UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Interacciones Ecológicas: Estudio de Microbiomas de Plantas y Genómica Comparativa de Bacterias del Suelo

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

ANGÉLICA MARIANA JARA SERVÍN

TUTOR PRINCIPAL

Dr. Luis David Alcaraz Peraza
Facultad de Ciencias

MIEMBROS DEL COMITÉ TUTOR

Dra. María del Rocío Cruz Ortega
Instituto de Ecología

Dr. Luis Servín González
Instituto de Investigaciones Biomédicas

Ciudad de México, marzo de 2024.

PROTESTA UNIVERSITARIA DE INTEGRIDAD Y HONESTIDAD ACADÉMICA Y PROFESIONAL

(Graduación con trabajo escrito)

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la Institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado:

Interacciones Ecológicas: Estudio de Microbiomas de Plantas y Genómica Comparativa de Bacterias del Suelo.

que presenté para obtener el grado de Doctorado es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi programa de posgrado, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de graduación.

Atentamente

Angélica Mariana Jara Servín

Num. Cuenta: 517026317

ÍNDICE

AGRADECIMIENTOS INSTITUCIONALES

Al programa de Maestría y Doctorado en Ciencias Bioquímicas de la Universidad Nacional Autónoma de México.

Este trabajo fue financiado por los proyectos DGAPA-PAPIIT-UNAM IN206824 e IN221420.

A la beca otorgada por el programa de Becas Nacionales CONACyT (CVU 765278).

Al Programa de Apoyo a los Estudios de Posgrado (PAEP) para participar en los congresos internacionales World Microbe Forum 2021 (modalidad virtual) y ASM Microbe 2023 (modalidad presencial en Houston, TX, USA).

A la M. en C. Laura Patricia Olguín Santos por su apoyo en el invernadero de la Facultad de Ciencias.

A la Dra. Mariana Peimbert Torres, Dra. Clara Tinoco-Ojanguren, Dr. Hugo R. Barajas de la Torre, Dr. Miguel F. Romero Gutierrez, M.en C. Adán Silva y M. en C. Gerardo Mejía Vázquez por su colaboración y asesoría experimental y conceptual.

A los miembros de mi comité tutor: la Dra. María del Rocío Cruz Ortega y el Dr. Luis Servín González.

A mi tutor principal, el Dr. Luis David Alcaraz Peraza.

AGRADECIMIENTOS A TÍTULO PERSONAL

A mis papás y hermano, por el apoyo a lo largo de estos cuatro años.

A Janat, por estar pendiente y presionar a la distancia.

A Jaqueline y Adrián, por todas las veces que me dejaron ensayar con ellos.

A los Blue Runners 365, que me ayudaron a liberar estrés con tantos kilómetros que hemos corrido.

A Cristóbal, por explicarme y ayudarme a incursionar en la bioinformática.

A Hugo, por su apoyo diario dentro y fuera del laboratorio. Incluso desde Suecia.

A Gerardo, por recordarme que existen los buenos amigos y por siempre estar dispuesto a abrir pista conmigo.

Al resto de integrantes del laboratorio de Genómica Ambiental con los que coincidí: Ana, Chora, Leo, Leslie, Miguel y Nathalia. *Terribilis est haec laboratorium*.

A Luis, por su voto de confianza para aceptarme al doctorado y por haberme acompañado y asesorado para sortear todos los obstáculos que impuso la pandemia.

1. RESUMEN

El microbioma de raíz de las plantas influye en el establecimiento y desarrollo de éstas. Este trabajo de tesis se enfoca en las interacciones planta-microorganismo y su papel en la distribución del pasto africano buffel (*Pennisetum ciliare)* y el género bacteriano *Solirubrobacter*, dos organismos abundantes en escalas geográficas amplias. El pasto buffel es una planta invasora introducida alrededor del mundo. Uno de los métodos a los que recurre para desplazar plantas nativas es la alelopatía: secreción de compuestos que afectan el crecimiento de otras plantas. A pesar de su amplia distribución, no se ha explorado su microbioma de raíz ni el efecto que los compuestos alelopáticos ejercen sobre éste. De manera análoga, *Solirubroabacter* es una Actinobacteriota abundante y ubicua en suelos y rizósferas mexicanos. Sin embargo, hasta la fecha existen sólo siete genomas y aislados y no se había realizado una descripción a nivel genético de este género.

El microbioma del pasto buffel se evaluó bajo diferentes tratamientos y tiempos y se identificaron taxa bacterianos asociados a su raíz mediante análisis del gen 16S rRNA. Obtuvimos un total de 2,164 *Amplicon Sequence Variants* (ASV) e identificamos 30 géneros conformando el microbioma núcleo, algunos de los cuales podrían producir compuestos antimicrobianos, promotores de crecimiento de plantas y contribuir a la expansión exitosa del buffel en varios entornos. A pesar de que se identificaron géneros significativamente más abundantes en presencia de compuestos alelopáticos, fue el desarrollo del buffel el factor que más influyó en la estructuración del microbioma de raíz. De acuerdo con esto, detectamos ocho géneros significativamente más abundantes en los distintos tiempos evaluados (*Nodosilinea_PCC-7104, Oceanibaculum, Pedobacter, Flavitalea, Ohtaekwangia, IS-44, Phytohabitans* y *Saccharothrix*).

Por otro lado, se construyó un pangenoma de los siete genomas de *Solirubrobacter* conformado por 19,645 familias de proteínas, 2,644 de las cuales conforman el genoma núcleo. *Solirubrobacter* tiene un amplio metabolismo para usar carbohidratos y tolerar estrés ambiental y, a pesar de estar reportado como un género no mótil, se identificaron proteínas y familias génicas relacionados con biosíntesis de flagelo. Realizando reclutamientos en metagenomas ambientales se evaluó la diversidad y prevalencia de *Solirubrobacter*, resultando estar asociada a diversos suelos mexicanos, pero ausente en otros seleccionados en una escala continental. En cambio, este género es ubicuo en rizósferas, independientemente de la especie o ubicación geográfica. De los metagenomas se obtuvieron 9,906 proteínas potencialmente nuevas para *Solirubrobacter*, incrementando así la descripción funcional y revelando diferencias en contenido génico de acuerdo a cada muestra ambiental. Mediante análisis del gen 16S rRNA de distintos suelos y rizósferas mexicanos obtuvimos 3,166 OTUs de *Solirubrobacter* y una red de co-ocurrencia que mostró la alta conectividad del género

con el resto de la comunidad. Esto demuestra que la diversidad de *Solirubrobacter* y el rol que juega en las comunidades microbianas ha sido escasamente estudiado y amerita desarrollarse en investigaciones futuras.

Conocer la diversidad y estructura de microbiomas de raíz de plantas invasoras, así como el material genético de los microorganismos puede proporcionar pistas sobre los ambientes en los que se distribuyen y prosperan, por lo que estudiar estas interacciones ecológicas puede mejorar nuestra comprensión de los procesos microbianos en el suelo y en el crecimiento de las plantas.

2. ABSTRACT

The root microbiome influences the establishment and growth of its plant host. This work focuses on plant-microorganism interactions and their role in distributing the African buffelgrass (*Pennisetum ciliare*) and the bacterium *Solirubrobacter*, two abundant and widely distributed organisms. Buffelgrass is an invasive plant introduced throughout the globe that uses allelopathy to displace native plants. Allelopathy is the secretion of compounds that hinder the growth of a recipient plant. Despite its distribution and impact on ecosystems, its root microbiome has not been thoroughly described, as well as the effects of the allelochemicals on its own microbiome. Similarly, *Solirubrobacter* is an abundant Actinobacteriota found in Mexican soils and rhizospheres from geographically distant locations. Only seven genomes are available in databases and up until now, its genetic content had not been described.

Using 16S rRNA sequences we evaluated the buffelgrass' root microbiome under different times and allelopathic treatments. A total of 2,164 Amplicon Sequence Variants (ASV) were obtained, and we identified 30 genera comprising the core microbiome. Among the core, several genera with the reported capacity to synthesize antimicrobials, vitamins, and plant-growth-promoting compounds are suitable candidates for contributing to buffelgrass' expansion capacity. Even though we identified a set of genera significantly more abundant depending on the allelopathic treatment, buffelgrass' development was the most influencing factor on microbiome composition. Accordingly, we detected a total of eight genera significantly more abundant in the evaluated times (*Nodosilinea_PCC-7104, Oceanibaculum, Pedobacter, Flavitalea, Ohtaekwangia, IS-44, Phytohabitans* y *Saccharothrix*).

On the other hand, a pangenome was constructed using the seven available *Solirubrobacter* genomes. Of the total 19,645 protein families, 2,644 comprised the core genome. *Solirubrobacter* showed a wide capacity to metabolize carbohydrates and tolerate environmental stress. Interestingly, proteins involved with flagellar biosynthesis were detected, although being reported as a non-motile genus. Metagenomic recruitment allowed us to explore diversity and prevalence in soils and rhizospheres, resulting in prevalence in Mexican soils but absence in other evaluated soils from a continental scale. In contrast, *Solirubrobacter* appeared ubiquitous in rhizospheres, regardless of plant species or geographic origin. From Mexican soils and rhizosphere metagenomes, 9,906 potential new proteins for *Solirubrobacter* were recruited, allowing us to expand the description of the genus while revealing genetic differences according to each environmental metagenome. Moreover, from Mexican soils and rhizosphere samples, we obtained a total of 3,166 OTUs assigned as *Solirubrobacter*. We used them to construct a co-occurring network where *Solirubrobacter* highly connected with the microbial community. The amount of OTUs reveals that *Solirubrobacter*'s diversity has yet to be

tackled, while its high connectivity highlights that the role *Solirubrobacter* plays in microbial communities should be further addressed.

Learning about root microbiome diversity and structure from invasive plants and the genetic content of microorganisms contained in root microbiomes may shed light on the environments where these organisms can thrive. Hence, studying ecological interactions between plants and microorganisms would improve our understanding of soil microbial processes and plant growth.

3. INTRODUCCIÓN

3.1. Las plantas y los microorganismos del suelo interactúan entre sí

El término microbioma hace alusión a un hábitat que incluye microorganismos (bacterias, hongos, arqueas, virus y protistas), sus genomas y las condiciones ambientales circundantes (Marchesi y Ravel, 2015). Los microbiomas pueden estar asociados a un hospedero, es decir, un organismo que alberga a otros mediante distintas interacciones. Las plantas hospedan comunidades de microorganismos en sus distintos tejidos y es precisamente el tejido al que se asocian el que determina el nombre de microbioma. De esta manera podemos tener microbiomas filosféricos, aquellos asociados con la superficie de la parte aérea de la planta, microbiomas endosféricos, las comunidades de microorganismos que habitan la parte interna de los tejidos, y microbiomas rizosféricos, aquellos que se asocian a la raíz por su parte externa (Dastogeer et al., 2020). La estructura y características de cada microbioma asociado a plantas varían como respuesta a un gran número de factores (Fitzpatrick et al., 2020), algunos de los cuales son comunes entre ellos, como la disponibilidad de nutrientes o el genotipo de la planta, mientras que otros impactan únicamente a algunos microbiomas; por ejemplo, los microbiomas filosféricos deben estar estructurados de tal manera que puedan prevalecer en un ambiente sujeto a corrientes de aire (Dastogeer et al., 2020).

En un contexto biológico, un organismo exitoso es aquel capaz de sobrevivir, reproducirse y adaptarse al ambiente en el que se desarrolla. Ésto depende en gran medida de las características intrínsecas de cada organismo, es decir, las características y propiedades inherentes al organismo y que no son fácilmente alteradas por factores externos, tales como su fisiología o formas de reproducción, y que le permiten desenvolverse en un ambiente determinado. No obstante, las interacciones con otros organismos también son parte fundamental para el desarrollo de un organismo, así como de la estructuración de ecosistemas y procesos ecológicos (Nemergut et al., 2013). Las interacciones dentro de la misma comunidad con microorganismos menos representados también contribuyen a los patrones de estructuración general. Por ejemplo, se ha reportado que la presencia de bacteriófagos específicos regula la abundancia de grupos bacterianos y algunos nemátodos son conocidos por depredar bacterias (Thakur y Geisen, 2019; Naureen at al., 2020). A pesar de que bajo ciertos contextos también se desarrollan interacciones negativas, la asociación hospedero-microbioma suele basarse en interacciones biológicas benéficas o neutrales (Braga et al., 2016; Miller et al., 2018). Este tipo de interacciones contribuyen a que las plantas puedan desarrollarse y obtener del medio de los elementos necesarios para su crecimiento.

Explorar la diversidad microbiana de suelos y raíces para estudiar microorganismos capaces de interactuar con las raíces de las plantas es uno de los enfoques necesarios para explicar parcialmente el establecimiento y desarrollo de plantas hospederas (Klock et al., 2015; Fonseca et al., 2018). El estudio de estas comunidades se ha abordado a través tanto de métodos tradicionales de cultivo como de métodos independientes de cultivo, entre los que se encuentran la secuenciación de genes marcadores como el 16S rRNA, secuenciación de metagenomas totales y secuenciación de genomas completos, entre otros (Loman and Pallen, 2015; Knight et al., 2018; Donachie et al., 2021).

3.2. *Pennisetum ciliare*: un pasto alelopático

El pasto buffel, *Pennisetum ciliare*, es un pasto invasor originario del este de África que se ha introducido para pastoreo en el sur de Estados Unidos, México, varias regiones de latinoamérica y en Australia (Marshall et al., 2012). Entre las ventajas competitivas que tiene este pasto se encuentran la viabilidad de sus semillas hasta por ocho meses, sistema radicular de más de 2 m de profundidad, contenido nutricional apropiado para ganado y resistencia a sequía (Hussain et al., 1982; Marshall et al., 2012; Lyons et al., 2013). Sin embargo, desde su introducción se ha extendido hasta cubrir un 12% del territorio mexicano y un 53% de la superficie del estado de Sonora (Arriaga et al., 2004). En el desierto de Sonora se han reportado disminuciones de hasta 53-73% de especies nativas, de las familias Fabaceae y Cactaceae, que han sido sustituidas por pastizales de buffel (Tinoco-Ojanguren et al., 2013). Uno de los mecanismos que utiliza esta planta para invadir es la alelopatía: síntesis y secreción de compuestos aleloquímicos (metabolitos secundarios) que influyen en el crecimiento de una planta receptora, independientemente de la disponibilidad de nutrientes (Saul-Tcherkas et al., 2013; Schandry y Becker, 2020). Los aleloquímicos del pasto buffel no están caracterizados en su totalidad pero se han identificado compuestos fenólicos, entre los que se encuentran los ácidos *p*-*OH*-benzoico, *p*-cumárico, caféico, vanílico, ferúlico, siríngico y gentísico (Hussain et al., 2011).

A pesar de la dispersión del pasto buffel y del desplazamiento de plantas nativas, la estructura y diversidad taxonómica de su microbioma no han sido exploradas. El microbioma de raíz de una planta juega un papel importante en el establecimiento y desarrollo de ésta, por lo que el perfil de exudados tan particulares de una planta invasora podría reclutar microorganismos con contenidos génicos específicos que contribuyan a la expansión de la misma (Klock et al., 2015; Jousset et al., 2008; Weston y Mathesius, 2014).

3.3. Microbioma rizosférico: estructuración y funciones

El proceso de reclutamiento inicia con la rizodeposición: secreción de compuestos orgánicos al suelo a través de la raíz. Estos compuestos suelen ser ricos en carbohidratos tales como celulosa, pectina, xilano y manosa, pero también contienen flavonoides, aminoácidos y ácidos orgánicos (Bulgarelli et al., 2013). Estos rizodepósitos actúan como señal para que los microorganismos se aproximen a la raíz mediante quimiotaxis: los exudados de raíz de la planta fungen como moléculas señal para los quimioreceptores de los microorganismos, los cuales desencadenan una respuesta de motilidad a través de falgelos o pilis (Knights et al., 2021). Una vez en la rizósfera, los microorganismos son capaces de adherirse y colonizar la raíz a través de proteínas como las adhesinas y las invasinas (Bishop y Rachwal, 2014), así como por la formación de biopelícula o *biofilms* a través del sistema de percepción de *quorum* o *quorum sensing* (Levy et al., 2018).

Se ha propuesto que el establecimiento de comunidades microbianas asociadas a las raíces está regida por el llamado modelo de dos pasos (Bulgarelli et al., 2013). Este modelo propone que los exudados de raíz atraen a microorganismos específicos a la zona de rizósfera (región del suelo inmediata a la raíz), llevando a cabo un primer filtrado de los microorganismos del suelo. En el segundo paso del modelo, las características de la pared celular de las plantas y del genotipo del hospedero seleccionan una porción de los microorganismos de la rizósfera que pasan a formar parte del microbioma de la endósfera, definida como la parte interna de la raíz de la planta (Lundberg et al., 2012; Bulgarelli et al., 2013). Estos microbiomas están sujetos a variaciones de la planta y a factores edáficos y derivados que determinan la composición de las comunidades microbianas alrededor y dentro de las raíces de la planta (Tkacz et al., 2015; Bulgarelli et al., 2013; Lundberg et al., 2012) (Figura 1).

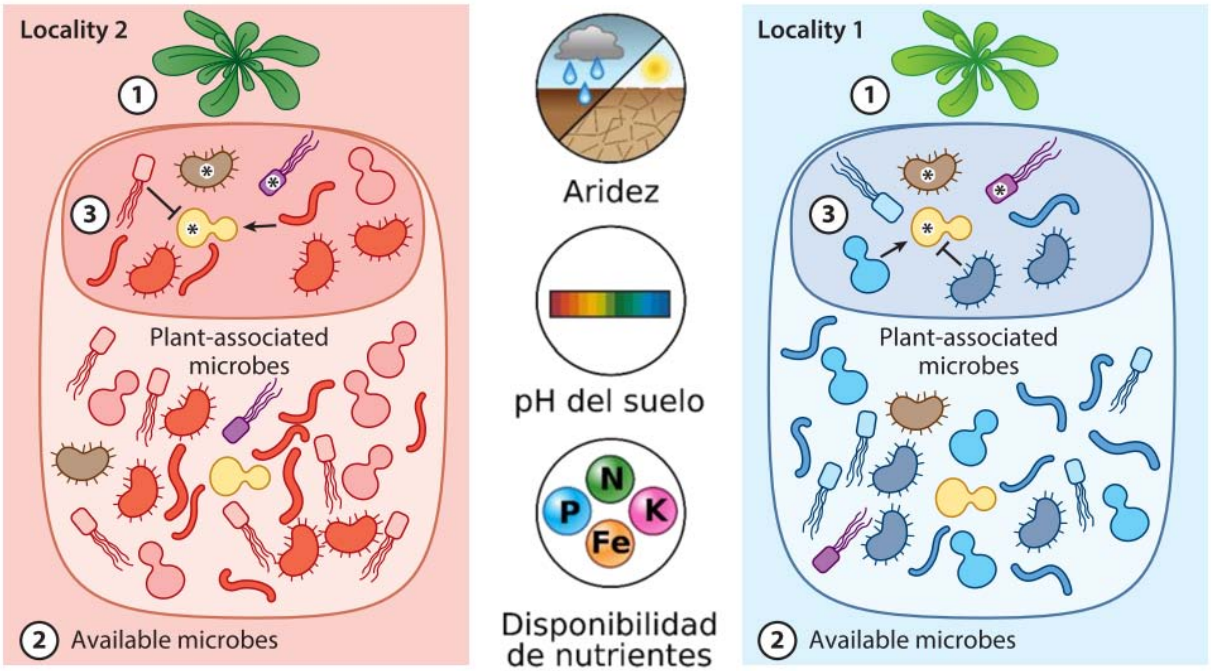


Figura 1. Factores que influyen en la estructuración del microbioma de raíz. Dentro del gradiente ambiental existen regiones o localidades que presentan comunidades microbianas particulares, con algunos miembros que pueden establecer interacciones con plantas. Factores abióticos como la aridez, pH y disponibilidad de nutrientes, crean gradientes ambientales que determinan la diversidad microbiana del suelo. Modificado de Fitzpatrick et al., 2020.

Cualquier microorganismo que se asocia a la planta hospedero puede ser benéfico, dañino o neutro para la planta y el rol que desempeñan puede variar de acuerdo a fluctuaciones ambientales o a la expresión de genes, tanto del hospedero como de otros miembros de la comunidad (Levy et al., 2018). Como ya se mencionó, los microorganismos contribuyen al desarrollo de la planta de diversas formas. Las fitohormonas son compuestos que influyen en el crecimiento de la planta y, entre las más relevantes, se encuentran las auxinas, citoquininas, giberelinas, ácido abscísico y etileno (Bulgarelli et al., 2013, Sabki et al., 2021). Cada uno de estos compuestos estimula el crecimiento de manera distinta. Por ejemplo, las auxinas, entre las que se incluye al ácido indol-acético, se han relacionado con el elongamiento de raíz (Park et al., 2021), mientras que las giberelinas activan genes involucrados con la floración (Salazar-Cerezo et al., 2018). Otra contribución que las bacterias le ofrecen a la planta es la protección ante patógenos mediante la síntesis y secreción de compuestos antimicrobianos, así como la capacidad de metabolizar compuestos dañinos (Levy et al., 2018; Trivedi et al., 2020), como compuestos aromáticos o metales pesados. Además, existen microorganismos benéficos que contribuyen al desarrollo de la resistencia sistémica inducida que protege las partes de

la planta no expuestas al ataque de patógenos o herbívoros (Pieterse et al., 2014, Fitzpatrick et al., 2020).

Por otro lado, los factores abióticos que contribuyen a la composición de las comunidades microbianas son varios, pero los que se han identificado ejerciendo un efecto más notorio son pH, contenido de nitrógeno, contenido de carbono orgánico, temperatura y estado redox (Fierer, 2017). En muestras geográficamente distantes se ha identificado la formación de grupos microbianos con estrategias de vida similares que les permiten desarrollarse en ambientes de características particulares, como ambientes con valores de pH altos (Delgado-Baquerizo et al., 2018). Esto indica que la composición de las comunidades microbianas está influenciada por las características del ambiente al que pertenecen, independientemente de su ubicación geográfica, y que podrían promover adaptaciones fisiológicas, así como diferencias en la diversidad y riqueza de las comunidades microbianas de acuerdo a su ubicación geográfica y hospedero (Malik et al., 2017; Fitzpatrick et al., 2020). Sin embargo, existen taxa bacterianos que son abundantes en suelos distintos y geográficamente distantes (Brewler et al., 2016), probablemente debido al potencial metabólico contenido en cada una de estos grupos.

3.4. Existen taxa ubicuos en suelos y rizósferas

Toda descripción taxonómica de una comunidad microbiana permite una mejor comprensión de las interacciones ecológicas dentro de la comunidad y con el hospedero. A pesar de la información que podemos obtener a partir de determinar la diversidad taxonómica microbiana, es importante contemplar el aspecto funcional. Esto contribuye a la comprensión de mecanismos moleculares y de interacción, ya que todos estos procesos están regulados por la expresión de genes como respuesta a señales ambientales, las cuales abarcan parámetros físicoquímicos del ambiente, señales generadas por la misma comunidad microbiana y señales producidas por el hospedero (Braga et al., 2016). Esta información está codificada en los genes contenidos en los genomas que se relacionan con las características del ambiente (Ranjard y Richaume, 2001); cuando un organismo se adapta genéticamente a un ambiente puede perder características que le permitan desarrollarse en otros contextos (Cooper y Lenski, 2000), por lo que pueden esperarse cambios en el genoma de una misma especie que habita diversos ecosistemas. Aunado a esto, existen taxa bacterianos ampliamente dispersos, tanto en términos de características ambientales como en términos de ubicación geográfica, cuya diversidad genética se encuentra pobremente estudiada y que se traduce en una falta de información respecto a adaptaciones ambientales y locales. Uno de estos grupos taxonómicos es Actinobacteriota, un phylum que parece ser ubicuo en suelos, independientemente de

su ubicación geográfica y características edáficas (Delgado-Baquerizo et al., 2018; Barajas et al., 2020). Dentro del phylum Actinobacteriota se encuentra el género *Solirubrobacter*, un microorganismo que se ha detectado en suelos con distintas características fisicoquímicas y asociada a la raíz de hospederos como pastos y tomates, además de estar entre las bacterias más abundantes en dichas muestras (Barajas et al., 2020).

3.5. *Solirubrobacter*: un género bacteriano desapercibido

Solirubrobacter es un género bacteriano perteneciente al phylum Actinobacteriota, clase Thermoleophilia, orden Solirubrobacterales y familia *Solirubrobacteraceae*. La primera cepa de este género fue aislada de lombricomposta en 2003, recibió el nombre de *Solirubrobacter pauli* (Singleton et al., 2003) y con ella se describió por primera vez el género: bacilos Gram positivos, aeróbicos y mesófilos que no forman esporas ni estructuras de motilidad y su contenido de G+C es de aproximadamente 70% (Singleton et al., 2003; Whitman et al., 2015). *Solirubrobacter pauli* presentó además poca resistencia a la desecación (Singleton et al., 2003). Actualmente el género contiene a las especies *S. pauli, S, soli, S. ginsenosidimutans, S. phytolaccae* y *S. taibaiensis* aisladas de lombricomposta, campos de gingseng y raíces de *Phytolacca acinosa* (Singleton et al., 2003; Kim et al., 2007; An et al., 2011; Wei et al., 2014; Zhang et al., 2014). A partir de ese año el número de aislados y genomas disponibles ha incrementado de forma mínima, de manera que actualmente hay disponibles un total de ocho genomas en las bases de datos (Tabla 1). Curiosamente, este género someramente descrito se encuentra ampliamente distribuido y es de los más abundantes en distintos suelos y hospederos. En un estudio previo, donde se buscó determinar el efecto de suelos geográficamente distantes y taxonómicamente diversos sobre la estructuración del microbioma, se reportó a *Solirubrobacter* como la Actinobacteriota más abundante en todas las muestras (Barajas et al. 2020). Aunado a esto, varios estudios ambientales han detectado a este género bacteriano mediante secuenciación del gen 16S rRNA, especialmente en suelos y rizósferas de diversos hospederos. Se han reportado correlaciones tanto positivas como negativas con el crecimiento de distintas plantas, lo que sugiere algún tipo de interacción entre la planta y *Solirubrobacter* que se ve reflejado en el desarrollo de la planta hospedero (Yang et al., 2012; Franke-Whittle et al., 2015; Hernández-Álvarez et al., 2022).

Tabla 1. Genomas de referencia de *Solirubrobacter* disponibles en NCBI. Modificado de Jara-Servín et al., 2023 b. *bioRxiv*.

Strain	Accession	Size (Mb)	Completeness	Country	Environment
S. soli DSM22325	ASM42366v1	9.31	98.71	South Korea	Ginseng field
S. sp URHD0082	ASM42594v1	6.64	99.14	N.d.	Mediterranean grassland
S. sp CPCC204708	ASM334462v2	7.59	98.71	China	Desert sand
S. pauli	ASM363375v1	7.13	98.71	USA	Vermicompost
S. ginsenosidimutans	ASM2758720v1	9.69	99.57	China	Ginseng field
S. phytolaccae	ASM2758719v1	7.73	98.71	China	Root of *Phytolacca acinosa*
S. taibaiensis	ASM2758722v1	8.51	98.71	China	Stem of *Phytolacca acinosa*
ASM999324v1	ASM999324v1	2.02	93.18	USA	Aquifer samples (geyser)

La premisa “Todo está en todos lados y el ambiente selecciona” considera que, debido a las características de las comunidades microbianas, éstas son propensas a dispersarse a distintos ambientes y, una vez que llegan a un nuevo ambiente es dicho ambiente el que filtra y selecciona qué microorganismos pueden desarrollarse en él (Martiny et al., 2006; Lindström y Langenheder, 2012). El uso de datos genómicos de diferentes cepas de la misma especie contribuye a responder preguntas sobre la fisiología bacteriana, identificación de genes esenciales para el metabolismo, procesos de patogénesis, procesos de adaptación al ambiente, patrones de dispersión y procesos evolutivos (Tettelin et al., 2005; Fitzpatrick et al., 2020). La capacidad de éxito de estos microorganismos debe estar contenida en los distintos genes que conforman sus genomas, definidos hasta cierto punto por el mismo ambiente (Fondi et al., 2016). Conocer el contenido genético de éstos y su relación con distintos factores ambientales y con su hospedero es necesario para discernir procesos evolutivos y sus interacciones tanto con el hospedero como con otros microorganismos de comunidades bacterianas (Fierer, 2017). Al estar presente en suelos contrastantes y en rizósferas de diversos hospederos, el contenido génico de *Solirubrobacter* podría variar de acuerdo al ambiente pero contener al mismo tiempo un arsenal genético que le permite colonizar suelos y raíces geográficamente distantes.

3.6. Pangenomas como herramienta de genómica comparativa

La genómica comparativa es una rama de la biología que consiste en comparar las secuencias del genoma de distintos organismos. Ésta juega un papel crucial en la descripción y clasificación de géneros o especies bacterianas, ya que involucra desarrollar comparaciones de genomas o secuencias de distintos organismos con el fin de identificar similitudes, diferencias y relaciones

evolutivas. Además, nos permite analizar la relación entre los genes codificados en el genoma de los microorganismos y las características del ambiente del que provienen (Ranjard y Richaume, 2001).

Al principio de la era genómica se pensaba que un genoma era suficiente para representar la diversidad genética de una especie. Sin embargo, con el paso del tiempo y con el uso de genómica comparativa, se hizo evidente que la variación interespecie es significativa. Por esta razón, el uso de datos genómicos de diferentes cepas de la misma especie nos permite responder preguntas sobre la fisiología bacteriana, identificación de genes esenciales para el metabolismo, procesos de patogénesis, procesos de adaptación al ambiente y procesos evolutivos (Tettelin et al., 2005; Fitzpatrick et al., 2020). Una herramienta utilizada para este fin es la construcción y análisis de pangenomas. El pangenoma es la suma de dos componentes: el genoma núcleo, que es el conjunto total de genes que comparten todas las cepas, y el genoma accesorio, compuesto por los genes que se encuentran en una o más cepas, así como los genes únicos de cada una de ellas. El genoma núcleo nos permite identificar los genes que deben estar presentes para poder definir por completo a una especie bacteriana, mientras que el genoma accesorio brinda información acerca de funciones que no son esenciales pero que representan una ventaja selectiva, como adaptación de nicho o resistencia a antibióticos (Tettelin et al., 2008). Una de las ventajas de realizar análisis del genoma núcleo de un grupo particular es que se pueden revelar grupos de genes conservados dentro de un rango taxonómico, de manera que se puedan encontrar funciones que determinen la cohesión del grupo o que brinden ciertas características fenotípicas.

Las características del pangenoma pueden relacionarse con el estilo de vida generalista o especialista de los microorganismos. De esta manera, especialistas que ocupan un nicho aislado, como los endosimbiontes, presentarán un pangenoma cerrado o que tenga una capacidad baja para obtener nuevos genes (Tettelin et al., 2005). Por el contrario, microorganismos que habitan en ecosistemas abiertos suelen tener pangenomas abiertos que le permitan integrar los genes necesarios para sobrevivir en un mayor número de condiciones ambientales (Bell y Bell, 2020). Sin embargo, es importante aclarar que los pangenomas abiertos o cerrados no están estrictamente asociados a su equivalente ambiental; en ambientes abiertos es posible también encontrar géneros con pangenoma cerrado, tal y como sucede con especies clonales o capaces de esporular (Rouli et al., 2014). El término especialista también se ha llegado a relacionar con el microbioma núcleo del hospedero, bajo la idea de que el microbioma núcleo es resultado de un proceso de especialización con respecto al hospedero (Lloyd-Price et al. 2017; Hamonts et al. 2018).

Esta distinción entre organismos generalistas y especialistas se ha empleado con el fin de identificar aquellos organismos que son sensibles a cambios ambientales (Bell y Bell, 2020). En relación con

eso, se ha planteado que los microorganismos generalistas suelen depender más de mecanismos de dispersión para colonizar distintas regiones, mientras que los organismos especialistas están sujetos a filtrado ambiental, porque tienen requerimientos ambientales que sólo pueden satisfacerse en ambientes específicos (Pandit et al., 2009). Aún así, este filtrado ambiental también se ha identificado en microorganismos clasificados como generalistas, revelando que los procesos de establecimiento y prevalencia de los organismos son eventos multifactoriales (Bissett et al., 2010; Székely y Langenheder, 2014; Custer et al., 2022).

4. PLANTEAMIENTO DEL PROBLEMA

Existen organismos que se encuentran distribuidos en escalas geográficas amplias y que son capaces de desarrollarse en ambientes de distintas características. Las interacciones que estos organismos establecen contribuyen al éxito tanto en su dispersión como en su prevalencia. Explorar estas interacciones proporciona información acerca de los factores implicados en el éxito de este tipo de organismos y, al mismo tiempo, nos permite analizar la biodiversidad taxonómica y metabólica de éstos.

Tanto el microbioma del pasto buffel como la genómica de *Solirubrobacter* se encuentran pobremente estudiados, lo que se traduce en una falta de información respecto a su distribución y adaptaciones. El estudio de microbiomas de raíces y contenido genético de microorganismos nos permite dilucidar las relaciones que existen entre plantas, microorganismos y ambiente. Bajo este contexto, este trabajo se enfoca en evaluar el papel que juegan las interacciones planta-microorganismo en dos organismos geográficamente dispersos: el primero en la planta *Pennisetum ciliare*, haciendo énfasis en la estructuración y diversidad de su microbioma de raíz, y el segundo en el género *Solirubrobacter*, enfocándose en un su diversidad genética y distribución ambiental.

5. HIPÓTESIS

Existirán interacciones planta-microorganismo que desempeñen un papel significativo en la distribución de plantas y bacterias en escalas geográficas amplias. Estas interacciones podrán abarcar una variedad de mecanismos, desde simbiosis mutualistas hasta dinámicas de adaptación, y su estudio podría contribuir a comprender la estructura y diversidad de los ecosistemas a nivel regional o incluso global.

5.1. Hipótesis particulares

- Los microorganismos reclutados en las raíces del pasto buffel estarán determinados por los aleloquímicos de la planta y podrían contribuir al establecimiento de ésta.
- *Solirubrobacter* tendrá un contenido génico vasto que le permitirá desarrollarse en distintos suelos y raíces.

6. OBJETIVO

Identificar las interacciones planta-microorganismo que contribuyen a la dispersión y prevalencia en escalas geográficas amplias de la planta *Pennisetum ciliare* y del género bacteriano *Solirubrobacter*.

6.1. Objetivos particulares

- Describir el microbioma rizosférico de *Pennisetum ciliare*.
- Evaluar cambios en el microbioma de *Pennisetum ciliare* bajo el efecto de exudados de raíz y lixiviados de la parte aérea.
- Describir el microbioma asociado a las raíces del pasto *Pennisetum ciliare* en una escala corta de tiempo.
- Discutir las contribuciones que el microbioma rizosférico podría proveer a *Pennisetum ciliare*.
- Construir un pangenoma de *Solirubrobacter* a partir de genomas de referencia.
- Realizar un análisis genómico comparativo de *Solirubrobacter* para explicar su ubicuidad.
- Evaluar la diversidad de *Solirubrobacter* en metagenomas ambientales.
- Reclutar nuevas proteínas de *Solirubrobacter* y construir un pangenoma ambiental extendido.
- Buscar genes de interacción planta-microorganismo en el pangenoma ambiental extendido.

7. METODOLOGÍA

7.1. Microbioma rizosférico de *Pennisetum ciliare*

7.1.1. Colecta de semillas, suelos y cultivo bajo condiciones alelopáticas.

Las semillas de *Pennisetum ciliare* fueron obtenidas de cinco plantas a finales del mes de septiembre de pastizales de buffel en Rancho Diamante (28˚ 41' N, 110˚ 15' W), Sonora, Mexico. Simultáneamente se colectaron muestras de suelo para usarlo en los experimentos de jardín común. La cubierta de las semillas se retiró y se germinaron en cajas Petri con 2% agar en una cámara de crecimiento Biotronette (20.2L, 25ºC, 12 h de fotoperiodo, 92 µmolm-2s-1). Dos días después las plántulas se trasplantaron a contenedores con *peat moss* donde crecieron por 70 días y fueron regadas cada dos semanas con solución Hoagland. Posteriormente se trasplantaron a un sistema de tubos PVC que contenían, en una proporción 1:1, arena sílica estéril y suelo de Rancho Damiante. Las plantas en estos tubos fueron sometidas a uno de tres tratamientos: exudados de raíz (C), lixiviados de la parte aérea (B) y control (A) (Figura 2). El tratamiento de exudados consistió en exponer a las raíces del pasto buffel a los exudados de raíz de otro individuo de pasto buffel, el tratamiento de exudados implicaba riego con una mezcla de agua y lixiviados de otro individuo buffel y el tratamiento control se basó en riego únicamente con agua.

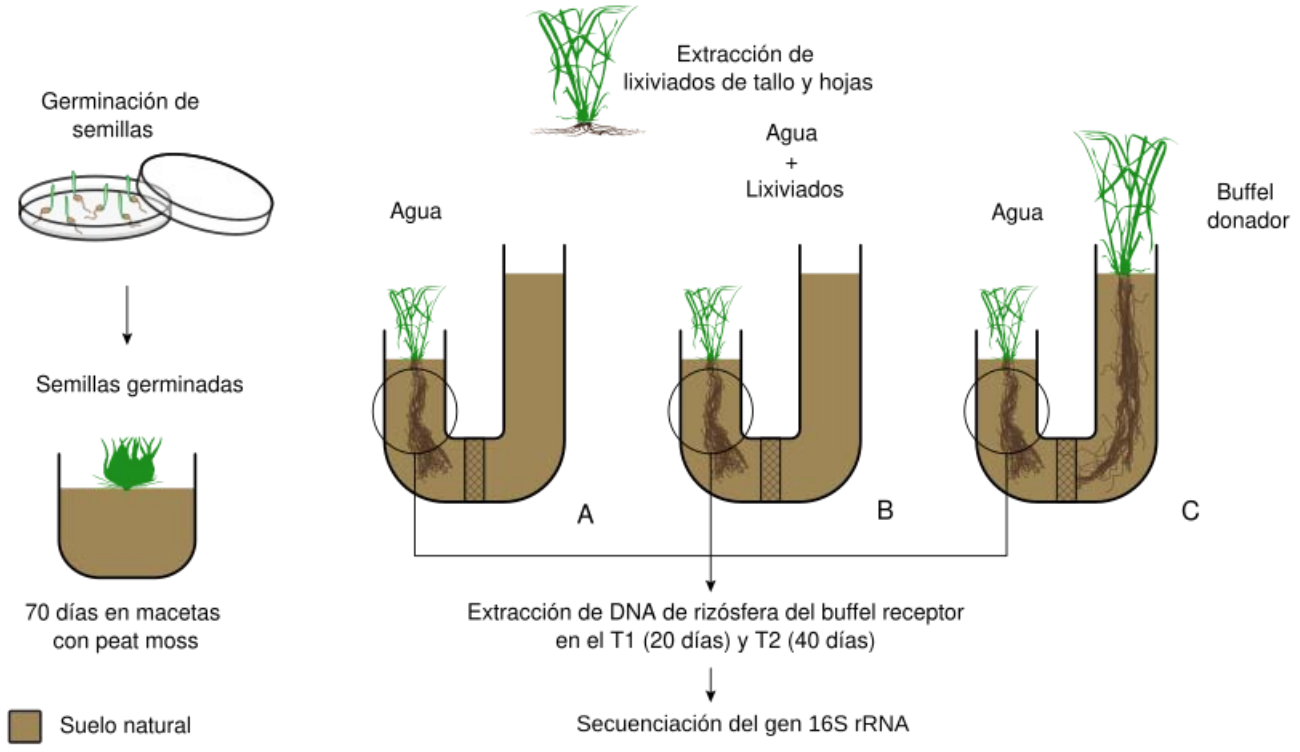


Figura 2. Diseño experimental. Las semillas de buffel se germinaron y transfirieron a recipientes con *peat moss*. Después de 70 días se trasladaron a tubos PVC con arena sílica y suelo de Rancho Diamante en una

proporción 1:1. Las plantas se sometieron a tres tratamientos con cuatro réplicas cada uno: (A) riego con agua destilada estéril, (B) riego con solución de lixiviados de la parte aérea del buffel y (C) riego con agua destilada esteril pero con exposición de las raíces del buffel a los exudados de raíz de otra planta buffel. Se tomaron muestras a los 20 días de crecimiento de dos réplicas y a los 40 días de las dos réplicas restantes. Tomado de Jara-Servín, 2023 a.

7.1.2. Obtención de muestras de rizósfera, procesamiento y secuenciación del gen 16S rRNA

Se tomaron muestras de rizósfera de las plantas buffel a los 20 días (T1) y a los 40 días (T2) de haber estado en el sistema de tubos PVC. Se obtuvo la rizósfera removiendo el suelo suelto cercano a la raíz para posteriormente lavar y someter las raíces a sonicación en buffer 1X de PBS (137 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4). Los pellets se recuperaron por centrifugación por 10 min a 1,300 g usando tubos de 50 mL y fueron mantenidos a -80ªC hasta el momento de la extracción.

La extracción de DNA se llevó a cabo con el kit para extracción de DNA Mobio PowerSoil (MoBio, Carlsbad, CA, United States) con las siguientes modificaciones: la solución C6 se calentó a 60ºC antes de la elución para incrementar el rendimiento de DNA y se realizaron dos rondas de elución con 30 µL en el mismo filtro. La amplificación del gen 16S rRNA se realizó por duplicado de acuerdo al protocolo de Illumina MiSeq (Illumina 2013) con los oligonucleótidos 341F/805R (región V3–V4) y los *overhangs* adaptadores 5’ (341F: 50 -CCTACGGGNGGCWGCAG-30; 805R: 50 -ACTACHVGGGTATCTAATCC 30). La reacción de PCR se realizó utilizando 2 µL de buffer, 1.2 µL de cada oligonucleótido (5 µM), 3 µL *enhancer*, 1.6 µL dNTPs (2.5 mM), 0.6 µL Mg_2S0_4 (1.5µM), 9.2 µL de agua grado molecular, 0.16 µL polimerasa Pfx (0.02 U/µL) (Invitrogen, Thermo Fisher Scientific, Waltham, MA) y 2 µL de DNA (volumen final de 20 µL). Las condiciones para la reacción de PCR fueron: temperatura inicial de 96ºC por 3 min seguido de 25 ciclos de 5s a 94ºC y 30s a 68ºC. Los productos amplificados se purificaron usando el kit SV Wizard (Promega, Madison, WI) y se secuenciaron con Illumina MiSeq (2 × 300 *paired-end*) en la Unidad se Secuenciación Masiva y Bioinformática del Instituto de Biotecnología de la UNAM.

7.1.3. Análisis de secuencias del gen 16S rRNA

Todas las librerías de amplicones pasaron por los filtros de calidad de Dada2 (v. 1.10.1) (Callahan et al., 2016), las primeras 17 pares de bases se removieron y las secuencias se recortaron hasta una

longitud de 250 pares de bases. Dado que las secuencias *reverse* no pasaron los filtros de calidad se utilizaron únicamente las secuencias *forward.* Todas las secuencias se procesaron para obtener Amplicon Sequence Variants (ASV) con Dada2 (v. 1.10.1). La asignación taxonómica se realizó con la base de datos Silva (v. 138) (Yilmaz et al., 2014) y con FastTreeMP (Price et al., 2009) se construyó un árbol filogenético.

La diversidad α y β se calcularon con los paquetes default de R (R Core Team, 2013) y phyloseq (McMurdie y Holmes, 2013), ggplot2 (Wickham y Chang, 2016) y vegan (Oksanen, 2017). La diversidad α se determinó con los índices de diversidad Shannon y Simpson, mientras que la diversidad β se evaluó con una agrupación jerárquica y un análisis de coordenadas principales (CAP) en una matriz de distancia de distancia UniFrac no ponderada (Lozupone y Knight, 2005). Las agrupaciones fueron evaluadas con la función estadística ANOSIM (Clarke, 1993) y se detectaron ASV significativamente abundantes por comparaciones entre tratamientos utilizando el paquete de R DESeq2 (Love et al., 2014).

7.2. Genómica comparativa de *Solirubrobacter*

7.2.1. Obtención de genomas y análisis filogenético

En el mes de febrero de 2023 se descargaron los ocho genomas de *Solirubrobacter* disponibles en NCBI y se evaluó su completitud utilizando Checkm (Parks et al., 2015) (Tabla 1). Utilizando GET_HOMOLOGUES (Contreras-Moreira y Vinuesa, 2013) se calculó un pangenoma de las cepas de *Solirubrobacter* y *Streptomyces coelicolor y Bacillus subtilis* como grupos externos (NCBI Accessions: GCA_013317105.1 y GCF_000009045.1). Los genes del core se concatenaron y alinearon con MAFFT (Katoh, 2002) para posteriormente construir una filogenia de máxima verosimilitud con 1,000 *bootstraps*. Con base en el criterio de información de Akaike el modelo evolutivo que se utilizó fue LG (Le y Gascuel, 2008). Simultáneamente, se calculó el Genome Similarity Score (GSS) (Moreno-Hagelsieb et al., 2013) por comparaciones pareadas y se construyó un dendrograma. Con base en los resultados obtenidos se retuvieron siete de los ocho genomas de referencia anotados como *Solirubrobacter*.

7.2.2. Diversidad y redes de co-ocurrencia de *Solirubrobacter* usando el gen 16S rRNA

Los OTUs correspondientes a *Solirubroabcter* de suelos, rizósferas y endósferas de México (NCBI Bioproject PRJNA603603) se extrajeron (3,166 de un total de 255,275 OTUs) y, junto son los siete genes 16S de las referencias (extraídos usando barrnap v. 0.7) se construyó un árbol filogenético usando FastTreeMP después de alinear con SSU-align (Nawrocki y Eddy, 2013). La red de co-ocurrencia se calculó con la librería igraph (Csardi y Nepusz, 2006) de R utilizando el conjunto de OTUs completo de México (255,275 OTUs) y colapsando a 1,748 géneros distintos. La red se visualizó y analizó con Gephi (v. 0.10) (Bastian et al., 2009).

7.2.3. Construcción y análisis del pangenoma de *Solirubrobacter*

Los genomas de referencia de *Solirubrobacter* se anotaron usando Prokka (v. 1.12) (Seeman, 2014) para obtener archivos gff3 que sirvieron de entrada para construir un pangenoma con Roary (Page et al., 2015). Con los archivos de salida se construyó un UpSet para visualizar la distribución de los genes núcleo, accesorio y únicos y se anotaron las secuencias utilizando las bases de datos Clusters of Orthologous Groups (COG) (Galperin et al., 2015) y Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al., 2016). Con la anotación de COG se construyó un mapa de calor usando los paquetes de R phyloseq, ggplot2 y gplots (Warnes et al., 2016) y haciendo un cluster jerárquico con una matriz de distancia de Bray-Curtis (Bray y Curtis, 1957).

Los aislados de *Solirubrobacter* se han reportado como no mótiles. Sin embargo, tras el análisis del perfil funcional obtenido se identificaron proteínas relacionadas con la biosíntesis de flagelo por lo que se decidió ahondar en este proceso. Se construyó un mapa de presencia-ausencia incluyendo las proteínas relacionadas con biosíntesis de flagelo presentes en *Conexibacter woesi* y *Patulibacter minatonensis* (Accession number: GCA_013317105.1 y ASM51932v1, respectivamente), ya que ambas son Actinobacteria flageladas.

7.2.4. Reclutamiento de secuencias de *Solirubrobacter* de metagenomas ambientales

Se seleccionaron 36 metagenomas de suelos y rizósferas de distintas ubicaciones geográficas. De éstos, 16 son metagenomas de suelos y rizósferas de México (NCBI Bioproject PRJNA603603), mientras que el resto fueron seleccionados si cumplían los siguientes requisitos: estar geográficamente distanciados de los metagenomas mexicanos y, en el caso de los suelos, provenir

de suelos agrícolas o, en el caso de las rizósferas, provenir de un hospedero distinto a los hospederos mexicanos. Como control se obtuvieron también metagenomas de sedimentos marinos, filósferas de distintas plantas y lagos. La información de todos los metagenomas usados se encuentra en la Tabla 2.

Utilizando tBLASTn (Altschul et al., 1990) se identificaron en los metagenomas todos los *reads* de *Solirubrobacter* y se sometieron a un análisis de calidad usando Trimmomatic (Bolger et al., 2014). A continuación, se utilizó el pangenoma para reclutar los *reads* de *Solirubrobacter* y obtener el porcentaje de identidad utilizando Promer (Delcher, 2002) y el *script promer_deid.py* para generar gráficos de porcentaje de identidad. Debido a la diferencia en cuanto a secuencias reclutadas, se decidió continuar únicamente con los metagenomas de suelos y rizósferas de México para la construcción del pangenoma ambiental extendido (EEP, por sus siglas en inglés *Environmental Extended Pangenome*).

Tabla 2. Metagenomas usados en este estudio. Tomado de Jara-Servín et al., 2023 b. *bioRxiv*.

Sample number	Metagenome tag	Acc. Num.	Environment	Host	Geographic origin
1	Mexico_1	SRR11092503	Soil	N.A.	Aguascalientes, Mexico
2	Mexico_2	SRR11092500	Soil	N.A.	Durango, Mexico
3	Mexico_3	SRR11092497	Soil	N.A.	Jalisco, Mexico
4	Mexico_4	SRR11092510	Soil	N.A.	Nayarit, Mexico
5	Mexico_5	SRR11092507	Soil	N.A.	Sinaloa, Mexico
6	Mexico_6	SRR11092504	Soil	N.A.	San Luis Potosí, Mexico
7	Canada_1	ERR1742273	Soil	N.A.	British Columbia, Canada
8	Canada_2	SRR6039452	Soil	N.A.	Ontario, Canada
9	Colombia_1	ERR5866886	Soil	N.A.	Amazon, Colombia
10	Colombia_2	SRR23012434	Soil	N.A.	Santander, Colombia
11	USA_1	SRR3989569	Soil	N.A.	Michigan, USA
12	USA_2	SRR3988848	Soil	N.A.	Wisconsin, USA
13	Poaceae_Mex_1	SRR11092511	Rhizosphere	*Poaceae*	Aguascalientes, Mexico
14	Poaceae_Mex_2	SRR11092501	Rhizosphere	*Poaceae*	Durango, Mexico
15	Asteraceae_Mex_3	SRR11092498	Rhizosphere	*Asteraceae*	Jalisco, Mexico
16	Fagales_Mex_4	SRR11092496	Rhizosphere	*Fagales*	Nayarit, Mexico
17	Poaceae_Mex_5	SRR11092505	Rhizosphere	*Poaceae*	San Luis Potosí, Mexico
18	Tomato_Mex_6	SRR11092512	Rhizosphere	*Solanum lycopersicum*	Aguascalientes, Mexico
19	Tomato_Mex_7	SRR11092502	Rhizosphere	*Solanum lycopersicum*	Durango, Mexico
20	Tomato_Mex_8	SRR11092499	Rhizosphere	*Solanum lycopersicum*	Jalisco, Mexico
21	Tomato_Mex_9	SRR11092509	Rhizosphere	*Solanum lycopersicum*	Sinaloa, Mexico
22	Tomato_Mex_10	SRR11092506	Rhizosphere	*Solanum lycopersicum*	San Luis Potosí, Mexico
23	Switchgrass_USA_11	SRR6957362	Rhizosphere	Switchgrass	Michigan, USA
25	Bean_USA_13	SRR15527069	Rhizosphere	*Phaseolus vulgaris*	Michigan, USA
26	Citrus_USA_14	SRR13588982	Rhizosphere	Citrus	California, USA
27	Sediments_1	SRR22177935	Sediments	N.A.	Indian Ocean
28	Sediments_2	SRR17063064	Sediments	N.A.	Pacific Ocean
29	Sediments_3	SRR3943820	Sediments	N.A.	Pacmanus Field
30	Sediments_4	DRR121220	Sediments	N.A.	North Western Pacific
31	Phyllosphere_1	SRR21860273	Phyllosphere	Pepper	Alabama, USA
32	Phyllosphere_2	SRR12131104	Phyllosphere	Tangerine	Zhejiang, China
33	Phyllosphere_3	DRR278742	Phyllosphere	Metrosideros	Hawaii, USA
34	Lake_1	ERR2681674	Lake	N.A.	Soyang, South Korea
35	Lake_2	SRR21524754	Lake	N.A.	Hells Canyon, USA
36	Lake_3	SRR18572986	Lake	N.A.	Xinjiang, China

7.2.5. Ensamblado de metagenomas de suelos y rizósferas mexicanos

Los 16 metagenomas de suelos y rizósferas de México del Bioproject PRJNA603603 en NCBI se sometieron a un control de calidad usando Trimmomatic. Para las secuencias retenidas de rizósferas se hizo un filtrado contra hospedero (*Solanum lycopersicum* para las muestras 18 a 22 y *Arabidopsis thaliana* para las muestras 13 a 17) utilizando Bowtie2 (Langmead y Salzberg, 2012). Las librerías metagenómicas se ensamblaron usando un ensamble híbrido que consistió en un primer ensamblado con SPADES (Nurk et al., 2017) y un segundo con Velvet (Zerbino y Birney, 2008). Los *contigs* de

ambos ensambles se unieron y se retuvieron únicamente aquellos de 100 pares de bases o más. Estos *contigs* se utilizaron para la construcción del EEP.

7.2.6. Construcción del Pangenoma Ambiental Extendido de *Solirubrobacter*

La estrategia empleada para construir el EEP está esquematizada en la Figura 3. Utilizando Bbmap (Bushnell, 2014) se buscaron las secuencias del pangenoma de referencia en los *contigs* de los metagenomas ensamblados. Todos los *contigs* que contenían secuencias del pangenoma de referencia se extrajeron y las proteínas codificadas alrededor de la secuencia identificada del pangenoma se consideraron potenciales proteínas nuevas para el género *Solirubrobacter*. Se realizó un segundo filtrado utilizando BLASTx (Gish y States, 1993) y al pangenoma de referencia como base de datos para la búsqueda y se determinó un valor de e mínimo de 1e-5. Posteriormente se usó Prodigal (Hyatt et al., 2010) para predecir ORFs y proteínas codificantes. Todas las secuencias se concatenaron y se analizaron con CD-HIT (Li y Godzik, 2006) para obtener el EEP conformado por las secuencias del pangenoma de referencia y las secuencias identificadas como posibles nuevas proteínas de *Solirubrobacter*. Al igual que con el pangenoma de referencia, el EEP se anotó utilizando COG y KEGG.

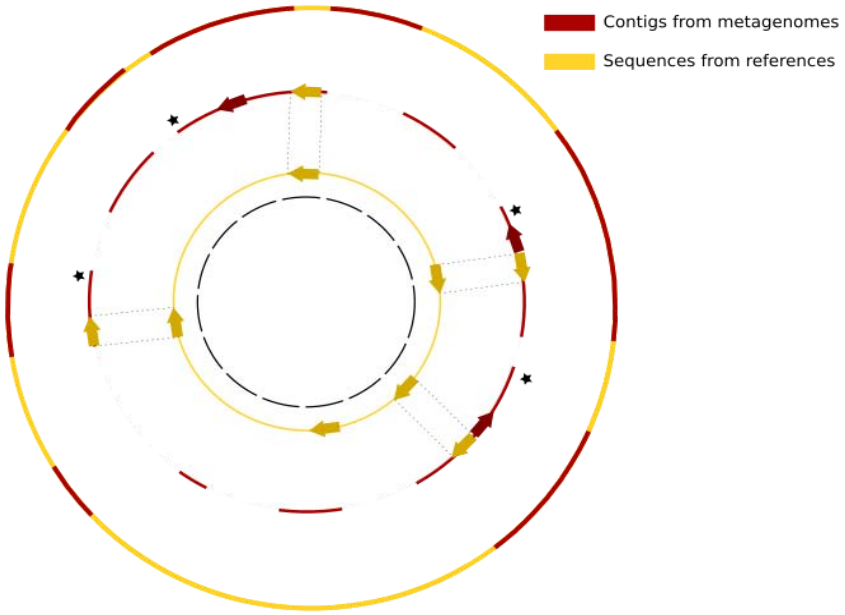


Figura 3. Estrategia seguida para construir el pangenoma ambiental extendido. El pangenoma de referencia (círculo amarillo interno) se busca en los *contigs* de metagenomas ambientales previamente ensamblados (secuencias rojas). Aquellos *contigs* que contienen la secuencia de algún gen de referencia y un gen contiguo

(estrellas) se consideraron como secuencias potencialmente nuevas para el género *Solirubrobacter*. Estos *contigs* se reclutaron y se fusionaron el pangenoma de referencia para obtener el pangenoma ambiental extendido (círculo rojo y amarillo externo).

7.2.7. Análisis del Pangenoma Ambiental Extendido de *Solirubrobacter*

Se construyó un perfil funcional con aquellas secuencias que se anotaron en el COG utilizando las mismas funciones que se emplearon para el perfil funcional del pangenoma de referencia. Además, se analizaron las diferencias entre las muestras mediante un Análisis de Coordenadas Principales (PcoA, por sus siglas en inglés *Principal Coordinate Analysis*) (Gower, 2015) y una matriz de disimilitud de Bray-Curtis. Los agrupamientos resultantes se evaluaron con la prueba estadística de ANOSIM. Por otro lado, se usó el archivo de salida de CD-HIT para construir una matriz de conteos y determinar familias génicas y, junto con los resultados de anotación funcional, determinar a qué funciones están asociadas las familias génicas presentes en *Solirubroabcter*.

Finalmente, se realizó en las secuencias del EEP una búsqueda manual de proteínas relacionadas con la interacción planta-microorganismo. Asimismo, se recurrió a Modelos Ocultos de Markov (HMM) de proteínas de este tipo de interacciones identificadas en Actinobacterias. Esta búsqueda se realizó con HMMER (v. 3.3.2) (Mistry et al., 2013) y se emplearon los genomas de *Streptomyces coelicolor* y *Streptomyces griseus* (Accession number: GCA_013317105.1 y ASM1060v1, respectivamente) como control positivo.

8. RESULTADOS

8.1. CAPÍTULO I: Microbioma de raíz y su posible rol en el desarrollo del pasto *Pennisetum ciliare*

El microbioma de raíz de una planta juega un papel importante en el establecimiento y desarrollo de ésta. Conocer qué microorganismos se adaptan a la rizósfera y cómo benefician a la planta nos permite desarrollar herramientas que promuevan o mermen el crecimiento de ésta. Bajo este contexto, conocer el microbioma rizosférico de plantas invasoras que desplazan a plantas nativas y alteran los ecosistemas es de suma importancia para entender los mecanismos de invasión y fomentar planes de acción para su control. En esta sección del trabajo se exploró el microbioma de raíz del pasto buffel, buscando determinar qué microorganismos se asocian a sus raíces, cómo cambia la estructura de la comunidad ante la exposición a compuestos aleloquímicos y las posibles interacciones entre los microorganismos y la planta.

A partir del análisis de las muestras logramos obtener el primer perfil taxonómico del microbioma rizosférico del pasto buffel conformado por un total de 2,164 ASVs. Este perfil resultó dominado por el phylum *Actinobacteria, Proteobacteria* y *Acidobacteria*, mientras que los géneros más abundantes fueron *RB41, Geodermatophilus, Gemmatirosa, Pseudonocardia, Krasilnikovia, Microvirga, Solirubrobacter, Gaiella* y *Angustibacter*. Identificamos un total de 30 géneros conformando el core del pasto buffel (presente en al menos 90% de las muestras), y cepas de varios de estos géneros se han reportado como capaces de sintetizar vitaminas, antibióticos y otros compuestos que podrían ayudar al establecimiento de la planta. Ciertos géneros, tales como *Planctomicrobium* y *Aurantimonas*, resultaron estadísticamente sobrerrepresentados dependiendo del tratamiento, demostrando que los compuestos aleloquímicos que produce el buffel influyen en la estructuración del microbioma. Aunado a esto, identificamos géneros capaces de metabolizar los mismos aleloquímicos que sintetiza el buffel (*Planctomicrobium, Aurantimonas, Saccharothrix* y *Tellurimicrobium*). Sin embargo, de acuerdo al ordenamiento construido, el factor más influyente resultó ser el tiempo en el que se tomaron las muestras (R: 0.2258; p = 0.0423, 9,999 permutaciones), es decir, el desarrollo de la planta. De acuerdo con esto, los resultados del análisis de DESeq2 identificaron géneros estadísticamente más abundantes en el T1 (*Nodosilinea_PCC-7104, Oceanibaculum, Pedobacter* y *Flavitalea*) y en el T2 (*Ohtaekwangia, IS-44, Phytohabitans* y *Saccharothrix*).

Estos resultados fueron publicados en la revista PLoS ONE (2023) y constituyen el artículo requisito. Para más detalles sobre el trabajo, se incluye a continuación la publicación de manera íntegra.

PLOS ONE

RESEARCH ARTICLE

Root microbiome diversity and structure of the Sonoran desert buffelgrass (*Pennisetum ciliare* L.)

Angélica Jara-Servin[1,2], Adán Silva[3], Hugo Barajas[1], Rocío Cruz-Ortega[4], Clara Tinoco-Ojanguren[3], Luis D. Alcaraz[1]*

1 Laboratorio de Genómica Ambiental, Departamento de Biología Celular, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico, **2** Posgrado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México, Mexico City, Mexico, **3** Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Universidad Nacional Autónoma de México, Hermosillo, Sonora, Mexico, **4** Departamento de Ecología Funcional, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico

* lalcaraz@ciencias.unam.mx

Citation: Jara-Servin A, Silva A, Barajas H, Cruz-Ortega R, Tinoco-Ojanguren C, Alcaraz LD (2023) Root microbiome diversity and structure of the Sonoran desert buffelgrass (*Pennisetum ciliare* L.). PLoS ONE 18(5): e0285978. https://doi.org/10.1371/journal.pone.0285978

Editor: Marcela Pagano, Universidade Federal de Minas Gerais, BRAZIL

Received: February 1, 2023

Accepted: April 28, 2023

Published: May 19, 2023

Data Availability Statement: All data used for this study can be accessed in the NCBI Sequence Read Archive (SRA) database under the Project ID PRJNA879420. All figures, tables, and datasets can be accessed in FigShare (https://doi.org/10.6084/m9.figshare.c.6605350.v2).

Funding: Consejo Nacional de Ciencia y Tecnología 765278 to Angélica Jara-Servin. Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México, DGAPA-PAPIIT-UNAM IN212914 to Clara Tinoco-Ojanguren.

Abstract

Buffelgrass (*Pennisetum ciliare*) is an invasive plant introduced into Mexico's Sonoran desert for cattle grazing and has converted large areas of native thorn scrub. One of the invasion mechanisms buffelgrass uses to invade is allelopathy, which consists of the production and secretion of allelochemicals that exert adverse effects on other plants' growth. The plant microbiome also plays a vital role in establishing invasive plants and host growth and development. However, little is known about the buffelgrass root-associated bacteria and the effects of allelochemicals on the microbiome. We used 16S rRNA gene amplicon sequencing to obtain the microbiome of buffelgrass and compare it between samples treated with root exacknudates and aqueous leachates as allelochemical exposure and samples without allelopathic exposure in two different periods. The Shannon diversity values were between $H' = 5.1811–5.5709$, with 2,164 reported bacterial Amplicon Sequence Variants (ASVs). A total of 24 phyla were found in the buffelgrass microbiome, predominantly *Actinobacteria*, *Proteobacteria*, and *Acidobacteria*. At the genus level, 30 different genera comprised the buffelgrass core microbiome. Our results show that buffelgrass recruits microorganisms capable of thriving under allelochemical conditions and may be able to metabolize them (*e.g.*, *Planctomicrobium*, *Aurantimonas*, and *Tellurimicrobium*). We also found that the community composition of the microbiome changes depending on the developmental state of buffelgrass ($p = 0.0366$; ANOSIM). These findings provide new insights into the role of the microbiome in the establishment of invasive plant species and offer potential targets for developing strategies to control buffelgrass invasion.

1. Introduction

Introducing exotic species to a new environment can alter the ecosystem and decrease biological diversity. Invasive species can transform ecosystem processes over spatial and temporal

Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México, DGAPA-PAPIIT-UNAM IN221420 to Luis D. Alcaraz. The funders had no role in the study design, data collection, analysis, decision to publish, or manuscript preparation.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: LDA is an Academic Editor for this journal. There are no other competing interests. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

scales, with various degrees of impact [1]. In many cases, invasive species have higher values of resource-acquisition traits, larger sizes, and higher growth rates, implying larger pool sizes of nutrients that allow them to outcompete native species [1]. Buffelgrass (*Pennisetum ciliare*, *Cenchrus ciliaris*) has been North America's most noxious invasive plant since its introduction from East Africa in the 1930s [2]. The germination, establishment, and subsequent seedling growth of buffelgrass depend on soil moisture and environmental temperature conditions [3]. Buffelgrass is preferred because of its high germinability, easy establishment, high seedling production and vigor, fast growth rate, and ability to withstand fires because of its massive root system [3–6]. Buffelgrass individual tussocks and their long lifespan allow them to re-sprout from established tufts following a fire [5]. Fire also temporarily increases available phosphorus levels in the soil, which are rapidly exploited by buffelgrass [7].

In northwestern Mexico, large tracts of desert and thorn scrub have been converted to buffelgrass pasture to improve rangelands for cattle production [8]. Buffelgrass coverage estimates are 53% of the Sonora state and 12% of Mexico [9]. In the Sonoran Desert, buffelgrass invasion can cause a decrease by 53%–73% of the number of native plant species compared to undisturbed thorn scrub, which has a higher species diversity and is commonly dominated by Fabaceae and Cactaceae [8]. In contrast, thorny legume trees such as *Acacia cochliacantha* and *Acacia farnesiana* dominate the succession of abandoned pastures in the thorn scrub of Mexico [10], and very few native species can regenerate in the ranges [8]. The approaches taken to control buffelgrass growth include applying herbicides, manual removal, prescribed burning, and controlled animal grazing, but controlling buffelgrass worldwide is still an urgent matter [5].

One of the mechanisms proposed as a strategy for invasive species control is allelopathy [8, 11], which is based on the active synthesis and release of allelochemicals (secondary metabolites) by one plant, influencing the growth of a recipient plant, regardless of resource availability [12, 13]. The production of biochemicals that natives in the invaded range do not produce plays in favor of the invasive species since these chemicals may affect native species that lack a coevolutionary-based tolerance [14, 15]. Allelopathy plays a vital role in invasion, vegetation patterning, the exclusion of associated species, and reduced plant productivity [16]. Previous reports have shown buffelgrass allelopathy through root exudates of 3-month-old individuals, decreasing the growth and germination of various species such as *Chrysopogon aucheri*, *Hyparrhenia rufa*, *Bothriochloa pertusa*, *Panicum antidotale*, *Setaria italica*, and *Pennisetum americanum* [4]. Buffelgrass leachates and root exudates obtained from 3-month-old buffelgrass plants and used to water *Brassica campestris*, *Lactuca sativa*, and *Setaria italica* provoked a reduction in radicle growth due to allelopathy [17]. The germination rates of these three species also decreased, but only when the leachates and exudates were concentrated through evaporation [18]. The decrease in wild species could result from the amount of forage buffelgrass produced, which is 4 to 10 times greater than the production of native species, leading to possible nutrient depletion, especially of N, from the soils it grows on [3]. Such characteristics may explain, to a certain degree, the observed self-declination of pastures of *Pennisetum ciliare* [3, 18].

Various approaches have been made to describe the chemical identity of buffelgrass allelochemicals affecting the growth of many plant species. So far, the identified compounds are phenolic compounds, including *p*-OH-benzoic acid, *p*-coumaric acid, caffeic acid, vanillic acid, ferulic acid, syringic acid, and gentisic acid [17]. Although phenolic compounds have a short half-life in soils [19, 20], their reversibly sorbed fractions contribute to the pool available for allelopathic interactions [19]. Hence, compounds such as ferulic acid, *p*-coumaric acid, *p*-OH-benzoic acid, and vanillic acid could accumulate in the soil and impede the germination and growth of susceptible species [17]. However, the fate and persistence of phytotoxins in soil are unclear, as are their phytotoxicity or biodisposition by microbes [17]. The mutualistic

association formed by the invader is presumed to contribute to a competitive advantage over native species [21, 22], and plants can impact their associated microbiomes as an adaptation strategy when confronted with biotic and abiotic challenges [15, 23, 24]. Moreover, allelopathic plants release compounds that might alter the composition of the microbial community, recruiting microorganisms involved in the metabolization of the allelochemicals produced [25, 26]. Hence, understanding the rhizosphere microbiomes of invasive plants and the elements that influence the recruitment of microorganisms could be particularly valuable for understanding the factors promoting plant invasiveness and the following impacts on the ecosystems. So far, there are no reports of the buffelgrass rhizosphere microbiome composition and the effects of its allelochemicals on the microbial communities associated with its roots.

This work investigated the buffelgrass root microbiome and its compositional changes under allelochemical exposure. Microbiome changes were tested at two-time points using plants exposed to buffelgrass root exudates, watered with leachates from the aerial part of buffelgrass, and with a distilled water regime. This experimental design allowed us to describe the buffelgrass microbiome throughout buffelgrass growth, obtain a core microbiome, and determine the effect that allelochemicals may have on taxa comprising the rhizosphere microbiome.

2. Methods

2.1 Seed and soil sampling

Pennisetum ciliare seeds were collected from five plants at the end of the summer (September 2013) from induced pastures at Rancho Diamante (28˚ 41' N, 110˚ 15' W), Sonora State, Mexico, where thorn scrub is the original vegetation. Seeds were stored in a paper bag and transported to the laboratory. We also collected soil samples from the same location for further plant growth in the greenhouse. We collected samples of Buffelgrass *Cenchrus ciliaris* (*Pennisetum ciliare*) from a location where we did not require special permissions, beyond the land owner. The land owner permitted us to collect the samples. It is important to note that while Buffelgrass is native to Tropical Africa and Asia, it is considered an invasive species in North America. However, it is not classified as an endangered or protected species and, therefore, does not require special permission for collection.

2.2 Plant growth under allelopathic conditions

We removed the seed covers of *P. ciliare* seeds for germination as naked caryopses have higher and more uniform germination [27]. Caryopses were germinated in Petri dishes with 2% agar in a growth chamber Biotronette (20.3L, 25˚C, 12-h photoperiod, 92 $\mu molm^{-2}s^{-1}$). Subsequently, 2-day-old seedlings were transferred to containers with peat moss, which grew for 70 days. The water regime consisted of watering with Hoagland solution every two weeks. After that, 70-day-old plants with a height of at least 10 cm were transplanted into curved PVC tubes containing sterilized silica sand and soil (at a proportion of 1:1) from Rancho Diamante. Plants in PVC tubes received a standard water regime (consisting of 25 mL of sterilized deionized water six times per week). They were divided into three different treatments: (1) root exudates, (2) aqueous leachates of the aerial plant part, and (3) control (Fig 1, https://doi.org/10.6084/m9.figshare.c.6605350.v2). The exudate treatment consisted of a second buffelgrass plant at the opposite end of the curved PVC tube, so the roots of the 70-day-old individual would be under the influence of the second buffelgrass plant, allowing to evaluate the effect of the root exudates on the rhizosphere microbiome. Physical interaction of the roots was avoided by installing a net at the middle of the tube. In the leachate treatment, one end of the PVC tube was planted with the 70-day-old buffelgrass, while the other was left unplanted. Plants under this treatment were watered with aqueous leachates extracted from buffelgrass green leaves

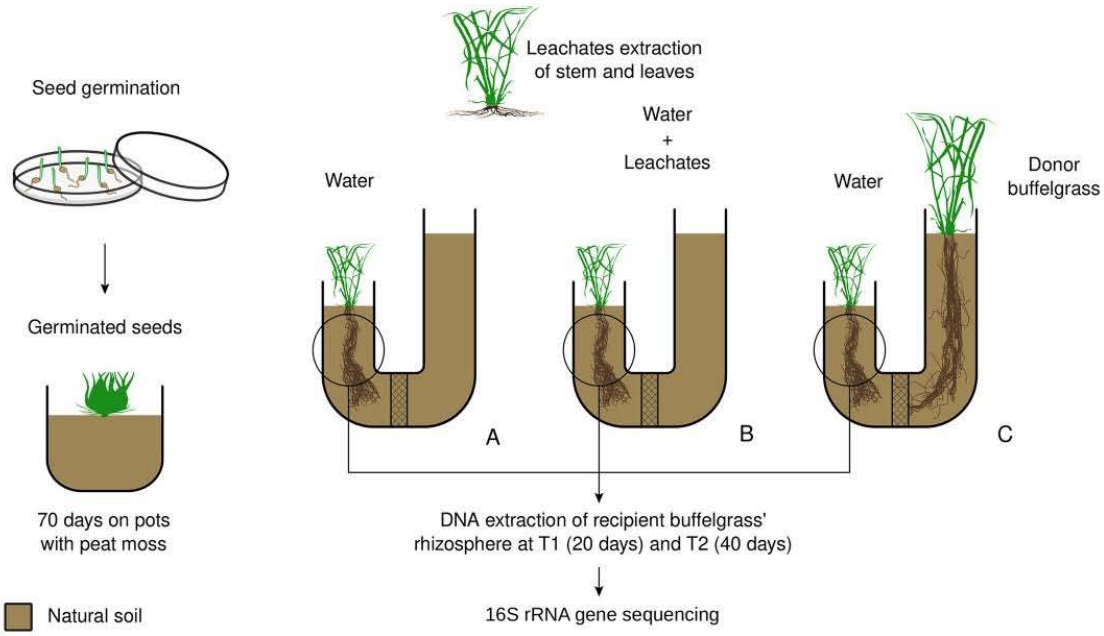


Fig 1. Experimental setup. Buffelgrass seeds germinated in a chamber and, after two days, transferred to peat moss-filled pots. After 70 days, the plants were transplanted into PVC pots containing silica and soil (1:1) from Rancho Diamante, Sonoran Desert. The plants were then divided into three groups, each with four replicates. Group (A) was watered with sterilized distilled water, Group (B) was watered with a solution of aqueous leachates from leaves and stems, and Group (C) were watered with distilled water but exposed to root exudates from another buffelgrass plant. Sampling was done at two different periods, 20 and 40 days, with two replicates from each group being analyzed at each time.

https://doi.org/10.1371/journal.pone.0285978.g001

and stems instead of water, but according to the volume and periodicity of the standard water regime. The control treatment consisted of only one end of the PVC planted with buffelgrass with the standard water regime. No Hoagland solution was added during the allelochemical treatments. The leachates were obtained as previously described [28] by collecting green leaves and stems of buffelgrass adult plants in Rancho El Diamante. Leaves were dried for 24 h at 24˚C and submerged in distilled water 1% (g/v) for 3 h, followed by filtering through Whatman #4 filtering paper to simulate rainfall. Two soil samples from Rancho Diamante were left unplanted and subjected to no treatment. These soil samples allowed us to determine the soil microbiome of the region without the influence of growing buffelgrass plants and the effect that buffelgrass has on the recruited microorganisms.

2.3 Rhizosphere sample collection

Rhizosphere samples were obtained at two different time points of growth from the PVC tubes: after 20 days (T1) and after 40 days (T2). In both cases, the roots were separated from the remaining parts of the plant to obtain the rhizosphere. Rhizosphere separation involved removing the loose soil and subsequently washing and submitting the roots to ultrasound in 1X PBS buffer (137 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4). All pellets obtained from buffelgrass rhizospheres were recovered through centrifugation (50-mL tubes, 1,300 g, 10 min) and kept at -80˚C until further analysis and DNA extraction.

2.4 Metagenomic DNA processing and 16S rRNA gene sequencing

The metagenomic DNA of soil samples and all the rhizosphere pellets was extracted using the Mobio PowerSoil DNA extraction kit (MoBio, Carlsbad, CA, United States) following the manufacturer's instructions, with minor modifications, heating the C6 elution solution to 60˚C before the elution step to increase the DNA yield. Two 30 μL elution processes were performed during the same spin filter.

The 16S rRNA gene amplification was performed in duplicates, following the Illumina ® MiSeq protocol for 16S metagenomic sequencing library preparation (Illumina 2013). The primer pair used for the PCR amplification was 341F/805R (targeting the V3–V4 regions), with the Illumina sequencing adaptors in 5′ overhangs (341F: 5′ –CCTACGGGNGGCWGCAG–3′; 805R: 5′ – ACTACHVGGGTATCTAATCC 3′). The PCR reaction mixture consisted of 2 μL buffer, 1.2 μL of each primer (5 μM), 3 μL enhancer, 1.6 μL dNTPs (2.5 mM), 0.6 μL Mg2S04 (1.5μM), 9.2 μL PCR-grade water, 0.16 μL Pfx polymerase (0.02 U/μL) (Invitrogen, Thermo Fisher Scientific, Waltham, MA), and 2 μL DNA template, adding up to a final volume of 20 μL.

The PCR conditions were as follows: initial 95˚C for 3 min, then 25 cycles at 94˚C for 5 s, followed by 68˚C for 30 s. The PCR products were pooled and purified with the SV Wizard PCR purification kit (Promega, Madison, WI). All samples were sequenced on the Illumina ® MiSeq platform (2 × 300 paired-end) at the University Unit of Massive Sequencing and Bioinformatics Facilities of the Biotechnology Institute, UNAM Mexico.

2.5 16S rRNA gene amplicon sequence analysis

The detailed protocol and bioinformatic methods used to process and analyze the 16S rRNA gene amplicon sequences are available on GitHub (https://github.com/genomica-fciencias-unam/buffelgrass). Briefly, all amplicon libraries were quality-checked using Dada2 [29], the first 17 bp were removed, and the sequences were trimmed to 250 bp. Only the forward reads were used since the quality profiles of the reverse reads were poor. All amplicon sequences were processed to obtain Amplicon Sequence Variants (ASV) using DADA2 (v. 1.10.1) [29] to denoise and remove chimeras. At the species level, taxonomy was assigned using the Silva database (v. 138) [30]. A phylogenetic tree was constructed using FastTreeMP [31] (S1 Dataset, https://doi.org/10.6084/m9.figshare.c.6605350.v2).

2.6 Diversity and statistical analysis

The α- and β-diversity of all samples were calculated using the phyloseq [32], ggplot2 [33], vegan [34], and R default packages [35]. We measured taxonomic α-diversity using Observed, Shannon, and Simpson diversity indices. Hierarchical clustering was performed using the hclust method on an unweighted UniFrac distance matrix [36]. The ASVs were clustered at the various taxonomic levels to perform abundance comparisons of the three treatments. Core microbiomes were analyzed using the upset function; ß-diversity was analyzed through a constrained analysis of principal coordinates (CAP) on an unweighted UniFrac distance matrix, based on the obtained ASV abundances per treatment; the clustering was evaluated through the ANOSIM statistical function [37]. Differential ASV abundances comparing treatments were calculated using DESeq2 of the R package [38]. Detailed statistical and bioinformatic methods are available on FigShare (https://doi.org/10.6084/m9.figshare.c.6605350.v2).

3. Results

We sequenced 686,078 paired-end reads, with a mean of 49,005.57 ± 10,041 sequences per sample (n = 8). We then clustered them into 2,164 Amplicon Sequence Variants (ASVs; 100%

Table 1. Alpha diversity for buffelgrass microbiomes.

Sample	Observed	Shannon	Simpson
C1_T1	308	5.350832	0.9934471
C2_T1	349	5.466695	0.9942502
C1_T2	270	5.268196	0.9935561
C2_T2	262	5.225775	0.9928567
E1_T1	291	5.298459	0.9932077
E2_T1	342	5.468742	0.9943344
E1_T2	384	5.570910	0.9947818
E2_T2	332	5.464093	0.9945279
L1_T1	126	4.433554	0.9849328
L2_T1	538	5.871980	0.9957321
L1_T2	262	5.181120	0.9923309
L2_T2	282	5.291607	0.9935689
S1	297	5.361435	0.9938688
S2	295	5.331116	0.9935597

Values calculated for Observed, Shannon, and Simpson diversity indices. C is for controls; E exudates; L leachate; S soil; T is for time.

https://doi.org/10.1371/journal.pone.0285978.t001

identity 16S rRNA gene OTUs) and described diversity by Observed ASVs, Simpson, and Shannon diversity indices (Table 1). The Shannon index (*H'*) was used to evaluate the **α**-diversity of our samples. Overall, the samples maintained a similar diversity (*H'* = 5.1811–5.5709), regardless of treatment or period, except for the outliers, corresponding to samples of the leachate treatment at the first period (*H'* = 4.4335 and *H'* = 5.8719). The diversity increased slightly with the development of buffelgrass treated with exudates (from *H'* = 5.2984 and 5.4687 in the first period to *H'* = 5.5709 and 5.4640 in the second one). On the contrary, control samples showed a diversity reduction as the experiment advanced (*H'* = 5.3508 and 5.4666 in the first period to *H'* = 5.2681 and 5.2257 in the second one). Species dominance, evaluated through the Simpson index (*D*), followed the same pattern as the one described for diversity, where the value for the exudate treatment was higher at the end of the experiment (*D* = 0.9947 and 0.9945) than at the beginning (*D* = 0.9932 and 0.9943), as opposed to the increase seen from the first to the second period in samples from the control treatment (*D* = 0.9934 and 0.9942 to *D* = 0.9935 and 0.9928). Soil samples had Shannon diversity (*H'* = 5.361425 and 5.331116) and Simpson (*D* = 0.9938688 and 0.9935597) indices not differing from the treatment values.

We detected 24 phyla in all our samples. Eleven phyla were ubiquitous and abundant: *Actinobacteria*, *Proteobacteria*, *Acidobacteria*, *Planctomycetes*, *Gemmatimonadetes*, *Chloroflexi*, *Verrucomicrobia*, *Bacteroidetes*, *Firmicutes*, *Cyanobacteria*, and *Armatimonadetes* (Fig 2, https://doi.org/10.6084/m9.figshare.c.6605350.v2). The heatmap also revealed differences in the microbiome composition among different periods. For example, *WPS-2* and *Hydrogenedentes*, whose abundances increased in exudates from T1 (relative abundance values of *WPS-2* = 0.00; *Hydrogenedentes* = 2.05E-05) to T2 (*WPS-2* = 7.29E-04; *Hydrogenedentes* = 2.64E-04). On the contrary, the abundance of *Deinococcus-Thermus* decreased from T1 (1.16E-04) to T2 (0.00) in the same treatment. The abundance of the phylum *Fibrobacteres* increased over time both in the exudate and the leachate treatments. Complete taxonomic annotation and abundances are available as Supplementary Material (S1 and S2 Tables, https://doi.org/10.6084/m9.figshare.c.6605350.v2).

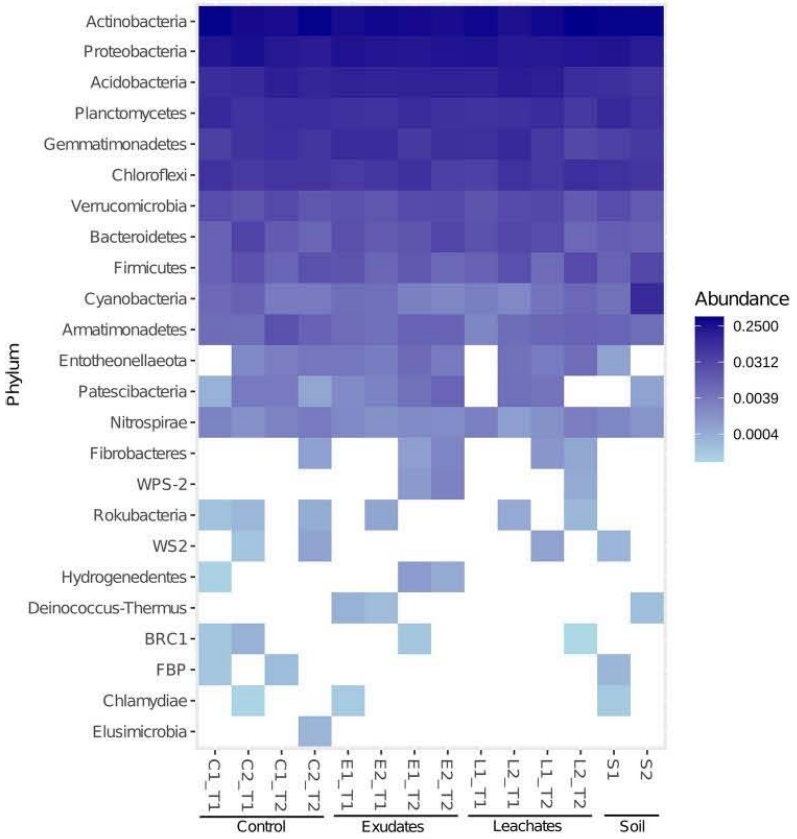


Fig 2. Buffelgrass microbiome at the phylum level. Each sample is labeled indicating control (C), exudates (E), leachates (L), soil (S), and sampling time (T1 and T2).

https://doi.org/10.1371/journal.pone.0285978.g002

The UniFrac dendrogram shows most communities clustered together according to time (Fig 3A, https://doi.org/10.6084/m9.figshare.c.6605350.v2). T2 showed larger dispersion in the dendrogram, suggesting community diversification over time. We further evaluated ß-diversity through a constrained analysis of principal coordinates (CAP) based on an unweighted UniFrac distance matrix (Fig 3B). The CAP analysis explained 20.9% of the observed variance. The x-axis described most of the ordination variance (10.8%) and separated the natural soil samples from the buffelgrass samples on the bottom right side of the ordination (Fig 3B). Control samples (C) are closer to the natural soils than the samples from allelochemical treatments (E and L). The CAP ordination clusters were evaluated with ANOSIM to test the treatment differences; no significant differences were found (R: 0.06579; p = 0.2902; 9,999 permutations). All T1 samples clustered at the upper half, as opposed to T2, located in the lower half of the ordination (Fig 3B). The ANOSIM test showed that time-based clustering was statistically significant (R: 0.2258; p = 0.0423, 9,999 permutations), meaning that the samples differed depending on time.

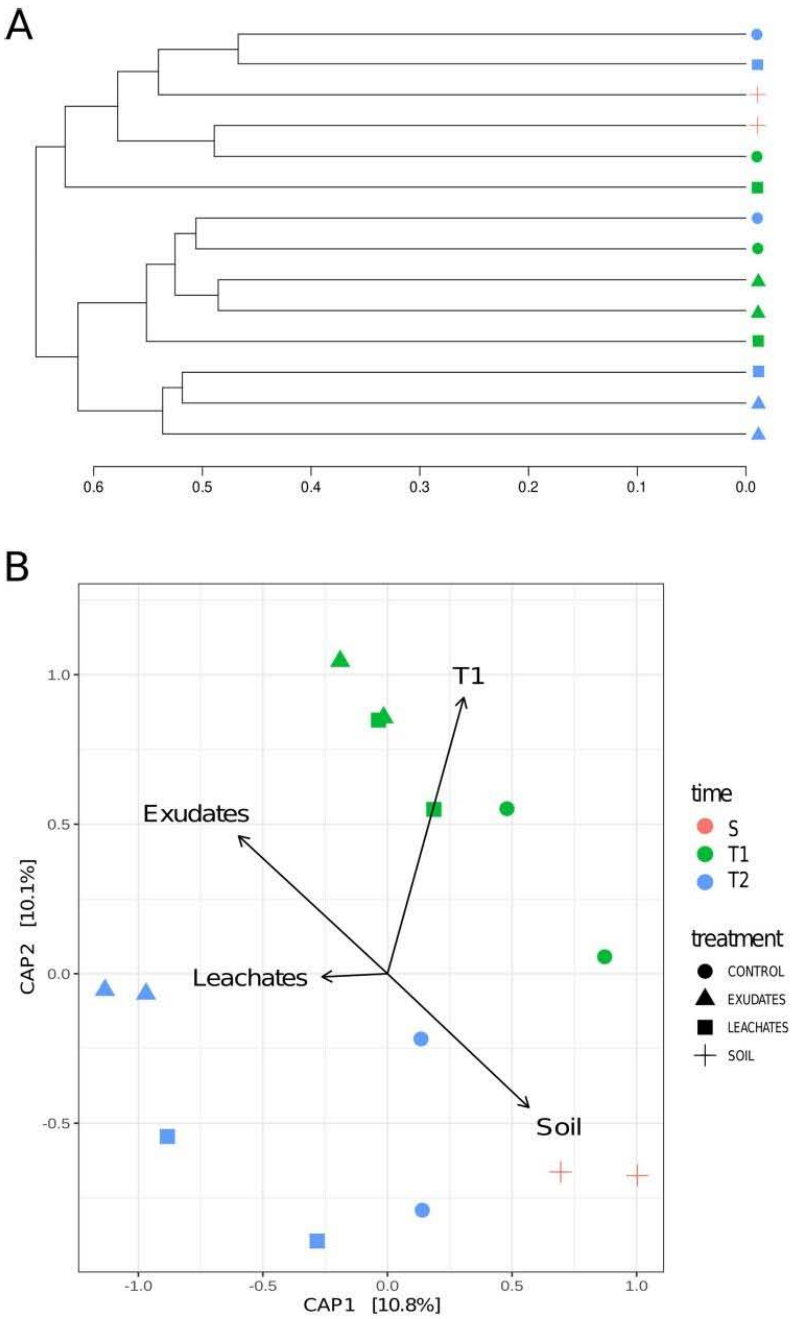


Fig 3. Beta diversity of the buffelgrass microbiome. (A) Dendrogram of microbiome relatedness based on an unweighted UniFrac distance matrix (A). (B) Beta diversity of the buffelgrass microbiome is represented as a constrained analysis of principal coordinates (CAP) based on an unweighted UniFrac distance matrix for all

treatments and natural soils (B). Vectors display the experiment's different treatments and time points. Statistical significance was evaluated using the ANOSIM test for treatments ($p = 0.092$) and time ($p = 0.0366$).

https://doi.org/10.1371/journal.pone.0285978.g003

We used a DESeq2 analysis to compare and identify abundant differential genera at each time point, irrespective of the treatment used (Fig 4A, S3 Table, https://doi.org/10.6084/m9.figshare.c.6605350.v2). In T1, we identified the following genera: *Nodosilinea_PCC-7104*, *Oceanibaculum*, *Pedobacter*, and *Flavitalea*, whereas *Ohtaekwangia*, *IS-44*, *Phytohabitans*, and *Saccharothrix* were enriched in T2. Subsequently, we compared the microbiome diversity at the genus level and found 14 genera representing 16 ASVs shared in all samples, including natural soils (S2 Fig, https://doi.org/10.6084/m9.figshare.c.6605350.v2). However, these genera represented only 0.78% of the whole ASVs dataset. Most ASVs were shared between all samples, with only 13 ASVs being unique to soil samples.

Actinobacteria were highly dominant across treatments and time. We found 30 genera in the treatment and control samples, comprising the buffelgrass core microbiome (Fig 5; S1 Fig, https://doi.org/10.6084/m9.figshare.c.6605350.v2). The core genera were *Geodermatophilus*, *RB41*, *Krasilnikovia*, *Angustibacter*, *Microvirga*, *Kribbella*, *Bradyrhizobium*, *Rubrobacter*, *Modestobacter*, *Pseudonocardia*, *Gemmatirosa*, *Bryobacter*, *Nitrospira*, *Nitrolancea*, *Solirubrobacter*, *Gaiella*, Candidatus_*Udaeobacter*, *Dactylosporangium*, *Altererythrobacter*, *Micromonospora*, *Roseomonas*, *Gemmata*, *Skermanella*, *Chthoniobacter*, *Gemmatimonas*, *Sphingomonas*, *JCM_18997*, Candidatus_*Alysiosphaera*, *Amycolatopsis*, and *Crossiella* (Fig 5). Most of them belong to the phylum *Actinobacteria*, except for *RB41* (*Acidobacteria*), *Gemmatirosa* and *Gemmatimonas* (*Gemmatimonadetes*), Candidatus_*Udaeobacter* and *Chthoniobacter* (*Verrucomicrobia*), *Gemmata* (*Planctomycetes*), and *Sphingomonas*, Candidatus_*Alysiosphaera*, *Roseomonas*, *Skermanella*, *Altererythrobacter*, and *Microvirga* (*Proteobacteria*). A complete list of the core microbiome species and a summary of where they had been isolated is available in S4 Table (https://doi.org/10.6084/m9.figshare.c.6605350.v2).

We identified enriched genera in the leachates compared to the controls, namely *Mle1-7*, *Caulobacter*, and *Phytohabitans* (Fig 4D, S3 Table). Compared to the controls, the genera enriched in the exudates were *Planctomicrobium*, *Aurantimonas*, *Cryptosporangium*, *Bauldia*, *Tellurimicrobium*, *Oxalicibacterium*, *Dokdonella*, *Nodosilinea-PCC-7104*, *Saccharothrix*, and *Caulobacter* (Fig 4B, S3 Table). The following genera were higher in the exudates than the leachates: *Saccharothrix*, *Jatrophihabitans*, *Tellurimicrobium*, *Bauldia*, *Aurantimonas*, *Oxalicibacterium*, *Planctomicrobium*, *Reyranella*, and *Singulisphaera* (Fig 4C, S3 Table). The genera enriched in the leachates were *Phytohabitans*, *Isoptericola*, *Pedobacter*, *Kouleothrix*, *Oceanibaculum*, *Yonghaparkia*, and *Pseudarthrobacter* (Fig 4C, S3 Table).

4. Discussion

4.1 The buffelgrass microbiome, clues for success as an invasive species?

Exploring the microbiomes of invasive species may shed light on the establishment and propagation of those species and their impacts on ecosystems, hence providing a turning point for developing new plant control strategies. Even though a high number of ASVs are shared between all treatment and soil samples, the microbiome composition of the rhizosphere of buffelgrass is distinguishable from that of the soil microbiome (Fig 3B, S1 Fig). Our results suggest that the overall bacterial communities of *Pennisetum ciliare* are structured by different factors. Overall, the Shannon values obtained for the samples in this study were within the range of Shannon values (H' = 5.1811–5.5709, Table 1) reported for microbiomes in desert soils [39]. Regarding the exudate treatment, previous reports have shown that the rhizosphere microbial

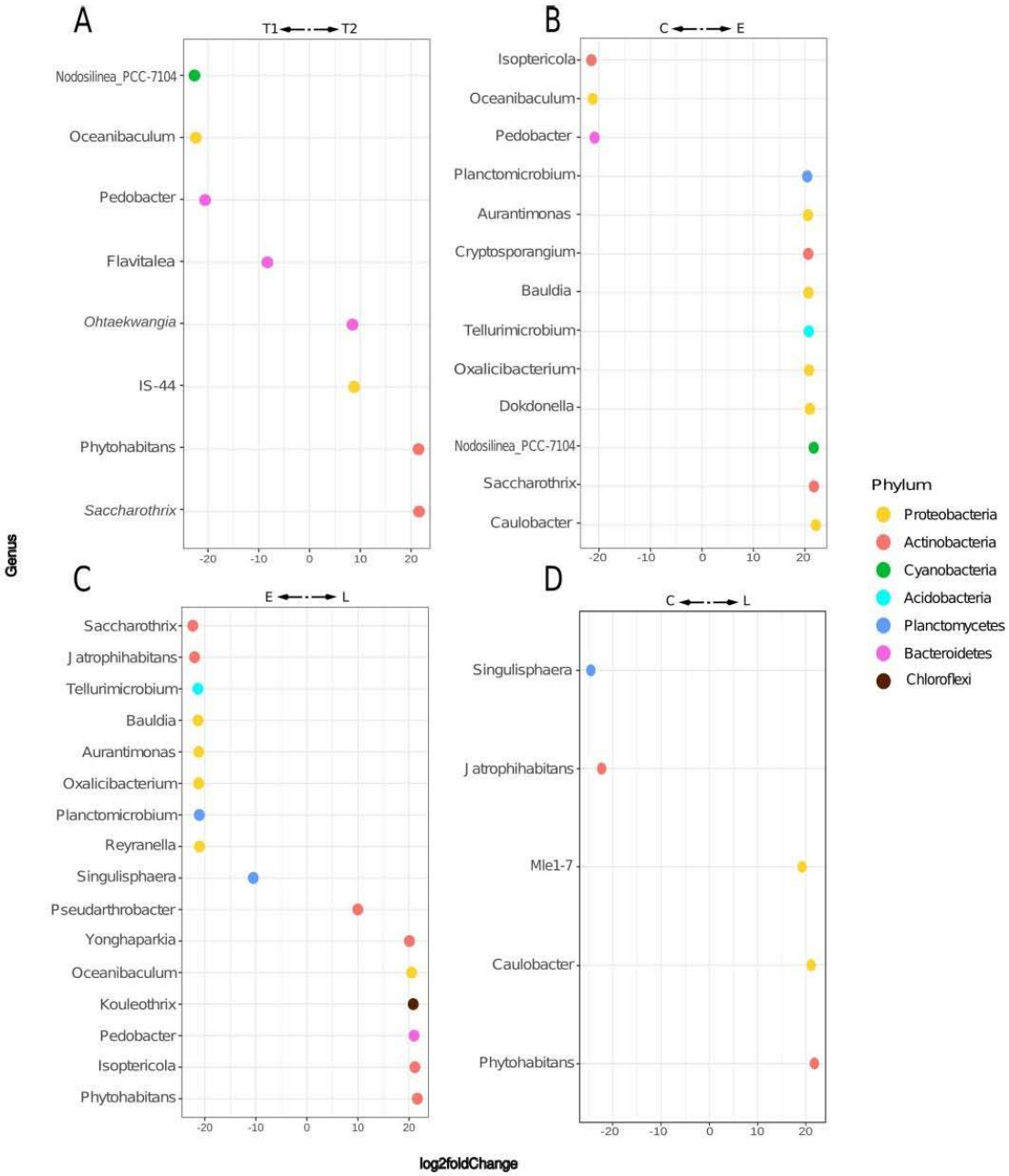


Fig 4. Enriched bacterial genera in allelochemical treatments. Shown are (A) the log2fold ratio between T1 and T2 (A), (B) exudates and leachates (B), (C) exudates and controls (C), and (D) leachates and controls (D). DESeq2 was used to get the significantly enriched genera in each condition using an $\alpha = 0.01$.

https://doi.org/10.1371/journal.pone.0285978.g004

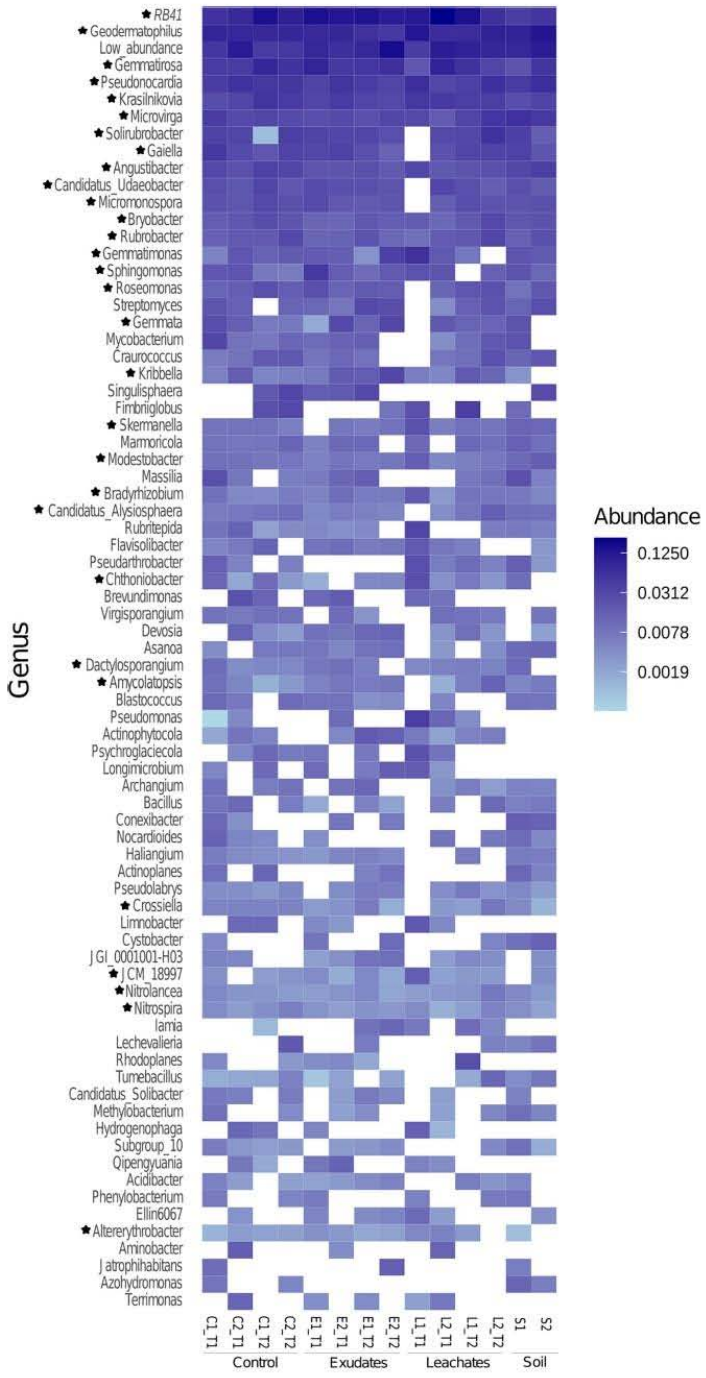

Genus
RB41
Geodermatophilus
Low_abundance
Gemmatirosa
Pseudonocardia
Krasilnikovia
Microvirga
Solirubrobacter
Gaiella
Angustibacter
Candidatus_Udaeobacter
Micromonospora
Bryobacter
Rubrobacter
Gemmatimonas
Sphingomonas
Roseomonas
Streptomyces
Gemmata
Mycobacterium
Craurococcus
Kribbella
Singulisphaera
Fimbriiglobus
Skermanella
Marmoricola
Modestobacter
Massilia
Bradyrhizobium
Candidatus_Alysiosphaera
Rubritepida
Flavisolibacter
Pseudarthrobacter
Chthoniobacter
Brevundimonas
Virgisporangium
Devosia
Asanoa
Dactylosporangium
Amycolatopsis
Blastococcus
Pseudomonas
Actinophytocola
Psychroglaciecola
Longimicrobium
Archangium
Bacillus
Conexibacter
Nocardioides
Haliangium
Actinoplanes
Pseudolabrys
Crossiella
Limnobacter
Cystobacter
JGI_0001001-H03
JCM_18997
Nitrolancea
Nitrospira
Iamia
Lechevalieria
Rhodoplanes
Tumebacillus
Candidatus_Solibacter
Methylobacterium
Hydrogenophaga
Subgroup_10
Qipengyuania
Acidibacter
Phenylobacterium
Ellin6067
Altererythrobacter
Aminobacter
Jatrophihabitans
Azohydromonas
Terrimonas
C1_T1
C2_T1
C1_T2
C2_T2
E1_T1
E2_T1
E1_T2
E2_T2
L1_T1
L2_T1
L1_T2
L2_T2
S1
S2
Control
Exudates
Leachates
Soil
Abundance
0.1250
0.0312
0.0078
0.0019

Fig 5. Buffelgrass core microbiome at the genus level. In the heatmap, genera marked with stars comprise the buffelgrass core microbiome. We collapsed genera with low abundances (< 0.03%).

https://doi.org/10.1371/journal.pone.0285978.g005

diversity tends to be higher when treated with root exudates, specifically when exposed to phenolic compounds [40]. These diversity observations align with the Shannon values obtained for exudates (H' = 5.51) and control treatments in T2 (H' = 5.24). The diversity for buffelgrass exposed to root exudates at the end of the experiment (H' = 5.51) was higher than the value from the bulk soil (H' = 5.34), a pattern already reported for ruderal plants, whose diversity in the rhizosphere tends to be higher than in bulk soil [41–43]. Our data showed that the rhizosphere microbiome of buffelgrass is dominated by the phyla *Actinobacteria*, *Proteobacteria*, *Acidobacteria*, *Planctomycetes*, *Gemmatimonadetes*, and *Chloroflexi* (Fig 2), in agreement with a previous report [2]. Likewise, the surrounding soil of the allelopathic desert shrub *Artemisia sieberi* shows similar phyla diversity [11]. The buffelgrass microbiome composition is also dominated by *Actinobacteria*, as reported in multiple arid areas [44, 45]. The ability of actinobacterial spores to germinate in environments with low water availability enables their adaptation to drought conditions [46] and shrub root zones of deserts [47].

Regarding the genus diversity, *RB41*, *Geodermatophilus*, *Gemmatirosa*, *Pseudonocardia*, *Krasilnikovia*, *Microvirga*, *Solirubrobacter*, *Gaiella*, *Angustibacter*, *Candidatus_Udaeobacter*, *Micromonospora*, *Bryobacter*, *Gemmatimonas*, *Sphingomonas*, and *Roseomonas* were highly abundant in buffelgrass, irrespective of the treatment or sampling time (Fig 5). Overall, 2,164 ASVs corresponded to 235 different genera of bacteria detected in our samples. Those 235 genera constitute the extended microbiome of buffelgrass roots under all tested conditions (S1 Fig). Those bacteria comprise what is referred to as the buffelgrass core microbiome. A core microbiome is a set of microorganisms forming cores of interactions that can be used to optimize microbial functions at the individual plant and ecosystem levels [23]. Diversity manipulation might be a key battleground where hosts and various hubs cooperate or compete, making them potential targets for plant biocontrol studies [48]. Rather than just promoting host plant growth, the core microbiome is vital in organizing the community assemblies mediating and organizing plant-microbe and microbe-microbe interactions by recruiting indigenous microorganisms with diverse functions and even suppressing high pathogen loads in the field [23]. Hence, the detailed study of core microbiomes should enable microbial species identification and functions on plant-microbe interactions that impact plant adaptation to arid environments and their featured plant-associated lifestyles [49, 50].

Among the core microbiome of buffelgrass (Fig 5, S2 Fig), some genera have been related to allelopathic conditions, such as *RB41*, *Bryobacter*, *Nitrospira*, *Gaiella*, and *Microvirga*, whose abundances changed depending on the amount of vanillic acid available [20, 51, 52]. Additionally, the genus *Nitrospira* has been reported in the soil surrounding buffelgrass and can oxidize ammonia to nitrate (commamox) [2]; commamox activity is frequent in oligotrophic habitats [53]. Hence, the presence of this genus in the core microbiome could help buffelgrass obtain nitrogen in deserts, where this element tends to be scarce. *Bradyrhizobium* is a common species in close association with plant roots, including known invasive species such as *Acacia dealbata*, and can enhance the competition ability of the tree [21, 54]. The potential for producing soluble vitamins, antimicrobials, and antibiotics detected in rhizosphere metagenomes and genomes of *Bradyrhizobium*, *Geodermatophilus*, *Pseudonocardia*, *Micromonospora*, *Crosiella*, *Amycolatopsis*, and *Kribbela*, may also be relevant in the context of plant invasions [55–60]. The genus *Sphingomonas* produces molecules that promote plant growth [61], whereas *Kribbella* isolates from allelopathic shrubs and other plant species show antifungal activity and contain genes related to the production of secondary metabolites [62, 63].

Gemmatimonas belong to the phylum *Gemmatimonadetes*, including taxa adapted to arid and oligotrophic conditions, and are among the most abundant bacteria in soils [64, 65]. The ability to thrive in replanted soils and to accumulate polyphosphates may be beneficial for the development of buffelgrass, which creates an allelopathic autotoxic environment and depends on phosphorus compounds to develop [5, 17, 66–68].

Solirubrobacter (*Actinobacteria*) is ubiquitous in buffelgrass. Only a few cultivated bacteria represent *Solirubrobacter*, with indirect evidence that showed them as a diverse group in nature [69]. *Solirubrobacter* can thrive in phenolic environments, using chlorogenic acid to grow [69], and can develop in the rhizosphere of cucumber treated with vanillin [20]. The relationship of *Solirubrobacter* with the plant host phenotype is ambiguous, with positive and negative correlations in multiple hosts [67, 70, 71]. Considering that several studies propose that certain fungi and bacteria contribute to the invasion success of plants introduced to novel habitats [25, 54], it is fair to suggest that the microorganisms recruited in the rhizosphere of buffelgrass contribute to a certain extent to the establishment and quick development of this grass in the habitats it invades.

4.2 Buffelgrass recruits taxa capable of thriving in an allelopathic environment

The allelochemicals exogenously amended during the experiment favored differential taxa abundance, according to our DESeq2 analysis (Fig 4). This effect of allelochemicals on microbiome structuring was observed in the CAP analysis, where the samples from the exudate and leachate treatments from T2 are further apart than they were in the first period (Fig 3B). Moreover, microbiomes from allelochemical treatments clustered apart from both soil and the controls. Even though the microbiomes under leachate treatment cluster closer to the controls, the distance is noticeable (Fig 3B). This strongly suggests that the presence of buffelgrass influences the root microbiome diversity.

Plants in desert ecosystems have developed ecophysiological adaptations, including salt and chemical compound excretion beneath the plant canopy, creating an allelopathic environment that affects organic matter and soil moisture levels [11]. Previously reported buffelgrass roots exudates comprised chlorogenic, ferulic, caffeic, *p*-OH-benzoic, *p*-coumaric, vanillic, syringic, and gentisic acids, all phenolic acids, and proven allelopathic agents [17]. Phenolic compounds recruit bacterial taxa more precisely than other rhizodeposits [72], and some compounds, such as coumaric acid, affect growth and microbial community composition in a concentration-dependent manner [19, 73]. Moreover, this concentration may increase through soil microbial production and using vanillic acid and *p*-OH-benzoic acid from ferulic acid and *p*-coumaric acid, respectively [19, 74, 75]. Since the buffelgrass microbiome changes across time and development, the root-recruited taxa are likely able to metabolize the phenolic compounds secreted by the roots at different growth stages [25].

We found 17 abundant differential bacterial genera in the treatments. As expected, some ASVs enriched in the exudates or leachate treatment are known for their capabilities to metabolize and produce phenolic compounds (S3 Table). Enriched taxa in allelochemical treatments were: *Caulobacter*, *Phytohabitans*, *Mle1-7*, *Planctomicrobium*, *Aurantimonas*, *Cryptosporangium*, *Bauldia*, *Tellurimicrobium*, *Oxalicibacterium*, *Dokdonella*, *Nodosilinea-PCC-7104*, and *Saccharothrix* (Fig 4B and 4D, S3 Table). *Caulobacter* (*Alphaproteobacteria*) was the only one enriched in both allelochemical treatments. *Caulobacter* acts as a microbial community hub on the leaf microbiome of *Arabidopsis* [48]; also as a plant growth-promoting bacterium (PGPB) through the production of plant hormones [76, 77]. Different species of *Caulobacter* can thrive in rhizosphere environments containing the same phenolic acids as those produced

by buffelgrass [18, 78–80]. *Cryptosporangium* and *Bauldia* are found in the roots of different allelopathic plants, such as *Eucalyptus* and *Andrographis* [81, 82].

Buffelgrass is a plant species preferred for grazing because of its high nutrient content. However, it can accumulate oxalate to potentially toxic concentrations for cattle [5, 17, 83]. *Oxalicibacterium* (*Burkholderiaceae)* comprises species isolated from soils and litter from oxalate-forming plants and collects oxalic acid or calcium oxalate as metabolic products, although oxalate can also be produced through the degradation of other compounds such as citrate [84–88]. Bacterial oxidation of calcium oxalate, which can be carried out by species such as *Oxalicibacterium* and *Bradyrhizobium*, entails soil alkalinization, which may act as a soil acidification buffer [88]. Similarly, *Tellurimicrobium* can grow on oxaloacetate [89]. The relevance of the ability to metabolize oxalic compounds is rooted in the importance of oxalate regarding plant physiology. Oxalate solubilizes insoluble phosphorus (in the form of aluminum phosphate) in the rhizosphere, is involved in pH regulation and calcium storage, and acts as a ligand for aluminum [85, 90]. Furthermore, oxalate degradation by some taxa, such as *Burkholderia*, lowers oxalate levels on plant surfaces and decreases the infection potential of pathogenic microorganisms attracted by oxalate [90].

Allelochemicals are compounds synthesized not only by plants but also by fungi and bacteria. The actinobacterium *Saccharothrix* was also enriched in the root exudate treatment compared to the controls. This genus can biosynthesize caffeic acid and cinnamic acid, the latter being a substrate for the biosynthesis of other phenolic compounds [91, 92]. The biosynthesis of phenolic compounds, such as vanillic acid from ferulic acid, through mechanisms involved with ß-oxidation, has been described in many bacteria [75]. The accumulation of allelochemicals, such as phenolic compounds, may have both plant and bacterial origins. Buffelgrass is a plant with a C4 metabolism, which enables it to thrive under drought conditions and high temperatures [6, 93]. Under those conditions, plants draw upon different strategies, such as regulating stomatal aperture [94]. Albeit in a concentration-dependent manner, phenolic acids influence the aperture and closure of stomata in various plant species [95–97] and may even alter photosynthetic performance [98]. Hence allelochemicals could act as signals in a cross-kingdom talk, and it is possible that the presence of phenolic compounds, either supplied by the plant or by the microorganisms, in the buffelgrass rhizosphere influences buffelgrass physiology through the regulation of stomatal aperture and closure, which in turn could diminish the growth of the plant while promoting its development in an arid environment. Further analysis using a metagenomic perspective may confirm the interpretations made in this study.

4.3 Time as a microbiome community-structuring factor

Rhizosphere microbiomes are composed of bacteria filtered by the same plant from the bulk soil, creating microbial communities in the rhizosphere, distinguishable from the soil communities [99–101]. The soil-to-rhizosphere effect was shown in the ordination (10.8% of the variance, Fig 3B), where the samples from bulk soil (S) cluster apart from all the treatment samples derived from buffelgrass plants. Interestingly, the Unifrac dendrogram shows a closer phylogenetic relationship between communities belonging to the same period (Fig 3A). Similarly, CAP analysis revealed that time influences the microbiome structure, shown through differences in the microbiome between T1 and T2 (Fig 3B). Studies have reported plant developmental effects on plant-associated microbiomes [102–104]. Even though the period evaluated is relatively short (20 days difference), specific phyla changed in their abundances concerning T1 and T2 (Fig 2). The *WPS-2* and *Hydrogenedentes* showed increased quantities from T1 to T2 in the root exudates treatment. In contrast, *Fibrobacteres* rose in the root exudate and leachate treatment. *Hydrogenedentes* have been reported in contaminated zones with

compounds such as phenanthrene [105] and associated with multiple plants [40,106]. Phylum *WPS-2* has been registered as dominant in the invasive tree *Eucalyptus*, with a declining abundance after the first year [107]. *Deinococcus-Thermus* abundance was reduced from T2 to T1 in the root exudate treatment; in a previous study, the absence of phenolic allelochemicals correlated with a decreased *Deinococcus-Thermus* abundance [40].

Additionally, plant age is the main factor shaping root microbiomes in maize and *Eucalyptus*, stabilizing over time [107, 108]. Regarding buffelgrass, the microbiome changes probably reflect plant aging. The plant's different development stages, including the allelochemicals exuded by the roots, cause changes in the compounds secreted by the roots [78]. Time differences were highlighted by differential taxa in buffelgrass (Fig 4A, S3 Table). During the first period (T1), the significantly enriched bacteria belonged to the phyla *Cyanobacteria* (*Nodosilinea PCC-7104*), *Proteobacteria* (*Oceanibaculum*), and *Bacteroidetes* (*Pedobacter* and *Flavitalea*). *Pedobacter* and *Oceanibaculum* have increased abundances during the intermediate plant growth stages [109]. Bacteria in the rhizosphere of younger plants tend to use simple amino acids instead of complex carbohydrates, as do older plants [110]. This young plant hypothesis is backed by previous studies, where the amendment of soils with a mixture of long-chain fatty acids and amino acids led to an enrichment of *Flavitalea* [111].

We observed that *Bacteroidetes* (*Ohtaekwangia*), *Proteobacteria* (*IS-44*), and *Actinobacteria* (*Phytohabitans* and *Saccharothrix*) were enriched in T2. This is consistent with previous findings that *Ohtaekwangia* dominates in the middle-growth stages of other plant microbiomes and decreases over time [104, 109]. Our results also agree with this pattern, as T2 buffelgrass plants were not at a young growth stage, and *Ohtaekwangia* was enriched in 112-day-old buffelgrass plants.

The changes in root exudates of a plant throughout its development can significantly impact the microbial communities in the surrounding soil, as highlighted by recent studies [78, 112]. This highlights the importance of considering the plant's phenology and physiological and chemical traits when examining the effects of invasive plant species on ecosystems [1, 73]. Additionally, the impact of phenolic compounds on the soil microbiome is known to be concentration-dependent [19, 73], and changes in root exudate patterns during the development of buffelgrass could provide new insights into the formation of rhizosphere microbiomes. Our experimental design, which included the use of PVC pots, may have also played a role in the accumulation of phenolics and leachates, leading to the selection of microorganisms that are capable of thriving in environments with high concentrations of phenolic compounds or can degrade these compounds as part of their metabolism. These microorganisms are probably selected based on the buffelgrass's growth stage and allelochemicals' fate and concentration. This interaction could allow the plant to benefit from the production of antimicrobial compounds and the degradation of contaminants, allowing the plant to expand its invasive range.

5. Conclusions

Actinobacteria dominated the microbiome of buffelgrass, resembling the microbiomes of other plants in deserts, arid zones, and other desert allelopathic shrubs. Our study found that rhizosphere microbiomes treated with allelochemicals had a higher abundance of specific taxa, highlighting the impact of buffelgrass on soil bacteria. Time was also found to be a significant factor in shaping the buffelgrass microbiome, and further research should focus on the effects of allelochemicals on the root exudate pattern throughout the plant's development. Additionally, we identified a core microbiome comprising microorganisms known for their antimicrobial or vitamin production capabilities, which likely played a role in shaping the remaining

microbial community associated with the roots of buffelgrass. Our findings suggest that certain recruited bacteria can metabolize allelochemical compounds excreted by the roots, potentially influencing buffelgrass physiology.

Supporting information

S1 Fig. Buffelgrass microbiome composition heatmap at the genus level.
(EPS)

S2 Fig. Shared genera between samples visualized in an UpSet plot. Shared taxa at the genus level between allelochemical and control treatments. The histogram shows the number of shared elements for each intersection set, ordered in a decreasing manner. Genera marked with stars comprise the buffelgrass core microbiome, considering taxa present in all treatment and control samples but allowing absence in one of the samples.
(TIFF)

S1 Table. Taxonomy table indicating the taxonomic assignment of each ASV.
(XLSX)

S2 Table. Counts table indicating the number of appearances of each ASV per sample.
(XLSX)

S3 Table. Genera significantly enriched in paired comparisons. The table shows significantly differentially abundant bacteria calculated by DESeq2 analysis. Information about the microbe genera, sources of isolation, and associated publications are also provided.
(XLSX)

S4 Table. Genera comprising the buffelgrass core microbiome. Bibliographical information about microbial sources of isolation and associated publications.
(XLSX)

S1 Dataset. Phylogenetic tree. Newick file for the phylogenetic tree constructed with FastTreeMP using 16S rRNA gene sequences obtained from buffelgrass microbiome samples.
(NWK)

Acknowledgments

We thank field work assistance by José F. Martínez.

Author Contributions

Conceptualization: Adán Silva, Rocío Cruz-Ortega, Clara Tinoco-Ojanguren.

Formal analysis: Angélica Jara-Servin, Luis D. Alcaraz.

Funding acquisition: Luis D. Alcaraz.

Methodology: Adán Silva, Hugo Barajas, Rocío Cruz-Ortega.

Resources: Rocío Cruz-Ortega, Clara Tinoco-Ojanguren.

Validation: Luis D. Alcaraz.

Writing – review & editing: Angélica Jara-Servin, Rocío Cruz-Ortega, Clara Tinoco-Ojanguren, Luis D. Alcaraz.

References

1. Ehrenfeld JG. Ecosystem Consequences of Biological Invasions. Annu Rev Ecol Evol Syst. 2010; 41: 59–80. https://doi.org/10.1146/annurev-ecolsys-102209-144650
2. Gornish ES, Franklin K, Rowe J, Barberán A. Buffelgrass invasion and glyphosate effects on desert soil microbiome communities. Biol Invasions. 2020; 22: 2587–2597. https://doi.org/10.1007/s10530-020-02268-8
3. Hanselka CW, Hussey MA FI. Buffelgrass. Warm-Season (C4) Grasses. 2015. pp. 477–502. https://doi.org/10.2134/agronmonogr45.c13
4. Hussain F, Naqvi HH, Ilahi I. Interference Exhibited by *Cenchrus ciliaris* L. and *Bothriochloa pertusa* (L.) A. Camus. Bull Torrey Bot Club. 1982; 109: 513. https://doi.org/10.2307/2996492
5. Marshall VM, Lewis MM, Ostendorf B. Buffel grass (*Cenchrus ciliaris*) as an invader and threat to biodiversity in arid environments: A review. J Arid Environ. 2012; 78: 1–12. https://doi.org/10.1016/j.jaridenv.2011.11.005
6. Lyons KG, Maldonado-Leal BG, Owen G. Community and Ecosystem Effects of Buffelgrass (*Pennisetum ciliare*) and Nitrogen Deposition in the Sonoran Desert. Invasive Plant Sci Manag. 2013; 6: 65–78. https://doi.org/10.1614/IPSM-D-11-00071.1
7. Miller G, Friedel M, Adam P, Chewings V. Ecological impacts of buffel grass (*Cenchrus ciliaris* L.) invasion in central Australia—does field evidence support a fire-invasion feedback? Rangel J. 2010; 32: 353. https://doi.org/10.1071/RJ09076
8. Tinoco-Ojanguren C, Díaz A, Martínez J, Molina-Freaner F. Species diversity and regeneration of native species in *Pennisetum ciliare* (buffelgrass) pastures from the thornscrub of Sonora, México. J Arid Environ. 2013; 97: 26–37. https://doi.org/10.1016/j.jaridenv.2013.05.009
9. Arriaga L, Castellanos V. AE, Moreno E, Alarcón J. Potential Ecological Distribution of Alien Invasive Species and Risk Assessment: a Case Study of Buffel Grass in Arid Regions of Mexico. Conserv Biol. 2004; 18: 1504–1514. https://doi.org/10.1111/j.1523-1739.2004.00166.x
10. Burgos A, Maass JM. Vegetation change associated with land-use in tropical dry forest areas of Western Mexico. Agric Ecosyst Environ. 2004; 104: 475–481. https://doi.org/10.1016/j.agee.2004.01.038
11. Saul-Tcherkas V, Unc A, Steinberger Y. Soil Microbial Diversity in the Vicinity of Desert Shrubs. Microb Ecol. 2013; 65: 689–699. https://doi.org/10.1007/s00248-012-0141-8 PMID: 23192699
12. Ridenour WM, Callaway RM. The relative importance of allelopathy in interference: the effects of an invasive weed on a native bunchgrass. Oecologia. 2001; 126: 444–450. https://doi.org/10.1007/s004420000533 PMID: 28547460
13. Schandry N, Becker C. Allelopathic Plants: Models for Studying Plant–Interkingdom Interactions. Trends Plant Sci. 2020; 25: 176–185. https://doi.org/10.1016/j.tplants.2019.11.004 PMID: 31837955
14. Kato-Noguchi H. Allelopathy and Allelochemicals of *Leucaena leucocephala* as an Invasive Plant Species. Plants. 2022; 11: 1672. https://doi.org/10.3390/plants11131672 PMID: 35807624
15. Thorpe AS, Thelen GC, Diaconu A, Callaway RM. Root exudate is allelopathic in invaded community but not in native community: field evidence for the novel weapons hypothesis. J Ecol. 2009; 97: 641–645. https://doi.org/10.1111/j.1365-2745.2009.01520.x
16. Rice E. L. Allelopathy. 2nd ed. Academic Press, Orlando, Florida, USA. 1984.
17. Hussain F, Ilahi I, Malik SA, Dasti AA, Ahmad B. Allelopathic effects of rain leachates and root exudates of *Cenchrus ciliaris* L. And *Bothriochloa pertusa* (L.) A. Camus. Pakistan J Bot. 2011; 43: 341–350.
18. Hussain MI, Reigosa MJ. Allelochemical stress inhibits growth, leaf water relations, PSII photochemistry, non-photochemical fluorescence quenching, and heat energy dissipation in three C3 perennial species. J Exp Bot. 2011; 62: 4533–4545. https://doi.org/10.1093/jxb/err161 PMID: 21659663
19. Blum U. Effects of microbial utilization of phenolic acids and their phenolic acid breakdown products on allelopathic interactions. J Chem Ecol. 1998; 24: 685–708. https://doi.org/10.1023/A:1022394203540
20. Song J, Qiu S, Zhao J, Han C, Wang Y, Sun X, et al. *Pseudonocardia tritici* sp. nov., a novel actinomycete isolated from rhizosphere soil of wheat (*Triticum aestivum* L.). Kleunen M, editor. Antonie Van Leeuwenhoek. 2019; 112: 765–773. https://doi.org/10.1007/s10482-018-01210-2 PMID: 30536164
21. Klock MM, Barrett LG, Thrall PH, Harms KE. Host promiscuity in symbiont associations can influence exotic legume establishment and colonization of novel ranges. Kleunen M, editor. Divers Distrib. 2015; 21: 1193–1203. https://doi.org/10.1111/ddi.12363
22. Fonseca E da S, Peixoto RS, Rosado AS, Balieiro F de C, Tiedje JM, Rachid CTC da C. The Microbiome of Eucalyptus Roots under Different Management Conditions and Its Potential for Biological

Nitrogen Fixation. Microb Ecol. 2018; 75: 183–191. https://doi.org/10.1007/s00248-017-1014-y PMID: 28634640

23. Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, et al. Core microbiomes for sustainable agroecosystems. Nat Plants. 2018; 4: 247–257. https://doi.org/10.1038/s41477-018-0139-4 PMID: 29725101

24. Ravanbakhsh M, Kowalchuk GA, Jousset A. Root-associated microorganisms reprogram plant life history along the growth–stress resistance tradeoff. ISME J. 2019; 13: 3093–3101. https://doi.org/10.1038/s41396-019-0501-1 PMID: 31511619

25. Trognitz F, Hackl E, Widhalm S, Sessitsch A. The role of plant–microbiome interactions in weed establishment and control. Muyzer G, editor. FEMS Microbiol Ecol. 2016;92: fiw138. https://doi.org/10.1093/femsec/fiw138 PMID: 27387910

26. Kamutando CN, Vikram S, Kamgan-Nkuekam G, Makhalanyane TP, Greve M, Roux JJ Le, et al. Soil nutritional status and biogeography influence rhizosphere microbial communities associated with the invasive tree *Acacia dealbata*. Sci Rep. 2017; 7: 6472. https://doi.org/10.1038/s41598-017-07018-w PMID: 28747705

27. Tinoco-Ojanguren C, Reyes-Ortega I, Sánchez-Coronado ME, Molina-Freaner F, Orozco-Segovia A. Germination of an invasive *Cenchrus ciliaris* L. (buffel grass) population of the Sonoran Desert under various environmental conditions. South African J Bot. 2016; 104: 112–117. https://doi.org/10.1016/j.sajb.2015.10.009

28. Cruz-Ortega R, Ayala-Cordero G, Anaya AL. Allelochemical stress produced by the aqueous leachate of *Callicarpa acuminata*: effects on roots of bean, maize, and tomato. Physiol Plant. 2002; 116: 20–27. https://doi.org/10.1034/j.1399-3054.2002.1160103.x PMID: 12207658

29. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016; 13: 581–583. https://doi.org/10.1038/nmeth.3869 PMID: 27214047

30. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 2014; 42: D643–D648. https://doi.org/10.1093/nar/gkt1209 PMID: 24293649

31. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol Biol Evol. 2009; 26: 1641–1650. https://doi.org/10.1093/molbev/msp077 PMID: 19377059

32. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Watson M, editor. PLoS One. 2013; 8: e61217. https://doi.org/10.1371/journal.pone.0061217 PMID: 23630581

33. Wickham H, Chang W. Package 'ggplot2'. R Foundation for Statistical Computing. 2016; Vienna, Austria. Available from https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf.

34. Oksanen J. Vegan: ecological diversity. In: R Package Version 2.4–4 [Internet]. 2017 p. 11. Available: https://cran.r-project.org/package=vegan=vegan

35. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2013 Vienna, Austria: R Foundation for Statistical Computing. Available: http://www.r-project.org/

36. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. Appl Environ Microbiol. 2005; 71: 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005 PMID: 16332807

37. Clarke KR. Non-parametric multivariate analyses of changes in community structure. Austral Ecol. 1993; 18: 117–143. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x

38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

39. Vásquez-Dean J, Maza F, Morel I, Pulgar R, González M. Microbial communities from arid environments on a global scale. A systematic review. Biol Res. 2020; 53: 29. https://doi.org/10.1186/s40659-020-00296-1 PMID: 32631429

40. Fang C, Zhuang Y, Xu T, Li Y, Li Y, Lin W. Changes in Rice Allelopathy and Rhizosphere Microflora by Inhibiting Rice Phenylalanine Ammonia-lyase Gene Expression. J Chem Ecol. 2013; 39: 204–212. https://doi.org/10.1007/s10886-013-0249-4 PMID: 23385369

41. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017; 551: 457–463. https://doi.org/10.1038/nature24621 PMID: 29088705

42. Marasco R, Mosqueira MJ, Fusi M, Ramond J-B, Merlino G, Booth JM, et al. Rhizosheath microbial community assembly of sympatric desert speargrasses is independent of the plant host. Microbiome. 2018; 6: 215. https://doi.org/10.1186/s40168-018-0597-y PMID: 30514367

43. Barajas HR, Martínez-Sánchez S, Romero MF, Álvarez CH, Servín-González L, Peimbert M, et al. Testing the Two-Step Model of Plant Root Microbiome Acquisition Under Multiple Plant Species and Soil Sources. Front Microbiol. 2020; 11: 1–17. https://doi.org/10.3389/fmicb.2020.542742 PMID: 33162946

44. Crits-Christoph A, Robinson CK, Barnum T, Fricke WF, Davila AF, Jedynak B, et al. Colonization patterns of soil microbial communities in the Atacama Desert. Microbiome. 2013; 1: 28. https://doi.org/10.1186/2049-2618-1-28 PMID: 24451153

45. Naylor D, Coleman-Derr D. Drought Stress and Root-Associated Bacterial Communities. Front Plant Sci. 2018; 8: 1–16. https://doi.org/10.3389/fpls.2017.02223 PMID: 29375600

46. Mohammadipanah F, Wink J. Actinobacteria from Arid and Desert Habitats: Diversity and Biological Activity. Front Microbiol. 2016; 6: 1–10. https://doi.org/10.3389/fmicb.2015.01541 PMID: 26858692

47. Steven B, Gallegos-Graves LV, Starkenburg SR, Chain PS, Kuske CR. Targeted and shotgun metagenomic approaches provide different descriptions of dryland soil microbial communities in a manipulated field study. Environ Microbiol Rep. 2012; 4: 248–256. https://doi.org/10.1111/j.1758-2229.2012.00328.x PMID: 23757280

48. Agler MT, Ruhe J, Kroll S, Morhenn C, Kim S-T, Weigel D, et al. Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. Waldor MK, editor. PLOS Biol. 2016; 14: e1002352. https://doi.org/10.1371/journal.pbio.1002352 PMID: 26788878

49. Busby PE, Soman C, Wagner MR, Friesen ML, Kremer J, Bennett A, et al. Research priorities for harnessing plant microbiomes in sustainable agriculture. PLOS Biol. 2017; 15: e2001793. https://doi.org/10.1371/journal.pbio.2001793 PMID: 28350798

50. Alcaraz LD, Peimbert M, Barajas HR, Dorantes-Acosta AE, Bowman JL, Arteaga-Vázquez MA. *Marchantia liverworts* as a proxy to plants' basal microbiomes. Sci Rep. 2018; 8: 12712. https://doi.org/10.1038/s41598-018-31168-0 PMID: 30140076

51. Zhou XG, Jia HT, Pan DD, Ge X, Qiao PL, Wu FZ. Effects of vanillin on cucumber rhizosphere bacterial community. Allelopath J. 2018; 44: 191–200. https://doi.org/10.26651/allelo.j./2018-44-2-1164

52. Zhou X, Wu F. Vanillic acid changed cucumber (*Cucumis sativus* L.) seedling rhizosphere total bacterial, *Pseudomonas* and *Bacillus* spp. communities. Sci Rep. 2018; 8: 4929. https://doi.org/10.1038/s41598-018-23406-2 PMID: 29563548

53. Podlesnaya G, Krasnopeev A, Potapov S, Tikhonova I, Shtykova Y, Suslova M, et al. Diversity of nitrifying bacteria in microbial communities from water and epilithic biofilms of the Lake Baikal littoral zone. Limnol Freshw Biol. 2020; 2020: 1008–1010. https://doi.org/10.31951/2658-3518-2020-A-4-1008

54. Kamutando CN, Vikram S, Kamgan-Nkuekam G, Makhalanyane TP, Greve M, Le Roux JJ, et al. The Functional Potential of the Rhizospheric Microbiome of an Invasive Tree Species, *Acacia dealbata*. Microb Ecol. 2019; 77: 191–200. https://doi.org/10.1007/s00248-018-1214-0 PMID: 29948018

55. Wardle DA, Nicholson KS, Ahmed M, Rahman A. Interference effects of the invasive plant Carduus nutans L. against the nitrogen fixation ability of *Trifolium repens* L. Plant Soil. 1994; 163: 287–297. https://doi.org/10.1007/BF00007978

56. Gao R, Liu C, Zhao J, Jia F, Yu C, Yang L, et al. *Micromonospora jinlongensis* sp nov, isolated from muddy soil in China and emended description of the genus *Micromonospora*. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol. 2014; 105: 307–315. https://doi.org/10.1007/s10482-013-0074-3 PMID: 24257967

57. Boumehira AZ, El-Enshasy HA, Hacène H, Elsayed EA, Aziz R, Park EY. Recent progress on the development of antibiotics from the genus *Micromonospora*. Biotechnology and Bioprocess Engineering. Korean Society for Biotechnology and Bioengineering; 2016. pp. 199–223. https://doi.org/10.1007/s12257-015-0574-2

58. Tanvir R, Sajid I, Hasnain S, Kulik A, Grond S. Rare actinomycetes *Nocardia caishijiensis* and *Pseudonocardia carboxydivorans* as endophytes, their bioactivity and metabolites evaluation. Microbiol Res. 2016; 185: 22–35. https://doi.org/10.1016/j.micres.2016.01.003 PMID: 26946375

59. Gonzalez-Pimentel JL, Dominguez-Moñino I, Jurado V, Laiz L, Caldeira AT, Saiz-Jimenez C. The Rare *Actinobacterium Crossiella* sp. Is a Potential Source of New Bioactive Compounds with Activity against Bacteria and Fungi. Microorganisms. 2022; 10: 1575. https://doi.org/10.3390/microorganisms10081575 PMID: 36013993

60. Song Z, Xu T, Wang J, Hou Y, Liu C, Liu S, et al. Secondary Metabolites of the Genus *Amycolatopsis*: Structures, Bioactivities and Biosynthesis. Molecules. 2021; 26: 1884. https://doi.org/10.3390/molecules26071884 PMID: 33810439

61. Khan AL, Waqas M, Kang SM, Al-Harrasi A, Hussain J, Al-Rawahi A, et al. Bacterial endophyte *Sphingomonas* sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. J Microbiol. 2014; 52: 689–695. https://doi.org/10.1007/s12275-014-4002-7 PMID: 24994010

62. Zhou Z, Gu J, Li Y-Q, Wang Y. Genome plasticity and systems evolution in *Streptomyces*. BMC Bioinformatics. 2012; 13: S8. https://doi.org/10.1186/1471-2105-13-S10-S8 PMID: 22759432

63. Gong X, Shi J, Zhou X, Yuan T, Gao D, Wu F. Crop Rotation With Cress Increases Cucumber Yields by Regulating the Composition of the Rhizosphere Soil Microbial Community. Front Microbiol. 2021; 12. https://doi.org/10.3389/fmicb.2021.631882 PMID: 33776961

64. DeBruyn JM, Nixon LT, Fawaz MN, Johnson AM, Radosevich M. Global Biogeography and Quantitative Seasonal Dynamics of *Gemmatimonadetes* in Soil. Appl Environ Microbiol. 2011; 77: 6295–6300. https://doi.org/10.1128/AEM.05005-11 PMID: 21764958

65. Zeng Y, Nupur Wu N, Madsen AM, Chen X, Gardiner AT, et al. *Gemmatimonas groenlandica* sp. nov. Is an Aerobic Anoxygenic Phototroph in the Phylum *Gemmatimonadetes*. Front Microbiol. 2021; 11: 1–18. https://doi.org/10.3389/fmicb.2020.606612 PMID: 33519753

66. Zhang Y, Cong J, Lu H, Yang C, Yang Y, Zhou J, et al. An Integrated Study to Analyze Soil Microbial Community Structure and Metabolic Potential in Two Forest Types. Ibekwe AM, editor. PLoS One. 2014; 9: e93773. https://doi.org/10.1371/journal.pone.0093773 PMID: 24743581

67. Franke-Whittle IH, Manici LM, Insam H, Stres B. Rhizosphere bacteria and fungi associated with plant growth in soils of three replanted apple orchards. Plant Soil. 2015; 395: 317–333. https://doi.org/10.1007/s11104-015-2562-x

68. Liu M-J, Jin C-Z, Ersiman A, Park D-J, Kim C-J. *Flavitalea flava* sp. nov., a bacterium isolated from a soil sample, and emended description of the genus *Flavitalea*. Antonie Van Leeuwenhoek. 2019; 112: 275–281. https://doi.org/10.1007/s10482-018-1156-z PMID: 30171433

69. Whitman WB. *Solirubrobacter*. Bergey's Man Syst Archaea Bact. 2015. https://doi.org/10.1002/9781118960608.gbm00227

70. Yang J, Ruegger PM, McKenry M V., Becker JO, Borneman J. Correlations between Root-Associated Microorganisms and Peach Replant Disease Symptoms in a California Soil. Yang C-H, editor. PLoS One. 2012; 7: e46420. https://doi.org/10.1371/journal.pone.0046420 PMID: 23071565

71. Hernández-Álvarez C, García-Oliva F, Cruz-Ortega R, Romero MF, Barajas HR, Piñero D, et al. Squash root microbiome transplants and metagenomic inspection for in situ arid adaptations. Sci Total Environ. 2022; 805: 150136. https://doi.org/10.1016/j.scitotenv.2021.150136 PMID: 34818799

72. Badri D V., Chaparro JM, Zhang R, Shen Q, Vivanco JM. Application of Natural Blends of Phytochemicals Derived from the Root Exudates of *Arabidopsis* to the Soil Reveal That Phenolic-related Compounds Predominantly Modulate the Soil Microbiome. J Biol Chem. 2013; 288: 4502–4512. https://doi.org/10.1074/jbc.M112.433300 PMID: 23293028

73. Zhou X, Wu F. p-Coumaric Acid Influenced Cucumber Rhizosphere Soil Microbial Communities and the Growth of *Fusarium oxysporum* f.sp. *cucumerinum* Owen. Gilbert JA, editor. PLoS One. 2012; 7: e48288. https://doi.org/10.1371/journal.pone.0048288 PMID: 23118972

74. Smith-Becker J, Marois E, Huguet EJ, Midland SL, Sims JJ, Keen NT. Accumulation of Salicylic Acid and 4-Hydroxybenzoic Acid in Phloem Fluids of Cucumber during Systemic Acquired Resistance Is Preceded by a Transient Increase in Phenylalanine Ammonia-Lyase Activity in Petioles and Stems1. Plant Physiol. 1998; 116: 231–238. https://doi.org/10.1104/pp.116.1.231 PMID: 9449843

75. Świzdor A, Panek A, Milecka-Tronina N, Kołek T. Biotransformations Utilizing β-Oxidation Cycle Reactions in the Synthesis of Natural Compounds and Medicines. Int J Mol Sci. 2012; 13: 16514–16543. https://doi.org/10.3390/ijms131216514 PMID: 23443116

76. Luo D, Langendries S, Mendez SG, De Ryck J, Liu D, Beirinckx S, et al. Plant Growth Promotion Driven by a Novel *Caulobacter Strain*. Mol Plant-Microbe Interact. 2019; 32: 1162–1174. https://doi.org/10.1094/MPMI-12-18-0347-R PMID: 30933667

77. Berrios L, Ely B. Plant growth enhancement is not a conserved feature in the *Caulobacter* genus. Plant Soil. 2020; 449: 81–95. https://doi.org/10.1007/s11104-020-04472-w

78. Hao ZP, Wang Q, Christie P, Li XL. Allelopathic potential of watermelon tissues and root exudates. Sci Hortic (Amsterdam). 2007; 112: 315–320. https://doi.org/10.1016/j.scienta.2006.12.030

79. Thanbichler M, Iniesta AA, Shapiro L. A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in *Caulobacter crescentus*. Nucleic Acids Res. 2007; 35: e137–e137. https://doi.org/10.1093/nar/gkm818 PMID: 17959646

80. Sun L-N, Yang E-D, Hou X-T, Wei J-C, Yuan Z-X, Wang W-Y. *Caulobacter rhizosphaerae* sp. nov., a stalked bacterium isolated from rhizosphere soil. Int J Syst Evol Microbiol. 2017; 67: 1771–1776. https://doi.org/10.1099/ijsem.0.001860 PMID: 28598303

81. Himaman W, Suksaard P, Mingma R, Matsumoto A, Duangmal K. *Cryptosporangium eucalypti* sp. nov., an actinomycete isolated from *Eucalyptus camaldulensis* roots. Int J Syst Evol Microbiol. 2017; 67: 3077–3082. https://doi.org/10.1099/ijsem.0.002086 PMID: 28857736

82. Li J, Chen X, Li S, Zuo Z, Zhan R, He R. Variations of rhizospheric soil microbial communities in response to continuous *Andrographis paniculata* cropping practices. Bot Stud. 2020; 61: 18. https://doi.org/10.1186/s40529-020-00295-1 PMID: 32542518

83. Rahman MM, Abdullah RB, Wan Khadijah WE. A review of oxalate poisoning in domestic animals: tolerance and performance aspects. J Anim Physiol Anim Nutr (Berl). 2013; 97: 605–614. https://doi.org/10.1111/j.1439-0396.2012.01309.x PMID: 22548678

84. Tamer AÜ, Aragno M, Şahin N. Isolation and Characterization of a New Type of Aerobic, Oxalic Acid Utilizing Bacteria, and Proposal of *Oxalicibacterium flavum* gen. nov., sp. nov. Syst Appl Microbiol. 2002; 25: 513–519. https://doi.org/10.1078/07232020260517643 PMID: 12583711

85. Watanabe T, Osaki M. Role of organic acids in aluminum accumulation and plant growth in *Melastoma malabathricum*. Tree Physiol. 2002; 22: 785–792. https://doi.org/10.1093/treephys/22.11.785 PMID: 12184982

86. Sahin N, Portillo MC, Kato Y, Schumann P. Description of *Oxalicibacterium horti* sp. nov. and *Oxalicibacterium faecigallinarum* sp. nov., new aerobic, yellow-pigmented, oxalotrophic bacteria. FEMS Microbiol Lett. 2009; 296: 198–202. https://doi.org/10.1111/j.1574-6968.2009.01636.x PMID: 19459944

87. Rahman MM, Kawamura O. Oxalate Accumulation in Forage Plants: Some Agronomic, Climatic and Genetic Aspects. Asian-Australasian J Anim Sci. 2011; 24: 439–448. https://doi.org/10.5713/ajas.2011.10208

88. Xu Q, Li K, Ruan Y, Kong Y, Liu M, Ling N, et al. Distinct response patterns of soil bacteria to oxalate imply their role in buffering soil acidification: Evidence from red soils with long-term fertilisation regimes. L Degrad Dev. 2019; 30: 1632–1641. https://doi.org/10.1002/ldr.3353

89. Pascual J, Huber KJ, Foesel BU, Overmann J. *Tellurimicrobium*. Bergey's Manual of Systematics of Archaea and Bacteria. Wiley; 2017. pp. 1–5. https://doi.org/10.1002/9781118960608.gbm01430

90. Kumar V, Irfan M, Datta A. Manipulation of oxalate metabolism in plants for improving food quality and productivity. Phytochemistry. 2019; 158: 103–109. https://doi.org/10.1016/j.phytochem.2018.10.029 PMID: 30500595

91. Berner M, Krug D, Bihlmaier C, Vente A, Müller R, Bechthold A. Genes and Enzymes Involved in Caffeic Acid Biosynthesis in the Actinomycete *Saccharothrix espanaensis*. J Bacteriol. 2006; 188: 2666–2673. https://doi.org/10.1128/JB.188.7.2666–2673.2006

92. Li Z-H, Wang Q, Ruan X, Pan C-D, Jiang D-A. Phenolics and Plant Allelopathy. Molecules. 2010; 15: 8933–8952. https://doi.org/10.3390/molecules15128933 PMID: 21139532

93. de Albuquerque FS, Macías-Rodríguez MÁ, Búrquez A, Astudillo-Scalia Y. Climate change and the potential expansion of buffelgrass (*Cenchrus ciliaris* L., *Poaceae*) in biotic communities of Southwest United States and northern Mexico. Biol Invasions. 2019; 21: 3335–3347. https://doi.org/10.1007/s10530-019-02050-5

94. Henry C, John GP, Pan R, Bartlett MK, Fletcher LR, Scoffoni C, et al. A stomatal safety-efficiency trade-off constrains responses to leaf dehydration. Nat Commun. 2019; 10: 3398. https://doi.org/10.1038/s41467-019-11006-1 PMID: 31363097

95. Einhellig FA, Kuan L-Y. Effects of Scopoletin and Chlorogenic Acid on Stomatal Aperture in Tobacco and Sunflower. Bull Torrey Bot Club. 1971; 98: 155. https://doi.org/10.2307/2484031

96. Rai VK, Sharma SS, Sharma S. Reversal of ABA-Induced Stomatal Closure by Phenolic Compounds. J Exp Bot. 1986; 37: 129–134. https://doi.org/10.1093/jxb/37.1.129

97. Tombesi S, Nardini A, Frioni T, Soccolini M, Zadra C, Farinelli D, et al. Stomatal closure is induced by hydraulic signals and maintained by ABA in drought-stressed grapevine. Sci Rep. 2015; 5: 12449. https://doi.org/10.1038/srep12449 PMID: 26207993

98. Yang CM, Chang IF, Lin SJ, Chou CH. Effects of three allelopathic phenolics on chlorophyll accumulation of rice (*Oryza sativa*) seedlings: II. Stimulation of consumption-orientation. Bot Bull Acad Sin. 2004; 45: 119–125. https://doi.org/10.7016/BBAS.200404.0119

99. Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, et al. Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. Nature. 2012; 488: 91–95. https://doi.org/10.1038/nature11336 PMID: 22859207

100. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, et al. Defining the core *Arabidopsis thaliana* root microbiome. Nature. 2012; 488: 86–90. https://doi.org/10.1038/nature11237 PMID: 22859206

101. Bulgarelli D, Schlaeppi K, Spaepen S, van Themaat EVL, Schulze-Lefert P. Structure and Functions of the Bacterial Microbiota of Plants. Annu Rev Plant Biol. 2013; 64: 807–838. https://doi.org/10.1146/annurev-arplant-050312-120106 PMID: 23373698

102. O'Brien FJM, Dumont MG, Webb JS, Poppy GM. Rhizosphere Bacterial Communities Differ According to Fertilizer Regimes and Cabbage (*Brassica oleracea* var. *capitata* L.) Harvest Time, but Not Aphid Herbivory. Front Microbiol. 2018; 9: 1–18. https://doi.org/10.3389/fmicb.2018.01620 PMID: 30083141

103. Tkacz A, Bestion E, Bo Z, Hortala M, Poole PS. Influence of Plant Fraction, Soil, and Plant Species on Microbiota: a Multikingdom Comparison. Bailey MJ, editor. MBio. 2020; 11: 1–17. https://doi.org/10.1128/mBio.02785-19 PMID: 32019791

104. Li X, Rui J, Mao Y, Yannarell A, Mackie R. Dynamics of the bacterial community structure in the rhizosphere of a maize cultivar. Soil Biol Biochem. 2014; 68: 392–401. https://doi.org/10.1016/j.soilbio.2013.10.017

105. Li X, Song Y, Yao S, Bian Y, Gu C, Yang X, et al. Can biochar and oxalic acid alleviate the toxicity stress caused by polycyclic aromatic hydrocarbons in soil microbial communities? Sci Total Environ. 2019; 695: 133879. https://doi.org/10.1016/j.scitotenv.2019.133879 PMID: 31425980

106. Liu J, Yu Z, Yao Q, Hu X, Zhang W, Mi G, et al. Distinct soil bacterial communities in response to the cropping system in a Mollisol of northeast China. Appl Soil Ecol. 2017; 119: 407–416. https://doi.org/10.1016/j.apsoil.2017.07.013

107. Qu F, ElOmari K, Wagner A, De Simone A, Beis K. Desolvation of the substrate-binding protein TauA dictates ligand specificity for the alkanesulfonate ABC importer TauABC. Biochem J. 2019; 476: 3649–3660. https://doi.org/10.1042/BCJ20190779 PMID: 31802112

108. Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, et al. Large-scale replicated field study of maize rhizosphere identifies heritable microbes. Proc Natl Acad Sci. 2018; 115: 7368–7373. https://doi.org/10.1073/pnas.1800918115 PMID: 29941552

109. Wang D, Lin J, Lin J, Wang W, Li S. Biodegradation of Petroleum Hydrocarbons by *Bacillus subtilis* BL-27, a Strain with Weak Hydrophobicity. Molecules. 2019; 24: 3021. https://doi.org/10.3390/molecules24173021 PMID: 31438460

110. Houlden A, Timms-Wilson TM, Day MJ, Bailey MJ. Influence of plant developmental stage on microbial community structure and activity in the rhizosphere of three field crops. FEMS Microbiol Ecol. 2008; 65: 193–201. https://doi.org/10.1111/j.1574-6941.2008.00535.x PMID: 18616582

111. Wen T, Zhao M, Yuan J, Kowalchuk GA, Shen Q. Root exudates mediate plant defense against foliar pathogens by recruiting beneficial microbes. Soil Ecol Lett. 2021; 3: 42–51. https://doi.org/10.1007/s42832-020-0057-z

112. Qiao Q, Wang F, Zhang J, Chen Y, Zhang C, Liu G, et al. The Variation in the Rhizosphere Microbiome of Cotton with Soil Type, Genotype and Developmental Stage. Sci Rep. 2017; 7: 3940. https://doi.org/10.1038/s41598-017-04213-7 PMID: 28638057

8.2. CAPÍTULO II: Diversidad ambiental y genómica comparativa de la bacteria *Solirubrobacter*

Considerando que *Solirubrobacter* es un género bacteriano presente en suelos contrastantes y en rizósferas de diversos hospederos, es probable que su genoma varíe de acuerdo al ambiente donde se encuentra pero que al mismo tiempo contenga un arsenal génico capaz de permitirle dispersarse a escalas geográficas amplias. *Solirubrobacter* nos permite hacer esta evaluación al estar presente en una gran variedad de suelos con distintas características y asociada a diversos hospederos de plantas, ambas en distintas locaciones geográficas. Esta sección se enfoca en realizar un análisis genómico comparativo de este género bacteriano.

Utilizando los genomas disponibles en bases de datos se realizó un análisis filogenético que nos permitió excluir uno de los ocho genomas, ya que no pertenece realmente al género *Solirubrobacter*: en la filogenia del gen 16S rRNA, en la filogenia del core y en el dendrograma generado con el análisis del GSS el genoma de ASM999324v1 se agrupa aparte del resto de *Solirubrobacter*, agrupándose incluso con los grupos externos *Bacillus subtitlis* y *Streptomyces coelicolor.* Utilizando los siete genomas que sí pertenecen al género se construyó un pangenoma conformado por 19,645 familias de proteínas, 2,644 de cuales conformaron el genoma núcleo del género. La adición secuencial de genomas incrementó continuamente el número nuevo de genes, indicando que el pangenoma de *Solirubrobacter* es un pangenoma abierto. Usando este pangenoma reclutamos secuencias de metagenomas de suelos y rizósferas, lo que nos permitió no solo determinar la presencia de las siete *Solirubrobater* de referencia sino también la posible existencia de nuevas cepas de esta bacteria. Este resultado se mantuvo para todos los metagenomas de rizósferas de distintas plantas y de distintas ubicaciones geográficas. La abundancia de secuencias reclutadas varió entre muestras de rizósfera, indicando que el hospedero y el ambiente que se genera en sus raíces influyen en las características genómicas de las cepas de *Solirubrobacter* presentes en estas muestras ambientales. Por el contrario, la presencia de *Solirubrobacter* en metagenomas de suelos fue dispar dependiendo de la proveniencia geográfica de la muestra. Los reclutamientos de suelos mexicanos presentaron un perfil similar al de las rizósferas, pero aquellos metagenomas de suelos canadienses, colombianos y estadounidenses mostraron una clara ausencia de este género. El suelo Canada_1 fue el único de los suelos no mexicanos que reclutó una cantidad considerable de secuencias del genoma núcleo pero menos de la mitad del genoma accesorio.

El pangenoma de referencia se anotó utilizando la base de datos de proteínas Clusters of Orthologous Groups (COGs) para obtener un perfil funcional y de familias génicas. Las categorías

más abundantes en el pangenoma fueron S (desconocido; 18.19%), K (transcripción; 11.53%), T (transducción; 6.44%), E (aminoácidos; 6.29%) y G (carbohidratos; 5.94%), mientras que en genoma núcleo las más abundantes fueron S (desconocido; 15.61%), E (aminoácidos; 8.01%), K (transcripción; 7.47%), C (energía; 6.81%), NA(-)(6.07%), J (ribosomal; 5.86%) y G (carbohidratos; 5.24%). Dentro de las proteínas identificadas en el pangenoma núcleo se encuentran proteínas de la familia ROK, involucradas en el metabolismo de carbohidratos (Bekiesch et al., 2016), y dominios de cupinas, asociadas a adquisición de nutrientes (Khuri et al., 2001). Se identificaron 25 factores sigma diferentes y diversos mecanismos de defensa, como los genes *drrA* y *drrB*, transportadores de la micotoxina tricoteceno, sistemas de toxina antitoxina, dioxigenasas para la degradación de compuestos aromáticos, bombas para arsénico, los genes del operón *uvrABC*, metiltransferasas, entre otros. Únicamente el 13.7% de las proteínas involucradas en metabolismo secundario se encontraron en el pangenoma, el restante 86.21% se detectaron en el genoma accesorio.

La expansión del genoma como consecuencia de familias génicas asociadas a ciertas funciones nos permite determinar a qué funciones está destinada una mayor porción del genoma de este género. Las familias más grandes detectadas están compuestas por 22 genes duplicados y pertenecen a la categoría de sensado ambiental, particularmente asociadas al regulón Lux, a una integrasa/transposasa/recombinasas tipo DDE y a flagelina, esta última conformada por 9 genes. El análisis enfocado en las proteínas de biosíntesis de flagelo reveló que en *Solirubroabcter* están presentes únicamente el 33.3% de las proteínas presentes en *Conexibacter* y *Patulibacter*. Algunas proteínas esenciales están ausentes, como MotB que forma parte del estator y FliQPR que conforman el sistema de secreción de flagelina para la construcción del flagelo. *Solirubrobacter soli* es la única de las siete cepas que contiene las secuencias de MotB y FliR en su genoma.

Con el EEP logramos identificar 9,906 proteínas potencialmente nuevas para *Solirubroabcter*. Con la anotación del EEP en la base de datos COG obtuvimos un perfil funcional donde algunas de las categorías incrementan su abundancia gracias a las proteínas reclutadas de los metagenomas. Esto, además de permitirnos expandir la descripción del género, nos ofrece nueva información acerca de adaptaciones locales de *Solirubrobacter* en cada ambiente. En el ordenamiento construido utilizando las muestras metagenómicas se forman dos agrupamientos, uno con todas las referencias y algunos metagenomas y otro únicamente con metagenomas, sugiriendo la existencia de nuevas cepas de *Solirubroabcter*. Por otro lado, la búsqueda de genes de interacción planta-microorganismo en el EEP con los HMM resultó en un total de 11 *hits* positivos, mientras que en los controles positivos se obtuvieron 173 *hits*. El análisis de la red de co-ocurrencia resultó en un valor de grado de conectividad de 97 (μ=40.08) para *Solirubrobacter*, mientras que el promedio de la red fue de 40.08.

La centralidad de intermediación fue bajo, con un valor de 0.003 mientras que la centralidad de cercanía fue de 0.376, mientras que el promedio para toda la red fue de 0.020.

El *preprint* de este trabajo se muestra a continuación y puede ser consultado en la dirección https://doi.org/10.1101/2023.10.31.564804. Actualmente el manuscrito se encuentra bajo revisión y en espera de ser publicado.

Unraveling the Genomic and Environmental Diversity of the Ubiquitous *Solirubrobacter*

1.1 Author names

Angélica Jara-Servín [1,2] https://orcid.org/0000-0002-0699-6289

Gerardo Mejia [1,2] https://orcid.org/0009-0001-3391-4939

Miguel F. Romero [3] https://orcid.org/0000-0002-3799-717X

Mariana Peimbert [4] https://orcid.org/0000-0002-4881-8587

Luis D. Alcaraz [1*] https://orcid.org/0000-0003-3284-0605

1.2 Affiliation(s)

[1] Laboratorio de Genómica Ambiental, Departamento de Biología Celular, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico

[2] Posgrado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México, Mexico City, Mexico

[3] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[4] Departamento de Ciencias Naturales, Unidad Cuajimalpa, Universidad Autónoma Metropolitana, Mexico City, Mexico

1.3 Corresponding author and email address

*Correspondence: Luis D. Alcaraz, lalcaraz@ciencias.unam.mx

1.4 Keywords

Solirubrobacter diversity, Phylogenomics of *Solirubrobacter*, Gene recruitment analysis, Pangenomic analysis, Environmental extended pangenome, Metagenomic diversity in soils.

2. Abstract

Solirubrobacter, a genus within the Actinobacteriota phylum, is commonly found in soils and rhizospheres yet remains unexplored despite its widespread presence and diversity, as revealed through metagenomic studies. Previously recognized as a prevalent soil bacterium, our study delved into phylogenomics, pangenomics, environmental diversity, and bacterial interactions of *Solirubrobacter*. Analyzing the limited genomic sequences available for this genus, we uncovered a pangenome consisting of 19,645 protein families, with 2,644 constituting a strict core genome. While reported isolates do not exhibit motility, we intriguingly discovered the presence of flagellin genes, albeit with an incomplete flagellum assembly pathway. Our examination of 16S ribosomal genes unveiled a considerable diversity (3,166 operational taxonomic units OTUs) of *Solirubrobacter* in Mexican soils, and co-occurrence network analysis indicated its extensive connectivity with other bacterial taxa. Through phylogenomic analysis, we delved into the relatedness of sequenced strains and notably dismissed ASM999324v1 as a member of this genus. Our investigation extended to the metagenomic diversity of *Solirubrobacter* across various environments, affirming its pervasive presence in rhizospheres and certain soils. This broader pangenomic view revealed genes linked to transcription, signal transduction, defense mechanisms, and carbohydrate metabolism, highlighting *Solirubrobacter's* adaptability. We observed that *Solirubrobacter's* prevalence in rhizospheres is geographically indiscriminate, prompting intriguing questions about its potential interactions with plants and the biotic and abiotic determinants of its soil occurrence. Given its richness and diversity, *Solirubrobacter* might be a versatile yet overlooked keystone species in its environments, meriting further recognition and study.

3. Impact statement

This study explored the enigmatic world of *Solirubrobacter*, a widespread microbe commonly found in soils and plants across various regions. Despite its prevalence, little is known about its genetic diversity and functionality and how it thrives in diverse environments. Our research unveils the genetic secrets of *Solirubrobacter*, shedding light on its adaptability and ecological

interactors and roles. We showed that *Solirubrobacter* environmental prevalence makes it a good candidate for studying the genetic basis of being a successful microbe associated with soil and plants.

4. Data summary

Data, scripts and statistical analysis available in GitHub:

https://github.com/genomica-fciencias-unam/Solirubrobacter

Sequences, phylogenetic analysis, raw data structures:

https://doi.org/10.6084/m9.figshare.24446521

16S rRNA gene raw data:

https://www.ncbi.nlm.nih.gov/sra/PRJNA603586

https://www.ncbi.nlm.nih.gov/sra/PRJNA603590

Shotgun metagenomes:

https://www.ncbi.nlm.nih.gov/bioproject/603603

All supporting data, code, and protocols are within the article, supplementary files, and described repositories.

5. Introduction

Solirubrobacter is an Actinobacteriota bacterial genus whose abundance appears high in soils worldwide [1]. The genus's first type strain and definition was isolated from vermicompost in 2003, *Solirubrobacter pauli* [2]. *S. pauli* was described as a genus of Gram-positive bacilli, pink-colored, aerobic and mesophilic, non-sporulating, without motile structures, low desiccation resistance, no growth at 1% NaCl, and with a ~70% G+C content [1,2]. Since then, *S. soli*, *S. ginsenosidimutans, S. phytolacca,* and *S. taibaiensis* have been isolated and described. Both *S. soli* and *S. ginsenosidimutans* were isolated from ginseng fields, while *S.*

phytolacca and *S. taibaiensis* from the roots and stem of the Indian pokeweed *Phytolacca acinosa* [3-5]. The carbohydrate utilization capabilities, growing conditions, and phenotype vary depending on the species (Table 1). The cultivated isolates of *Solirubrobacter* are very limited, probably due to their slow growth and the lack of a specific selective medium, even though poor nutrient media seem to contribute to their isolation [6]. *Solirubrobacter* is also related to harsh environments, such as black chickpea fertilized culture soils [7] and coal and soil samples from opencast coal mines [8]. This bacterial genus has been linked with metabolizing contaminants such as detergents [9] and positively correlated with bioaccumulation of Cd and Zn in plants [10]. Despite being widely reported in agricultural soils, *Solirubrobacter* seems sensitive to antibiotics, decreasing its abundance by more than 50% in agricultural soils contaminated with antibiotics of manure origin [11]. Even though some strains have been reported as non-resistant to desiccation [2], in studies analyzing UV exposure, *Solirubrobacter* appeared in slightly increased abundance [12] and even as one of the key genera in microbial communities from sand desert exposed to intense solar UV radiation [13].

Numerous environmental studies have detected this bacterium through 16S rRNA gene sequencing, especially in soils and rhizospheres of diverse plants. Our group has examined soil and rhizosphere microbiome samples across Mexico, from soils with very diverse physicochemical properties and rhizospheres of wild and cultivated plants. In these hundreds of samples we have observed the prevalence of *Solirubrobacter* [14-16]. The limited data on this genus obstructs elucidating the precise role *Solirubrobacter* plays in the rhizosphere. However, both favorable and adverse correlations with plant growth have been documented [15, 17, 18], implying potential gene presence promoting host plant development in *Solirbrobacter* genomes. Interestingly, *S. soli* DSM 22325 was sequenced as a directed effort of the Genome Encyclopedia of Bacteria and Archaea (GEBA) [19]. The GEBA consortium also identified *S. soli* as the top recruiting genome when aligning it against metagenomes, with preferences for terrestrial (50%), plant (34%), and aquatic environments (6.5%) [19].

The widespread prevalence of this bacterial genus resonates with the assertion, "Everything is everywhere, but the environment selects," a notion sparking ecological debates since the eighteenth century [20]. This theory posits that due to the swift population expansion and vast size of microbial communities, microbial taxa can disperse rapidly, and their distributions are shaped by environmental factors—thus negating the roles of historical biogeography and dispersal events in community formation [21, 22]. Among these environmental determinants are established factors like pH, alongside more intricate ones such as pedogenesis, trophic

112 status, available plant debris, and vegetative cover [23, 24]. However, some research
113 indicates that environmental influences wane beyond a specific geographic extent (10 km),
114 which suggests that the same biogeographic and historical processes of macroorganisms can
115 apply to bacterial populations and communities [21, 25]. Geographical barriers foster endemic
116 microorganisms, primarily as they prevent a significant dispersion [21, 26]. Even so, instances
117 of microorganisms from the same taxonomic group identified across disparate geographical
118 locations have been recorded, with the underexplored genus *Solirubrobacter* being one such
119 case. Despite its ubiquity across diverse soils and plant hosts, the genetic diversity enabling
120 *Solirubrobacter* to flourish in varied and contrasting environments remains overlooked.
121 Delving into the genetic makeup of *Solirubrobacter*, in light of the "Everything is everywhere,
122 but the environment selects" concept, not only elucidates this genus but also sheds light on
123 how geographically distant locales and differing environmental attributes might influence the
124 genetic distribution and adaptive strategies of a singular bacterial genus.

125 *Solirubrobacter* is a common and prevalent bacterium in soils and rhizospheres, yet its biology
126 remains largely unexplored. This study aims to unravel the diversity and functionality of
127 *Solirubrobacter*, striving to elucidate its ecological roles, adaptive capabilities, and interactions
128 with other microorganisms. To achieve this, we adopted a comprehensive strategy
129 encompassing phylogenomics, pangenomics, and metagenomic analysis. Initially, we
130 collected and examined the available genomic sequences of the *Solirubrobacter* genus.
131 Additionally, we analyzed environmental 16S ribosomal gene sequences to discern the
132 diversity and distribution of *Solirubrobacter* across various environments. Subsequently, we
133 constructed an Environmental Extended Pangenome (EEP) by recruiting potential new
134 *Solirubrobacter* proteins from soil and rhizospheric metagenomes to investigate if similar
135 environments select for comparable features. These methodologies enabled us to determine
136 the prevalence of *Solirubrobacter* in soil and rhizospheres, regardless of geographic location.
137 Our integrated approach enhanced the understanding of *Solirubrobacter* genomic diversity
138 and its significance in terrestrial ecosystems, laying a foundation for further investigation and
139 continuous study of this understudied microbial genus.

140

141 6. Methods

142 *Solirubrobacter* core genome retrieval and phylogenetic analysis

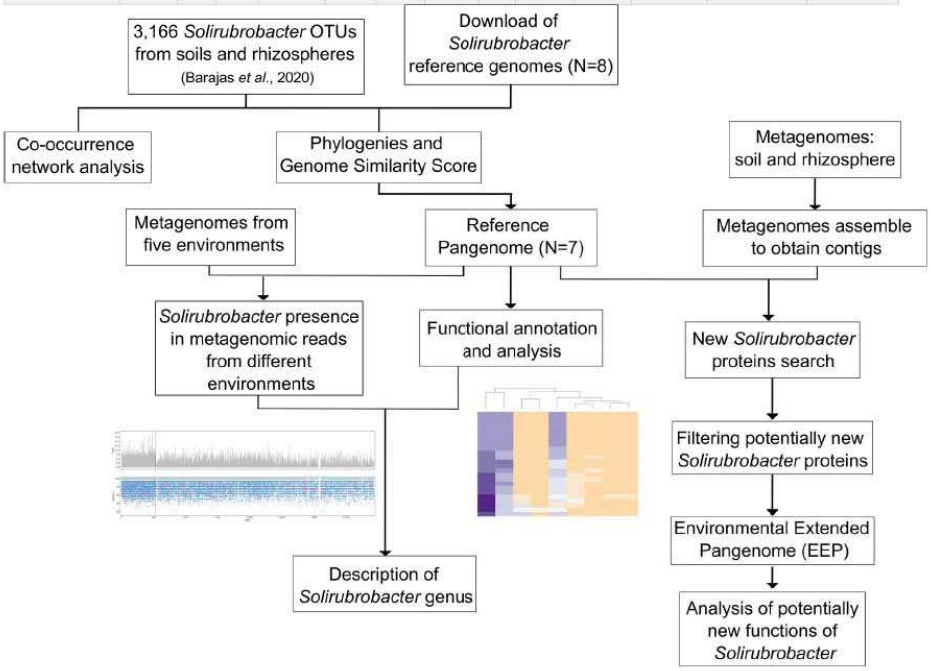
143 We summarized the main methodological steps (Fig. 1). All available genomes and draft
144 assemblies from the genus *Solirubrobacter* were retrieved from NCBI in February 2023. In

total, eight *Solirubrobacter* species were downloaded (Table 1). Genome quality and completeness were evaluated using Checkm [27]. Using the GET_HOMOLOGUES pipeline [28], we calculated a pangenome of all *Solirubrobacter* strains while using *Streptomyces coelicolor* and *Bacillus subtilis* as outgroups (NCBI Accessions: GCA_013317105.1 and GCF_000009045.1). Core genes were retrieved and aligned individually with MAFFT [29] for further concatenation into a core genome alignment. Then, using R v 4.2.2 [30] ape [31] and phangorn [32] libraries, we calculated the best evolutionary model of the core genome alignment with ten taxa, a length of 105,813 characters, and 48,460 site patterns. The best Akaike information criterion (AIC) model was LG [33], with a gamma-distributed rate (k = 4). The model was used to build a consensus maximum likelihood phylogeny with 1,000 bootstraps. The complete analysis protocol is available on GitHub (https://github.com/genomica-fciencias-unam/Solirubrobacter). Additionally, we carried out the genome similarity score (GSS) analysis [34-36]. A set of seven *Solirubrobacter* genomes was retained and used for further analysis and are henceforth called reference genomes or reference species.

Table 1. *Solirubrobacter* reference genome information and traits.

Strain	*Accession*	*Size (Mb)*	*%GC*	*Completeness*	*Country*	*Color*	*Growth T (°C)*	*Growth >1% NaCl*	*Motility*	*Sporulation*	*Environment*	*Genome submitter*	*Associated publications*
S. soli DSM22325	ASM42366v1	9.31	71.4	98.71	South Korea	White	15-33	-	-	-	Ginseng field	DOE Joint Genome Institute	Kim et al., 2007, Mukherjee et al., 2017
S. sp URHD0082	ASM42594v1	6.64	72.3	99.14	N.d.	N.d.	N.d.	N.d.	N.d	N.d	Mediterranean grassland	DOE Joint Genome Institute	N.A.
S. sp CPCC204708	ASM334462v2	7.59	69.9	98.71	China	N.d.	N.d.	N.d.	N.d	N.d	Desert sand	Chinese Academy of Medical Sciences & Peking Union Medical College	N.A.
S. pauli	ASM363375v1	7.13	72.1	98.71	USA	Pink	28-30	-	-	-	Vermicompost	DOE Joint Genome Institute	Singleton et al., 2003
S. ginsenosidimutans	ASM2758720v1	9.69	71	99.57	China	Pale yellow	18-37	-	-	-	Ginseng field	Chinese Academy of Medical Sciences & Peking Union Medical College	An et al., 2011
S. phytolaccae	ASM2758719v1	7.73	71.5	98.71	China	White	25-30	+	-	-	Root of *Phytolacca acinosa*	Chinese Academy of Medical Sciences & Peking Union Medical College	Wei et al., 2014

S. taibaiensis	ASM2758722v1	8.51	68.62	98.71	China	White	7-33	-	-	-	Stem of *Phytolacca acinosa*	Chinese Academy of Medical Sciences & Peking Union Medical College	Zhang et al., 2014
ASM999324v1	ASM999324v1	2.02	69.9	93.18	USA	N.d.	N.d.	N.d.	N.d	N.d	Aquifer samples (geyser)	DOE Joint Genome Institute	Probst et al., 2018



161 **Fig. 1.** Overview of this work.

162 ***Solirubrobacter* diversity and networks using 16S rRNA gene**

163 We reanalyzed 107 microbiomes of rhizosphere, endosphere, and soils from Mexico [14] to
164 extract the sequences of the OTUs corresponding to *Solirubrobacter*. The dataset included
165 255,275 16S OTUs (97% identity). For genus-level analysis, the dataset was managed with
166 R's phyloseq [37], vegan [38], and ggplot2 [39] libraries. We also extracted the 16S rRNA
167 genes from the reference genome sequences using barrnap (v. 0.7) [40] and concatenated
168 them into the 16S microbiome sequences. We aligned all the 16S using SSU-align [41],
169 followed by a phylogeny using FastTreeMP [42].

For the network analysis, the 16S OTUs sequences were merged to the genus level, resulting in 1,748 bacteria genera. With the R igraph library [43], we calculated a co-occurrence, undirected network. We exported it as a graphml object to be analyzed by Gephi graph visualization and manipulation software (v 0.10) [44]. Complete protocols and datasets are available on GitHub (https://github.com/genomica-fciencias-unam/Solirubrobacter) and FigShare (https://doi.org/10.6084/m9.figshare.2444652).

Solirubrobacter pangenome analysis

All *Solirubrobacter* genomes were annotated using Prokka (v.1.12) [45] to generate gff3 output files that served as input for Roary [46]. Roary is designed to construct pangenomes from related species. An UpSet diagram [47] was constructed with R to illustrate the shared and unique gene distribution. The resulting output files with the sequences comprising the pangenome were annotated using the Clusters of Orthologous Groups (COG) [48] and Kyoto Encyclopaedia of Genes and Genomes (KEGG) [49] database. The resulting annotations were used to create a matrix with counts of the number of proteins of each COG in the reference pangenome, the core genome, and each *Solirubrobacter* strain. The matrix was standardized using Z values. Using only the sequences successfully annotated in COG, a functional profile was constructed using the phyloseq, ggplot2, gplots [50], and R default packages. Hierarchical clustering was performed using the *hclust* method on a Bray-Curtis dissimilarity matrix [51].

After annotation, we were able to detect proteins involved with flagellar biosynthesis. Considering that *Solirubrobacter* isolates are described as non-motile, we retrieved all proteins involved with flagellar assembly (KEGG and COG databases). All selected proteins were used to construct a presence-absence matrix. For comparison, the reference genomes of *Conexibacter woesi* and *Patulibacter minatonensis* were downloaded from NCBI (Accession number: GCA_013317105.1 and ASM51932v1, respectively) and annotated against the same databases. We chose these bacteria because they are flagellated and belong to the same order as *Solirubrobacter*.

Solirubrobacter recruitment with metagenomes from multiple environments

To evaluate the presence of the reported *Solirubrobacter* species in environmental samples, we used a total of 36 metagenomes from rhizospheres and soils from different geographic locations. Mexican soils and rhizospheres comprised 16 of the 36 selected metagenomes and were retrieved from a previous study evaluating soil and rhizosphere microbial composition (NCBI Bioproject PRJNA603603) [14]. The remaining soil metagenomes were selected if the sample came from an agricultural field and the location was geographically distant from our Mexican soils. The criteria for rhizospheres was also a geographical separation and plant host different from the Mexican ones. Lake water, phyllosphere, and sea sediment metagenomes were used as negative controls. All metagenomes used for this section are listed in Table S1. Reads associated with *Solirubrobacter* from the 36 metagenomes were identified using the reference pangenome sequences and tBLASTn [52]. The resulting hits were filtered according to bitscore value and quality-checked using Trimmomatic [53]. Later, the constructed pangenome and the *Solirubrobacter* metagenomic reads were aligned using Promer [54]. We then generated identity graphs showing the percentage identity and coverage of the alignment through the *promer_deid.py* in-house script available on GitHub. After analysis of the Promer alignments, the rest of the study was carried out using only the metagenomes from Mexican soils (six metagenomes. Samples 1 to 6; Table S1) and Mexican rhizospheres (ten metagenomes. Samples 13 to 22) were selected for the construction of the Environmental Extended Pangenome (EEP).

Metagenomic data assembly

The sixteen selected Mexican soils and rhizosphere shotgun metagenomes from the NCBI Bioproject PRJNA603603 [14] were subjected to quality control using Trimmomatic, and paired-end matched reads were retained for subsequent analysis. Metagenomic reads from samples 18 to 22 (rhizospheres; Table S1) were filtered out, matching *Solanum lycopersicum* genome (Accession number: GCF_000188115.5), while samples 13 to 17 (rhizospheres; Table S1) were filtered against the *Arabidopsis thaliana* genome (Accession number: GCF_000001735.4) with Bowtie2 [55]. Metagenomic libraries were assembled using a hybrid assembly approach. First, SPADES [56] with –a meta function was used to generate contigs, and high-quality reads were mapped against the SPADES contigs. Second, all unmapped reads were subjected to a second assembly with Velvet [57]. The resulting contigs from both assemblies were merged, and only a minimum of 100 bp length contigs were retained. All obtained contigs were used to obtain an EEP for *Solirubrobacter*.

Solirubrobacter Environmental Extended Pangenome

To recruit potentially new *Solirubrobacter* proteins, we used the pangenome sequences as a reference and, using Bbmap [58], mapped them against the previously assembled contigs of the 16 selected metagenomes. All contigs containing *Solirubrobacter* proteins were extracted, and coded proteins surrounding the already known *Solirubrobacter* protein were considered potentially new proteins of this genus. We performed a filtering step using BLASTx [59] with the pangenome as a database and filtering by an e-value of 1e-5. We then used Prodigal [60] to predict ORFs and coding proteins for the new sequences and the reference sequences of *Solirubrobacter*. All sequences were concatenated in a single file and ran through CD-HIT [61] using the following parameters: -n 3 -c 0.50 -aS 0.7 -d 0 -M 30000 -T 10. These settings allowed us to create an EEP composed of the potentially new *Solirubrobacter* proteins and the proteins from the reference species. The EEP was annotated using COG and KEGG as carried out for the reference pangenome.

Functional and diversity analysis of the *Solirubrobacter* environmental extended pangenome (EEP)

Using only the sequences successfully annotated in COG, a functional profile of the EEP was constructed using the same methods listed for the reference pangenome functional profile. Additionally, for the annotated sequences, differences between samples were analyzed through an unconstrained Principal Coordinate Analysis (PcoA) [62] on Bray-Curtis dissimilarity matrices; the clustering was evaluated through the ANOSIM statistical function [63]. To further describe the metabolic capability of the genus *Solirubrobacter*, the number of duplicated sequences per sample was counted using the output file of the CD-HIT analysis. This analysis allowed us to determine gene families and evaluate which metabolic activities *Solirubrobacter* assigns a higher portion of its genome. Detailed statistical and bioinformatic methods for all the methods used in this study can be accessed on GitHub.

Annotated sequences from the reference and EEP were manually analyzed to identify any proteins related to plant-microorganism interactions. Additionally, Hidden Markov Models (HMM) for Actinobacteria proteins involved in plant-microorganism interactions [64] were used to search for proteins of this type of interaction in *Solirubrobacter* metagenomic proteins from Mexican soils and rhizospheres. This search used HMMER v. 3.3.2 [65], and only the three best hits were requested. *Streptomyces coelicolor* and *Streptomyces griseus* genomes (Accession number: GCA_013317105.1 and ASM1060v1, respectively) were used for comparison since strains from these species are known to interact with plants.

7. Results

Functional composition of *Solirubrobacter* references core genome

A total of 13,160 proteins were annotated (66.98% of the pangenome, Table S3) using the COG ontology, with 2,421 of them belonging to the core genome. Hypothetical proteins accounted for 6,484 of the pangenome, of which 223 were from the core genome. We visualized the abundance of the COGs per species, as well as the core genome and pangenome (Fig. 2). The most abundant COGs in the pangenome were S (unknown; 18.19%), K (transcription; 11.53%), T (transduction; 6.44%), E (aminoacids; 6.29%), and G (carbohydrate; 5.94%). Regarding annotated sequences from the core genome, the most abundant COGs were S (unknown; 15.61%), E (aminoacids; 8.01%), K (transcription; 7.47%), C (energy; 6.81%), NA(-)(6.07%), J (ribosomal; 5.86%), and G (carbohydrate; 5.24%).

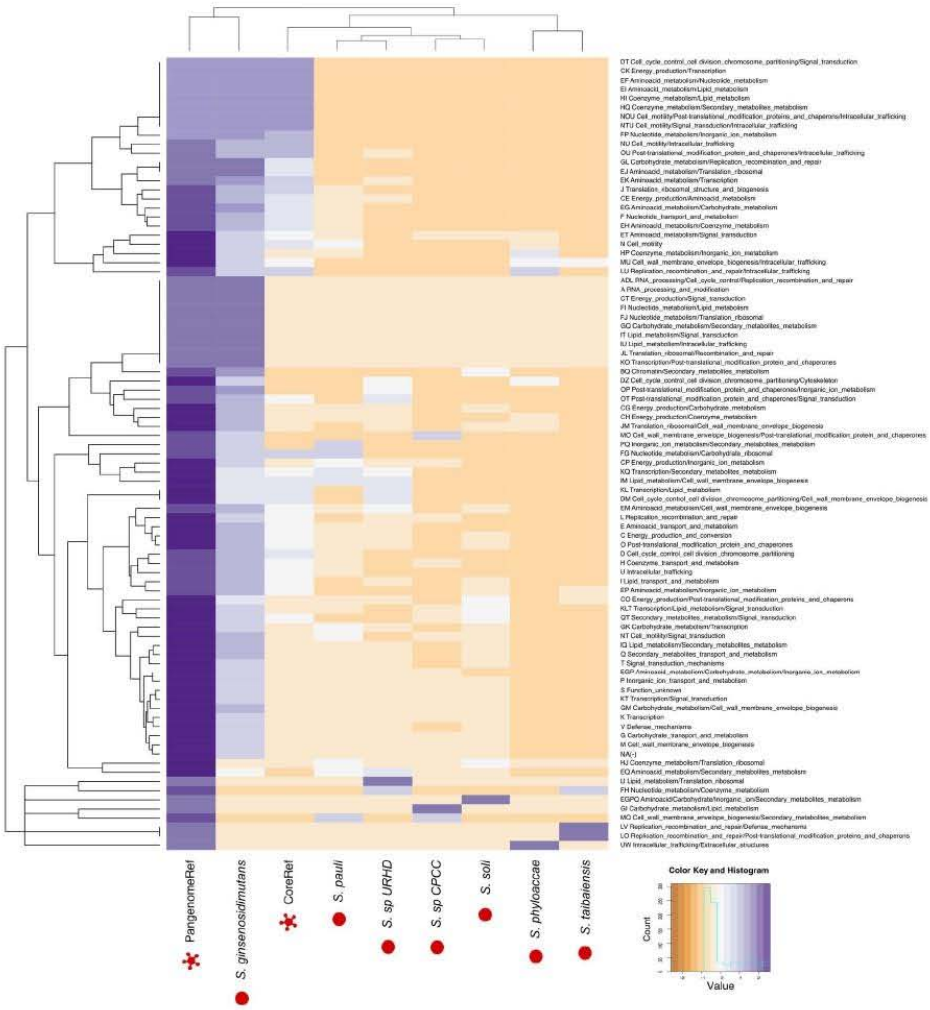


Fig. 2. Functional profile of *Solirubrobacter* references pangenome. Heatmap showing normalized Cluster of Orthologous Groups (COGs) frequencies within each group are shown as a two way clustering. Circles refer to reference strains and stars to the constructed pangenome and core genome of *Solirubrobacter*.

The ability to exploit different carbohydrate sources is reflected in the core COG G (carbohydrate). In this COG we found five protein families identified as ROK family proteins, as well as enzymes for xylose isomerases (8), alpha amylase domains (2), starch synthase

(1), hydrolases (12), HpcH/HpaI aldolase/citrate lyase family (1), and polysaccharide deacetylase (2). Moreover, a total of 8 cupin proteins or cupin domains were identified between COGs G and S (carbohydrate and unknown, respectively).

Sigma factors lead to a wide gene expression diversity by modifying the specificity of RNA polymerase [66]. Among COGs K (transcription), T (transduction), and KT (transcription/transduction) of the core genome we detected 25 sigma factors, of which three were ECF sigma factors and 13 sigma-70 factors. Histidine kinases are an important part of two-component systems environmental sensing [67] and a total of 18 annotated protein families were histidine kinases. Additionally, 12 protein families were identified as the helix_turn_helix domain of the Lux regulon or as LuxR regulators.

Regarding defense mechanisms (COG V), ABC transporters represented 26.47%, including daunorubicin resistance (4) and chromate (1) transporters. We detected homologs to genes *drrA* and *drrB*, conferring resistance to daunorubicin, an antibiotic produced by bacteria from the genus *Streptomyces* [68, 69]. Two family proteins were annotated as fungal trichothecene efflux pump (TRI12), contributing to the secretion of this mycotoxin from the cell. The AAA domain from the putative AbiEii Type IV Toxin-Antitoxin system was also identified in the core genome. AbiE TA systems confer phage resistance to the cell-inducing bacteriostasis: AbiEii acts as a GTP-binding NTase toxin that is neutralized by the expression of AbiEi [70]. A total of 18 protein families were annotated as dioxygenases, including glyoxalase/bleomycin resistance and protocatechuate dioxygenases, key enzymes in the degradation of various aromatic compounds [71]. ArsB, an arsenic pump membrane along cognate regulators, was identified in the core genome. Stress-related proteins were detected. Three families of cold shock proteins are involved in cellular responses against low-temperature stress, while four protein families are involved in the use of trehalose as a compatible solute (COG G; carbohydrate). The protein YaaA is involved in peroxide stress (COG S; unknown). From COG P (inorganic ions), the iron-storage gene *bfrB* gene was also found in the core genome. Finally, multiple repair systems for damaged DNA were found in 20 family proteins from the COG L (replication and repair), including the *uvrABC* operon. Secondary metabolism (COG Q) in the core genome is just 13.79%, while 86.21% of detected genes were in the accessory genome. In the core genome, three protein families were identified as cytochrome P450, while ten were identified as methyltransferases involved with secondary metabolism.

Upon annotating the core genome using KEGG, 52.8% of the core sequences were successfully annotated, enabling the reconstruction of biosynthetic pathways for tryptophan,

phenylalanine, tyrosine, histidine, lysine, and the branched-chain amino acids valine, leucine, and isoleucine (Fig. S1). The metabolic pathway for arginine degradation to ornithine was identified, contrasting with the absence of an arginine biosynthetic pathway. The reference core genome lacked enzymes for nitrate reduction, yet enzymes and an ammonium transporter necessary for glutamate synthesis from ammonia were present. The *argFGH* operon for arginine biosynthesis was missing in the core (only *argH* was detected through COG annotation). Transporters for the missing amino acids were not identified in the core genome (except for the branched-chain amino acid transporter). However, a methionine transporter was found in the pangenome, along with portions of cystine and urea transporters. This suggests that for other essential amino acids, *Solirubrobacter* likely relies on membrane transporters for uptake.

Gene families expansions in *Solirubrobacter*

We clustered all the proteins into single-copy genes, gene families from two to five members, and equal or larger than six within each *Solirubrobacter*-predicted proteome (Table S4). The most prominent gene families accounted for a maximum of 22 genes in *S. ginsenosidimutans* and *S.* sp. CPCC. They were annotated as COGs K (transcription) and L (replication and repair) (Fig. 3a). COG K (transcription) proteins were identified as helix_turn_helix Lux Regulon and COG L (replication and repair) as RNA-mediated transposition proteins. Gene families with nine genes coding for flagellin from COG N (motility) are present in strains *S. pauli, S. phytolaccae*, and *S. soli*. All strains, except for *S. phytolaccae* and *S.* sp. CPCC, have gene families from COG KT (transcription/transduction) of sizes ranging from seven to up to 15 and were identified as LuxR family transcriptional regulator genes. Interestingly, six families containing six or more hypothetical proteins were found. Of these hypothetical proteins, we analyzed family number 528 through a structural alignment that suggests that this family is homologous to the antimicrobial peptide Brevinin-1JDa (Fig. S2).

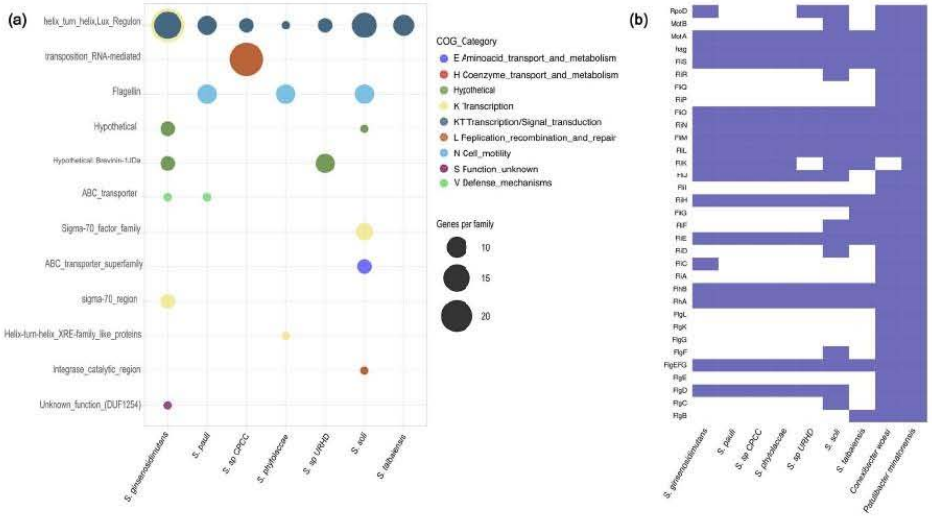


Fig. 3. Extended families in *Solirubrobacter* references. (a) Bubble plot showing gene families with six or more duplicated genes and their annotation according to Cluster of Orthologous Groups (COGs) for each reference strain. (b) Detected proteins involved in flagellar biosynthesis in *Solirubrobacter*. The presence (blue) or absence (white) of each protein per *Solirubrobacter* reference is displayed. Flagellar biosynthesis proteins present in *Conexibacter woesi* and *Patulibacter minatonensis* are shown as comparison, because they are *Actinobacteriota* reported to swim. All flagellar proteins shown were selected based on the annotation using COG and KEGG databases.

Proteins involved with motility and flagellar assembly were detected among *Solirubrobacter* references (Fig. 3b). A total of 33 proteins bound to flagellar biosynthesis were identified in the Actinobacteriota flagellated species *Conexibacter woesi* or *Patulibacter minatonensis*, from which only 11 (33.33%) were present in all *Solirubrobacter* references. The rest of the proteins were detected in either *C. woesi* or *P. minatonensis* but were absent in most of the references. All *Solirubrobacter* and Actinobacteriota controls contained flagellin synthesis genes (*hag*). Moreover, gene families with nine genes coding for flagellin from COG N (motility) are present in strains *S. pauli, S. phytolaccae*, and *S. soli* (Fig. 3a). Among the genes detected in *Solirubrobacter* are structural components such as the hook (FlgE), along with regulatory elements like FliK, responsible for controlling hook size, and FliS, a chaperone facilitating flagellin transport [72]. Prime proteins such as the cytoplasmic membrane ring (FliF) and the motor component (MotB) are absent in *Solirubrobacter*.

366 ***Solirubrobacter* comparative genomics**

367 We used the eight NCBI available genomes of *Solirubrobacter pauli* (type strain), *S. soli*
368 DSM22325 (*S. soli*), *S.* sp. CPCC204708 (*S.* sp. CPCC), *S.* sp. URHD0082 (*S.* sp. URHD),
369 *S. ginsenosidimutans, S. phytolaccae, S. taibaiensis,* and ASM999324v1. The main genomic
370 features and reported phenotypic descriptors were summarized in Table 1. The genome size
371 of ASM999324v1, annotated as a *Solirubrobacter* in GenBank, is considerably smaller (2.02
372 Mb) than the average genome size of the rest of the *Solirubrobacter* (8.08 ± 1.12 Mb). Using
373 a 16S rRNA gene recovered from the genome sequences we did a phylogenetic
374 reconstruction using *Streptomyces coelicolor* and *Bacillus subtilis* as outgroups (Fig. 4a).
375 From the 16S of the genomes, the basal clades have bootstrap support (>0.6). In contrast, the
376 top clade formed by *S. pauli* and *S. phytolaccae* had 0.5 support. We noted that ASM999324v1
377 is within the outgroup clade at all times (bootstrap = 1), with low support of the inner branching
378 (Fig. 4a). Then, using a core genome maximum-likelihood phylogeny, we observed complete
379 support (bootstrap = 1) in the consensus tree (1,000 bootstraps) for all branches of
380 *Solirbubrobacter* and the outgroup clade by *B. subtilis* and *S. coelicolor* but with a single
381 branch of ASM999324v1 (Fig. 4b).

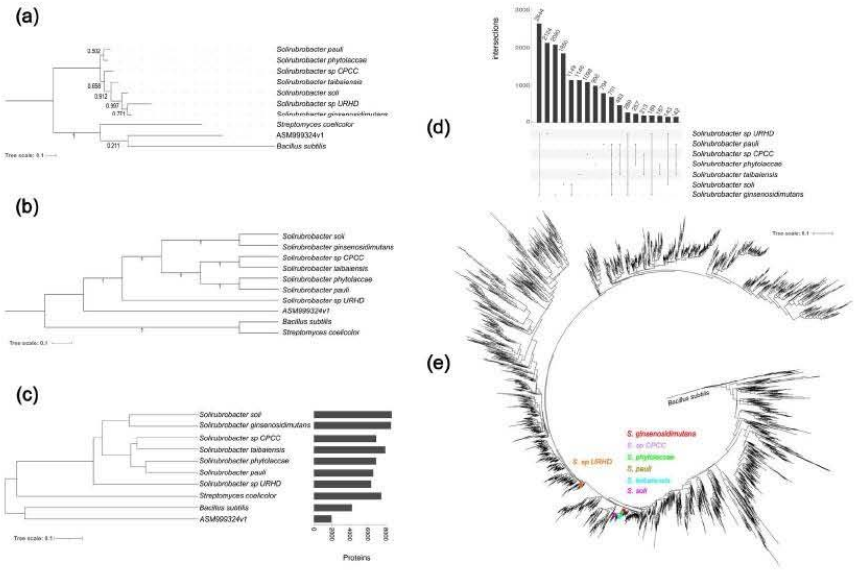
382 The core genome consensus phylogeny showed *S.* sp. URHD as the basal clade, sister to
383 two other clades, one formed by *S.* sp. CPCC, *S. taibaiensis, S. phytolaccae*, and *S. pauli*,
384 and a top clade branched into *S. soli* and *S. gingsenosidimutans* (Fig. 2b). Both 16S and core
385 genome phylogenies located ASM999324v1 in a separate branch than the rest of species
386 from *Solirubrobacter*. Additionally, we conducted a genome similarity score (GSS) analysis
387 because it is a metric that allows pairwise orthologs beyond the core genome, thus including
388 vertically and horizontally transmitted elements and pangenomic information in the analysis.
389 Briefly, the GSS used paired comparisons based on the sum of bit-scores of shared orthologs,
390 detected as reciprocal best hits (RBH), and normalized against the sum of bit-scores of the
391 compared genes against themselves (self-bit-scores). Its values range from 0 to 1, where 0
392 indicates the genomic content of compared species is the same, and a value of 1 indicates
393 the absence of homologous proteins. Then, the GSS can be plotted as a distance dendrogram,
394 where the maximum pairwise protein comparisons are plotted to the right of the graph (Fig.
395 4c). All *Solirubrobacter* strains had GSS values between 0.1896 - 0.3794, but ASM999324v1
396 presented an average value of 0.6615 when compared to the other strains of the genus (Table
397 S2).
398 Additionally, the core genome and GSS provide valuable gene function information and allow
399 us to identify shared functions and metabolic pathways among species, as well as provide

information about the ecological roles and adaptations of the species. The GSS dendrogram was, not surprisingly, consistent with the core genome phylogeny, showing the same clade arrangement and relationships for the *Solirubrobacter* (Fig 4c). However, the genome assembly ASM999324v1 reported as *Solirubrobacter* in NCBI, is again located in a clade with *B. subtilis*. The smaller genome of ASM999324v1, along with the phylogenetic inconsistencies, helped us to conclude that it is not a strain of *Solirubrobacter* and was excluded from the rest of the analysis.

Solirubrobacter reference genomes harbor an average of 7,423 protein-coding genes, with *S.* sp. URHD having the lowest number and *S. ginsenosidimutans* the highest (6,353 and 8,541, respectively; Fig. 4c). The average GC content was 69.81% and a total of 2,644 proteins were shared among all *Solirubrobacter* species, a strict core genome (Fig. 4d, Fig. S3).

Solirubrobacter diversity and interactions

Because of the large prevalence of *Solirubrobacter* in soil and rhizosphere microbiomes, we started delving into its diversity through sequence retrieval and 16S gene annotation to describe the environmental diversity. We used 107 samples from previous work involving sampling comprehensive Mexican soil diversity and its role in rhizosphere microbiome structuring [14]. *Solirubrobacter* environmental diversity is represented in the phylogeny constructed using the 3,166 *Solirubrobacter* OTUs identified in our samples; we also included the reference genomes showing that they map a few clades from the whole 16S phylogeny (Fig. 4e). *Solirubrobacter*'s diversity shows that a great effort is yet to be carried to capture the diversity of this bacterial genus.



421

422 **Fig. 4.** Phylogenomic diversity and shared genes of *Solirubrobacter*. (a) A 16S rRNA gene
423 Neighbour-Joining (NJ) phylogeny of the available genome references. Bootstrap values are
424 shown. (b) Core genome maximum likelihood phylogeny using 275 ortholog proteins and
425 1,000 bootstraps. (c) *Solirubrobacter* references dendrogram constructed using genomic
426 similarity score (GSS) distance (d) UpSet diagram showing *Solirubrobacter* shared genes and
427 pairwise intersections, in the first column, the core genome intersection is shown with 2,644
428 genes shared in all *Solirubrobacter* genomes. (e) *Solirubrobacter* reference genomes against
429 environmental diversity, a 16S NJ phylogeny of the operational taxonomic units (OTUs)
430 annotated as *Solirubrobacter* from Mexican soils and rhizospheres microbiomes, the
431 reference genomes 16S are shown in colored dots. In all the phylogenetic and GSS analysis,
432 *Streptomyces coelicolor* and *Bacillus subtilis* are used as outgroups.

433 *Solirubrobacter* interactions with other members of the microbial community were evaluated
434 through a network analysis using a 16S co-occurrence network (Fig. 5) from the 107 root-
435 associated and Mexican soil microbiomes. We got a network with 1,748 genera (nodes) and
436 35,034 edges, representing a large and complex amount of interactions among the bacteria.

The node degree centrality had an average of 40.08 and a range from 0 to 228. *Solirubobacter* had a degree centrality of 97, indicating a larger number of connections to other taxa. Network closeness centrality was approximately 0.020, with *Solirubrobacter* having a larger value of 0.376, suggesting a keystone role. However, *Solirubrobacter* was not a network bridge, as normalized betweenness centrality is low (0.003). To understand community interactions, *Solirubrobacter*'s neighbors according to co-occurrence and higher correlations ($r^2 > 0.8$) were filtered (Fig. 5b). The *Solirubrobacter* co-occurring genus are: *Gemmata, Longispora, Conexibacter, Planctopirus, Prausurella, Paludisphera, Thermobispora, Sterolibacterium, Thermoleophilum, Tundrisphera, Telmatocola,* and ca. *Nostocoida* (Fig 5b).

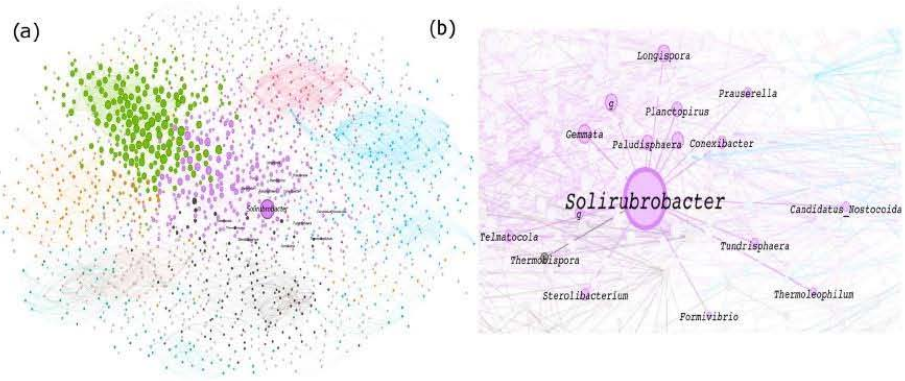


Fig. 5. Identifying ecological neighbors for *Solirubrobacter*. (a) Co-occurrence network based in 107 soil and rhizosphere 16S microbiomes highlighting *Solirubrobacter* node. Each node represents a bacterial genus, and each edge a correlation ($r^2 > 0.8$). Colors represent modularity class. (b) Close-up of the network highlighting the co-occurring genera with *Solirubrobacter*.

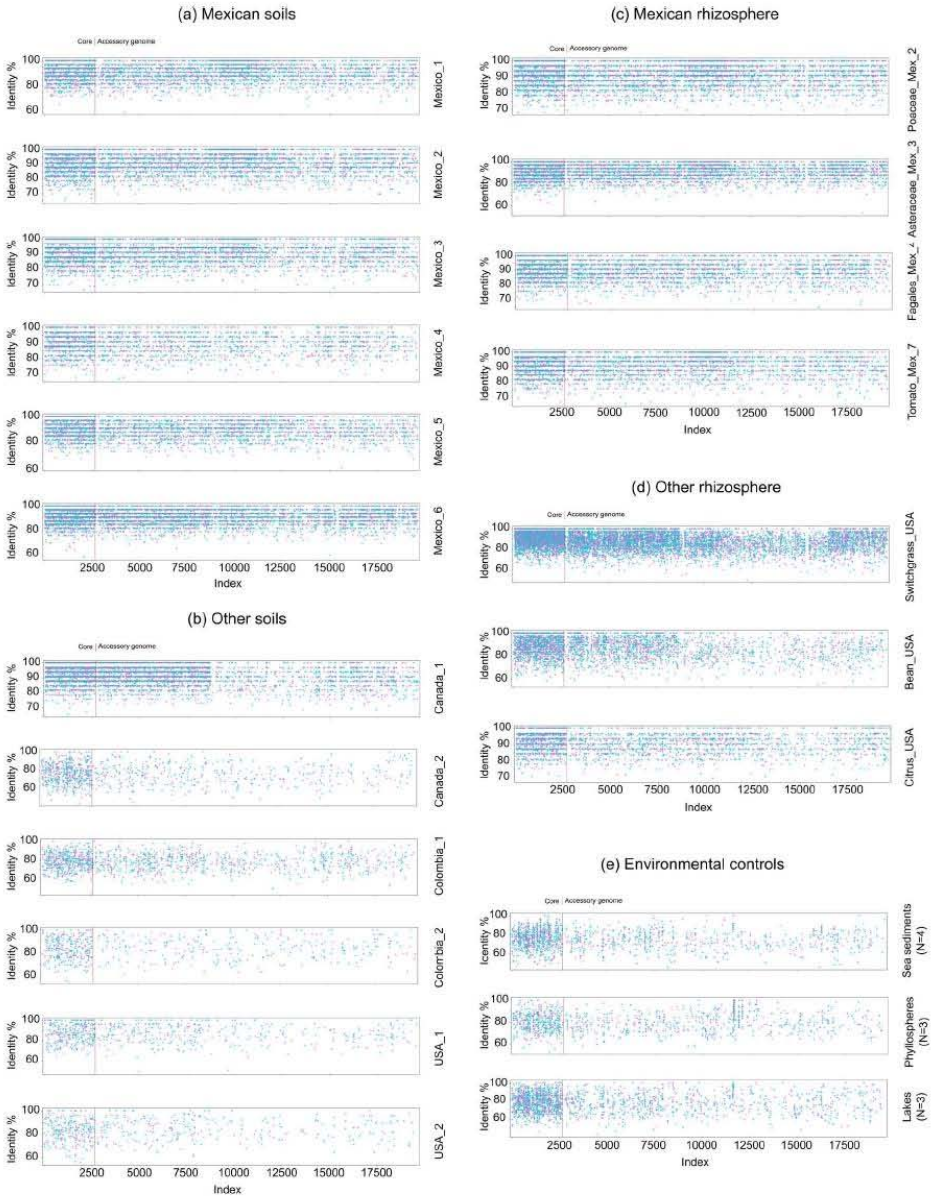
The pangenome as a tool to map *Solirubrobacter* into its naturally occurring niches

We calculated a *Solirubrobacter* pangenome from the genome references, which was composed of 19,645 family proteins, of which 2,644 of them were the genus core genome (Fig. 4d, Fig. S3). A total of 6,917 family proteins were the shell genome and 10,084 the cloud genome, adding up to an accessory genome of 17,001 family proteins. *Solirubrobacter* has an open pangenome, with each new genome adding unique genes (Fig. S4).

Further, we used *Solirubrobacter* pangenome as an anchor to recruit metagenomic reads to assess its environmental distribution in multiple environments. Metagenomic recruitments

showed the six-frame translated amino acid identity (AAI) percentage and coverage between *Solirubrobacter* pangenome to metagenomic reads from Mexican soils and rhizospheres (Fig. 6a and 6c, Table S5 and S6). The variation in AAI of translated genes ranges from ~70-100%, not only in the core genome sequences but throughout the entire pangenome, indicating they mainly belong to the *Solirubrobacter* genus and their ubiquitous presence in Mexican soils and rhizospheres.

The Mexican rhizosphere encompasses a dataset of tomato *(Solanum lycopersicum*) and wild plants belonging to multiple plant families (*Asteraceae, Fagales, Fabaceae, Lamiaceae*, and *Poaceae*) [14]. Contrary to what is observed for Mexican soils, the metagenomic recruitment with non-Mexican soils was scarce (Fig. 6a and 6b and Table S5 and S7). The percentage of identity recruitments in non-Mexican soils ranged between ~65 and 100% AAI, with a drastic 50% reduction in recruitments against the accessory genome (Fig. 6b and Table S7). Most of the recruited sequences were from Canadian soil metagenomes, whereas American and Colombian soils presented poor recruitment of *Solirubrobacter* sequences (Fig. 6b and Table S7). On the other hand, non-Mexican rhizospheres presented a higher recruitment and identity percentage throughout the entire pangenome, similar to the profile obtained for Mexican rhizospheres (Fig. 6c and 6d and Table S6 and S8). Nonetheless, rhizospheres from Mexican tomatoes and American citrus and beans presented lower sequence recruitment when compared to the rest of the rhizosphere samples (Fig. 6c and 6d). These recruitment differences suggested that the host genotype may influence the diversity and presence of *Solirubroabcter*.



481

482 **Fig. 6.** *Solirubrobacter* pangenomic recruitment from metagenomes. Coverage and average
483 amino acid identity (AAI) graphs of (a) Mexican soils, (b) Canadian, Colombian, and American

soils, (c) Mexican rhizospheres, (d) American rhizospheres, and (e) control environments. Except for the control environments, each recruitment graph corresponds to a single metagenome of the corresponding environment. Only a representative subset of four Mexican rhizosphere metagenomes with the highest sequencing depth is shown. The recruitment graph for all Mexican rhizospheres is found in Fig. S5. Blue and magenta dots indicate matching six-frame translated sequences in forward or reverse DNA strands. The core genome is to the left of the plots, delimited from the accessory genomes by a vertical brown bar in each graph.

Regarding lakes, phyllospheres, and sea sediments, the identity percentage ranged between 60 and 90%, with few reads of 100% AAI (Fig. 6e). In these three environments, the sequences recruited were lower than the sequences detected in soils and rhizospheres (Fig. 6). Hence, we can conclude that *Solirubrobacter* is a bacterium associated with rhizospheres and some soils.

Accessory genome sequences with a recruitment coverage ≥ 11 for soils and ≥ 18 for rhizospheres were considered high coverage. This coverage resulted in 13 sequences for Mexican soils (Table S5), including helix_turn_helix, Lux Regulon, MmpL family, and AAA domain of the putative AbiEii Type IV Toxin-Antitoxin system. MmpL proteins have been associated in *Mycobacterium* with translocating siderophores [73]. Concerning rhizospheres, the six high-coverage sequences (Table S6) were the F5/8 type C domain, FGGY family of carbohydrate kinases, histidine kinase HAMP region domain protein, peptidase S8 family involved in serine metabolism, pyridoxal-dependent decarboxylase conserved domain, and short-chain dehydrogenases reductases (SDR) family proteins.

A set of pangenome proteins resulted in gaps in the accessory genome recruitments (gap around the 15,500 index; Fig. 6). These 186 pangenome proteins absent in the metagenomes belong to *S.* sp. URHD, the basal group in both core and GSS phylogenies. Only 14 of these proteins are annotated; the rest are hypothetical proteins. These accessory proteins could be a specialized set of *S.* sp. URHD. Despite being the genome with fewer proteins (6,353; Fig. 4c), *S.* sp. URHD hosted a higher number of unique family proteins (2,124; Fig. 4d). So, it is suggested that *S.* sp. URHD may have undergone an overall genome size reduction, with particular gene expansions and acquisitions to adapt itself to a unique niche. On the contrary, the larger genome sizes of *S. ginsenosidimutans* or *S. taibaiensis* suggest generalist strategies.

517

518 ***Solirubrobacter* Environmental Extended Pangenome**

519 We calculated 29,330 protein families comprising the environmental extended pangenome
520 (EEP), meaning 9,906 potentially new environmental *Solirubrobacter* proteins were detected.
521 After COG assignments, 17,189 protein sequences were matched (58.60% of the EEP, Table
522 S9), with the remaining 12,141 as hypothetical proteins. The most abundant COGs were S
523 (unknown; 18.73%), K (transcription; 9.33%), NA(-) (6.74%), E (amino acids; 6.49%), C
524 (energy; 6.19%), T (transduction; 5.71%). We visualized the different COGs per sample as
525 well as the reference core genome and the EEP (Fig. S6). When evaluating the differences
526 among the annotated proteins from the references and our metagenomic annotated proteins,
527 the PCoA ordination revealed two main clusters: an upper cluster containing most of the
528 reference genomes and some metagenomic rhizospheres and a second cluster with almost
529 all soils and the rest of the rhizosphere metagenomes (Fig. S7).

530 A closer examination of the EPP revealed COGs that increased their abundance compared to
531 the reference pangenome, such as the 23 COGs contained in the lower clusters of the EEP
532 COGs heatmap (Fig. S6). These COGs contain different hydrolases, proteins from the CarA
533 family for carbapenem synthesis (COG EF; amino acids/nucleotide), ethanolamine utilization
534 proteins (COG CQ; energy/secondary metabolism), and phenylacetic acid degradation
535 proteins (COG CI; energy/lipids). Additionally, some COGs were mainly enriched in specific
536 metagenomes. Such is the case of Tomato_Mex_10 COGs MQ (cell wall/secondary
537 metabolism), FI (nucleotides/lipids), DJ (cell cycle/ribosomal), BDLTU (chromatin/cell
538 cycle/replication/transduction/intracellular trafficking), and CI (energy/lipids) (z = 3.39 for all of
539 them). The proteins contained in these COGs include methionine biosynthesis protein MetW,
540 ParE protein from the type II toxin-antitoxin system, phosphatidylinositol kinase, and
541 phenylacetic acid degradation proteins. Similarly, samples Mexico_6 and Tomato_Mex_8
542 contain permease for allantoin in COG FH (nucleotides/coenzymes) (z = 1.82), a plant-derived
543 metabolite that influences microbial community structure in the rhizosphere (Wang et al.,
544 2010). COGs CQ (energy/secondary metabolism) (z = 3.39) and EU (amino acids/intracellular
545 trafficking) (z = 3.39 for both COGs) from Fagales_Mex_4 contain carboxysome and
546 dienelactone hydrolase proteins, the latter being involved in the catabolism of chlorochatecol.

547

Solirubrobacter interactions with plants and other microorganisms

We observed a few protein families involved in plant interactions in the reference pangenome. From COG I (lipids), squalene/phytoene synthase was detected, as well as chorismate binding enzyme from COG HQ (coenzymes/secondary metabolism). Conversely, in the EEP, we found protein families involved in chorismate biosynthesis. Regarding motility, 20 protein families are flagellum-related, most found in COG N (motility). Phytase-related protein families were identified in *S. ginsenosidimutans, S. phytolaccae, S*. sp. CPCC, and *S.* sp. URHD. Phytase proteins were identified in *S. ginsenosidimutans, S.* sp. CPCC, and *S. soli*. *S. ginsenosidimutans*, while *S. soli* presented auxin-binding proteins. *S. phytolaccae* contains two unique sequences for phenazine biosynthesis (PhzF and phenazine biosynthesis protein A/B) while sharing three phenazine biosynthesis-like protein sequences with the rest of the species, except for *S.* sp. URHD. *S. phytolaccae* and *S. taibaiensis* contain farnesyl diphosphate synthase (COG MU; cell wall/intracellular trafficking), codified in the *ispA* gene, and involved in carotenoid biosynthesis. Mainly, *S. phytolaccae* contains a Hep/Hag repeat protein domain in COG UW (intracellular trafficking/extracellular structures) (z = 1.7638). The Hep/Hag repeat has been associated with bacteria root attachment [74, 75]. However, only 11 sequences from Mexican rhizosphere and soil metagenomes matched a previously reported gene set for plant-microorganism interaction (HMM) (Table S10) [64]. Contrastingly, *S. coelicolor* and *S. griseus* had 142 and 31 positive matches, respectively. The sequences that had a positive match with the HMM were all hypothetical. This matching pattern suggested that *Solirubrobacter* lacks most Actinobacteriota known proteins that foster plant interactions.

8. Discussion

Reference core genome summary of the genus *Solirubrobacter*

The analysis of all the available genomes revealed that strain ASM999324v1 genome did not belong to the genus *Solirubrobacter* (Fig. 4), contradicting the NCBI genome database. The remaining seven *Solirubrobacter* genomes allowed us to construct a 19,645 family protein pangenome. As previously mentioned, *Solirubrobacter* reference pangenome is open (Fig. S4), which is consistent with the fact that *Solirubrobacter* is found in open environments such as soils, in contrast with species with closed pangenomes, which tend to occupy isolated niches [76]. Since the analysis of multiple and independent isolates contributes to the

understanding of the global complexity of a bacterial species [76], a more accurate description of this genus can be achieved once the number of available genomes increases.

Solirubrobacter ubiquitousness can be explained by its genomic content (Fig. 2). A comprehensive carbohydrate metabolism translates as plasticity when thriving in different environments. Accordingly, COG G (carbohydrate) was among the most abundant in the pangenome and core genome, containing proteins to metabolize simple and complex carbohydrates. Moreover, ROK proteins and cupin domains contribute to *Solirubrobacter*'s metabolic capabilities. The former can act as sugar kinases or as transcriptional regulators involved in carbon and sugar metabolism [77], while the latter is involved in a wide variety of processes, such as acquisition of nutrients, synthesis of antibiotics, and catabolism of different organic compounds [78, 79].

An assortment of sigma factors present in a genome is a proxy of physiological and developmental characteristics, primarily since accessory sigma factors (such as ECF) usually activate the transcription of specific gene sets in response to environmental signals [80]. Selfsame, this bacterium possesses proteins involved in the Lux operon, which regulates the expression of different genes involved in virulence factor expression, exoenzyme secretion, biofilm formation, motility, oxidative stress response, cellulase synthesis, carbon metabolism, RNA processing, sugar transport, and vitamin and biosynthesis of secondary metabolites [81].

Some core proteins allow *Solirubrobacter* to resist different types of environmental stresses, such as oxidative, osmotic, or temperature shock. The presence of antibiotic, chromate, and mycotoxin transporters are defense mechanisms that reduce the intracellular concentration of these hazardous compounds. Glutamate decarboxylase enzyme is involved in the glutamate-dependent acid resistance (GDAR) system, a system described in different bacteria that confers resistance to acidity through scavenging protons by GadA/GadB during GABA formation to elevate internal pH [82, 83]. Finally, the presence of methyltransferases is worth mentioning since they are related to the catabolism of xenobiotic compounds [84]. In agreement with this, *Solirubrobacter* has been associated with degrading different hazardous compounds [85, 86].

The largest gene families are linked with environmental sensing, defense, and potential motility

Gene families arise from duplicated genes the organism retains because they result in advantageous capacities facing specific environments [87]. *Solirubrobacter* favors gene families involved in environmental sensing, such as the LuxR regulators, and defense, such as antimicrobial compounds (Fig. 3a).

Regarding defense mechanisms, the sequence identified as RNA-mediated transposition proteins was a DDE-type integrase/transposase/recombinase, codified in the *tnsB* gene in *E. coli*. This is a sequence-specific DNA-binding protein required for Tn7 transposition. Tn7 transposon inserts itself at high frequencies and codifies for antibiotic-resistance genes [88].

In Actinobacteriota, LuxR regulators are associated with different domains, and such combinations of domains result in modulation of the expression of different genes, including genes for the biosynthesis of secondary metabolites [81, 89]. Moreover, the specific combination of the LuxR+domain is dependent on the environmental origin of the bacterium, with a positive correlation between the number of LuxR proteins and the association with plants [89]. In Gram-negative bacteria, LuxR is involved in detecting environmental signals, while LuxI contributes to the production of signaling molecules; in Gram-positive bacteria, the presence of LuxR and absence of LuxI could be a technique used by Gram-positive bacteria to interpret Gram-negative quorum-sensing signals [90]. Besides conforming gene families, regulon Lux-related proteins were part of the high-coverage proteins identified in soils. The capacity to sense and interpret environmental stimuli is crucial in highly changing environments, such as soil, especially for a non-motile bacterium like *Solirubrobacter*. This environmental sensing is further sustained by the maintenance of gene families devoted to environmental sensing in the genome of *S.* sp. URHD (Fig. 3a). This URHD strain could be suffering a genome-reduction process.

Regardless of all *Solirubrobacter* isolates being described as non-motile, attention is drawn to the presence of proteins involved in flagellar biosynthesis and flagellin gene families. Despite all *Solirubrobacter* strains harboring flagellin-coding genes, the absence of the cytoplasmic membrane ring (FliF) is noteworthy. This structure plays a pivotal role not only in the flagellum but also in its homologous superstructure, the type III secretion system, and, consequently, the injectisome [91]. Furthermore, the protein MotB, integral to the proton channel and acting as the flagellar motor stator, is conspicuously absent. However, the presence of MotA, the other protein comprising the stator, suggests that all these flagellar genes are not associated with other types of secretion systems [92]. Essential proteins for flagellar biosynthesis vary depending on the species, and the presence of flagellar proteins in non-motile bacteria has been linked to the remains of transport systems [93, 94]. Altogether, these observations lead to the possibility that we may be witnessing vestiges of a flagellum that are no longer in use or that have recruited other proteins in the formation of a novel type of flagellum.

647 **_Solirubrobacter_ is a bacterium associated with soils at a regional scale and with**
648 **rhizospheres of specific plant hosts**

649 The diversity of *Solirubrobacter*, as revealed by the 16S gene (Fig. 4e), underscores the
650 uncharacterized *Solirubrobacter* species in soils and rhizospheres. This diversity extends to
651 the coding gene level, evidenced by the percentage identity variation of the recruitment graphs
652 for rhizosphere and soil metagenomes (Fig. 6). The quantity and distribution of metagenomic
653 reads recruited to any given genome indicates the abundance of closely related organisms
654 [95]. Conversely, when recruiting metagenomes from lakes, phyllosphere, and sediments, the
655 lower amount and sequence identity of recruited reads suggest a lesser presence of this
656 bacterium in these environments, used as negative environmental controls (Fig. 6e). However,
657 *Solirubrobacter* has been detected as a prevalent endophytic bacteria in the ginseng plant
658 (*Panax notoginseng*) and as one of the two most-abundant genera associated with
659 *Trachymyrmex septentrionalis* ants [96, 97], indicating other environments may also harbor
660 *Solirubrobacter* communities.
661 The presence of *Solirubrobacter* in soil appears to vary by region: it is found in all Mexican
662 soils, scarcely in Colombian and American soils, while Canadian soils contain
663 *Solirubrobacter's* core genome but less than 50% of the accessory genome (Fig. 6a and 6b).
664 This variation may indicate the presence of distant *Solirubrobacter* species or closely related
665 bacteria from another possibly unknown genus. As proposed, microbial species distribution is
666 also influenced by local conditions upon immigration to new habitats, much like
667 microorganisms [21, 22]. Therefore, even if *Solirubrobacter* bacteria could disperse from
668 Mexican to foreign soils, varying environmental factors would dictate the persistence or
669 disappearance. Further research is required to explore its presence in other soils and
670 environments, considering abiotic, pedogenic, and regional parameters.
671 Similarly to soil variations, plant roots appear to affect *Solirubrobacter* abundance and
672 unidentified strains of the genus. Interestingly, *Solirubrobacter* was more abundant and
673 diverse in wild plants' rhizospheres than in cultivated crops (Fig. 6c and 6d, and Fig. S5).
674 Although not as apparent as soil recruitments, wild plants recruited 9,565 sequences and only
675 7,910 from tomatoes (Table S6). The higher prevalence of *Solirubrobacter* in wild plants may
676 be explained by changes crops underwent along the domestication process. However, these
677 variations in recruitment may be due to other factors associated with the plant genotype and
678 should be further addressed.

679

Environmental Extended Pangenome: Hints for new *Solirubrobacter* species and new metabolic functions

The presence of new *Solirubrobacter* strains was analyzed by comparing the genomic content between reference species and environmental metagenomic samples. The PCoA ordination allowed us to evaluate the difference between the genomic content present in each sample (Fig. S7). The formation of two clusters indicates the presence of two different sets of *Solirubrobacter*: one containing all reference species and some environmental samples and the other one containing species similar between them but different from the references. Hence, it is fair to suggest that these metagenomic samples foster unknown *Solirubroabcter* strains. This idea is further sustained through the percentage identity graphs (Fig. 6) that show enough variation to suggest the presence of new species and through the functional profile of the EEP (Fig. S6), which bears proteins with new functions for the genus.

The construction of a *Solirubrobacter* EEP allowed us not only to identify proteins that potentially belong to this genus but also to suggest a genetic content related to local adaptations while simultaneously broadening our view regarding its metabolic potential. We detected in our metagenomic samples proteins involved in assimilatory sulfate reduction an incomplete route in the reference genomes. Assimilatory sulfate reduction reduces sulfate to sulfide, which can ultimately be used to synthesize cysteine [98]. Sulfur metabolism contributes to soil fertility and the bioavailability of nutrients such as P, Fe, K, and Zn for plants and other microorganisms [99-102]. *Solirubrobacter* from samples Mexico_1, Mexico_4, Asteraceae_Mex_3, and Tomato_Mex_9 seem to reduce sulfate (Fig. S8) partially. The enzyme CisIJ that carries out the last reduction step is present in Mexico_1 and Mexico_4, suggesting that even though *Solirubrobacter*-associated proteins cannot reduce sulfate ultimately, other microorganisms in the environment can carry out the final step, creating a syntrophic process.

***Solirubrobacter* ecological interactions**

In both the reference pangenome and the EEP, permeases for allantoin, chorismate, phytase, and a few other plant-microorganism interaction proteins were identified. Similarly, COG N (motility) was among the COGs with higher representation in the reference pangenome (z = 2.2670). The root-isolated *S. phytolaccae* has a high proportion of the Hep/Hag repeat protein (COG UW; intracellular trafficking/extracellular structures), whose domain has been associated with the binding activity of adhesins, invasins, and agglutinins in soil and plant-associated bacteria [74, 75]. However, *Solirubrobacter* presented ~30% or less of the plant-

714 microorganism interaction gene sequences [64] that *S. coelicolor* and *S. griseus* bear in their
715 genomes. Moreover, among the high-coverage proteins identified in rhizospheres are proteins
716 that indicate a possible indirect interaction with the plant host. FGGY kinases, a type of high-
717 coverage protein identified in rhizosphere metagenomes, are highly plastic proteins involved
718 in bacterial adaptation to exploit a plethora of carbohydrates using different metabolic
719 pathways [103]. Another high-coverage identified protein in rhizospheres is the pyridoxal-5'-
720 phosphate-dependent enzyme, corresponding to a glutamate decarboxylase. This enzyme is
721 responsible for the first step of GABA synthesis, a compound associated with interactions
722 between plants and bacteria [104].
723 Despite being associated with the rhizosphere of different plants, the total plant-
724 microorganism interaction genes were relatively low (Table S10) [64]. Hence, even though
725 *Solirubrobacter* has been detected in the rhizosphere of different plants, it appears not to
726 interact with the plant directly. Nevertheless, since we detected a tremendous amount of
727 percentage identity variance between the reference and the metagenomic-recruited proteins,
728 the genetic content of future isolates from rhizospheric environments should be thoroughly
729 analyzed.
730 Its proximity to the roots could derive from a need to interact with the rest of the bacterial
731 community. Given the absence of biosynthetic pathways for essential amino acids and the
732 lack of numerous enzymes involved in the nitrogen cycle, it is reasonable to surmise that
733 *Solirubrobacter* relies on other microorganisms to fulfill critical metabolic needs. Our network
734 analysis (Fig. 5) revealed that five out of the twelve co-occurring bacteria genera had reported
735 capabilities for nitrate reduction (*Gemmata, Longispora, Conexibacter, Prausurella,* and
736 *Sterolibacterium*) and six of them produce mycelia (*Gemmata, Longispora, Planctopirus,*
737 *Prausurella*, *Thermobispora,* and ca. *Nostocoida*) [105-112]. The mycelial structures offer a
738 protective environment for *Solirubrobacter* and other embedded microorganisms.
739 *Planctopirus*, known for its antibiotic production and attachment to eukaryotes, including plants
740 [110], may act as a conduit between *Solirubrobacter* and the plant host. Notably, the antibiotic-
741 producing and multidrug-resistant genus *Gemmata* [105, 113], previously linked to
742 *Solirubrobacter*, was found alongside *Solirubrobacter* as one of the most abundant genera
743 within the core rhizospheric microbiome of invasive buffelgrass (*Pennisetum ciliare*) in a prior
744 study [16]. This suggests a potentially tighter interaction necessary for root colonization and
745 recruitment of the broader microbial community. Moreover, given *Gemmata's* apparent lack of
746 conventional iron metabolic pathways [105], it may rely on other microorganisms for iron
747 acquisition —the co-occurrence of *Solirubrobacter* and *Gemmata* in a plant-interaction context
748 warrants further investigation.

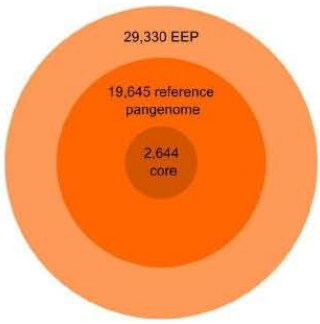


Fig. 7. Venn diagram representing protein families in the core genome, the reference pangenome, and the environmental extended pangenome (EEP).

Our study pioneers an approach beyond employing genomes assembled from metagenomes (MAGs). Initially, we carried out an inventory of genomic diversity based on available databases and characterized the pangenome. This approach expedited comparative genomics studies, aided by a reference. Further, guided by the pangenomic reference, we identified genes within metagenomic assemblies, expanding the pangenomic repertoire of *Solirubrobacter* and shedding light on the dynamics of the communities they inhabit. The methodology employed in this study facilitated a tri-faceted examination of *Solirubrobacter*: phylogenomics, pangenomics, and metagenomic analysis. These perspectives gave us a broader understanding of this bacterial genus, addressing taxonomic diversity and elucidating the plethora of functions inherent in *Solirubrobacter*. Each analysis augmented the number of gene functions associated with *Solirubrobacter* (Fig. 7), underlining the open nature of *Solirubrobacter*'s pangenome while concurrently indicating that specific environments compel their members to possess particular genetic content. Following the discourse "Everything is everywhere, but the environment selects" concerning *Solirubrobacter*, we probed into it on a vast geographical scale, examining North and South American soil and rhizosphere metagenomes. Based on its occurrence patterns, we propose that *Solirubrobacter* is present in every soil, with the rhizosphere being its preferred habitat. We also scrutinized these large-sized genomes, regarding them as generalists [114], aiming to elucidate the mechanisms underlying their successful distribution. *Solirubrobacter*, a non-sporulating Actinobacteriota in all reported isolates, showcases an advantageous trait for environmental endurance and distribution compared to other bacteria like *Bacillus* [115]. *Solirubrobacter* exists as vegetative

775 active cells in the environment, exhibiting drought resistance and compensating for the lack of
776 spore-related DNA damage protection through expanded repair systems like the *uvrABC* gene
777 families reported herein. Notably, despite being a prevalent bacterium, there are no reports of
778 swimming cells, and the pangenomic evidence hints at a deteriorated flagellum system or
779 possibly a novel variant; this warrants further investigation.

780 In conclusion, our analysis facilitated the first comparative genomics description of the
781 *Solirubrobacter* genus. The genetic repertoire of *Solirubrobacter* equips it with the capability
782 to metabolize various compounds and endure challenging environments, ranging from
783 defense against microbial adversaries to the degradation of hazardous compounds and
784 adaptation to other environmental abiotic alterations. Intriguingly, its ubiquity is governed not
785 only by its genetic arsenal but also by regional-scale factors. *Solirubrobacter* exhibits a more
786 pronounced prevalence in Mexican soils than in other nations. This pattern insinuates that
787 regional or historical elements may play a role in the dispersal and establishment of
788 *Solirubrobacter*. However, its presence in rhizospheres seems to transcend the geographical
789 origin of the sample, albeit being partially influenced by the host species. These results evoke
790 questions regarding whether bacteria from this genus interact with the plants or confine their
791 interactions to microorganisms associated with the roots. Through this investigation, we
792 augmented the known coding gene functions via the expanded pangenome. Additionally, the
793 pangenomic identity variations observed in pangenome recruitments against metagenomes
794 suggest that the detected *Solirubrobacter* strains possess distinct metabolic capacities,
795 warranting further characterization.

796

797 # 9. Author statements

798 ## 9.1 Author contributions

799 Conceptualization: Angélica Jara-Servín, Luis D. Alcaraz

800 Data curation: Angélica Jara-Servín, Gerardo Mejia, Miguel F. Romero, Luis D. Alcaraz

801 Formal analysis: Angélica Jara-Servín, Luis D. Alcaraz

802 Funding acquisition: Luis D. Alcaraz

803 Investigation: Angélica Jara-Servín, Luis D. Alcaraz

Methodology: Angélica Jara-Servín, Luis D. Alcaraz

Project Administration: Luis D. Alcaraz

Resources: Mariana Peimbert, Luis D. Alcaraz

Validation: Mariana Peimbert, Luis D. Alcaraz

Visualization: Angélica Jara-Servín, Gerardo Mejia, Luis D. Alcaraz

Writing - original draft: Angélica Jara-Servín, Luis D. Alcaraz

Writing - review & editing: Mariana Peimbert, Luis D. Alcaraz

9.2 Conflicts of interest

The authors declare that they do not have a conflict of interest.

9.3 Funding information

This work was supported by Universidad Nacional Autónoma de México by the project DGAPA-PAPIIT-UNAM IN206824 to LDA. Conachyt Ph.D. scholarship (CVU 765278) to AJS.

9.4 Ethical approval

Not required.

10.References

1. **Whitman WB**. *Solirubrobacter*. 2015. Epub ahead of print 2015. DOI: 10.1002/9781118960608.gbm00227. 10.1002/9781118960608.gbm00227.

2. **Singleton DR, Furlong MA, Peacock AD, White DC, Coleman DC, *et al.*** *Solirubrobacter pauli* gen. nov., sp. nov., a mesophilic bacterium within the

827 Rubrobacteridae related to common soil clones. *Int J Syst Evol Microbiol* 2003;53:485–
828 490. 10.1099/ijs.0.02438-0.

829 3. **Kim MK, Na JR, Lee TH, Im WT, Soung NK, *et al.*** *Solirubrobacter soli* sp. nov.,
830 isolated from soil of a ginseng field. *Int J Syst Evol Microbiol* 2007;57:1453–1455.
831 10.1099/ijs.0.64715-0.

832 4. **An DS, Wang L, Kim MS, Bae HM, Lee ST, *et al.*** *Solirubrobacter*
833 *ginsenosidimutans* sp. nov., isolated from soil of a ginseng field. *Int J Syst Evol*
834 *Microbiol* 2011;61:2606–2609. 10.1099/ijs.0.028431-0.

835 5. **Zhang L, Zhu L, Si M, Li C, Zhao L, *et al.*** *Solirubrobacter taibaiensis* sp. nov.,
836 isolated from a stem of Roxb. *Antonie Van Leeuwenhoek* 2014;106:279–285.
837 10.1007/s10482-014-0194-4.

838 6. **Seki T, Matsumoto A, Ōmura S, Takahashi Y**. Distribution and isolation of
839 strains belonging to the order Solirubrobacterales. *J Antibiot (Tokyo)* 2015;68:763–766.
840 10.1038/ja.2015.67.

841 7. **Malhotra J, Aparna K, Dua A, Sangwan N, Trimurtulu N, *et al.*** Microbial and
842 genetic ecology of tropical Vertisols under intensive chemical farming. *Environ Monit*
843 *Assess* 2015;187:4081. 10.1007/s10661-014-4081-2.

844 8. **Wang B, Wang Y, Cui X, Zhang Y, Yu Z**. Bioconversion of coal to methane by
845 microbial communities from soil and from an opencast mine in the Xilingol grassland of
846 northeast China. *Biotechnol Biofuels* 2019;12:236. 10.1186/s13068-019-1572-y.

847 9. **Hung C-M, Chen C-W, Huang C-P, Dong C-D**. Degradation of 4-nonylphenol in
848 marine sediments using calcium peroxide activated by water hyacinth (Eichhornia
849 crassipes)-derived biochar. *Environ Res* 2022;211:113076.
850 10.1016/j.envres.2022.113076.

851 10. **Wu Y, Santos SS, Vestergård M, Martín González AM, Ma L, *et al.*** A field
852 study reveals links between hyperaccumulating Sedum plants-associated bacterial
853 communities and Cd/Zn uptake and translocation. *Sci Total Environ* 2022;805:150400.
854 10.1016/j.scitotenv.2021.150400.

855 11. **Zhao L, Pan Z, Sun B, Sun Y, Weng L, *et al.*** Responses of soil microbial
856 communities to concentration gradients of antibiotic residues in typical greenhouse
857 vegetable soils. *Sci Total Environ* 2023;855:158587. 10.1016/j.scitotenv.2022.158587.

12. **Bañeras L, Llorens L, Díaz□Guerra L, Gispert M, Hernández□del Amo E, *et al.*** Resilience of microbial communities in Mediterranean soil after induced drought and manipulated <scp>UV</scp> radiation. *Eur J Soil Sci*;73. Epub ahead of print 16 January 2022. DOI: 10.1111/ejss.13218. 10.1111/ejss.13218.

13. **Sun Y, Shi Y-L, Wang H, Zhang T, Yu L-Y, *et al.*** Diversity of Bacteria and the Characteristics of Actinobacteria Community Structure in Badain Jaran Desert and Tengger Desert of China. *Front Microbiol* 2018;9:1–14. 10.3389/fmicb.2018.01068.

14. **Barajas HR, Martínez-Sánchez S, Romero MF, Álvarez CH, Servín-González L, *et al.*** Testing the Two-Step Model of Plant Root Microbiome Acquisition Under Multiple Plant Species and Soil Sources. *Front Microbiol* 2020;11:1–17. 10.3389/fmicb.2020.542742.

15. **Hernández-Álvarez C, García-Oliva F, Cruz-Ortega R, Romero MF, Barajas HR, *et al.*** Squash root microbiome transplants and metagenomic inspection for in situ arid adaptations. *Sci Total Environ* 2022;805:150136. 10.1016/j.scitotenv.2021.150136.

16. **Jara-Servin A, Silva A, Barajas H, Cruz-Ortega R, Tinoco-Ojanguren C, *et al.*** Root microbiome diversity and structure of the Sonoran desert buffelgrass (*Pennisetum ciliare* L.). *PLoS One* 2023;18:e0285978. 10.1371/journal.pone.0285978.

17. **Yang J, Ruegger PM, McKenry M V., Becker JO, Borneman J**. Correlations between Root-Associated Microorganisms and Peach Replant Disease Symptoms in a California Soil. *PLoS One* 2012;7:e46420. 10.1371/journal.pone.0046420.

18. **Franke-Whittle IH, Manici LM, Insam H, Stres B**. Rhizosphere bacteria and fungi associated with plant growth in soils of three replanted apple orchards. *Plant Soil* 2015;395:317–333. 10.1007/s11104-015-2562-x.

19. **Mukherjee S, Seshadri R, Varghese NJ, Eloe-Fadrosh EA, Meier-Kolthoff JP, *et al.*** 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol* 2017;35:676–683. 10.1038/nbt.3886.

20. **O'Malley MA**. The nineteenth century roots of 'everything is everywhere'. *Nat Rev Microbiol* 2007;5:647–651. 10.1038/nrmicro1711.

21. **Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, *et al.*** Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 2006;4:102–112. 10.1038/nrmicro1341.

22. **Lindström ES, Langenheder S**. Local and regional factors influencing bacterial community assembly. *Environ Microbiol Rep* 2012;4:1–9. 10.1111/j.1758-2229.2011.00257.x.

23. **Sánchez-Marañón M, Miralles I, Aguirre-Garrido JF, Anguita-Maeso M, Millán V, *et al.*** Changes in the soil bacterial community along a pedogenic gradient. *Sci Rep* 2017;7:14593. 10.1038/s41598-017-15133-x.

24. **Zhang B, Wu X, Tai X, Sun L, Wu M, *et al.*** Variation in Actinobacterial Community Composition and Potential Function in Different Soil Ecosystems Belonging to the Arid Heihe River Basin of Northwest China. *Front Microbiol* 2019;10:1–11. 10.3389/fmicb.2019.02209.

25. **Reno ML, Held NL, Fields CJ, Burke P V., Whitaker RJ**. Correction for Reno et al., Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc Natl Acad Sci* 2009;106:18873–18873. 10.1073/pnas.0911400106.

26. **Louca S**. The rates of global bacterial and archaeal dispersal. *ISME J* 2022;16:159–167. 10.1038/s41396-021-01069-8.

27. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW**. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–1055. 10.1101/gr.186072.114.

28. **Contreras-Moreira B, Vinuesa P**. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. *Appl Environ Microbiol* 2013;79:7696–7701. 10.1128/AEM.02411-13.

29. **Katoh K**. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30:3059–3066. 10.1093/nar/gkf436.

30. **R Core Team.** R: A Language and Environment for Statistical Computing [Internet]. 2013 Vienna, Austria: R Foundation for Statistical Computing. Available: http://www.r-project.org/

31. **Paradis E, Schliep K**. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;35:526–528. 10.1093/bioinformatics/bty633.

32. **Schliep KP**. phangorn: phylogenetic analysis in R. *Bioinformatics* 2011;27:592–593. 10.1093/bioinformatics/btq706.

33. **Le SQ, Gascuel O**. An Improved General Amino Acid Replacement Matrix. *Mol Biol Evol* 2008;25:1307–1320. 10.1093/molbev/msn067.

34. **Moreno-Hagelsieb G, Collado-Vides J**. A powerful non-homology method for the prediction of operons in prokaryotes. *Bioinformatics* 2002;18:S329–S336. 10.1093/bioinformatics/18.suppl_1.S329.

35. **Moreno-Hagelsieb G, Wang Z, Walsh S, ElSherbiny A**. Phylogenomic clustering for selecting non-redundant genomes for comparative genomics. *Bioinformatics* 2013;29:947–949. 10.1093/bioinformatics/btt064.

36. **Hernández-González IL, Moreno-Hagelsieb G, Olmedo-Álvarez G**. Environmentally-driven gene content convergence and the *Bacillus* phylogeny. *BMC Evol Biol* 2018;18:148. 10.1186/s12862-018-1261-7.

37. **McMurdie PJ, Holmes S**. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013;8:e61217. 10.1371/journal.pone.0061217.

38. **Oksanen J**. Vegan: ecological diversity. *R Package Version 2.4-4* 2017;1:11.

39. **Wickham MH**. Package 'ggplot2'.

40. **Seemann, T.** Barrnap 0.8: Rapid Ribosomal RNA Prediction. 2013. https://github.com/tseemann/barrnap.

41. **Nawrocki EP, Eddy SR**. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 2013;29:2933–2935. 10.1093/bioinformatics/btt509.

42. **Souza V, Moreno-Letelier A, Travisano M, Alcaraz LD, Olmedo G, *et al.*** The lost world of cuatro ciénegas basin, a relictual bacterial niche in a desert oasis. *Elife* 2018;7:1–17. 10.7554/eLife.38278.

43. **Csardi G, Nepusz T**. "The igraph software package for complex network research." *InterJournal, Complex Systems* 2006, 1695. https://igraph.org.

44. **Bastian M., Heymann S., Jacomy M.** Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media 2009.

45. **Seemann T**. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069. 10.1093/bioinformatics/btu153.

951 46. **Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, *et al.*** Roary: rapid large-
952 scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
953 10.1093/bioinformatics/btv421.

954 47. **Love MI, Huber W, Anders S**. Moderated estimation of fold change and
955 dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
956 10.1186/s13059-014-0550-8.

957 48. **Galperin MY, Makarova KS, Wolf YI, Koonin E V.** Expanded microbial genome
958 coverage and improved protein family annotation in the COG database. *Nucleic Acids*
959 *Res* 2015;43:D261–D269. 10.1093/nar/gku1223.

960 49. **Kanehisa M, Sato Y, Morishima K**. BlastKOALA and GhostKOALA: KEGG Tools for
961 Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*
962 2016;428:726–731. 10.1016/j.jmb.2015.11.006.

963 50. **Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley,**
964 **T., et al**. gplots: Various R Programming Tools for Plotting Data 2016. Available online
965 at: https://CRAN.R-project.org/package=gplots

966 51. **Bray JR, Curtis JT**. An Ordination of the Upland Forest Communities of Southern
967 Wisconsin. *Ecol Monogr* 1957;27:325–349. 10.2307/1942268.

968 52. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. Basic local alignment
969 search tool. *J Mol Biol* 1990;215:403–410. 10.1016/S0022-2836(05)80360-2.

970 53. **Bolger AM, Lohse M, Usadel B**. Trimmomatic: a flexible trimmer for Illumina
971 sequence data. *Bioinformatics* 2014;30:2114–2120. 10.1093/bioinformatics/btu170.

972 54. **Delcher AL**. Fast algorithms for large-scale genome alignment and comparison.
973 *Nucleic Acids Res* 2002;30:2478–2483. 10.1093/nar/30.11.2478.

974 55. **Langmead B, Salzberg SL**. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
975 2012;9:357–359. 10.1038/nmeth.1923.

976 56. **Nurk S, Meleshko D, Korobeynikov A, Pevzner PA**. metaSPAdes: a new
977 versatile metagenomic assembler. *Genome Res* 2017;27:824–834.
978 10.1101/gr.213959.116.

979 57. **Zerbino DR, Birney E**. Velvet: Algorithms for de novo short read assembly
980 using de Bruijn graphs. *Genome Res* 2008;18:821–829. 10.1101/gr.074492.107.

981 58. **Bushnell B.** "BBMap: A Fast, Accurate, Splice-Aware Aligner". United States 2014.
982 https://www.osti.gov/servlets/purl/1241166.

983 59. **Gish W, States DJ**. Identification of protein coding regions by database similarity
984 search. *Nat Genet* 1993;3:266–272. 10.1038/ng0393-266.

985 60. **Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, *et al.*** Prodigal:
986 prokaryotic gene recognition and translation initiation site identification. *BMC*
987 *Bioinformatics* 2010;11:119. 10.1186/1471-2105-11-119.

988 61. **Li W, Godzik A**. Cd-hit: a fast program for clustering and comparing large sets of
989 protein or nucleotide sequences. *Bioinformatics* 2006;22:1658–1659.
990 10.1093/bioinformatics/btl158.

991 62. **Gower JC**. Principal Coordinates Analysis. In: *Wiley StatsRef: Statistics Reference*
992 *Online*. Wiley. pp. 1–7. 10.1002/9781118445112.stat05670.pub2.

993 63. **Clarke KR**. Non-parametric multivariate analyses of changes in community
994 structure. *Austral Ecol* 1993;18:117–143. 10.1111/j.1442-9993.1993.tb00438.x.

995 64. **Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, *et***
996 ***al.*** Genomic features of bacterial adaptation to plants. *Nat Genet* 2018;50:138–
997 150. 10.1038/s41588-017-0012-9.

998 65. **Mistry J, Finn RD, Eddy SR, Bateman A, Punta M**. Challenges in homology
999 search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*
1000 2013;41:e121–e121. 10.1093/nar/gkt263.

1001 66. **Kim J-N, Kim Y, Jeong Y, Roe J-H, Kim B-G, *et al.*** Comparative Genomics
1002 Reveals the Core and Accessory Genomes of *Streptomyces* Species. *J Microbiol*
1003 *Biotechnol* 2015;25:1599–1605. 10.4014/jmb.1504.04008.

1004 67. **Lipa P, Janczarek M**. Phosphorylation systems in symbiotic nitrogen-fixing bacteria
1005 and their role in bacterial adaptation to various environmental stresses. *PeerJ*
1006 2020;8:e8466. 10.7717/peerj.8466.

1007 68. **Kaur P**. Expression and characterization of DrrA and DrrB proteins of
1008 *Streptomyces peucetius* in Escherichia coli: DrrA is an ATP binding protein. *J Bacteriol*
1009 1997;179:569–575. 10.1128/jb.179.3.569-575.1997.

69. **Malla S, Niraula NP, Liou K, Sohng JK**. Self-resistance mechanism in *Streptomyces peucetius*: Overexpression of *drrA*, *drrB* and *drrC* for doxorubicin enhancement. *Microbiol Res* 2010;165:259–267. 10.1016/j.micres.2009.04.002.

70. **Dy RL, Przybilski R, Semeijn K, Salmond GPC, Fineran PC**. A widespread bacteriophage abortive infection system functions through a Type IV toxin–antitoxin mechanism. *Nucleic Acids Res* 2014;42:4590–4605. 10.1093/nar/gkt1419.

71. **Kamimura N., Masai E.** The Protocatechuate 4,5-Cleavage Pathway: Overview and New Findings. In: Nojiri, H., Tsuda, M., Fukuda, M., Kamagata, Y. (eds) Biodegradative Bacteria 2014. Springer, Tokyo.

72. **Liu R, Ochman H**. Stepwise formation of the bacterial flagellar system. *Proc Natl Acad Sci* 2007;104:7116–7121. 10.1073/pnas.0700266104.

73. **Melly G, Purdy G**. MmpL Proteins in Physiology and Pathogenesis of *M. tuberculosis*. *Microorganisms* 2019;7:70. 10.3390/microorganisms7030070.

74. **Bishop AH, Rachwal PA**. Identification of Genes Required for Soil Survival in *Burkholderia thailandensis* by Transposon-Directed Insertion Site Sequencing. *Curr Microbiol* 2014;68:693–701. 10.1007/s00284-014-0526-7.

75. **Mhedbi-Hajri N, Darrasse A, Pigné S, Durand K, Fouteau S, *et al.*** Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evol Biol* 2011;11:67. 10.1186/1471-2148-11-67.

76. **Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, *et al.*** Erratum: Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial 'pan-genome' (Proceedings of the National Academy of Sciences of the United States of America (September 27, 2005) 102, 39 (13950-13955)). *Proc Natl Acad Sci U S A* 2005;102:16530. 10.1073/pnas.0508532102.

77. **Bekiesch P, Forchhammer K, Apel AK**. Characterization of DNA Binding Sites of RokB, a ROK-Family Regulator from *Streptomyces coelicolor* Reveals the RokB Regulon. *PLoS One* 2016;11:e0153249. 10.1371/journal.pone.0153249.

78. **Khuri S, Bakker FT, Dunwell JM**. Phylogeny, Function, and Evolution of the Cupins, a Structurally Conserved, Functionally Diverse Superfamily of Proteins. *Mol Biol Evol* 2001;18:593–605. 10.1093/oxfordjournals.molbev.a003840.

79. **Dunwell JM, Purvis A, Khuri S**. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 2004;65:7–17. 10.1016/j.phytochem.2003.08.016.

80. **Paget M, Helmann J**. Protein family review: The σ70 family of sigma factors. *Genome Biol* 2003;4:1–6. https://doi.org/10.1186/gb-2003-4-1-203.

81. **Li Z, Li X, Xia H**. Roles of LuxR-family regulators in the biosynthesis of secondary metabolites in Actinobacteria. *World J Microbiol Biotechnol* 2022;38:250. 10.1007/s11274-022-03414-1.

82. **Riordan JT, Mitra A**. Regulation of *Escherichia coli* Pathogenesis by Alternative Sigma Factor N. *EcoSal Plus*;7. Epub ahead of print 25 April 2017. DOI: 10.1128/ecosalplus.ESP-0016-2016. 10.1128/ecosalplus.ESP-0016-2016.

83. **Mallick S, Das S**. Acid-tolerant bacteria and prospects in industrial and environmental applications. *Appl Microbiol Biotechnol* 2023;107:3355–3374. 10.1007/s00253-023-12529-w.

84. **Lennard L**. Methyltransferases. In: *Comprehensive Toxicology*. Elsevier. pp. 435–457. 10.1016/B978-0-08-046884-6.00421-8.

85. **Martínez-Rabelo F, Gómez-Guzmán LA, García-Segura DR, Villegas-García E, Rodriguez-Campos J, *et al.*** Hydrocarbon bioremediation in a pilot-scale: A combination of bioaugmentation, phytoremediation, and vermiremediation. *Environ Technol Innov* 2023;31:103210. 10.1016/j.eti.2023.103210.

86. **Wang W, Gao Y, Du J, Zheng L, Kong X, *et al.*** Dose–effect of nitrogen regulation on the bioremediation of diesel contaminated soil. *Environ Technol Innov* 2023;32:103245. 10.1016/j.eti.2023.103245.

87. **Pushker R, Mira A, Rodríguez-Valera F**. Comparative genomics of gene-family size in closely related bacteria. *Genome Biol*;5. Epub ahead of print 2004. DOI: https://doi.org/10.1186/gb-2004-5-4-r27. https://doi.org/10.1186/gb-2004-5-4-r27.

88. **Craig NL**. Tn7: a target site-specific transposon. *Mol Microbiol* 1991;5:2569–2573. 10.1111/j.1365-2958.1991.tb01964.x.

89. **Santos CL, Correia-Neves M, Moradas-Ferreira P, Mendes MV**. A Walk into the LuxR Regulators of Actinobacteria: Phylogenomic Distribution and Functional Diversity. *PLoS One* 2012;7:e46758. 10.1371/journal.pone.0046758.

90. **Rajput A, Kumar M**. In silico analyses of conservational, functional and phylogenetic distribution of the LuxI and LuxR homologs in Gram-positive bacteria. *Sci Rep* 2017;7:6969. 10.1038/s41598-017-07241-5.

91. **Abby SS, Rocha EPC**. The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems. *PLoS Genet* 2012;8:e1002983. 10.1371/journal.pgen.1002983.

92. **Worrall LJ, Majewski DD, Strynadka NCJ**. Structural Insights into Type III Secretion Systems of the Bacterial Flagellum and Injectisome. *Annu Rev Microbiol* 2023;77:669–698. 10.1146/annurev-micro-032521-025503.

93. **Snyder LAS, Loman NJ, Fütterer K, Pallen MJ**. Bacterial flagellar diversity and evolution: seek simplicity and distrust it? *Trends Microbiol* 2009;17:1–5. 10.1016/j.tim.2008.10.002.

94. **Schuhmacher JS, Thormann KM, Bange G**. How bacteria maintain location and number of flagella? *FEMS Microbiol Rev* 2015;39:812–822. 10.1093/femsre/fuv034.

95. **Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, *et al.*** The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* 2007;5:e77. 10.1371/journal.pbio.0050077.

96. **Ishak HD, Miller JL, Sen R, Dowd SE, Meyer E, *et al.*** Microbiomes of ant castes implicate new microbial roles in the fungus-growing ant *Trachymyrmex septentrionalis*. *Sci Rep* 2011;1:204. 10.1038/srep00204.

97. **Dong L, Cheng R, Xiao L, Wei F, Wei G, *et al.*** Diversity and composition of bacterial endophytes among plant parts of *Panax notoginseng*. *Chin Med* 2018;13:41. 10.1186/s13020-018-0198-5.

98. **Simon J, Kroneck PMH**. Microbial Sulfite Respiration. In: *Advances in Microbial Physiology*. pp. 45–117. 10.1016/B978-0-12-410515-7.00002-0.

99. **Sahu A, Bhattacharjya S, Mandal A, Thakur JK, Atoliya N, *et al.*** Microbes: A Sustainable Approach for Enhancing Nutrient Availability in Agricultural Soils. In: *Role of Rhizospheric Microbes in Soil*. Singapore: Springer Singapore. pp. 47–75. 10.1007/978-981-13-0044-8_2.

100. **Chaudhary S, Sindhu SS, Dhanker R, Kumari A**. Microbes-mediated sulphur cycling in soil: Impact on soil fertility, crop production and environmental sustainability. *Microbiol Res* 2023;271:127340. 10.1016/j.micres.2023.127340.

101. **Gupta VVSR, Germida JJ**. Microbial transformations of sulfur in soil. In: *Principles and Applications of Soil Microbiology*. Elsevier. pp. 489–522. 10.1016/B978-0-12-820202-9.00018-6.

102. **Narayan OP, Kumar P, Yadav B, Dua M, Johri AK**. Sulfur nutrition and its role in plant growth and development. *Plant Signal Behav*;00. Epub ahead of print 7 February 2022. DOI: 10.1080/15592324.2022.2030082. 10.1080/15592324.2022.2030082.

103. **Zhang Y, Zagnitko O, Rodionova I, Osterman A, Godzik A**. The FGGY Carbohydrate Kinase Family: Insights into the Evolution of Functional Specificities. *PLoS Comput Biol* 2011;7:e1002318. 10.1371/journal.pcbi.1002318.

104. **Shelp BJ, Bown AW, Faure D**. Extracellular γ -Aminobutyrate Mediates Communication between Plants and Other Organisms. *Plant Physiol* 2006;142:1350–1352. 10.1104/pp.106.088955.

105. **Aghnatios R, Drancourt M**. Gemmata species: Planctomycetes of medical interest. *Future Microbiol* 2016;11:659–667. 10.2217/fmb-2015-0001.

106. **Schumann P**. *Conexibacter*. In: *Bergey's Manual of Systematics of Archaea and Bacteria*. Wiley. pp. 1–6. 10.1002/9781118960608.gbm00225.

107. **Matsumoto A, Takahashi Y, Shinose M, Seino A, Iwai Y, *et al.*** *Longispora albida* gen. nov., sp. nov., a novel genus of the family *Micromonosporaceae*. *Int J Syst Evol Microbiol* 2003;53:1553–1559. 10.1099/ijs.0.02595-0.

108. **Li J, Chen X, Li S, Zuo Z, Zhan R, *et al.*** Variations of rhizospheric soil microbial communities in response to continuous *Andrographis paniculata* cropping practices. *Bot Stud* 2020;61:18. 10.1186/s40529-020-00295-1.

109. **Tarlera S, Denner EBM**. Sterolibacterium denitrificans gen. nov., sp. nov., a novel cholesterol-oxidizing, denitrifying member of the β-Proteobacteria. *Int J Syst Evol Microbiol* 2003;53:1085–1091. 10.1099/ijs.0.02039-0.

110. **Kohn T, Wiegand S, Boedeker C, Rast P, Heuer A, *et al.*** *Planctopirus ephydatiae*, a novel Planctomycete isolated from a freshwater sponge. *Syst Appl Microbiol* 2020;43:126022. 10.1016/j.syapm.2019.126022.

111. **Liolios K., Sikorski J., Jando M. et al.** Complete genome sequence of *Thermobispora bispora* type strain (R51T). Stand in Genomic Sci 2, 2010; 318–326 https://doi.org/10.4056/sigs.962171

112. **Blackall LL, Seviour EM, Bradford D, Rossetti S, Tandoi V, *et al.*** 'Candidatus *Nostocoida limicola*', a filamentous bacterium from activated sludge. *Int J Syst Evol Microbiol* 2000;50:703–709. 10.1099/00207713-50-2-703.

113. **Fuerst J. A.; Lee K.-C., Butler M. K.** *Gemmata*. In Bergey's Manual of Systematics of Archaea and Bacteria; John Wiley & Sons, Ltd.: Chichester, U.K., 2015; pp 1– 5 https://doi.org/10.1002/9781118960608.gbm00783.pub2

114. **Bell TH, Bell T**. Many roads to bacterial generalism. *FEMS Microbiology Ecology*;97. Epub ahead of print 1 January 2021. DOI: 10.1093/femsec/fiaa240. 10.1093/femsec/fiaa240.

115. **Alcaraz LD, Moreno-Hagelsieb G, Eguiarte LE, Souza V, Herrera-Estrella L, *et al.*** Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* 2010;11:332. 10.1186/1471-2164-11-332.

All high-resolution figures can be downloaded from the following URL:

https://doi.org/10.6084/m9.figshare.24446521

9. DISCUSIÓN

El establecimiento de organismos en un ambiente específico es un proceso que se ve influenciado por interacciones con el ambiente y con otros organismos. En el caso del pasto buffel, otras características tales como su germinación, crecimiento, capacidad de rebrote, tamaño radicular y uso de la alelopatía eran ya conocidas (Hussain et al., 1982; Marshall et al., 2012; Lyons et al., 2013). En este trabajo reportamos la estructura y diversidad de su microbioma de raíz, una comunidad de microorganismos con la que interactúa el buffel y que no había sido previamente descrita.

El microbioma de raíz contribuye al establecimiento y desarrollo de plantas invasoras cuando éstas son introducidas a un nuevo ambiente (Trognitz et al., 2016; Kamutando et al., 2019). La selección y reclutamiento de microorganismos específicos por parte de la planta le permite a ésta establecer interacciones que le proporcionen ventajas ante el resto de organismos. En el microbioma núcleo del buffel identificamos a los géneros *Bradyrhizobium, Geodermatophilus, Pseudonocardia, Micromonospora, Crosiella, Amycolatopsis* y *Kribbela*, cepas de los cuales se han reportado con el potencial de sintetizar vitaminas, antimicrobianos y antibióticos (Wardle et al., 1994; Gao et al., 2014; Boumehira et al., 2016; Tanvir et al., 2016; Gonzalez-Pimentel et al., 2022; Gong et al., 2021). Los géneros *Sphingomonas* y *Nitrospira* son conocidos por promover el crecimiento de la planta y por reducir amonio a nitrato, respectivamente (Khan et al., 2014; Gornish et al., 2020). Dentro de las bacterias identificadas en el microbioma se encuentra el género *Saccharothrix*, previamente reportado como capaz de sintetizar compuestos fenólicos (Li et al., 2010), pudiendo reforzar el potencial alelopático del buffel. Tomando esto en consideración, estos microorganismos podrían contribuir al establecimiento del buffel, supliendo nutrientes, hormonas y defensa durante su desarrollo.

El patrón de exudados de raíz varía con el crecimiento de la planta, provocando cambios en las comunidades microbianas que se asocian a las raíces (O'Brien et al., 2018; Tkacz et al., 2020). Esto concuerda con nuestros resultados: identificamos ocho géneros significativamente sobrerrepresentados en los distintos tiempos evaluados. De estos cuatro, incrementos en la abundancia de *Pedobacter* y *Oceanibaculum* ya se han reportado en plantas en un estado intermedio de crecimiento (Wang et al., 2019). De manera similar, *Ohtaekwangia*, sobrerrepresentado en el segundo tiempo, incrementa su abundancia en plantas más viejas (Li et al., 2014).

La identificación tanto a nivel taxonómico como a nivel funcional depende de la comparación de las secuencias obtenidas contra bases de datos confiables (Robinson et al., 2021). Estas bases de datos requieren ser revisadas periódicamente con el fin de mantenerlas actualizadas. El genoma

ASM999324v1 se encontraba anotado como *Solirubrobacter* sp. y fue sometido recientemente (Probst et al., 2018) pero considerando los análisis filogenómicos de este trabajo resulta que el genoma no pertenece al género en cuestión. Es por esto que a pesar de los esfuerzos que se realizan para curar las bases de datos y mantenerlas actualizadas, es necesario corroborar, en la medida de lo posible, que las secuencias y asignaciones realizadas sean correctas para evitar ejecutar análisis con información errónea.

Dentro del microbioma del buffel se identificó al género *Solirubrobacter*, que se ha asociado a muchas especies distintas de plantas y a un rango taxonómico amplio de suelos (Barajas et al., 2020). A partir de nuestros resultados parece que la capacidad de desarrollarse en ambientes de características diversas está dada en gran medida por su contenido génico, que le permite metabolizar distintos carbohidratos, interpretar señales ambientales y resistir diferentes tipos de estrés ambiental. La facultad de utilizar múltiples carbohidratos le confiere a *Solirubrobacter* flexibilidad en cuanto a los ambientes en los que se puede desarrollar, ya que su metabolismo no está limitado a ciertos compuestos. Dentro del genoma núcleo de las referencias de *Solirubrobacter* se identificaron secuencias que además le proveen de mecanismos de defensa y tolerancia a estrés ambiental, tales como los genes *drrA* y *drrB* para resistencia a daunorrubicina, transportadores de micotoxinas y solutos compatibles y el operón *uvrABC* para reparación de DNA. En las mismas cepas de *Solirubrobacter* utilizadas en este estudio también se ha reportado resistencia a otros tipos de estrés ambiental, como resistencia a estrés osmótico, y la biosíntesis de algunos antimicrobianos, como acromicina A (Jiang et al., 2023). A través de una búsqueda manual de la anotación del pangenoma se identificaron las enzimas feruloil esterasa y la 4-cumarato—CoA ligasa, involucradas con el metabolismo de ácidos fenólicos como los que se han reportado en las raíces del pasto buffel (Hussain et al., 2011). Por otro lado, la presencia de familias génicas de hasta 22 genes relacionados con el regulón Lux sugiere que *Solirubrobacter* cuenta con distintos mecanismos de respuesta ante diversos estímulos ambientales. En el caso particular de las Actinobacteria el regulón Lux se ha vinculado a biosíntesis de metabolitos secundarios (Li et al., 2022). Sorprendentemente, junto con varias proteínas involucradas en biosíntesis del flagelo, se detectaron familias génicas grandes destinadas a la síntesis de flagelina, a pesar de que todos los aislados disponibles de este género se han reportado como no mótiles. La presencia de estos genes podrían ser remanentes de un flagelo que *Solirubrobacter* solía tener y ha ido perdiendo gradualmente conforme pasa el tiempo, aunque estas proteínas también podrían formar parte de otro tipo de estructura similar que aún no se determina. Un caso especial es el de *S. soli*, que parece tener la mayoría de las proteínas presentes en las Actinobacterias utilizadas como control. Es posible que durante su caracterización fisiológica no se haya evaluado su motilidad bajo las condiciones apropiadas para que *S. soli* sintetizara un

flagelo. Es probable que al incluir genomas nuevos en este análisis genómico comparativo se logre ahondar más en las posibles funciones de este conjunto de proteínas.

Los reclutamientos de metagenomas nos permitieron evaluar la diversidad ambiental de *Solirubrobacter*. En los metagenomas de rizósfera se recuperó prácticamente completo el pangenoma de referencia y las variaciones en porcentaje de identidad de las secuencias sugieren la presencia de nuevas cepas. Esto se mantuvo de manera constante en todas las rizósferas, mientras que en los suelos hubo variaciones dependiendo del origen geográfico del metagenoma, siendo mucho más prevalentes las cepas de *Solirubrobacter* en suelos mexicanos. Estas variaciones en los suelos podrían deberse a factores regionales abióticos, cobertura vegetal o incluso a factores históricos como los procesos de pedogénesis (Sánchez-Marañón et al., 2017). Contrariamente a lo que sucede en suelos, *Solirubrobacter* parece asociarse a las rizósferas de todas las plantas evaluadas, independientemente de la ubicación geográfica y de la especie de planta. La capacidad de asociarse a un amplio rango de plantas ya se ha reportado en otros taxa bacterianos (Fierer, 2017). A pesar de estar presente en todas las rizósferas analizadas parece haber un efecto del hospedero, ya que en algunas plantas como los pastos fue mayor el número de secuencias reclutadas en comparación con las rizósferas de tomate. Esto concuerda con reportes previos, en los que se observó que el genotipo de la planta puede jugar un papel más importante que el del suelo para la estructuración de comunidades microbianas (Lundberg et al., 2012). Incluso se han reportado diferencias en las comunidades de raíz dependiendo de si la planta hospedero es salvaje o es una planta domesticada (Barajas et al., 2020).

Curiosamente, la cantidad de genes involucrados con procesos de interacción planta-microorganismo identificados en el EEP, tanto de forma manual como a través de los HMM, parece indicar que *Solirubrobacter* no interactúa directamente con la planta a pesar de preferir estar inmersa en las comunidades rizosféricas. Sin embargo, se ha reportado la identificación de la vía de síntesis de ácido-3-indol acético en las mismas siete cepas de *Solirubrobacter* utilizadas en este estudio (Jiang et al., 2023), por lo que su potencial de interacción con plantas debe continuar evaluándose. Para evaluar su interacción con el resto de la comunidad microbiana realizamos un análisis de red de co-ocurrencia, donde *Solirubrobacter* resultó presentar un alto grado de conectividad (97) en comparación con el promedio de la red (40.08). Esto indica que *Solirubrobacter* es co-ocurrente con 97 géneros bacterianos presentes en la comunidad, sugiriendo que podría interpretar un papel central en la estructuración y funcionamiento de la comunidad. Uno de los géneros con los que *Solirubrobacter* co-ocurre de acuerdo a la red es *Gemmata*, microorganismo que también se encuentra conformando el microbioma núcleo del buffel. Si una especie se adapta a la presencia y cualidades de otra, las interacciones que se dan entre ellos llegarían a incentivar procesos de

co-evolución (Barraclough, 2015), por lo que evaluar las posibles interacciones entre *Gemmata* y *Solirubroabcter* podría considerarse para estudios futuros.

Finalmente, el bajo número de secuencias reclutadas de los metagenomas de sedimentos marinos, filósferas y lagos sugiere que *Solirubrobacter* es una bacteria asociada a suelos y rizósferas. No obstante, la presencia y abundancia de este género ha sido reportada en otros otros entornos, como el microbioma de las hormigas *Trachymyrmex septentrionalis* (Ishak et al., 2011), por lo que el rango de ambientes en los que puede desenvolverse *Solirubrobacter* debe explorarse con más detalle. Aquellos géneros presentes en ambientes heterogéneos resultan ser propensos a desarrollar fenotipos que se adapten más fácilmente a cambios ambientales (Büchi y Vuilleumier 2014), así como adquirir nuevas funciones por procesos de transferencia horizontal al estar en contacto con una mayor diversidad de microorganismos (Bell y Bell, 2020). Esto podría explicar en cierta medida la ubicuidad y distribución de *Solirubrobacter*, aunque es necesario recalcar que la caracterización fisiológica y la secuenciación de genomas de nuevas cepas obtenidas de ambientes contrastantes es necesaria para poder ampliar el potencial funcional de este género bacteriano.

Solirubrobacter no solo está presente en el microbioma del buffel, sino que forma parte del microbioma núcleo de este pasto. Todos los géneros bacterianos que se identifican como microbioma núcleo de un hospedero podrían estar adaptados a la planta, considerándose especialistas del ambiente generado por la misma (Barraclough, 2015; Bell y Bell, 2020). Sin embargo, *Solirubroabcter* no se ha reportado únicamente asociada a *Pennisetum ciliare*, sino que tiene la capacidad de asociarse con distintas plantas que van desde pastos como *Poaceae, Asteraceae y Fagales,* hasta plantas de cultivo, como tomate, frijol y cítricos. Todos los metagenomas de suelos mexicanos que se utilizaron para este estudio son de diferentes tipos de acuerdo a la clasificación taxonómica de suelos de la FAO (IUSS, 2015) y esto implica que las características fisicoquímicas de cada uno de los suelos sean contrastantes entre sí. Bajo este contexto e independientemente de su presencia en el microbioma núcleo del pasto buffel, *Solirubrobacter* podría considerarse generalista por su presencia en suelos y raíces con características y genotipos variables entre sí, además de contener el potencial genético para tolerar diferentes tipos de estrés ambiental y la capacidad de explotar distintos recursos energéticos. Es importante mencionar que estas características no impiden que *Solirubroabcter* tenga un contenido génico distinto en cada uno de los diversos ambientes en los que se desenvuelve. Esto se ve reflejado en las proteínas que se reclutaron de cada uno de los metagenomas y que no necesariamente están presentes en el resto de las muestras, indicando que el genoma accesorio de la cepa de *Solirubroabcter* presente en una muestra particular le permiten a esta cepa interactuar y

desarrollarse en las condiciones creadas por el hospedero o suelo al que pertenece la muestra de metagenoma.

10. CONCLUSIONES

A través de los análisis y resultados obtenidos en este trabajo se pudieron identificar interacciones y elementos genéticos que contribuyen al establecimiento y prevalencia de la planta *Pennisetum ciliare* y del género bacteriano *Solirubrobacter*.

Hay un efecto de selección del buffel sobre el microbioma de suelo que se refleja en la composición del microbioma de raíz que difiere de la composición del microbioma de suelo. El factor que más influyó en la estructuración del microbioma rizosférico fue el tiempo, relacionado con el desarrollo de la planta buffel. Acorde con esto, identificamos distintos géneros bacterianos más abundantes en cada uno de los periodos evaluados. Curiosamente, los aleloquímicos del buffel no influyeron significativamente en la estructuración del microbioma de raíz, a pesar de que logramos identificar microorganismos significativamente más abundantes en presencia de exudados y lixiviados del buffel. La estructura y diversidad del microbioma de raíz del pasto buffel contiene géneros bacterianos potencialmente capaces de sintetizar compuestos como vitaminas, compuestos promotores de crecimiento, antibióticos y antifúngicos que estimulan el desarrollo y la prevalencia del buffel en los distintos ambientes en los que se dispersa, permitiendo que se expanda de manera exitosa.

Solirubrobacter, un género bacteriano ubicuo en ambientes de rizósfera y en suelos, encierra un potencial metabólico amplio, tanto para uso de carbohidratos como para tolerar estrés ambiental. Las familias génicas más grandes identificadas en los genomas corresponden a funciones de transducción de señales, revelando que la capacidad de sensar e interpretar estímulos ambientales de particular importancia para este género, considerando que las cepas identificadas hasta el momento no tienen la capacidad de desplazarse mediante flagelos. Identificamos una gran diversidad ambiental de *Solirubrobacter*, tanto a nivel de cepas como a nivel funcional, evidenciando la gran cantidad de información sobre el género que permanece aún inexplorada. Su distribución geográfica parece estar influenciada por factores regionales, siendo más prevalente en suelos mexicanos que en otros suelos evaluados a escala continental. En cambio, su presencia en ambientes de rizósfera es ubicua y las variaciones detectables en cuanto a abundancia y diversidad parecen no depender de la ubicación geográfica sino del tipo de hospedero. Paradójicamente, *Solirubrobacter* parece no interactuar directamente con la planta sino con los microorganismos que forman parte del microbioma de raíz.

A través de la construcción de un pangenoma ambiental extendido identificamos funciones presentes en condiciones ambientales particulares. Es decir, que su presencia en suelos y raíces,

independientemente de la distancia geográfica, conlleva a que su contenido genético esté adaptado a las características particulares de cada nicho que habita. Sin embargo, considerando su amplia distribución geográfica y su prevalencia en rizósferas y en suelos contrastantes entre sí consideramos que *Solirubroabcter* es un microorganismo generalista. Los esfuerzos futuros por incrementar el número disponible de genomas permitirán ahondar más en la descripción del género y el rol que juega en los ambientes que habita.

Conocer los mecanismos y características que un organismo requiere para dispersarse, sobrevivir y reproducirse en un ambiente determinado contribuye a nuestra comprensión del funcionamiento de los ecosistemas. El papel que desempeña el microbioma de raíz con respecto a las plantas invasoras nos permitiría proponer estrategias de control que reduzcan las ventajas competitivas que tienen este tipo de plantas. De manera similar, conocer el contenido génico de los microorganismos nos ayuda a inferir el tipo de ambientes en el que cada microorganismo se desenvuelve de manera exitosa. El estudio de estas interacciones ecológicas podría finalmente contribuir a la comprensión de los procesos microbianos que se llevan a cabo en el suelo, así como al establecimiento y desarrollo de plantas.

11. PERSPECTIVAS

Este trabajo logró hacer una aportación sólida para solventar la falta de información acerca del microbioma de raíz de *Pennisetum ciliare* y del contenido genético y de las características de *Solirubrobacter* como género bacteriano. Sin embargo, aún existen numerosas áreas de estudio pendientes que requieren investigación adicional.

La caracterización de los aleloquímicos que sintetiza y secreta el buffel a través de la raíz, así como aquellos que acumula en la parte aérea y que se lixivian con la lluvia, no se ha evaluado a detalle, por lo que sería necesario identificar su naturaleza química y su concentración para poder determinar de manera más específica el efecto de dichos compuestos sobre el microbioma rizosférico. De igual modo, la secuenciación de un metagenoma total del microbioma del pasto buffel contribuirá a describir las funciones contenidas en los microorganismos que se asocian al buffel, expandiendo así la información respecto a las interacciones plantas microorganismo que en este caso se mantuvieron limitadas a inferencias metabólicas a partir de estudios de cepas de los géneros identificados.

Con respecto a *Solirubrobacter* también hay múltiples aspectos que demandan un análisis más exhaustivo. La ubicuidad de *Solirubrobacter* y el escaso número de aislados y genomas disponibles en las bases de datos reflejan la necesidad de aislar cepas de diferentes regiones geográficas. La caracterización fisiológica de estas cepas y análisis genómicos podrá contribuir a identificar cepas nuevas que nos permitan entender mejor el rol que juega esta bacteria en los ambientes en los que se desarrolla. Asimismo, estudiar la relación entre el contenido génico de *Solirubrobacter* y las características fisicoquímicas de suelos mostrará de manera más íntegra las adaptaciones al ambiente de esta bacteria. Los metagenomas de suelos mexicanos fueron analizados fisicoquímicamente al momento de la toma de muestra, por lo que el análisis que se propone podría iniciarse. Determinar la abundancia de elementos móviles podría reflejar la capacidad de adaptación a distintos ambientes, así como su potencial para interactuar con otros microorganismos. De igual manera, la construcción de genomas de *Solirubrobacter* a partir de los metagenomas con suficiente profundidad de secuenciación podría brindar información nueva respecto al potencial genético. Finalmente, tomando en cuenta que *Solirubrobacter* se ha identificado asociada a insectos, consideramos que sería conveniente evaluar su presencia en ambientes distintos a los ya examinados en este trabajo.

12. REFERENCIAS

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

An, D.-S., Wang, L., Kim, M. S., Bae, H.-M., Lee, S.-T., & Im, W.-T. (2011). *Solirubrobacter ginsenosidimutans* sp. nov., isolated from soil of a ginseng field. *International Journal of Systematic and Evolutionary Microbiology*, *61*(11), 2606–2609. https://doi.org/10.1099/ijs.0.028431-0

Arriaga, L., Castellanos V., A. E., Moreno, E., & Alarcón, J. (2004). Potential Ecological Distribution of Alien Invasive Species and Risk Assessment: a Case Study of Buffel Grass in Arid Regions of Mexico. *Conservation Biology*, *18*(6), 1504–1514. https://doi.org/10.1111/j.1523-1739.2004.00166.x

Barajas, H. R., Martínez-Sánchez, S., Romero, M. F., Álvarez, C. H., Servín-González, L., Peimbert, M., Cruz-Ortega, R., García-Oliva, F., & Alcaraz, L. D. (2020). Testing the Two-Step Model of Plant Root Microbiome Acquisition Under Multiple Plant Species and Soil Sources. *Frontiers in Microbiology*, *11*(October), 1–17. https://doi.org/10.3389/fmicb.2020.542742

Barraclough, T. G. (2015). How Do Species Interactions Affect Evolutionary Dynamics Across Whole Communities? *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 25–48. https://doi.org/10.1146/annurev-ecolsys-112414-054030

Bastian M., Heymann S., Jacomy M. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media 2009.

Bekiesch P, Forchhammer K, Apel AK. Characterization of DNA Binding Sites of RokB, a ROK-Family Regulator from *Streptomyces coelicolor* Reveals the RokB Regulon. PLoS One 2016;11:e0153249. 10.1371/journal.pone.0153249.

Bell, T. H., & Bell, T. (2021). Many roads to bacterial generalism. In *FEMS Microbiology Ecology* (Vol. 97, Issue 1). Oxford University Press. https://doi.org/10.1093/femsec/fiaa240

Bishop, A. H., & Rachwal, P. A. (2014). Identification of Genes Required for Soil Survival in *Burkholderia thailandensis* by Transposon-Directed Insertion Site Sequencing. *Current Microbiology*, *68*(6), 693–701. https://doi.org/10.1007/s00284-014-0526-7

Bissett, A., Richardson, A. E., Baker, G., Wakelin, S., & Thrall, P. H. (2010). Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Molecular Ecology*, *19*(19), 4315–4327. https://doi.org/10.1111/j.1365-294X.2010.04804.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boumehira, A. Z., El-Enshasy, H. A., Hacène, H., Elsayed, E. A., Aziz, R., & Park, E. Y. (2016). Recent progress on the development of antibiotics from the genus *Micromonospora*. In *Biotechnology and Bioprocess Engineering* (Vol. 21, Issue 2, pp. 199–223). Korean Society for Biotechnology and Bioengineering. https://doi.org/10.1007/s12257-015-0574-2

Braga, R. M., Dourado, M. N., & Araújo, W. L. (2016). Microbial interactions: ecology in a molecular perspective. *Brazilian Journal of Microbiology*, *47*, 86–98. https://doi.org/10.1016/j.bjm.2016.10.005

Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349. https://doi.org/10.2307/1942268

Brewer, T. E., Handley, K. M., Carini, P., Gilbert, J. A., & Fierer, N. (2016). Genome reduction in an abundant and ubiquitous soil bacterium 'Candidatus *Udaeobacter copiosus*.' *Nature Microbiology*, *2*(2), 16198. https://doi.org/10.1038/nmicrobiol.2016.198

Büchi, L., & Vuilleumier, S. (2014). Coexistence of Specialist and Generalist Species Is Shaped by Dispersal and Environmental Factors. *The American Naturalist*, *183*(5), 612–624. https://doi.org/10.1086/675756

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

Bushnell B. "BBMap: A Fast, Accurate, Splice-Aware Aligner". United States 2014. https://www.osti.gov/servlets/purl/1241166.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecology*, *18*(1), 117–143. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x

Contreras-Moreira, B., & Vinuesa, P. (2013). GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. *Applied and Environmental Microbiology*, *79*(24), 7696–7701. https://doi.org/10.1128/AEM.02411-13

Cooper, V. S., & Lenski, R. E. (2000). The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature*, *407*(6805), 736–739. https://doi.org/10.1038/35037572

Csardi G, Nepusz T. "The igraph software package for complex network research."InterJournal, Complex Systems 2006, 1695. https://igraph.org.

Custer, G. F., Bresciani, L., & Dini-Andreote, F. (2022). Ecological and Evolutionary Implications of Microbial Dispersal. *Frontiers in Microbiology*, *13*(April), 1–13. https://doi.org/10.3389/fmicb.2022.855859

Dastogeer, K. M. G., Tumpa, F. H., Sultana, A., Akter, M. A., & Chakraborty, A. (2020). Plant microbiome–an account of the factors that shape community composition and diversity. *Current Plant Biology*, *23*, 100161. https://doi.org/10.1016/j.cpb.2020.100161

Delcher, A. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*, *30*(11), 2478–2483. https://doi.org/10.1093/nar/30.11.2478

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K., & Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science*, *359*(6373), 320–325. https://doi.org/10.1126/science.aap9516

Donachie, S. P., Fraser, C. J., Hill, E. C., & Butler, M. A. (2021). The Problem with 'Microbiome.' *Diversity*, *13*(4), 138. https://doi.org/10.3390/d13040138

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, *15*(10), 579–590. https://doi.org/10.1038/nrmicro.2017.87

Fitzpatrick, C. R., Salas-González, I., Conway, J. M., Finkel, O. M., Gilbert, S., Russ, D., Teixeira, P. J. P. L., & Dangl, J. L. (2020). The Plant Microbiome: From Ecology to Reductionism and Beyond. *Annual Review of Microbiology*, *74*(1), 81–100. https://doi.org/10.1146/annurev-micro-022620-014327

Fondi, M., Karkman, A., Tamminen, M. v., Bosi, E., Virta, M., Fani, R., Alm, E., & McInerney, J. O. (2016). "Every gene is everywhere but the environment selects": Global geolocalization of gene sharing in environmental samples through network analysis. *Genome Biology and Evolution*, *8*(5), 1388–1400. https://doi.org/10.1093/gbe/evw077

Fonseca, E. da S., Peixoto, R. S., Rosado, A. S., Balieiro, F. de C., Tiedje, J. M., & Rachid, C. T. C. da C. (2018). The Microbiome of Eucalyptus Roots under Different Management Conditions and Its Potential for Biological Nitrogen Fixation. *Microbial Ecology*, *75*(1), 183–191. https://doi.org/10.1007/s00248-017-1014-y

Franke-Whittle IH, Manici LM, Insam H, Stres B. Rhizosphere bacteria and fungi associated with plant growth in soils of three replanted apple orchards. *Plant Soil* 2015;395:317–333. 10.1007/s11104-015-2562-x

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. v. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*, *43*(D1), D261–D269. https://doi.org/10.1093/nar/gku1223

Gao, R., Liu, C., Zhao, J., Jia, F., Yu, C., Yang, L., Wang, X., & Xiang, W. (2014). *Micromonospora jinlongensis* sp nov, isolated from muddy soil in China and emended description of the genus

Micromonospora. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *105*(2), 307–315. https://doi.org/10.1007/s10482-013-0074-3

Gish, W., & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nature Genetics*, *3*(3), 266–272. https://doi.org/10.1038/ng0393-266

Gong, X., Shi, J., Zhou, X., Yuan, T., Gao, D., & Wu, F. (2021). Crop Rotation With Cress Increases Cucumber Yields by Regulating the Composition of the Rhizosphere Soil Microbial Community. *Frontiers in Microbiology*, *12*(March). https://doi.org/10.3389/fmicb.2021.631882

Gonzalez-Pimentel, J. L., Dominguez-Moñino, I., Jurado, V., Laiz, L., Caldeira, A. T., & Saiz-Jimenez, C. (2022). The Rare Actinobacterium *Crossiella* sp. Is a Potential Source of New Bioactive Compounds with Activity against Bacteria and Fungi. *Microorganisms*, *10*(8), 1575. https://doi.org/10.3390/microorganisms10081575

Gornish, E. S., Franklin, K., Rowe, J., & Barberán, A. (2020). Buffelgrass invasion and glyphosate effects on desert soil microbiome communities. *Biological Invasions*, *22*(8), 2587–2597. https://doi.org/10.1007/s10530-020-02268-8

Gower, J. C. (2015). Principal Coordinates Analysis. In *Wiley StatsRef: Statistics Reference Online* (pp. 1–7). Wiley. https://doi.org/10.1002/9781118445112.stat05670.pub2

Hamonts, K., Trivedi, P., Garg, A., Janitz, C., Grinyer, J., Holford, P., Botha, F. C., Anderson, I. C., & Singh, B. K. (2018). Field study reveals core plant microbiota and relative importance of their drivers. *Environmental Microbiology*, *20*(1), 124–140. https://doi.org/10.1111/1462-2920.14031

Hernández-Álvarez C, García-Oliva F, Cruz-Ortega R, Romero MF, Barajas HR, et al. Squash root microbiome transplants and metagenomic inspection for in situ arid adaptations. *Sci Total Environ* 2022;805:150136. 87110.1016/j.scitotenv.2021.150136.

Hussain, F., Naqvi, H. H., & Ilahi, I. (1982). Interference Exhibited by *Cenchrus ciliaris* L. and *Bothriochloa pertusa* (L.) A. Camus. *Bulletin of the Torrey Botanical Club*, *109*(4), 513. https://doi.org/10.2307/2996492

Hussain, F., Ilahi, I., Malik, S. A., Dasti, A. A., & Ahmad, B. (2011). Allelopathic effects of rain leachates and root exudates of *Cenchrus ciliaris* L. And *Bothriochloa pertusa* (L.) A. Camus. *Pakistan Journal of Botany*, *43*(1), 341–350.

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Ishak, H. D., Miller, J. L., Sen, R., Dowd, S. E., Meyer, E., & Mueller, U. G. (2011). Microbiomes of ant castes implicate new microbial roles in the fungus-growing ant *Trachymyrmex septentrionalis*. *Scientific Reports*, *1*(1), 204. https://doi.org/10.1038/srep00204

Jara-Servin, A., Silva, A., Barajas, H., Cruz-Ortega, R., Tinoco-Ojanguren, C., & Alcaraz, L. D. (2023 a). Root microbiome diversity and structure of the Sonoran desert buffelgrass (*Pennisetum ciliare* L.). *PLOS ONE*, *18*(5), e0285978. https://doi.org/10.1371/journal.pone.0285978

Jara-Servin, A., Mejia, G., Romero, M. F., Peimbert, M. & Alcaraz, L. D. (2023 b) Unraveling the Genomic and Environmental Diversity of the Ubiquitous *Solirubrobacter* (n.d.) [Under review]. *bioarXiv* https://doi.org/https://doi.org/10.1101/2023.10.31.564804

Jiang, Z., Mou, T., Sun, Y., Su, J., Yu, L., & Zhang, Y.-Q. (2023). Environmental distribution and genomic characteristics of *Solirubrobacter*, with proposal of two novel species. *Frontiers in Microbiology*, *14*(December). https://doi.org/10.3389/fmicb.2023.1267771

Jousset, A., Scheu, S., & Bonkowski, M. (2008). Secondary metabolite production facilitates establishment of rhizobacteria by reducing both protozoan predation and the competitive effects of indigenous bacteria. *Functional Ecology*, *22*(4), 714–719. https://doi.org/10.1111/j.1365-2435.2008.01411.x

Kamutando, C. N., Vikram, S., Kamgan-Nkuekam, G., Makhalanyane, T. P., Greve, M., le Roux, J. J., Richardson, D. M., Cowan, D. A., & Valverde, A. (2019). The Functional Potential of the Rhizospheric Microbiome of an Invasive Tree Species, *Acacia dealbata*. *Microbial Ecology*, *77*(1), 191–200. https://doi.org/10.1007/s00248-018-1214-0

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Khan, A. L., Waqas, M., Kang, S. M., Al-Harrasi, A., Hussain, J., Al-Rawahi, A., Al-Khiziri, S., Ullah, I., Ali, L., Jung, H. Y., & Lee, I. J. (2014). Bacterial endophyte *Sphingomonas* sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. *Journal of Microbiology*, *52*(8), 689–695. https://doi.org/10.1007/s12275-014-4002-7

Khuri S, Bakker FT, Dunwell JM. Phylogeny, Function, and Evolution of the Cupins, a Structurally Conserved, Functionally Diverse Superfamily of Proteins. Mol Biol Evol 2001;18:593–605. 10.1093/oxfordjournals.molbev.a003840.

Kim, M. K., Na, J. R., Lee, T. H., Im, W. T., Soung, N. K., & Yang, D. C. (2007). *Solirubrobacter soli* sp. nov., isolated from soil of a ginseng field. *International Journal of Systematic and Evolutionary Microbiology*, *57*(7), 1453–1455. https://doi.org/10.1099/ijs.0.64715-0

Klock, M. M., Barrett, L. G., Thrall, P. H., & Harms, K. E. (2015). Host promiscuity in symbiont associations can influence exotic legume establishment and colonization of novel ranges. *Diversity and Distributions*, *21*(10), 1193–1203. https://doi.org/10.1111/ddi.12363

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.-I., McDonald, D., Melnik, A. v., Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson, L. R., Tripathi, A., Xu, Z. Z., … Dorrestein, P. C. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, *16*(7), 410–422. https://doi.org/10.1038/s41579-018-0029-9

Knights, H. E., Jorrin, B., Haskett, T. L., & Poole, P. S. (2021). Deciphering bacterial mechanisms of root colonization. *Environmental Microbiology Reports*, *13*(4), 428–444. https://doi.org/10.1111/1758-2229.12934

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Le, S. Q., & Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, *25*(7), 1307–1320. https://doi.org/10.1093/molbev/msn067

Levy, A., Salas Gonzalez, I., Mittelviefhaus, 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

Li, X. G., Ding, C. F., Hua, K., Zhang, T. L., Zhang, Y. N., Zhao, L., Yang, Y. R., Liu, J. G., & Wang, X. X. (2014). Soil sickness of peanuts is attributable to modifications in soil microbes induced by peanut root exudates rather than to direct allelopathy. *Soil Biology and Biochemistry*, *78*, 149–159. https://doi.org/10.1016/j.soilbio.2014.07.019

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li Z, Li X, Xia H. Roles of LuxR-family regulators in the biosynthesis of secondary metabolites in Actinobacteria. World J Microbiol Biotechnol 2022;38:250. 10.1007/s11274-022-03414-1.

Li, Z.-H., Wang, Q., Ruan, X., Pan, C.-D., & Jiang, D.-A. (2010). Phenolics and Plant Allelopathy. *Molecules*, *15*(12), 8933–8952. https://doi.org/10.3390/molecules15128933

Lindström, E. S., & Langenheder, S. (2012). Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports*, *4*(1), 1–9. https://doi.org/10.1111/j.1758-2229.2011.00257.x

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., Brady, A., Creasy, H. H., McCracken, C., Giglio, M. G., McDonald, D., Franzosa, E. A., Knight, R., White, O., & Huttenhower, C. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, *550*(7674), 61–66. https://doi.org/10.1038/nature23889

Loman, N. J., & Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nature Reviews Microbiology*, *13*(12), 787–794. https://doi.org/10.1038/nrmicro3565

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lozupone, C., & Knight, R. (2005). UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, 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

Lyons, K. G., Maldonado-Leal, B. G., & Owen, G. (2013). Community and Ecosystem Effects of Buffelgrass (*Pennisetum ciliare*) and Nitrogen Deposition in the Sonoran Desert. *Invasive Plant Science and Management*, *6*(1), 65–78. https://doi.org/10.1614/IPSM-D-11-00071.1

Malik, A. A., Thomson, B. C., Whiteley, A. S., Bailey, M., & Griffiths, R. I. (2017). Bacterial Physiological Adaptations to Contrasting Edaphic Conditions Identified Using Landscape Scale Metagenomics. *MBio*, *8*(4). https://doi.org/10.1128/mBio.00799-17

Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*, *3*(1), 31. https://doi.org/10.1186/s40168-015-0094-5

Marshall, V. M., Lewis, M. M., & Ostendorf, B. (2012). Buffel grass (*Cenchrus ciliaris*) as an invader and threat to biodiversity in arid environments: A review. *Journal of Arid Environments*, *78*, 1–12. https://doi.org/10.1016/j.jaridenv.2011.11.005

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Øvreås, L., Reysenbach, A.-L., Smith, V. H., & Staley, J. T. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, *4*(2), 102–112. https://doi.org/10.1038/nrmicro1341

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Miller, E. T., Svanbäck, R., & Bohannan, B. J. M. (2018). Microbiomes as Metacommunities: Understanding Host-Associated Microbes through Metacommunity Ecology. *Trends in Ecology & Evolution*, *33*(12), 926–935. https://doi.org/10.1016/j.tree.2018.09.002

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, *41*(12), e121–e121. https://doi.org/10.1093/nar/gkt263

Moreno-Hagelsieb, G., Wang, Z., Walsh, S., & ElSherbiny, A. (2013). Phylogenomic clustering for selecting non-redundant genomes for comparative genomics. *Bioinformatics*, *29*(7), 947–949. https://doi.org/10.1093/bioinformatics/btt064

Naureen, Z., Dautaj, A., Anpilogov, K., Camilleri, G., Dhuli, K., Tanzi, B., Maltese, P. E., Cristofoli, F., Antoni, L. de, Beccari, T., Dundar, M., & Bertelli, M. (2020). Bacteriophages presence in nature and their role in the natural selection of bacterial populations. In *Acta Biomedica* (Vol. 91, pp. 1–13). Mattioli 1885. https://doi.org/10.23750/abm.v91i13-S.10819

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., Darcy, J. L., Lynch, R. C., Wickey, P., & Ferrenberg, S. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews*, *77*(3), 342–356. https://doi.org/10.1128/MMBR.00051-12

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a newversatile metagenomic assembler. Genome Res 2017;27:824–834. 10.1101/gr.213959.116.

O'Brien, F. J. M., Dumont, M. G., Webb, J. S., & Poppy, G. M. (2018). Rhizosphere Bacterial Communities Differ According to Fertilizer Regimes and Cabbage (*Brassica oleracea* var. capitata L.) Harvest Time, but Not Aphid Herbivory. *Frontiers in Microbiology*, *9*(JUL), 1–18. https://doi.org/10.3389/fmicb.2018.01620

Oksanen, J. (2017). *Vegan: ecological diversity*. R Package Version 2.4-4. https://cran.r-project.org/package=vegan

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Park, S., Kim, A.-L., Hong, Y.-K., Shin, J.-H., & Joo, S.-H. (2021). A highly efficient auxin-producing bacterial strain and its effect on plant growth. *Journal of Genetic Engineering and Biotechnology*, *19*(1), 179. https://doi.org/10.1186/s43141-021-00252-w

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Pieterse, C. M. J., Zamioudis, C., Berendsen, R. L., Weller, D. M., van Wees, S. C. M., & Bakker, P. A. H. M. (2014). Induced Systemic Resistance by Beneficial Microbes. *Annual Review of Phytopathology*, *52*(1), 347–375. https://doi.org/10.1146/annurev-phyto-082712-102340

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650. https://doi.org/10.1093/molbev/msp077

Probst, A. J., Ladd, B., Jarett, J. K., Geller-Mcgrath, D. E., Sieber, C. M. K., Emerson, J. B., Anantharaman, K., Thomas, B. C., Malmstrom, R. R., Stieglmeier, M., Klingl, A., Woyke, T., Ryan, M. C., & Banfield, J. F. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nature Microbiology*, *3*(3), 328–336. https://doi.org/10.1038/s41564-017-0098-y

R Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2013 Vienna, Austria: R Foundation for Statistical Computing. Available: http://www.r-project.org/

Ranjard, L., & Richaume, A. (2001). Quantitative and qualitative microscale distribution of bacteria in soil. *Research in Microbiology*, *152*(8), 707–716. https://doi.org/10.1016/S0923-2508(01)01251-7

Rouli, L., Mbengue, M., Robert, C., Ndiaye, M., la Scola, B., & Raoult, D. (2014). Genomic analysis of three African strains of *Bacillus anthracis* demonstrates that they are part of the clonal expansion of an exclusively pathogenic bacterium. *New Microbes and New Infections*, *2*(6), 161–169. https://doi.org/10.1002/nmi2.62

Sabki, M. H., Ong, P. Y., Ibrahim, N., Lee, C. T., Klemeš, J. J., Li, C., & Gao, Y. (2021). A review on abiotic stress tolerance and plant growth metabolite framework by plant growth-promoting bacteria for sustainable agriculture. *Chemical Engineering Transactions*, *83*, 367–372. https://doi.org/10.3303/CET2183062

Salazar-Cerezo, S., Martínez-Montiel, N., García-Sánchez, J., Pérez-y-Terrón, R., & Martínez-Contreras, R. D. (2018). Gibberellin biosynthesis and metabolism: A convergent route for plants, fungi and bacteria. *Microbiological Research*, *208*, 85–98. https://doi.org/10.1016/j.micres.2018.01.010

Sánchez-Marañón, M., Miralles, I., Aguirre-Garrido, J. F., Anguita-Maeso, M., Millán, V., Ortega, R., García-Salcedo, J. A., Martínez-Abarca, F., & Soriano, M. (2017). Changes in the soil bacterial community along a pedogenic gradient. *Scientific Reports*, *7*(1), 14593. https://doi.org/10.1038/s41598-017-15133-x

Saul-Tcherkas, V., Unc, A., & Steinberger, Y. (2013). Soil Microbial Diversity in the Vicinity of Desert Shrubs. *Microbial Ecology*, *65*(3), 689–699. https://doi.org/10.1007/s00248-012-0141-8

Schandry, N., & Becker, C. (2020). Allelopathic Plants: Models for Studying Plant–Interkingdom Interactions. *Trends in Plant Science*, *25*(2), 176–185. https://doi.org/10.1016/j.tplants.2019.11.004

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Singleton, D. R., Furlong, M. A., Peacock, A. D., White, D. C., Coleman, D. C., & Whitman, W. B. (2003). *Solirubrobacter pauli* gen. nov., sp. nov., a mesophilic bacterium within the *Rubrobacteridae* related to common soil clones. *International Journal of Systematic and Evolutionary Microbiology*, *53*(2), 485–490. https://doi.org/10.1099/ijs.0.02438-0

Székely, A. J., & Langenheder, S. (2014). The importance of species sorting differs between habitat generalists and specialists in bacterial communities. *FEMS Microbiology Ecology*, *87*(1), 102–112. https://doi.org/10.1111/1574-6941.12195

Tanvir, R., Sajid, I., Hasnain, S., Kulik, A., & Grond, S. (2016). Rare actinomycetes *Nocardia caishijiensis* and *Pseudonocardia carboxydivorans* as endophytes, their bioactivity and metabolites evaluation. *Microbiological Research*, *185*, 22–35. https://doi.org/10.1016/j.micres.2016.01.003

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. v., Crabtree, J., Jones, A. L., Durkin, A. S., Deboy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit Y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., … Fraser, C. M. (2005). Erratum: Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome” (Proceedings of the National Academy of Sciences of the United States of America (September 27, 2005) 102, 39 (13950-13955)). *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16530. https://doi.org/10.1073/pnas.0508532102

Tettelin H., Riley D., Cattuto C., & Medini D. (2008). Comparative genomics: the bacterial pan-genome. Current Opinion in Microbiology, 11(5), 472–477. doi:10.1016/j.mib.2008.09.006

Thakur, M. P., & Geisen, S. (2019). Trophic Regulations of the Soil Microbiome. *Trends in Microbiology*, *27*(9), 771–780. https://doi.org/10.1016/j.tim.2019.04.008

Tinoco-Ojanguren, C., Díaz, A., Martínez, J., & Molina-Freaner, F. (2013). Species diversity and regeneration of native species in *Pennisetum ciliare* (buffelgrass) pastures from the thornscrub of Sonora, México. *Journal of Arid Environments*, *97*, 26–37. https://doi.org/10.1016/j.jaridenv.2013.05.009

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

Tkacz, A., Bestion, E., Bo, Z., Hortala, M., & Poole, P. S. (2020). Influence of Plant Fraction, Soil, and Plant Species on Microbiota: a Multikingdom Comparison. *MBio*, *11*(1), 1–17. https://doi.org/10.1128/mBio.02785-19

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

Trognitz, F., Hackl, E., Widhalm, S., & Sessitsch, A. (2016). The role of plant–microbiome interactions in weed establishment and control. *FEMS Microbiology Ecology*, *92*(10), fiw138. https://doi.org/10.1093/femsec/fiw138

Wang, D. sheng, Erihemu, & Yang, B. (2019). Diversity of bacterial communities in the rhizosphere of the endangered plant, *Paeonia jishanensis*. *Archives of Biological Sciences*, *71*(3), 525–531. https://doi.org/10.2298/ABS190203036W

Wardle, D. A., Nicholson, K. S., Ahmed, M., & Rahman, A. (1994). Interference effects of the invasive plant *Carduus nutans* L. against the nitrogen fixation ability of *Trifolium repens* L. *Plant and Soil*, *163*(2), 287–297. https://doi.org/10.1007/BF00007978

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., et al. gplots: Various R Programming Tools for Plotting Data 2016. Available online at: https://CRAN.R-project.org/package=gplots

Wei, L., Ouyang, S., Wang, Y., Shen, X., & Zhang, L. (2014). *Solirubrobacter phytolaccae* sp. nov., an endophytic bacterium isolated from roots of *Phytolacca acinosa* Roxb. *International Journal of Systematic and Evolutionary Microbiology*, *64*(PART 3), 858–862. https://doi.org/10.1099/ijs.0.057554-0

Weston, L. A., & Mathesius, U. (2014). *Root Exudation: The Role of Secondary Metabolites, Their Localisation in Roots and Transport into the Rhizosphere* (pp. 221–247). https://doi.org/10.1007/978-3-642-54276-3_11

Whitman, W. B. (2015). *Solirubrobacter*. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1–5). Wiley. https://doi.org/10.1002/9781118960608.gbm00227

Wickham, M. H. (2014). *Package 'ggplot2.'*

Yang J, Ruegger PM, McKenry M V., Becker JO, Borneman J. Correlations between Root-Associated Microorganisms and Peach Replant Disease Symptoms in a California Soil. *PLoS One* 2012;7:e46420. 10.1371/journal.pone.0046420.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., & Glöckner, F. O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, *42*(D1), D643–D648. https://doi.org/10.1093/nar/gkt1209

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, L., Zhu, L., Si, M., Li, C., Zhao, L., Wei, Y., & Shen, X. (2014). *Solirubrobacter taibaiensis* sp. nov., isolated from a stem of *Phytolacca acinosa* Roxb. *Antonie van Leeuwenhoek*, *106*(2), 279–285. https://doi.org/10.1007/s10482-014-0194-4

13. ANEXOS

Los protocolos y métodos bioinformáticos utilizados para este trabajo de tesis están disponibles en las siguientes ligas:

https://github.com/genomica-fciencias- unam/buffelgrass
https://github.com/genomica-fciencias-unam/Solirubrobacter