71*(11), 7271–7278. <https://doi.org/10.1128/AEM.71.11.7271-7278.2005>
- Chiquito-Contreras, C. J., Meza-Menchaca, T., Guzmán-López, O., Vásquez, E. C., & Ricaño-Rodríguez, J. (2024). Molecular Insights into Plant–Microbe Interactions: A Comprehensive Review of Key Mechanisms. *Frontiers in Bioscience-Elite*, *16*(1), 9. <https://doi.org/10.31083/j.fbe1601009>
- Cho, H.-S., Park, S.-Y., Ryu, C.-M., Kim, J. F., Kim, J.-G., & Park, S.-H. (2007). Interference of quorum sensing and virulence of the rice pathogen *Burkholderia glumae* by an engineered endophytic bacterium. *FEMS Microbiology Ecology*, *60*(1), 14–23. <https://doi.org/10.1111/j.1574-6941.2007.00280.x>
- Christen, B., Abeliuk, E., Collier, J. M., Kalogeraki, V. S., Passarelli, B., Collier, J. A., Fero, M. J., McAdams, H. H., & Shapiro, L. (2011). The essential genome of a bacterium. *Molecular Systems Biology*, *7*(1). <https://doi.org/10.1038/msb.2011.58>
- Chuang, H.-Y., Hofree, M., & Ideker, T. (2010). A Decade of Systems Biology. *Annual Review of Cell and Developmental Biology*, *26*(1), 721–744. <https://doi.org/10.1146/annurev-cellbio-100109-104122>
- Cole, B. J., Feltcher, M. E., Waters, R. J., Wetmore, K. M., Mucyn, T. S., Ryan, E. M., Wang, G., Ul-Hasan, S., McDonald, M., Yoshikuni, Y., Malmstrom, R. R., Deutschbauer, A. M., Dangl, J. L., & Visel, A. (2017). Genome-wide identification of bacterial plant colonization genes. *PLOS Biology*, *15*(9), e2002860. <https://doi.org/10.1371/journal.pbio.2002860>
- Compant, S., Clément, C., & Sessitsch, A. (2010). Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biology and Biochemistry*, *42*(5), 669–678. <https://doi.org/10.1016/j.soilbio.2009.11.024>
- Cooksey, D. A. (1986). *Galls of Gypsophila paniculata caused by Erwinia herbicola*.
- Corpe, W. A., & Rheem, S. (1989). Ecology of the methylotrophic bacteria on living leaf surfaces. *FEMS Microbiology Letters*, *62*(4), 243–249. <https://doi.org/10.1111/j.1574-6968.1989.tb03698.x>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Craig, N. L. (1997). TARGET SITE SELECTION IN TRANSPOSITION. *Annual Review of Biochemistry*, *66*(1), 437–474. <https://doi.org/10.1146/annurev.biochem.66.1.437>
- Czeczuga, B., & Godlewska, A. (2009). *Aquatic fungi and fungus-like organisms growing on seeds of 131 plant taxa by November*, 451–467. <https://doi.org/10.1127/0029-5035/2009/0089-0451>
- De Maayer, P., Chan, W. Y., Rezzonico, F., Bühlmann, A., Venter, S. N., Blom, J., Goesmann, A., Frey, J. E., Smits, T. H. M., Duffy, B., & Coutinho, T. A. (2012). Complete Genome Sequence of Clinical

- Isolate *Pantoea ananatis* LMG 5342. *Journal of Bacteriology*, 194(6), 1615–1616. <https://doi.org/10.1128/JB.06715-11>
- De Maayer, P., Venter, S. N., Kamber, T., Duffy, B., Coutinho, T. A., & Smits, T. H. (2011). Comparative genomics of the type VI secretion systems of *Pantoea* and *Erwinia* species reveals the presence of putative effector islands that may be translocated by the VgrG and Hcp proteins. *BMC Genomics*, 12(1), 576. <https://doi.org/10.1186/1471-2164-12-576>
- de Weert, S., Vermeiren, H., Mulders, I. H. M., Kuiper, I., Hendrickx, N., Bloemberg, G. V., Vanderleyden, J., De Mot, R., & Lugtenberg, B. J. J. (2002). Flagella-Driven Chemotaxis Towards Exudate Components Is an Important Trait for Tomato Root Colonization by *Pseudomonas fluorescens*. *Molecular Plant-Microbe Interactions*®, 15(11), 1173–1180. <https://doi.org/10.1094/MPMI.2002.15.11.1173>
- DeJesus, M. A., Ambadipudi, C., Baker, R., Sasseti, C., & Ioerger, T. R. (2015). TRANSIT - A Software Tool for Himar1 TnSeq Analysis. *PLOS Computational Biology*, 11(10), e1004401. <https://doi.org/10.1371/journal.pcbi.1004401>
- Delaux, P.-M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., Rothfels, C. J., Sederoff, H. W., Stevenson, D. W., Surek, B., Zhang, Y., Sussman, M. R., Dunand, C., Morris, R. J., Roux, C., ... Ané, J.-M. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences*, 112(43), 13390–13395. <https://doi.org/10.1073/pnas.1515426112>
- Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., von Mering, C., & Vorholt, J. A. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proceedings of the National Academy of Sciences*, 106(38), 16428–16433. <https://doi.org/10.1073/pnas.0905240106>
- Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1), 3. <https://doi.org/10.1186/1471-2105-7-3>
- Dong, C.-J., Wang, L.-L., Li, Q., & Shang, Q.-M. (2019). Bacterial communities in the rhizosphere, phyllosphere and endosphere of tomato plants. *PLOS ONE*, 14(11), e0223847. <https://doi.org/10.1371/journal.pone.0223847>
- Downie, J. A. (2010). The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiology Reviews*, 34(2), 150–170. <https://doi.org/10.1111/j.1574-6976.2009.00205.x>
- Duong, D. A., Jensen, R. V., & Stevens, A. M. (2018). Discovery of *Pantoea stewartii* ssp. *stewartii* genes important for survival in corn xylem through a Tn-Seq analysis. *Molecular Plant Pathology*, 19(8), 1929–1941. <https://doi.org/10.1111/mpp.12669>
- Dutkiewicz, J., Mackiewicz, B., Kinga Lemieszek, M., Golec, M., & Milanowski, J. (2016). <i>Pantoea agglomerans</i>: a mysterious bacterium of evil and good. Part III. Deleterious effects: infections of humans, animals and plants. *Annals of Agricultural and Environmental Medicine*, 23(2), 197–205. <https://doi.org/10.5604/12321966.1203878>

- Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., Eisen, J. A., & Sundaresan, V. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences*, *112*(8). <https://doi.org/10.1073/pnas.1414592112>
- Elbeltagy, A., Nishioka, K., Suzuki, H., Sato, T., Sato, Y.-I., Morisaki, H., Mitsui, H., & Minamisawa, K. (2000). Isolation and characterization of endophytic bacteria from wild and traditionally cultivated rice varieties. *Soil Science and Plant Nutrition*, *46*(3), 617–629. <https://doi.org/10.1080/00380768.2000.10409127>
- Ewald, P. W. (1983). *Transmission Modes and Evolution of the Parasitism-Mutualism Continuum* ". 295–306.
- Ezra, D., Barash, I., Weinthal, D. M., Gaba, V., & Manulis, S. (2004). *pthG* from *Pantoea agglomerans* pv. *gypsophilae* encodes an avirulence effector that determines incompatibility in multiple beet species. *Molecular Plant Pathology*, *5*(2), 105–113. <https://doi.org/10.1111/j.1364-3703.2004.00211.x>
- Fabian, B. K., Tetu, S. G., & Paulsen, I. T. (2020). Application of Transposon Insertion Sequencing to Agricultural Science. *Frontiers in Plant Science*, *11*. <https://doi.org/10.3389/fpls.2020.00291>
- FÅHRAEUS, G. (1957). The Infection of Clover Root Hairs by Nodule Bacteria Studied by a Simple Glass Slide Technique. *Microbiology*, *16*(2), 374–381. <https://doi.org/10.1099/00221287-16-2-374>
- Falk, R. (2010). Mutagenesis as a genetic research strategy. *Genetics*, *185*(4), 1135–1139. <https://doi.org/10.1534/genetics.110.120469>
- Fan, X. , Y. R. , Q. S. , H. F. . (2013). Over-expression of pectinase gene in endophytic Bacillus strains and its effect on colonization. *Environ. Biol.*, 805–810.
- FIALA, V., GLAD, C., MARTIN, M., JOLIVET, E., & DERRIDJ, S. (1990). Occurrence of soluble carbohydrates on the phylloplane of maize (*Zea mays* L.): variations in relation to leaf heterogeneity and position on the plant. *New Phytologist*, *115*(4), 609–615. <https://doi.org/10.1111/j.1469-8137.1990.tb00492.x>
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning*, *29*(2/3), 131–163. <https://doi.org/10.1023/A:1007465528199>
- Fürnkranz, M., Wanek, W., Richter, A., Abell, G., Rasche, F., & Sessitsch, A. (2008). Nitrogen fixation by phyllosphere bacteria associated with higher plants and their colonizing epiphytes of a tropical lowland rainforest of Costa Rica. *The ISME Journal*, *2*(5), 561–570. <https://doi.org/10.1038/ismej.2008.14>
- Fusaro, V. A., Mani, D. R., Mesirov, J. P., & Carr, S. A. (2009). Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nature Biotechnology*, *27*(2), 190–198. <https://doi.org/10.1038/nbt.1524>
- Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., & Koonin, E. V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread

- pathogens. *Nucleic Acids Research*, 49(D1), D274–D281. <https://doi.org/10.1093/nar/gkaa1018>
- Gandolfi, I., Bertolini, V., Ambrosini, R., Bestetti, G., & Franzetti, A. (2013). *Unravelling the bacterial diversity in the atmosphere*. 4727–4736. <https://doi.org/10.1007/s00253-013-4901-2>
- Gawronski, J. D., Wong, S. M. S., Giannoukos, G., Ward, D. V., & Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proceedings of the National Academy of Sciences*, 106(38), 16422–16427. <https://doi.org/10.1073/pnas.0906627106>
- Gepts, P., Beavis, W. D., Brummer, E. C., Shoemaker, R. C., Stalker, H. T., Weeden, N. F., & Young, N. D. (2005). Legumes as a Model Plant Family. Genomics for Food and Feed Report of the Cross-Legume Advances through Genomics Conference. *Plant Physiology*, 137(4), 1228–1235. <https://doi.org/10.1104/pp.105.060871>
- Gibson, K. E., Kobayashi, H., & Walker, G. C. (2008). Molecular Determinants of a Symbiotic Chronic Infection. *Annual Review of Genetics*, 42(1), 413–441. <https://doi.org/10.1146/annurev.genet.42.110807.091427>
- Giddens, S. R., Houlston, G. J., & Mahanty, H. K. (2003). The influence of antibiotic production and pre-emptive colonization on the population dynamics of *Pantoea agglomerans* (*Erwinia herbicola*) Eh1087 and *Erwinia amylovora* in planta. *Environmental Microbiology*, 5(10), 1016–1021. <https://doi.org/10.1046/j.1462-2920.2003.00506.x>
- Gieco, J. , M. M. V. and B. D. (2007). Enfermedades de la alfalfa y abordaje molecular de la selección por resistencia. In: D. Basigalup (Ed.), *El cultivo de la alfalfa en la Argentina . El Cultivo de La Alfalfa En La Argentina, Ediciones INTA*, 451–475.
- Glick, B. R. (1995). The enhancement of plant growth by free-living bacteria. *Canadian Journal of Microbiology*, 41(2), 109–117. <https://doi.org/10.1139/m95-015>
- Glick, B. R. (2012). Plant Growth-Promoting Bacteria: Mechanisms and Applications. *Scientifica*, 2012, 1–15. <https://doi.org/10.6064/2012/963401>
- Glick, B. R., Penrose, D. M., & Li, J. (1998). A Model For the Lowering of Plant Ethylene Concentrations by Plant Growth-promoting Bacteria. *Journal of Theoretical Biology*, 190(1), 63–68. <https://doi.org/10.1006/jtbi.1997.0532>
- Glick, B. R., Todorovic, B., Czarny, J., Cheng, Z., Duan, J., & McConkey, B. (2007). Promotion of Plant Growth by Bacterial ACC Deaminase. *Critical Reviews in Plant Sciences*, 26(5–6), 227–242. <https://doi.org/10.1080/07352680701572966>
- Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R., & Gordon, J. I. (2009a). Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host & Microbe*, 6(3), 279–289. <https://doi.org/10.1016/j.chom.2009.08.003>
- Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R., & Gordon, J. I. (2009b). Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host & Microbe*, 6(3), 279–289. <https://doi.org/10.1016/j.chom.2009.08.003>

- Gourion, B., Rossignol, M., & Vorholt, J. A. (2006). A proteomic study of *Methylobacterium extorquens* reveals a response regulator essential for epiphytic growth. *Proceedings of the National Academy of Sciences*, *103*(35), 13186–13191. <https://doi.org/10.1073/pnas.0603530103>
- Grant, A. J., Farris, M., Alefounder, P., Williams, P. H., Woodward, M. J., & O'Connor, C. D. (2003). Co-ordination of pathogenicity island expression by the BipA GTPase in enteropathogenic *Escherichia coli* (EPEC). *Molecular Microbiology*, *48*(2), 507–521. <https://doi.org/10.1046/j.1365-2958.2003.t01-1-03447.x>
- Guerrero, R., Margulis, L., & Berlanga, M. (2013). Symbiogenesis: the holobiont as a unit of evolution. *International Microbiology: The Official Journal of the Spanish Society for Microbiology*, *16*(3), 133–143. <https://doi.org/10.2436/20.1501.01.188>
- Guevarra, R. B., Magez, S., Peeters, E., Chung, M. S., Kim, K. H., & Radwanska, M. (2021). Comprehensive genomic analysis reveals virulence factors and antibiotic resistance genes in *Pantoea agglomerans* KM1, a potential opportunistic pathogen. *PLOS ONE*, *16*(1), e0239792. <https://doi.org/10.1371/journal.pone.0239792>
- Haefele, D. M., & Lindow, S. E. (1987). Flagellar Motility Confers Epiphytic Fitness Advantages upon *Pseudomonas syringae*. *Applied and Environmental Microbiology*, *53*(10), 2528–2533. <https://doi.org/10.1128/aem.53.10.2528-2533.1987>
- Hagen, M. J., Stockwell, V. O., Whistler, C. A., Johnson, K. B., & Loper, J. E. (2009). Stress Tolerance and Environmental Fitness of *Pseudomonas fluorescens* A506, Which Has a Mutation in *rpoS*. *Phytopathology*, *99*(6), 679–688. <https://doi.org/10.1094/PHYTO-99-6-0679>
- Hallmann, J., & Berg, G. (n.d.). Spectrum and Population Dynamics of Bacterial Root Endophytes. In *Microbial Root Endophytes* (pp. 15–31). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-33526-9_2
- Handelsman, J., & Brill, W. J. (1985). *Erwinia herbicola* isolates from alfalfa plants may play a role in nodulation of alfalfa by *Rhizobium meliloti*. *Applied and Environmental Microbiology*, *49*(4), 818–821. <https://doi.org/10.1128/aem.49.4.818-821.1985>
- Hardoim, P. R., Hardoim, C. C. P., Overbeek, L. S. Van, & Elsas, J. D. Van. (2012). *Dynamics of Seed-Borne Rice Endophytes on Early Plant Growth Stages*. *7*(2). <https://doi.org/10.1371/journal.pone.0030438>
- Hardoim, P. R., van Overbeek, L. S., & Elsas, J. D. van. (2008). Properties of bacterial endophytes and their proposed role in plant growth. *Trends in Microbiology*, *16*(10), 463–471. <https://doi.org/10.1016/j.tim.2008.07.008>
- Harwood, C. R., & C. S. M. (1990). *Molecular biological methods for Bacillus*.
- Helmann, T. C., Deutschbauer, A. M., & Lindow, S. E. (2019). Genome-wide identification of *Pseudomonas syringae* genes required for fitness during colonization of the leaf surface and apoplast. *Proceedings of the National Academy of Sciences*, *116*(38), 18900–18910. <https://doi.org/10.1073/pnas.1908858116>

- Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., & Holden, D. W. (1995). Simultaneous Identification of Bacterial Virulence Genes by Negative Selection. *Science*, *269*(5222), 400–403. <https://doi.org/10.1126/science.7618105>
- Herigstad, B., Hamilton, M., & Heersink, J. (2001). How to optimize the drop plate method for enumerating bacteria. *Journal of Microbiological Methods*, *44*(2), 121–129. [https://doi.org/10.1016/S0167-7012\(00\)00241-4](https://doi.org/10.1016/S0167-7012(00)00241-4)
- Hernandez, D., François, P., Farinelli, L., Østerås, M., & Schrenzel, J. (2008). De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Research*, *18*(5), 802–809. <https://doi.org/10.1101/gr.072033.107>
- Herrera, C. M., Koutsoudis, M. D., Wang, X., & von Bodman, S. B. (2008). *Pantoea stewartii* subsp. *stewartii* Exhibits Surface Motility, Which is a Critical Aspect of Stewart's Wilt Disease Development on Maize. *Molecular Plant-Microbe Interactions*[®], *21*(10), 1359–1370. <https://doi.org/10.1094/MPMI-21-10-1359>
- Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, *8*(1), 15–25. <https://doi.org/10.1038/nrmicro2259>
- Holt, K. E., Parkhill, J., Mazzoni, C. J., Roumagnac, P., Weill, F.-X., Goodhead, I., Rance, R., Baker, S., Maskell, D. J., Wain, J., Dolecek, C., Achtman, M., & Dougan, G. (2008). High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nature Genetics*, *40*(8), 987–993. <https://doi.org/10.1038/ng.195>
- Hurek, T., Reinhold-Hurek, B., Van Montagu, M., & Kellenberger, E. (1994). Root colonization and systemic spreading of *Azoarcus* sp. strain BH72 in grasses. *Journal of Bacteriology*, *176*(7), 1913–1923. <https://doi.org/10.1128/jb.176.7.1913-1923.1994>
- Izutsu, M., Lake, D. M., Matson, Z. W. D., Dodson, J. P., & Lenski, R. E. (2024). Effects of periodic bottlenecks on the dynamics of adaptive evolution in microbial populations. *Microbiology*, *170*(9). <https://doi.org/10.1099/mic.0.001494>
- 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>
- James A. Duke. (1983). *Handbook of energy crops*.
- James, E. K., Gyaneshwar, P., Mathan, N., Barraquio, W. L., Reddy, P. M., Iannetta, P. P. M., Olivares, F. L., & Ladha, J. K. (2002). Infection and Colonization of Rice Seedlings by the Plant Growth-Promoting Bacterium *Herbaspirillum seropedicae* Z67. *Molecular Plant-Microbe Interactions*[®], *15*(9), 894–906. <https://doi.org/10.1094/MPMI.2002.15.9.894>
- Johnston-monje, D., & Raizada, M. N. (2011). *Conservation and Diversity of Seed Associated Endophytes in Zea across Boundaries of Evolution , Ethnography and Ecology*. *6*(6). <https://doi.org/10.1371/journal.pone.0020396>

- Jones, D. L., Nguyen, C., & Finlay, R. D. (2009). Carbon flow in the rhizosphere: carbon trading at the soil–root interface. *Plant and Soil*, 321(1–2), 5–33. <https://doi.org/10.1007/s11104-009-9925-0>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070>
- Kidwell, M. G., & Lisch, D. R. (2001). PERSPECTIVE: TRANSPOSABLE ELEMENTS, PARASITIC DNA, AND GENOME EVOLUTION. *Evolution*, 55(1), 1–24. <https://doi.org/10.1111/j.0014-3820.2001.tb01268.x>
- Kim, H. J., Lee, J. H., Kang, B. R., Rong, X., McSpadden Gardener, B. B., Ji, H. J., Park, C.-S., & Kim, Y. C. (2012). Draft Genome Sequence of *Pantoea ananatis* B1-9, a Nonpathogenic Plant Growth-Promoting Bacterium. *Journal of Bacteriology*, 194(3), 729–729. <https://doi.org/10.1128/JB.06484-11>
- Kiss, E., Huguet, T., Poinot, V., & Batut, J. (2004). The *typA* Gene is Required for Stress Adaptation as Well as for Symbiosis of *Sinorhizobium meliloti* 1021 with Certain *Medicago truncatula* Lines. *Molecular Plant-Microbe Interactions*[®], 17(3), 235–244. <https://doi.org/10.1094/MPMI.2004.17.3.235>
- Klaedtke, S., Jacques, M., Raggi, L., Prévieux, A., Bonneau, S., Negri, V., Chable, V., & Barret, M. (2016). Terroir is a key driver of seed-associated microbial assemblages. *Environmental Microbiology*, 18(6), 1792–1804. <https://doi.org/10.1111/1462-2920.12977>
- Kleckner, N., Chan, R. K., Tye, B.-K., & Botstein, D. (1975). Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. *Journal of Molecular Biology*, 97(4), 561–575. [https://doi.org/10.1016/S0022-2836\(75\)80059-3](https://doi.org/10.1016/S0022-2836(75)80059-3)
- Kleckner, N., Roth, J., & Botstein, D. (1977). Genetic engineering in Vivo using translocatable drug-resistance elements. *Journal of Molecular Biology*, 116(1), 125–159. [https://doi.org/10.1016/0022-2836\(77\)90123-1](https://doi.org/10.1016/0022-2836(77)90123-1)
- Kleinbaum, D. G., Kupper, L. L., & Chambless, L. E. (1982). Logistic regression analysis of epidemiologic data: theory and practice. *Communications in Statistics - Theory and Methods*, 11(5), 485–547. <https://doi.org/10.1080/03610928208828251>
- Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., von Mering, C., & Vorholt, J. A. (2012). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *The ISME Journal*, 6(7), 1378–1390. <https://doi.org/10.1038/ismej.2011.192>
- Knief, C., Ramette, A., Frances, L., Alonso-Blanco, C., & Vorholt, J. A. (2010). Site and plant species are important determinants of the *Methylobacterium* community composition in the plant phyllosphere. *The ISME Journal*, 4(6), 719–728. <https://doi.org/10.1038/ismej.2010.9>

- Koutsoudis, M. D., Tsaltas, D., Minogue, T. D., & von Bodman, S. B. (2006). Quorum-sensing regulation governs bacterial adhesion, biofilm development, and host colonization in *Pantoea stewartii* subspecies *stewartii*. *Proceedings of the National Academy of Sciences*, *103*(15), 5983–5988. <https://doi.org/10.1073/pnas.0509860103>
- Lacroute, F. (1975). *Chapter 13 The Use of Mutants in Metabolic Studies* (pp. 235–245). [https://doi.org/10.1016/S0091-679X\(08\)60326-X](https://doi.org/10.1016/S0091-679X(08)60326-X)
- Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J., & Robertson, H. M. (1999). Hyperactive transposase mutants of the *Himar 1* mariner transposon. *Proceedings of the National Academy of Sciences*, *96*(20), 11428–11433. <https://doi.org/10.1073/pnas.96.20.11428>
- Langridge, G. C., Phan, M.-D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J., & Turner, A. K. (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Research*, *19*(12), 2308–2316. <https://doi.org/10.1101/gr.097097.109>
- Larrañaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J. A., Armañanzas, R., Santafé, G., Pérez, A., & Robles, V. (2006). Machine learning in bioinformatics. *Briefings in Bioinformatics*, *7*(1), 86–112. <https://doi.org/10.1093/bib/bbk007>
- Lenski, R. E., Rose, M. R., Simpson, S. C., & Tadler, S. C. (1991). Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist*, *138*(6), 1315–1341. <https://doi.org/10.1086/285289>
- Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>
- Levy, A., Salas Gonzalez, I., Mittelviehhaus, M., Clingenpeel, S., Herrera Paredes, S., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., Rangel Alvarez, B., Lundberg, D. S., Lu, T.-Y., Lebeis, S., Jin, Z., McDonald, M., Klein, A. P., Feltcher, M. E., Rio, T. G., ... Dangl, J. L. (2018). Genomic features of bacterial adaptation to plants. *Nature Genetics*, *50*(1), 138–150. <https://doi.org/10.1038/s41588-017-0012-9>
- Lin Tung, W., & Chow, K.-C. (1995). A modified medium for efficient electrotransformation of *E. coli*. *Trends in Genetics*, *11*(4), 128–129. [https://doi.org/10.1016/S0168-9525\(00\)89022-8](https://doi.org/10.1016/S0168-9525(00)89022-8)
- López, J. L., Alvarez, F., Príncipe, A., Salas, M. E., Lozano, M. J., Draghi, W. O., Jofré, E., & Lagares, A. (2018). Isolation, taxonomic analysis, and phenotypic characterization of bacterial endophytes present in alfalfa (*Medicago sativa*) seeds. *Journal of Biotechnology*, *267*, 55–62. <https://doi.org/10.1016/j.jbiotec.2017.12.020>
- López, J. L., Fourie, A., Poppeliers, S. W. M., Pappas, N., Sánchez-Gil, J. J., de Jonge, R., & Dutilh, B. E. (2023). Growth rate is a dominant factor predicting the rhizosphere effect. *The ISME Journal*, *17*(9), 1396–1405. <https://doi.org/10.1038/s41396-023-01453-6>
- López-López, A., Rogel, M. A., Ormeño-Orrillo, E., Martínez-Romero, J., & Martínez-Romero, E. (2010). *Phaseolus vulgaris* seed-borne endophytic community with novel bacterial species

- such as *Rhizobium endophyticum* sp. nov. *Systematic and Applied Microbiology*, 33(6), 322–327. <https://doi.org/10.1016/j.syapm.2010.07.005>
- Lorenzi, A. S., Bonatelli, M. L., Chia, M. A., Peressim, L., & Quecine, M. C. (2022). Opposite Sides of *Pantoea* agglomerans and Its Associated Commercial Outlook. *Microorganisms*, 10(10), 2072. <https://doi.org/10.3390/microorganisms10102072>
- Luchetti Abril. (2019). “*Construcción y evaluación de un sistema reportero para evaluar la transferencia conjugativa.*” Universidad Nacional de La Plata.
- Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektsen, A., Kunin, V., Rio, T. G. del, Edgar, R. C., Eickhorst, T., Ley, R. E., Hugenholtz, P., Tringe, S. G., & Dangl, J. L. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature*, 488(7409), 86–90. <https://doi.org/10.1038/nature11237>
- Luneau, J. S., Baudin, M., Quiroz Monnens, T., Carrère, S., Bouchez, O., Jardinaud, M., Gris, C., François, J., Ray, J., Torralba, B., Arlat, M., Lewis, J. D., Lauber, E., Deutschbauer, A. M., Noël, L. D., & Boulanger, A. (2022). Genome-wide identification of fitness determinants in the *Xanthomonas campestris* bacterial pathogen during early stages of plant infection. *New Phytologist*, 236(1), 235–248. <https://doi.org/10.1111/nph.18313>
- Luo, H., Lin, Y., Liu, T., Lai, F.-L., Zhang, C.-T., Gao, F., & Zhang, R. (2021). DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. *Nucleic Acids Research*, 49(D1), D677–D686. <https://doi.org/10.1093/nar/gkaa917>
- Lv, L., Luo, J., Ahmed, T., Zaki, H. E. M., Tian, Y., Shahid, M. S., Chen, J., & Li, B. (2022). Beneficial Effect and Potential Risk of *Pantoea* on Rice Production. *Plants*, 11(19), 2608. <https://doi.org/10.3390/plants11192608>
- Maayer, P. De, Aliyu, H., Vikram, S., Blom, J., Duffy, B., Cowan, D. A., Smits, T. H. M., Venter, S. N., Coutinho, T. A., & Guimarães, L. C. (2017). *Pathogenomic and Evolutionary Genomic Insights into the Agronomically Relevant Enterobacteria Pantoea ananatis and Pantoea stewartii*. 8(September), 1–14. <https://doi.org/10.3389/fmicb.2017.01755>
- Mano, H., Tanaka, F., Watanabe, A., Kaga, H., Okunishi, S., & Morisaki, H. (2006). Culturable Surface and Endophytic Bacterial Flora of the Maturing Seeds of Rice Plants (*Oryza sativa*) Cultivated in a Paddy Field. *Microbes and Environments*. <https://doi.org/10.1264/jsme2.21.86>
- Marco, M. L., Legac, J., & Lindow, S. E. (2005). *Pseudomonas syringae* genes induced during colonization of leaf surfaces. *Environmental Microbiology*, 7(9), 1379–1391. <https://doi.org/10.1111/j.1462-2920.2005.00825.x>
- Martínez-Viveros, O., Jorquera, M. A., Crowley, D. E., Gajardo, G., & Mora, M. L. (2010). MECHANISMS AND PRACTICAL CONSIDERATIONS INVOLVED IN PLANT GROWTH PROMOTION BY RHIZOBACTERIA. *Journal of Soil Science and Plant Nutrition*, 10(3). <https://doi.org/10.4067/S0718-95162010000100006>
- McCormack, P. J., Wildman, H. G., & Jeffries, P. (1994). Production of antibacterial compounds by phylloplane-inhabiting yeasts and yeastlike fungi. *Applied and Environmental Microbiology*, 60(3), 927–931. <https://doi.org/10.1128/aem.60.3.927-931.1994>

- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. <https://doi.org/10.1007/BF02478259>
- McCully, M. E. (2001). Niches for bacterial endophytes in crop plants: a plant biologist's view. *Functional Plant Biology*, 28(9), 983. <https://doi.org/10.1071/PP01101>
- Medrano, E. G., & Bell, A. A. (2012). Genome Sequence of *Pantoea* sp. Strain Sc 1, an Opportunistic Cotton Pathogen. *Journal of Bacteriology*, 194(11), 3019–3019. <https://doi.org/10.1128/JB.00450-12>
- Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, 10(1), 2182. <https://doi.org/10.1038/s41467-019-10210-3>
- Meng, J., Young, G., & Chen, J. (2021). The Rcs System in Enterobacteriaceae: Envelope Stress Responses and Virulence Regulation. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.627104>
- Michaud, R., Lehman, W. F., & Rumbaugh, M. D. (2015). *World Distribution and Historical Development* (pp. 25–91). <https://doi.org/10.2134/agronmonogr29.c2>
- Minogue, T. D., Carlier, A. L., Koutsoudis, M. D., & Von Bodman, S. B. (2005). The cell density-dependent expression of stewartan exopolysaccharide in *Pantoea stewartii* ssp. *stewartii* is a function of EsaR-mediated repression of the *rcsA* gene. *Molecular Microbiology*, 56(1), 189–203. <https://doi.org/10.1111/j.1365-2958.2004.04529.x>
- Minogue, T. D., Trebra, M. W., Bernhard, F., & Bodman, S. B. von. (2002). The autoregulatory role of EsaR, a quorum-sensing regulator in *Pantoea stewartii* ssp. *stewartii*: evidence for a repressor function. *Molecular Microbiology*, 44(6), 1625–1635. <https://doi.org/10.1046/j.1365-2958.2002.02987.x>
- 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>
- Mohammadi, M., Burbank, L., & Roper, M. C. (2012). *Pantoea stewartii* subsp. *stewartii* Produces an Endoglucanase That Is Required for Full Virulence in Sweet Corn. *Molecular Plant-Microbe Interactions*®, 25(4), 463–470. <https://doi.org/10.1094/MPMI-09-11-0226>
- Monier, J.-M., & Lindow, S. E. (2003). Differential survival of solitary and aggregated bacterial cells promotes aggregate formation on leaf surfaces. *Proceedings of the National Academy of Sciences*, 100(26), 15977–15982. <https://doi.org/10.1073/pnas.2436560100>
- Moreno, M. V., Arolfo, V., López, J. L., Erdozain, S. A., Bigatton, E. D., Ayoub, I., Lucini, E. I., Lagares, A., & Odorizzi, A. (2024). Selección de cepas bacterianas con capacidad antifúngica contra fitopatógenos de alfalfa para constituir un consorcio bacteriano. *AgriScientia*, 41(1), 73–82. <https://doi.org/10.31047/1668.298x.v41.n1.41895>
- Neidig, A., Yeung, A. T., Rosay, T., Tettmann, B., Stempel, N., Rueger, M., Lesouhaitier, O., & Overhage, J. (2013). TypA is involved in virulence, antimicrobial resistance and biofilm

- formation in *Pseudomonas aeruginosa*. *BMC Microbiology*, 13(1), 77. <https://doi.org/10.1186/1471-2180-13-77>
- Nelson, E. B. (2004). *MICROBIAL DYNAMICS AND INTERACTIONS IN THE SPERMOSPHERE*. 28, 271–309. <https://doi.org/10.1146/annurev.phyto.42.121603.131041>
- Nelson, E. B. (2018). The seed microbiome: Origins, interactions, and impacts. *Plant and Soil*, 422(1–2), 7–34. <https://doi.org/10.1007/s11104-017-3289-7>
- Nguyen, C. (2009). Rhizodeposition of Organic C by Plant: Mechanisms and Controls. In *Sustainable Agriculture* (pp. 97–123). Springer Netherlands. https://doi.org/10.1007/978-90-481-2666-8_9
- Nick, T. G., & Campbell, K. M. (2007). *Logistic Regression* (pp. 273–301). https://doi.org/10.1007/978-1-59745-530-5_14
- Nizan, R., Barash, I., Valinsky, L., Lichter, A., & Manulis, S. (1997). The Presence of *hrp* Genes on the Pathogenicity-Associated Plasmid of the Tumorigenic Bacterium *Erwinia herbicola* pv. *gypsophila*. *Molecular Plant-Microbe Interactions*®, 10(5), 677–682. <https://doi.org/10.1094/MPMI.1997.10.5.677>
- Nowak, J. (1998). Benefits of in vitro “biotization” of plant tissue cultures with microbial inoculants. *In Vitro Cellular & Developmental Biology - Plant*, 34(2), 122–130. <https://doi.org/10.1007/BF02822776>
- Oldroyd, G. E. D., & Downie, J. A. (2008). Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. *Annual Review of Plant Biology*, 59(1), 519–546. <https://doi.org/10.1146/annurev.arplant.59.032607.092839>
- Otero-Bravo, A., & Sabree, Z. L. (2021). Multiple concurrent and convergent stages of genome reduction in bacterial symbionts across a stink bug family. *Scientific Reports*, 11(1), 7731. <https://doi.org/10.1038/s41598-021-86574-8>
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarszewski, A., Chaumeil, P.-A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, 36(10), 996–1004. <https://doi.org/10.1038/nbt.4229>
- Parmar, N., & Dufresne, J. (2011). *Beneficial Interactions of Plant Growth Promoting Rhizosphere Microorganisms* (pp. 27–42). https://doi.org/10.1007/978-3-642-19769-7_2
- Pedregosa, F., V. G. G. A., M. V., T. B., G. O., ... others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2825–2830.
- Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangl, J. L., Buckler, E. S., & Ley, R. E. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proceedings of the National Academy of Sciences*, 110(16), 6548–6553. <https://doi.org/10.1073/pnas.1302837110>
- Perry, B. J., & Yost, C. K. (2014). Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae. *BMC Microbiology*, 14(1). <https://doi.org/10.1186/s12866-014-0298-z>

- PERSELLO-CARTIEAUX, F., NUSSAUME, L., & ROBAGLIA, C. (2003). Tales from the underground: molecular plant–rhizobacteria interactions. *Plant, Cell & Environment*, 26(2), 189–199. <https://doi.org/10.1046/j.1365-3040.2003.00956.x>
- Philippot, L., Raaijmakers, J. M., Lemanceau, P., & van der Putten, W. H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiology*, 11(11), 789–799. <https://doi.org/10.1038/nrmicro3109>
- Pini, F., Frascella, A., Santopolo, L., Bazzicalupo, M., Biondi, E. G., Scotti, C., & Mengoni, A. (2012). Exploring the plant-associated bacterial communities in *Medicago sativa* L. *Plant, Cell & Environment*, 35(1), 1–12. <https://doi.org/10.1186/1471-2180-12-78>
- Prashar, P., Kapoor, N., & Sachdeva, S. (2014). Rhizosphere: its structure, bacterial diversity and significance. *Reviews in Environmental Science and Bio/Technology*, 13(1), 63–77. <https://doi.org/10.1007/s11157-013-9317-z>
- Prieto, P., Schilirò, E., Maldonado-González, M. M., Valderrama, R., Barroso-Albarracín, J. B., & Mercado-Blanco, J. (2011). Root Hairs Play a Key Role in the Endophytic Colonization of Olive Roots by *Pseudomonas* spp. with Biocontrol Activity. *Microbial Ecology*, 62(2), 435–445. <https://doi.org/10.1007/s00248-011-9827-6>
- Proietti, S., Giangrande, C., Amoresano, A., Pucci, P., Molinaro, A., Bertini, L., Caporale, C., & Caruso, C. (2014). *Xanthomonas campestris* lipooligosaccharides trigger innate immunity and oxidative burst in *Arabidopsis*. *Plant Physiology and Biochemistry*, 85, 51–62. <https://doi.org/10.1016/j.plaphy.2014.10.011>
- Ramonedá, J., Fan, K., Lucas, J. M., Chu, H., Bissett, A., Strickland, M. S., & Fierer, N. (2024). Ecological relevance of flagellar motility in soil bacterial communities. *The ISME Journal*, 18(1). <https://doi.org/10.1093/ismejo/wrae067>
- Rastogi, G., Coaker, G. L., & Leveau, J. H. J. (2013). *New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches*. <https://doi.org/10.1111/1574-6968.12225>
- Rediers, H., Bonnacarrère, V., Rainey, P. B., Hamonts, K., Vanderleyden, J., & De Mot, R. (2003). Development and Application of a *dapB*-Based In Vivo Expression Technology System To Study Colonization of Rice by the Endophytic Nitrogen-Fixing Bacterium *Pseudomonas stutzeri* A15. *Applied and Environmental Microbiology*, 69(11), 6864–6874. <https://doi.org/10.1128/AEM.69.11.6864-6874.2003>
- Remus-Emsermann, M. N. P., Kowalchuk, G. A., & Leveau, J. H. J. (2013). Single-cell versus population-level reproductive success of bacterial immigrants to pre-colonized leaf surfaces. *Environmental Microbiology Reports*, 5(3), 387–392. <https://doi.org/10.1111/1758-2229.12040>
- Remus-Emsermann, M. N. P., Tecon, R., Kowalchuk, G. A., & Leveau, J. H. J. (2012). Variation in local carrying capacity and the individual fate of bacterial colonizers in the phyllosphere. *The ISME Journal*, 6(4), 756–765. <https://doi.org/10.1038/ismej.2011.209>

- Rigano, L. A., Siciliano, F., Enrique, R., Sendín, L., Filippone, P., Torres, P. S., Qüesta, J., Dow, J. M., Castagnaro, A. P., Vojnov, A. A., & Marano, M. R. (2007). Biofilm Formation, Epiphytic Fitness, and Canker Development in *Xanthomonas axonopodis* pv. *citri*. *Molecular Plant-Microbe Interactions*[®], 20(10), 1222–1230. <https://doi.org/10.1094/MPMI-20-10-1222>
- Roncato-Maccari, L. D. B., Ramos, H. J. O., Pedrosa, F. O., Alquini, Y., Chubatsu, L. S., Yates, M. G., Rigo, L. U., Steffens, M. B. R., & Souza, E. M. (2003). Endophytic *Herbaspirillum seropedicae* expresses *nif* genes in gramineous plants. *FEMS Microbiology Ecology*, 45(1), 39–47. [https://doi.org/10.1016/S0168-6496\(03\)00108-9](https://doi.org/10.1016/S0168-6496(03)00108-9)
- ROPER, M. C. (2011). *Pantoea stewartii* subsp. *stewartii*: lessons learned from a xylem-dwelling pathogen of sweet corn. *Molecular Plant Pathology*, 12(7), 628–637. <https://doi.org/10.1111/j.1364-3703.2010.00698.x>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- Rosenblueth, M., & Martínez-Romero, E. (2006). Bacterial Endophytes and Their Interactions with Hosts. *Molecular Plant-Microbe Interactions*[®], 19(8), 827–837. <https://doi.org/10.1094/MPMI-19-0827>
- Royet, K., Parisot, N., Rodrigue, A., Gueguen, E., & Condemine, G. (2019). Identification by Tn-seq of *Dickeya dadantii* genes required for survival in chicory plants. *Molecular Plant Pathology*, 20(2), 287–306. <https://doi.org/10.1111/mpp.12754>
- Rudgers, J. A., Afkhami, M. E., Rúa, M. A., Davitt, A. J., Hammer, S., & Huguet, V. M. (2009). A fungus among us: broad patterns of endophyte distribution in the grasses. *Ecology*, 90(6), 1531–1539. <https://doi.org/10.1890/08-0116.1>
- Ryan, R. P., & Dow, J. M. (2011). Communication with a growing family: diffusible signal factor (DSF) signaling in bacteria. *Trends in Microbiology*, 19(3), 145–152. <https://doi.org/10.1016/j.tim.2010.12.003>
- Salas, M. E., Lozano, M. J., López, J. L., Draghi, W. O., Serrania, J., Torres Tejerizo, G. A., Albicoro, F. J., Nilsson, J. F., Pistorio, M., Del Papa, M. F., Parisi, G., Becker, A., & Lagares, A. (2017). Specificity traits consistent with legume-rhizobia coevolution displayed by *Ensifer meliloti* rhizosphere colonization. *Environmental Microbiology*, 19(9), 3423–3438. <https://doi.org/10.1111/1462-2920.13820>
- Salas Maria Eugenia. (2015). “La simbiosis fijadora de nitrógeno *Sinorhizobium meliloti* - alfalfa: Aproximaciones ómicas aplicadas a la identificación y caracterización de determinantes genéticos del rizobio asociados a la colonización temprana de la raíz de alfalfa (*Medicago Sativa*).” Instituto de Biotecnología y Biología Molecular.
- Salvucci, E. (2014). Microbiome, holobiont and the net of life. *Critical Reviews in Microbiology*, 1–10. <https://doi.org/10.3109/1040841X.2014.962478>
- Sambrook, J. , F. E. R. , & M. T. (1989). *Molecular Cloning: A Laboratory Manual: Vol. (2nd ed.)* (second).

- Sanderson, M. A., & Adler, P. R. (2008). Perennial Forages as Second Generation Bioenergy Crops. *International Journal of Molecular Sciences*, 9(5), 768–788. <https://doi.org/10.3390/ijms9050768>
- Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, 2020. <https://doi.org/10.1093/database/baaa062>
- Schreiber, L., Krimm, U., Knoll, D., Sayed, M., Auling, G., & Kroppenstedt, R. M. (2005). Plant–microbe interactions: identification of epiphytic bacteria and their ability to alter leaf surface permeability. *New Phytologist*, 166(2), 589–594. <https://doi.org/10.1111/j.1469-8137.2005.01343.x>
- Sessitsch, A., Hardoim, P., Döring, J., Weilharter, A., Krause, A., Woyke, T., Mitter, B., Hauberg-Lotte, L., Friedrich, F., Rahalkar, M., Hurek, T., Sarkar, A., Bodrossy, L., van Overbeek, L., Brar, D., van Elsas, J. D., & Reinhold-Hurek, B. (2012). Functional Characteristics of an Endophyte Community Colonizing Rice Roots as Revealed by Metagenomic Analysis. *Molecular Plant-Microbe Interactions*[®], 25(1), 28–36. <https://doi.org/10.1094/MPMI-08-11-0204>
- Sheibani-Tezerji, R., Naveed, M., Jehl, M.-A., Sessitsch, A., Rattei, T., & Mitter, B. (2015). The genomes of closely related *Pantoea* ananatis maize seed endophytes having different effects on the host plant differ in secretion system genes and mobile genetic elements. *Frontiers in Microbiology*, 6. <https://doi.org/10.3389/fmicb.2015.00440>
- Simon, R., Priefer, U., & Pühler, A. (1983a). A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Bio/Technology*, 1(9), 784–791. <https://doi.org/10.1038/nbt1183-784>
- Simon, R., Priefer, U., & Pühler, A. (1983b). A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Bio/Technology*, 1(9), 784–791. <https://doi.org/10.1038/nbt1183-784>
- Sørensen, J., & Sessitsch, A. (2006). Plant-associated bacteria-lifestyle and molecular interactions. In *Modern Soil Microbiology, Second Edition* (pp. 211–236).
- Spaepen, S., Vanderleyden, J., & Remans, R. (2007). Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiology Reviews*, 31(4), 425–448. <https://doi.org/10.1111/j.1574-6976.2007.00072.x>
- Sprent, J. I., & de Faria, S. M. (1988). Mechanisms of infection of plants by nitrogen fixing organisms. *Plant and Soil*, 110(2), 157–165. <https://doi.org/10.1007/BF02226795>
- Stoltzfus, J. C. (2011). Logistic Regression: A Brief Primer. *Academic Emergency Medicine*, 18(10), 1099–1104. <https://doi.org/10.1111/j.1553-2712.2011.01185.x>
- Stout, V., & Gottesman, S. (1990). RcsB and RcsC: a two-component regulator of capsule synthesis in *Escherichia coli*. *Journal of Bacteriology*, 172(2), 659–669. <https://doi.org/10.1128/jb.172.2.659-669.1990>

- Sugiyama, A., Ueda, Y., Zushi, T., Takase, H., & Yazaki, K. (2014). Changes in the Bacterial Community of Soybean Rhizospheres during Growth in the Field. *PLoS ONE*, *9*(6), e100709. <https://doi.org/10.1371/journal.pone.0100709>
- Sun, Y., Cheng, Z., & Glick, B. R. (2009). The presence of a 1-aminocyclopropane-1-carboxylate (ACC) deaminase deletion mutation alters the physiology of the endophytic plant growth-promoting bacterium *Burkholderia phytofirmans* PsJN. *FEMS Microbiology Letters*, *296*(1), 131–136. <https://doi.org/10.1111/j.1574-6968.2009.01625.x>
- Tenaillon, O., Barrick, J. E., Ribeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., Wu, G. C., Wielgoss, S., Cruveiller, S., Médigue, C., Schneider, D., & Lenski, R. E. (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*, *536*(7615), 165–170. <https://doi.org/10.1038/nature18959>
- Tkacz, A., Cheema, J., Chandra, G., Grant, A., & Poole, P. S. (2015). Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. *The ISME Journal*, *9*(11), 2349–2359. <https://doi.org/10.1038/ismej.2015.41>
- Touw, W. G., Bayjanov, J. R., Overmars, L., Backus, L., Boekhorst, J., Wels, M., & van Hijum, S. A. F. T. (2013). Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? *Briefings in Bioinformatics*, *14*(3), 315–326. <https://doi.org/10.1093/bib/bbs034>
- Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T., & Singh, B. K. (2020). Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology*, *18*(11), 607–621. <https://doi.org/10.1038/s41579-020-0412-1>
- Truyens, S., Weyens, N., Cuypers, A., & Vangronsveld, J. (2014). *Minireview Bacterial seed endophytes: genera, vertical transmission and interaction with plants.* <https://doi.org/10.1111/1758-2229.12181>
- Turner, T. R., James, E. K., & Poole, P. S. (2013). The plant microbiome. *Genome Biology*, *14*(6), 209. <https://doi.org/10.1186/gb-2013-14-6-209>
- van Opijnen, T., Bodi, K. L., & Camilli, A. (2009a). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, *6*(10), 767–772. <https://doi.org/10.1038/nmeth.1377>
- van Opijnen, T., Bodi, K. L., & Camilli, A. (2009b). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, *6*(10), 767–772. <https://doi.org/10.1038/nmeth.1377>
- van Opijnen, T., & Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*, *11*(7), 435–442. <https://doi.org/10.1038/nrmicro3033>
- Van Overbeek, L., & Van Elsas, J. D. (2008). Effects of plant genotype and growth stage on the structure of bacterial communities associated with potato (*Solanum tuberosum* L.). *FEMS Microbiology Ecology*, *64*(2), 283–296. <https://doi.org/10.1111/j.1574-6941.2008.00469.x>

- Vandenkoornhuysse, P., Quaiser, A., Duhamel, M., Le Van, A., & Dufresne, A. (2015). The importance of the microbiome of the plant holobiont. *New Phytologist*, *206*(4), 1196–1206. <https://doi.org/10.1111/nph.13312>
- Verikas, A., Gelzinis, A., & Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, *44*(2), 330–349. <https://doi.org/10.1016/j.patcog.2010.08.011>
- von Bodman, S. B., Bauer, W. D., & Coplin, D. L. (2003). Quorum sensing in plant-pathogenic bacteria. *Annual Review of Phytopathology*, *41*(1), 455–482. <https://doi.org/10.1146/annurev.phyto.41.052002.095652>
- Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Publishing Group*, *10*(12), 828–840. <https://doi.org/10.1038/nrmicro2910>
- Walterson, A. M., & Stavrinides, J. (2015). *Pantoea*: insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiology Reviews*, *39*(6), 968–984. <https://doi.org/10.1093/femsre/fuv027>
- Webster, G., Jain, V., Davey, M. R., Gough, C., Vasse, J., Dénarié, J., & Cocking, E. C. (1998). The flavonoid naringenin stimulates the intercellular colonization of wheat roots by *Azorhizobium caulinodans*. *Plant, Cell & Environment*, *21*(4), 373–383. <https://doi.org/10.1046/j.1365-3040.1998.00278.x>
- Weibull, J., Ronquist, F., & Brishammar, S. (1990). Free Amino Acid Composition of Leaf Exudates and Phloem Sap. *Plant Physiology*, *92*(1), 222–226. <https://doi.org/10.1104/pp.92.1.222>
- Wheatley, R. M., Ford, B. L., Li, L., Aroney, S. T. N., Knights, H. E., Ledermann, R., East, A. K., Ramachandran, V. K., & Poole, P. S. (2020). Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proceedings of the National Academy of Sciences*, *117*(38), 23823–23834. <https://doi.org/10.1073/pnas.2009094117>
- Whipps, J. M., Hand, P., Pink, D., & Bending, G. D. (2008). Phyllosphere microbiology with special reference to diversity and plant genotype. *Journal of Applied Microbiology*, *105*(6), 1744–1755. <https://doi.org/10.1111/j.1365-2672.2008.03906.x>
- Wilson, M., & Lindow, S. E. (1994). Coexistence among Epiphytic Bacterial Populations Mediated through Nutritional Resource Partitioning. *Applied and Environmental Microbiology*, *60*(12), 4468–4477. <https://doi.org/10.1128/aem.60.12.4468-4477.1994>
- Wiser, M. J., Ribbeck, N., & Lenski, R. E. (2013). Long-Term Dynamics of Adaptation in Asexual Populations. *Science*, *342*(6164), 1364–1367. <https://doi.org/10.1126/science.1243357>
- Wu, Y., & Gadsden, S. A. (2023). Machine learning algorithms in microbial classification: a comparative analysis. *Frontiers in Artificial Intelligence*, *6*. <https://doi.org/10.3389/frai.2023.1200994>
- Xie, H., Pasternak, J. J., & Glick, B. R. (1996). Isolation and Characterization of Mutants of the Plant Growth-Promoting Rhizobacterium *Pseudomonas putida* GR12-2 That Overproduce Indoleacetic Acid. *Current Microbiology*, *32*(2), 67–71. <https://doi.org/10.1007/s002849900012>

- Yang, C.-H., Crowley, D. E., Borneman, J., & Keen, N. T. (2001). Microbial phyllosphere populations are more complex than previously realized. *Proceedings of the National Academy of Sciences*, *98*(7), 3889–3894. <https://doi.org/10.1073/pnas.051633898>
- Yeoh, Y. K., Dennis, P. G., Paungfoo-Lonhienne, C., Weber, L., Brackin, R., Ragan, M. A., Schmidt, S., & Hugenholtz, P. (2017). Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nature Communications*, *8*(1), 215. <https://doi.org/10.1038/s41467-017-00262-8>
- Yousaf, S., Afzal, M., Reichenauer, T. G., Brady, C. L., & Sessitsch, A. (2011). Hydrocarbon degradation, plant colonization and gene expression of alkane degradation genes by endophytic *Enterobacter ludwigii* strains. *Environmental Pollution*, *159*(10), 2675–2683. <https://doi.org/10.1016/j.envpol.2011.05.031>
- Zgadżaj, R., Garrido-Oter, R., Jensen, D. B., Koprivova, A., Schulze-Lefert, P., & Radutoiu, S. (2016). Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. *Proceedings of the National Academy of Sciences*, *113*(49). <https://doi.org/10.1073/pnas.1616564113>
- Zhao, S., Wei, H., Lin, C.-Y., Zeng, Y., Tucker, M. P., Himmel, M. E., & Ding, S.-Y. (2016). Burkholderia phytofirmans Inoculation-Induced Changes on the Shoot Cell Anatomy and Iron Accumulation Reveal Novel Components of Arabidopsis-Endophyte Interaction that Can Benefit Downstream Biomass Deconstruction. *Frontiers in Plant Science*, *7*. <https://doi.org/10.3389/fpls.2016.00024>
- Zilber-Rosenberg, I., & Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews*, *32*(5), 723–735. <https://doi.org/10.1111/j.1574-6976.2008.00123.x>
- Zúñiga, A., Poupin, M. J., Donoso, R., Ledger, T., Guilianni, N., Gutiérrez, R. A., & González, B. (2013). Quorum Sensing and Indole-3-Acetic Acid Degradation Play a Role in Colonization and Plant Growth Promotion of *Arabidopsis thaliana* by *Burkholderia phytofirmans* PsJN. *Molecular Plant-Microbe Interactions*[®], *26*(5), 546–553. <https://doi.org/10.1094/MPMI-10-12-0241-R>

