



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

PANGENÓMICA DEL *PHYLUM* THERMOPLASMATOTA

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

PRESENTA :

HILDA PALACIOS ÁLVAREZ

DIRECTOR DE TESIS:

DR. ARTURO CARLOS II BECERRA BRACHO

CD. MX. 2023





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

"The time will come I believe, though I shall not live to see it, when we shall have very fairly true genealogical trees of each great kingdom of nature."

Charles Darwin, carta a Thomas Huxley. Sep 1857

Agradecimientos

A mi familia, por acompañarme en cada paso y cada sueño.

A Yaru por pasar todas las noches de desvelos en mi regazo, acompañándome desde que era un pequeño cachorro y yo una niña pequeña hasta el momento en que partió.

A Liz, Emmi y Mondra porque desde aquel primer día ocuparon y seguirán ocupando un lugar especial en mi vida y en mi corazón.

A Ale, por compartir estos años conmigo.

Imgrin y Abelardo por ser mi lugar seguro en todo momento.

A los macaques, por las enseñanzas y risas que siempre acompañan en el laboratorio.

A mi tutor, Arturo, por guiarme en este camino pero al mismo tiempo darme la libertad de seguir imaginando y proponiendo.

A mis sinodales, por que cada uno formó una parte importante en mi formación como bióloga, siempre les agradeceré todas sus enseñanzas.

Índice general

Resumen	1
Introducción	2
Archaea	2
Thermoplasmatota	3
Thermoplasmata	5
Poseidoniia	6
Evolución y filogenias	9
Problemática en las inferencias filogenéticas	9
Transferencia horizontal y reducción del genoma	10
Pangenómica	13
Hipótesis	15
Objetivos	15
Metodología	16
Resultados	20
Discusión	28
Pangenoma arqueano	28
Inferencia filogenética	29
Ventajas y desventajas de la metodología	31
Conclusiones	33
Perspectivas	34
Bibliografía	35
Figuras suplementarias	47

Resumen

A pesar del incremento exponencial de la cantidad de secuencias disponibles pertenecientes a *Archaea*, sus relaciones evolutivas continúan siendo ambiguas. En este trabajo se infirió una filogenia a partir de datos pangenómicos (repertorio genético de un linaje) obtenidos mediante la identificación de genes ortólogos presentes en genomas completos del dominio Archaea, esto con el fin de conocer las relaciones evolutivas del *phylum Thermoplasmatota*. Se utilizaron caracteres discretos, tratando la ausencia/presencia de proteínas que conforman los proteomas arqueanos como si se tratara de la ausencia/presencia de caracteres morfológicos, un enfoque usualmente restringido al estudio de caracteres morfológicos en microorganismos. La mayoría de las filogenias inferidas para el Dominio Archaea o clados pertenecientes a éste, se basan en genes como 16S ó en la concatenación de genes, por lo que el uso de caracteres discretos para inferir la historia evolutiva en Archaea es un método innovador que puede permitirnos resolver relaciones ancestría-descendencia basales. Ésto dió como resultado la división del *phylum* en dos grandes grupos: uno conformado por la clase *Thermoplasmata* y otro por la clase *Poseidoniiia*, además de mostrarnos que, en su mayoría, son genes ribosomales los que están presentes y más conservados dentro y entre estos linajes.

Introducción

Archaea

Inicialmente, todos los microorganismos unicelulares no eucariotas se clasificaron dentro del grupo '*Prokaryota*', posteriormente se agruparon en el grupo '*Monera*'. Estas dos clasificaciones agrupan tanto a Bacteria como a Archaea en un solo grupo basado principalmente en la morfología celular (Baker *et al.*, 2020). Fue hasta 1977 cuando C.R. Woese y G. Fox propusieron que los seres vivos se habían diversificado a partir de tres líneas de descendencia: *Eubacteria*, *Archaea* y *Eukarya*, basado en la caracterización de secuencias de RNA ribosomal (rRNA) generando una asociación filogenética entre ellas (Woese & Fox, 1977).

Las Arqueas son microorganismos procariontes que poseen características moleculares distintivas tanto de bacterias como de eucariotas (Blohs *et al.*, 2019). En *Archaea*, la pared celular carece de peptidoglicano y la membrana celular está compuesta de éter-lípidos que son mucho más estables que los éter-lípidos específicos de *Eubacteria*. Además, el Dominio *Archaea* tiene algunas estructuras externas específicas, como *archellum*, pili, hami y cánulas, pero las funciones de estas estructuras aún no están claras (Efenberger *et al.*, 2014). Originalmente, estos organismos se descubrieron y aislaron de ecosistemas extremos, incluidos entornos volcánicos, lagos salados y otros biotopos que se caracterizan por temperaturas extremas, valores bajos de pH o altas concentraciones de hierro.

Durante las últimas décadas, los estudios han revelado que las arqueas están distribuidas universalmente y podrían estar entre los microorganismos más abundantes y activos en ciertos ambientes como la columna de agua del océano (Borrel *et al.*, 2020). Se ha sugerido que las arqueas representan más del 20% de los procariontes oceánicos y probablemente son los organismos más abundantes en los sedimentos del subsuelo marino y los hábitats geotérmicos (Eme & Doolittle, 2015). Pueden habitar ambientes extremos, típicamente caracterizados por alta temperatura o presión, como ventilas hidrotermales, o por alta salinidad, alcalinidad o acidez extremas. Estos microorganismos versátiles también residen en el tracto digestivo de los rumiantes y termitas (donde producen metano) y en el

suelo. Pueden vivir en condiciones anóxicas en el lodo e incluso en depósitos de petróleo (Esko, *et al.* 2009).

Las arqueas también son capaces de formar biopelículas interactuando con algunas otras bacterias (Efenberger *et al.*, 2014). Son organismos metabólicamente diversos que coexisten con Bacteria y Eukarya en la mayoría de los entornos, incluidos los extremos. Su predominio en el plancton marino, apunta a que tienen un papel crucial y aún poco conocido en los ciclos biogeoquímicos de nuestro planeta. Además, las arqueas son, hasta ahora, los únicos organismos capaces de realizar el proceso de metanogénesis (producción de metano a partir de H₂ y CO₂), y por esta razón son fundamentales para los estudios de paleoambientes y paleobiología (Gribaldo & Brochier-Armanet, 2006). En los últimos años se ha logrado un gran progreso en la comprensión de Archaea, pero aún se desconoce mucho de sus características biológicas (Efenberger *et al.*, 2014).

La clasificación taxonómica de *Archaea* enfrenta un problema. Inicialmente, la clasificación se basó en filogenias basadas en 16S rRNA, que dividía a Archaea en dos grandes grupos, Euryarchaeota y Crenarchaeota; sin embargo, el descubrimiento de nuevos linajes desafió esta clasificación bipartita. La posición filogenética de estos linajes aún es controversial, y la falta de secuencias de grupos externos en análisis previos dificulta determinar su estado taxonómico (Petitjean *et al.*, 2014). Estos problemas resaltan la necesidad de análisis filogenéticos más completos y el desarrollo de nuevos métodos para abordar los desafíos en la clasificación taxonómica de *Archaea* (Bromberg *et al.*, 2016).

Thermoplasmatota

Thermoplasmatota es un *phylum* propuesto en 2019, tras realizar un análisis filogenético mediante alineaciones concatenadas de marcadores genéticos. Está conformado por distintos clados previamente no caracterizados (Rinke *et al.*, 2019). *Thermoplasmatota* es un linaje de arqueas ecológicamente significativo y distribuido globalmente (Hu *et al.* 2021), incluye organismos metanógenos (*Methanomassiliicoccales*), fotoheterótrofos *Candidatus Poseidoniales* (anteriormente Marine Group II) y otros heterótrofos (Marine Group III y DHVEG-6). Además, recientemente, se ha demostrado que los miembros no clasificados de

Thermoplasmatota tienen el potencial genético para la reducción de sulfato, oxidación de hidrógeno y desnitrificación, así como la capacidad de fermentación (Flood *et al.*, 2021).

Debido a la variedad de hábitats y metabolismos observados dentro de este grupo *Thermoplasmatota* representa un modelo clave para estudiar los procesos subyacentes a la evolución y adaptación de las arqueas en entornos contrastantes. Los clados recientemente descritos de *Thermoplasmatota*, brindan la oportunidad de estudiar la compleja historia evolutiva de este grupo (Adam *et al.*, 2017; Sheridan *et al.*, 2022).

Debido a su amplia gama de estilos de vida y características celulares y genómicas específicas, este grupo único podría servir como un excelente modelo para el estudio de la adaptación ambiental y los procesos evolutivos dentro de las arqueas. Todos los miembros del orden *Thermoplasmatales* tienen un pH óptimo para el crecimiento extremadamente ácido (0.7-3) por lo tanto, la caracterización de estas arqueas deficientes en una pared celular puede proporcionar información importante sobre los mecanismos de tolerancia al ácido de la vida microbiana. Del mismo modo, la capacidad de reducir azufre vuelve a algunos miembros un modelo para estudios sobre la proteólisis y la acción consecutiva de proteasas y peptidasas. Por otro lado, la facultad de sobrevivir en entornos altamente ácidos, ricos en metales o metaloides es un foco de estudio de la adaptabilidad para prosperar en esos entornos (Justice *et al.*, 2012; Kletzin, 2007; Gupta *et al.*, 2021). La existencia de miembros tanto anaeróbicos como aeróbicos dentro de *Thermoplasmata* lo convierte en un grupo interesante para investigar los detalles de la transición evolutiva del metabolismo anaeróbico al aeróbico estricto. Además, los genomas de *Thermoplasmatota* contienen el tercer número más alto de genes desconocidos, lo que invita a pensar que los linajes pertenecientes a este phylum deben explorarse más a fondo (Zheng *et al.*, 2022). El limitado número de órdenes de *Thermoplasmatota* caracterizados hasta el momento dificulta nuestra comprensión del potencial metabólico, ecológico y de la historia evolutiva de este importante clado (Hu *et al.*, 2021; Yuan *et al.*, 2021).

Clasificación taxonómica

Dominio	<i>Archaea</i>
Phylum	<i>Thermoplasmatota</i>
Clases	<i>Thermoplasmata, Poseidonii</i>

Recuadro 1. Clasificación taxonómica de *Thermoplasmatota*.

Thermoplasmata

En estudios anteriores ya se ha descrito a *Thermoplasmata* como una clase profundamente ramificada (Compte-Port *et al.*, 2017) y fenotípicamente versátil, sin embargo, el conocimiento referente a su diversidad filogenética continúa siendo poco estudiado (Lino *et al.*, 2013).

Este clado se ha encontrado en una miríada de hábitats, tanto extremos como moderados, y es uno de los linajes más abundantes en los sedimentos globales, mostrando una distribución generalizada (Zhou *et al.*, 2019), entre los distintos hábitats en los que se han encontrado están los sedimentos de lagos de agua dulce, pantanos y el subsuelo marino, lagos kársticos estratificados (Compte-Port *et al.*, 2017), drenaje ácido de minas (AMD, por sus siglas en inglés), *Smoldering coal-refuse piles*, *Geothermally heated, acid soil*, (Kletzin, 2007), campos ácidos y solfatáricos (Lino *et al.*, 2013), rumen de distintos animales (Pinnell *et al.*, 2022) e incluso en partículas de plásticos del océano Pacífico (Tora *et al.*, 2022).

Son pocas las interacciones ecológicas que se han documentado dentro de este grupo, todas ellas con *phyla* arqueanos: con el *phylum Bathyarchaeota*, con quién muestran fuertes co-distribuciones en hábitats sedimentarios y cumplen con funciones involucradas en el reciclaje de carbono orgánico (Compte-Port *et al.*, 2017), con el *phylum Nitrososphaerota* en biopelículas de minas de hierro en donde se cree cumplen con una complementariedad metabólica (Gupta *et al.*, 2021) y con los *phyla* Lokiarchaeota y Hadarchaeota en sedimentos salinos y de agua dulce (Zhou *et al.*, 2019).

Capacidades metabólicas

Las capacidades metabólicas dentro de Thermoplasmata son sumamente diversas, entre sus distintos órdenes hay organismos heterótrofos, quimioheterótrofos, mixótrofos, litoautotróficos, acidófilos, aerobios, anaerobios, desde mesófilos a termófilos, con la capacidad de realizar fermentación alcohólica y acética, distintas vías autotróficas como la vía Wood-Ljungdahl o ciclo incompleto de dicarboxilato/4-hidroxiacetato, metanogénesis hidrogenotrófica y metilotrófica. Además, algunos clados están involucrados en rutas particulares como la reducción-oxidación de Fe/S, resistencia a metales pesados, o juegan un papel crucial en el ciclo del carbono (Gupta *et al.*, 2021; Lang *et al.*, 2015; Lino *et al.*, 2013; Zhou *et al.*, 2019).

A pesar de que este clado de arqueas está ampliamente distribuido alrededor del mundo y cumple con roles ecológicamente relevantes, continúa siendo poco estudiado, especialmente los miembros que habitan en ambientes moderados, esto es debido principalmente a la falta de representantes cultivados (Compte-Port *et al.*, 2017; Zhou *et al.* 2019).

La limitada información sobre la diversidad filogenética de este grupo también sesga las inferencias sobre estados ancestrales, en donde continúa el debate sobre si el ancestro era o no metanógeno y/o si esta capacidad fue adquirida a través de la transferencia horizontal de genes después de la divergencia de los linajes basales (Lino *et al.*, 2013; Zinke *et al.*, 2020).

Poseidoniiia

Ésta clase se ha identificado en muestras marinas de todo el mundo, siendo las arqueas planctónicas marinas más comunes en la superficie del océano (Pereira *et al.*, 2019). Habitan en la zona fótica oligotrófica, hábitats más costeros o incluso estuarinos (salobres), por lo que los subgrupos de esta clase pueden representar diferentes ecotipos (Zhang *et al.*, 2015).

Desde su descubrimiento por DeLong en 1992, a pesar de su distribución global y de representar una porción significativa del plancton microbiano en la zona fótica, el MGII (ahora clase Poseidoniiia) continúa siendo un grupo enigmático de organismos en el medio ambiente marino (Tully, 2019). Hasta la fecha los miembros de MG II no han sido cultivados; sin embargo, se ha obtenido información sobre su estilo de vida a través de estudios

metagenómicos (Ma *et al.*, 2020; Pereira *et al.*, 2019) pero su función ecológica precisa sigue siendo difícil de alcanzar (Zhang *et al.*, 2015).

Capacidades metabólicas

Los genes secuenciados incluyen funciones de codificación asociadas con metabolismos heterótrofos aeróbicos basados en diferentes sustratos como proteínas, carbohidratos, ácidos grasos y lípidos, respiración anaeróbica dentro de microambientes con bajo contenido de oxígeno, otros organismos muestran estilos de vida fotoheterótrofos a través de proteorodopsinas, estilos de vida defensivo, típico de los organismos expuestos a toxinas naturales, en este caso las floraciones de cianobacterias que producen biotoxinas marinas, reducción de sulfato, hidrolasas de glucósidos, que indican la degradación de sustratos de algas (Ma *et al.*, 2020; Orellana *et al.*; 2019, Pereira *et al.*, 2019; Tully, 2019; Zhang *et al.*, 2015)

El potencial metabólico y la distribución de los géneros MGII revelan roles distintos en el medio ambiente, identificando subclados costeros que degradan sacáridos de algas, subclados oceánicos superficiales oligotróficos que degradan proteínas y subclados mesopelágicos que carecen de proteorodopsinas, comunes en todos los demás subclados (Tully, 2019). El gran potencial de hidrólisis de polisacáridos podría desempeñar un papel importante en el ciclo del carbono del ecosistema marino (Qu *et al.*, 2022). En la clase Poseidoniia, el análisis de funciones metabólicas clave sugieren que el ancestro del grupo era un organismo fotoheterótrofo que habitaba aguas superficiales (basado en genes de proteorodopsinas), sin embargo la capacidad fototrófica se perdió completamente en distintos géneros (Rinke *et al.*, 2019).

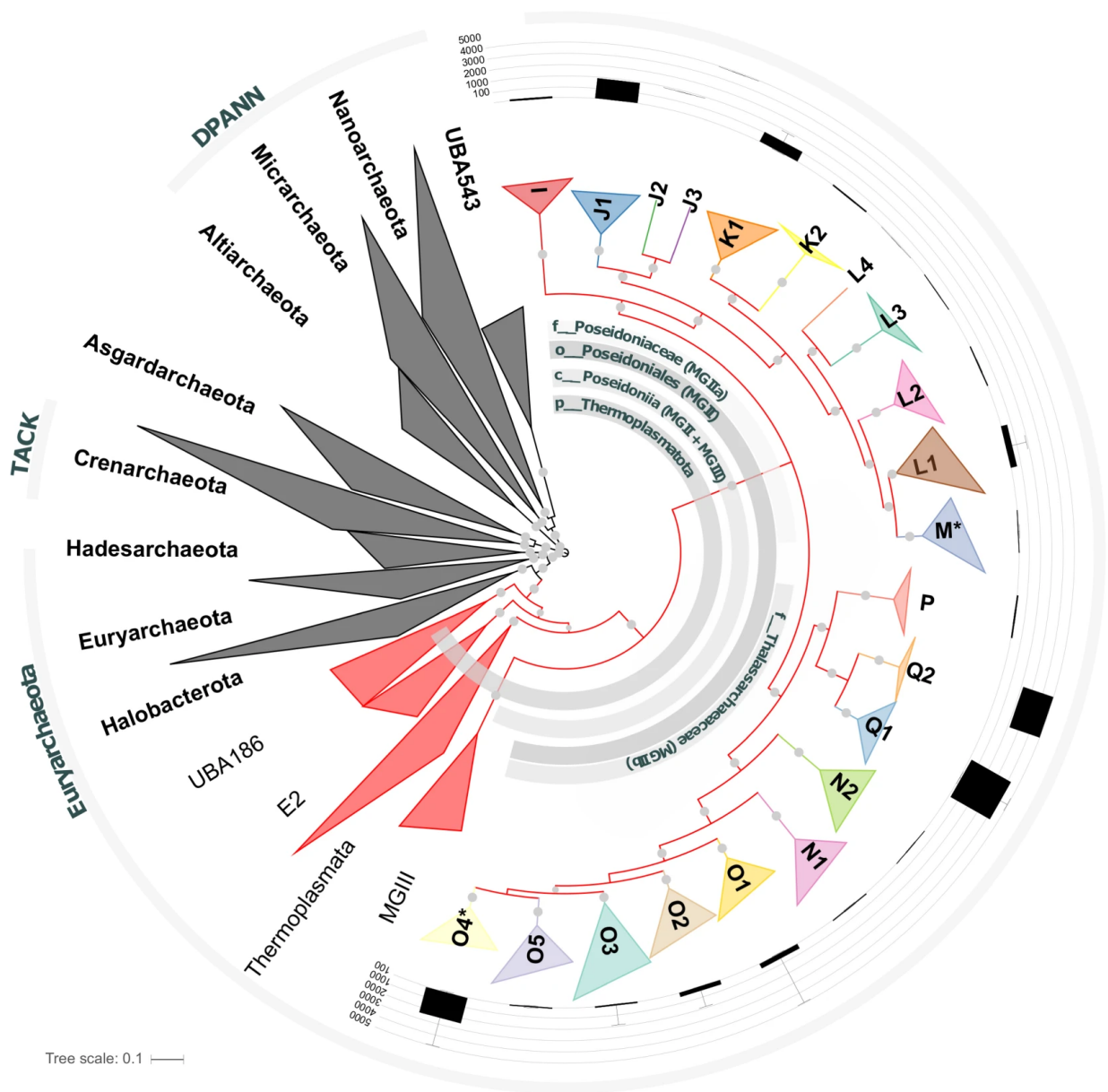


Figura 1. *Thermoplasmatota* como nuevo *phylum* del Dominio *Archaea*. En esta Inferencia filogenómica del MGII se propone la reclasificación a nivel taxonomico de *phylum* de la clase *Thermoplasmatota*, proponiendo el nombre *Thermoplasmatota* para el mismo. Dentro de este clado se incluyen los clados MGII, MGIII (Modificado de Rinke, 2019).

Evolución y filogenias

Podemos entender la evolución como el cambio en los caracteres heredables en las poblaciones a través de las generaciones. Con frecuencia, las relaciones evolutivas de las especies (o, más generalmente, taxa) se reconstruyen mediante un árbol filogenético basado en caracteres ancestrales de una naturaleza diversa, por ejemplo: moleculares, fenotípicas, geográficas o ecológicas. Éstas reconstrucciones involucran diferentes escalas de tiempo, que van desde unos pocos años para virus de evolución rápida (p. ej., ébola) a cientos de millones de años para los eucariontes pluricelulares (p. ej., las plantas) (Bordewich *et al.* 2018; Gascuel & Steel, 2020).

A lo largo de la historia se han utilizado distintos enfoques para dilucidar estas relaciones evolutivas entre los organismos. Algunos genes individuales como el 16S rRNA han tenido un enorme valor para el estudio de la filogenia de los procariontes en las últimas décadas. Se ha secuenciado ampliamente y existe una gran biblioteca de secuencias de referencia; sin embargo, los mecanismos de transporte lateral de genes entre los procariontes a lo largo del tiempo evolutivo son múltiples y diferentes a los de los eucariotas. Estos mecanismos incluyen transformación, transducción, conjugación, agentes de transferencia de genes e integrones. Como consecuencia, las historias genéticas de la gran mayoría de sus genes son discordantes, lo que significa que el modelo tradicional del árbol de la vida es un marco problemático para estudiar la evolución microbiana (Baptiste *et al.*, 2009).

Problemática en las inferencias filogenéticas de Procariontes

Los análisis comparativos de secuencias de genes y genomas indican que el intercambio de información genética dentro y entre especies procariontes, como quiera que se defina, es mucho más frecuente y general de lo que se pensaba anteriormente. Aunque el intercambio por recombinación homóloga está limitado por la divergencia de secuencia y debería disminuir notablemente con la "distancia filogenética", el intercambio por los diversos procesos de recombinación no homóloga, designados colectivamente como Transferencia Horizontal de Genes (HGT por sus siglas en inglés) no está tan limitado. La nueva comprensión de ambos fenómenos y su interacción potencial sugiere que los modelos

evolutivos tradicionales y las filogenias en forma de árbol, son inadecuados para describir los procesos y patrones evolutivos en procariontes (Gogarten *et al.* 2002).

El rápido recambio genético debido a la HGT y la pérdida de genes, pueden conducir a grandes diferencias genómicas entre bacterias y arqueas estrechamente relacionadas. Incluso organismos casi idénticos en la mayor parte del genoma pueden tener cientos de genes únicos, y una comparación reciente de redes de transferencia de genes entre cientos de genomas secuenciados ha sugerido que, en promedio, el 20% de los genes se han adquirido recientemente. Debido a esta gran diversidad, es difícil estimar el acervo genético total incluso para grupos estrechamente relacionados (Polz *et al.*, 2013).

Transferencia horizontal de genes y reducción de los genomas

La HGT es una de las fuerzas evolutivas más frecuentes e importantes pero menos entendidas. La HGT es el movimiento de material genético a través de los límites de cepas y especies (Baltrus, 2013), es decir, que se da de manera no genealógica, también es conocida como transferencia lateral de genes, o LTG (por sus siglas en inglés) (Boto, 2015). Los genes transferidos horizontalmente pueden pertenecer incluso a taxones superiores distintos (Mikhailovsky & Gordon, 2020), por ejemplo, entre arqueas y bacterias anaeróbicas y/o termófilas que coexisten en hábitats como fuentes termales, sedimentos marinos y pozos petroleros (Fuchsman *et al.*, 2017). Este mecanismo permite la adquisición de innovaciones evolutivas (Boto, 2010) y es de gran interés porque puede impulsar la innovación funcional a través de la introducción de nuevos genes y rutas metabólicas, además de que las adaptaciones impulsadas por la HGT podrían ser parte integral de las principales transiciones evolutivas (Fournier *et al.* 2015; Van Etten & Bhattacharya, 2020).

La HGT es tan común que la capacidad para inferir relaciones entre organismos a través del análisis de árboles filogenéticos ha sido cuestionada para bacterias y arqueas (Husnik & McCutcheon, 2018). Existe un debate bastante sólido entre quienes creen que la HGT en la evolución temprana pudo haber sido tan dominante que no es posible determinar un árbol de la vida para los clados más profundos y quienes sostienen que aún se puede

detectar algún patrón central (Syvanen, 2012). Esto presenta un desafío filosófico al esfuerzo de describir la historia evolutiva del mundo viviente utilizando la metáfora del Árbol de la vida, que excluye la participación de procesos distintos a la modificación gradual por descendencia (Boto, 2015).

Es posible que la evolución de *Archaea* haya sido impactada en gran medida por HGT de linajes distantes. Además, todas las arqueas que se han estudiado intensamente en laboratorio de forma experimental tienen mecanismos que permiten un intercambio de genes continuo y eficiente entre especies o cepas genéticamente relacionadas. La HGT entre arqueas y bacterias, aunque mucho más rara que la transferencia de genes entre arqueas, ha ocurrido a lo largo de la evolución, a menudo con consecuencias dramáticas. Se cree que las adquisiciones de genes bacterianos facilitó la transición de un estilo de vida hipertermófilo a temperaturas más bajas, lo que se refleja en la adquisición convergente de las mismas familias de genes por diferentes linajes de arqueas. De acuerdo con la naturaleza adaptativa de estos genes, las funciones de las proteínas codificadas tienden a involucrar el metabolismo energético, el transporte y metabolismo de aminoácidos y la biogénesis de lípidos o membranas (Gophna & Altman-Price, 2022).

La mayoría de los genomas de bacterias y arqueas son pequeños, al menos en comparación con los genomas de eucariontes. Los genomas procariontes son muy compactos, con genes codificadores de proteínas densamente empaquetados y una fracción baja de secuencias no codificantes. Se cree que el tamaño pequeño del genoma se selecciona para una replicación rápida, mientras que la alta densidad de genes también facilita la correulación de la expresión génica a través de la organización en operones (Sela *et al.*, 2016).

El rango de tamaños del genoma de las arqueas es mucho más estrecho que el de las bacterias. Los genomas más pequeños de *Archaea* y *Eubacteria* pertenecen a parásitos o simbioses, pero los organismos de vida libre de ambos grupos de hábitats con nutrientes limitados, como aguas marinas abiertas, también poseen genomas pequeños (Kellner *et al.* 2018). La eliminación de genes bajo la presión de la selección purificadora¹ no es el factor

¹ La selección natural negativa (también llamada selección purificadora) fija y mantiene rasgos biológicos adaptativos, proporciona estabilidad a largo plazo y elimina mutaciones nocivas (Estrada *et al.* 2022).

dominante de la evolución microbiana. Por el contrario, la adquisición de genes parece ser en gran parte un proceso adaptativo, aunque la selección positiva que gobierna la dinámica del genoma, en promedio, probablemente sea débil. Esta conclusión de ninguna manera contradice la teoría de la genética de poblaciones como tal, pero es incompatible con la suposición de que los genes recién adquiridos (y fijados en la población) son neutrales en promedio (Sela *et al.*, 2016; Wolf & Koonin, 2013). Recientemente, un número creciente de evidencias respaldan la importancia evolutiva de la reducción, en lugar de la amplificación, del tamaño del genoma. Autores como Bosi y Mascagni en el 2019, sugieren que en procariontes, la reducción del genoma también va acompañada de una simplificación real, en términos de complejidad del organismo (Bosi & Mascagni, 2019).

La secuenciación masiva ha permitido la recuperación de una gran cantidad de datos genómicos que han mejorado el muestreo taxonómico entre bacterias y arqueas lo que ha permitido realizar “filogenias de genomas completos” (árboles inferidos de la concatenación de numerosos genes marcadores), sin embargo, los genomas son sumamente dinámicos en contenido, tamaño y tasas evolutivas. Los genes pueden perderse, las regiones genómicas no codificantes o egoístas pueden expandirse o contraerse en escalas de tiempo cortas, diferentes genes tienen distintas tasas evolutivas debido a presiones selectivas desiguales, y los genes pueden obtenerse a través de la duplicación dentro de los genomas y la HGT, estas razones ponen en duda la idea de que el uso de más genes y genomas mejore necesariamente la calidad de los árboles resultantes (Husnik & McCutcheon, 2018; Martínez-Gutiérrez & Aylward, 2021). La concatenación de múltiples genes puede mejorar la precisión debido a un aumento en el número de caracteres filogenéticamente informativos en relación con los sitios de ruido. A pesar de esto, no está claro si la inclusión de más genes y genomas necesariamente mejora la calidad de los árboles resultantes y, de no ser así, qué conjuntos de genes marcadores y estrategias de muestreo de taxones producen las filogenias más sólidas. (Gadagkar *et al.* 2005; de Queiroz y Gatesy 2007).

Pangenómica

Los pangenomas surgen como consecuencia de la ganancia y pérdida constante de genes, siendo la primera comúnmente el resultado de la HGT en procariontes (Domingo-Sananes & McInerney, 2021). El pangenoma es definido como la colección de todos los grupos de genes ortólogos (OGGs) de un conjunto o muestra de genomas. Un pangenoma se puede dividir en tres categorías de genes ortólogos formadas por genes con diferente grado de presencia: (i) *core*, que incluye los OGGs representados en todos los genomas de la muestra; (ii) *shell*, que comprende genes presentes en una fracción considerable de genomas; y (iii) *cloud* OGGs, que contienen genes presentes en solo una pequeña fracción de los genomas (Moldovan & Gelfand, 2018).

En general, los genomas centrales (*core*) se encargan de las funciones constitutivas. Estos incluyen producción de energía, metabolismo de aminoácidos, metabolismo de nucleótidos, transporte de lípidos y maquinaria de traducción. Los genomas accesorios a menudo codifican genes implicados en el transporte de proteínas y defensa, así como en muchas funciones nicho- específicas. Además, los plásmidos, fagos y transposones también se asocian a menudo con genomas accesorios (Tettelin & Medini, 2020). La ganancia y pérdida de genes están sujetas a deriva y selección, lo que da como resultado los patrones típicos que observamos en los pangenomas. Estos patrones incluyen un aumento en el número de genes accesorios observados y una disminución en el número de genes *core* observados a medida que secuenciamos más genomas de la misma especie (Domingo-Sananes & McInerney, 2021).

El estilo de vida de las especies de interés es uno de los parámetros que dicta con fuerza la forma de distribución del pangenoma; por ejemplo, si por la adición recurrente de miembros del grupo, el pangenoma continúa creciendo, estamos analizando un pangenoma abierto (tales ejemplos incluyen patógenos humanos y bacterias ambientales). Por otro lado, si la complejidad del grupo se agota muy rápido, incluso a partir de los análisis de un puñado de miembros del grupo, entonces estamos ante un pangenoma cerrado en el que sólo necesitamos unos pocos representantes para describir la totalidad de la variabilidad de la secuencia (Tettelin & Medini, 2020). Hay una relación negativa general entre el tamaño del pangenoma y la proporción de genes *core*: los pangenomas “abiertos” son de mayor tamaño, tienen una

proporción menor de genes *core* y tasas más altas de ganancia genética por transferencia horizontal de genes (HGT), mientras que los pangenomas “cerrados” son de tamaño más pequeño, tienen una mayor proporción de genes *core* y tasas más bajas de ganancia genética por HGT. El equilibrio entre la ganancia y la pérdida de genes contribuye a la formación de un pangenoma. Es probable que las especies con pangenomas abiertos grandes ocupen nichos más variados y comunidades más complejas, y tengan poblaciones efectivas de mayor tamaño en comparación con las especies con pangenomas más pequeños (Brockhurst *et al.* 2019) .

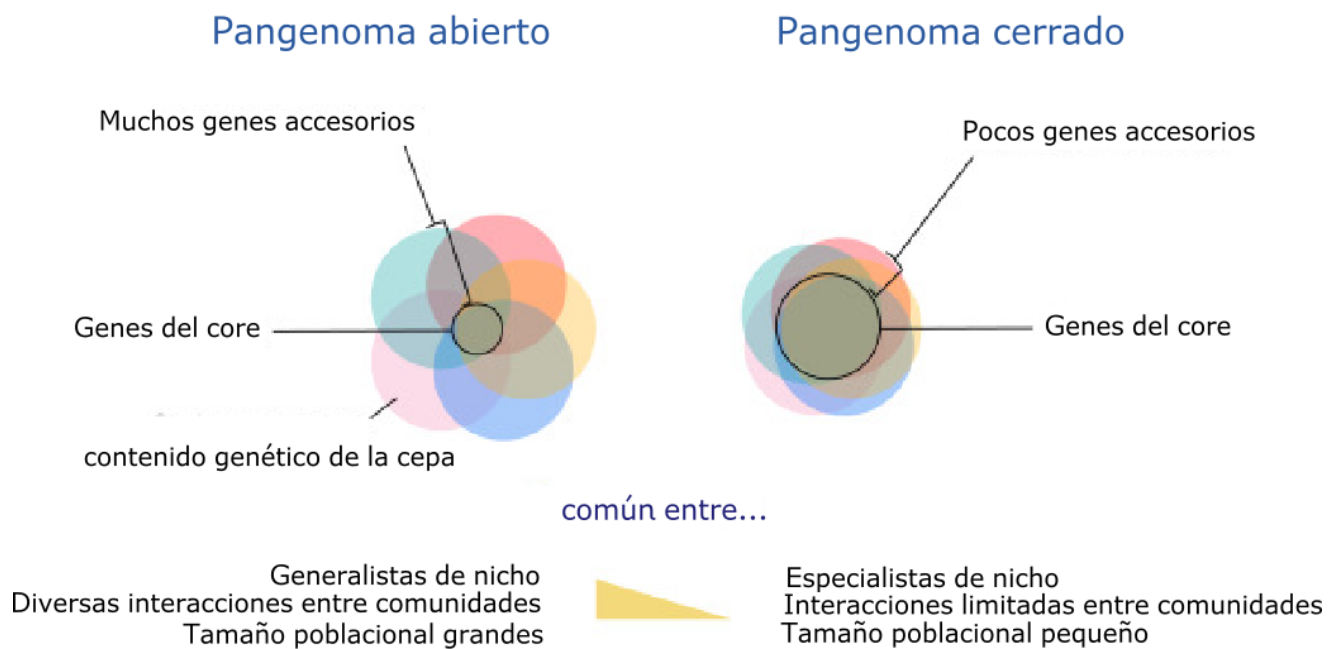


Figura 2. Componentes de los pangenomas. Diferencias entre pangenomas abiertos y cerrados, además de su relación con modos de vida de los organismos y poblaciones (Modificado de Brockhurst *et al.*, 2019).

Hipótesis

El estudio de proteomas arqueanos mediante análisis de la presencia/ausencia de proteínas a modo de caracteres discretos permitirá inferir algunas relaciones filogenéticas dentro del *phylum Thermoplasmatota*.

Objetivos

- Obtener el pangenoma de los principales *phyla* del dominio *Archaea*.
- Inferir una filogenia de *Archaea* a partir de datos pangenómicos.
- Resolver las relaciones ancestro-descendiente del *phylum Thermoplasmatota*.

Metodología

Obtención de datos

Selección de genomas

A través de Genome (versión *beta*) disponible en los servicios y bases de datos del *National Center for Biotechnology Information* (NCBI) (Sayers *et al.*, 2021) se recuperó una lista con todos los genomas disponibles del dominio *Archaea* utilizando su identificador taxonómico [2157]. Dado que no todos los genomas son adecuados para realizar estudios pangenómicos, se usó el nivel de ensamblaje (al menos en la categoría de *scaffold*) como criterio para seleccionar sólo aquellos genomas que fueran informativos en este estudio.

Extracción de metadatos y reducción de redundancia

Tras obtener una lista con los genomas pertenecientes a *Archaea*, ésta fue enriquecida con la taxonomía completa para cada especie de acuerdo a la clasificación de NCBI (Schoch *et al.*, 2020). Debido a que muchas especies contaban con más de un genoma se consideraron tres criterios para reducir esta redundancia:

1. Nivel de ensamblaje: en donde, *complete* > *chromosome* > *scaffold*
2. Conteo total de proteínas en el genoma: seleccionando los genomas con el mayor número de proteínas.
3. Fecha: seleccionando los genomas más recientes.

Debido a la sobrerrepresentación a nivel de clados, se empleó un enfoque progresivo, reduciendo al 60% clados con más de 500 genomas, 70% clados con 100 a 500 genomas y 80% clados con menos de 100 genomas, esta selección se hizo nuevamente bajo los tres criterios anteriores. Este proceso de selección dio como resultado una lista final de 429 genomas, que se descargaron utilizando NCBI *batch entrez*.

Análisis pangenómico

Para obtener el pangenoma de los 429 genomas seleccionados se utilizó el software Get_Homologues (Contreras-Moreira & Vinuesa, 2013). El genoma de referencia utilizado fue el de la especie *Methanosarcina acetivorans*. La siguiente tabla describe los parámetros utilizados:

Parámetro	Valor	Descripción
-t	0	Reportar grupos de secuencias que incluyan al menos t taxa (t=0 reporta todos los grupos [OMCL COGS]).
-M	N/A	Utilizar el algoritmo orthoMCL (OMCL, PubMed=12952885)
-C	75	Porcentaje mínimo de cobertura en alineaciones por pares de BLAST (rango [1-100])
-E	1e-05	Valor E máximo
-r	N/A	Proteoma de referencia. Archivo .faa/.gbk
-n	10	Número de subprocesos para BLAST/HMMER/MCL en modo de ejecución 'local'

Matriz pangenómica

La matriz que se construyó a partir de los grupos de genes ortólogos obtenidos con Get_homologues es una matriz con valores binarios que denotan la presencia (1) o ausencia (0) de una proteína en el genoma de una especie determinada. En esta matriz cada fila representa una especie distinta y cada columna una proteína.

La matriz pangenómica se filtra eliminando los grupos de ortólogos presentes en menos del 10% de los genomas, esto con el fin de realizar los análisis posteriores sólo con datos informativos (Anexo 2).

Análisis filogenético

Inferencia Bayesiana

El modelo Mk de evolución de rasgos discretos es una versión del modelo Jukes-Cantor de sustitución de nucleótidos generalizado para adaptarse a un número variable de estados de carácter (Jukes & Cantor, 1969). La matriz Q para un modelo Mk de dos estados es:

$$Q = \begin{pmatrix} -\mu_0 & \mu_{01} \\ \mu_{10} & -\mu_1 \end{pmatrix}$$

Utilizando la matriz binaria, se realizó un análisis filogenético basado en el modelo Mk implementado en RevBayes (Höhna *et al.*, 2016). Este enfoque supone que la ganancia y la pérdida de grupos ortólogos dentro de los linajes siguen un proceso de Markov reversible en el tiempo. La distribución de probabilidad posterior del árbol filogenético y los parámetros asociados se determinaron utilizando algoritmos Markov Chain Monte Carlo (MCMC).

Para éste análisis se emplearon cuatro cadenas, un período de calentamiento relativo del 25 %, el modelo LG y sitios invariables más modelos gamma para la variación de la tasa entre los sitios de la matriz. Se definió una distribución previa exponencial para longitudes de rama con un parámetro de 0.2 y los movimientos de topología de árbol se propusieron utilizando NNI (mvNNI) y SPR (mvSPR), con pesos proporcionales al número de ramas. Fue utilizada una distribución previa uniforme para la topología del árbol y las longitudes de las ramas, incorporando una distribución exponencial para las longitudes de las ramas. A la matriz de tasas de sustitución se le aplicó el modelo de Jukes-Cantor (Erickson, 2010),

mientras que para el parámetro alfa de la distribución *gamma* se estableció una distribución previa uniforme entre 0 y 1E6. Discretizando la distribución *gamma* en cuatro categorías y un movimiento de escala con un peso de 2. El modelo se ajustó a la matriz binaria y se utilizó MCMC con dos ejecuciones y un esquema de combinación "mixto" para las inferencias.

Se realizaron 20,000 iteraciones para asegurar suficiente convergencia y muestreo de la distribución posterior. Se extrajo el árbol máximo *a posteriori* (MAP), que representa el árbol filogenético con la probabilidad posterior más alta, para ofrecer una comprensión integral de las relaciones evolutivas entre los linajes de arqueas en función de la presencia y ausencia de grupos de ortólogos en sus genomas.

Resultados

Pangenoma

En este trabajo se utilizaron los genomas de 429 especies distintas, seleccionadas de forma que estuvieran representados todos los clados principales del Dominio *Archaea*, incluyendo grupos problemáticos como *Methanonatronarchaeia* o *Nanoarchaeota*. Tras el análisis pangenómico se obtuvo una matriz binaria de presencia-ausencia con un total de 7,695 grupos de proteínas ortólogas.

Las especies quedaron distribuidas en 28 clados, concordantes con clasificaciones taxonómicas actuales. Para representar gráficamente la diversidad y prevalencia de las proteínas que conforman el pangenoma arqueano, éste se dividió de modo que cada uno de los 28 clados conforman un sub-pangenoma. Hay proteínas que se encuentran presentes en todos los clados con una prevalencia del 100% o cercana a este valor, por otro lado existen proteínas que parecen ser exclusivas de clados específicos, un ejemplo es la gran diversidad de proteínas que conforman el pangenoma de *Halobacteria* (Figura 3).

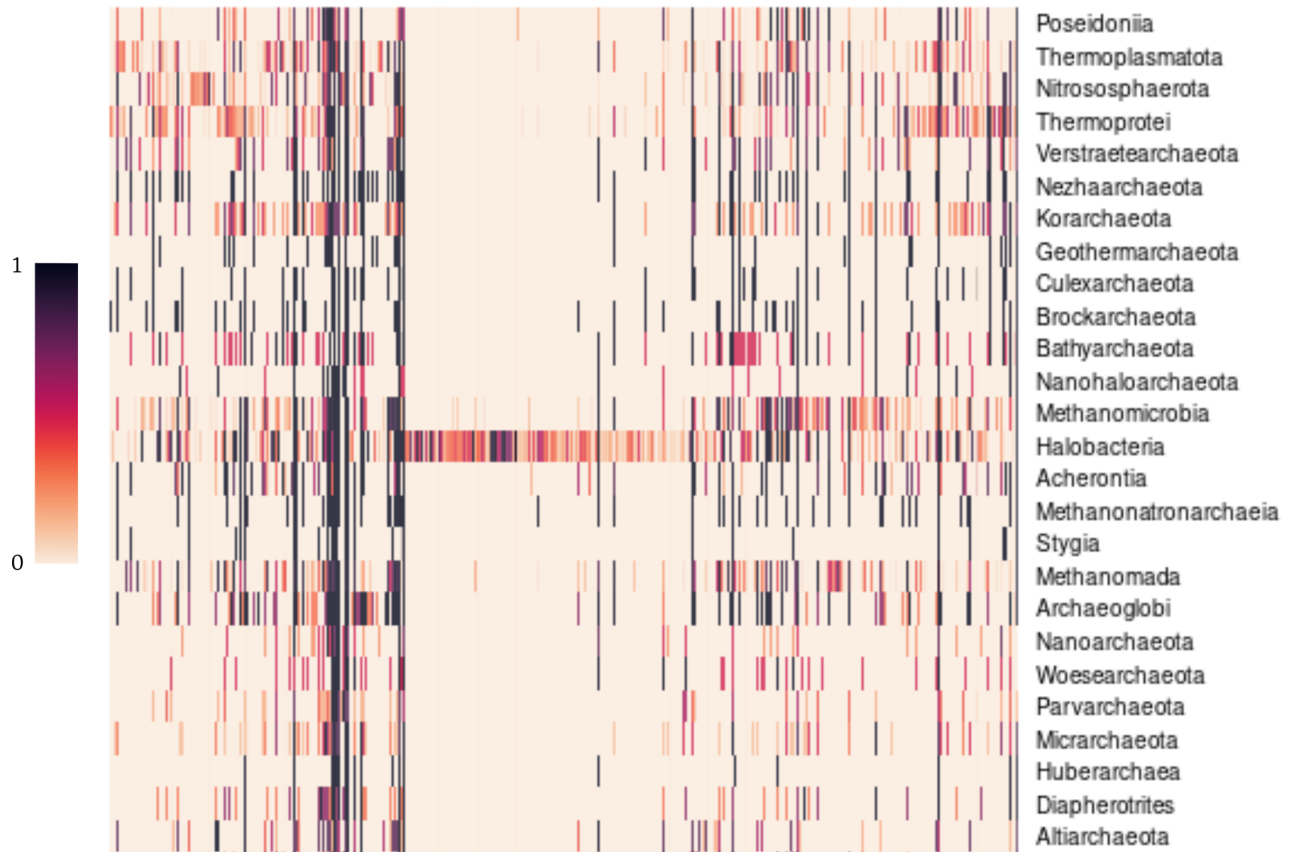


Figura 3. Pangenoma arqueano: En este mapa de calor el eje “y” corresponde a los clados arqueanos, mientras que el eje “x” representa los grupos de proteínas que conforman el pangenoma total. A medida que oscurece el color representa una prevalencia mayor de una proteína dentro de un clado, valores de 1 esquematizados en color negro representan a proteínas presentes en el 100% de organismos de un clado, por otro lado valores de 0 esquematizados en color crema indica la ausencia de una proteína.

Inferencia filogenética

A partir de la matriz pangenómica se realizó una inferencia filogenética mediante análisis bayesiano (Figura 4).

El *phylum Thermoplasmatota* se separa en dos grupos, uno conformado por la clase *Thermoplasmata* y otro por la clase *Poseidoniiia*, ambos clados cercanos al *superphylum* DPANN.

En DPANN los *phyla Candidatus Woesearchaeota*, *Ca. Nanohaloarchaea*, *Ca. Parvarchaeota*, *Nanoarchaeota*, *Ca. Undinarchaeota*, *Ca. Huberarchaea*, *Ca. Aenigmarchaeota* y *Ca. Diapherotrites* continúan agrupándose como un gran grupo monofilético. En el caso particular del *phylum Ca. Altiarchaeota*, también clasificado como parte de DPANN queda, no sólo fuera de este grupo, sino como un grupo independiente en la topología de la filogenia.

Dentro del *phylum Euryarchaeota*, la mayoría de los taxa pertenecientes continúan formando un clado monofilético, estos son: *Halobacteria*, *Methanonatronarchaeia*, *Archaeoglobi*, *Methanomada* y *Methanomicrobia*, sin embargo, los grupos *Stygia* (*Hadesarchaea*) y *Acherontia* forman clados independientes evolutivamente más cercanos a TACK.

Los *phyla* presentes del *superphylum* TACK son *Nezhaarchaeota*, *Verstraetearchaeota*, *Brockarchaeota*, *Nitrososphaerota* (antes *Thaumarchaeota*), *Culexarchaeota*, *Geothermarchaeota*, *Bathyarchaeota*, *Asgard*, *Korarchaeota*, *Thermoprotei* (antes *Crenarchaeota*).

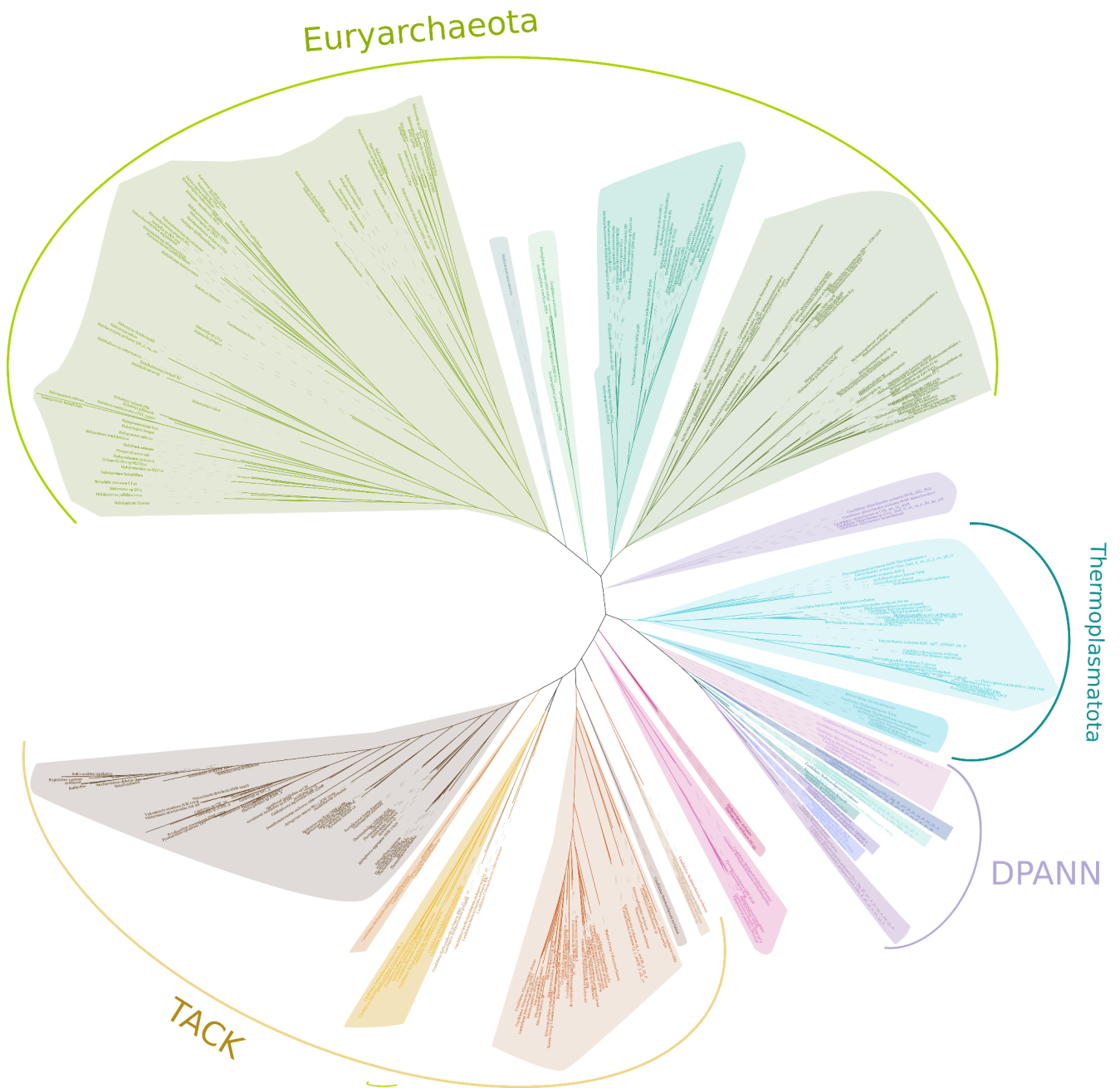


Figura 4. Árbol sin enraizar del dominio *Archaea*. Esta filogenia se generó a partir de una matriz pangenómica con 429 especies y ~ 7,600 grupos de proteínas ortólogas. Se utilizó el programa RevBayes para la inferencia bayesiana.

Thermoplasmatota

Analizando más detalladamente la filogenia, observamos como el *phylum* *Thermoplasmatota* queda segregado en dos grupos monofiléticos. Uno de estos grupos está conformado por especies de la clase *Poseidoniiia*, mientras que el segundo grupo, está conformado por especies de la clase *Thermoplasmata*. Dentro de esta topología, *Poseidoniiia* es grupo hermano del *superphylum* DPANN.

La clase *Poseidoniiia* incluye a un único orden: *Candidatus Poseidoniales*. En el otro grupo, representado por la clase *Thermoplasmata* se separa en tres clados distintos; uno en donde están representados los órdenes *Thermoprofundales*, *Ca. Sysuiplasmatales*, *Ca. Thermoprofundales* y *Thermoplasmatales*. En el siguiente clado se encuentra el orden *Methanomassiliicoccales*, en donde se encuentran los únicos organismos del *phylum* *Thermoplasmatota* con capacidad metanogénica, y en el último grupo se encuentra *Aciduliprofundales*, además de organismos que no han sido clasificados en algún orden hasta el momento.

Los nodos más basales están soportados por valores posteriores altos (1, el cual es el máximo valor posible), mientras que en nodos cercanos a ramas más terminales los valores posteriores son menores (Figura 5).

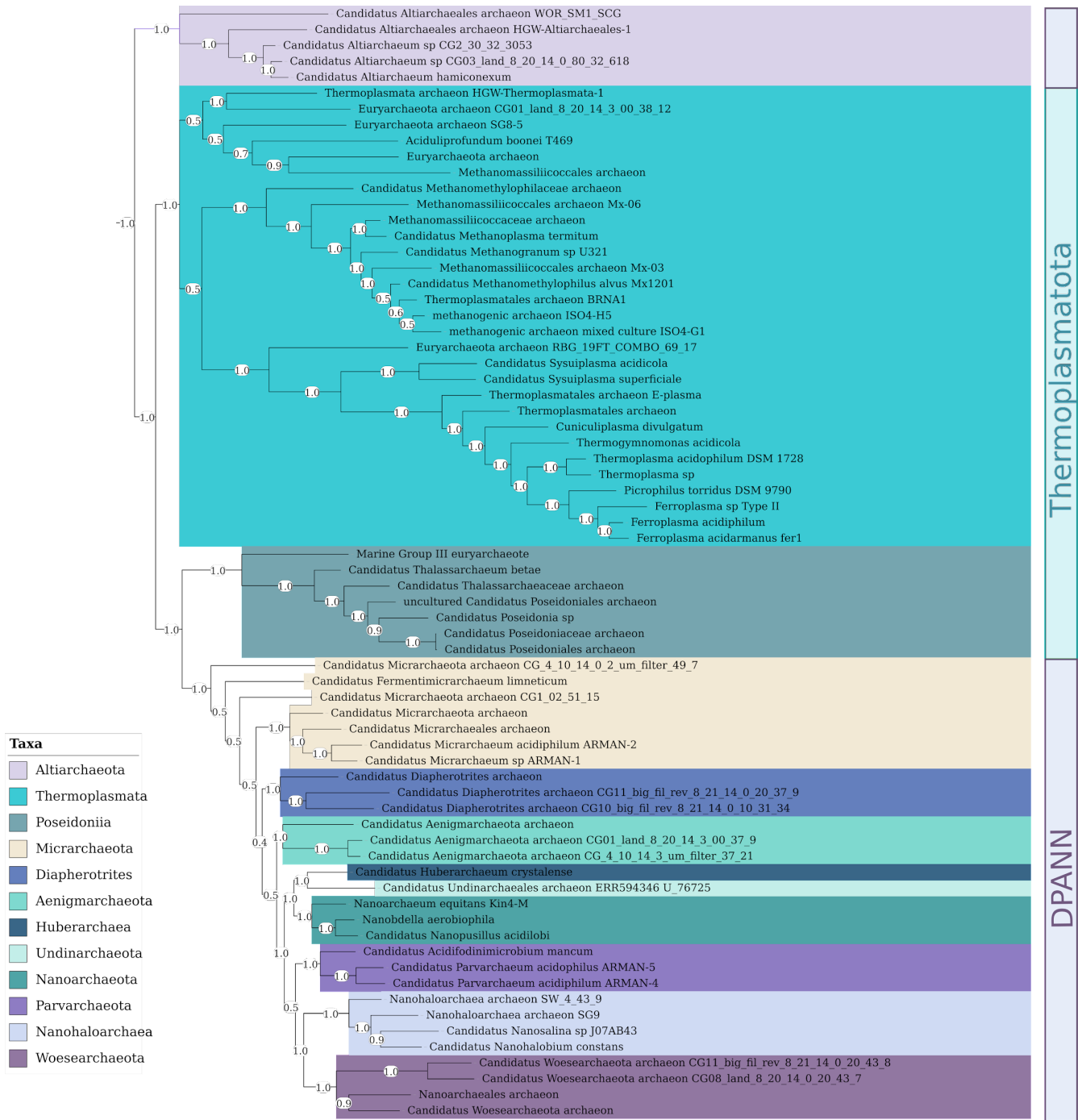


Figura 5. Relaciones evolutivas del *phylum Thermoplasmata*. Esta filogenia muestra la posición de las clases Poseidoniia y Thermoplasmata (*phylum Thermoplasmata*) en relación con el *superphylum* DPANN. Incluye los valores posteriores calculados para cada nodo.

A pesar de comenzar con una muestra de todas las arqueas, nuestro trabajo se centra en el *phylum Thermoplasmatota*. Ya que en la inferencia filogenética este *phylum* quedó separado, realizamos una comparación de los pangenomas de acuerdo a las dos clases en los que se separó (Figura 6).

El pangenoma de la clase *Poseidoniiia* se conforma por un total de 1063 proteínas, de éstas 465 están a nivel de *core* y 887 son parte del nivel “accesorio”. El pangenoma de la clase *Thermoplasmata* es de mayor tamaño, conformado por un total de 2726 proteínas, 450 de éstas están a nivel de *core*, y a nivel “accesorio” muestran una gran diversidad con un total de 2276 proteínas. Del total de proteínas que conforman los dos pangenomas, 910 están presentes en ambos pangenomas, lo que representa un ~85% del pangenoma de *Poseidoniiia* pero sólo el ~33% del pangenoma de *Thermoplasmata* (Figura 6b). La discrepancia entre estas dos clases se debe a la diversidad de proteínas accesorias presentes sólo en *Thermoplasmata*.

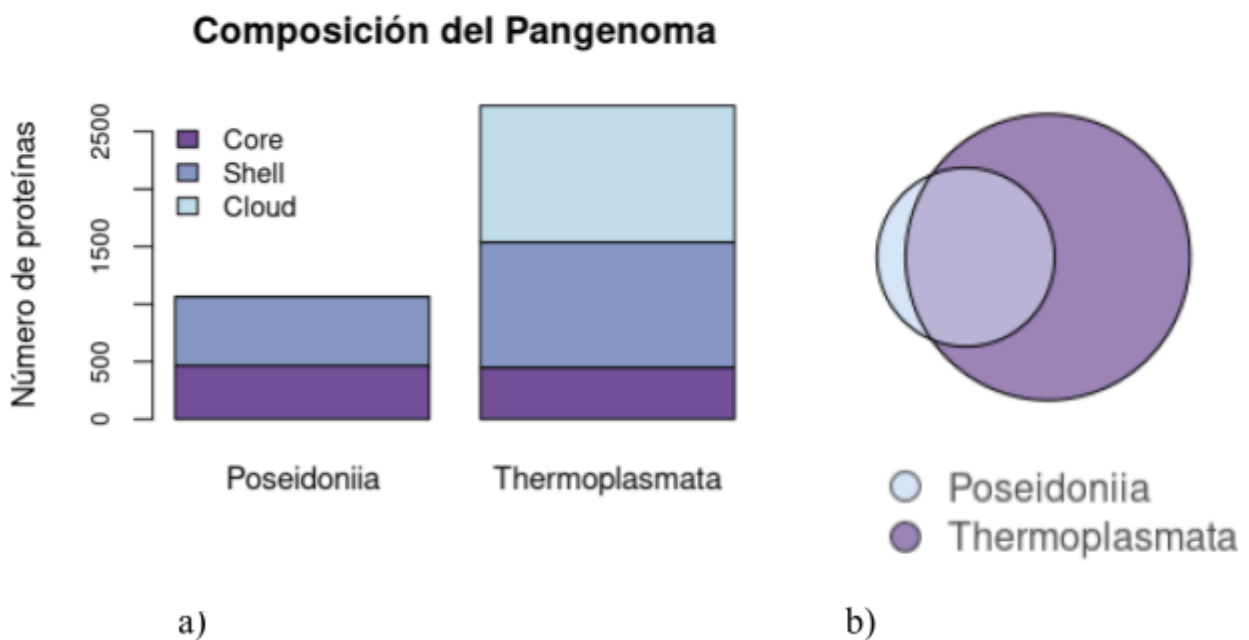


Figura 6. Comparación de pangenomas. a) Composición de pangenomas. b) Proporción de proteínas totales compartidas entre los pangenomas de *Thermoplasmata* y *Poseidoniiia*.

Los tamaños del *core* de ambas clases son bastante similares, y comparten el ~50% de estas proteínas, aunque en proporción, ~43% de las proteínas presentes en *Poseidoniiia* pertenecen al *core* mientras que en *Thermoplasmata* representan sólo el ~16% de las proteínas totales (Figura 7a). Por otro lado, del total de proteínas accesorias de los pangenomas, comparten un total de 352, representando un ~40% de las proteínas accesorias de *Poseidoniiia* y ~15% de las proteínas accesorias de *Thermoplasmata* (Figura 7b). El *core* se compone principalmente de proteínas ribosomales y proteínas hipotéticas (Figura 7c), en cuanto a las demás funciones, las proteínas se relacionan con funciones como actividad catalítica, unión, funciones estructurales, actividades dependientes de ATP, transportadores y regulación de la traducción y transcripción (Figura suplementaria 1). A nivel de clases de proteínas, se dividen en enzimas de interconversión de metabolitos, proteínas relacionadas con la traducción, reguladores transcripcionales, señalización intercelular, metabolismo de RNA, proteínas de unión, metabolismo de DNA y proteínas de tráfico de membrana (Figura suplementaria 2).

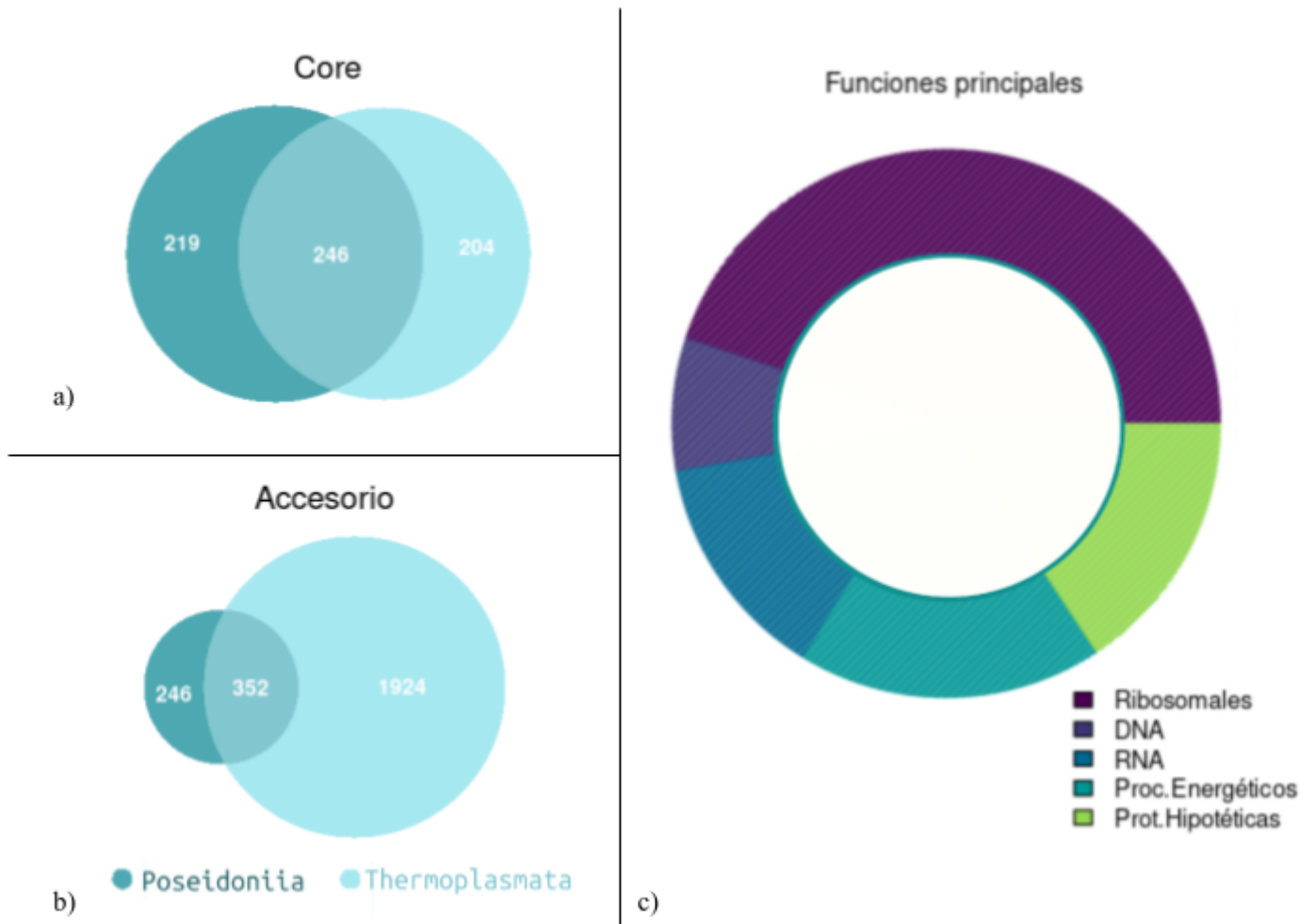


Figura 7. Proteínas compartidas entre las clases *Thermoplasmata* y *Poseidonii*.
a) Proteínas compartidas a nivel de *core*. b) Proteínas accesorias compartidas. c) Principales funciones de las proteínas compartidas por ambos grupos a nivel de *core*.

Discusión

Pangenoma arqueano

A pesar de que existen diversos estudios en donde se emplea un enfoque pangenómico en Archaeas, la mayoría se centran en grupos específicos como *Halobacteria* (Gaba *et al.*, 2020, Wang *et al.*, 2022), archaeas metanogénicas (Hansen *et al.*, 2011, Prondzinsky *et al.*, 2022), clados pertenecientes a *Euryarchaeota* (Deschamps *et al.*, 2014) o a clados clasificados dentro del grupo TACK (Banerjee *et al.*, 2021, González-Castillo *et al.*, 2022, Zhong *et al.*, 2021). Wolf y colaboradores realizaron un análisis pangenómico que abarca muchos de los clados principales del dominio, aunque la representación del grupo DPANN se limita a un único clado (Wolf *et al.*, 2012). Por esto, parece necesario realizar un análisis pangenómico en dónde están representados la mayoría de los clados arqueanos.

A medida que aumenta la cantidad de secuencias analizadas para obtener un pangenoma el *core* va reduciéndose y los genes presentes a nivel accesorio aumentan. En el pangenoma obtenido existe una gran diversidad en cuanto a tamaño de los sub-pangenomas de cada clado como en la prevalencia de las proteínas que los conforman, sin embargo, existen proteínas que se encuentran distribuidas y con prevalencias del 100% o valores cercanos en todos los clados arqueanos, por lo que son de especial interés para estudios futuros.

El clado con un mayor número de representantes dentro de este trabajo es *Halobacteria*, en donde también se aprecia que es el clado con la mayor diversidad de proteínas que conforman un sub-pangenoma, dentro de éstas, algunas pertenecientes tanto a su *core* como a un nivel accesorio no se encuentran en ningún otro clado arqueano. A pesar de que se requiere de futuros análisis para determinar la función exacta de éstas proteínas, el nivel accesorio suele relacionarse a distintas capacidades metabólicas por lo que éstas proteínas podrían estar relacionadas y jugar un papel clave en el estilo de vida halófilo de este clado.

Nezhaarchaeota, *Geothermarchaeota*, *Culexarchaeota*, *Brockarchaeota*, *Methanonatronarchaeia*, *Stygia* y *Huberarchaea* son clados propuestos recientemente y existen pocos genomas clasificados dentro de cada uno de estos clados. En sus sub-pangenomas respectivos todas las proteínas forman parte del *core*, esto puede deberse a que los pangenomas de éstos clados posiblemente sean pangenomas cerrados y no existan altas tasas de transferencia horizontal. Sin embargo, no puede descartarse que sea debido a un sesgo por el número reducido de genomas disponibles, problema que sólo podrá resolverse cuando existan más genomas secuenciados de especies pertenecientes a estos clados.

Debido a que el interés de este estudio se centra en el *phylum* Thermoplasmata se analizaron más detalladamente los sub-pangenomas obtenidos para éste grupo. La clase *Thermoplasmata* tiene un pangenoma abierto, esto quiere decir que muestra una gran diversidad de proteínas presentes a nivel accesorio. De acuerdo con estudios previos esta clase muestra una gran diversidad en cuanto a capacidades metabólicas y hábitats en los que está presente.

Por otra parte, *Poseidonii* muestra un pangenoma cerrado, la diversidad de proteínas que lo conforman es mucho menor comparándolo con la clase *Thermoplasmata*. Un gran porcentaje de las proteínas de éste sub-pangeoma conforman el *core*, además de que ni siquiera tiene proteínas clasificadas a nivel cloud (presentes en menos del 20% de organismos). En estudios previos se han aislado genomas pertenecientes a esta clase en hábitats restringidos, lo cual podría ser una explicación de la diversidad reducida de proteínas presentes en este grupo.

La conformación del *core* entre estas clases coincide con la propuesta de que los genes que lo conforman, se encuentran bajo presiones selectivas muy altas, tal es el caso de los genes ribosomales. Las proteínas presentes a nivel accesorio estarían más relacionadas con capacidades metabólicas determinadas para distintos hábitats y nichos.

Inferencia filogenética

La mayoría de los métodos utilizados en inferencias filogenéticas se enfocan en el uso de uno o pocos genes para dilucidar las relaciones evolutivas entre los organismos, el uso de genes concatenados puede ser una herramienta útil aunque complicada de utilizar, ya que distintos genes tienen distintas tasas evolutivas lo que puede causar un sesgo en los estudios.

El gen de RNA ribosomal de la subunidad pequeña 16S se ha utilizado con éxito para catalogar y estudiar la diversidad de especies y comunidades procarióticas, pero ofrece una resolución limitada a nivel de especie y niveles más finos, y no puede representar la diversidad y fluidez del genoma completo (Rodríguez-R *et al.*, 2018). Además, su uso suele ser problemático debido al sesgo de amplificación por PCR, la saturación derivada del uso de secuencias de nucleótidos y un número limitado de posiciones de alineación, que puede ser insuficiente para resolver las relaciones evolutivas entre linajes divergentes (Martínez-Gutiérrez & Aylward, 2021).

Aunque recientemente se han descubierto numerosos *phyla* nuevos de arqueas, habitualmente se reportan relaciones filogenéticas inconsistentes y muchas relaciones evolutivas entre *phyla* y entre dominios continúan siendo ambiguas (Martínez-Gutiérrez & Aylward, 2021). Esta disponibilidad de secuencias genómicas de una gran diversidad de arqueas ha desdibujado la línea tradicional entre *Euryarchaeota* y *Crenarchaeota* tal como la definieron G. Fox y otros pioneros de la investigación de arqueas a principios de la década de 1980 (Raymann *et al.*, 2015). La disponibilidad de nuevas secuencias continuará modificando estas topologías, generando conflictos taxonómicos y enriqueciendo nuestro conocimiento sobre las archaeas.

Por éstos motivos decidimos utilizar un método que fuera capaz de representar la diversidad y prevalencia proteómica dentro del Dominio *Archaea*, además de brindar un árbol filogenético a partir del conjunto entero de datos obtenidos.

El árbol filogenético obtenido al utilizar la información pangenómica del Dominio a modo de caracteres discretos dio como resultado un árbol en donde casi todos los grupos formados coinciden con clasificaciones recientes, aunque difiere en cuanto a la topología. El

único grupo que difiere de éstas clasificaciones es el phylum *Thermoplasmatota*, que quedó segregado en las dos grandes clases que lo conforman: *Thermoplasmata* y *Poseidoniiia*.

En filogenias anteriores se posiciona a *Thermoplasmatota* como parte de Euryarchaeota, aunque como otros clados, ha ido cambiando su posición dentro de las filogenias conforme aumenta el número de secuencias disponibles. En la topología obtenida, el *superphylum* DPANN (con excepción de *Altiarchaeaota*) forma un grupo con las clases *Thermoplasmata* y *Poseidoniiia*. DPANN se caracteriza por especies con genomas reducidos, y cabe mencionar que el *phylum* *Altiarchaeaota*, que queda excluido de este gran grupo, es el clado clasificado dentro de DPANN con los proteomas más numerosos. Por otro lado, los proteomas de *Poseidoniiia* tienen tamaños similares a los otros grupos que conforman DPANN.

La separación del *phylum* *Thermoplasmatota* puede no ser debida estrictamente a las relaciones filogenéticas, sino, ser sólo un reflejo de la gran diferencia entre sub-pangenomas antes mencionada, en donde la gran diversidad de proteínas accesorias de *Thermoplasmata* lo separaría de la clase *Poseidoniiia* en la topología de nuestro árbol. El mismo caso aplicaría para el *phylum* *Altiarchaeaota*, que cuenta con la mayor diversidad de proteínas dentro de DPANN y queda excluido de este *superphylum*.

El uso de caracteres discretos empleado en este trabajo, a pesar de ser un enfoque clásico, se ha restringido a estudios basados en caracteres morfológicos de distintos grupos eucariontes, por lo tanto, no había sido utilizado para estudios basados en caracteres moleculares, ni tampoco en estudios de grupos procariontes, por consiguiente, en *Archaea*. Sin embargo, resultó ser útil para elucidar las relaciones evolutivas de *Thermoplasmatota* y podría emplearse para resolver filogenias discordantes en otros grupos de organismos.

Ventajas y desventajas de la metodología

Al traducir el pangenoma a una matriz binaria se reduce la complejidad computacional, lo que permite realizar análisis más amplios, y reducir el sesgo informativo al incorporar un número mayor de secuencias pertenecientes a la mayoría de clados arqueanos al estudio. Además, el número relativamente alto de secuencias actualmente disponibles ha brindado una mayor oportunidad para controlar posibles sesgos como la transferencia horizontal de genes y las diferencias de tasas evolutivas entre linajes (Battistuzzi *et al.*, 2004).

Una de las desventajas al utilizar este enfoque es que grupos que difieran mucho en el tamaño pangenómico son propensos a ser separados, a pesar de posiblemente pertenecer al mismo clado. Este escenario haría necesario comparar algunos resultados con otro método que permita generar un consenso y determinar si dicho grupo es un grupo polifilético o monofilético.

Utilizar información pangenómica permite no sólo realizar inferencias filogenéticas, sino, también brinda información sobre la diversidad y prevalencia de todas las proteínas presentes a diferentes niveles taxonómicos, desde especie hasta *phyla* o incluso dominios. Esta información proteómica puede permitirnos conocer las características biológicas de las especies que sólo conocemos por secuenciación, desde inferir posibles metabolismos hasta interacciones ecológicas.

Conclusiones

La falta de representantes cultivados y datos genómicos curados para la mayoría de especies de *Archaea* continúa siendo un obstáculo en el estudio evolutivo y biológico de los grupos.

La inferencia filogenética realizada en este trabajo sugiere que el *phylum* Candidatus *Thermoplasmatota*, en realidad está conformado por dos clados independientes (clases *Thermoplasmata* y *Poseidonii*), pudiendo considerarse así como un grupo polifilético.

El *core* pangenómico de *Thermoplasmatota* se conforma por proteínas independientes a factores ambientales que se encuentran bajo grandes presiones selectivas, tal es el caso de proteínas ribosomales.

La inferencia filogenética presentada en este trabajo estuvo basada en el estudio de proteomas, mediante la interpretación de su contenido como caracteres discretos. Esto resultó ser un método útil no sólo para inferir relaciones filogenéticas entre los organismos, sino también para conocer sobre las características biológicas de los mismos, desde estilos de vida, habilidades metabólicas, interacciones ecológicas o incluso probables roles dentro de los ciclos biogeoquímicos globales.

Perspectivas

El pangenoma arqueano es sumamente diverso, estudios futuros pueden centrarse en las autopomorfias de distintos clados que nos permitan conocer y comprender los mecanismos que permiten sobrevivir y prosperar a estos organismos en distintos hábitats y bajo condiciones sumamente adversas (desde un punto de vista antropocéntrico). Una aproximación sería la determinación de funciones de proteínas clasificadas como hipotéticas dentro del grupo, además de determinar las rutas metabólicas características y sus posibles variaciones.

Bibliografia

1. Adam, P. S., Borrel, G., Brochier-Armanet, C., & Gribaldo, S. (2017). The growing tree of Archaea: New perspectives on their diversity, evolution and ecology. *The ISME Journal*, 11(11), 2407–2425. <https://doi.org/10.1038/ismej.2017.122>
2. Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., & Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nature Microbiology*, 5(7), 887–900. <https://doi.org/10.1038/s41564-020-0715-z>
3. Baltrus D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in ecology & evolution*, 28(8), 489–495. <https://doi.org/10.1016/j.tree.2013.04.002>
4. Banerjee, R., Chaudhari, N. M., Lahiri, A., Gautam, A., Bhowmik, D., Dutta, C., Chattopadhyay, S., Huson, D. H., & Paul, S. (2021). *Interplay of Various Evolutionary Modes in Genome Diversification and Adaptive Evolution of the Family Sulfolobaceae*. *Frontiers in Microbiology*, 12, 639995. <https://doi.org/10.3389/fmicb.2021.639995>
5. Baptiste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F.-J., Dupré, J., Dagan, T., Boucher, Y., & Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. *Biology Direct*, 4, 34. <https://doi.org/10.1186/1745-6150-4-34>
6. Battistuzzi, F. U., Feijao, A., & Hedges, S. B. (2004). A genomic timescale of prokaryote evolution: Insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evolutionary Biology*, 4, 44. <https://doi.org/10.1186/1471-2148-4-44>
7. Blohs, M., Moissl-Eichinger, C., Mahnert, A., Spang A., Dombrowski, N., Krupovic, M., Klingl, A. (2019). *Archaea- An Introduction*. *Encyclopedia of Microbiology*. 4 ed. 243-252. <https://doi.org/10.1016/B978-0-12-809633-8.20884-4>
8. Borrel, G., Brugère, J.F., Gribaldo, S., Schmitz, R., Moissi -Eichinger, C. (2020). *The host- associated archaeome*. *Nature Rev Microbiol* 18(11): 622-636. [doi: 10.1038/s41579-020-0407-y](https://doi.org/10.1038/s41579-020-0407-y)
9. Bordewich, M., Deutschmann, I. M., Fischer, M., Kasbohm, E., Semple, C., & Steel,

- M. (2018). On the information content of discrete phylogenetic characters. *Journal of mathematical biology*, 77(3), 527–544. <https://doi.org/10.1007/s00285-017-1198-2>
10. Bosi, E., & Mascagni, F. (2019). Less Is More: Genome Reduction and the Emergence of Cooperatio —Implications into the Coevolution of Microbial Communities. *International Journal of Genomics*, 2019, 2659175. <https://doi.org/10.1155/2019/2659175>
 11. Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal Society B: Biological Sciences*, 277(1683), 819-827
 12. Boto L. (2015). Evolutionary change and phylogenetic relationships in light of horizontal gene transfer. *Journal of biosciences*, 40(2), 465–472. <https://doi.org/10.1007/s12038-015-9514-8>
 13. Brockhurst, M., Harrison, E., Hall, J., Richards, T., McNally, A., MacLean, C. (2019). The Ecology and Evolution of Pangenomes. *Current Biology* 29, R1094–R1103. <https://doi.org/10.1016/j.cub.2019.08.012>
 14. Bromberg, R., Grishin, N. V., & Otwinowski, Z. (2016). Phylogeny Reconstruction with Alignment-Free Method That Corrects for Horizontal Gene Transfer. *PLOS Computational Biology*, 12(6), e1004985. <https://doi.org/10.1371/journal.pcbi.1004985>
 15. Cavicchioli, R. (2011). Archaea—Timeline of the third domain. *Nature Reviews Microbiology*, 9(1), Article 1. <https://doi.org/10.1038/nrmicro2482>
 16. Compte-Port, S., Subirats, J., Fillol, M., Sánchez-Melsió, A., Marcé, R., Rivas-Ruiz, P., Rosell-Melé, A., & Borrego, C. M. (2017). Abundance and Co-Distribution of Widespread Marine Archaeal Lineages in Surface Sediments of Freshwater Water Bodies across the Iberian Peninsula. *Microbial Ecology*, 74(4), 776–787. <https://doi.org/10.1007/s00248-017-0989-8>
 17. Contreras-Moreira, B., & Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology*, 79(24), 7696–7701. <https://doi.org/10.1128/AEM.02411-13>
 18. DeLong, E.F.. (1992). Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12).

- <https://doi.org/10.1073/pnas.89.12.5685>
19. Domingo-Sananes, M., McInerney, J. (2021). Mechanisms That Shape Microbial Pangenomes. *Trends in Microbiology* 29 (6). 493-503. <https://doi.org/10.1016/j.tim.2020.12.004>
 20. Deschamps, P., Zivanovic, Y., Moreira, D., Rodriguez-Valera, F., & López-García, P. (2014). Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic thaumarchaeota and euryarchaeota. *Genome Biology and Evolution*, 6(7), 1549–1563. <https://doi.org/10.1093/gbe/evu127>
 21. Efenberger, M., Brzezińska-Błaszczuk, E., & Wódz, K. (2014). [Archaeons—Still unknown microorganisms]. *Postepy Higieny I Medycyny Doswiadczalnej* (Online), 68, 1452–1463. <https://doi.org/10.5604/17322693.1131697>
 22. Eme, L., Doolittle, F. (2015). Archaea. *Current Biology* 25 (19). R851-R855. <https://doi.org/10.1016/j.cub.2015.05.025>
 23. Erickson, K. (2010). The Jukes-Cantor Model of Molecular Evolution. *PRIMUS*, 20(5), 438–445. <https://doi.org/10.1080/10511970903487705>
 24. Esko, J.D., Doering, T., Raetz, C. (2009). *Chapter 20: Eubacteria and Archaea*. *Essentials of Glycobiology*. 2nd ed.
 25. Estrada, A., Suárez-Díaz, E., & Becerra, A. (2022). Reconstructing the Last Common Ancestor: Epistemological and Empirical Challenges. *Acta Biotheoretica*, 70(2), 15. <https://doi.org/10.1007/s10441-022-09439-1>
 26. Flood, B. E., Louw, D. C., Plas, A. K. V. der, & Bailey, J. V. (2021). Giant sulfur bacteria (Beggiatoaceae) from sediments underlying the Benguela upwelling system host diverse microbiomes. *PLOS ONE*, 16(11), e0258124. <https://doi.org/10.1371/journal.pone.0258124>
 27. Fournier, G. P., Andam, C. P., & Gogarten, J. P. (2015). Ancient horizontal gene transfer and the last common ancestors. *BMC evolutionary biology*, 15, 70. <https://doi.org/10.1186/s12862-015-0350-0>
 28. Fuchsman, C., Collins, R., Rocap, G., Brazelton, W. (2017). *Effect of the*

environment on horizontal gene transfer between bacteria and archaea. PeerJ 29;5:e3865. doi: 10.7717/peerj.3865

29. Gaba, S., Kumari, A., Medema, M., & Kaushik, R. (2020). Pan-genome analysis and ancestral state reconstruction of class halobacteria: Probability of a new super-order. *Scientific Reports*, 10, 21205. <https://doi.org/10.1038/s41598-020-77723-6>
30. Gascuel, O., & Steel, M. (2020). A Darwinian Uncertainty Principle. *Systematic biology*, 69(3), 521–529. <https://doi.org/10.1093/sysbio/syz054>
31. Gogarten, J. P., Doolittle, W. F., & Lawrence, J. G. (2002). Prokaryotic Evolution in Light of Gene Transfer. *Molecular Biology and Evolution*, 19(12), 2226–2238. <https://doi.org/10.1093/oxfordjournals.molbev.a004046>
32. González-Castillo, A., Carballo, J. L., & Bautista-Guerrero, E. (2022). Genomics, Phylogeny, and in Silico Phenotyping of Nitrosopumilus Genus. *Current Microbiology*, 80(1), 3. <https://doi.org/10.1007/s00284-022-03121-y>
33. Gophna, U., & Altman-Price, N. (2022). Horizontal Gene Transfer in Archaea-From Mechanisms to Genome Evolution. *Annual Review of Microbiology*, 76, 481–502. <https://doi.org/10.1146/annurev-micro-040820-124627>
34. Grbaldo, S., Brochier-Armanet, C. (2006). *The origin and evolution of Archaea: a state of the art*. *Phil. Trans. R. Soc. B*. 361, 1007-1022 doi:10.1098/rstb.2006.1841
35. Gupta, A., Saha, A., & Sar, P. (2021). Thermoplasmata and Nitrososphaeria as dominant archaeal members in acid mine drainage sediment of Malanjkhand Copper Project, India. *Archives of Microbiology*, 203(4), 1833–1841. <https://doi.org/10.1007/s00203-020-02130-4>
36. Hansen, E. E., Lozupone, C. A., Rey, F. E., Wu, M., Guruge, J. L., Narra, A., Goodfellow, J., Zaneveld, J. R., McDonald, D. T., Goodrich, J. A., Heath, A. C., Knight, R., & Gordon, J. I. (2011). Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proceedings of the National Academy of Sciences of the United States of America*, 108(Suppl 1), 4599–4606. <https://doi.org/10.1073/pnas.1000071108>
37. Höhna, Landis, Heath, Boussau, Lartillot, Moore, Huelsenbeck, Ronquist. (2016).

- [RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language](#). *Systematic Biology*, 65:726-736.
38. Hu W, Pan J, Wang B, Guo J, Li M, Xu M. (2021). Metagenomic insights into the metabolism and evolution of a new Thermoplasmata order (Candidatus Gimiplasmatales). *Environ Microbiol.* 2021 23(7):3695-3709. <https://doi.org/10.1111/1462-2920.15349>
39. Husnik, F., & McCutcheon, J. P. (2018). Functional horizontal gene transfer from bacteria to eukaryotes. *Nature reviews. Microbiology*, 16 (2), 67–79. <https://doi.org/10.1038/nrmicro.2017.137>
40. Jain, A., & Krishnan, K. P. (2021). Marine Group-II archaea dominate particle-attached as well as free-living archaeal assemblages in the surface waters of Kongsfjorden, Svalbard, Arctic Ocean. *Antonie Van Leeuwenhoek*, 114(5), 633. <https://doi.org/10.1007/s10482-021-01547-1>
41. Jendoubi, T. (2021). Approaches to Integrating Metabolomics and Multi-Omics Data: A Primer. *Metabolites*, 11(3), 184. <https://doi.org/10.3390/metabo11030184>
42. Joshua, C. J. (2019). Metabolomics: A Microbial Physiology and Metabolism Perspective. En E.
43. E. K. Baidoo (Ed.), *Microbial Metabolomics: Methods and Protocols* (pp. 71–94). Springer. <https://doi.org/10.1007/978-1-4939-8757-3>
44. Judge, A., & Dodd, M. S. (2020). Metabolism. *Essays in Biochemistry*, 64(4), 607–647. <https://doi.org/10.1042/EBC20190041>
45. Justice, N. B., Pan, C., Mueller, R., Spaulding, S. E., Shah, V., Sun, C. L., Yelton, A. P., Miller, C. S., Thomas, B. C., Shah, M., VerBerkmoes, N., Hettich, R., & Banfield, J. F. (2012). Heterotrophic Archaea Contribute to Carbon Cycling in Low-pH, Suboxic Biofilm Communities. *Applied and Environmental Microbiology*, 78(23), 8321–8330. <https://doi.org/10.1128/AEM.01938-12>
46. Kellner, S., Spang, A., Offre, P., Szöλλösi, G. J., Petitjean, C., & Williams, T. A. (2018). Genome size evolution in the Archaea. *Emerging Topics in Life Sciences*, 2(4), 595–605. <https://doi.org/10.1042/ETLS20180021>
47. Kletzin, A. (2007). General Characteristics and Important Model Organisms. En *Archaea* (pp. 14–92). John Wiley & Sons, Ltd. <https://doi.org/10.1128/9781555815516.ch2>

48. Lerat, E., Daubin, V., & Moran, N. A. (2003). From Gene Trees to Organismal Phylogeny in Prokaryotes: The Case of the γ -Proteobacteria. *PLoS Biology*, 1(1), e19. <https://doi.org/10.1371/journal.pbio.0000019>
49. Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13(9), 2178–2189. <https://doi.org/10.1101/gr.1224503>
50. Lin, X., Handley, K. M., Gilbert, J. A., & Kostka, J. E. (2015). Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota and Thermoplasmata in deep anoxic peat. *The ISME Journal*, 9(12), 2740–2744. <https://doi.org/10.1038/ismej.2015.77>
51. Iino, T., Tamaki, H., Tamazawa, S., Ueno, Y., Ohkuma, M., Suzuki, K., Igarashi, Y., & Haruta, S. (2013). Candidatus Methanogranum caenicola: A Novel Methanogen from the Anaerobic Digested Sludge, and Proposal of Methanomassiliicoccaceae fam. nov. and Methanomassiliicoccales ord. nov., for a Methanogenic Lineage of the Class Thermoplasmata. *Microbes and Environments*, 28(2), 244–250. <https://doi.org/10.1264/jsme2.ME12189>
52. Ma, C., Coffinet, S., Lipp, J. S., Hinrichs, K.-U., & Zhang, C. (2020). Marine Group II Euryarchaeota Contribute to the Archaeal Lipid Pool in Northwestern Pacific Ocean Surface Waters. *Frontiers in Microbiology*, 11, 1034. <https://doi.org/10.3389/fmicb.2020.01034>
53. Martinez-Gutierrez, C. A., & Aylward, F. O. (2021). Phylogenetic Signal, Congruence, and Uncertainty across Bacteria and Archaea. *Molecular Biology and Evolution*, 38(12), 5514–5527. <https://doi.org/10.1093/molbev/msab254>
54. McInerney, J. O., McNally, A., & O'Connell, M. J. (2017). Why prokaryotes have pangenomes. *Nature Microbiology*, 2(4), Article 4. <https://doi.org/10.1038/nmicrobiol.2017.40>
55. Mikhailovsky, G., & Gordon, R. (2020). Shuffling type of biological evolution based on horizontal gene transfer and the biosphere gene pool hypothesis. *Bio Systems*, 193-194, 104131. <https://doi.org/10.1016/j.biosystems.2020.104131>
56. Moldovan, M., Gelfand, M. (2018). Pangenomic Definition of Prokaryotic Species and the Phylogenetic Structure of *Prochlorococcus* spp. *Front. Microbiol.* 9; 428.

- <https://doi.org/10.3389/fmicb.2018.00428>
57. Orellana, L. H., Francis, T. B., Krüger, K., Teeling, H., Müller, M.-C., Fuchs, B. M., Konstantinidis, K. T., & Amann, R. I. (2019). Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota. *The ISME Journal*, 13(12), 3024. <https://doi.org/10.1038/s41396-019-0491-z>
58. Pereira, O., Hochart, C., Auguet, J. C., Debroas, D., & Galand, P. E. (2019). Genomic ecology of Marine Group II, the most common marine planktonic Archaea across the surface ocean. *MicrobiologyOpen*, 8(9), e00852. <https://doi.org/10.1002/mbo3.852>
59. Petitjean, C., Deschamps, P., López-García, P., & Moreira, D. (2015). Rooting the Domain Archaea by Phylogenomic Analysis Supports the Foundation of the New Kingdom Proteoarchaeota. *Genome Biology and Evolution*, 7(1), 191–204. <https://doi.org/10.1093/gbe/evu274>
60. Pinnell, L., Reyes, A., Wolfe, C., Weinroth, M., Metcalf, J., Delmore, R., Belk, K., Morley, P., Engle, T. (2022). Bacteroidetes and Firmicutes Drive Differing Microbial Diversity and Community Composition Among Micro-Environments in the Bovine Rumen. *Front. Vet. Sci.*, 19. <https://doi.org/10.3389/fvets.2022.897996>
61. Polz, M. F., Alm, E. J., & Hanage, W. P. (2013). Horizontal Gene Transfer and the Evolution of Bacterial and Archaeal Population Structure. *Trends in genetics : TIG*, 29(3), 170–175. <https://doi.org/10.1016/j.tig.2012.12.006>
62. Prondzinsky, P., Toyoda, S., & McGlynn, S. E. (2022). The methanogen core and pangenome: Conservation and variability across biology's growth temperature extremes. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 30(1), dsac048. <https://doi.org/10.1093/dnares/dsac048>
63. Psomopoulos, F. E., van Helden, J., Médigue, C., Chasapi, A., & Ouzounis, C. A. (2020). *Ancestral state reconstruction of metabolic pathways across pangenome ensembles*. *Microbial genomics*, 6(11), mgen000429. <https://doi.org/10.1099/mgen.0.000429>
64. Qu, L., Cai, R., Hu, Z., & Wang, H. (2022). Metagenomic assemblage genomes

- analyses reveal the polysaccharides hydrolyzing potential of marine group II euryarchaea. *Environmental Research*, 209, 112865. <https://doi.org/10.1016/j.envres.2022.112865>
65. Raymann, K., Brochier-Armanet, C., & Gribaldo, S. (2015). The two-domain tree of life is linked to a new root for the Archaea. *Proceedings of the National Academy of Sciences of the United States of America*, 112(21), 6670–6675. <https://doi.org/10.1073/pnas.1420858112>
66. Rinke, C., Rubino, F., Messer, L. F., Youssef, N., Parks, D. H., Chuvochina, M., Brown, M., Jeffries, T., Tyson, G. W., Seymour, J. R., & Hugenholtz, P. (2019). A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. Nov.). *The ISME Journal*, 13(3), Art. 3. <https://doi.org/10.1038/s41396-018-0282-y>
67. Rodriguez-R, L. M., Gunturu, S., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., Cole, J. R., & Konstantinidis, K. T. (2018). The Microbial Genomes Atlas (MiGA) webserver: Taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Research*, 46, W282–W288. <https://doi.org/10.1093/nar/gky467>
68. Sambamoorthy, G., Sinha, H., & Raman, K. (2019). Evolutionary design principles in metabolism. *Proceedings of the Royal Society B: Biological Sciences*, 286(1898), 20190098. <https://doi.org/10.1098/rspb.2019.0098>
69. Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., ... Sherry, S. T. (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 50(D1), D20–D26. <https://doi.org/10.1093/nar/gkab1112>
70. Schoch, C. L., Ciufu, S., Domrachev, M., Hottot, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database: The Journal of Biological Databases and Curation*, 2020, baaa062.

<https://doi.org/10.1093/database/baaa062>

71. Sela, I., Wolf, Y. I., & Koonin, E. V. (2016). Theory of prokaryotic genome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 113(41), 11399–11407. <https://doi.org/10.1073/pnas.1614083113>
72. Sheridan, P. O., Meng, Y., Williams, T. A., & Gubry-Rangin, C. (2022). Recovery of Lutacidiplasmatales archaeal order genomes suggests convergent evolution in Thermoplasmatota. *Nature Communications*, 13(1), Art. 1. <https://doi.org/10.1038/s41467-022-31847-7>
73. Syvanen M. (2012). Evolutionary implications of horizontal gene transfer. *Annual review of genetics*, 46, 341–358. <https://doi.org/10.1146/annurev-genet-110711-155529>
74. Tettelin, H., Medini, D. (2020). The Pangenome: Diversity, Dynamics and Evolution of Genomes. Springer <https://doi.org/10.1007/978-3-030-38281-0>
75. Tora, D., Hentschel, U., Lips, S., Schmitt-Jansen, M., & Borchert, E. (2022). 16s rRNA gene sequence analysis of the microbial community on microplastic samples from the North Atlantic and Great Pacific Garbage Patches (p. 2022.12.22.521553). bioRxiv. <https://doi.org/10.1101/2022.12.22.521553>
76. Tully, B. J. (2019). Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. *Nature Communications*, 10, 271. <https://doi.org/10.1038/s41467-018-07840-4>
77. Van Etten, J., & Bhattacharya, D. (2020). Horizontal Gene Transfer in Eukaryotes: Not if, but How Much?. *Trends in genetics : TIG*, 36(12), 915–925. <https://doi.org/10.1016/j.tig.2020.08.006>
78. Wang, S., Narsing Rao, M. P., Wei, D., Sun, L., Fang, B.-Z., Li, W.-Q., Yu, L.-H., & Li, W.-J. (2022). Complete genome sequencing and comparative genome analysis of the extremely halophilic archaea, *Haloterrigena daqingensis*. *Biotechnology and Applied Biochemistry*, 69(4), 1482–1488. <https://doi.org/10.1002/bab.2220>
79. Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, 74(11), 5088–5090.

<https://doi.org/10.1073/pnas.74.11.5088>

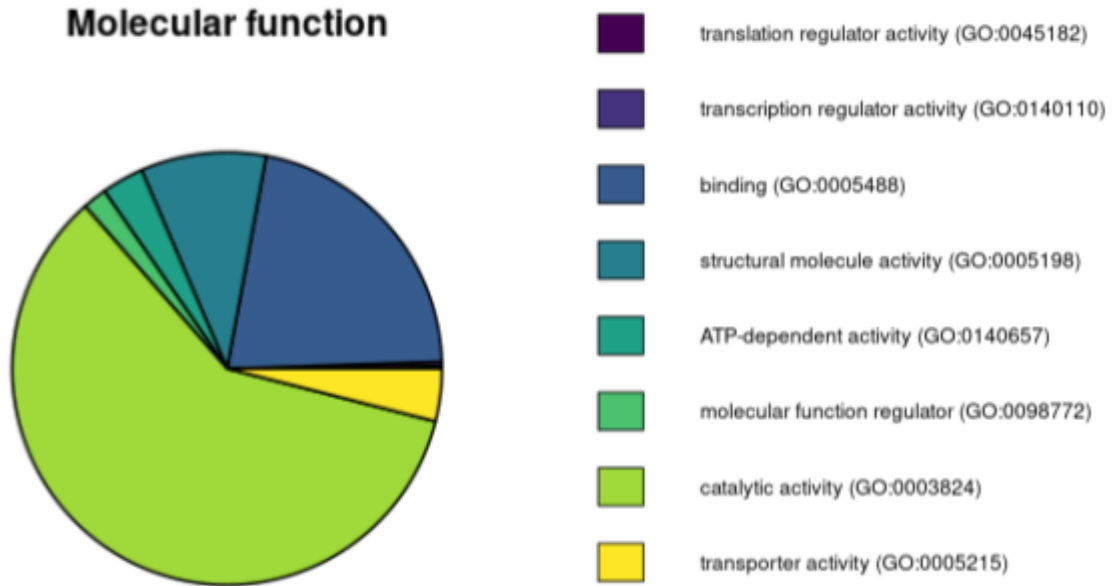
80. Wolf, Y. I., Makarova, K. S., Yutin, N., & Koonin, E. V. (2012). Updated clusters of orthologous genes for Archaea: A complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biology Direct*, 7, 46. <https://doi.org/10.1186/1745-6150-7-46>
81. Wolf, Y. I., & Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. *Bioessays*, 35(9), 829–837. <https://doi.org/10.1002/bies.201300037>
82. Xavier, J. C., Gerhards, R. E., Wimmer, J. L. E., Brueckner, J., Tria, F. D. K., & Martin, W. F. (2021). The metabolic network of the last bacterial common ancestor. *Communications Biology*, 4, 413. <https://doi.org/10.1038/s42003-021-01918-4>
83. Xu, B., Fan, L., Wang, W., Zhu, Y., & Zhang, C. (2023). Diversity, distribution, and functional potentials of magroviruses from marine and brackish waters. *Frontiers in Microbiology*, 14, 1151034. <https://doi.org/10.3389/fmicb.2023.1151034>
84. Yuan, Y., Liu, J., Yang, T.-T., Gao, S.-M., Liao, B., & Huang, L.-N. (2021). Genomic Insights into the Ecological Role and Evolution of a Novel Thermoplasmata Order, “Candidatus Sysuiplasmatales”. *Applied and Environmental Microbiology*, 87(22), e01065-21. <https://doi.org/10.1128/AEM.01065-21>
85. Zhang, C. L., Xie, W., Martin-Cuadrado, A.-B., & Rodriguez-Valera, F. (2015). Marine Group II Archaea, potentially important players in the global ocean carbon cycle. *Frontiers in Microbiology*, 6, 1108. <https://doi.org/10.3389/fmicb.2015.01108>
86. Zheng, P.-F., Wei, Z., Zhou, Y., Li, Q., Qi, Z., Diao, X., & Wang, Y. (2022). Genomic Evidence for the Recycling of Complex Organic Carbon by Novel Thermoplasmata Clades in Deep-Sea Sediments. *mSystems*, 7(3), e00077-22. <https://doi.org/10.1128/msystems.00077-22pollutants>
87. Zhong, C., Wang, L., & Ning, K. (2021). Pan-genome study of Thermococcales reveals extensive genetic diversity and genetic evidence of thermophilic adaption. *Environmental Microbiology*, 23(7), 3599–3613. <https://doi.org/10.1111/1462-2920.15234>
88. Zhou, Z., Liu, Y., Lloyd, K. G., Pan, J., Yang, Y., Gu, J.-D., & Li, M. (2019).

Genomic and transcriptomic insights into the ecology and metabolism of benthic archaeal cosmopolitan, Thermoprofundales (MBG-D archaea). *The ISME Journal*, 13(4), Article 4. <https://doi.org/10.1038/s41396-018-0321-8>

89. Zinke, L. A., Evans, P. N., Santos-Medellín, C., Schroeder, A. L., Parks, D. H., Varner, R. K., Rich, V. I., Tyson, G. W., & Emerson, J. B. (2021). Evidence for non-methanogenic metabolisms in globally distributed archaeal clades basal to the Methanomassiliicoccales. *Environmental Microbiology*, 23(1), 340–357. <https://doi.org/10.1111/1462-2920.15316>

Figuras suplementarias

1. Funciones moleculares de las proteínas *core*



2. Clase de proteínas del *core*

