

UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS



**Evolución del estilo de vida acidófilo en el filo ancestral
Verrucomicrobia**

**Tesis presentada a la Universidad de Chile para optar al grado de
Magíster en Bioquímica área de Especialización en Bioquímica
Ambiental y Memoria para optar al Título de Bioquímico por:**

GONZALO ANDRÉS NEIRA MORALES

Director de Tesis: Dr. David S. Holmes

Santiago-CHILE

Julio 2019

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS

INFORME DE APROBACIÓN DE TESIS DE MAGISTER

Se informa a la Dirección de la Escuela de Graduados de la Facultad de Ciencias Químicas y Farmacéuticas que la Tesis de Magíster y Memoria de Título presentada por el candidato

GONZALO ANDRÉS NEIRA MORALES

Ha sido aprobada por la Comisión de Evaluadora de Tesis como requisito para optar al grado de Magíster en Bioquímica, Área de Especialización en Bioquímica Ambiental y Título de Bioquímico, en el examen público rendido el día

Director de Tesis:

Dr. David S. Holmes

Co-director de Tesis:

Dr. Sergio Lobos

Comisión Evaluadora de Tesis:

Dr. Sergio Álvarez

Dra. Jenny Blamey

Dr. Vinicius Maracaja

AGRADECIMIENTOS

En primer lugar, agradecer a mi director de Tesis, Dr. David Holmes por su constante apoyo durante la realización de este proyecto, por las estimulantes discusiones científicas y la confianza a pesar de mi desconocimiento inicial en esta área del conocimiento. Agradecer a mis compañeras de laboratorio Eva y Carolina por su paciencia en enseñarme las distintas herramientas bioinformáticas y por los consejos durante todo el proceso. A mi codirector Dr. Sergio Lobos por sus prontas revisiones de cada escrito de la tesis y por sus comentarios de ánimo en cada etapa.

A la comisión revisora, Dra. Jenny Blamey, Dr. Vinicius Maracaja y Dr. Sergio Álvarez, por sus comentarios y disponibilidad en las distintas etapas de este proceso que funcionaron como una retroalimentación muy valiosa al momento de finalizar esta tesis.

A mi familia, mi madre Pola por todos sus años de trabajo, cariño y apoyo que fueron fundamentales para poder desarrollarme como profesional. Mis abuelos Hipolito y Ana por todos los valores que me entregaron en la crianza. A mis tías Mirtha y Gilda, gracias por permitirme concentrarme en mis estudios a pesar de todas las situaciones vividas a través de los años. A mi padre Pedro por su apoyo incondicional en cada decisión que tomé.

A mis compañeros de la Universidad, porque atrasarme un año fue una de las mejores cosas que pudo pasarme mirando en retrospectiva. A mis amigos del colegio, Sebastián, Joaquín, Oscar y Julio, porque sin todas esas tardes de juego durante la carrera no hubiera podido sobrevivir. A Estefany, aunque cambiaste de camino, tu apoyo se sintió todo el tiempo. Carolina por la amistad de largos años y la influencia para tomar esta ruta en la vida.

A mi pareja Macarena, porque sin su compañía, afecto y contención durante todos estos años no hubiera conseguido seguir hasta el final.

FINANCIAMIENTO

Esta tesis fue desarrollada en el Center for Bioinformatics and Genome Biology, de la Fundación Ciencia & Vida, bajo la dirección del Dr. David Holmes y la codirección del Dr. Sergio Lobos.

El trabajo contó con el financiamiento de los proyectos: Programa de Apoyo a Centros con Financiamiento Basal AFB 170004 y Fondecyt 1181717.

PRESENTACIONES A CONGRESOS

David S. Holmes, Carolina González, Eva Vergara, **Gonzalo Neira**. The Case for the Emergence of Life in Thermo-Acidic Conditions. Astrobiology 2017 (International Astronomical Union), Coyhaique, Chile, Nov. 26-Dec. 1, 2017.

Carolina González, Eva Vergara, **Gonzalo Neira**, Jorge Valdés and David S. Holmes. Forces Shaping Microbial Speciation in Extremely Acidic Environments. 12th International Congress of Extremophiles. Napoles, Italy, Sept. 16-20, 2018.

Gonzalo Neira, Eva Vergara, Carolina González and David Holmes. Phylogenomic and comparative genomic analyses shed light on the evolution of acidophilic traits in the Phylum Verrucomicrobia. ISCB-LA SOIBIO EMBnet 2018. Viña del Mar, Chile, November 5 - 9, 2018.

Carolina González, Eva Vergara, **Gonzalo Neira**, Diego Cortez, Jorge Valdés and David Holmes. Forces Shaping Microbial Speciation In Extremely Acidic Environments. ISCB-LA SOIBIO EMBnet 2018. Viña del Mar, Chile, November 5 - 9, 2018.

Gonzalo Neira, Eva Vergara, Carolina González and David Holmes. Integrative Genomics Insights into the Evolution of the Deeply Rooted, Extreme Acidophiles of the Verrucomicrobia Phylum. XXIV ALAM 2018 - XL SOMICH - II ASOCHIN – IX SLAMTB. Santiago, Chile, Nov. 13-16, 2018.

ÍNDICE GENERAL

AGRADECIMIENTOS	i
FINANCIAMIENTO	ii
PRESENTACIONES A CONGRESOS	iii
ÍNDICE GENERAL	iv
ÍNDICE DE FIGURAS	vii
ÍNDICE DE TABLAS	viii
ÍNDICE DE TABLAS SUPLEMENTARIAS	ix
ABREVIATURAS	x
ABSTRACT	xii
RESUMEN	xiv
1. INTRODUCCIÓN	1
1.1. Extremófilos.	1
1.2. Herramientas Bioinformáticas y Microbiología.	2
1.3. Acidófilos.	4
1.4. Verrucomicrobia.	8
2. HIPÓTESIS	12
3. OBJETIVO GENERAL	12
4. OBJETIVOS ESPECÍFICOS	12
5. METODOLOGÍA	13
5.1. Recopilación de genomas y análisis de calidad.	13
5.2. Obtención de marcadores.	13

5.3. Construcción de árboles filogenéticos.	15
5.4. Análisis de similitud de secuencias.	15
5.5. Predicción de elementos móviles.	16
5.6. Anotación funcional de genes codificantes para proteínas.	16
5.7. Identificación familias de proteínas y definición de pangenoma.	17
5.8. Identificación y análisis de propiedades acidófilas en genomas.	17
5.9. Reconstrucción de la historia evolutiva del grupo acidófilo del filo Verrucomicrobia.	18
6.- RESULTADOS	19
6.1. Análisis filogenéticos, filogenómicos y de similitud de secuencias del filo Verrucomicrobia.	19
6.1.1. Determinación de calidad de los genomas.	19
6.1.2. Marcadores filogenéticos y filogenómicos.	19
6.1.3. Árboles filogenéticos y filogenómicos.	22
6.1.4. Análisis de similitud de secuencias.	24
6.2. Predicción de propiedades funcionales relacionadas a acidofilia y genómica comparativa.	26
6.2.1. Predicción de elementos móviles, definición de estabiloma y moviloma.	26
6.2.2. Identificación de propiedades funcionales de genes codificantes para proteínas.	28
6.2.3. Identificación de familias de proteínas y definición de pangenoma.	29
6.2.4. Identificación y análisis de propiedades acidófilas en <i>Methylacidiphilaceae</i>.	31
6.3. Reconstrucción evolutiva de propiedades acidófilas en <i>Methylacidiphilaceae</i>.	33
6.3.1. Transferencia horizontal de genes.	33

6.3.2. Fusión de secuencias.	37
6.3.3. Duplicación de secuencias.	39
6.3.4. Reconstrucción evolutiva propiedades acidófilas.	40
7. DISCUSIÓN	42
7.1. Nueva clasificación taxonómica para Verrucomicrobia.	42
7.2. Definición de moviloma y estabiloma.	44
7.3. Análisis funcional de <i>Methylacidiphilaceae</i>.	45
7.4. Definición del pangenoma de <i>Methylacidiphilaceae</i>.	46
7.5. Mecanismos de resistencia a pH bajo.	47
7.6. Transferencia horizontal de genes.	48
7.7. Reconstrucción de historia evolutiva.	49
8. CONCLUSIONES	51
9. MATERIAL SUPLEMENTARIO	52
10. BIBLIOGRAFÍA	57

ÍNDICE DE FIGURAS

Figura 1	Mecanismos de adaptación a pH bajo de microorganismos acidófilos.	7
Figura 2	Árbol filogenético del gen ARN ribosomal 16S de los miembros de Verrucomicrobia, mostrando sus 7 subdivisiones clásicas.	9
Figura 3	Árbol filogenético de 16S rRNA Verrucomicrobia. Árbol filogenético con 27 organismos del filo Verrucomicrobia.	24
Figura 4	Árbol filogenéticos de marcadores concatenados Verrucomicrobia.	25
Figura 5	Clasificación taxonómica basada en similitud de secuencias.	27
Figura 6	<i>Heatmap</i> de valores ANIb-TETRA.	27
Figura 7	Predicción elementos móviles.	29
Figura 8	<i>Heatmap</i> distribución de COG.	31
Figura 9	Diagrama de Venn del pangenoma de organismos acidófilos.	32
Figura 10	Mecanismos de resistencia a pH bajo identificados en <i>Methylacidiphilaceae</i>	34
Figura 11	Contexto genómico glutamato descarboxilasa y antiporter	38
Figura 12	Contexto genómico arginina/agmatina antiporter y transportador de fosfato.	39
Figura 13	Contexto genómico transportador de potasio y cianasa.	40
Figura 14	Contexto genómico AguA y AguB.	41
Figura 15	Fusión hpnD y hpnE <i>Methylacidiphilaceae</i> .	44
Figura 16	Reconstrucción evolutiva predicha de genes asociados a resistencia a pH extremadamente bajo en <i>Methylacidiphilaceae</i> .	46

ÍNDICE DE TABLAS

Tabla 1	Genomas de Verrucomicrobia	14
Tabla 2	Genomas de Verrucomicrobia y valores de calidad	21
Tabla 3	Marcadores para árboles filogenéticos y filogenómicos	23
Tabla 4	<i>Moviloma</i> y <i>estabiloma</i> en organismos acidófilos	29

ÍNDICE DE TABLAS SUPLEMENTARIAS

Tabla suplementaria 1. Genes asociados a resistencia a acidez.	59
Tabla suplementaria 2. Marcadores utilizados para árbol 31cic.	61
Tabla suplementaria 3. Marcadores utilizados para árbol bac120.	62

ABREVIATURAS

AAI	<i>Average Aminoacid identity</i>
ADN	Ácido desoxirribonucleico
AdiA	Arginina descarboxilasa
AdiC	Arginina/agmatina antiporter
AguA	Agmatina deiminasa
AguB	N-carbamoilputrescina amidohidrolasa
ANI	<i>Average nucleotide identity</i>
ARN	Ácido ribonucleico
ARNr	Ácido ribonucleico ribosomal
ATP	Adenosín trifosfato
BBBH	<i>Bidirectional best blast hit</i>
Can	Anhidrasa carbónica
COG	<i>Clusters of orthologous groups</i>
CynS	Cianato hidratasa
FTP	<i>File transfer protocol</i>
GadA	Glutamato descarboxilasa
GadC	Glutamato/GABA antiporter
HMM	Modelo oculto de Márkov
HpnC	hidroxiescualeno sintasa
HpnD	Prescualeno difosfato sintasa
HpnE	Hidroxiescualeno dehydroxilasa
HupB	Proteína de unión a DNA homóloga a HU

Kch	Transportador de potasio asociado a voltaje
Kdp	ATPasa asociada a transporte de potasio
ME	<i>Minimum evolution</i>
MFUM	<i>Methylacidiphilum fumariolicum</i>
MINF	<i>Methylacidiphilum infernorum</i>
MKAM	<i>Methylacidiphilum Kamchatkense</i>
ML	<i>Maximum likelihood</i>
NaH	Antiporter de sodio/protones
NCBI	<i>National Center for Biotechnology Information</i>
NJ	<i>Neighbour joining</i>
NR	<i>Non redundant</i>
PVC	Planctomycetes-Verrucomicrobia-Chlamidya
Pst	Complejo transportador de fosfato
Refseq	<i>Reference sequence database</i>
Slp	Lipoproteina asociada a privación de carbono
SpeA	Arginina descarboxilasa
TETRA	<i>Tetranucleotide frequencies</i>
VLP2	<i>Verrucomicrobia bacterium LP2A</i>
VS3C	<i>Verrucomicrobium sp. 3C</i>

ABSTRACT

Extremophiles are defined as organisms capable of living in extreme conditions that had previously been considered too extreme to support life. Acidophiles are an interesting group of extremophiles. They grow optimally at pHs <3.0. Acidophiles are present in the three domains of life (Archaea, Bacteria and Eukaryote) and are represented in many different branches within these domains. The phylum Verrucomicrobia (Bacteria) include lineages of extreme acidophiles.

Verrucomicrobia are considered to be an ancient phylum, rooting near the base of the tree of life. Studies based on 16S rRNA gene phylogeny have classified the phylum into 7 subdivisions that include microorganisms with sequenced genomes. One of these divisions is a clade of acidophilic organisms (optimum growth pH between 2.0 and 3.0) that has not been studied in detail. This acidophilic clade is separated into two lineages: the genus *Methylacidiphilum* (thermophiles with optimum growth around 60 °C) and the genus *Methylacidimicrobium* (mesophiles with optimal growth around 30 °C). The nearest clades to these acidophiles are neutrophils, posing the question as to how the acidophiles gained their ability to live at extremely low pH. Importantly, from the perspective of deep evolution, they are the only known methanotrophic organisms that inhabit extremely low pH conditions, organisms that play a key role in balancing the atmospheric methane.

Difficulties in experimental manipulation of these microorganisms, associated with growth dependence on rare metals and the lack of genetic manipulation techniques, pose challenges to understand the mechanisms that allow them to survive in low pH environments. With the dramatic increase of available genomic sequences in recent years, the use of various bioinformatics tools has become essential as part of the study of organisms that are difficult to manipulate experimentally.

In this thesis project, I propose an evolutionary model that suggests how acidophilic properties arose within the Verrucomicrobia using bioinformatics tools including phylogenetic studies and comparative genomics. Mechanisms related to resistance at low pH were identified and their distribution within the pangenome, core-genome, disposable genome and unique genome for each member of the *Methylacidiphilaceae* family was studied. These mechanisms include: decarboxylation/antiporter mechanisms of glutamate and arginine; F₀F₁-ATP synthase complex; cation transporters and membrane molecules related to conferring homeostasis of protons.

The proposed model of genomic evolution could explain the origin of the acidophilic properties within the phylum Verrucomicrobia. Analysis of gain and loss of genes based on the maximum parsimony method infers a common ancestor of all members of *Methylacidiphilaceae*. Signals of horizontal transfer were identified in the genomic neighborhood of genes such as *gadC* and *adiC*, multiple copies of genes such as *slp*, a second complex F₀F₁-ATP synthase and modifications such as fusion of *hpnD* and *hpnE* in the hopanoid biosynthetic pathway were identified as important changes that potentially contribute to the gain of acidophilic property within Verrucomicrobia.

RESUMEN

Los extremófilos se encuentran definidos como organismos capaces de vivir en condiciones consideradas extremas, en las cuales, previamente se pensaba no podría existir vida. Dentro de éstos, un grupo de interés corresponde a los denominados acidófilos, los cuales poseen un crecimiento óptimo a pH bajo el neutro. Son definidos dentro de este grupo como acidófilos extremos los organismos cuyo óptimo crecimiento es a pH bajo 3,0. Los acidófilos se encuentran presentes en los tres dominios de la vida (Archaea, Bacteria y Eukarya) divergiendo en distintas ramas dentro de estos dominios. El filo Verrucomicrobia (Bacteria) presenta organismos con esta característica.

Considerado como un filo ancestral por su posición cerca de la raíz en el árbol de la vida, los miembros del filo Verrucomicrobia han sido encontrados de forma ubicua, presentes en un 99% en muestras de suelo ambientales y un 98% en muestras marinas. Estudios basados en el gen ARN ribosomal 16S han dividido el filo en 7 subdivisiones, para las cuales una serie de organismos cuentan con genomas secuenciados. Sin clasificación taxonómica a nivel de subdivisión se encuentra un clado de organismos acidófilos, característica que no ha sido estudiada en detalle en estos organismos. El clado acidófilo se separa en dos linajes de organismos acidófilos extremos diferentes (pH óptimo de crecimiento entre 2,0 y 3,0) y clasificados en los géneros *Methylacidiphilum* (termófilos con óptimo crecimiento alrededor de 60 °C) y *Methylacidimicrobium* (mesófilos con óptimo crecimiento alrededor de 30 °C). Características comunes a ambos es ser bacterias Gram negativo y ser los únicos organismos conocidos metanótrofos que habitan en condiciones de pH extremadamente bajo, adicionalmente juegan un rol clave en la mantención de los niveles de concentración de metano ambiental.

Dificultades en la manipulación experimental de estos microorganismos, asociado al crecimiento dependiente de metales raros y la ausencia de técnicas de manipulación genética en acidófilos extremos, no han permitido comprender por completo cuáles son los mecanismos que les permiten sobrevivir en ambientes de bajo pH. Con el aumento de secuencias genómicas en los últimos años, el uso de variadas herramientas bioinformáticas se ha vuelto esencial como parte del estudio de organismos de difícil manipulación experimental.

En esta tesis, se propone un modelo de evolutivo sobre el origen de la propiedad acidófila dentro del filo Verrucomicrobia, mediante el uso de herramientas de bioinformática como estudios filogenéticos y genómica comparativa. Los organismos acidófilos del filo se clasifican dentro de una familia única denominada *Methylacidiphilaceae* dividida en dos géneros *Methylacidiphilum* y *Methylacidimicrobium*, cuya característica diferenciadora es la temperatura óptima de crecimiento siendo termófilos y mesófilos respectivamente.

Posteriormente se identifica los mecanismos relacionados a la resistencia a bajo pH presentes en estos organismos acidófilos y fue definido el pangenoma, core-genoma, genoma prescindible y genoma único para cada miembro de la familia *Methylacidiphilaceae*. Se identificó la presencia de mecanismos de descarboxilación/antiporter de glutamato y arginina, el complejo F₀F₁-ATP sintasa, transportadores de cationes y moléculas de membrana relacionados a conferir esta característica homeostasis de protones como parte del core genoma de *Methylacidiphilaceae*.

Finalmente, se propone un modelo de evolución genómica que podría explicar el origen de las propiedades acidófilas dentro del filo Verrucomicrobia. Análisis de ganancia y pérdida de genes basado en el método de la máxima parsimonia permite proponer que la propiedad acidófila fue obtenida en un ancestro inferido común de todos los miembros de *Methylacidiphilaceae*. Indicios de transferencia horizontal de genes fueron identificados en el vecindario genómico de genes como *gadC* y *adiC*. Múltiples copias de genes como *slp*, un segundo complejo F₀F₁-ATP sintasa y modificaciones como la fusión de *hpnD* y *hpnE* en la vía de biosíntesis de hopanoides fueron identificados como fuerzas que contribuyeron en la ganancia de la propiedad acidófila dentro Verrucomicrobia.

1. INTRODUCCIÓN

1.1. Organismos extremófilos.

Los organismos que viven en ambientes que normalmente consideramos no habitables, son denominados extremófilos (Rampelotto, 2013). Los extremófilos han sido identificados en los tres dominios de la vida (Archaea, Bacteria y Eukarya, con mayor abundancia en los dos primeros) y han permitido ampliar nuestro espectro de entendimiento sobre en qué condiciones la vida puede existir dentro de nuestro planeta (Gupta et al., 2014). Los extremófilos se han encontrado en ambientes con temperaturas cercanas a los 100°C (Mehta *et al.*, 2016), en agua de mar a temperaturas bajo 5°C (de Maayer *et al.*, 2014), alta radiación ionizante del orden de los 10.000 Gray (Cox y Batista, 2005) y pH extremadamente bajos (cerca de 0) y altos (cerca de 12) (Pikuta *et al.*, 2007). Debido a estas propiedades los extremófilos han sido propuestos como posibles modelos de estudios de organismos ancestrales (Di Giulio, 2000).

Los extremófilos son clasificados según la condición particular en la que son capaces de crecer. Encontramos termófilos y psicrófilos (adaptados a altas y bajas temperaturas respectivamente), halófilos (resistentes a altas concentraciones de sales), barófilos (asociados a altas presiones), acidófilos y alcalófilos (capaces de crecer a bajo y alto pH, respectivamente), entre otros. Organismos capaces de crecer en más de una de estas condiciones extremas son clasificados como poli-extremófilos (Gerday, C. y Glansdorff, N. 2007).

Las condiciones extremas en las que logran crecer de manera óptima estos organismos los hacen interesantes en diferentes aplicaciones tales como:

(i) Optimización de procesos biotecnológicos que ocurren en condiciones ambientales necesariamente extremas (Coker, 2016).

(ii) Modelos de organismos ancestrales que nos puedan entregar indicios sobre el origen de la vida en nuestro planeta (Delaye *et al.*, 2005).

(iii) Modelos para la búsqueda de vida en otros planetas, tanto de nuestro sistema solar como exoplanetas (Cavicchioli, 2002).

(iv) En biología sintética, mediante la construcción o rediseño de nuevos sistemas biológicos para estudios de evolución o construcción de sistemas biológicos artificiales (Rothschild, 2010).

(v) Aplicaciones médicas, mediante la producción de diferentes moléculas antitumorales, antibióticas o antifúngicas (Coker, 2016).

Dadas las diferentes aplicaciones de los microorganismos extremófilos, es importante comprender los mecanismos que permiten a estos organismos crecer en sus nichos ecológicos. En la actualidad existen una serie de complicaciones en el cultivo de estos organismos extremófilos en el laboratorio. Entre estas dificultades es importante destacar la diferencia existente entre las condiciones de cultivo en laboratorio y las verdaderas condiciones ambientales en las que estos prosperan (Vartoukian *et al.*, 2010). Además, factores como las diversas interacciones entre microorganismos en su nicho ecológico, los diferentes procesos químicos que ocurren y la interacción entre procesos bióticos y abióticos son claves para entender la dificultad en simular un ambiente similar al original en condiciones de laboratorio (Nichols *et al.*, 2008). Otra complicación son los largos tiempos de incubación necesarios para el crecimiento de algunos de estos organismos (Alain, 2009). Actualmente gracias a los avances en genética molecular y modernas herramientas de secuenciación de ADN genómico, ya es posible mediante análisis bioinformáticos, responder preguntas sobre el funcionamiento y la evolución de los organismos extremófilos.

1.2. Herramientas Bioinformáticas y Microbiología.

Estudios basados en secuencias genómicas mediante la utilización de herramientas bioinformáticas nos permiten obtener información con relevancia biológica para organismos los cuales no se encuentran aislados en cultivos puros (Tettelin *et al.*, 2008). Esto es fundamental teniendo en consideración que cerca de un 98% del total de las bacterias no pueden ser manipuladas mediante técnicas clásicas experimentales debido a la imposibilidad de su cultivo (Amann *et al.*, 1995). Los avances en tecnologías de secuenciación y la disminución en sus costos han aumentado la disponibilidad de genomas secuenciados completos en bases de datos accesibles (Hall, 2007).

Las disciplinas que más se han beneficiado de la tecnología de secuenciación masiva son los campo de la filogenética y filogenómica (Posada, 2016). La idea de comprender las relaciones entre organismos comenzó después del trabajo de Charles Darwin en “El origen de las especies”, donde por primera vez se presentó un árbol que representa relaciones entre las especies (*Charles*

Darwin and the Tree of Life). Posteriormente estas reconstrucciones fueron realizadas utilizando modelos de reconstrucción basados en características homólogas compartidas entre diferentes organismos, por ejemplo, características anatómicas en fósiles (Donoghue *et al.*, 1989). Con la aparición de herramientas de biología molecular y la obtención de secuencias de ADN, una serie de genes comenzaron a ser utilizados como marcadores. En el caso de los organismos procariontes, se comenzó con genes codificantes para la subunidad pequeña ribosomal, como el gen del ARN ribosomal 16S (Ludwig y Schleifer, 1994). La utilización de estos genes altamente conservados significó un aumento en la capacidad de realizar inferencias evolutivas, pero a su vez al aumentar la cantidad de marcadores utilizados provocó dificultades, como la obtención de distintas topologías en árboles filogenéticos (Philippe *et al.*, 2011). Por esta problemática, y debido al aumento en la cantidad de datos generados se comenzó a utilizar no solo un gen marcador, sino una serie de secuencias tanto nucleotídicas como aminoacídicas para la construcción de filogenias dando inicio al campo de la filogenómica (Delsuc *et al.*, 2005).

Actualmente existen tres métodos principales de construcciones de árboles filogenéticos basados en la utilización de matrices construidas con caracteres homólogos:

- (i) Métodos basados en distancia, que convierten la matriz de caracteres en una de distancia y donde se puede ver la distancia entre cada una de las especies en estudio. El árbol filogenético se infiere desde esta matriz usando algoritmos como *neighbour joining* (NJ) o *minimum evolution* (ME).
- (ii) Máxima parsimonia, en que se selecciona el árbol en el cual el mínimo número de cambios entre caracteres explique la información observada.
- (iii) Función *likelihood*, o simplemente *likelihood*, que se basa en calcular la probabilidad de que un árbol dado pueda producir la información entregada. *Maximum likelihood* (ML) selecciona el árbol que maximiza la probabilidad de observar la información basada en un modelo específico y los métodos bayesianos utilizan la probabilidad posterior usando una fórmula matemática (Yang y Rannala, 2012).

Otro de los campos beneficiados con la llegada de tecnología de secuenciación masiva es la genómica comparativa que analiza las diferencias entre múltiples genomas (Valdes *et al.*, 2010). Las preguntas iniciales fueron relacionadas al número de genes que forman el repertorio de una especie en particular. Uno de los conceptos más utilizados que se desprenden de la genómica

comparativa es el pangenoma, definido como la suma de todos los genes de los genomas en estudio. Adicionalmente, fue definido el *core-genoma* como el grupo de genes que están presentes en todos los genomas analizados, el cual normalmente codifica para proteínas en relación con funciones esenciales para la célula. Finalmente, el genoma dispensable fue definido como los genes que estaban compartidos en solo algunas de las especies pero no en todas, y es el encargado de otorgar diversidad a las diferentes especies (Medini *et al.*, 2015).

1.3. Acidófilos.

Los organismos acidófilos se encuentran definidos como organismos capaces de crecer de manera óptima a un pH menor a 5,0. Adicionalmente, se encuentran clasificados como acidófilos extremos los organismos cuyo crecimiento óptimo es a un pH menor a 3,0 y acidófilos moderados los cuales tienen su óptimo de crecimiento entre pH 3 y 5 (Johnson, 2007). Organismos capaces de crecer a un pH bajo 5,0 pero que tienen su óptimo de crecimiento fuera de este rango son clasificados como tolerantes a acidez y no propiamente como acidófilos. En el árbol de la vida los acidófilos se encuentran de forma ubicua, presentes tanto en Archaea como en Bacteria (Johnson, 1998) y también en Eukarya donde algunas algas y levaduras son capaces de tolerar pH cercanos a 0 (Baker *et al.*, 2004).

Estos organismos se encuentran presentes en ambientes con alta acidez, la cual pueden ser producto de procesos naturales como la producción de ácido carbónico formado desde agua y CO₂ (Rowell y Wild, 1985) o la oxidación del azufre elemental y minerales sulfurados, produciendo ácido sulfúrico, el cual es responsable de la gran mayoría de los ambientes extremadamente ácidos en el planeta (Johnson, 2007). Ejemplos de estos ambientes donde organismos acidófilos han sido aislados y estudiados son:

- i) Áreas volcánicas, como las existentes en Yellowstone National Park, Nueva Zelanda o Islandia, donde el azufre elemental formado por la condensación de H₂S y SO₂ (asociado a las emanaciones volcánicas) es oxidado por arqueas y bacterias produciendo altos niveles de acidificación (Dopson y Johnson, 2012).
- ii) Drenajes ácidos mineros generados por la actividad humana, donde minerales asociados a la actividad minera (principalmente minerales que contienen azufre dentro de su

composición) al ser expuestos tanto a agua como al aire, son oxidados produciendo la acidez ambiental, proceso acelerado por la presencia de bacterias y arqueas las cuales son capaces de catalizar las reacciones de oxidación de estos materiales (Akcil y Koldas, 2006).

En la actualidad la comprensión sobre los mecanismos que permiten el crecimiento y la supervivencia de los acidófilos sigue en expansión gracias al descubrimiento constante de nuevas vías asociadas. De manera general, para crecer de manera óptima en condiciones ambientales de alta acidez, estos organismos mantienen un pH interno cercano al neutro a pesar de que el pH exterior sea extremadamente bajo (Matin, 1999). Los mecanismos de resistencia a acidez actualmente descritos son: (i) alta impermeabilidad de la membrana a protones; (ii) tamaño de poro reducido en canales de membrana; (iii) inhibición en el flujo de protones debido al aumento en la carga positiva de la membrana interna; (iv) expulsión del exceso de protones fuera de la célula; (v) desacoplamiento de protones por ácidos orgánicos. Un resumen esquemático de algunas de estas propiedades se observa en la **Figura 1** (Baker-Austin y Dopson, 2007).

Los microorganismos acidófilos presentan un rango variado de metabolismos, estos pueden ser tanto quimiótrofos (obteniendo energía desde compuestos químicos) como fotótrofos (obtención de energía desde la luz), mientras que la asimilación del carbono puede ser autótrofa (obtención desde CO₂), heterótrofa (obtención de compuestos orgánicos) o mixotrófica (ambos mecanismos) (Johnson y Hallberg, 2008). Los donadores de electrones inorgánicos tienen un rol importante como fuentes de energía en muchos de los acidófilos extremos, esto asociado a la mayor concentración en ambientes ácidos del ión ferroso y algunas formas de azufre reducido, superando al carbono orgánico (Golyshina *et al.*, 2000; Rohwerder y Sand, 2007). Producto de las restricciones ambientales, el rango de donadores de electrones orgánicos usados por acidófilos es pequeño, siendo mayormente capaces de utilizar compuestos monoméricos de bajo peso molecular como algunos azúcares o alcoholes (Johnson, 2007). Una gran oportunidad derivada de estos ambientes para los organismos acidófilos es la gradiente de protones existente debido a la alta acidificación de los ambientes, la cual es utilizada para la generación de ATP, funcionando como una ventaja adaptativa (Ingledeew, 1982).

Hasta el mes de marzo de 2015, 161 genomas de acidófilos fueron identificados, provenientes de organismos aislados, muestras ambientales y reconstrucciones metagenómicas, en estado finalizado o sin finalizar, entregando una amplia fuente de información para estudios genómicos (Cardenas *et al*, 2016). Estudios basados en genomas de microorganismos acidófilos han sido realizados en procariontes, en arqueas y en bacterias. En el dominio Archaea existen genomas representativos de diferentes ramas, como Crenarchaeota y Euryarchaeota (Slonczewski, 2013). En el dominio Bacteria, genomas representativos de filos como Firmicutes y Proteobacteria han sido ampliamente estudiados mediante métodos bioinformáticos debido a sus aplicaciones en procesos industriales, donde un ejemplo son los organismos del género *Acidithiobacillus*. Sin embargo, a la fecha existen filos profundamente enraizados en el árbol de la vida, como Verrucomicrobia, en los cuales no se han realizados estudios exhaustivos sobre la propiedad acidófila, dado que está compuesto principalmente por bacterias ambientales sin aplicaciones industriales ya determinadas. El estudio de filos profundamente enraizados es muy importante, ya que nos puede entregar indicios sobre el posible origen esta característica (Johnson, 1998).

Los acidófilos son un grupo importante de microorganismos tanto desde el punto de vista ambiental como económico, por lo tanto, es fundamental comprender los mecanismos por los cuales estos organismos logran sobrevivir en ambientes ácidos y cuál es la participación de éstos en sus nichos ecológicos. Gracias a los avances en tecnologías de secuenciación, hoy se pueden realizar estudios genómicos en algunos grupos de bacterias poco estudiados como el filo Verrucomicrobia.

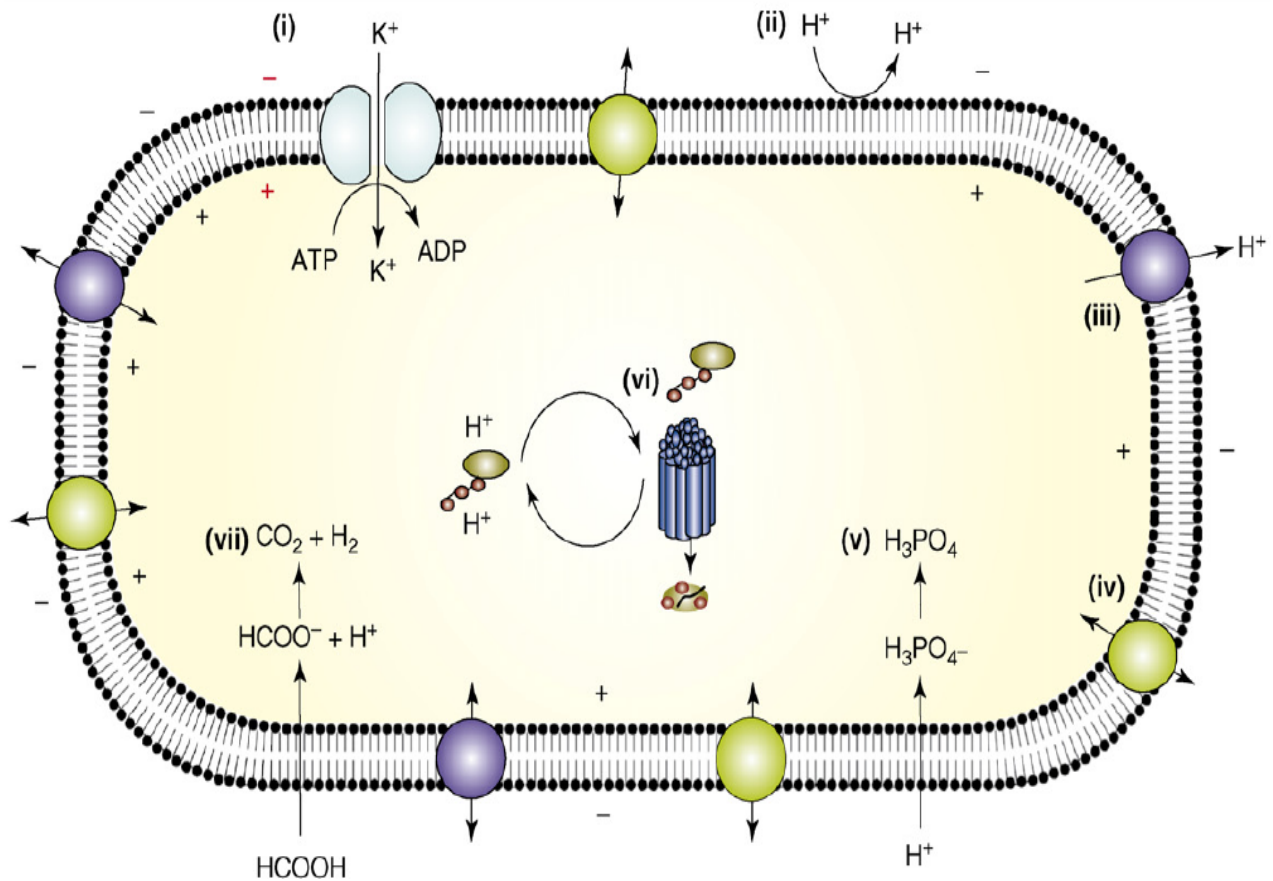


Figura 1. Mecanismos de adaptación a pH bajo de microorganismos acidófilos. i) Disminución del flujo de protones. Presencia de transportadores de cationes los cuales aumentan la carga positiva en la cara interna de la membrana ii) disminución permeabilidad de membrana, modificaciones en las estructuras lipídicas disminuyen la entrada de protones; iii) expulsión de protones, la diferencia de pH es controlada con la expulsión desde el espacio intracelular; iv) transportadores secundarios, se ve un aumento en relación a la presencia de estos en organismos neutrófilos; v) buffer intracelulares, moléculas capaces de secuestrar protones en el medio citoplasmático; vi) sistemas de reparación de proteínas, aumento en la proporción de proteínas asociadas a estos sistemas en comparación a organismos neutrófilos; vii) ácidos orgánicos funcionan como desacopladores que posteriormente podrían ser degradados (adaptado de Baker-Austin, C., y Dopson, M. 2007).

1.4. Verrucomicrobia.

El filo Verrucomicrobia fue inicialmente propuesto como una división dentro del dominio Bacteria (Hedlund *et al.*, 1996) y posteriormente, con nuevos estudios de su taxonomía, fue clasificado como un nuevo filo (Garrity y Holt, 2001). Estas especies fueron observadas por primera vez en ambientes marinos (Henrici y Johnson, 1935), pero a la fecha se sabe que existen en variados ambientes naturales, llegando a estar presentes en un 99% de las muestras de suelo (Bergmann *et al.*, 2011) y un 98% de las muestras marinas (Freitas *et al.*, 2012). El filo está principalmente compuesto por especies identificadas a partir de secuencias genómicas ambientales y por especies aisladas y cultivadas (Davis *et al.*, 2005; Janssen *et al.*, 1997). Estudios filogenéticos utilizando el gen ARN ribosomal 16S de microorganismos del filo los han clasificado en el árbol de la vida dentro del súper-filo “PVC”, formando un clado en conjunto con Planctomycetes, Chlamydia, Lentisphaerae, OP3 y Poribacteria (Wagner y Horn, 2006). La clasificación de “súper-filo” ha supuesto una excepción a las clasificaciones tradicionales debido a que muchos de sus miembros presentan características muy variadas desde punto de vista de la variedad de sus metabolismos, así como de su morfología (Fuerst, 2006). Estudios filogenéticos al interior de Verrucomicrobia basados en el gen ARN ribosomal 16S, describieron inicialmente 5 subdivisiones diferentes (Hugenholtz *et al.*, 1998), lo que fue ampliado a 7 subdivisiones con el aumento de secuencias (Schlesner *et al.*, 2006). La clasificación clásica del filo Verrucomicrobia puede ser observada en la **Figura 2**.

De particular interés para nuestro grupo de estudio son los representantes acidófilos en Verrucomicrobia, los cuales no se encuentran clasificados en ninguna subdivisión tradicional. Las características generales de estos organismos son: (i) Gram negativo (ii) metanótrofos, en particular realizan oxidación aeróbica del metano; (iii) autótrofos, utilizando CO₂ como fuente de carbono mediante el uso del ciclo de Calvin; (iv) el haber sido aislados desde fuente geotermales; (v) presentar un óptimo de crecimiento a pH bajos entre 1 y 3,5.

Con el aumento de secuencias y de especies aisladas desde distintos ambientes, han surgido problemáticas en la clasificación dentro del filo Verrucomicrobia. Dos ejemplos son los casos de la subdivisión número 5 y la subdivisión número 7, para las cuales una vez realizados nuevos análisis fenotípicos, genómicos y ambientales la primera subdivisión fue clasificada como un filo nuevo denominado Kiritimatiellaota (Spring *et al.*, 2016) y la segunda, actualmente es considerada como parte del filo Lentisphaerae (Cho *et al.*, 2004). Debido a estas inconsistencias

en las clasificaciones del filo es fundamental realizar nuevos estudios utilizando más marcadores genéticos (no sólo el gen ARN ribosomal 16S), para obtener filogenias que expliquen de manera más robusta las relaciones en Verrucomicrobia.

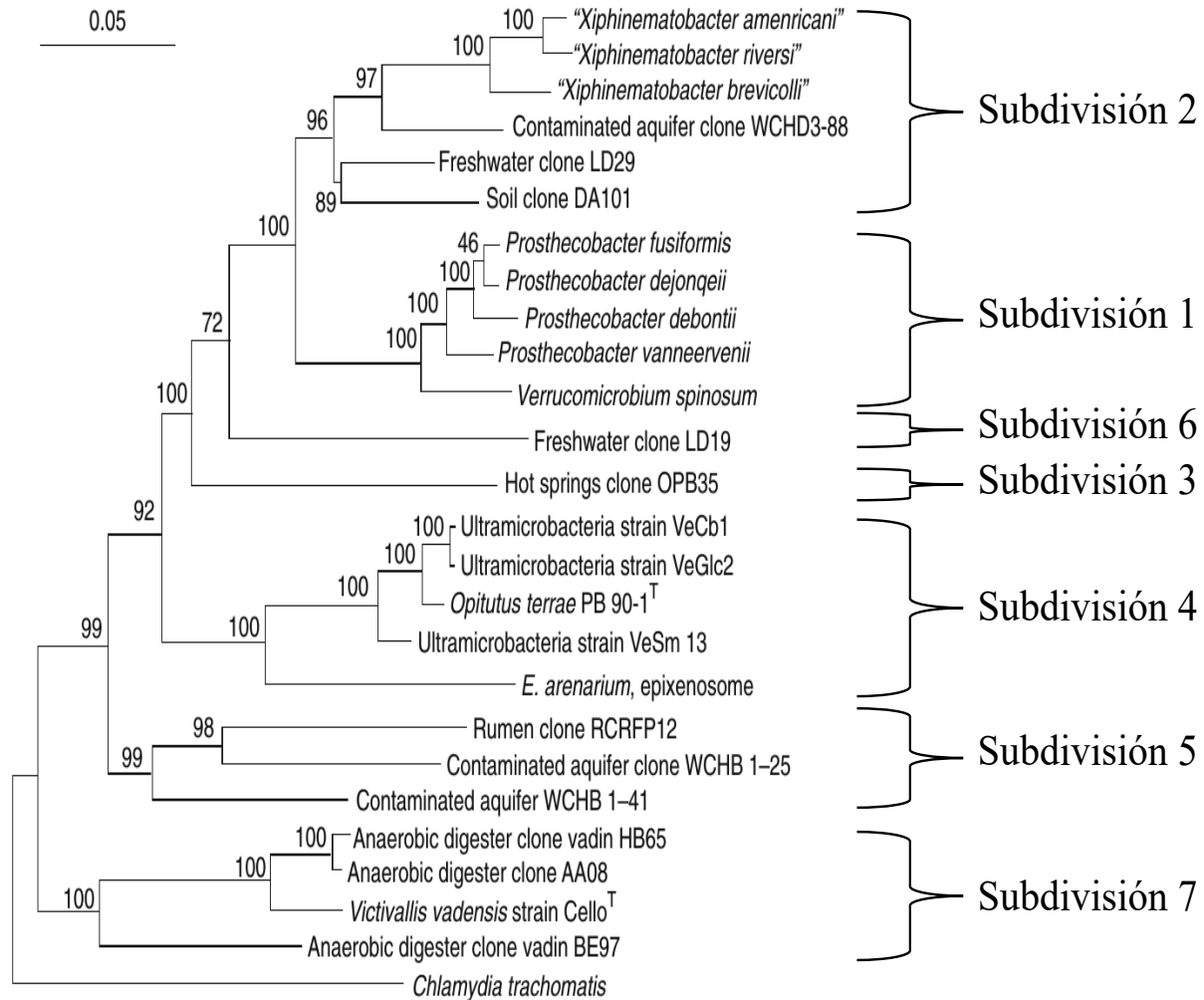


Figura 2. Árbol filogenético del gen ARN ribosomal 16S de los miembros de Verrucomicrobia, mostrando sus 7 subdivisiones clásicas. (1) *Verrucomicrobiae*, con organismos de los géneros *Prostheco bacter* y *Verrucomicrobium*; (2) "Spartobacteria", mayormente encontradas en muestras de suelo (Janssen, 2006); (3) Subdivisión 3, donde el único organismo aislado es *Pedosphaera parvula*; (4) *Opitutae*, con dos clados formados por *Opitutus terrae* y *Puneicoccales*; (5) Subdivisión 5, donde sólo está aislado *Kiritimatiella glycovorans*; (6) Subdivisión 6. Sólo con secuencias de ARN ribosomal 16S; (7) Subdivisión 7, donde encontramos a *Victivallis vadensis*. Valores de *bootstrap* en los nodos (Adaptado de Schlesner *et al.*, 2006).

Estudios filogenéticos han mostrado la existencia de dos clados claramente demarcados de organismos acidófilos en el filo Verrucomicrobia, el primero correspondiente al género *Methylacidiphilum* donde los organismos comparten la característica de ser poli-extremófilos, pudiendo resistir altas temperaturas con óptimos de crecimiento entre 55 °C a 60 °C y el segundo, correspondiente al género propuesto *Methylacidimicrobium*, donde los organismos a diferencia de los anteriores mencionados son mesófilos.

Para el género *Methylacidiphilum* encontramos las siguientes especies descritas:

- i) *Methylacidiphilum fumariolicum*, aislado desde la región volcánica de Campi Flegrei (Región de Campania, Italia). La cepa tipo (única existente) es *M. fumariolicum* SoIV, la que cuenta con 2 genomas secuenciados uno sin finalizar y otro completado. Esta bacteria tiene un óptimo de crecimiento a 55 °C, a pH 2,0, utiliza el ciclo de Calvin-Benson-Bassham para la fijación de carbono, y obtiene energía mediante la oxidación de metano en condiciones aeróbicas (Khadem *et al.*, 2012).
- ii) *Methylacidiphilum kamchatkense*, aislado desde fuentes termales en la región de Kamchatka, Rusia. La cepa tipo (y única existente) es *M. kamchatkense* Kam1, la que cuenta con un solo genoma secuenciado sin finalizar. Esta bacteria tiene un óptimo de crecimiento a temperaturas entre 55 °C- 60 °C y pH entre 2,0 – 3,5. Además, utiliza el ciclo de Calvin-Benson-Basshan para fijación de carbono (Erikstad y Birkeland, 2015).
- iii) *Methylacidiphilun inferorum*, aislado desde las fuentes termales en “Hell’s Gate”, Nueva Zelanda. La cepa tipo es *M. inferorum* V4, la que cuenta con un genoma completamente secuenciado. Esta especie tiene un óptimo de crecimiento a 60 °C y pH entre 2,0 - 2,5. También utiliza el ciclo de Calvin-Benson-Basshan y se le han determinado genes para la fijación de nitrógeno desde el ambiente o amonio (Hou *et al.*, 2008).

El grupo correspondiente al género propuesto como *Methylacidimicrobium*, clasificación no validada de forma taxonómica hasta la fecha, la que corresponde a un clado de organismos mesófilos que necesitan lantánidos para crecer fijando carbono mediante el ciclo de Calvin. Para este género encontramos 2 especies sugeridas que poseen algún genoma en estado de borrador:

- i) *Verrucomicrobium* sp. 3C**, aislado en suelos Solfatara de Pozzuoli, Italia. La cepa tipo es *Verrucomicrobium* sp. 3C y cuenta con un genoma sin finalizar. Tiene un óptimo de crecimiento a 35 °C y pH de 1,5 – 3,0. No existen estudios detallados sobre su metabolismo central (van Teeseling *et al.*, 2014).
- ii) *Verrucomicrobia bacterium strain LP2A***, aislado en suelos Solfatara de Pozzuoli, Italia. Esta cepa cuenta con un genoma sin finalizar descrito como *Verrucomicrobia bacterium* LP2A. Presenta un crecimiento óptimo a 37 °C y pH de 3,0. No existen estudios detallados sobre su metabolismo central (van Teeseling *et al.*, 2014).

La existencia de organismos acidófilos del filo Verrucomicrobia con genomas secuenciados, presenta la oportunidad de realizar estudios exhaustivos, utilizando herramientas bioinformáticas sobre los mecanismos que éstos utilizan para crecer de manera óptima en pH extremadamente bajo en conjunto con la evolución de estos acidófilos de un filo profundamente enraizado en el árbol de la vida.

2. HIPÓTESIS

El análisis bioinformático (*in silico*) de la información genómica nos permite proponer un modelo evolutivo genómico sobre el origen de la acidofilia en el filo **Verrucomicrobia**.

3. OBJETIVO GENERAL

Proponer un modelo de evolución genómica integrado que nos entregue información sobre el origen de la propiedad acidófila en representantes del filo **Verrucomicrobia**.

4. OBJETIVOS ESPECÍFICOS

1.- Realizar análisis filogenéticos y filogenómicos para comprender la relación entre los miembros del filo **Verrucomicrobia** proponiendo una taxonomía definitiva para éstos.

2.- Predecir propiedades funcionales de representantes acidófilos del filo **Verrucomicrobia** utilizando genómica comparativa.

3.- Proponer un modelo evolutivo integrado para los miembros de **Verrucomicrobia** con el fin de descifrar el origen de la acidofilia en este grupo.

5. METODOLOGÍA

5.1. Recopilación de genomas y análisis de calidad.

Se utilizó la base de datos NCBI *taxonomy* con fecha de abril de 2018 para identificar todos los organismos del filo Verrucomicrobia que contaran con genomas secuenciados y especies aisladas (con el objetivo de obtener genomas de mayor calidad con relación a los provenientes de muestras ambientales o metagenómicas) (Federhen, 2011). Los organismos que cumplieran con estos parámetros (genomas asociados a especies aisladas) fueron descargados desde la base de datos ubicada en el ftp de *Refseq* (la cual cuenta con genomas y diferentes anotaciones) (Pruitt *et al.*, 2006). Adicionalmente se incluyó un genoma correspondiente al filo Lentisphaerae para su utilización como grupo externo al filo en estudio (**Tabla 1**). Para cada uno de los organismos fueron descargados los archivos correspondientes al genoma (secuencia a nivel nucleotídico) y las proteínas predichas anotadas (secuencias a nivel aminoacídicos). Durante este trabajo se utilizará el término gen al trabajar con secuencias a nivel nucleotídico y proteína al trabajar con la secuencia predicha a nivel aminoacídico de cada gen. La calidad general, completitud y contaminación de los genomas fue estimada utilizando el programa checkM versión 1.0.6 con parámetros predeterminados (Parks *et al.*, 2015).

<i>Nombre organismo</i>	<i>Estado ensamble</i>	<i>Refseq</i>	<i>Tamaño (Mb)</i>	<i>GC (%)</i>
<i>Akkermansia glycaniphila</i>	Completo	NZ_LT629973	3.07	57.6
<i>Akkermansia muciniphila</i> ATCC BAA-835	Completo	NC_010655	2.66	55.8
<i>Akkermansia muciniphila</i> BSH05	Borrador	GCF_002885105	2.69	57.8
<i>Akkermansia muciniphila</i> YL44	Completo	NZ_CP015409	2.75	55.7
<i>Candidatus Xiphinematobacter</i> sp. Idaho Grape	Completo	NZ_CP012665	0.92	47.7
<i>Cephalotococcus capnophilus</i>	Borrador	NZ_LSZP00000000	2.09	60.5
<i>Cephalotococcus primus</i>	Borrador	NZ_LSZQ00000000	2.36	63
<i>Chthoniobacter flavus</i> Ellin428	Borrador	NZ_ABVL00000000	7.85	61.1
<i>Coraliomargarita akajimensis</i> DSM 45221	Completo	NC_014008	3.75	53.6
<i>Didymococcus colitermitum</i> TAV2	Borrador	NZ_ABEA00000000	5.67	60.9
<i>Lacunisphaera limnophila</i>	Completo	NZ_CP016094	4.2	66.5
<i>Lentisphaera araneosa</i> HTCC2155	Borrador	NZ_ABCK00000000	6.02	41
<i>Methylococcus furax</i> SolV	Completo	NZ_LM997411	2.48	41.5
<i>Methylococcus inferorum</i> V4	Completo	NC_010794	2.29	45.5
<i>Methylococcus kamchatkense</i> Kam1	Borrador	NZ_JQNX00000000	2.21	40.4
<i>Opitutaceae bacterium</i> TAV3	Borrador	NZ_LXWT00000000	5.62	60.6
<i>Opitutaceae bacterium</i> TAV4	Borrador	NZ_LXWU00000000	5.61	60.7
<i>Opitutaceae bacterium</i> TSB47	Borrador	NZ_LRRQ00000000	7.82	62.9
<i>Opitutaceae</i> sp. TAV1	Borrador	NZ_AHKS00000000	7.07	63.2
<i>Opitutaceae</i> sp. TAV5	Completo	NZ_CP007053	7.32	63.6
<i>Opitutus</i> sp. GAS368	Completo	NZ_LT629735	4.15	65.7
<i>Opitutus terrae</i> PB90-1	Completo	NC_010571	5.96	65.3
<i>Pedosphaera parvulla</i> Ellin514	Borrador	NZ_ABOX00000000	7.41	52.6
<i>Prostheco bacterium</i> debontii ATCC 700200	Borrador	NZ_FUYE00000000	6.29	56.4
<i>Rubritalea marina</i> DSM 17716	Borrador	NZ_ARJP00000000	3.01	51.6
<i>Terrimicrobium saccharophilum</i> strain NM-5	Borrador	NZ_BDCO00000000	4.75	60.2
<i>Verrucomicrobia bacterium</i> SbV1	Borrador	NZ_OKRK00000000	8.51	59.7
<i>Verrucomicrobia bacterium</i> SCGC	Borrador	NZ_CAGO00000000	4.85	46.4
<i>Verrucomicrobia bacterium</i> AAA164-O14	Borrador	GCA_000264605.3	3.35	48.4
<i>Verrucomicrobia bacterium</i> AAA164-L15	Borrador	GCA_000285795.3	2.64	48.4
<i>Verrucomicrobium</i> sp. 3C	Borrador	GCF_000379365	2.77	60.8
<i>Verrucomicrobia bacterium</i> LP2A	Borrador	NZ_JAFS00000000	2.47	62.7
<i>Verrucomicrobium</i> sp. BvORR034	Borrador	GCA_000739635	8.57	60.5

Tabla 1 | Genomas de Verrucomicrobia. Nombre del organismo según *NCBI taxonomy*, estado de completitud del ensamblaje, número de acceso base de datos *Refseq*, tamaño en mega pares de bases, contenido de GC como porcentaje.

5.2. Obtención de marcadores.

Las inferencias filogenéticas fueron realizadas utilizando diferentes marcadores. Para todos los genomas en estudio se seleccionaron secuencias completas (o casi completas) del gen ARN ribosomal 16S (tamaños mayores a 1,300 pares de bases) desde la base de datos SILVA (High Quality Ribosomal RNA Database; <https://www.arb-silva.de/>) accedida en abril del año 2018 (Quast C. et al., 2013). El set de marcadores denominado 31cic fue identificado en cada uno de los genomas utilizando modelos ocultos de markov (HMMs). Este set de marcadores corresponde a 31 proteínas de copia única que se encuentran universalmente conservadas en el dominio Bacteria según lo descrito por Ciccarelli et al. (2006). Un *bitscore* de 60 fue seleccionado como línea de corte para la identificación de cada uno de los marcadores. Estos análisis fueron realizados utilizando el programa FetchMG versión 1.0 (Sunagawa et al., 2013).

Finalmente, se identificó un set de 120 proteínas (bac120) descritos por Rinke et al. (2017) en cada uno de los genomas. Secuencias revisadas de Uniprot (The UniProt Consortium, 2018) fueron utilizadas para entrenar perfiles de HMMs mediante una modificación del programa FetchMG, los valores de corte generados por esta calibración fueron utilizados como corte para la identificación de los marcadores.

5.3. Construcción de árboles filogenéticos.

Cada set de marcadores (gen ARN ribosomal 16S, 31cic y bac120) fueron analizados de manera independiente mediante el siguiente protocolo: Las secuencias fueron alineadas con *MAFFT* versión 7.407 (Kato y Standley, 2013) utilizando la opción *linsi* (opción que disminuye la velocidad del proceso, pero entrega mayor precisión en los alineamientos realizados) y parámetros predeterminados. Los alineamientos fueron observados y curados de forma manual con el programa *Aliview* versión 1.25 (Larsson, 2014). Las regiones conservadas en los alineamientos múltiples fueron seleccionadas con el programa *GBLOCKS* versión 0.91b (Talavera y Castresana, 2007) utilizando parámetros predefinidos (Ciccarelli *et al.*, 2006). Las secuencias asociadas a cada uno de los grupos de proteínas fueron concatenadas utilizando un *script in house*. Los modelos de sustitución (nucleotídica o aminoacídica según correspondiera en cada grupo) fueron definidos mediante *Modelfinder* incluido dentro de *IQ-Tree* (Kalyaanamoorthy *et al.*, 2017). Finalmente, los árboles filogenéticos y filogenómicos fueron construidos en base al método de máxima verosimilitud utilizando el programa *IQ-Tree* versión 1.6.8 (Nguyen *et al.*, 2014), seleccionando

los modelos de sustitución adecuados en base a los resultados entregados por *Modelfinder* y con un método de *ultrafast bootstrap* de 1000 réplicas. Los árboles resultantes fueron visualizados con *Figtree* versión 1.4.2 (Rambaut, 2012).

5.4. Análisis de similitud de secuencias.

Para proponer una clasificación taxonómica definitiva en los miembros del filo Verrucomicrobia se utilizaron diferentes tipos de análisis de similitud de secuencias como ANI y TETRA. Se calculó el valor de ANIb (Average Nucleotide Identity basado en BLAST) (Goris et al 2007). Valores sobre el 30% de identidad, alineamientos sobre el 70% de cobertura y un largo mínimo de secuencia de 1020 pares de bases fueron aplicados a los cálculos de ANIb. Las frecuencias de TETRA y coeficientes de correlación fueron calculados con los algoritmos descritos en Tetteling et al, 2004. Ambos valores fueron utilizados para hacer una clasificación a nivel de especie de los genomas en estudio. Los cálculos mencionados fueron realizados con el programa Jspecies versión 1.2.1 (Richter y Rossello 2009). Fue comparada la identidad entre las secuencias del gen ARN ribosomal 16S (obtenida mediante blastn), generando una matriz de alineamientos entre todas las secuencias en estudio. El valor de identidad de las secuencias del gen ARN ribosomal 16S fue utilizado para hacer la definición taxonómica a nivel de género y clasificaciones superiores, donde los valores de corte para género son del 94.5% y familia del 86.5% (Yarza et al., 2014). El valor de AAI (Amino Acid Identity) fue calculado utilizando *Diamond* con comparaciones en pares utilizando el mejor hit bidireccional mediante el programa *CompareM* versión 0.0.21. Valores de corte para género fueron entre 65 y 95% y familia entre 45% y 65% (Konstantidinis, 2017).

5.5. Predicción de elementos móviles.

Los elementos de móviles fueron identificados usando un *pipeline* de anotación en Bioperl (*in-house pipeline*, Center for Bioinformatics and Genome Biology, Fundación Ciencia & Vida, Santiago, Chile) con los parámetros por defecto (Ullrich *et al.*, 2016). Para la identificación específica de transposones fue utilizada la base de datos *TnpPred* accedida en julio del año 2018 (Riadi *et al.*, 2012). La base de datos *ISsaga* versión 2.0 fue utilizada para la detección de secuencias de inserción, utilizando valores de corte de E-value < 0.0001 (Varani *et al.*, 2011). Finalmente, utilizando *blastp* contra la base de datos *ACLAME* versión 0.4 se realizó una búsqueda

de genes asociados a virus o profagos que pudieran estar presentes en los genomas utilizando valores de $E\text{-value} < 0.0001$ como corte (Leplae *et al.*, 2009).

Secuencias que fueran identificadas por más de un programa como elementos móviles fueron seleccionadas. Adicionalmente, los resultados tanto de *TnpPred* como *ISSaga* fueron revisados de forma manual y con una nueva búsqueda mediante *blastp* contra la base de datos NR de *NCBI* para ser clasificados de forma definitiva como elemento móvil. Finalmente fueron definidos el *estabiloma* (genoma sin los elementos móviles) y el *moviloma* (elementos móviles identificados) para cada uno de los genomas (Siefert, 2009).

5.6. Anotación funcional de los genomas.

La predicción funcional de los genes fue realizada utilizando las secuencias de proteínas provenientes de la anotación entregada por *NCBI RefSeq* para los 5 genomas acidófilos (*Methylophilum fumariolicum*, *Methylophilum kamchatkense*, *Methylophilum inferorum*, *Verrucomicrobium* sp. 3C, *Verrucomicrobia bacterium* LP2A) y el grupo externo (*Pedosphaera parvula*). Se utilizó la base de datos COG (*Clusters of Orthologous Groups of proteins*) (Tatusov *et al.*, 2003), en su versión actualizada el año 2014. La clasificación fue realizada mediante la comparación de las proteínas con la base de datos usando *blastp* con un valor de corte de $E\text{-value} < 0.0001$. La asociación a una categoría de COG fue utilizando el *hit* con mayor cobertura usando un *script in house* basado en Bioperl (Stajich *et al.*, 2002). Los valores para cada una de las categorías fueron normalizados con relación al número de proteínas totales por genoma para obtener el porcentaje correspondiente a cada una de las categorías por genoma.

5.7. Identificación de familias de proteínas y definición del pangenoma.

La determinación de familias de proteínas entre los 5 organismos acidófilos del filo *Verrucomicrobia* fue realizado utilizando todas las proteínas del *estabiloma* (de esta forma excluyendo los elementos móviles identificados). Las proteínas seleccionadas fueron ordenadas utilizando un *script in house* basado en *blastp* comparando todas las proteínas de cada genoma entre sí con el método de *Bidirectional Best Blast Hit* (BBBH). Las familias de proteínas fueron construidas mediante la regla de 50/50 (correspondiente a alineamientos con valores sobre el 50% de identidad y 50% de cobertura) y cada proteína fue asignada a una familia (Snipen y Ussery,

2010). Las familias de proteínas fueron clasificadas en core-genoma, genoma-dispensable y genoma-único de acuerdo con su distribución en los genomas en estudio. El core-genoma corresponde a las familias de proteínas que son compartidas por todos los genomas en estudio, el genoma dispensable corresponde a las familias de proteínas asignadas a más de una especie, pero no compartidas en todas y el genoma único corresponde a las familias de proteínas asignadas a un solo genoma. La suma de todos estos grupos corresponde al pangenoma (unión de todas las familias de proteínas en estudio; Tettelin et al., 2005).

5.8. Predicción y análisis de propiedades acidófilas en genomas.

Se construyó una base de datos de proteínas asociadas a mecanismos de resistencia a pH bajo basado en una búsqueda bibliográfica. Las proteínas fueron categorizadas en sistemas de resistencia a acidez canónicos (21 proteínas), sistemas de chaperonas o reparación de ADN (10 proteínas), isla génica de resistencia a acidez (4), sistema de desaminasas y ruptura de urea (12 proteínas), biosíntesis de hopanoides (14 proteínas) y transportadores de cationes inorgánicos (8 proteínas) (**Tabla suplementaria 1**). Las proteínas fueron identificadas en los genomas en estudio mediante *blastp* utilizando como valor de corte un *E-value* < 0.0001 y posteriormente fueron confirmados de manera individual utilizando *blastp* contra la base de datos NR de *NCBI*. Finalmente, fueron identificadas dentro de las familias de proteínas del pangenoma, es decir, en core-genoma, genoma-dispensable o genoma único.

5.9. Predicción de la historia evolutiva del grupo acidófilo del filo Verrucomicrobia.

Se observó el contexto genético de cada una de las proteínas de interés utilizando el programa *Artemis* (Brettin et al. 2015) y se identificaron señales de transferencia horizontal de genes (presencia de secuencias asociadas a elementos móviles). La conservación de regiones genómicas, asociadas a las proteínas de resistencia a pH bajo fueron analizadas con el programa *MAUVE* (Darling et al., 2010) (alineamientos progresivos de todos los genomas en estudio). Predicción de *cluster* de genes fueron realizadas utilizando la base de datos *DOOR2* (Mao et al., 2013).

Una propuesta del origen de las proteínas de interés identificadas, las cuales estaban asociadas a transferencia horizontal de genes, fue realizada mediante una curación manual de los resultados de *blastp* contra la base de datos de NR. Los posibles donadores fueron identificados sólo para los casos cuyas proteínas presentaban similitud con secuencias de filos diferentes a Verrucomicrobia.

Una matriz de presencia y ausencia de genes fue construida, basado en la información de las familias de proteínas identificadas en el pangenoma mediante un *script in house*. Una predicción, basada en la máxima parsimonia para la ganancia y pérdida de genes fue realizada mediante el programa *GLOOME* utilizando el árbol generado con los marcadores bac120 como referencia (Cohen *et al.*, 2010). Fueron analizadas de manera particular las familias de proteínas asociadas a las proteínas de resistencia a pH bajo, observando en qué nodos sucedieron eventos de ganancia y pérdida, con el fin de construir su historia evolutiva. Eventos de duplicación fueron propuestos basados en las proteínas identificadas en múltiples copias en cada una de las familias de proteínas y confirmando su presencia de manera única en el organismo mediante *blastp* contra la base de datos NR de *NCBI* utilizando como valor de corte un E-value < 0.0001. La fusión de proteínas relevantes fue analizada mediante alineamiento múltiple entre todas las secuencias identificadas utilizando el programa *MAFFT* con opción *linsi* y parámetros predeterminados. Los alineamientos fueron visualizados mediante *Aliview* y se analizó la conservación de dominios claves o posiciones específicas entre las secuencias en estudio.

6.- RESULTADOS

6.1. Análisis filogenéticos, filogenómicos y de similitud de secuencias del filo Verrucomicrobia.

6.1.1. Determinación de calidad de los genomas.

Los 33 genomas descargados desde *NCBI* fueron analizados para determinar su calidad general, nivel de contaminación y completitud con el programa *Checkm* (**Tabla 2**). Un genoma asociado al grupo externo del filo *Lentisphaerae* fue seleccionado y un total de 27 genomas asociados al filo *Verrucomicrobia* fueron seleccionados al contar con una completitud mayor a 90%, contaminación menor al 10% y calidad general sobre 50. En los 27 genomas seleccionados de *Verrucomicrobia* se encuentran representantes de todas las subdivisiones identificadas por estudios filogenéticos tradicionales, excluyendo organismos de las subdivisiones 5 y 7, actualmente clasificadas de manera oficial como otros filos. El mayor número de genomas fueron identificados en la subdivisión número 4 con 12 genomas y el menor en la subdivisión número 3 con 1 genoma. Se identificaron 5 genomas de organismos acidófilos (según descripción en literatura), los cuales no cuentan con una clasificación taxonómica a nivel de subdivisión en grupos tradicionales. Fueron descartados 5 genomas por no cumplir con los parámetros de corte: un genoma por obtener un valor menor a 50 en calidad general (*Verrucomicrobia bacterium* SbV1) y 4 genomas por tener una completitud menor a 90 (*Akkermansia muciniphila* BSH05 y *Verrucomicrobia bacterium* SCGC/AAA164-014/AAA164-L15).

<i>Nombre organismo</i>	<i>Compleitud</i>	<i>Contaminación</i>	<i>Calidad General</i>
<i>Akkermansia glycaniphila</i>	91,84	4,76	68,04
<i>Akkermansia muciniphila</i> ATCC BAA-835	97,96	0	97,96
<i>Akkermansia muciniphila</i> BSH05	86,85	0,68	83,45
<i>Akkermansia muciniphila</i> YL44	97,96	0	97,96
<i>Candidatus Xiphinematobacter</i> sp. Idaho Grape	90,86	0	90,86
<i>Cephalotococcus capnophilus</i>	97,95	0	97,95
<i>Cephalotococcus primus</i>	97,26	0	97,26
<i>Chthoniobacter flavus</i> Ellin428	97,3	2,87	82,95
<i>Coralimargarita akajimensis</i> DSM 45221	99,77	0	99,77
<i>Didymococcus colitermitum</i> TAV2	98,06	0,5	95,56
<i>Lacunisphaera limnophila</i>	99,32	2,7	85,82
<i>Lentisphaera araneosa</i> HTCC2155	97,12	5,21	71,07
<i>Methylacidiphilum fumariolicum</i> SolV	99,32	1,01	94,27
<i>Methylacidiphilum infernorum</i> V4	98,65	0,34	96,95
<i>Methylacidiphilum kamchatkense</i> Kam1	98,42	1,01	93,37
<i>Opitutaceae bacterium</i> TAV3	94,01	0,34	92,31
<i>Opitutaceae bacterium</i> TAV4	92,98	0,48	90,58
<i>Opitutaceae bacterium</i> TSB47	97,95	4,32	76,35
<i>Opitutaceae</i> sp. TAV1	98,63	0,23	97,48
<i>Opitutaceae</i> sp. TAV5	100	0,23	98,85
<i>Opitutus</i> sp. GAS368	99,32	3,38	82,42
<i>Opitutus terrae</i> PB90-1	99,32	3,42	82,22
<i>Pedosphaera parvula</i> Ellin514	99,32	2,7	85,82
<i>Prostheco bacter debontii</i> ATCC 700200	100	1,36	93,2
<i>Rubritalea marina</i> DSM 17716	100	0,68	96,6
<i>Terrimicrobium sacchariphilum</i> strain NM-5	100	1,36	93,2
<i>Verrucomicrobia bacterium</i> SbV1	93,24	9,76	44,44
<i>Verrucomicrobia bacterium</i> SCGC	84,29	35,49	-93,45
<i>Verrucomicrobia bacterium</i> AAA164-O14	61,29	2,59	48,34
<i>Verrucomicrobia bacterium</i> AAA164-L15	52,71	8,44	10,51
<i>Verrucomicrobium</i> sp. 3C	97,93	0,68	94,53
<i>Verrucomicrobia bacterium</i> LP2A	98,65	0,68	95,25
<i>Verrucomicrobium</i> sp. BvORR034	100	4,9	75,5

Tabla 2. Genomas de Verrucomicrobia y valores de calidad. Nombre del organismo según *NCBI taxonomy*, porcentaje de completitud, nivel de contaminación y calidad general de los genomas.

6.1.2. Marcadores filogenéticos y filogenómicos.

Todos los genomas en estudio contaban con el gen ARN ribosomal 16S con un tamaño superior a 1300 pares de bases. Para el set de proteínas conservadas 31cic, el número total de marcadores utilizados para la construcción del árbol filogenómico fue de 26 marcadores, correspondiente a los compartidos por todos los organismos en estudio (**Tabla suplementaria 2**). Para el set proteínas conservadas bac120, el número final de marcadores utilizados en todos los genomas fue de 44 (**Tabla suplementaria 3**). Para ambos casos el número final de marcadores utilizados corresponde a los marcadores compartidos por todos los organismos en estudio. El marcador *atpB* (ATP sintasa subunidad beta) fue descartado debido a encontrarse en más de una copia en varios genomas del filo. Un resumen del número de marcadores identificados en cada uno de los genomas se encuentra en la **Tabla 3**.

6.1.3. Árboles filogenéticos y filogenómicos.

El primer árbol filogenético fue construido utilizando el gen ARN ribosomal 16S (**Figura 3**). La topología obtenida muestra a la subdivisión 1 como la más ancestral dentro de Verrucomicrobia. La subdivisión número 2 aparece como un clado monofilético después de la subdivisión número 1. La subdivisión 3 aparece posteriormente con el genoma del único representante *Pedosphaera parvula*. Los organismos acidófilos se encuentran agrupados y separados en dos clados. El primer clado corresponde a *Verrucomicrobium* sp. 3C y *Verrucomicrobia bacterium* LP2A, organismos mesófilos y acidófilos. El segundo clado corresponde a *Methylacidiphilum kamchatkense*, *Methylacidiphilum infernorum* y *Methylacidiphilum fumariolicum*, organismos termófilos y acidófilos. Finalmente, aparece la subdivisión 4 agrupada también como clado monofilético. Valores de *bootstrap* bajo 70% fueron observados en las separaciones entre las subdivisiones y valores de 100% fueron observados en la separación dentro del clado acidófilo.

Los árboles filogenómicos basados en marcadores concatenados (31cic y bac120) muestran la misma topología entre sí (**Figura 4**). La subdivisión número 4 aparece como la más ancestral del filo a diferencia de lo encontrado con el árbol de ARN ribosomal 16S. Los valores de *bootstrap* en la separación de las subdivisiones presentan valores mayores en comparación al de 16S rRNA en todos los nodos. Los acidófilos mantienen su topología de dos clados divididos entre organismos mesófilos y termófilos. La única diferencia entre los árboles de ambos grupos de

marcadores (31cic y bac120) es en los valores de *bootstrap* observados en la separación entre el clado acidófilo y las subdivisiones número 1 y 2, separación que presenta un valor de soporte sobre 70 en el árbol realizado con 31cic y un soporte de 100 para el árbol realizado con bac120.

<i>Nombre</i>	<i>Silva</i>	<i>Largo 16S</i>	<i>31cic</i>	<i>bac120</i>
<i>Akkermansia glycaniphila</i>	LT629973	1502	31	98
<i>Akkermansia muciniphila</i> ATCC BAA-835	CP001071	1505	31	101
<i>Akkermansia muciniphila</i> YL44	CP015409	1516	31	102
<i>Candidatus Xiphinematobacter</i> sp. Idaho Grape	CP012665	1531	31	92
<i>Cephalotococcus capnophilus</i>	KM066003	1552	31	95
<i>Cephalotococcus primus</i>	KM066004	1552	31	93
<i>Chthoniobacter flavus</i> Ellin428	ABVL01000001	1467	31	96
<i>Coralimargarita akajimensis</i> DSM 45221	CP001998	1542	31	99
<i>Didymococcus colitermitum</i> TAV2	ABEA03000104	1539	30	91
<i>Lacunisphaera limnophila</i>	CP016094	1558	31	94
<i>Lentisphaera araneosa</i> HTCC2155	ABCK01000018	1497	31	99
<i>Methylacidiphilum fumariolicum</i> SolV	EF591088	1543	31	101
<i>Methylacidiphilum infernorum</i> V4	AM900833	1517	31	99
<i>Methylacidiphilum kamchatkense</i> Kam1	JQNX01000005	1530	29	93
<i>Opitutaceae bacterium</i> TAV3	AY587231	1491	31	95
<i>Opitutaceae bacterium</i> TAV4	LXWT01000067	1565	28	80
<i>Opitutaceae bacterium</i> TSB47	AY587234	1454	29	85
<i>Opitutaceae</i> sp. TAV1	CP007053	1564	31	95
<i>Opitutaceae</i> sp. TAV5	JQ864375	1522	31	93
<i>Opitutus</i> sp. GAS368	LT629735	1564	31	91
<i>Opitutus terrae</i> PB90-1	CP001032	1551	31	96
<i>Pedosphaera parvula</i> Ellin514	ABOX02000035	1562	30	96
<i>Prostheco bacterium debontii</i> ATCC 700200	U60014	1468	31	102
<i>Rubritalea marina</i> DSM 17716	DQ302104	1429	31	105
<i>Terrimicrobium sacchariphilum</i> strain NM-5	GU129926	1418	31	102
<i>Verrucomicrobium</i> sp. 3C	JAFS01000001	1511	31	95
<i>Verrucomicrobia bacterium</i> LP2A	ARAS01000058	1511	31	98
<i>Verrucomicrobium</i> sp. BvORR034	BATQ01000187	1519	31	102

Tabla 3. Marcadores para árboles filogenéticos y filogenómicos. Nombre del organismo, referencia base de datos SILVA, largo de la secuencia del gen ARN ribosomal 16s, número de marcadores 31cic por genoma, número de marcadores bac120 por genoma.

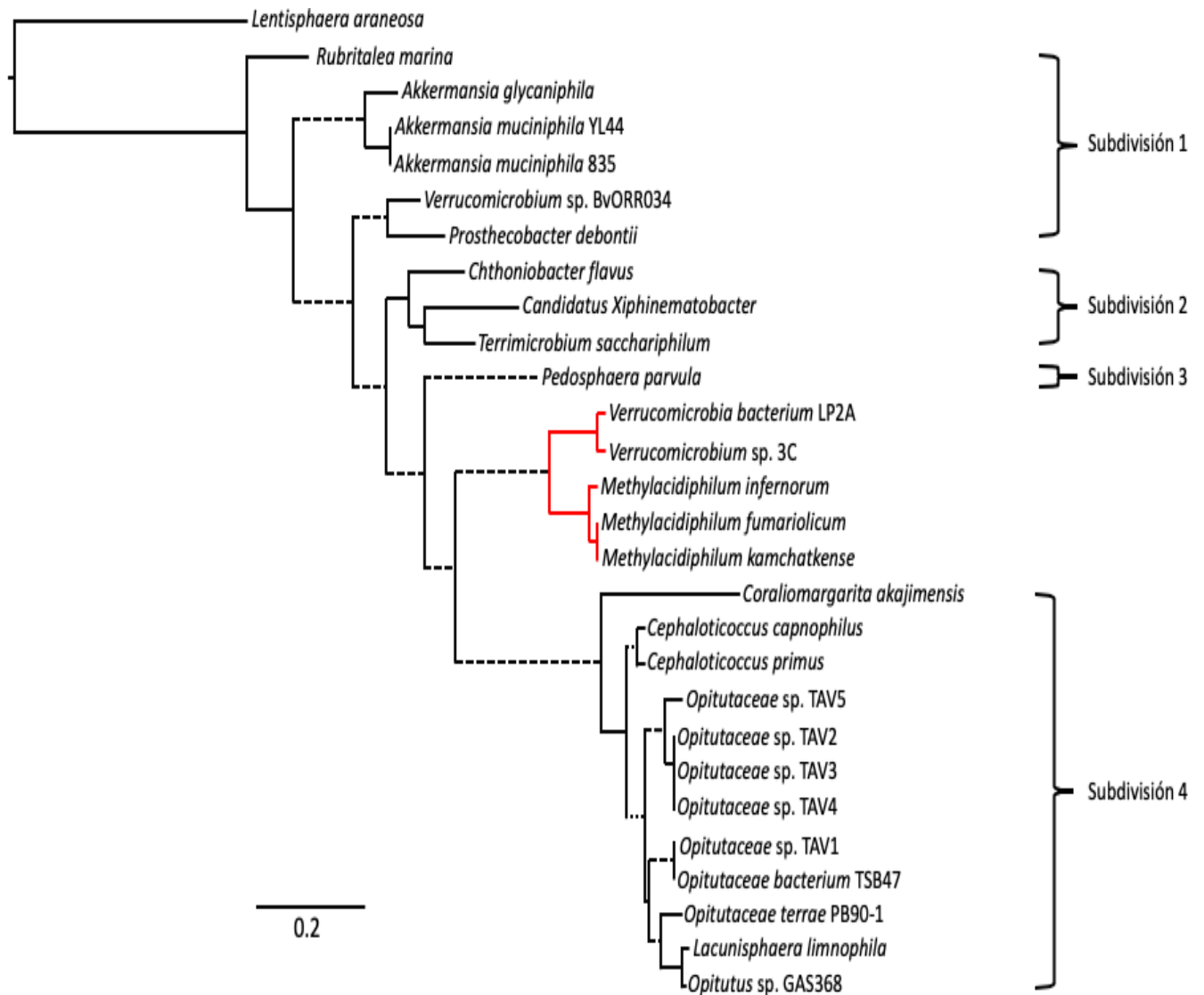


Figura 3. Árbol filogenético de 16S rRNA Verrucomicrobia. Árbol filogenético 27 genomas en estudio de *Verrucomicrobia* y el grupo externo *Lentisphaera araneosa*. Las líneas punteadas muestran ramas con soporte de *bootstrap* menor a 70. En rojo las ramas correspondientes a organismos acidófilos (sin clasificación en subdivisión tradicional). En llaves se señalan las subdivisiones según la clasificación tradicional del filo.

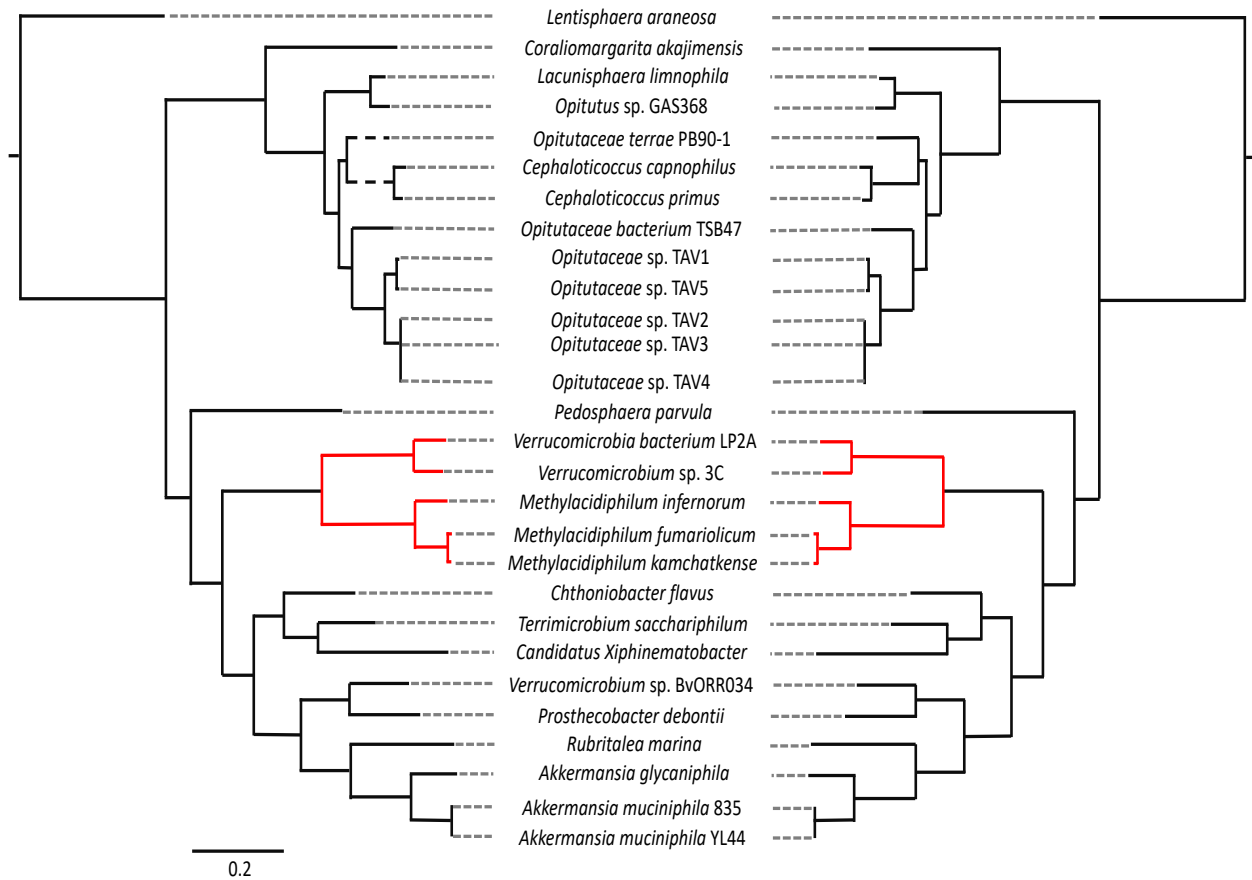


Figura 4. Árboles filogenómicos de marcadores concatenados Verrucomicrobia.

El árbol con 26 marcadores de 31ic concatenados para los 27 genomas en estudio de Verrucomicrobia y grupo externo *Lentisphaera araneosa* se muestra al costado izquierdo. El árbol con 45 marcadores de bac120 concatenados para los 28 genomas en estudio se muestra al costado derecho. Las líneas punteadas muestran ramas con soporte de *bootstrap* menor a 70. En rojo se ven las ramas correspondientes a organismos acidófilos.

6.1.4. Análisis de similitud de secuencias.

La identidad del gen ARN ribosomal 16S fue calculada mediante *blastN* y fue utilizada para proponer la clasificación taxonómica a nivel de género y familia en los organismos acidófilos de Verrucomicrobia (**Figura 5**). Ambos organismos mesófilos fueron agrupados dentro de un mismo género (*Methylacidimicrobium*) con un valor de 98% de identidad; de igual manera los tres organismos termófilos fueron agrupados dentro de un mismo género (*Methylacidiphilum*) con valores de identidad entre 98 y 99% (valor de corte para género es 94,5%). A su vez, todos los organismos acidófilos se agrupan dentro de una misma familia (*Methylacidiphilaceae*), presentando valores de identidad sobre 90% (valor de corte para familia es 86,5%) (Yarza et al., 2014) (**Tabla suplementaria 3**).

El valor de AAI fue calculado con el programa *CompareM* y utilizado para confirmar la clasificación a nivel de género de los organismos acidófilos (**Figura 5**). Los organismos mesófilos fueron agrupados en un mismo género (*Methylacidimicrobium*), al igual que con la identidad de 16S, con un valor de 79% y en un diferente género con relación a los organismos termófilos con un valor de ~56%. Los organismos termófilos fueron agrupados en un mismo género (*Methylacidiphilum*) con valores de ~94%. Los resultados de la clasificación taxonómica a este nivel fueron representados utilizando como base el árbol filogenómico generado con los marcadores bac120 (Konstantidinis, 2017).

Los valores de ANIb y TETRA fueron calculados con el programa *Jspecies* (Richter y Rossello 2009) y fueron utilizados para la clasificación taxonómica a nivel de especie (**Figura 6**). Todos los organismos acidófilos además del grupo externo fueron clasificados como especies diferentes en base al valor de ANIb bajo el valor de corte de 96% donde los genomas más cercanos fueron los de *Methylacidiphilum kamchatkense* y *Methylacidiphilum fumariolicum* con un valor de 92%. Ambos organismos fueron clasificados como una misma especie según el valor de TETRA con un valor de 0,991 el cual se encuentra sobre el valor de corte de 0,99, todos los demás genomas fueron clasificados como especies diferentes según este valor.

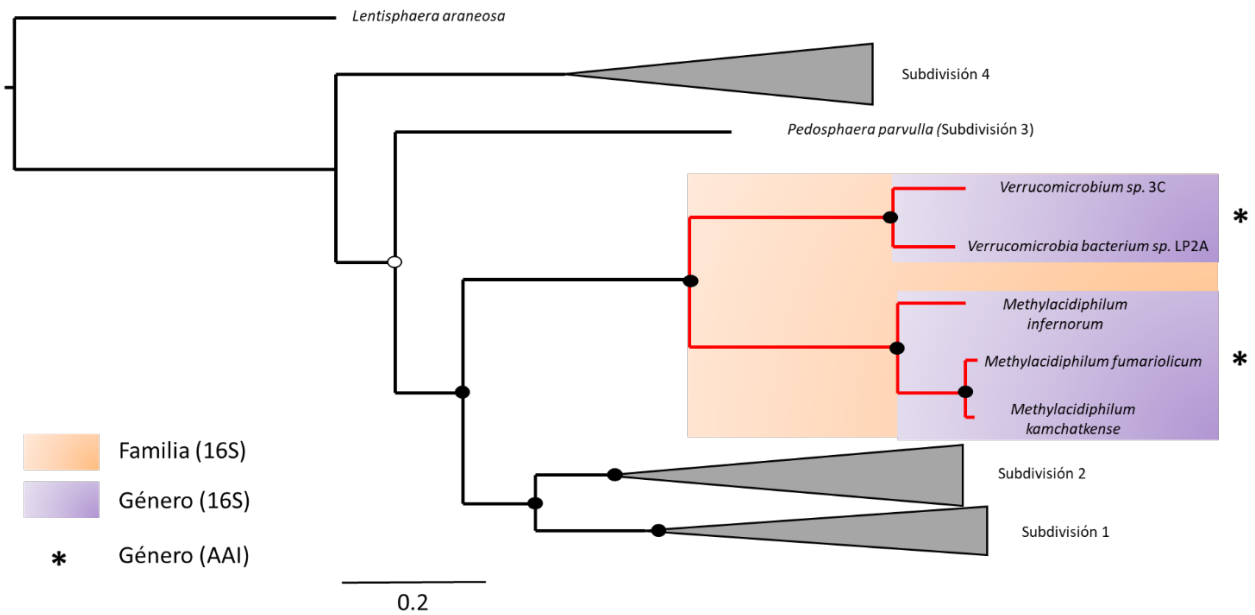


Figura 5. Clasificación taxonómica basada en similitud de secuencias. Árbol filogenómico basado en marcadores bac120. Los rectángulos de colores muestran resultados de identidad ARN ribosomal 16S (naranja para familia y morado para género). Asteriscos son predicciones de género apoyadas por AAI.



Figura 6 Heatmap de valores ANIb-TETRA. Valores de ANIb expresados como porcentaje de similitud en el rango de 60 - 100%. Los valores de TETRA sobre 0,99 son representados con asteriscos en las casillas correspondientes.

6.2. Análisis de genómica comparativa en acidófilos y grupo externo neutrófilo.

6.2.1. Predicción de elementos móviles, definición de estabiloma y moviloma.

Los elementos móviles fueron identificados en cada uno de los genomas mediante 3 bases de datos diferentes. La base de datos *TnpPred* fue utilizada para la predicción de transposasas, obteniendo como resultado 2 secuencias identificadas como transposasas en *Methyacidiphilum inferorum* (organismo con menor número identificado) y 114 secuencias identificadas para *Verrucomicrobium sp. 3C* (organismo con mayor número identificado). *ISsaga* fue utilizado para la predicción de secuencias de inserción, obteniendo como resultado 2 secuencias identificadas como mínimo en el genoma de *Methyacidiphilum kamchatkense* y 167 secuencias identificadas como máximo en el organismo *Verrucomicrobium sp. 3C*. Finalmente *ACLAME* fue utilizado para la predicción de elementos móviles asociados a plasmidio, fagos, transposasas y secuencias de inserción en general, como resultado se obtuvieron 1023 secuencias identificadas como mínimo en el genoma de *Methyacidiphilum kamchatkense* y 2966 secuencias identificadas como máximo para el organismo *Pedosphaera parvula* (**Figura 7**).

Secuencias identificadas por más de dos programas fueron seleccionadas automáticamente para definir el moviloma (conjunto de elementos móviles identificados en un genoma), el número más bajo de secuencias identificadas fue de 4 en *Methyacidiphilum fumariolicum* y el máximo identificado fue de 170 para *Verrucomicrobium sp. 3*. El estabiloma fue definido como todas las secuencias no identificadas como elementos móviles y para todos los genomas en estudio 90% del total del genoma (**Tabla 4**).

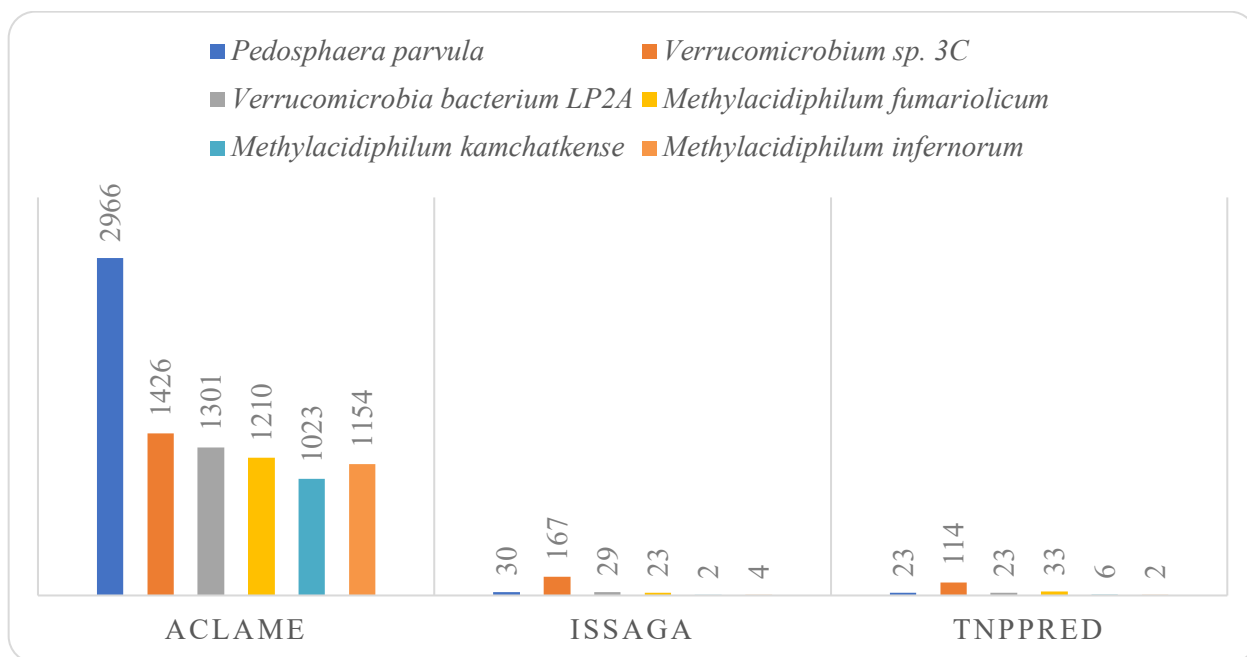


Figura 7. Predicción elementos móviles. Porcentaje por genoma de elementos móviles predichos por los programas *TnpPred*, *ISsaga* y *ACLAME*.

Organismos	Proteínas	Estabiloma	Estabiloma (%)	Moviloma	Moviloma (%)
<i>Pedosphaera parvula</i>	6100	6069	99,49	31	0,51
<i>Verrucomicrobium sp. 3C</i>	2473	2303	93,13	170	6,87
<i>Verrucomicrobia bacterium LP2A</i>	2231	2196	98,43	35	1,57
<i>Methylacidiphilum infernorum</i>	2221	2171	97,75	50	2,25
<i>Methylacidiphilum kamchatkense</i>	1851	1844	99,62	7	0,38
<i>Methylacidiphilum fumariolicum</i>	2107	2103	99,81	4	0,19

Tabla 4. Moviloma y estabiloma en organismos acidófilos. Valores totales de proteínas por genoma. Proteínas totales separadas en estabiloma y moviloma con sus porcentajes calculados en base al total de proteínas predichas en el genoma.

6.2.2. Predicción de propiedades funcionales de genes codificantes para proteínas.

Los estabilomas de los organismos acidófilos y del grupo externo fueron anotados funcionalmente utilizando la base de datos COG mediante una búsqueda de *blastp*. Al analizar los organismos acidófilos se obtuvo una distribución en la anotación funcional similar en cada una de las categorías entre los organismos acidófilos. La categoría con mayor porcentaje de proteínas asignadas, sin considerar las categorías de función desconocida o general, fue la de pared celular y membrana celular (M) con valores entre 6,9% hasta 8,5% para todos los genomas. Proteínas clasificadas como función general (R) estuvieron en un rango entre 6,26% y 7,44%. Las proteínas con función desconocida (S) se encontraron en un rango entre 2,97% y 4,3%, adicionalmente las proteínas que no fueron identificadas dentro de ninguna categoría alcanzaron rangos con un mínimo de 16,7% para *Methylacidiphilum kamchatkense* hasta 25,56% y máximo para *Verrucomicrobia bacterium* LP2A, rango en el que se vio la mayor variación entre los genomas. La anotación funcional para el organismo utilizado como grupo externo *Pedosphaera parvula*, muestra diferencias en diferentes categorías cuando es comparada con las secuencias de acidófilos. Categorías que presentan un mayor número de secuencias asignadas en *Pedosphaera parvula* son las relacionadas con movilidad (N) con una mayor proporción, 2% superior en promedio comparado con los organismos acidófilos, transducción de señales (T) donde existe una proporción 4% mayor, alcanzando un valor de 7,25% del genoma identificado dentro de esta categoría y la categoría de función general siendo la categoría con mayor porcentaje, alcanzando un 9,26%. Categorías que tienen mayor parte del genoma identificado como organismos acidófilos son, membrana y pared celular (M) con un 1,5% de asignación mayor en comparación al grupo externo y, producción de energía (C) y transportes de aminoácidos (E), presentando un perfil similar con un porcentaje mayor de 2% y 3% en acidófilos (**Figura 8**).

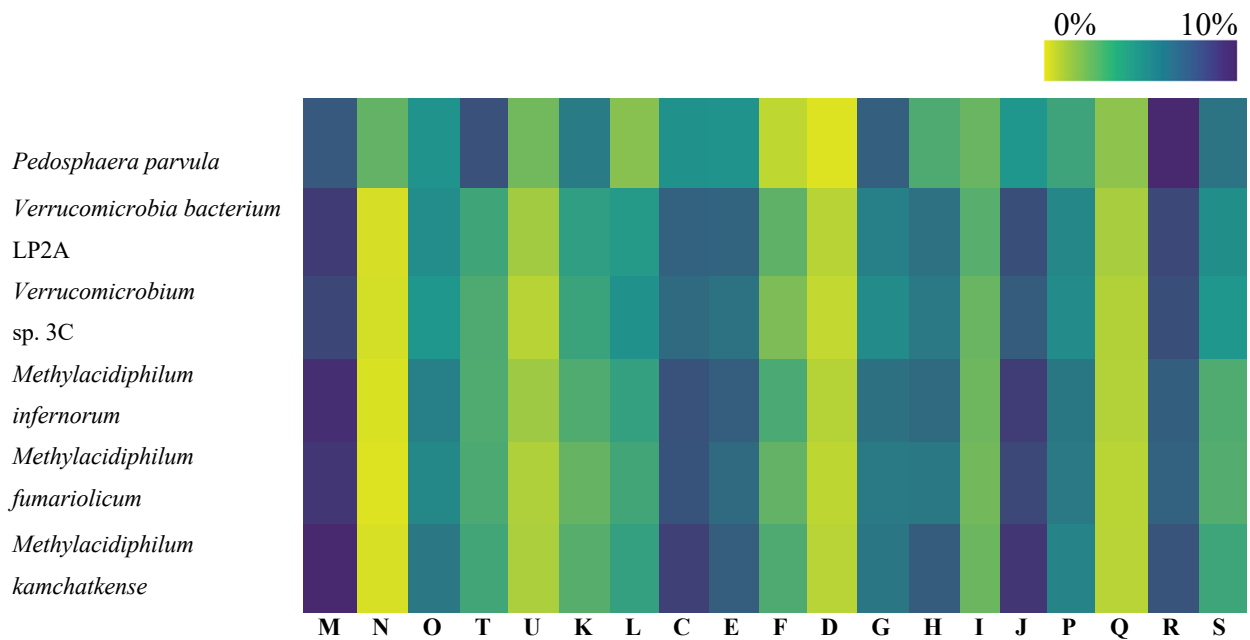


Figura 8. Heatmap distribución de COG. Porcentajes de proteínas anotadas en cada categoría funcional de COG en relación con el total de proteínas en el estabiloma de cada organismo. Categorías de COG: (M) Pared celular y membrana; (N) Movilidad; (O) Modificación post-traducciona y chaperonas; (T) Transducción de señales; (U) Tráfico intracelular y secreción; (J) Traducción; (K) Transcripción; (L) Replicación, recombinación y reparación; (C) Producción de energía; (E) Transporte de aminoácidos; (F) Transporte de nucleótidos; (G) Transporte de carbohidratos; (H) Transporte de coenzimas; (I) Transporte de lípidos; (P) Transporte de iones inorgánicos; (Q) Biosíntesis metabolitos secundarios; (R) Función general; (S) Función desconocida.

6.2.3. Identificación de familias de proteínas y definición de pangenoma.

Las familias de proteínas fueron identificadas en todos los organismos acidófilos basándose en el mejor hit bidireccional de *blastP*, utilizando la regla de 50% de identidad y 50% de cobertura en los alineamientos (Snipen y Ussery, 2010).

El pangenoma (total de familias de proteínas de todos los genomas) fue definido por un total de 4302 familias de proteínas, donde el organismo con menor número de familias fue *Methyacidiphilum kamchatkense* con 1844 familias y el con mayor número de familias fue *Verrucomicrobium* sp. 3C con 2303 familias de proteínas.

El core genoma está compuesto por un total de 801 familias de proteínas, lo cual corresponde a un total de 18,6% de las familias del pangenoma. El genoma dispensable corresponde a porcentajes desde el 19,6% hasta un 23,2% del pangenoma (dependiendo del organismo) con el menor número de familias clasificadas como genoma dispensable en *Methyacidiphilum fumariolicum* y el mayor para *Verrucomicrobia bacterium* LP2A. La mayoría de las familias de proteínas que fueron identificadas como parte del genoma dispensable son compartidas entre los organismos pertenecientes a un mismo género, 642 familias son compartidas por los tres organismos del género *Methyacidiphilum* correspondiendo a un 14,9% con relación al pangenoma y 702 familias son compartidas por los dos organismos del género *Methyacidimicrobium* correspondiente a un 16,2% con relación al pangenoma. Las familias únicas van desde el 4,1% hasta un 12,1% del pangenoma. Los dos organismos del género *Methyacidimicrobium* presentan los menores porcentajes con un 4,1% y un 5,7%, mientras los organismos del género *Methyacidiphilum* se encuentran entre un 8% hasta el máximo de 12,1% de familias identificadas como únicas (**Figura 9**).

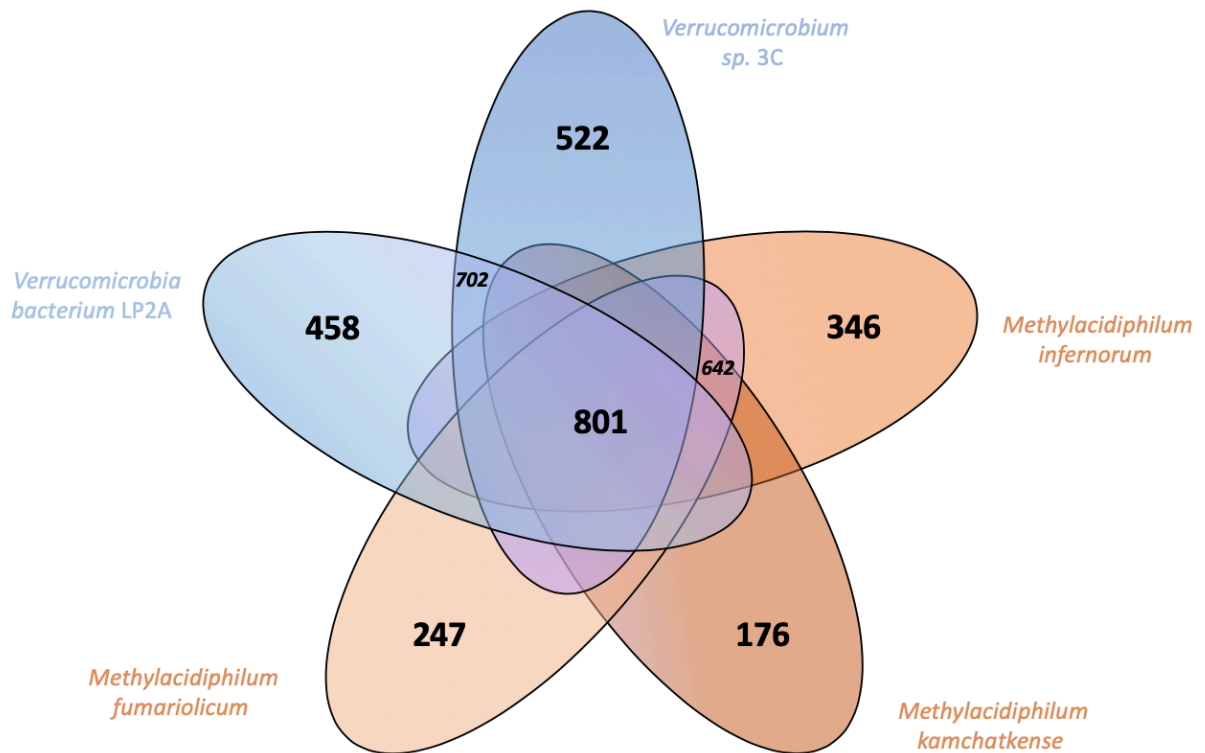


Figura 9. Diagrama de Venn del pangenoma de organismos acidófilos. Familia de proteínas totales por cada organismo se encuentran bajo los nombres respectivos. Core-genoma en el centro, familias de proteínas dispensables en la en espacios compartidos por más de un genoma y únicas en espacios no compartidos.

6.3 Predicción de proteínas relacionadas a la resistencia a pH extremadamente bajo.

6.3.1. Identificación y análisis de propiedades acidófilas en *Methylacidiphilaceae*.

Utilizando una base de datos construida a partir de datos de la literatura con proteínas relacionadas a resistencia a pH extremadamente ácido, estas proteínas fueron identificadas en las diferentes familias de proteínas dentro del pangenoma de *Methylacidiphilaceae*.

Dos sistemas de descarboxilación y antiporter fueron identificados en el core genoma de organismos acidófilos. En primer lugar, el sistema de descarboxilación de glutamato fue identificado, formado por las proteínas GadA (glutamato descarboxilasa) y GadC (glutamato/GABA antiporter), ambas proteínas fueron identificadas de forma exclusiva en los organismos acidófilos dentro del filo Verrucomicrobia. La proteína del sistema de descarboxilación de arginina SpeA (arginina descarboxilasa) fue identificada en todos los organismos acidófilos y también en *Pedosphaera parvula*. AdiC (arginina/agmatina antiporter) fue identificada solamente en los organismos acidófilos. El sistema de resistencia basado en la expulsión de protones asociado al complejo ATPasa F₀F₁ (compuesto por 8 proteínas diferentes) fue identificado en el core genoma de los organismos acidófilos, además del grupo externo neutrófilo. Una segunda copia del complejo con 8 proteínas asociadas fue identificada de forma exclusiva en todos los organismos acidófilos. Transportadores de cationes como Kch (transportador de potasio asociado a voltaje) y el antiporter NaH (antiporter de sodio/protones) fueron encontrados exclusivamente en los organismos acidófilos. El complejo transportador de potasio Kdp también fue identificado, donde las subunidades KdpA (ATPasa relacionada a potasio subunidad A) y KdpC (ATPasa relacionada a potasio subunidad C) fueron identificadas en todos los acidófilos además del grupo externo neutrófilo, pero la tercera subunidad KdpB (ATPasa relacionada a potasio subunidad B) sólo fue encontrada en el core genoma de organismos acidófilos. La mayoría de las proteínas chaperonas de la base de datos identificadas estaban en todos los organismos en estudio incluyendo a *Pedosphaera parvula*, a excepción de HupB (Proteína de unión a DNA homóloga a HU), la cual fue identificada exclusivamente en los organismos acidófilos. Para la ruta de agmatina deiminasa, la primera proteína AguA (Agmatina

deiminasa) encargada del paso de agmatina a N-Carbamoilputrescine se encuentra en todos los organismos incluyendo el grupo externo y la segunda proteína AguB (N-carbamoilputrescina amidohidrolasa), encargada al paso a putrescina, sólo fue identificada en el core genoma de los organismos acidófilos.

El complejo transportador de fosfato Pst (con sus cuatro proteínas asociadas) fue identificado de manera exclusiva en los organismos ácidosfilos. Las proteínas CynS (cianato hidratasa) y Can (anhidrasa carbónica) utilizadas en la producción de amonio para aumentar el pH citoplasmático fueron encontradas sólo en organismos acidófilos, donde adicionalmente en el organismo *Verrucomicrobium* sp. 3C se identificaron 4 copias adicionales de la proteína Can (alcanzando un total de 5 copias). La proteína de membrana Slp (lipoproteína asociada a privación de carbono) también es exclusiva de organismos acidófilos y 2 copias adicionales fueron identificadas en *Verrucomicrobia bacterium* LP2A. Finalmente, varias proteínas asociadas a la ruta de biosíntesis de hopanoides fueron identificadas, como las primeras tres proteínas fundamentales (HpnC/D/E) involucrados en el paso de (2E,6E)-farnesyl difosfato hasta escualeno. Además, fue identificada una fusión de las proteínas HpnD y HpnE en todos los organismos acidófilos. La proteína HpnF encargada de la transformación de escualeno fue identificada en todos los organismos incluyendo el grupo externo. Las proteínas encargadas de la ramificación a diferentes hopanoides (HpnH/G/O) fueron identificadas sólo en organismos acidófilos.

Un resumen de todas las proteínas asociadas a resistencia a pH extremadamente ácido identificadas en *Methylacidiphilaceae* se encuentra en la **Figura 10**.

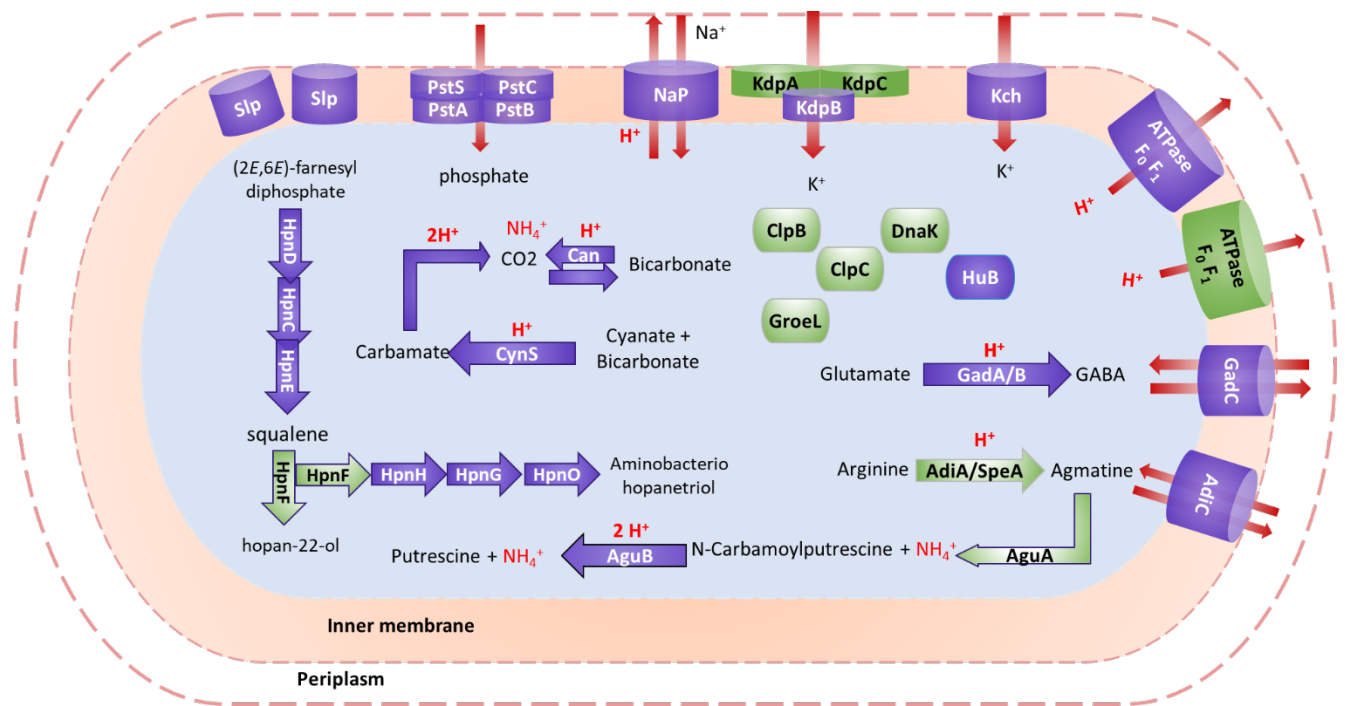


Figura 10. Mecanismos de resistencia a pH bajo identificados en *Methylophilaceae*. En morado, proteínas presentes en el core-genoma de organismos acidófilos y ausentes en el grupo externo neutrófilo. Reacciones que consumen protones y transportadores que expulsan protones indicados en rojo con el número de protones consumidos o expulsados. En verde, proteínas presentes tanto en acidófilos como en el grupo externo neutrófilo.

6.4. Reconstrucción evolutiva de propiedades acidófilas en *Methylacidiphilaceae*.

6.4.1. Transferencia horizontal de genes.

El contexto genómico de cada una de las proteínas de interés fue analizado mediante *Artemis* y *Mauve*, los eventos de transferencia horizontal fueron revisados en base a una búsqueda con *blastP* contra la base de datos *NR* de *NCBI*.

La proteína encargada de la descarboxilación de glutamato GadA se encuentra en una familia en el pangenoma de *Methylacidiphilaceae*. El contexto genético es altamente conservado para todos los organismos en la familia sin modificaciones en ambos organismos del género *Methylacidimicrobium*. En los organismos del género *Methylacidiphilum* fue identificada la inserción de dos secuencias (una L, D transpeptidasa y una nitroreductasa) separando la región analizada. En *Methylacidiphilum fumariolicum* se encuentra una inserción con una serie de proteínas no encontradas en los otros miembros. En el vecindario genómico de los acidófilos fue identificada la secuencia de ClpB, chaperona que ha sido asociada a resistencia a pH bajo, ésta a su vez también fue identificada en el grupo externo neutrófilo, pero en un contexto en el cual no existen otros genes de resistencia a acidez (**Figura 11A**).

El antiporter GadC se encuentra en una familia para todos los organismos de *Methylacidiphilaceae*. El contexto genético se encuentra conservado para los dos miembros del género *Methylacidimicrobium* y presenta una serie de señales de transferencia horizontal con secuencias de inserción cercanas en *Verrucomicrobium* sp. 3C, presencia de transposasas e integrasas en *Verrucomicrobia bacterium* LP2A. Para los organismos del género *Methylacidiphilum* también fue identificado un contexto genético altamente conservado, *Methylacidiphilum fumariolicum* presenta indicios que nos permiten proponer la transferencia horizontal con secuencias de inserción cercanas a GadC (**Figura 11B**).

El antiporter AdiC se encuentra dividido en dos familias de proteínas, una para los organismos del género *Methylacidimicrobium* los que comparten el mismo contexto genómico rodeado de una serie de proteínas hipotéticas de función desconocida. La secuencia con mayor similitud pertenece a *Hyphomicrobium denitrificans* del filo Proteobacteria. La segunda familia de proteínas está asociada al género *Methylacidiphilum*, ésta también comparte un contexto genético

entre todos sus miembros. Fueron identificadas señales de transferencia horizontal de genes en dos de los organismos, en *Methylacidiphilum infernorum* fue identificada una transposasa y en *Methylacidiphilum fumariolicum* una secuencia de inserción dividiendo la región conservada con los otros organismos (**Figura 12A**). Para la proteína AdiC en *Methylacidiphilum* la secuencia más cercana en similitud está asociada a *Matigocoleus testarum* del filo Cyanobacteria.

El complejo transportador de fosfato (Pst) fue identificado en todos los organismos de *Methylacidiphilaceae*. El contexto genómico está conservado en todos los genomas de la familia a excepción de *Methylacidiphilum fumariolicum*, en el cual fue identificada una secuencia de inserción y varias proteínas hipotéticas. En el contexto genómico para todos los organismos fue identificado adicionalmente el antiporter de sodio/protones río arriba de las secuencias asociadas al transportador de fosfato (**Figura 12B**). La subunidad PstS tiene como secuencia más cercana en similitud a *Geobacter metallireducens* del filo proteobacteria. La subunidad PstC tiene como secuencia más cercana en similitud a *Geobacter bemidjiensis* del filo proteobacteria. La subunidad PstA tiene como secuencia más cercana en similitud a *Geobacter uraniireducens* del filo proteobacteria. La subunidad PstB tiene como secuencia más cercana en similitud a *Alicyclobacillus acidocaldarius* del filo Firmicutes

El transportador de potasio asociado a voltaje Kch fue identificado en el core genoma de *Methylacidiphilaceae*. El contexto genético está conservado en todos los acidófilos sin estar presente en el grupo externo neutrófilo (**Figura 13A**). La secuencia más cercana en similitud corresponde a *Chloroherpeton thalassium* del filo Chlorobi.

La proteína CynS fue identificada en todos los organismos acidófilos, divide en dos familias de proteínas dentro del pangenoma, una de ellas asociada al género *Methylacidiphilum*, la cual cuenta con un contexto genético conservado, pero con una serie de inserciones. Para *Methylacidiphilum infernorum* fue identificada río arriba una proteína hipotética seguida a la secuencia de interés. Para *Methylacidiphilum kamchatkense* y *Methylacidiphilum fumariolicum* fue reconocida una inserción en la misma zona río arriba de CynS donde para el primer organismo, fueron identificadas una proteína hipotética y una secuencia con el dominio conservado DUF2971. En el segundo organismo fue reconocida una proteína asociada a nitroreductasa (**Figura 13B**). Para CynS en *Methylacidiphilum* la secuencia más cercana en similitud es *Reinekea blandensis* del filo Proteobacteria. En la segunda familia de proteínas asignada al género *Methylacidimicrobium* fue observado un contexto genético conservado entre ambos miembros del género diferente al del

género *Methylacidiphilum*, la secuencia más cercana en similitud es *Thiomonas delicata* del filo Proteobacteria.

Finalmente, AguA proteína capaz de utilizar agmatina (producto del sistema de descarboxilación de arginina) fue identificada en todos los organismos incluyendo el grupo externo neutrófilo. Esta secuencia se encuentra conservada entre todos los organismos. En el contexto genético para todos los organismos acidófilos fue identificada adicionalmente la proteína AguB la cual no fue identificada dentro del grupo externo acidófilo. Ambas proteínas fueron identificadas dentro de un cluster con otras proteínas hipotéticas además de la proteína de membrana Slp también utilizada como mecanismo de resistencia a pH extremadamente ácido. En los organismos de *Methylacidimicrobium* adicionalmente dentro del mismo contexto genético fue identificada la proteína HpnR asociada a la biosíntesis de hopanoides (**Figura 14**). Para AguB en ambos géneros la secuencia con mayor similitud corresponde a la encontrada en el organismo *Persephonella marina* del filo Aquificae.

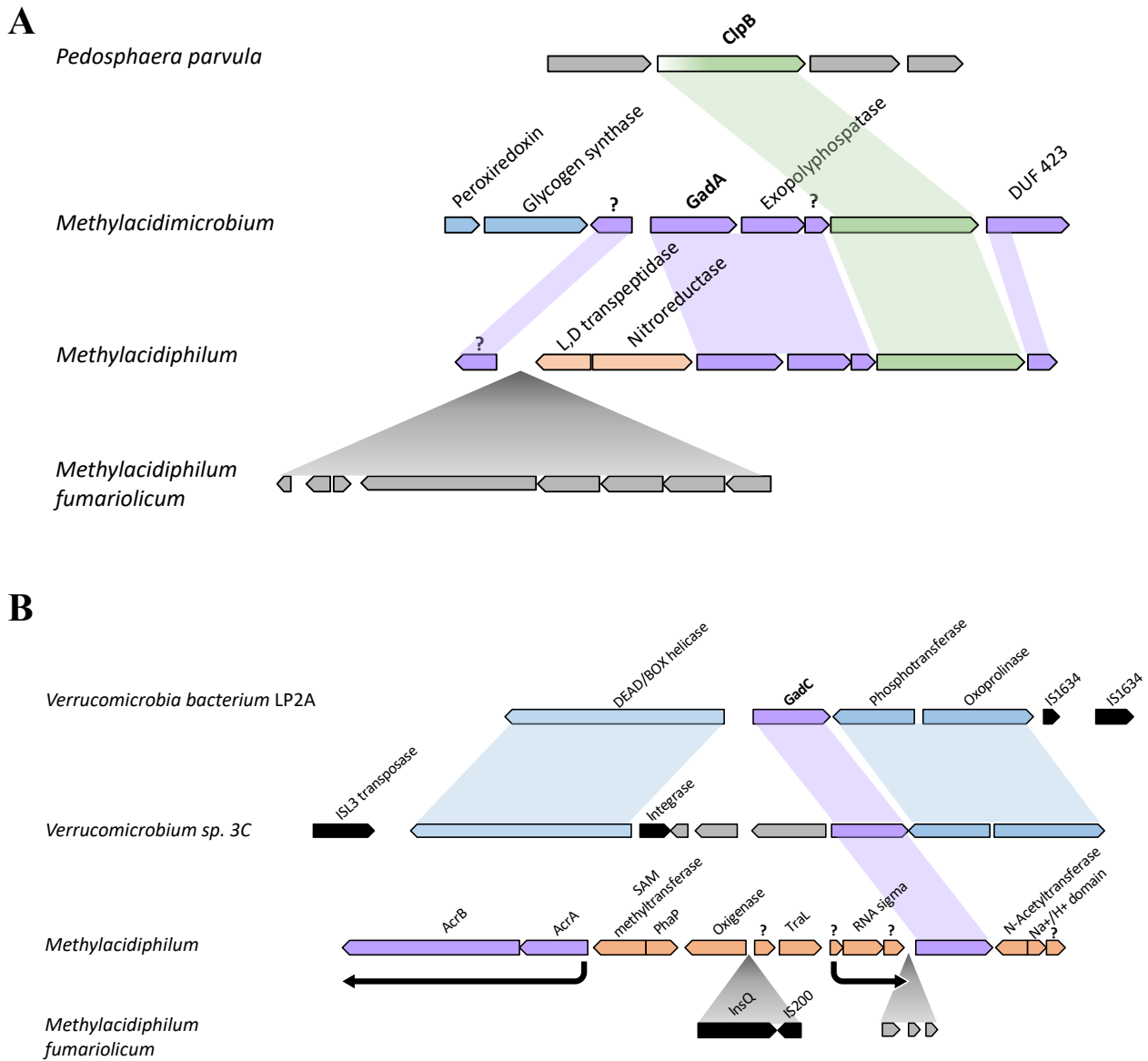


Figura 11. Contexto genómico glutamato descarboxilasa y antiporter. A. Contexto genómico de glutamato descarboxilasa (**GadA**). B. Contexto genómico glutamato/GABA antiporter (**GadC**). En color verde se muestran secuencias del core genoma acidófilos y grupo externo, en color morado core genoma de acidófilos, en azul genoma dispensable *Methylacidimicrobium*, y en naranja secuencias del genoma dispensable *Methylacidiphilum*. En color gris proteínas no conservadas. En color negro se muestran secuencias identificadas como elementos móviles. Se señala nombre de proteína, dominios conservados o proteínas hipotéticas (signo de interrogación) según fueran identificadas. Flechas negras muestran *cluster* predichos por DOOR2.

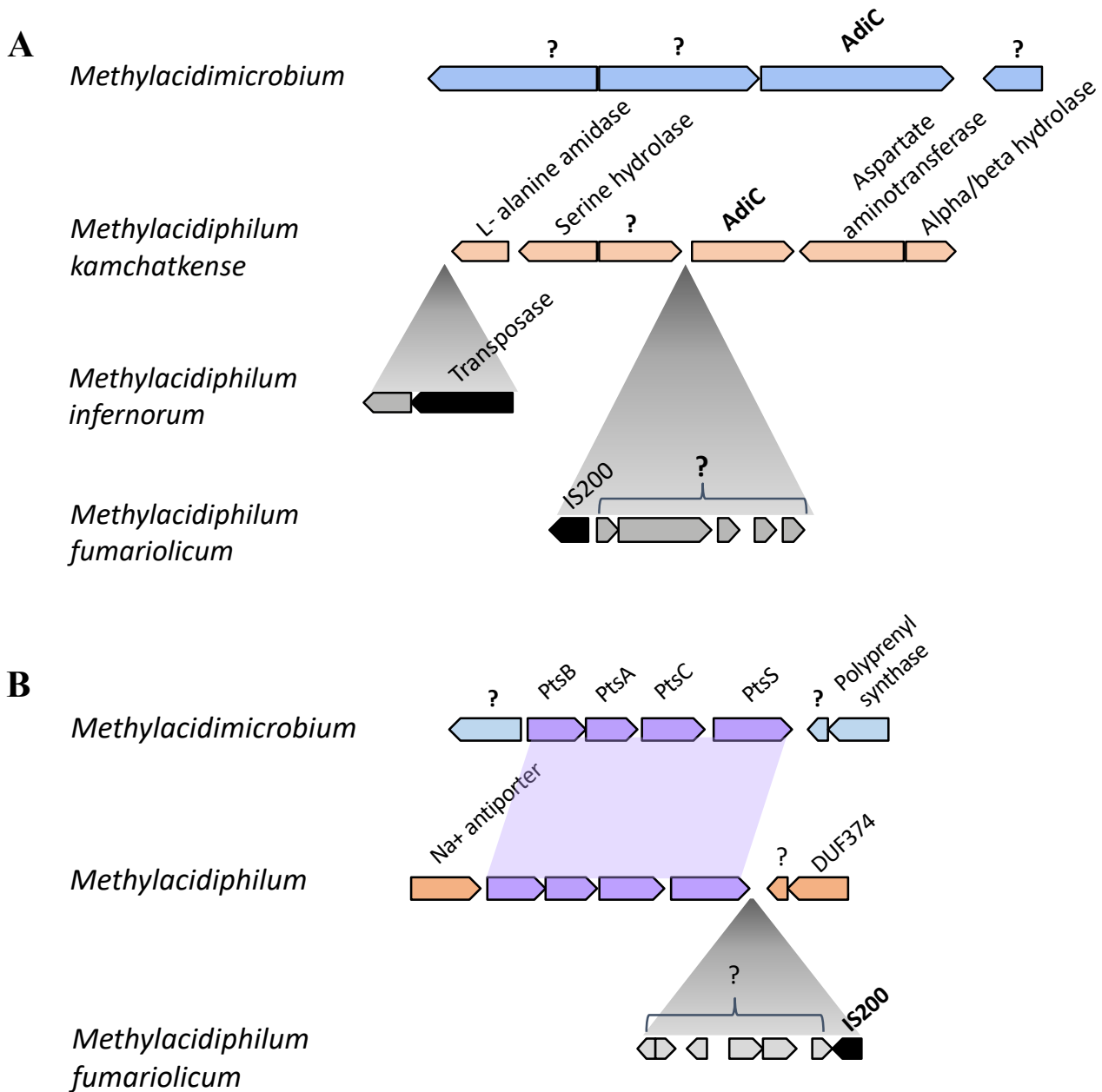


Figura 12. Contexto genómico arginina/agmatina antiporter y transportador de fosfato. A. Contexto genómico de arginina/agmatina antiporter (**AdiC**). **B.** Contexto genómico transportador de fosfato (**PstA/B/C/S**). En color morado core genoma de acidófilos, en azul genoma dispensable *Methylacidimicrobium*, y en naranja secuencias del genoma dispensable *Methylacidiphilum*. En color gris proteínas no conservadas. En color negro se muestran secuencias identificadas como elementos móviles. Se señala nombre de proteína, dominios conservados o proteínas hipotéticas (signo de interrogación) según fueran identificadas. Flechas negras muestran *cluster* predichos por DOOR2.

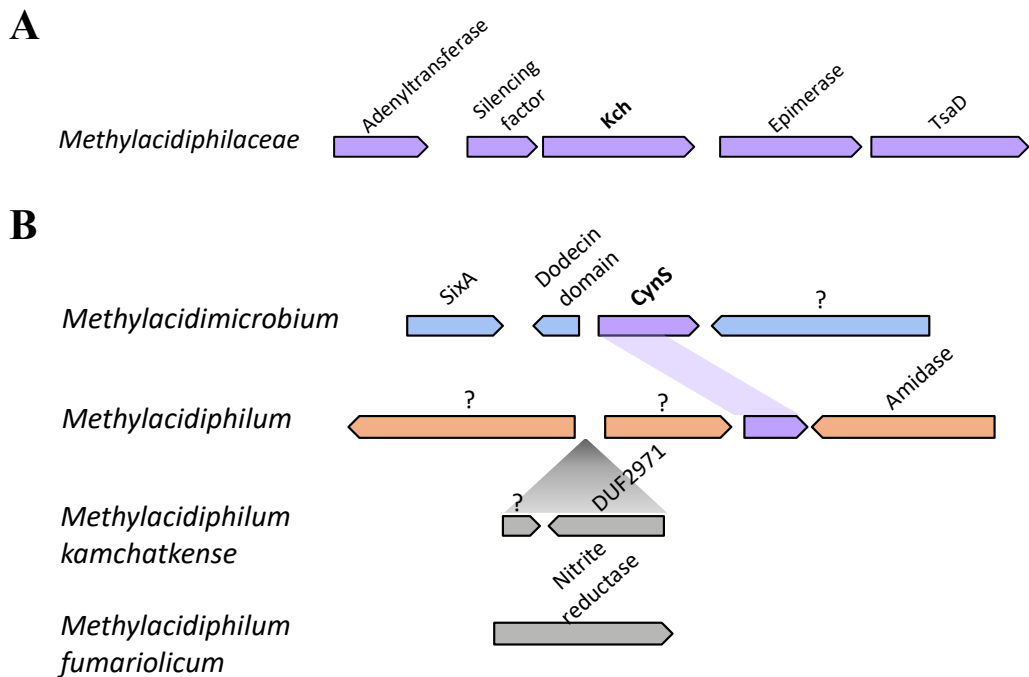


Figura 13 Contexto genómico transportador de potasio (Kch). B. Contexto genómico cianasa (CynS). En color morado core genoma de acidófilos, en azul genoma dispensable *Methylophilum*, y en naranja secuencias del genoma dispensable *Methylophilum*. En color gris proteínas no conservadas. En color negro se muestran secuencias identificadas como elementos móviles. Se señala nombre de proteína, dominios conservados o proteínas hipotéticas (signo de interrogación) según fueran identificadas. Flechas negras muestran *cluster* predichos por DOOR2.

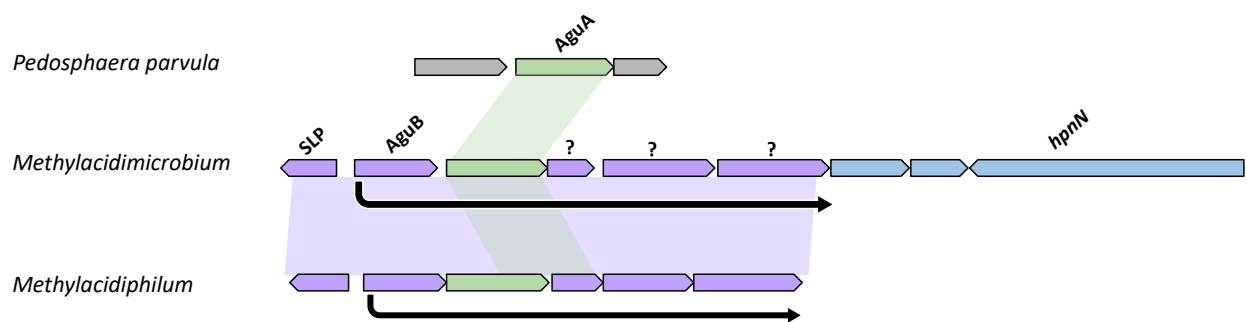


Figura 14 Contexto genómico AguA y AguB. En color morado core genoma de acidófilos, en azul genoma dispensable *Methylophilum*, y en naranja secuencias del genoma dispensable *Methylophilum*. En color gris proteínas no conservadas. En color negro se muestran secuencias identificadas como elementos móviles. Se señala nombre de proteína, dominios conservados o proteínas hipotéticas (signo de interrogación) según fueran identificadas. Flechas negras muestran *cluster* predichos por DOOR2.

6.4.2. Fusión de secuencias.

La biosíntesis de hopanoides usa una serie de enzimas para el paso desde farnesil difosfato hasta el escualeno. Los genes codificantes de estas proteínas se encuentran en un mismo contexto genético formado por las proteínas HpnC (hidroxiescualeno sintasa), HpnD (preescualeno difosfato sintasa) y HpnE (hidroxiescualeno dehydroxilasa). Dentro del filo Verrucomicrobia, las secuencias solo fueron identificadas en la familia *Methylacidiphilaceae*, no se encuentran en un clúster único teniendo un contexto conservado para HpnC, pero separado de HpnD y HpnE.

Para todos los organismos de la familia *Methylacidiphilaceae* fue detectada una fusión de HpnD y HpnE en una misma secuencia de largo cercano a 750. La primera parte de la secuencia en el intervalo desde 25 hasta 282 aminoácidos presenta el dominio de HpnD identificado con el número de acceso TIGR03465 y un valor de E-value de $7,41e-82$. La segunda parte de la secuencia en el intervalo entre 319 hasta los 739 aminoácidos presenta el dominio de HpnE identificado con el número de acceso TIGR03467 y un valor de E-value de $2,22e-104$ (**Figura 14**).

Esta fusión de proteínas fue identificada adicionalmente en un organismo del filo, *Verrucomicrobia bacterium* LW23, genoma derivado desde muestras ambientales y sin clasificación taxonómica. El único organismo fuera de los mencionados en el estudio que presenta esta fusión es *Saccharopolyspora spinosa* del filo Actinobacteria.

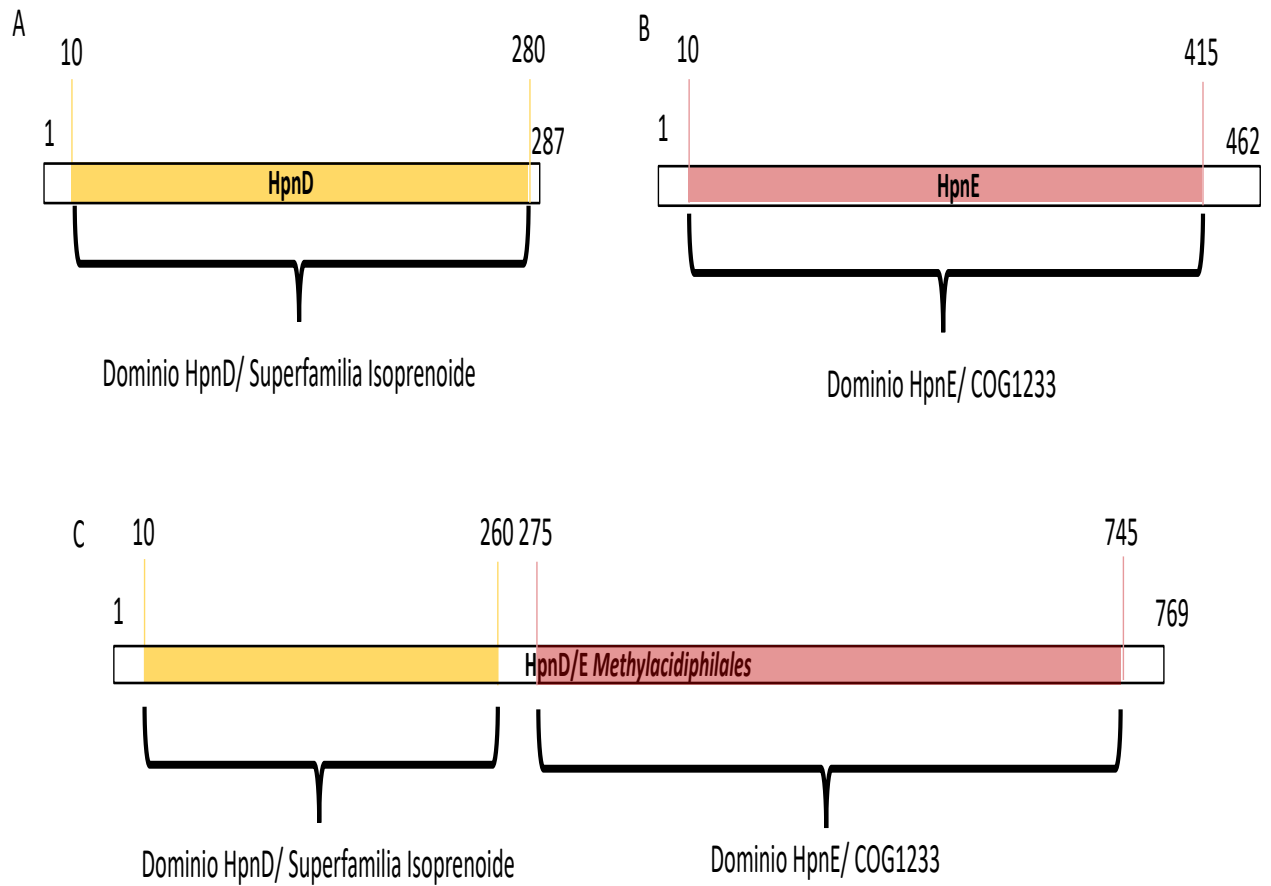


Figura 15. Fusión HpnD y HpnE en *Methylocidiphilaceae*. A. Secuencia de HpnD tradicional, en color amarillo aparece representado el dominio conservado identificado por NCBI CDD correspondiente a HpnD y la superfamilia de isoprenoides. B. Secuencia de HpnE tradicional, en color rojo aparece representado el dominio conservado identificado por NCBI CDD para HpnE y el COG1233. C. Secuencia fusionada presente en *Methylocidiphilaceae* donde se muestran regiones conservadas de HpnD y HpnE.

6.4.3. Duplicación de secuencias.

Utilizando la base de datos de proteínas asociadas a resistencia a pH extremadamente bajo y mediante un análisis de las familias de proteínas identificadas del pangenoma, fueron identificadas las diferentes copias de cada una de las proteínas.

La proteína Can fue identificada como una familia de proteínas presente en el core genoma de *Methyacidiphilaceae* y ausente en el resto de los organismos del filo. Fue identificada una copia única en el género *Methyacidiphilum* y en *Verrucomicrobia bacterium* LP2A, pero en múltiples copias dentro de *Verrucomicrobium* sp 3C. Dos copias (WP_018290803.1 y WP_018290881.1) aparecen anotadas dentro del genoma como Can y presentan un alto nivel de identidad con valores de 88 y 89% al ser comparadas con la secuencia de *Verucomicrobia bacterium* LP2A, otro miembro del género. Las otras tres secuencias (WP_018289550.1, WP_081616280.1 y WP_018290803.1) identificadas como Can, aparecen anotadas como proteínas hipotéticas y presentan valores de cobertura de 98% e identidad de 88% con las otras copias dentro del mismo genoma.

La proteína de membrana Slp fue identificada en todos los organismos acidófilos de forma exclusiva en los cuales adicionalmente fueron identificadas varias copias por genoma. En *Methyacidiphilum fumariolicum* fueron identificadas 5 copias (WP_009058835.1, WP_009060460.1, WP_009059773.1, WP_009059287.1 y WP_009061061.1) anotadas como Slp, con coberturas en su alineamiento sobre el 80% y valores de identidad entre 32 y 36%. La mayor similitud fue encontrada con secuencias del mismo género seguido de *Brenneria salicis* del filo Proteobacteria. En *Methyacidiphilum kamchatkense* fueron identificadas 2 copias (WP_079254243.1 y WP_039721964.1), secuencias que tienen una cobertura del 80% y identidad de 36% al ser comparadas, la secuencia con mayor similitud fuera de los organismos de la misma familia fue *Nitrococcus mobilis* del filo Proteobacteria. Para *Methyacidiphilum infernorum* fueron identificadas cuatro copias (WP_012464109.1, WP_079200445.1, WP_012464346.1 y WP_048810202.1) las cuales tienen niveles de cobertura desde 76 a 96% y porcentajes de identidad entre 32 y 32%, la secuencia con mayor similitud fuera de la misma familia fue *Pectobacterium punjabense* del filo Proteobacteria.

El complejo ATPasa F_0F_1 está compuesto por dos dominios diferentes: el dominio F_0 compuesto por tres subunidades (a, b y c) y el dominio F_1 compuesto por cinco subunidades diferentes (alfa, beta, delta, épsilon y gama). Una copia del *cluster* con las ocho proteínas se

encuentra altamente conservado en todos los organismos acidófilos además del organismo neutrófilo. De forma adicional fue identificada una segunda copia del complejo ATPasa F₀F₁ compuesta por las ocho proteínas de forma exclusiva dentro de los 5 organismos acidófilos, esta segunda copia no fue identificada en ningún otro organismo del filo Verrucomicrobia. Para la subunidad A la secuencia más cercana fuera de la familia de proteínas corresponde al organismo *Legionella pneumophila* con valores de 50% de identidad y 94% de cobertura. Para la subunidad B la secuencia más cercana es *Desulfopila aestuarii* con valores de 33% de identidad y un 98% de cobertura. Para la subunidad C la secuencia más cercana es *Methylocaldum* sp. 14B con valores de 69% de identidad y 100% de cobertura. Para la subunidad alfa la secuencia más cercana es *Legionella hackeliae* con valores de 65% de identidad y 95% de cobertura. Para la subunidad beta la secuencia más cercana es *Microbulbifer marinus* con valores de 69% de identidad y 95% de cobertura. Para la subunidad épsilon la secuencia más cercana es *Methylosarcina fibrata* con valores de 39% de identidad y 97% de cobertura. Para la subunidad gama la secuencia más cercana es *Acidithiobacillus* sp. CJ-2 con valores de 35% de identidad y 96% de cobertura.

6.4 .4. Predicción trayectoria evolutiva de las propiedades acidófilas.

Un análisis de ganancia y pérdida de genes fue realizado utilizando el programa *GLOOME*. Adicionalmente fue integrada la información relacionada a la duplicación de proteínas y el análisis de transferencia horizontal de genes para la generación del modelo evolutivo (**Figura 15**).

En el nodo N2, definido como la separación entre el grupo externo *Pedosphaera párvula* (neutrófilo) y los organismos acidófilos, fue identificada la mayor parte de la ganancia de genes con la presencia del sistema de descarboxilación de glutamato *gadA* y *gadC* en este nodo; la segunda enzima en el sistema de agmatina deiminasa *aguB*; las cuatro subunidades del transportador de fosfato *pstA*, *pstB*, *pstC* y *pstS*; la chaperona *huB* y una serie de proteínas asociadas a la biosíntesis de hopanoides, como las encargadas del paso de farnesil difosfato a escualeno *hpnC* y la fusión *hpnDE*, además de las encargadas de la ramificación de los hopanoides como *hpnO*, *hpnG* y *hpnI*. La duplicación del *clúster* de genes asociado a la segunda copia de la ATPasa F₀F₁ en los organismos acidófilos también ocurrió en el nodo N2. En la separación entre ambos géneros identificadas por el nodo N3 para *Methylacidimicrobium* fue identificada la ganancia de genes como el antiporter de arginina/agmatina *adiC* y el antiporter de sodio/protones

naH, adicionalmente en este nodo también fue identificada la duplicación de *huB*. En el nodo N4 que marca al género *Methylacidiphilum* se ve repetida la ganancia de *adiC* y *naH*.

Finalmente, los nodos terminales para los 5 acidófilos presentan duplicación de la proteína de membrana *slp* con un número diferente de copias para cada organismo, adicionalmente en *Verrucomicrobium* sp. 3C fue identificada la duplicación en múltiples copias de la enzima anhidrasa carbónica *can*.

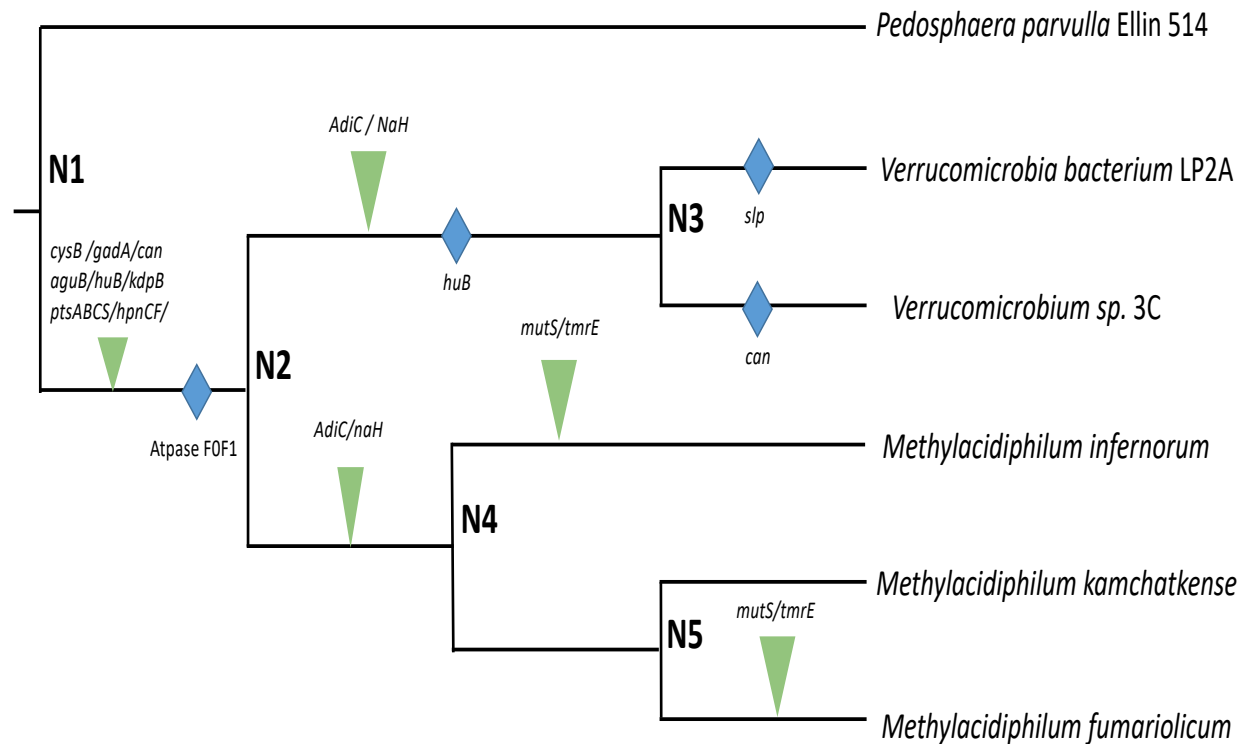


Figura 16. Reconstrucción evolutiva predicha de genes asociados a resistencia a pH extremadamente bajo en *Methylacidiphilaceae*. Representados desde N1 hasta N5 los nodos de separación entre diferentes clados y organismos. En color verde ganancia de genes escritos en la parte superior. Diamantes azules representa duplicación de secuencias.

7. DISCUSIÓN

Con la llegada de nuevas técnicas de secuenciación masiva y la utilización de marcadores conservados como el ARN ribosomal 16S, se ha descubierto y comenzado el estudio de una serie de organismos denominados como “materia negra microbiana”, donde técnicas como la genómica comparativa además nos permiten inferir aspectos funcionales de los organismos y proponer nuevos mecanismos moleculares (Rinke *et al.*, 2013).

Verrucomicrobia ha sido poco estudiado debido a las dificultades en el cultivo de algunos de sus organismos (Bergmann *et al.*, 2011), pero al ser un clado ancestral, nos puede entregar información importante sobre el origen de diferentes propiedades como la acidofilia dentro del dominio Bacteria.

7.1. Nueva clasificación taxonómica para Verrucomicrobia.

En la actualidad la gran mayoría de las especies dentro del filo se mantienen sin clasificación taxonómica clara, contando sólo con la clasificación a nivel de especie. En el filo Verrucomicrobia, la clasificación utilizada tradicionalmente corresponde al nivel de “Subdivisión”, nivel taxonómico no utilizado comúnmente para definir organismos procariontes (siendo este más ampliamente utilizado en la clasificación de plantas), lo que genera un primer nivel de dificultad en su estudio. Dos de estas subdivisiones, en años recientes, han sido clasificadas como nuevos filos (correspondiendo actualmente a Kiritimatiellaeota y Lentisphaerae), lo cual apoya la necesidad de realizar un análisis exhaustivo de la clasificación del filo (Spring *et al.*, 2016, Cho *et al.*, 2004).

En este trabajo fueron utilizados una serie de marcadores para realizar la reconstrucción filogenética de Verrucomicrobia. En primer lugar, utilizando la secuencia del gen ARN ribosomal 16S problemas fueron identificados, donde la subdivisión uno no pudo ser agrupada como un clado monofilético, problema observado en trabajos anteriores utilizando este mismo marcador, derivado del bajo nivel de resolución que nos entrega el gen ARN ribosomal 16S (Op den Camp *et al.*, 2009). Además, una serie de nodos presentaron valores de *Bootstrap* bajo 70, lo que no nos permite proponer con seguridad, según los valores de confianza una topología definitiva con la utilización única de este marcador. Limitaciones en la utilización del gen ARN ribosomal 16S, como las

observadas en este trabajo han sido analizadas previamente en otros estudios (Koeppel *et al.*, 2008).

Como solución, una aproximación común es utilizar otros marcadores, como proteínas conservadas dentro del dominio Bacteria y de copia única (Ciccarelli *et al.*, 2008; Parks *et al.*, 2018) para proponer topologías definitivas, lo cual es principalmente importante en clados menos estudiados con técnicas tradicionales de cultivo, donde existe un gran número de genomas proveniente desde muestras ambientales (Rajendhran y Gunasekaran, 2011). Los árboles construidos con las proteínas concatenadas de copia única (31cic y bac120) presentaron la misma topología, donde la subdivisión número cuatro aparece como la más ancestral del filo. Esta topología es diferente a la observada en el árbol construido con el gen ARN ribosomal 16S, en el cual la subdivisión número uno era la más ancestral. Los resultados obtenidos con ambos marcadores de proteínas concatenadas presentan soportes de *bootstrap* en la separación de la rama ancestral (subdivisión número cuatro), además son concordantes con los resultados obtenidos por reconstrucciones recientes del árbol de la vida (Hug *et al.*, 2015; Parks *et al.*, 2018). Otro problema resuelto encontrado en el árbol del marcador ARN ribosomal 16S fue que todos los organismos de la subdivisión número uno, se agruparon en un clado monofilético. Los valores de *bootstrap* observados para ambos marcadores de proteínas concatenadas fueron mayores en comparación al marcador ARN ribosomal 16S. Entre ambos sets el de bac120 presentó mejores valores de *Bootstrap*, resultado en concordancia a lo esperado por contar con un mayor número de proteínas concatenadas (44 en comparación a las 26 del set 31cic). En todos los árboles construidos (16S rRNA, 31cic y bac120) el grupo externo más cercano a los organismos acidófilos identificado fue el neutrófilo *Pedosphaera parvula*, correspondiente al único genoma de la subdivisión número seis. De igual manera, la topología de los organismos acidófilos se conservó entre los diferentes métodos, con dos clados diferentes, uno con los organismos acidófilos y termófilos (*Methyacidiphilum*) y otro con acidófilos y mesófilos.

Una serie de métodos fueron utilizados para obtener una clasificación taxonómica concreta de los organismos acidófilos. La identidad del marcador ARN ribosomal 16S ha sido propuesta como un valor válido para la clasificación a niveles taxonómicos desde especie hasta filo (Yarza *et al.*, 2014). El valor de AAI (Average Aminoacid Identity) fue utilizado para la clasificación a nivel de género y familia. Los valores de ANI (Average Nucleotide Identity) y TETRA (frecuencia de tetranucleótidos) en conjunto son utilizados para la clasificación a nivel de especie. A la fecha

la clasificación taxonómica para los organismos acidófilos a nivel de familia corresponde a *Methylacidiphilaceae*, agrupando a los tres organismos clasificados dentro del género *Methylacidiphilum* (organismos acidófilos y termófilos). Adicionalmente, dos organismos acidófilos cuentan con clasificación propuesta a nivel de género como *Methylacidimicrobium* (organismos acidófilos y mesófilos), sin clasificación a nivel de familia. Esta clasificación es basada en sus características fenotípicas y no han sido realizados estudios de genómica exhaustivos (van Teeseling *et al.*, 2014). Los valores de identidad del marcador ARN ribosomal 16S y AAI concuerdan, agrupando a ambos géneros de organismos acidófilos dentro de una misma familia y confirman la división en los dos géneros existentes. En base a estos resultados se propone agregar al género *Methylacidimicrobium* a la familia *Methylacidiphilaceae*, incluyendo de esta manera a todos los organismos acidófilos del filo Verrucomicrobia dentro de una familia. Los resultados de clasificación a nivel de especie con ANI y TETRA confirman lo previamente reportado en la literatura clasificando a los cinco organismos acidófilos como especies únicas.

7.2. Definición de moviloma y estabiloma.

El moviloma se encuentra definido como el conjunto de todos los elementos móviles dentro de un genoma. Estos de manera general pueden ser divididos en cuatro amplias categorías: (i) Transposones; (ii) Plásmidos; (iii) Bacteriófagos; (iv) Secuencias de inserción (Siefert, 2009). Los elementos del moviloma son capaces de entregarnos información útil tanto a niveles de ecología como evolución, pero significan una complicación adicional para los análisis de genómica comparativa a nivel funcional. Al estar involucrados con procesos de transferencia horizontal de genes, sucesivos eventos de transposición pueden provocar el truncamiento de estas secuencias, produciendo un bajo nivel de conservación entre ellas al quedar como fragmentos de las secuencias originales dentro del genoma, característica que los hace difícil de identificar por algoritmos bioinformáticos (Frost *et al.*, 2005). Este problema aumenta la complejidad al analizar el número de familias de proteínas generadas en la construcción del pangenoma, ya que al realizar este análisis, los diferentes fragmentos serían identificados como múltiples familias únicas en los genomas, lo cual puede llevar a interpretaciones erróneas en la comparación de diferentes genomas.

En este trabajo fueron utilizadas tres bases de datos especializadas en la detección de los diferentes tipos de elementos móviles. Resultados con amplia diferencia en el número de

elementos móviles por genoma fueron observados. La base de datos enfocada en la detección de Transposones (*TnpPred*) y la enfocada en detección de secuencias de inserción (*ISsaga*) mostraron resultados similares en el número de secuencias identificadas donde *Verrucomicrobium* sp. 3C muestra el mayor porcentaje del genoma identificado como elemento móvil. La tercera base de datos (ACLAME) integra información de todas las categorías de elementos móviles y presentó valores sobre el 50% de secuencias identificadas como elemento móvil para cada uno de los genomas. Esta tendencia es totalmente diferente al número de elementos móviles detectados por las dos bases de datos mencionadas anteriormente.

La diferencia en el número de secuencias identificadas como elementos móviles entre las dos primeras bases de datos y la última resultado puede explicarse por diferentes factores: En primer lugar, la inclusión de secuencias aportadas por usuarios, secuencias que no necesariamente poseen una curación manual exhaustiva provocando la aparición de numerosos falsos positivos. Adicionalmente en la inclusión de secuencias de plásmidos y transposasas pudieron ser incluidas secuencias que son transportadas por estos elementos móviles pero que no son realmente elementos móviles.

Este resultado destaca la importancia de la curación de la información en estudios bioinformáticos, donde la automatización de procesos completos puede llevar a errores de sobre estimación en la identificación de secuencias, lo cual a su vez puede producir una serie de errores sistemáticos en análisis de genómica comparativa.

7.3. Análisis funcional de *Methylacidiphilaceae*.

Análisis de las predicciones funcionales utilizando la base de datos COG para todos los organismos de la familia fueron realizados. No se observaron diferencias entre todos los organismos acidófilos, los cuales presentaban un perfil similar en los porcentajes de proteínas identificadas dentro de cada una de las categorías funcionales. Este resultado se encuentra acorde a lo esperado ya que todos los organismos en estudio que han sido descritos presentan una morfología similar además de características bioquímicas similares (Op den Camp *et al.*, 2009). Todos los organismos son metanótrofos aeróbicos, utilizan el ciclo de Calvin-Benson-Bashamm para la fijación de carbono y son capaces de fijar nitrógeno desde amonio o nitrógeno ambiental (van Teeseling *et al.*, 2014). La mayor diferencia a nivel fisiológico es la capacidad del género *Methylacidiphilum* a crecer en temperaturas entre 45 °C a 55 °C (mayores que las del género

Methylacidimicrobium cuyo crecimiento es a 30 °C). A pesar de esta diferencia ninguna categoría se vio enriquecida para dar cuenta de este cambio en el estilo de vida.

Al comparar el número de proteínas asignadas en cada categoría funcional en los organismos acidófilos con el grupo externo neutrófilo *Pedosphaera parvula*, fueron observadas diferencias. La categoría de movilidad (M), cuenta con un mayor número de secuencias asignadas, lo cual puede ser explicado por la presencia de mecanismos de motilidad en esta bacteria (presencia de flagelo), mientras que ningún mecanismo de movilidad ha sido reportado en los organismos acidófilos. Las categorías asociadas a transducción de señales (T) y traducción y biogénesis (J), tienen el doble de porcentaje de secuencias a asignadas en *Pedosphaera parvula*, lo cual ha sido reportado anteriormente en trabajo de con organismos acidófilos los cuales no poseen mecanismos de motilidad además de contar con menos sistemas de dos componentes al ser comparados con los organismos neutrófilos más cercanos (Gonzalez *et al.*, 2017). Al existir un compartimiento intracelular adicional, es posible que una mayor cantidad de proteínas estén involucradas en la transducción de señales entre citoplasma y pirelulosoma, produciendo la mayor asignación en el organismo neutrófilo en comparación a los organismos acidófilos que no tienen este compartimiento intracelular.

Existen categorías que presentaron una mayor cantidad de proteínas asignadas para los organismos acidófilos:

La categoría de producción de energía (C), cuenta con valores entre 2 a 3% de mayor asignación acidófilos en comparación al neutrófilo, lo cual se puede relacionar a la adaptación que realizan organismos acidófilos utilizando la gradiente de protones generadas por el ambiente extremadamente ácido (Baker-Austen y Dopson, 2007). En este contexto la identificación de dos copias del complejo ATPasa F₀F₁ apoyan la hipótesis de que parte de la adaptación de estos organismos puede estar asociada a la utilización de la gradiente de protones como una ventaja adaptativa.

La categoría de transporte de aminoácidos (S) presenta una asignación 2% mayor en organismos acidófilos, esto puede relacionarse con la utilización de mecanismos de antiporter de proteínas y exportación de proteínas hacia la membrana, como resistencia a pH extremadamente ácido (Lund *et al.*, 2014).

Para todos los genomas incluido el grupo externo, cerca de un 15% de las proteínas de cada uno de los genomas no fue asignado a ninguna categoría funcional. Al incluir dentro de este análisis

las categorías de proteínas pobremente caracterizadas (S) y proteínas de función desconocida (R), las cuales tienen valores entre 3% a 5% de proteínas asignadas por genoma y la categoría de función general con valores entre 6% a 9% de proteínas asignadas por genoma, se alcanza un total cercano a 25% de proteínas asignadas por genoma en categorías que no nos entregan información funcional al no contar con función definida.

Este resultado puede estar asociado al carácter ambiental de todas las especies en estudio, las cuales además no están relacionadas a procesos biotecnológicos o patologías humanas, lo que genera un sesgo en la cantidad de información disponible en bases de datos sobre este tipo de organismos. Debido a esto es una proyección fundamental el estudio de proteínas no caracterizadas en estos organismos, ya que estas son capaces de entregarnos información nueva asociada al estilo de vida acidófilo.

7.4. Definición del pangenoma de *Methylacidiphilaceae*.

El pangenoma de los 5 organismos acidófilos de la familia *Methylacidiphilaceae* está conformado por 4,302 familias de proteínas. Este resultado es comparable con los obtenidos en otros estudios de genómica comparativa en organismos acidófilos que utilizan una cantidad similar de genomas, como en la clase *Acidithiobacillia* (González *et al.*, 2016).

El core genoma está compuesto por 801 familias de proteínas anotadas, las cuales en su mayoría corresponden a proteínas ribosomales o asociadas a procesos fundamentales para la supervivencia de la bacteria. En el core genoma fueron identificadas la mayoría de las proteínas asociadas a la resistencia a pH extremadamente ácido, lo cual nos indica que la mayoría de los mecanismos utilizados para vivir en esta condición son altamente conservados entre los distintos organismos acidófilos del filo.

El genoma dispensable para cada uno de los organismos se encuentra compartido mayormente con los representantes del mismo género. Para los organismos acidófilos y mesófilos del género *Methylacidimicrobium*, un total de 742 familias de proteínas anotadas fueron identificadas. Para los organismos acidófilos y termófilos del género *Methylacidiphilum*, un total de 642 familias de proteínas anotadas fueron identificadas. Este resultado concuerda con los resultados de estudios filogenéticos y clasificación taxonómica, mostrando una clara separación entre ambos clados de organismos acidófilos. Además, estudiar el genoma dispensable de estos

organismos es de particular interés para estudios posteriores, debido a que las proteínas anotadas que están asociadas en la adaptación de *Methyacidiphilum* para crecer en temperaturas entre 50 °C a 60 °C se encuentran dentro de esta categoría del pangenoma.

En las familias de proteínas anotadas como únicas por genoma destaca la mayor cantidad de familias para el organismo *Verrucomicrobium* sp. 3C. Esta tendencia es interesante, ya que este organismo es el con genoma de mayor tamaño (2,4 Mb en comparación a el resto que esta entre 2 - 2,2 Mb de tamaño) y también fue el que presentó un mayor número secuencias identificadas como elementos móviles. Es posible que el mayor número de familias de proteínas anotadas como únicas sea producto de diferentes eventos de transferencia horizontal o infección por fagos, lo que puede otorgar características particulares para este genoma en comparación al resto.

7.5. Mecanismos de resistencia a pH extremadamente ácido.

Una serie proteínas anotadas asociadas a resistencia a pH extremadamente ácido fueron identificadas en *Methyacidiphilaceae*, la mayoría de estas proteínas anotadas se encontraron conservadas dentro de la familia acidófila.

Fueron identificados dos sistemas de descarboxilación y antiporter (glutamato y arginina), de los 4 sistemas de descarboxilación y antiporter tradicionalmente descritos en bacterias enteropatógenicas. Los de glutamato y arginina son los expresados de forma particular en la adaptación en pH extremadamente ácidos, a diferencia de los de ornitina y lisina que se expresan a pH moderadamente ácido (Foster, 2004).

Las dos copias identificadas del complejo ATPasa F_0F_1 , son de particular interés debido a que la duplicación de este complejo ha sido relacionada con adaptación a condiciones ambientales, como es el caso de organismos halófilos, donde la segunda copia del complejo es utilizada en la expulsión de iones de sodio siendo un mecanismo de adaptación a las altas concentraciones de cationes (Dibrova *et al.*, 2010). En otros casos, la duplicación de este complejo también ha sido relacionada con producción de energía mediante vías alternativas (Koumandou y Kossida, 2014). La duplicación identificada en los organismos de *Methyacidiphilaceae* sólo presenta similitud con secuencias asociadas a otros organismos acidófilos, este resultado nos indica que esta segunda copia del complejo puede funcionar como una adaptación específica al ambiente de alta concentración de protones, tanto utilizando la gradiente de protones como un método para la

generación de energía o como un mecanismo adicional de expulsión de protones desde el medio citoplasmático.

Proteínas anotadas asociadas a la biosíntesis de hopanoides también fueron identificadas de manera exclusiva en organismos acidófilos del filo Verrucomicrobia. La fusión de las proteínas HpnD y HpnE fue identificada de manera conservada en todos los organismos de *Methylacidiphilaceae*. Fusiones en la vía de biosíntesis de hopanoides han sido previamente identificadas en otros organismos ácidos como en el género *Acidithiobacillus*, donde HpnC, HpnD y HpnE se encontraron fusionadas y en género *Leptosprillum*, donde HpnG cuenta con un dominio adicional. La fusión identificada en *Methylacidiphilaceae* no fue identificada en ningún organismo fuera de este grupo, por lo cual se propone que la fusión funciona como adaptación particular en los acidófilos de Verrucomicrobia al ambiente extremadamente ácido, al poder producir con mayor eficacia escualeno uniendo dos de las enzimas necesarias del proceso. Una serie de proteínas anotadas asociadas a la ramificación de los hopanoides (HpnG, HpnP) fueron identificadas. Los hopanoides ramificados son moléculas que aumentan la rigidez de la membrana y que han sido identificadas como moléculas relacionadas en la tolerancia a pH extremadamente ácido (Schmerk *et al.*, 2011).

Finalmente, es importante destacar las limitaciones en la interpretación de estos resultados, ya que, al estar trabajando con proteínas anotadas desde información genómica, sólo es posible proponer la potencialidad de un organismo de estar expresando los diferentes sistemas mencionados. La ventaja de estudios bioinformáticos y la identificación de estas proteínas anotadas es que nos permite proponer objetivos específicos para estudios mediante técnicas tradicionales de laboratorio.

7.6. Transferencia horizontal de genes.

Fueron identificadas diferentes proteínas anotadas asociadas a transferencia horizontal de genes en el contexto genómico de varias proteínas anotadas de interés, incluyendo transposasas, integrasas y proteínas anotadas asociadas a fagos en los genomas de la familia *Methylacidiphilaceae*. Un ejemplo de esta situación es el contexto genómico de genes como *gadC* o *adiC*, las cuales corresponden a antiporter de sistemas de descarboxilación, transposasas o integrasas, las cuales fueron identificadas tanto río abajo como río arriba. Adicionalmente, la gran mayoría de las proteínas anotadas asociadas a resistencia a pH extremadamente ácido identificadas

presentaron una mayor similitud con proteínas anotadas pertenecientes a filos diferentes a Verrucomicrobia. La presencia de transposasas, integrasas y proteínas asociadas a fagos nos permiten proponer que la adquisición de proteínas anotadas relacionadas a resistencia a pH extremadamente ácido ocurrió mediante transferencia horizontal de genes, posteriormente siendo heredadas de manera vertical desde un ancestro común acidófilo.

El estudio de las secuencias con mayor similitud a las proteínas anotadas no encontradas dentro del filo Verrucomicrobia nos permite proponer posibles donadores de estas proteínas anotadas asociadas a resistencia a acidez. Las proteínas anotadas con mayor similitud en la mayoría de los casos estaban asociados a organismos del filo Proteobacteria y no pertenecían a organismos acidófilos. Este resultado es difícil de interpretar, ya que puede estar asociado al sesgo existente en la información genómica en las bases de datos, donde Proteobacteria es el filo con mayor cantidad de organismos aislados, lo que lleva a tener una mayor cantidad de genomas disponibles (Yarza *et al.*, 2014). Problemas en la anotación de proteínas en los organismos depositados es otro factor que puede afectar la detección de genes en filos más cercanos al de este estudio. Con la información que existente a la fecha es posible hacer una propuesta certera que dé cuenta de la trayectoria evolutiva de estos genes que otorgan la resistencia a acidez en los organismos de *Methylacidiphilaceae*. A pesar de que Verrucomicrobia aparece en el árbol de la vida como un filo ancestral cercano a la bifurcación entre los dominios Bacteria y Archaea, no es posible identificar mediante el presente estudio las fuentes originales de estas proteínas, siendo una proyección del trabajo la búsqueda del origen de estas proteínas.

7.7. Predicción de la historia evolutiva.

La mayoría de los eventos de ganancia de genes ocurrieron en N2, nodo que representa la separación del grupo externo (organismo neutrófilo) y los organismos acidófilos. Este resultado da cuenta en primer lugar, de que el estado ancestral de los organismos del filo era el de neutrofilia y que, mediante la ganancia de genes por una serie de mecanismos, un ancestro común a ambos géneros obtuvo la capacidad de crecer en pH extremadamente ácido dentro de Verrucomicrobia.

En nodos de la separación entre ambos géneros N3 y N4, aparecen genes particulares asociados a antiporter de membrana como *adiC* y *naH*. Estos eventos son interesantes de estudiar ya que al estar presentes en todos los organismos acidófilos en estudio, fue fundamental la

realización de una confirmación manual de que en ambos casos estos no hubieran sido obtenidos evolutivamente en un nodo ancestral N2, como la mayoría de los genes relacionados acidofilia. El contexto genético para estos genes es conservado entre los organismos de un mismo género (*Methylacidiphilum* o *Methylacidimicrobium*), pero diferente comparados entre los de diferente género, información que es observada en las secuencias con mayor similitud que mantienen el mismo patrón de conservación en organismos intra-género pero diferencias inter-género, apoyando la hipótesis de que ambos genes fueron obtenidos en eventos particulares en ancestros de cada uno de los géneros y no en un ancestro común a todos como es el caso de otros genes que se encuentran presentes en todos los acidófilos.

En nodos terminales, la mayoría de los eventos identificados fueron eventos de duplicación de genes como *slp*, evento observado tanto en *Methylacidiphilum* como *Methylacidimicrobium*. Las secuencias de *slp* cuentan con un alto grado de similitud entre sí, pero no con secuencias de otros filos, este resultado nos permite proponer que las diferentes copias del gen son producto de una duplicación y no diferentes eventos de transferencia horizontal. Otro caso de duplicación identificado fue con el gen *can*, el cual presenta diferentes copias con alta similitud en *Verrucomicrobium* sp. 3C.

Estos resultados nos permiten proponer que eventos de transferencia horizontal en un ancestro común a todos los acidófilos fue la principal fuente de genes asociados a la resistencia a pH extremadamente ácido. Eventos como la duplicación de genes fueron identificados mayormente en nodos terminales, dando cuenta de eventos que ocurrieron de manera particular en especies específicas.

8. CONCLUSIONES

- Los organismos acidófilos del filo Verrucomicrobia se encuentran agrupados en la familia *Methylacidiphilaceae* y su organismo más cercano es *Pedosphaera parvula*, un organismo neutrófilo y mesófilo.
- La familia *Methylacidiphilaceae* se encuentra dividida en dos géneros, *Methylacidiphilum* con organismos termófilos y *Methylacidimicrobium*, con organismos mesófilos.
- El pangenoma de *Methylacidiphilaceae* está definido por 4302 familias de proteínas, de las cuales 801 familias de proteínas fueron identificadas como parte del core genoma.
- Una serie de mecanismos de resistencia a pH bajo, como los de descarboxilación y antiporter de glutamato y arginina, biosíntesis de hopanoides, transportadores de cationes (sodio y potasio) fueron identificados en el core genoma de organismos acidófilos y ausentes en el ancestro neutrófilo más cercano taxonómicamente.
- Mecanismos de transferencia horizontal de genes fueron identificados en el vecindario de genes como *gadC*, *adiC*, *pstABCD*, siendo un mecanismo de ganancia de estos genes.
- La duplicación de genes como el complejo ATPasaF₀F₁, *can* y *slp* fueron identificados como mecanismos de adaptación al pH extremadamente ácido en *Methylacidiphilaceae*.
- La propiedad acidófila fue obtenida mediante la ganancia de genes dentro del filo Verrucomicrobia en un ancestro común a todos los acidófilos el cual presentaba la mayoría de los genes identificados.

9. MATERIAL SUPLEMENTARIO

<i>Nombre gen</i>	<i>Nombre proteína</i>	<i>ID (Uniprot)</i>	<i>Referencia</i>
<i>adiA</i>	Arginina descarboxilasa	P28629	Foster, 2004
<i>AdiC</i>	Arginina:agmatina antiporter	P60061	Foster, 2004
<i>cadA</i>	Lisina descarboxilasa	P0A9F3	Soksawatmaekhin <i>et al.</i> , 2004
<i>cadB</i>	Cadaverina:lisina antiporter	P33234	Soksawatmaekhin <i>et al.</i> , 2004
<i>SpeC</i>	Ornitina descarboxilasa	P21169	Romano <i>et al.</i> , 2014
<i>SpeF</i>	Ornitina descarboxilasa	P24169	Romano <i>et al.</i> , 2014
<i>potE</i>	Putrescina:H ⁺ / putrescina/orinitina antiporter	P0AAF1	Romano <i>et al.</i> , 2014
<i>gada</i>	Glutamato descarboxilasa	P69908	Bearson <i>et al.</i> , 2006
<i>gadB</i>	Glutamato descarboxilasa	P69910	Bearson <i>et al.</i> , 2006
<i>gadC</i>	Glutamato/4-aminobutirato antiporter	P63235	Bearson <i>et al.</i> , 2006
<i>dps</i>	Proteína de proteccion de DNA	P0ABT2	Jeong <i>et al.</i> , 2008
<i>hdeA</i>	Chaperona de estrés periplasmático	P0AES9	Gajiwala y Burley, 2000
<i>hdeB</i>	Chaperona de estrés periplasmático	P0AET2	Gajiwala y Burley, 2000
<i>hchA</i>	Deglicasa ácida 1	P31658	Mujacic y Baneyx, 2006
<i>dnaK</i>	Chaperona DnaK	P0A6Y8	Lund <i>et al.</i> , 2014
<i>groL</i>	Chaperonina GroEL	P0A6F5	Lund <i>et al.</i> , 2014
<i>clpB</i>	Chaperona ClpB	P63285	Leverrier <i>et al.</i> , 2004
<i>clpC</i>	Proteasa dependiente de ATP ClpC	Q2G0P5	Leverrier <i>et al.</i> , 2004
<i>yhiD</i>	Transportador de magnesio YhiD	P0AFV2	Mates <i>et al.</i> , 2007
<i>slp</i>	Lipoproteína asociada a deprivación de carbono	P37194	Mates <i>et al.</i> , 2007
<i>hdeD</i>	Proteína de resistencia acidez de membrana	P0AET5	Mates <i>et al.</i> , 2007
<i>arcA</i>	Arginina deiminasa	P75218	Casiano-Colon y Marquis, 1988
<i>argI</i>	Ornitina carbamoiltransferasa	P04391	Casiano-Colon y Marquis, 1988
<i>arcC</i>	Carbamato kinasa	Q48295	Casiano-Colon y Marquis, 1988
<i>aguA</i>	Agmatina deiminasa	Q9I6J9	Lund <i>et al.</i> , 2014
<i>aguB</i>	N-carbamoilputrescina amidohidrolasa	Q9I6J8	Lund <i>et al.</i> , 2014
<i>ureA</i>	Ureasa subunidad gama	Q55053	Krulwich <i>et al.</i> , 2011
<i>ureB</i>	Ureasa subunidad beta	Q55054	Krulwich <i>et al.</i> , 2011
<i>ureC</i>	Ureasa subunidad alfa	P50047	Krulwich <i>et al.</i> , 2011
<i>ureD</i>	Ureasa proteína accesoria UreD	Q09063	Krulwich <i>et al.</i> , 2011
<i>ureE</i>	Ureasa proteína accesoria UreE	P18317	Krulwich <i>et al.</i> , 2011
<i>ureF</i>	Ureasa proteína accesoria UreF	P18318	Krulwich <i>et al.</i> , 2011
<i>ureG</i>	Ureasa proteína accesoria UreG	P18319	Krulwich <i>et al.</i> , 2011
<i>atpE</i>	ATP sintasa subunidad C	P68699	Foster, 2004
<i>atpF</i>	ATP sintasa subunidad B	P0ABA0	Foster, 2004
<i>atpB</i>	ATP sintasa subunidad A	P0AB98	Foster, 2004

<i>atpA</i>	ATP sintasa subunidad alfa	P0ABB0	Foster, 2004
<i>atpD</i>	ATP sintasa subunidad beta	P0ABB4	Foster, 2004
<i>atpG</i>	ATP sintasa subunidad gama	P0ABA6	Foster, 2004
<i>atpH</i>	ATP sintasa subunidad delta	P0ABA4	Foster, 2004
<i>atpC</i>	ATP sintasa subunidad epsilon	P0A6E6	Foster, 2004
<i>hpnD</i>	Prescualeno difosfatasa sintasa	Q6N3F2	Kulkarni et al., 2015
<i>hpnC</i>	Hidroiescualeno sintasa	Q6N3F1	Kulkarni et al., 2015
<i>hpnE</i>	Hidroiescualeno dehydroxylase	Q6N3F3	Kulkarni et al., 2015
<i>hpnP</i>	Hopanoide C-2 metilasa	B3QHD1	Kulkarni et al., 2015
<i>hpnF</i>	Escualeno hopano ciclasa	G0JPV0	Kulkarni et al., 2015
<i>hpnB</i>	Glicosiltransferasa asociada a hopano	Q5NP68	Geiger, 2018
<i>hpnH</i>	Proteína SAM asociada a hopanoides	Q5NP62	Kulkarni et al., 2015
<i>hpnG</i>	Fosforilasa asociada a hopanoides	Q5NP63	Kulkarni et al., 2015
<i>hpnR</i>	Hopanoide C-3 metilasa	Q60AV6	Geiger, 2018
<i>hpnI</i>	Glycosiltransferasa asociada a hopanoides	Q5NNW4	Geiger, 2018
<i>hpnK</i>	Hydrolasa asociada a hopanoides	Q5NNW2	Geiger, 2018
<i>hpnJ</i>	Radical SAM HpnJ	Q5NNW3	Geiger, 2018
<i>hpnM</i>	Proteína de biosíntesis hopanoides HpnM	Q5NP60	Geiger, 2018
<i>hpnA</i>	Epimerasa asociada a hopanoides	Q5NP69	Geiger, 2018
<i>hpnL</i>	Proteína putativa de membrana	A0A0B6S2B0	Geiger, 2018
<i>hpnN</i>	Transportador RND	G8MDC1	Geiger, 2018
<i>hpnO</i>	Aminotransferasa HpnO	M9U6E8	Geiger, 2018
<i>kdpC</i>	ATPasa transportadora de potasio KdpC	P03961	Buetti-Dinh et al., 2016
<i>kdpA</i>	ATPasa transportadora de potasio KdpA	P03959	Buetti-Dinh et al., 2016
<i>kdpB</i>	ATPasa transportadora de potasio KdpB	P03960	Buetti-Dinh et al., 2016
<i>kdpD</i>	Proteína sensora KdpD	P21865	Buetti-Dinh et al., 2016
<i>kch</i>	Canal de potasio asociado a voltaje	P31069	Buetti-Dinh et al., 2016
<i>trkA</i>	Sistema de captación de potasio Trk	P0AGI8	Buetti-Dinh et al., 2016
<i>nhaA</i>	Na ⁺ /H ⁺ antiporter	P13738	Buetti-Dinh et al., 2016
<i>pstS</i>	Proteína de unión a fosfato	P0AG82	Chen et al., 2015
<i>pstC</i>	Proteína de unión a fosfato	P0AGH8	Chen et al., 2015
<i>pstA</i>	Permeasa asociada a fosfato	P07654	Chen et al., 2015
<i>pstB</i>	Proteína de importación de fosfato	P0AAH0	Chen et al., 2015
<i>clpX</i>	Proteasa dependiente de ATP Clp	P0A6H1	Chen et al., 2015
<i>clpP</i>	Proteasa dependiente de ATP Clp	P0A6G7	Chen et al., 2015
<i>hupB</i>	Proteína de unión a ácidos nucleicos	P0ACF4	Chen et al., 2015

Tabla suplementaria 1. Genes asociados a resistencia a acidez.

<i>Grupo de ortologos</i>	<i>Tamaño Promedio</i>	<i>Anotación</i>
COG0012	380	ATPasa YchF
COG0048	137	Proteína ribosomal S12
COG0052	240	Proteína ribosomal S2
COG0080	154	Proteína ribosomal L11
COG0081	230	Proteína ribosomal L1
COG0087	288	Proteína ribosomal L3
COG0091	157	Proteína ribosomal L22
COG0092	240	Proteína ribosomal S3
COG0093	130	Proteína ribosomal L14
COG0094	182	Proteína ribosomal L5
COG0096	131	Proteína ribosomal S8
COG0097	177	Proteína ribosomal L6P/L9E
COG0098	220	Proteína ribosomal S5
COG0099	133	Proteína ribosomal S13
COG0100	145	Proteína ribosomal S11
COG0102	167	Proteína ribosomal L13
COG0103	172	Proteína ribosomal S9
COG0172	422	Seril tRNA sintetasa
COG0186	122	Proteína ribosomal S17
COG0197	175	Proteína ribosomal L16/L10AE
COG0200	166	Proteína ribosomal L15
COG0201	445	Translocasa SecY
COG0202	323	ARN polimerasa, subunidad alfa
COG0256	178	Proteína ribosomal L18
COG0522	199	Proteína ribosomal S4
COG0533	380	tRNA A37

Tabla suplementaria 2. Marcadores utilizados para árbol 31cic.

<i>Grupo de ortólogos</i>	<i>Tamaño Promedio</i>	<i>Nombre Proteína</i>
PF00380	122	Proteína ribosomal S9
PF00410	126	Proteína ribosomal S8
PF01025	176	GrpE
TIGR00019	361	PrfA
TIGR00061	101	Proteína ribosomal L21
TIGR00086	144	SmpB
TIGR00088	233	TrmD
TIGR00092	368	YchF
TIGR00116	293	Tsf
TIGR00158	148	Proteína ribosomal L9
TIGR00337	526	PyrG
TIGR00414	418	SerS
TIGR00420	351	TrmU
TIGR00431	210	TruB
TIGR00445	321	MraY
TIGR00496	176	Frr
TIGR00615	196	RecR
TIGR00635	305	RuvB
TIGR00963	787	SecA
TIGR00967	414	SecY
TIGR01009	212	Proteína ribosomal S3
TIGR01011	225	Proteína ribosomal S2
TIGR01017	200	Proteína ribosomal S4
TIGR01021	156	Proteína ribosomal S5
TIGR01032	114	Proteína ribosomal L20
TIGR01044	103	Proteína ribosomal L22
TIGR01066	141	Proteína ribosomal L13
TIGR01071	144	Proteína ribosomal L15
TIGR01146	286	Atp sintasa, subunidad gama
TIGR01164	126	Proteína ribosomal L16
TIGR01169	227	Proteína ribosomal L1
TIGR01171	275	Proteína ribosomal L2
TIGR01393	595	LepA
TIGR01632	140	Proteína ribosomal L11
TIGR01953	340	NusA
TIGR02012	321	RecA
TIGR02013	1238	RpoB
TIGR02027	298	RpoA

<i>TIGR02075</i>	233	PyrH
<i>TIGR02729</i>	329	CgtA
<i>TIGR03625</i>	202	Proteína ribosomal L3
<i>TIGR03632</i>	117	Proteína ribosomal S11
<i>TIGR03654</i>	175	Proteína ribosomal L6
<i>TIGR03723</i>	314	TsaD
<i>TIGR03953</i>	188	Proteína ribosomal L4

Tabla suplementaria 3. Marcadores utilizados para árbol bac120.

10. BIBLIOGRAFÍA

- Akcil, A., & Koldas, S. (2006). Acid Mine Drainage (AMD): causes, treatment and case studies. *Journal of Cleaner Production*, 14(12), 1139-1145.
- Alain, K., & Querellou, J. (2009). Cultivating the uncultured: limits, advances and future challenges. *Extremophiles*, 13(4), 583-594.
- Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*, 59(1), 143-169.
- Anvar, S. Y., Frank, J., Pol, A., Schmitz, A., Kraaijeveld, K., den Dunnen, J. T., & den Camp, H. J. O. (2014). The genomic landscape of the verrucomicrobial methanotroph *Methylacidiphilum fumariolicum* SolV. *BMC genomics*, 15(1), 914.
- Baker-Austin, C., & Dopson, M. (2007). Life in acid: pH homeostasis in acidophiles. *Trends in microbiology*, 15(4), 165-171.
- Baker, B. J., Lutz, M. A., Dawson, S. C., Bond, P. L., & Banfield, J. F. (2004). Metabolically active eukaryotic communities in extremely acidic mine drainage. *Applied and Environmental Microbiology*, 70(10), 6264-6271.
- Bearson, S., Bearson, B., & Foster, J. W. (1997). Acid stress responses in enterobacteria. *FEMS microbiology letters*, 147(2), 173-180.
- Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., ... & Fierer, N. (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biology and Biochemistry*, 43(7), 1450-1455.
- Buetti-Dinh, A., Dethlefsen, O., Friedman, R., & Dopson, M. (2016). Transcriptomic analysis reveals how a lack of potassium ions increases *Sulfolobus acidocaldarius* sensitivity to pH changes. *Microbiology*, 162(8), 1422-1434.
- Cárdenas, J. P., Quatrini, R., & Holmes, D. S. (2016). Progress in acidophile genomics. *The Acidophiles*, 179-198.
- Casiano-Colón, A. I. D. A., & Marquis, R. E. (1988). Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. *Appl. Environ. Microbiol.*, 54(6), 1318-1324.
- Cavicchioli, R. (2002). Extremophiles and the search for extraterrestrial life. *Astrobiology*, 2(3), 281-292.
- Chen, L. X., Hu, M., Huang, L. N., Hua, Z. S., Kuang, J. L., Li, S. J., & Shu, W. S. (2015). Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. *The ISME journal*, 9(7), 1579.
- Cho, J. C., Vergin, K. L., Morris, R. M., & Giovannoni, S. J. (2004). *Lentisphaera araneosa* gen. nov., sp. nov., a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, *Lentisphaerae*. *Environmental Microbiology*, 6(6), 611-621.
- Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *science*, 311(5765), 1283-1287.
- Cohen, O., Ashkenazy, H., Belinky, F., Huchon, D., & Pupko, T. (2010). GLOOME: gain loss mapping engine. *Bioinformatics*, 26(22), 2914-2915.
- Coker, J. A. (2016). Extremophiles and biotechnology: current uses and prospects. *F1000Research*, 5.

- Cox, M. M., & Battista, J. R. (2005). *Deinococcus radiodurans*—the consummate survivor. *Nature Reviews Microbiology*, 3(11), 882.
- Davis, K. E., Joseph, S. J., & Janssen, P. H. (2005). Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Applied and environmental microbiology*, 71(2), 826-834.
- De Maayer, P., Anderson, D., Cary, C., & Cowan, D. A. (2014). Some like it cold: understanding the survival strategies of psychrophiles. *EMBO reports*, 15(5), 508-517.
- Delage, L., Becerra, A., & Lazcano, A. (2005). The last common ancestor: What's in a name?. *Origins of Life and Evolution of Biospheres*, 35(6), 537-554.
- Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5), 361-375.
- Di Giulio, M. (2000). The universal ancestor lived in a thermophilic or hyperthermophilic environment. *Journal of theoretical biology*, 203(3), 203-213.
- Dibrova, D. V., Galperin, M. Y., & Mulkidjanian, A. Y. (2010). Characterization of the N-ATPase, a distinct, laterally transferred Na⁺-translocating form of the bacterial F-type membrane ATPase. *Bioinformatics*, 26(12), 1473-1476.
- Donoghue, M. J., Doyle, J. A., Gauthier, J., Kluge, A. G., & Rowe, T. (1989). The importance of fossils in phylogeny reconstruction. *Annual review of Ecology and Systematics*, 20(1), 431-460.
- Dopson, M., & Johnson, D. B. (2012). Biodiversity, metabolism and applications of acidophilic sulfur-metabolizing microorganisms. *Environmental microbiology*, 14(10), 2620-2631.
- Dunfield, P. F., Yuryev, A., Senin, P., Smirnova, A. V., Stott, M. B., Hou, S., ... & Wang, J. (2007). Methane oxidation by an extremely acidophilic bacterium of the phylum *Verrucomicrobia*. *Nature*, 450(7168).
- Erikstad, H. A., Jensen, S., Keen, T. J., & Birkeland, N. K. (2012). Differential expression of particulate methane monooxygenase genes in the verrucomicrobial methanotroph '*Methylacidiphilum kamchatkense*' Kam1. *Extremophiles*, 16(3), 405-409.
- Erikstad, H. A., & Birkeland, N. K. (2015). Draft genome sequence of "*Candidatus Methylacidiphilum kamchatkense*" Strain Kam1, a thermoacidophilic methanotrophic Verrucomicrobium. *Genome announcements*, 3(2), e00065-15.
- Federhen, S. (2011). The NCBI taxonomy database. *Nucleic acids research*, 40(D1), D136-D143.
- Foster, J. W. (2004). *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nature Reviews Microbiology*, 2(11), 898.
- Freitas, S., Hatosy, S., Fuhrman, J. A., Huse, S. M., Welch, D. B. M., Sogin, M. L., & Martiny, A. C. (2012). Global distribution and diversity of marine Verrucomicrobia. *The ISME journal*, 6(8), 1499.
- Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*, 3(9), 722.
- Fuerst, J. A., & Sagulenko, E. (2011). Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. *Nature Reviews Microbiology*, 9(6), 403-413.
- Fuerst, J. A. (2013). The PVC superphylum: exceptions to the bacterial definition?. *Antonie Van Leeuwenhoek*, 104(4), 451-466.

- Gajiwala, K. S., & Burley, S. K. (2000). HDEA, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. *Journal of molecular biology*, 295(3), 605-612.
- Geiger, O. (Ed.). (2018). *Biogenesis of Fatty Acids, Lipids and Membranes*. Springer.
- Gerday, C., & Glansdorff, N. (2007). *Physiology and biochemistry of extremophiles*. ASM Press.
- Golyshina, O. V., Pivovarova, T. A., Karavaiko, G. I., Kondratéva, T. F., Moore, E. R., Abraham, W. R., ... & Golyshin, P. N. (2000). *Ferroplasma acidiphilum* gen. nov., sp. nov., an acidophilic, autotrophic, ferrous-iron-oxidizing, cell-wall-lacking, mesophilic member of the Ferroplasmaceae fam. nov., comprising a distinct lineage of the Archaea. *International journal of systematic and evolutionary microbiology*, 50(3), 997-1006.
- González, C., Lazcano, M., Valdés, J., & Holmes, D. S. (2016). Bioinformatic Analyses of Unique (Orphan) Core Genes of the Genus *Acidithiobacillus*: Functional Inferences and Use As Molecular Probes for Genomic and Metagenomic/Transcriptomic Interrogation. *Frontiers in microbiology*, 7.
- Gupta, G. N., Srivastava, S., Khare, S. K., & Prakash, V. (2014). Extremophiles: an overview of microorganism from extreme environment. *International Journal of Agriculture, Environment and Biotechnology*, 7(2), 371.
- Hedlund, B. P., Gosink, J. J., & Staley, J. T. (1996). Phylogeny of Prosthecobacter, the fusiform caulobacters: members of a recently discovered division of the Bacteria. *International Journal of Systematic and Evolutionary Microbiology*, 46(4), 960-966.
- Henrici, A. T., & Johnson, D. E. (1935). Studies of Freshwater Bacteria: II. Stalked Bacteria, a New Order of Schizomycetes 1. *Journal of Bacteriology*, 30(1), 61.
- Hou, S., Makarova, K. S., Saw, J. H., Senin, P., Ly, B. V., Zhou, Z., ... & Wolf, Y. I. (2008). Complete genome sequence of the extremely acidophilic methanotroph isolate V4, *Methylacidiphilum infernorum*, a representative of the bacterial phylum Verrucomicrobia. *Biology Direct*, 3(1), 26.
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., ... & Suzuki, Y. (2016). A new view of the tree of life. *Nature microbiology*, 1(5), 16048.
- Ingledeu, W. J. (1982). Thiobacillus ferrooxidans the bioenergetics of an acidophilic chemolithotroph. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics*, 683(2), 89-117.
- Islam, T., Jensen, S., Reigstad, L. J., Larsen, Ø., & Birkeland, N. K. (2008). Methane oxidation at 55 C and pH 2 by a thermoacidophilic bacterium belonging to the Verrucomicrobia phylum. *Proceedings of the National Academy of Sciences*, 105(1), 300-304.
- Janssen, P. H., Schuhmann, A., Mörschel, E., & Rainey, F. A. (1997). Novel anaerobic ultramicrobacteria belonging to the Verrucomicrobiales lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. *Applied and Environmental Microbiology*, 63(4), 1382-1388.
- Jeong, K. C., Hung, K. F., Baumler, D. J., Byrd, J. J., & Kaspar, C. W. (2008). Acid stress damage of DNA is prevented by Dps binding in *Escherichia coli* O157: H7. *BMC microbiology*, 8(1), 181.
- Jiang, H., Chen, Y., Jiang, P., Zhang, C., Smith, T. J., Murrell, J. C., & Xing, X. H. (2010). Methanotrophs: multifunctional bacteria with promising applications in environmental bioengineering. *Biochemical Engineering Journal*, 49(3), 277-288.

- Johnson, D. B. (1998). Biodiversity and ecology of acidophilic microorganisms. *FEMS microbiology ecology*, 27(4), 307-317.
- Johnson, D. B. (2007). Physiology and ecology of acidophilic microorganisms. In *Physiology and biochemistry of extremophiles* (pp. 257-270). American Society of Microbiology.
- Johnson, D. B., & Hallberg, K. B. (2008). Carbon, iron and sulfur metabolism in acidophilic micro-organisms. *Advances in microbial physiology*, 54, 201-255.
- Johnson, D. B., & Quatrini, R. (2016). Acidophile microbiology in space and time. *Acidophiles*, 3.
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*.
- Kant, R., Van Passel, M. W., Sangwan, P., Palva, A., Lucas, S., Copeland, A., ... & Bruce, D. (2011). Genome sequence of *Pedospira parvula* Ellin514, an aerobic verrucomicrobial isolate from pasture soil. *Journal of bacteriology*.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.
- Khadem, A. F., Pol, A., Jetten, M. S., & den Camp, H. J. O. (2010). Nitrogen fixation by the verrucomicrobial methanotroph '*Methylacidiphilum fumariolicum*' SolV. *Microbiology*, 156(4), 1052-1059.
- Khadem, A. F., Pol, A., Wiczorek, A., Mohammadi, S. S., Francoijs, K. J., Stunnenberg, H. G., ... & den Camp, H. J. O. (2011). Autotrophic methanotrophy in Verrucomicrobia: *Methylacidiphilum fumariolicum* SolV uses the Calvin-Benson-Bassham cycle for carbon dioxide fixation. *Journal of bacteriology*, 193(17), 4438-4446.
- Khadem, A. F., Wiczorek, A. S., Pol, A., Vuilleumier, S., Harhangi, H. R., Dunfield, P. F., ... & Stein, L. Y. (2012). Draft genome sequence of the volcano-inhabiting thermoacidophilic methanotroph *Methylacidiphilum fumariolicum* strain SolV. *Journal of bacteriology*, 194(14), 3729-3730.
- Kolaczkowski, B., & Thornton, J. W. (2004). Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, 431(7011), 980-984.
- Krulwich, T. A., Sachs, G., & Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. *Nature Reviews Microbiology*, 9(5), 330.
- Kulkarni, G., Busset, N., Molinaro, A., Gargani, D., Chaintreuil, C., Silipo, A., ... & Newman, D. K. (2015). Specific hopanoid classes differentially affect free-living and symbiotic states of *Bradyrhizobium diazoefficiens*. *MBio*, 6(5), e01251-15.
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22), 3276-3278.
- Koeppel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., ... & Nevo, E. (2008). Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences*, 105(7), 2504-2509.
- Koumandou, V. L., & Kossida, S. (2014). Evolution of the F0F1 ATP synthase complex in light of the patchy distribution of different bioenergetic pathways across prokaryotes. *PLoS computational biology*, 10(9), e1003821.

- Leplae, R., Lima-Mendez, G., & Toussaint, A. (2009). ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic acids research*, 38(suppl_1), D57-D61.
- Leverrier, P., Vissers, J. P., Rouault, A., Boyaval, P., & Jan, G. (2004). Mass spectrometry proteomic analysis of stress adaptation reveals both common and distinct response pathways in *Propionibacterium freudenreichii*. *Archives of microbiology*, 181(3), 215-230.
- Ludwig, W., & Schleifer, K. H. (1994). Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS microbiology reviews*, 15(2-3), 155-173.
- Lund, P., Tramonti, A., & De Biase, D. (2014). Coping with low pH: molecular strategies in neutralophilic bacteria. *FEMS microbiology reviews*, 38(6), 1091-1125.
- Mao, X., Ma, Q., Zhou, C., Chen, X., Zhang, H., Yang, J., ... & Xu, Y. (2013). DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic acids research*, 42(D1), D654-D659.
- Mates, A. K., Sayed, A. K., & Foster, J. W. (2007). Products of the *Escherichia coli* acid fitness island attenuate metabolite stress at extremely low pH and mediate a cell density-dependent acid resistance. *Journal of bacteriology*, 189(7), 2759-2768.
- Medini, D., Donati, C., Tettelin, H., Massignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current opinion in genetics & development*, 15(6), 589-594.
- Mehta, R., Singhal, P., Singh, H., Damle, D., & Sharma, A. K. (2016). Insight into thermophiles and their wide-spectrum applications. *3 Biotech*, 6(1), 81.
- Mols, M., Van Kranenburg, R., Van Melis, C. C., Moezelaar, R., & Abee, T. (2010). Analysis of acid-stressed *Bacillus cereus* reveals a major oxidative response and inactivation-associated radical formation. *Environmental microbiology*, 12(4), 873-885.
- Mujacic, M., & Baneyx, F. (2006). Regulation of *Escherichia coli* hchA, a stress-inducible gene encoding molecular chaperone Hsp31. *Molecular microbiology*, 60(6), 1576-1589.
- Nichols, D., Lewis, K., Orjala, J., Mo, S., Ortenberg, R., O'connor, P., ... & Epstein, S. S. (2008). Short peptide induces an "uncultivable" microorganism to grow in vitro. *Applied and environmental microbiology*, 74(15), 4889-4897.
- Op den Camp, H. J., Islam, T., Stott, M. B., Harhangi, H. R., Hynes, A., Schouten, S., ... & Dunfield, P. F. (2009). Environmental, genomic and taxonomic perspectives on methanotrophic *Verrucomicrobia*. *Environmental Microbiology Reports*, 1(5), 293-306.
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarszewski, A., Chaumeil, P. A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology*.
- Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS biology*, 9(3), e1000602.
- Pikuta, E. V., Hoover, R. B., & Tang, J. (2007). Microbial extremophiles at the limits of life. *Critical reviews in microbiology*, 33(3), 183-209.
- Pol, A., Heijmans, K., Harhangi, H. R., Tedesco, D., Jetten, M. S., & Den Camp, H. J. O. (2007). Methanotrophy below pH 1 by a new *Verrucomicrobia* species. *Nature*, 450(7171), 874-878.

- Pol, A., Barends, T. R., Dietl, A., Khadem, A. F., Eygensteyn, J., Jetten, M. S., & Op den Camp, H. J. (2014). Rare earth metals are essential for methanotrophic life in volcanic mudpots. *Environmental microbiology*, *16*(1), 255-264.
- Posada, D. (2016). Phylogenomics for systematic biology. *Systematic biology*, *65*(3), 353-356.
- Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2006). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, *35*(suppl_1), D61-D65.
- Rampelotto, P. H. (2013). Extremophiles and extreme environments.
- Riadi, G., Medina-Moenne, C., & Holmes, D. S. (2012). TnpPred: a web service for the robust prediction of prokaryotic transposases. *Comparative and functional genomics*, *2012*.
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., ... & Dodsworth, J. A. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, *499*(7459), 431.
- Rohwerder, T., & Sand, W. (2007). Oxidation of inorganic sulfur compounds in acidophilic prokaryotes. *Engineering in Life Sciences*, *7*(4), 301-309.
- Romano, A., Ladero, V., Alvarez, M. A., & Lucas, P. M. (2014). Putrescine production via the ornithine decarboxylation pathway improves the acid stress survival of *Lactobacillus brevis* and is part of a horizontally transferred acid resistance locus. *International journal of food microbiology*, *175*, 14-19.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, *61*(3), 539-542.
- Rowell, D. L., & Wild, A. (1985). Causes of soil acidification: a summary. *Soil Use and Management*, *1*(1), 32-33.
- Rothschild, L. J. (2010). A powerful toolkit for synthetic biology: over 3.8 billion years of evolution. *Bioessays*, *32*(4), 304-313.
- Schmerk, C. L., Bernards, M. A., & Valvano, M. A. (2011). Hopanoid production is required for low pH tolerance, antimicrobial resistance, and motility in *Burkholderia cenocepacia*. *Journal of bacteriology*, JB-05979.
- Sharp, C. E., Smirnova, A. V., Graham, J. M., Stott, M. B., Khadka, R., Moore, T. R., ... & Dunfield, P. F. (2014). Distribution and diversity of Verrucomicrobia methanotrophs in geothermal and acidic environments. *Environmental microbiology*, *16*(6), 1867-1878.
- Siefert, J. L. (2009). Defining the mobilome. In *Horizontal Gene Transfer* (pp. 13-27). Humana Press.
- Slonczewski, J. L., & Foster, J. W. (2013). *Microbiology: An Evolving Science: Third International Student Edition*. WW Norton & Company.
- Soksawatmaekhin, W., Kuraishi, A., Sakata, K., Kashiwagi, K., & Igarashi, K. (2004). Excretion and uptake of cadaverine by CadB and its physiological functions in *Escherichia coli*. *Molecular microbiology*, *51*(5), 1401-1412.
- Spring, S., Bunk, B., Spröer, C., Schumann, P., Rohde, M., Tindall, B. J., & Klenk, H. P. (2016). Characterization of the first cultured representative of Verrucomicrobia subdivision 5 indicates the proposal of a novel phylum. *The ISME journal*, *10*(12), 2801-2816.

- Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology*, *56*(4), 564-577.
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., ... & Rao, B. S. (2003). The COG database: an updated version includes eukaryotes. *BMC bioinformatics*, *4*(1), 41.
- Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current opinion in microbiology*, *11*(5), 472-477.
- Travisany, D., Di Genova, A., Sepúlveda, A., Bobadilla-Fazzini, R. A., Parada, P., & Maass, A. (2012). Draft genome sequence of the *Sulfobacillus thermosulfidooxidans* Cutipay strain, an indigenous bacterium isolated from a naturally extreme mining environment in Northern Chile. *Journal of bacteriology*, *194*(22), 6327-6328.
- Ullrich, S. R., González, C., Poehlein, A., Tischler, J. S., Daniel, R., Schlömann, M., ... & Mühling, M. (2016). Gene loss and horizontal gene transfer contributed to the genome evolution of the extreme acidophile “Ferrovum”. *Frontiers in microbiology*, *7*, 797.
- Valdes, J., Quatrini, R., Hallberg, K., Dopson, M., Valenzuela, P. D., & Holmes, D. S. (2009). Draft genome sequence of the extremely acidophilic bacterium *Acidithiobacillus caldus* ATCC 51756 reveals metabolic versatility in the genus *Acidithiobacillus*. *Journal of bacteriology*, *191*(18), 5877-5878.
- Valdés, J., Cárdenas, J. P., Quatrini, R., Esparza, M., Osorio, H., Duarte, F., ... & Holmes, D. S. (2010). Comparative genomics begins to unravel the ecophysiology of bioleaching. *Hydrometallurgy*, *104*(3-4), 471-476.
- van Teeseling, M. C., Pol, A., Harhangi, H. R., van der Zwart, S., Jetten, M. S., den Camp, H. J. O., & van Niftrik, L. (2014). Expanding the verrucomicrobial methanotrophic world: description of three novel species of *Methylacidimicrobium* gen. nov. *Applied and environmental microbiology*, *80*(21), 6782-6791.
- Varani, A. M., Siguier, P., Gourbeyre, E., Charneau, V., & Chandler, M. (2011). ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome biology*, *12*(3), R30.
- Wagner, M., & Horn, M. (2006). The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Current opinion in biotechnology*, *17*(3), 241-249.
- Vartoukian, S. R., Palmer, R. M., & Wade, W. G. (2010). Strategies for culture of ‘unculturable’ bacteria. *FEMS microbiology letters*, *309*(1), 1-7.
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, *13*(5), 303.
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., ... & Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, *12*(9), 635.