



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

GENÓMICA COMPARATIVA DE LA ADQUISICIÓN DE HIERRO POR SIDERÓFOROS EN BACTERIAS

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:
LESLIE NUÑEZ LÓPEZ

TUTOR PRINCIPAL
DR. LUIS DAVID ALCARAZ PERAZA
[Facultad de Ciencias, UNAM](#)

MIEMBROS DEL COMITÉ TUTOR
DR. SAMUEL CANIZALES QUINTEROS
[Facultad de Química \(Unidad periférica INMEGEN\), UNAM](#)
DRA. MARÍA SOLEDAD FUNES ARGÜELLO
[Instituto de Fisiología Celular, UNAM](#)

Ciudad Universitaria, Ciudad de México. Enero, 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

Índice

Índice	2
Agradecimientos académicos	4
Resumen	5
Introducción	6
Importancia y disponibilidad del hierro en los sistemas biológicos.....	6
Sistemas bacterianos de adquisición de hierro.....	7
Sideróforos.....	8
Transporte de sideróforos en bacterias	11
Consecuencias fisiológicas de la producción de sideróforos	15
Genómica comparativa	16
Recursos bioinformáticos y bases de datos	17
Búsqueda bioinformática de proteínas homólogas	18
Perfiles filogenéticos y genómica comparativa	21
Justificación	24
Hipótesis	24
Objetivo general	24
Objetivos particulares.....	24
Metodología	25
Base de datos de proteínas que participan en la adquisición del hierro.....	25
Identificación de proteínas homólogas de síntesis y transporte de sideróforos	25
Clasificación taxonómica de las proteínas de síntesis y transporte de sideróforos	26
Construcción de modelos estadísticos para proteínas asociadas a la adquisición de hierro a partir de alineamientos múltiples de secuencias.....	26
Obtención de proteomas bacterianos para la búsqueda de sideróforos	27
Búsqueda de proteínas homólogas de síntesis y transporte de sideróforos en proteomas bacterianos	27
Obtención de matriz de datos con las proteínas homólogas de síntesis y transporte de sideróforos	28
Análisis de la abundancia y distribución de las proteínas homólogas asociadas a la adquisición de hierro	28
Resultados	30
Construcción de la base de datos de referencia para proteínas involucradas en la biosíntesis y transporte de sideróforos	30
Distribución filogenética de los sideróforos en Bacteria	35

Búsqueda masiva de perfiles ocultos de Markov, para generar perfiles filogenéticos	
Identificación, abundancia y distribución de las proteínas homólogas	41
Discusión.....	57
Conclusiones	71
Perspectivas.....	72
Material suplementario	73
Literatura consultada.....	74

Agradecimientos académicos

Al Posgrado en Ciencias Bioquímicas de la Universidad Nacional Autónoma de México.

El presente trabajo se realizó gracias al apoyo económico de los proyectos: DGAPA-PAPIIT-UNAM TA2001171 y SEP-CONACyT Ciencia Básica 237387.

Agradezco al CONACyT (CVU 904516) por haberme otorgado apoyo económico durante mis estudios de maestría.

Este proyecto de maestría fue elaborado bajo la dirección del Dr. Luis David Alcaraz Peraza en el Laboratorio de Genómica Ambiental de la Facultad de Ciencias de la Universidad Nacional Autónoma de México.

Resumen

El aumento de oxígeno en la atmósfera primitiva repercutió en la disponibilidad de hierro, el ion ferroso era abundante en el océano hasta que la oxigenación de la atmósfera terrestre provocó su oxidación y precipitación. Este cambio en la biodisponibilidad del hierro probablemente ejerció una presión selectiva sobre las estrategias que las bacterias desarrollaron para su absorción altamente selectiva. Debido a que la mayoría de las bacterias estudiadas hasta el momento, bajo condiciones experimentales controladas, tienen la capacidad de producir sideróforos y es posible que esto esté asociado con una gran diversidad de mecanismos moleculares, tanto para la síntesis, como para el transporte de hierro mediado por los sideróforos a través de las membranas celulares. En la presente tesis, se hicieron búsquedas de proteínas relacionadas con el metabolismo de sideróforos desde la síntesis, el transporte, la respuesta al estrés oxidativo y la regulación de la transcripción de los sideróforos. La identificación se realizó mediante la construcción y búsqueda de perfiles de modelos ocultos de Markov (pHMM). Para la búsqueda inicial de proteínas, se utilizó la base datos SEED *subsystems*, la cual se simplificó en 1,007 familias de proteínas que representan la variabilidad de proteínas bacterianas conocidas para el aprovechamiento de hierro a través de sideróforos. Los pHMMs generados fueron capaces de predecir proteínas homólogas en una selección de 116 proteomas bacterianos. De esta forma exploramos la distribución filogenética de los sideróforos, desconociendo si era un patrón ubicuo o si existía especialización de algunas bacterias. Se pudo identificar a distintos organismos especialistas en la producción de sideróforos, así como analizar otros patrones basados en la frecuencia y distribución filogenética de estas proteínas. Observamos una distribución discreta de las proteínas, así como expansiones génicas en algunos organismos y con ello se pudo identificar a las bacterias *Pseudomonas syringae* y *Photorhabdus luminescens* como organismos especialistas en la adquisición de hierro mediante sideróforos. La mayor parte de las bacterias estudiadas dedica al menos 10% de su proteoma al aprovechamiento de hierro a través de sideróforos; sin embargo, los patrones asociados al tipo de proteínas que posiblemente emplea cada bacteria correlaciona estrechamente con la biología y ecología de cada taxón en particular.

Introducción

Importancia y disponibilidad del hierro en los sistemas biológicos

El hierro es un micronutriente esencial para el crecimiento y el metabolismo bacteriano, primordialmente por su actividad como cofactor en las enzimas para catalizar las reacciones redox involucradas en procesos celulares como la respiración, la síntesis de DNA y la protección de las especies reactivas de oxígeno (*ROS*, por sus siglas en inglés). Sin embargo, su biodisponibilidad ha sido determinante en la historia evolutiva de los organismos porque a pesar de ser de los principales tipos de metales que conforman la corteza terrestre, la demanda de hierro a menudo excede las concentraciones de hierro disponibles para los microorganismos (Gu et al., 2020; Kaplan & Ward, 2013; Kramer, Özkaya, & Kümmerli, 2020)□.

Generalmente, se puede encontrar al hierro en sus dos estados redox: la forma férrica (Fe^{3+}) u oxidada que es la más abundante en la naturaleza, así como la forma ferrosa (Fe^{2+}) o reducida, son estos estados redox alternativos lo que permiten que el hierro desempeñe un papel versátil dentro de los sistemas biológicos (S. Andrews et al., 2013)□. Se ha estimado que la concentración total de especies inorgánicas de hierro en el suelo, en condiciones aeróbicas y pH de 7.0, es de aproximadamente 1×10^{-10} M (Colombo, Palumbo, He, Pinton, & Cesco, 2014)□. Bajo estas condiciones, el Fe^{2+} se oxida rápidamente a Fe^{3+} y en ausencia de un ligando orgánico fuerte puede precipitarse rápidamente como un óxido férrico hidratado. La solubilidad de los óxidos de hierro en sistemas aeróbicos depende de las propiedades de los sólidos, del tamaño de partícula y del pH, la fuerza iónica y la concentración de ligandos orgánicos en solución (Kraemer, 2004)□.

La concentración limitante de hierro para una célula bacteriana es variable y depende de la especie y de la condición de crecimiento; sin embargo, se ha estimado que la concentración crítica intracelular para el crecimiento de algunas bacterias es de 1×10^{-6} M (Miethke & Marahiel, 2007)□. En cuanto a la concentración extracelular que representa un estado de estrés o tóxico para la célula es de 1×10^{-3} M de hierro en sus formas Fe^{2+} y Fe^{3+} , provocando la inhibición del crecimiento bacteriano (N Kalantari & Ghaffari, 2008; Narges Kalantari, 2008)□. Se sabe también que la mayor parte de ese hierro intracelular, al menos

para *Escherichia coli* creciendo en medio rico y en fase estacionaria, corresponde a hierro férrico, lo cual puede tener coincidencia con que el hierro ferroso es altamente reactivo y debe almacenarse en su forma oxidada (S. Andrews et al., 2013)□.

En conjunto, esto nos indica que a pesar de ser un elemento abundante, el hierro tiene baja biodisponibilidad, por lo que los organismos han desarrollado intrincados mecanismos para absorber, metabolizar y almacenar hierro, así como regular estos procesos para mantener su homeostasis (Connorton, Balk, & Rodríguez-Celma, 2017)□.

Sistemas bacterianos de adquisición de hierro

Por una parte el hierro es un nutriente indispensable para la supervivencia, mientras que por otro lado puede ser nocivo en concentraciones limitantes o excesivas debido a su toxicidad y su capacidad para generar estrés oxidativo (S. C. Andrews, Robinson, & Rodríguez-Quñones, 2003; Golonka, Yeoh, & Vijay-Kumar, 2019)□. Para controlar esta relación dicotómica, las células bacterianas regulan estrechamente la producción y la actividad de los sistemas de captación y almacenamiento de hierro. Se sugiere que estos sistemas de regulación son más complejos que su desinhibición a consecuencia de la disminución de la disponibilidad del hierro en el citosol de la célula (Crosa & Walsh, 2002).

Con frecuencia se hace referencia a que existen dos principales estrategias para aprovechar las fuentes de hierro en el medio. La primera es a través de moléculas específicas y de alta afinidad que facilitan el transporte del hierro al interior de la célula bacteriana. La segunda involucra transportadores específicos de tipo ABC (*ATP-binding cassette transporters*, por sus siglas en inglés) y que incluye la interacción con receptores de alta afinidad que se encuentran en la superficie celular con la capacidad de transferir complejos de hierro o hierro libre al interior de la célula (Hayrapetyan, Siezen, Abee, & Groot, 2016)□.

La captación del hierro puede tener inicio con su asociación con sideróforos en el medio; sucede a través de diferentes estrategias de regulación en donde activadores y represores génicos pueden mediar la actividad de diferentes genes que codifican proteínas para el ensamblaje y exporte de sideróforos, así como el consumo de complejos ferri-sideróforos (Crosa & Walsh, 2002)□. La importancia de esta primera estrategia recae no sólo en que los sideróforos son un grupo químicamente diverso de metabolitos secundarios, sino que casi todas las especies bacterianas conocidas producen sideróforos, lo que hace que su secreción sea el principal mecanismo de adquisición de hierro en bacterias (Kramer et al.,

2020)□.

Sideróforos

Sideróforo (del griego *-sideros*, *-phore*: <<transportador de hierro>>) hace referencia a pequeñas (150-2000 Da) moléculas quelantes orgánicas producidas por las bacterias, poseen una afinidad muy alta por el hierro y se añaden con mayor facilidad en su forma Fe^{3+} , por lo que participan de manera activa en el transporte de este metal a través de las membranas celulares y su biosíntesis es esencialmente regulada por el nivel del hierro intracelular (Matzanke, 2011; Schalk & Guillon, 2013)□.

Los sideróforos tienen esqueleto peptídico con cadenas laterales de aminoácidos que les proporcionan mayor afinidad y a menudo se modifican para generar ligandos de coordinación con el hierro (S. Andrews et al., 2013)□. A pesar de las diferencias estructurales que existen entre los sideróforos, se reconocen algunas características en común como la capacidad de formar seis enlaces complejos coordinados con Fe^{3+} de gran estabilidad termodinámica. Además, los grupos ligandos incluyen átomos de oxígeno del hidroxamato, catecolato, ácido α -hidroxicarboxílico o del ácido α -cetocarboxílico (Matzanke, 2011)□.

De acuerdo con los grupos funcionales que poseen los sideróforos y que permiten su unión al hierro, éstos pueden ser clasificados en tres principales familias: hidroxamatos, catecolatos y carboxilatos (Figura A). A veces, estos grupos funcionales encargados de quelar al hierro están combinados en el mismo sideróforo y se refieren a ellos como de tipo mixto. Los sideróforos utilizan la carga negativa del oxígeno para coordinarse con el ión férrico; sin embargo, cada familia posee diferentes características que pueden afectar la afinidad con el hierro (Holden & Bachman, 2015)□ y su interacción con proteínas transportadoras que ayudan a la asimilación de estos complejos ferri-sideróforos (S. Andrews et al., 2013; Crosa & Walsh, 2002)

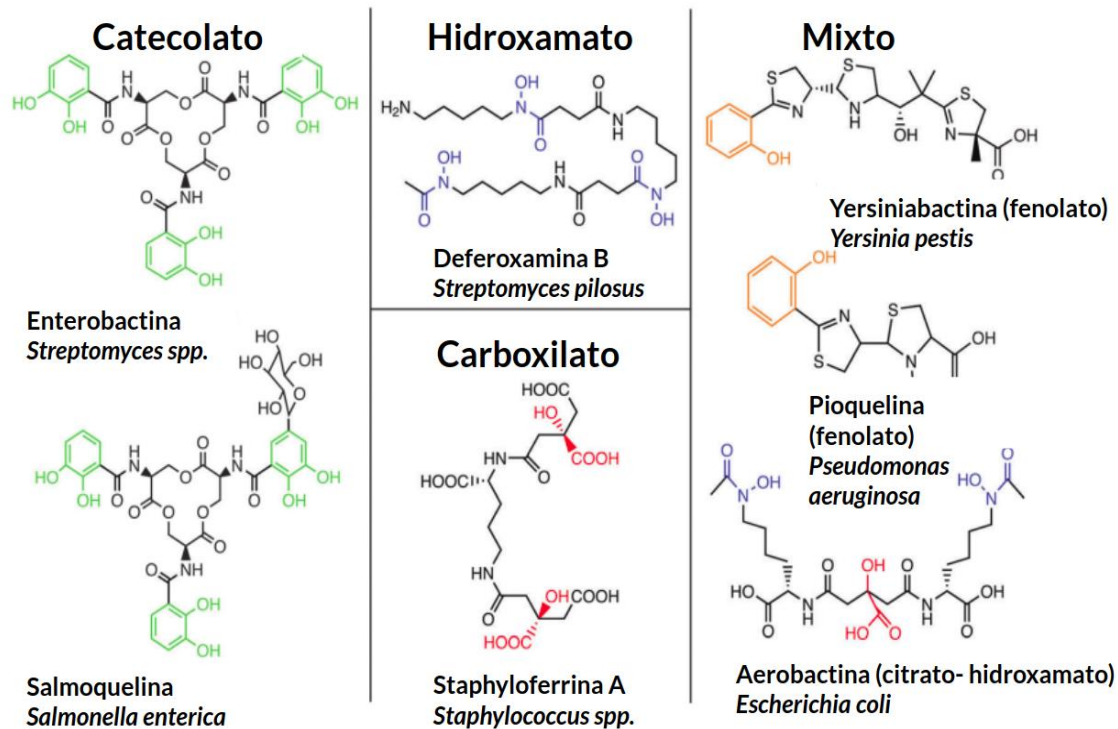


Figura A. Los sideróforos se pueden clasificar en familias estructurales según los grupos funcionales que poseen. Las principales familias estructurales son: catecolatos, hidroxamatos y carboxilatos. Los grupos funcionales responsables de la unión al hierro se resaltan en verde (catecolato), azul (hidroxamato) y rojo (carboxilato). Los sideróforos de tipo mixto se consideran así porque tienen elementos de dos o más familias de sideróforos. Los restos de unión para los sideróforos de tipo mixto están resaltados en naranja (fenolato) o en los colores anteriormente mencionados. Para cada familia se hace referencia a un ejemplo y el organismo del que procede. Modificado de Holden & Bachman, 2015.

La familia de los catecolatos es la que tiene mejor afinidad por el ión férrico. Su estructura se distingue por poseer grupos hidroxilos de manera adyacente a 5 anillos de catecol, que generalmente derivan de un grupo 2,3-dihidroxibenzoato (DHB). Por mencionar algunos ejemplos, se tiene a la enterobactina (también conocida como enteroquelina), la anguibactina, la vibriobactina, la salmoquelina, la vulnibactina y la acinetobactina; que fueron inicialmente descritos en *Escherichia coli*, *Vibrio anguillarum*, *Vibrio cholerae*, *Salmonella enterica*, *Vibrio vulnificus*, *Acinetobacter baumannii*, respectivamente. En particular, la enterobactina es capaz de disociar el hierro de proteínas de su hospedero, incluso de aquellas proteínas que poseen alta afinidad por este ión como la transferrina (S. Andrews et al., 2013). La salmoquelina también presenta características únicas, ya que su esqueleto de DHB también está conformado por glucosa (Matzanke, 2011) □.

La estructura de los hidroxamatos tiene menor afinidad por el hierro que la de los catecolatos. Son sintetizados principalmente por bacterias Gram positivas. En esta categoría se encuentra el ferricromo, la alcaligina y la deferoxamina (S. Andrews et al., 2013; Crosa & Walsh, 2002). Los sideróforos de tipo carboxilato, en general, quelan con menor facilidad al ión férrico en condiciones de pH fisiológico; sin embargo, son más eficientes cuando el pH del medio es ácido. Algunos sideróforos que encontramos en esta familia son la estafiloferrina, parabactina, petrobactina, pioverdina, yersiniabactina y el citrato. Otro ejemplo más es la aerobactina, que posee los grupos hidroxamato y carboxilato (S. Andrews et al., 2013)□.

Por otra parte, también encontramos a sideróforos como la pioquelina, a la cual frecuentemente se hace referencia como de tipo fenolato por poseer grupos fenoles en su estructura; sin embargo, al tener una combinación de las principales familias de sideróforos, generalmente se les denomina como de tipo mixto (S. Andrews et al., 2013; Crosa & Walsh, 2002)□. Por mencionar otros ejemplos de sideróforos mixtos que combinan ligandos carboxilatos con hidroxamatos, se tiene a la aerobactina y acinetoferrina (Matzanke, 2011)□.

La biosíntesis de sideróforos ocurre mediante dos rutas principales, la primera es llevada a cabo por miembros de la familia multienzimática de sintetetas de péptidos no ribosomales (NRPS, por sus siglas en inglés) y que también son responsables de la síntesis de los principales antibióticos de origen peptídico bacteriano. Las NRPS son enzimas multimodulares que producen péptidos a partir de una secuencia particular, pero no requieren de RNA como molde ni adaptador, lo que hacen es ordenar los monómeros de aminoácidos de acuerdo al orden de los dominios de las NRPS, siguiendo un proceso enzimático de iniciación, elongación, modificación y terminación del sideróforo (Crosa & Walsh, 2002)□.

Un módulo mínimo de NRPS contiene tres dominios. El primero es el dominio de adenilación (A), que se encarga de reconocer específicamente el sustrato del módulo y catalizar la adenilación de su grupo carboxilo. El dominio de tiolación (T) utiliza el grupo tiol terminal para capturar el grupo carboxilo activado del adenilato. Por último, el dominio de condensación (C) cataliza la acilación del grupo tioéster resultante con el grupo acil activado y que está unido al dominio. Los dominios encargados de la adenilación, tiolación y condensación de las enzimas NRPS funcionan de manera coordinada para unir aminoácidos por medio de grupos tiol. Los dominios ATC son altamente selectivos por el tipo de sustratos que emplean para la síntesis de péptidos, en particular, el dominio de la

adenilación. Las enzimas NRPS también poseen la capacidad de unir ácidos grasos, α -cetoácidos y α -hidroxiácidos de manera similar a la síntesis de otros metabolitos (Kadi & Challis, 2009; Lamb, 2015)□.

Algunos de los sideróforos sintetizados bajo esta vía están conformados básicamente por aminoácidos proteicos y no proteicos, unidos por enlaces peptídicos (Lamb, 2015)□. Como ejemplos de este tipo de sideróforos se tiene a la yersiniabactina, la vibriobactina, la micobactina y la enterobactina (Carroll & Moore, 2018)□.

Por otra parte, algunos sideróforos bacterianos no son polipéptidos y se producen mediante sintetetas independientes de péptidos no ribosomales (*NIS*, por sus siglas en inglés), las cuales sólo han sido caracterizadas en bacterias y en algunas especies de hongos. Las NIS sintetetas llevan a cabo una sola reacción enzimática que condensa ácido cítrico con un grupo amina o un grupo alcohol por medio de una reacción de adenilación. Las enzimas NIS poseen un dominio N-terminal conservado que se encarga de quelar el hierro y un dominio C-terminal que está relacionado con el transporte y metabolismo del hierro (Carroll & Moore, 2018)□. Algunos ejemplos de sideróforos producidos por medio de NIS son la aerobactina, rizobactina, alcaligina, deferoxamina, vibrioferrina, estafilobactina, antraquelina y la acromobactina (Challis, 2005; Himpsl et al., 2010)□.

Transporte de sideróforos en bacterias

El consumo de hierro en bacterias es un proceso dependiente de receptores y de energía que media la translocación del complejo al interior de la célula mediante proteínas receptoras localizadas en la membrana celular (Matzanke, 2011)□. El transporte o difusión de estos complejos ferri-sideróforos a través de las membranas de las bacterias representa cierto reto para las células, teniendo que impedir la entrada de sustancias nocivas como detergentes y antibióticos. Por lo anterior, es muy probable que las proteínas receptoras surgieron de manera paralela a los sideróforos durante el curso de la evolución, es así que exhiben gran especificidad para los sideróforos individuales o un pequeño grupo de sideróforos parecidos (Hannauer et al., 2012; Matzanke, 2011)□.

Generalmente, las bacterias Gram negativas absorben los complejos ferri-sideróforos a través de receptores específicos de la pared celular en un proceso que es impulsado por el potencial de la membrana citoplasmática y mediado por el sistema TonB-ExbB-ExbD (familia de proteínas TBDT) de transducción de energía. Sin embargo, en condiciones anaeróbicas o reductoras, el Fe^{2+} es estable y soluble y puede difundir libremente a través

de las porinas de la pared celular y con ayuda del sistema de proteínas FeoABC (Lim, Hassan, Tetu, Loper, & Paulsen, 2012)□. La familia de proteínas TBDT (transportador dependiente de TonB) tienen un dominio en forma de barril-β transmembranal estructurado por 22 láminas β anfífilicas antiparalelas (β1-β22) y un dominio globular formado por aproximadamente 150 residuos altamente conservados del extremo amino-terminal, que se pliega dentro del barril obturando la cara interior del receptor. El sitio de unión siempre está ubicado en la cara extracelular del receptor, que consta de grandes *loops* expuestos sobre la superficie celular. La unión entre el ferri-sideróforo y el receptor de membrana externa es de gran afinidad ($K_d=0.1\mu\text{M}$) por lo que cada TBDT es capaz de reconocer y transportar un sideróforo específico, o en algunos casos, unos pocos sideróforos relacionados estructuralmente, pero nunca sideróforos con diferentes estructuras químicas (Hannauer et al., 2012)□.

A partir de estructuras obtenidas de rayos X se han logrado obtener algunos detalles conformacionales del receptor en cuanto a su interacción con el complejo ferri-sideróforo, y es que aparentemente este contacto no ocasiona grandes alteraciones en la estructura expuesta al medio, sólo son unos pequeños cambios los que se han podido notar en la porción que corresponde al periplasma. Por otro lado, los sideróforos sufren cambios conformacionales esencialmente en su extremo N-terminal (Matzanke, 2011)□.

Las proteínas ExbB y ExbD tienen dos estados funcionales. El primero es el no energizado, donde el complejo ferri-sideróforo se une al receptor, pero no es transportado al interior de la célula. El segundo es cuando los receptores utilizan el gradiente de carga para producir una forma activada de TonB, que media un cambio conformacional del receptor de la membrana externa. Esto conduce a la translocación del complejo ferri-sideróforo al periplasma y la desenergización de TonB. Se asume que el cambio del extremo N-terminal del sideróforo, a consecuencia de su interacción con el receptor, facilita esta transferencia de energía con la proteína TonB. La forma desenergizada de TonB se recicla después (Matzanke, 2011; Wooldridge, Morrissey, & Williams, 1992)□.

Una vez en el periplasma, las proteínas de unión periplásmica transportan los complejos desde los receptores de la pared celular hasta transportadores de tipo ABC embebidos en la membrana citoplasmática y que permiten el paso del complejo ferri-sideróforo al citoplasma en un proceso dependiente de protones (Mosbahi, Wojnowska, Albalat, & Walker, 2018)□. Cada proteína de unión (*PBP*, por sus siglas en inglés) acepta un complejo ferri-sideróforo a la vez (Köster, 2001)□. Por ejemplo, en *E. coli*, la proteína de unión periplásmica FecB es responsable de la transferencia de ferricitrato, la proteína FepB transfiere sideróforos de

catecolato y FhuD es responsable del transporte de sideróforos de hidroxamato y puede interactuar con una amplia gama de complejos ferri-sideróforos como ferricromo y deferoxamina B (S. Andrews et al., 2013)□.

Las proteínas de unión periplásmica de diferentes sistemas presentan una baja similitud de secuencia entre ellas, sin embargo, comparten la misma estructura de dos dominios globulares interconectados por una cadena polipeptídica que facilita el movimiento de los dos lóbulos mediante un mecanismo bisagra. No está claro si la proteína periplásmica recoge el complejo ferri-sideróforo directamente del receptor, o si simplemente lo extrae del periplasma (Clarke, Tari, & Vogel, 2005)□. La proteína periplásmica unida al complejo ferri-sideróforo, lo entrega específicamente a una permeasa de membrana interna y es translocado a través de la membrana citoplasmática gracias al aporte energético proporcionado por el componente ATPasa del sistema. Es así como se puede decir que la especificidad del transporte del complejo ferri-sideróforo reside en el reconocimiento de los receptores TBDT, que son de mayor afinidad y especificidad que las proteínas periplásmicas (Köster, 2001)□.

Las bacterias Gram positivas, que carecen de una doble membrana celular y no requieren receptores ni sistemas TonB-ExbB-ExbD, también utilizan permeasas de tipo ABC dependientes de proteínas de unión que son análogas a las empleadas por las Gram negativas y que les permiten internalizar el complejo ferri-sideróforo (S. C. Andrews et al., 2003)□. Los sistemas transportadores ABC forman complejos proteínicos de dos subunidades idénticas o permeasas integrales homólogas de membrana que forman un canal a través del cual pasa el complejo ferri-sideróforo y dos subunidades que hidrolizan ATP, que se encargan de aportar la energía requerida para el transporte del complejo a través de la membrana citoplasmática. Como se mencionó anteriormente, muchas bacterias poseen un sistema alternativo de transporte de hierro a cargo del operón *feoABC* que consta de tres genes, regulados por la proteína Fur, y se encarga del transporte del ión ferroso (Fe^{2+}). La función de transporte del hierro parece ser de particular importancia en condiciones anaeróbicas o microaerofílicas (menos de 5 ppm de oxígeno) (Cartron, Maddocks, Gillingham, Craven, & Andrews, 2006; Hsueh et al., 2013)□, cuando el hierro ferroso (Fe^{2+}) permanece estable y predomina sobre el hierro férrico (Fe^{3+}). Otro sistema de transporte de hierro alterno son los sistemas ABC dependientes de proteínas de unión a metales que tienen especificidad por el hierro, pero que no requieren necesariamente receptores en la membrana externa o sideróforos. Tales sistemas incluyen los transportadores SfuABC, SitABCD, YfeABCD, FbpABC y FutABC (S. C. Andrews et al., 2003)□.

El destino de los sideróforos después de la liberación del hierro mediante una reacción de reducción, por medio de reductasas y esterasas, en el citosol o periplasma no se conoce bien. En algunos casos, se pueden volver a utilizar mediante mecanismos de reciclaje específicos (Kramer et al., 2020)□.

Los mecanismos de síntesis, secreción y adquisición de sideróforos descritos anteriormente están bajo un estricto control que mantiene la homeostasis del hierro en la célula. La principal proteína reguladora es Fur (*Ferric uptake regulator*, por sus siglas en inglés), capaz de controlar la expresión dependiente de hierro de más de 90 genes. Fur es un regulador global de la absorción de hierro que puede funcionar como activador transcripcional o represor, no sólo a través de mecanismos directos (interacción con el promotor) sino que también puede hacerlo de manera indirecta (pequeños RNAs reguladores) (Braun & Braun, 2002)□. Otro nivel de regulación ocurre cuando el complejo ferri-sideróforo ingresa a la célula y los factores sigma de función extracitoplasmática (*ECF*, por sus siglas en inglés), o los sistemas de dos componentes sirven como puntos de control, lo que desencadena una mayor síntesis de sideróforos (Figura B) (Kramer et al., 2020).

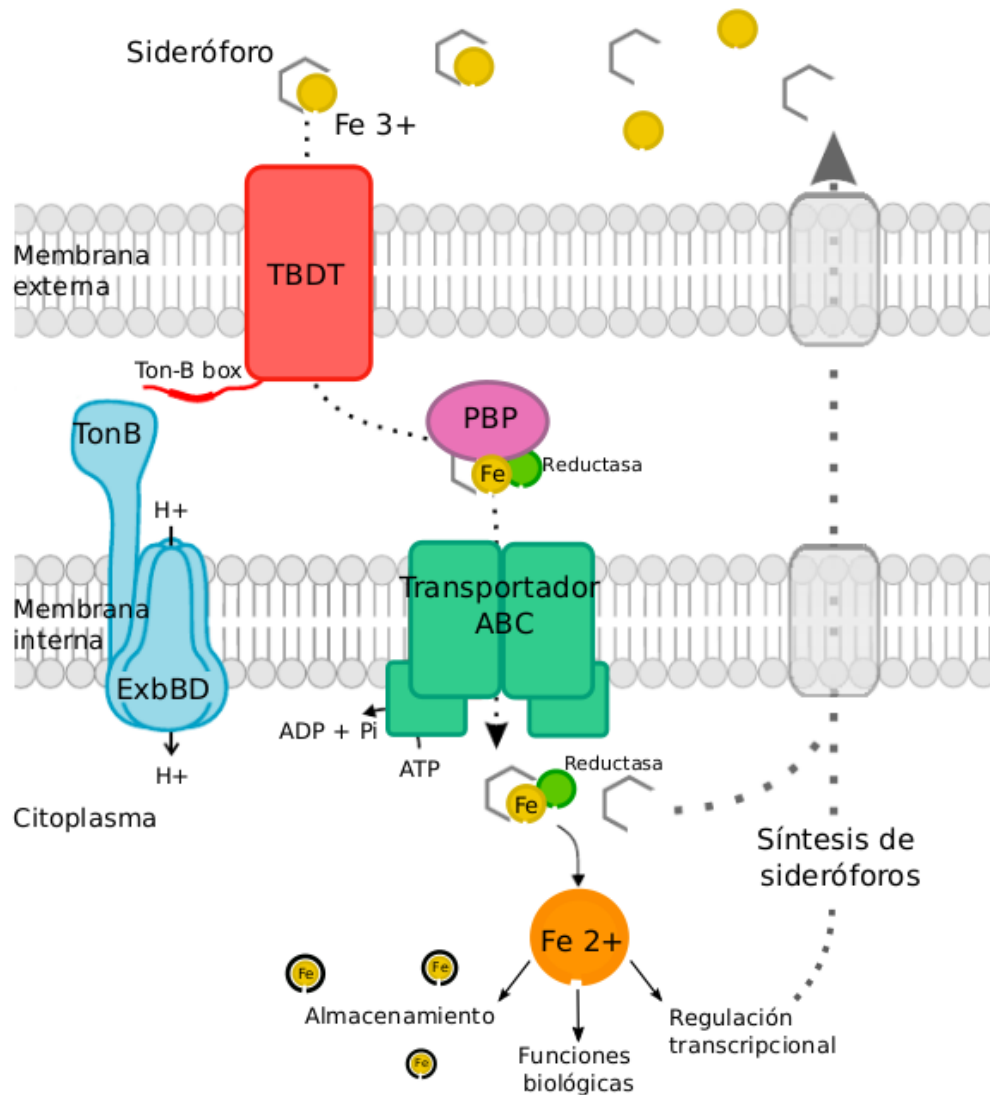


Figura B. Representación esquemática de la captación general de hierro mediada por sideróforos. Modificado de (Lemos & Balado, 2020)

Consecuencias fisiológicas de la producción de sideróforos

Los sideróforos no sólo son importantes por su capacidad de transportar hierro y otros metales, cumplir funciones de señalización para regular su propia producción, inducir la expresión de factores de virulencia, proteger a la célula contra ROS (especies reactivas de oxígeno), o tener actividad antibiótica; sino que son mediadores de las interacciones entre los miembros de las comunidades bacterianas y sus huéspedes eucariotas (Kramer et al., 2020)□.

Los sideróforos son elementos costosos de secretar, porque se requiere de un proceso

dependiente de energía y mecanismos complejos específicos en la membrana celular. Una vez fuera de la célula, los sideróforos difunden y se pierden para la célula productora y sus aminoácidos no pueden reciclarse fácilmente. Paralelamente, otro aspecto a considerarse es que las células bacterianas casi nunca están solas, sino que crecen en grupos, frecuentemente son biopelículas unidas a superficies en el suelo, huéspedes y partículas en ambientes acuáticos. Esto repercute en el uso de los sideróforos y los hace propensos a la explotación social, por lo que son considerados bienes públicos (García-Garcera & Rocha, 2020a; Kramer et al., 2020)□.

De esta forma, las células no dependen exclusivamente de sus propios sideróforos para sobrellevar condiciones limitantes de hierro, sino que pueden beneficiarse de sideróforos producidos por otras células. El costo de producción y beneficio que estas proteínas pueden brindar son determinantes muy importantes de la selección para el mantenimiento y expresión de los genes que los codifican. Por lo tanto, el intercambio de sideróforos es un mecanismo importante que minimiza su pérdida a través de la difusión, es determinante en las interacciones sociales de las bacterias a nivel molecular y también está influenciado por la diversidad de la comunidad, así como la estructura del hábitat (García-Garcera & Rocha, 2020a; Jin et al., 2018a; Kramer et al., 2020)□.

Genómica comparativa

La genómica comparativa nos permite analizar secuencias de genes de genomas completos y compararlos entre diferentes organismos, para dilucidar sus características únicas y compartidas, además de proporcionar información sobre su abundancia y distribución filogenética e incluso permite su interpretación funcional mediante la anotación de genomas secuenciados (Hardison, 2003; Sivashankari & Shanmughavel, 2007)□. Los análisis comparativos proporcionan información poderosa dado que las características de los organismos se ven reflejadas en las secuencias de DNA. Dicho de otra manera, las características compartidas a nivel de secuencia entre dos organismos se consideran propiedades que se originaron en el genoma de su ancestro común, mientras que sus diferencias reflejan las distintas historias de cada genoma trazado desde el ancestro común. Teniendo en cuenta que la secuencia de DNA refleja patrones de conservación a lo largo del tiempo evolutivo, se pueden abordar diferentes preguntas comparando genomas con diferentes distancias filogenéticas (Kawashima, 2018)□. Al realizar análisis genómicos comparativos, se deben considerar los siguientes aspectos biológicos: estructura del genoma, tipos de cambio estructural y distancia evolutiva. La estructura del genoma hace referencia principalmente a la elección del tipo de datos y el método propuesto para

analizarlos, el cambio estructural y distancia evolutiva tiene que ver principalmente con la homología de los genes y su análisis comparativo, respecto a su historia evolutiva. Una de las ventajas que proporciona la búsqueda de genes homólogos es la predicción de la función del gen. Esto tiene como fundamento el que los genes homólogos generalmente comparten funciones similares, también es posible que sean similares a nivel de secuencia. La importancia de este tipo de búsquedas reside en que el número de genes para los cuales se ha dilucidado su función de forma experimental es reducido a comparación del número de genes secuenciados (Gabaldón & Koonin, 2013; Toh, Hayashida, & Miyata, 1983)□.

Recursos bioinformáticos y bases de datos

La creciente disponibilidad de secuencias genómicas de múltiples organismos ha hecho indispensables las bases de datos informáticas como un recurso necesario para recopilar, almacenar y organizar la gran cantidad de datos biológicos. A la fecha son al menos 3,000 recursos genómicos, herramientas y bases de datos distintas disponibles públicamente en Internet. Tan sólo la revisión anual de la revista *Nucleic Acids Research*, 2019 reportó 1,613 bases de datos disponibles y activas actualmente (Rigden & Fernández, 2019)□. Una de las principales bases de datos que contiene múltiples recursos de acceso libre sobre proteínas es UniProt (Universal Protein Resource) y es el resultado de la colaboración entre diferentes institutos de investigación, como lo es el Instituto europeo de bioinformática (EBI) y el Instituto suizo de bioinformática (SIB). UniProt es una base de datos de mínima redundancia, es anotada manualmente por expertos y considera la evidencia experimental como parte de la curación de los datos contenidos. La anotación manual consiste en una revisión crítica de los datos predichos computacionalmente, integración e interpretación crítica de información relevante biológicamente, se atribuye evidencia científica. UniProt es revisada y actualizada constantemente por los curadores (Apweiler, 2008)□. Otro recurso de gran importancia para el análisis de genomas procariontes es SEED *subsystems*, una base de datos que integra anotaciones genómicas y herramientas de análisis, por lo que también es utilizado como una plataforma para desarrollar anotaciones *de novo*. Está organizado de manera jerárquica, en donde los subsistemas son conjuntos de roles funcionales que forman parte de un proceso biológico específico. Los roles funcionales comprenden diferentes familias de proteínas (FIGfam) relacionadas funcionalmente, por lo que todos los miembros de una FIGfam comparten la misma función biológica (Overbeek et al., 2005)□. La base de datos SEED *subsystems* proporciona la capacidad de asociar reacciones con roles funcionales de manera independiente del organismo (DeJongh et al., 2007)□. Además concentra información de

aproximadamente 5,000 genomas procariontes diferentes, 30,000 anotaciones funcionales, 11,000 modelos metabólicos, 178,000 familias de proteínas, cerca de 10,250 roles funcionales y 1,060 subsistemas (Aziz et al., 2012; Rodriguez-Brito, Rohwer, & Edwards, 2006)□.

Los curadores de esta base de datos no tienen restricciones en cómo agrupan las proteínas relacionadas funcionalmente dentro de un subsistema. Los servidores de SEED están disponibles para uso público, igualmente sus datos son actualizados regularmente y además de facilitar la anotación de genomas, posibilitan realizar reconstrucciones metabólicas y acceder a modelos metabólicos ya existentes (Overbeek et al., 2014)□.

Búsqueda bioinformática de proteínas homólogas

Los alineamientos múltiples de secuencias (*MSA*, por sus siglas en inglés) son un recurso bioinformático que se emplea para entender las relaciones evolutivas entre secuencias de proteínas, comparando e identificando regiones conservadas u homólogas de una familia de proteínas y su información puede ser incorporada a modelos estadísticos como son las las matrices de puntuación de posiciones específicas (*PSSM*, por sus siglas en inglés), los perfiles y modelos ocultos de Markov (*pHMM*, por sus siglas en inglés) (Finn et al., 2016; Pundir, Martin, & O'Donovan, 2017)□.

Lo que permite visualizar un MSA son los residuos de aminoácidos que pueden estar altamente conservados en algunas posiciones, qué posiciones pueden tolerar ciertas sustituciones o deleciones y si esto no compromete sus propiedades fisicoquímicas, así como las posiciones que pueden ser evolutivamente neutrales o variables (Eddy, 1998)□. Estas características son añadidas a las PSSMs y entonces es posible obtener una descripción cuantitativa de la frecuencia de los residuos de aminoácidos en un MSA. Es importante añadir que en una PSSM se asume la independencia entre las posiciones de los residuos de aminoácidos en el MSA. Sin embargo, cuando una PSSM considera las penalizaciones por abrir y extender huecos a consecuencia del alineamiento, se le puede referir como un perfil y puede ser considerado como un representante del grupo de secuencias de proteínas en cuestión. Los perfiles son empleados para realizar búsquedas de secuencias homólogas en bases de datos, o bien, para detectar relaciones distantes entre proteínas (Chatzou et al., 2016; Madera & Gough, 2002; Stojmirović, Gertz, Altschul, & Yu, 2008)□.

Para el análisis de estos perfiles fue propuesto el uso complementario de los modelos ocultos de Markov (*HMM*, por sus siglas en inglés) porque igualmente son capaces de producir sistemas de puntuación de posición específica a partir de un MSA, por lo que son altamente sensibles para representar inserciones y eliminaciones de aminoácidos, además de predecir las frecuencias de caracteres o residuos no observados (Choo, Tong, & Zhang, 2004; Eddy, 1998, 2018) □. La teoría de los HMM fue desarrollada a finales de la década de 1960 por Baum y colaboradores (“Baum , Eagon : An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology,” n.d.; Baum & Petrie, 1966) □. Inicialmente fueron empleados en el campo del reconocimiento automático de patrones en el habla (Rabiner, 1989) □. Desde entonces su aplicación ha sido relevante en múltiples disciplinas como la bioinformática.

Un HMM, también conocido como cadena oculta de Markov, describe una secuencia de eventos y cada evento determina la probabilidad del siguiente. El proceso que describe el HMM tiene direccionalidad y cada estado sucede con cierta probabilidad de emisión y transición. Al igual que los perfiles, los HMM son utilizados para modelar las características estadísticas de toda una familia de proteínas y pueden ser usados para buscar en bases de datos secuencias relacionadas con la familia de proteínas modelada. Es así que, los pHMMs incorporan información probabilística basada en las características estructurales de las secuencias de proteínas, permitiendo modelar familias, predecir o buscar dominios, así como proteínas homólogas (Eddy, 1998; Finn et al., 2016) □.

Un HMM se compone de varios estados interconectados y cada uno emite un símbolo de salida que corresponde a un aminoácido. Como la evolución del DNA, RNA y las secuencias de proteínas está influenciada por reordenamientos genómicos como las sustituciones, inserciones y deleciones (Eddy, 1995; Ezawa, 2016) □ designa los estados del HMM como coincidencias, inserciones y deleciones. Cada estado tiene dos tipos de parámetros, las probabilidades de emisión y las probabilidades de transición.

Las probabilidades de emisión hacen referencia a la probabilidad que existe de encontrar un aminoácido en un estado específico. Por otra parte, las probabilidades de transición de estado son las probabilidades de pasar del estado actual a un nuevo estado. Una secuencia considera un estado inicial que va pasando de un estado a otro hasta que se alcanza un estado final, emitiendo símbolos de cada estado por el que se pasa. Esta secuencia de estados está oculta y sólo la secuencia de símbolos emitidos es observable, a esto se debe el término de modelo oculto de Markov. Las secuencias generadas representan el consenso de las características que incorpora el modelo a través de la variación en las probabilidades

de transición y emisión (Figura C) (Eddy, 1995) □.

La eficiencia de los pHMMs dependerá esencialmente del alineamiento múltiple de secuencias con el que se entrene. Otra ventaja que presentan los pHMMs ante la detección de proteínas homólogas es que pueden construirse a partir de matrices de sustitución basadas en modelos evolutivos (e.g. BLOSUM62) (Eddy, 1998; Anoop Kumar & Cowen, 2009; Nahas, Kassim, & Shikoun, 2012) □□.

La base probabilística de los HMMs ha facilitado su construcción de manera automatizable, además ha permitido establecer e interpretar diferentes parámetros en un perfil, como el *E-value* y el *bit score*. El *E-value* es la expectativa del número de errores de un pHMM por búsqueda en una base de datos. Un *E-value* sólo es verdadero en el contexto de cada búsqueda, por ejemplo, una consulta en una base de datos grande tendrá más errores que una consulta en una base de datos pequeña buscada con el mismo modelo (Finn, Clements, & Eddy, 2011) □. Ya que este valor depende del tamaño de la base de datos, alternativamente se puede usar de referencia el *bit score*, que es un valor logarítmico asociado al sistema de puntuación empleado para construir el perfil y hace referencia a si una secuencia es homóloga al pHMM. Es una puntuación independiente del tamaño de la base de datos (Eddy, 1998; Finn et al., 2011) □.

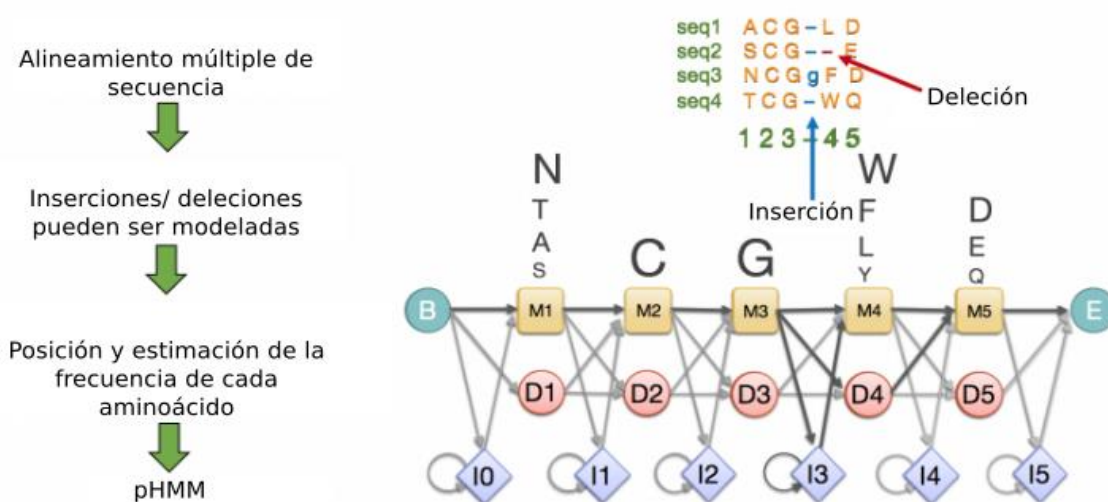


Figura C. Representación de un perfil de Modelo Oculto de Markov, que encapsula los cambios evolutivos que se han producido en un conjunto de secuencias de proteínas relacionadas. El entrenamiento del HMM se basa en un AMS. Las probabilidades de emisión de los aminoácidos en la secuencia se representan con letras apiladas de diferentes alturas y está relacionada con la frecuencia que tiene un aminoácido en un AMS.

La secuencia considera un estado inicial (B) que va pasando de un estado a otro, hasta llegar a uno final (E). Los recuadros en color amarillo representan los estados de *match* o coincidencia (M). Los diamantes representan el estado de inserción (I) que se utilizan para modelar regiones altamente variables en el alineamiento. Los círculos representan el estado de eliminación (D) y no coinciden con ningún residuo. El modelo probabilístico final transmite la estimación de las frecuencias observadas de los aminoácidos en cada posición, así como las transiciones entre los aminoácidos observadas de cada posición en un alineamiento múltiple de secuencia. Modificado de EMBL-EBI, 2020.

Perfiles filogenéticos y genómica comparativa

La homología o similitud entre secuencias, permite inferir o asignar una función a una proteína en donde se desconoce, esto se basa también en la evidencia experimental de su función molecular. Si dos proteínas tienen similitudes significativas entre sus secuencias, es probable que descendan de un ancestro común y compartan una función común. Aunque diferentes organismos pueden haber desarrollado soluciones distintas para sus requisitos fisiológicos a través de la evolución, se espera que los genes involucrados tengan un grado similar de conservación □(Pellegrini, 2012)□. A medida que aumenta la evidencia experimental de las funciones de las proteínas, este enfoque permite asignar la anotación funcional de una proteína caracterizada a sus homólogos no caracterizados, y así anotar un número cada vez mayor de genes secuenciados. Es así que el análisis comparativo de los patrones de distribución de genes entre los genomas ha sido importante para inferir vínculos funcionales, aunque puede haber problemas derivado de esto como la propagación de errores de anotación en las bases de datos (Dey & Meyer, 2015; Martín, Herrero, Mateos, & Dopazo, 2003; Pellegrini, 2012)□.

Con el enfoque bioinformático antes descrito, el perfil filogenético de una proteína es una representación bidimensional que contiene la información de presencia o ausencia de homólogos de esa proteína en diferentes genomas de interés. El agrupamiento de los patrones de presencia o ausencia constituye una herramienta de visualización que brinda información inmediata sobre la similitud de los genomas estudiados. Además permite realizar análisis cuantitativos con base en los diferentes grados relativos de conservación y el patrón de distribución de proteínas homólogas a través de los genomas. De esta forma, el perfil de conservación del genoma o perfil filogenético, puede proporcionar una representación de la conservación relativa de proteínas y familias de proteínas en las diferentes categorías taxonómicas de los organismos. El siguiente aspecto a analizar es la relación de esta información obtenida con los rasgos fenotípicos, porque los fenotipos de los

distintos genomas son consecuencia de las diferentes distribuciones de genes (Jothi, Przytycka, & Aravind, 2007; Martín et al., 2003) □.

La elección del conjunto de referencia de genomas en los que se perfila una proteína determinada afecta el potencial predictivo de los perfiles filogenéticos. Aunque se considera como una observación intuitiva es importante resaltar que, más allá de la cantidad de genomas, la selección cuidadosa de genomas informativos y que representen la diversidad filogenética de la vida es esencial para un buen rendimiento (Jothi et al., 2007) □.

Las investigaciones realizadas por los grupos de Barker y Zhou demostraron que la incorporación de información filogenética en la comparación de perfiles filogenéticos, puede ayudar a distinguir genes que se han ganado o perdido en diferentes ocasiones. Además, esta perspectiva mejora la precisión de la predicción y no se ve afectado por el número de genomas relacionados o redundantes en el conjunto de referencia. Dado que es una aproximación que se hace a partir de un árbol filogenético, es necesario tomar en cuenta la calidad de este (Figura D) (Barker & Pagel, 2005; Jothi et al., 2007; Yin et al., 2018) □.

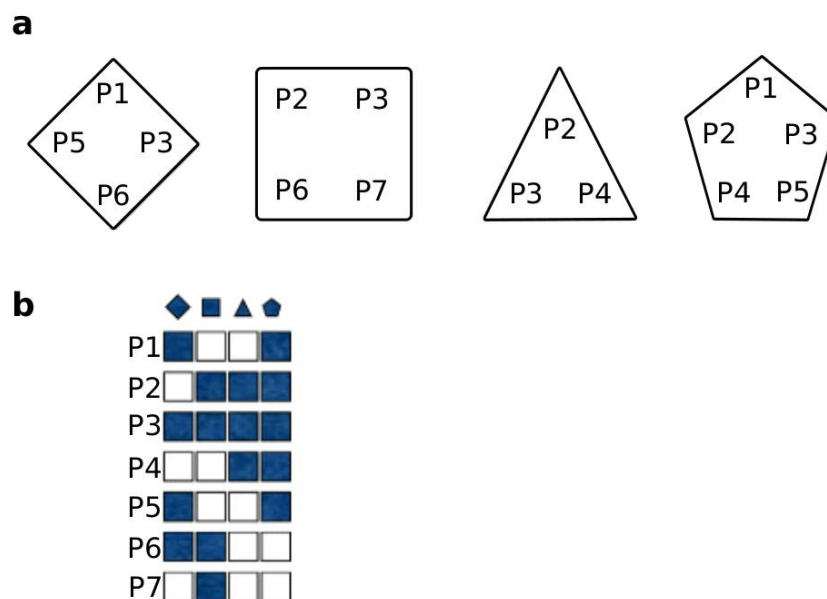


Figura D. Representación esquemática de un perfil filogenético. En el panel (a) se observan cuatro muestras de genomas que incluyen diferentes familias de proteínas (P1-P7). El panel (b) muestra los perfiles filogenéticos que indican la presencia o ausencia de estas familias de proteínas en los genomas. Modificado de Pellegrini, 2012.

Reconstruir el árbol filogenético de la vida ha sido uno de los problemas más fundamentales en la biología. Las complicaciones derivan de los diferentes enfoques bajo los que se ha

abordado su reconstrucción, como el taxonómico o los crecientes datos moleculares y genómicos. Ante la necesidad de una filogenia reproducible y actualizada con una selección de genomas informativos y completamente secuenciados, Ciccarelli (et al., 2006) propuso un procedimiento automático para reconstruir el árbol de la vida. Dicho trabajo se basó en la concatenación de 31 genes ortólogos involucrados en el proceso de la traducción, presentes en 191 especies representativas de los tres dominios de la vida con genomas completamente anotados (Figura E).

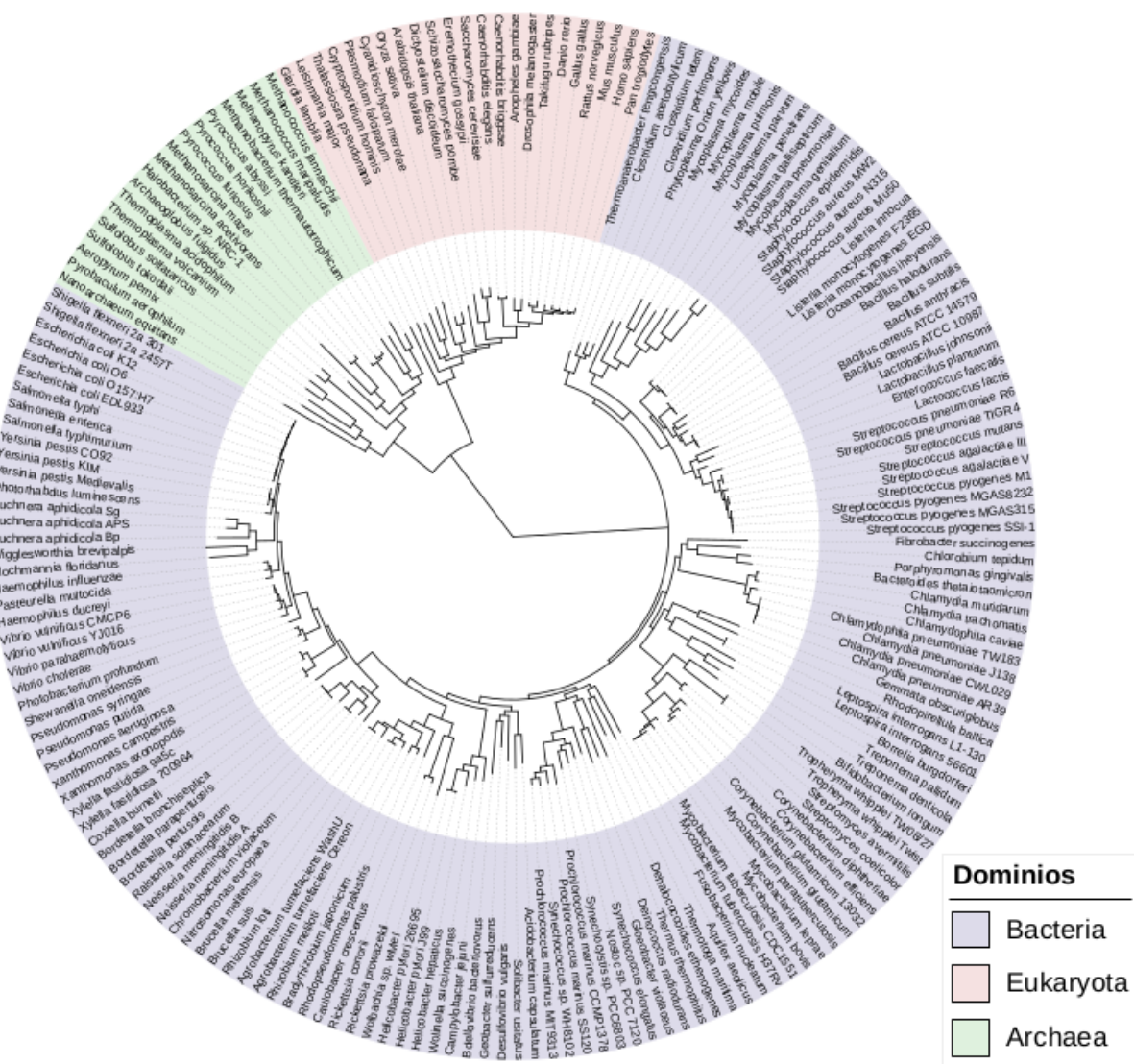


Figura E. Árbol filogenético postulado por Ciccarelli (et al., 2006). En morado se señala el dominio Bacteria, en rosa a Eukaryota y en verde a Archaea. Disponible en <http://itol.embl.de/itol.cgi> (Letunic & Bork, 2011) □

Justificación

El hierro biodisponible es un micronutriente escaso, pero esencial para el metabolismo de las bacterias. Se han investigado las estrategias de adquisición de hierro de diferentes bacterias seleccionadas, pero la reciente explosión en la secuenciación de genomas procariotas ofrece la oportunidad de tener un panorama más completo de las vías de absorción microbiana de hierro. Es por ello, que buscamos en estos conjuntos de datos los homólogos de genes de biosíntesis y transporte de sideróforos, analizándolos desde la perspectiva de la genómica comparativa.

Hipótesis

Si las proteínas que están asociadas a la adquisición de hierro por medio de sideróforos se encuentran distribuidas de manera restringida entre las bacterias, entonces se podrán identificar patrones según su abundancia y distribución filogenética en los principales phyla bacterianos.

Objetivo general

Predecir y analizar las proteínas responsables asociadas a la adquisición de hierro mediante la síntesis y transporte de sideróforos en un grupo representativo de phyla bacterianos.

Objetivos particulares

Elaborar una base de datos con proteínas involucradas en la biosíntesis y transporte de sideróforos.

Elaborar modelos predictivos que permitan el reconocimiento de proteínas homólogas en proteomas bacterianos.

Describir la abundancia y distribución filogenética de las proteínas asociadas a la adquisición de hierro en cada uno de los proteomas bacterianos analizados.

Metodología

Base de datos de proteínas que participan en la adquisición del hierro

Con el objetivo de recuperar las proteínas que están involucradas con la adquisición de hierro del medio en bacterias, se realizó una búsqueda en la base de datos SEED *subsystems*. Se buscó primero por anotación, incluyendo todas aquellas que estuvieran anotadas como parte del subsistema de adquisición y metabolismo del hierro, así como las que incluyeran en su anotación alguna referencia al metabolismo del hierro. De este primer filtro, se obtuvieron un total de 93,516 secuencias de proteínas redundantes. El grupo de datos representa 59,168 proteínas con anotaciones distintas.

Para facilitar el análisis, este conjunto de datos se organizó en familias de proteínas con ayuda del programa CD-HIT (Fu, Niu, Zhu, Wu, & Li, 2012)□. Brevemente, las secuencias de proteínas se clasificaron en orden descendente respecto a su longitud; de manera que, la secuencia de mayor longitud se estableció como la proteína de referencia o representativa de la primera familia. El resto de las secuencias se comparó con la proteína representativa, si la similitud de cualquiera cumplió con un umbral de 40% de identidad y 70% de cobertura (Pevsner, 2015) fue agrupada como parte de la misma familia de proteínas. En caso contrario, esa secuencia de proteína definió una nueva familia y posteriormente se compararon contra ésta las secuencias restantes.

El agrupamiento de proteínas fue inspeccionado manualmente, conservando únicamente aquellas que son responsables de la síntesis de sideróforos, así como las que reconocen y permiten el transporte de los complejos ferri-sideróforos a través de las membranas celulares. Por lo tanto, el conjunto original de proteínas se redujo a un total de 1007 familias de sideróforos. Cada familia de proteínas fue clasificada de acuerdo con el tipo y familia de sideróforo según su estructura química, ruta de biosíntesis, función biológica, clasificación taxonómica asignada a la secuencia de proteína, organismo en el que fue inicialmente descrito, anotación funcional de SEED y número de representantes por familia.

Identificación de proteínas homólogas de síntesis y transporte de sideróforos

Para identificar proteínas homólogas encargadas de ensamblar y transportar

sideróforos, se utilizaron como secuencias semillas las 1,007 proteínas representativas obtenidas de SEED y se realizó una búsqueda con el programa blastp (Camacho et al., 2009) contra la base de datos de RefSeq de proteínas, seleccionando el algoritmo de búsqueda PSI-BLAST (Position-Specific Iterated BLAST). Para cada una de las búsquedas, se seleccionaron los mejores 100 *hits* de tres ciclos de iteraciones, que cumplieran con un valor esperado máximo $1e^{-6}$ y que representaran a más de una especie. Las secuencias homólogas fueron descargadas en archivos FASTA individuales.

Clasificación taxonómica de las proteínas de síntesis y transporte de sideróforos

Se realizó la asignación taxonómica para cada una de las 1007 secuencias representativas de proteínas, para ello se empleó el programa Kaiju (Menzel, Ng, & Krogh, 2016). Este programa proporciona asignación taxonómica a las secuencias especificadas con base en la información contenida en la base de datos no redundante de taxonomía RefSeq NCBI. La ventaja de emplear este método es que puede tener mayor sensibilidad cuando se asigna taxonomía a secuencias de aminoácidos.

Construcción de modelos estadísticos para proteínas asociadas a la adquisición de hierro a partir de alineamientos múltiples de secuencias

Con el propósito de recuperar información posicional de los residuos de aminoácidos para cada familia de sideróforos, se realizaron alineamientos múltiples de las secuencias de proteínas incluidas en cada familia con el programa MUSCLE (Edgar, 2004) con los parámetros asignados por default. De manera paralela, se realizó un alineamiento múltiple con el mismo programa y parámetros para 10 secuencias anotadas por SEED como proteínas receptoras de aerobactina. La visualización de este último análisis fue mediante la plataforma Jalview (Martin et al., 2019), que además permite analizar y editar alineamientos múltiples de secuencias.

Posteriormente, se construyeron 995 perfiles de modelos ocultos de Markov (pHMM) a partir de los MSA de las diferentes familias de proteínas consideradas. Para ello, se empleó la opción *hmmbuild* del programa HMMER3 (Eddy, 2018, 2019). Con la opción *hmmsearch* fueron inicialmente entrenados los pHMM contra sí mismos, de esta manera se recuperó el valor de *bit score* mínimo, que es un valor de selección, para que cada modelo

pueda encontrar a un representante de la misma familia de proteínas en cualquier conjunto de datos. El *bit score* fue tomado en cuenta debido a que su estimación no depende del tamaño de la base de datos y le permite a un modelo oculto de Markov (HMM) discriminar entre secuencias de proteínas (Eddy, 2018)□.

Obtención de proteomas bacterianos para la búsqueda de sideróforos

Se seleccionaron 116 genomas bacterianos de phyla representativos del dominio Bacteria, propuestos en la reconstrucción filogenética del árbol de la vida que realizó Ciccarelli (et al., 2006). Posteriormente, se descargaron los archivos de secuencias de aminoácidos totales (proteomas) disponibles en la base de datos del NCBI por medio de su servidor FTP (<ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria>). Se dio prioridad a aquellos descritos como genomas de referencia o representativo. En caso de encontrar más de un proyecto de secuenciación, se seleccionaron los proteomas con la fecha de actualización más reciente. Cada proteoma fue descargado (Octubre 2019) en archivos de texto individuales con extensión .faa

Búsqueda de proteínas homólogas de síntesis y transporte de sideróforos en proteomas bacterianos

Los 995 pHMM generados se usaron para identificar proteínas involucradas en la síntesis de sideróforos y transporte de complejos ferri-sideróforos en los 116 proteomas bacterianos seleccionados previamente. Se usó la opción `hmmsearch` del programa HMMER3 (Eddy, 2018, 2019)□. Se realizaron cuatro diferentes ensayos tomando en cuenta distintos valores de corte de selección; de este modo, la primera condición no consideró un valor mínimo de *bit score*, la segunda respetó el mínimo valor de *bit score* establecido para cada modelo, la tercera consideró el valor de corte mínimo menos un valor de desviación estándar y en la cuarta condición se disminuyeron dos valores de desviación estándar al valor de *bit score* mínimo. Se realizaron un total de 115,652 búsquedas de proteínas homólogas por cada ensayo y se recolectaron los mejores aciertos que cumplieron con los valores de corte de selección.

Obtención de matriz de datos con las proteínas homólogas de síntesis y transporte de sideróforos

Para analizar la abundancia y distribución de las proteínas homólogas de adquisición de hierro encontradas con los pHMM se generaron las correspondientes matrices de datos que contenían esta información (Tabla S1). En las filas se organizaron los proteomas bacterianos ordenados filogenéticamente y en las columnas se ordenaron las proteínas analizadas. En las celdas intermedias se colocaron los valores correspondientes a la abundancia de las proteínas encontradas con los pHMM en cada proteoma. Estas matrices fueron usadas para los análisis de abundancia y distribución filogenética de las proteínas homólogas de síntesis y transporte de sideróforos.

Análisis de la abundancia y distribución de las proteínas homólogas asociadas a la adquisición de hierro

Con la intención de darle carácter cualitativo a los datos de abundancia y fácilmente vislumbrar la presencia o ausencia de los principales mecanismos asociados a la síntesis y transporte de sideróforos, se construyó una matriz binaria. De manera complementaria, se abordó el carácter cuantitativo haciendo una comparación de la abundancia de las proteínas analizadas en los diferentes proteomas. Para ello, fue necesario normalizar los datos de abundancia de las proteínas encontradas en cada proteoma, considerando el contenido total de proteínas predichas por cada genoma completo. Así, se obtuvo un porcentaje por cada organismo comparado que representa la proporción del genoma destinado a los mecanismos de adquisición de hierro por medio de sideróforos.

Para analizar los patrones de distribución de las proteínas homólogas encontradas con los pHMMs se empleó el lenguaje de programación estadístico R v. 3.6.1 (R Core Team, 2018) con la biblioteca ggplot2 (Wickham, 2016). Las matrices obtenidas previamente se graficaron con *heatmaps* ordenados de acuerdo con la filogenia de referencia (Ciccarelli et al., 2006). Este tipo de gráfico es una representación codificada por colores de las proteínas homólogas encontradas en cada uno de los genomas analizados. En la imagen obtenida es posible distinguir patrones de distribución entre grupos taxonómicos. El árbol filogenético fue editado desde el servidor de iTOL *Interactive tree of life* (<http://itol.embl.de/itol.cgi>) y posteriormente en el programa InkScape (Harrington, 2004-2005) para añadir datos sobre patogenicidad.

La distribución de las estrategias destinadas para adquirir hierro por cada genoma se graficó con un diagrama de violín, igualmente disponible desde la biblioteca ggplot2 (Wickham, 2016). Con la misma biblioteca de visualización de datos, se calculó la frecuencia relativa de las proteínas homólogas asociadas al aprovechamiento del hierro mediante sideróforos y se obtuvo un gráfico de barras apiladas con los genomas ordenados filogenéticamente. Asimismo, se editó el programa Inkscape (Harrington, 2004-2005) para añadir datos sobre patogenicidad.

Resultados

Construcción de la base de datos de referencia para proteínas involucradas en la biosíntesis y transporte de sideróforos

Mediante una exploración a la base de datos SEED *subsystems* identificamos 61 diferentes subsistemas asociados a la adquisición y el metabolismo del hierro. Por mencionar algunos se tiene a: Adquisición de hierro en *Streptococcus*, adquisición de hierro en *Vibrio*, sistema de transporte de hierro de tipo ABC, metabolismo del hierro de *Campylobacter*, sistema de transporte de hierro EfeUOB inducido por bajo pH, sistema de transporte de dicitrato férrico Fec, clúster de adquisición de hierro en *Thermus*, sistema receptor y sensor de sideróforos, transporte de hierro, adquisición de hierro mediada por salmoquelina, sistema encargado de quelar hierro mediado por petrobactina, cápsula de antígeno O importante para la persistencia ambiental, transporte de hierro ferroso EfeUOB inducido por bajo pH, ferredoxinas, ferredoxinas reducidas dependientes de NADH:NADDP⁺ oxidorreductasa, piruvato: ferredoxina oxidorreductasa, sideróforo deferoxamina E, síntesis de vibrioferrina, biosíntesis del grupo hemo, ensamblaje de cluster Fe-S, entre otros.

Las propiedades fundamentales de los subsistemas y la estructuración de la información de SEED *subsystems* constituyeron los cimientos de nuestro enfoque para construir una base de datos que incluyera las principales proteínas que están a cargo de la adquisición del hierro mediante sideróforos.

Para ello, recuperamos todo el conjunto de datos de proteínas disponibles en el servidor FTP SEED (<ftp://ftp.theseed.org/>) en un directorio que contiene 5 archivos de texto que integran información con al menos 178,000 familias de proteínas. Esta información fue sometida a un proceso de filtrado para conservar únicamente aquellas proteínas que estuvieran anotadas con funciones relacionadas con metabolizar hierro. Por lo que el conjunto de datos se redujo a 93,516 proteínas redundantes, que corresponden 6,566 proteínas codificantes (PEG) distintas.

Para disminuir la redundancia, el conjunto de datos se sometió a un proceso de agrupamiento (*clustering*) tomando como referencia el porcentaje de similitud (30% mínimo) y cobertura (70% longitud) entre las secuencias, obteniendo un total de 3,763 familias de proteínas. Debido a que el enfoque de este trabajo fue identificar aquellas proteínas que se

encargan de captar hierro del exterior y facilitar su internalización, las 3,763 familias de proteínas fueron curadas manualmente para eliminar aquellas que no intervinieran directamente con estos procesos. Por lo tanto, el conjunto original de 93,516 proteínas redundantes se redujo a un total de 1,007 familias de proteínas asociadas a la síntesis y transporte de sideróforos (Tabla 1, S2.1). La familia de proteínas mejor representada con 995 proteínas homólogas fue una parte del sistema de transporte de la enterobactina férrica (FepC) involucrada con la translocación del complejo ferri-sideróforo a través de la membrana citoplasmática. Los estudios disponibles sobre la regulación, la síntesis, el reconocimiento y el transporte de la enterobactina sugieren que pueda ser el mejor entendido de los sistemas de absorción de hierro mediados por sideróforos (Raymond, Dertz, & Kim, 2003)□.

Tabla 1. Familias de proteínas de síntesis y transporte de sideróforos mejor representadas en la base de datos construida a partir de SEED *subsystems*.

No. de proteínas	Familia de proteínas
995	Ferric enterobactin transport ATP-binding protein FepC (TC 3.A.1.14.2)
577	Ferric vibriobactin, enterobactin transport system, permease protein VctD (TC 3.A.1.14.6)
466	Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), permease component FhuB
418	2,3-dihydroxybenzoate-AMP ligase (EC 2.7.7.58) [enterobactin] siderophore
411	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EC 1.3.1.28) [enterobactin] siderophore
380	Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), periplasmic substrate binding protein FhuD
336	Siderophore staphylobactin biosynthesis protein SbnB
314	ABC-type Fe ³⁺ -siderophore transport system, permease component
290	Fe-bacillibactin uptake system FeuB
286	Ferric enterobactin transport system permease protein FepD (TC 3.A.1.14.2)
281	Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), ATP-binding protein FhuC
267	Ferric enterobactin transport system permease protein FepD (TC 3.A.1.14.2)
266	Aerobactin siderophore receptor lutA
255	Ferric enterobactin-binding periplasmic protein FepB (TC 3.A.1.14.2)
253	Citrate:6-N-acetyl-6-N-hydroxy-L-lysine ligase, alpha subunit (EC 6.3.2.27), aerobactin biosynthesis protein lucA
231	Siderophore [Alcaligin-like] biosynthesis complex, long chain
220	iron-chelator utilization protein
203	TonB-dependent hemin , ferrichrome receptor
195	Bacillibactin synthetase component F (EC 2.7.7.-)
194	Enterobactin exporter EntS

Además de la anotación del SEED, a partir de una revisión bibliográfica sobre las principales de los grupos funcionales y características de los sideróforos, se clasificó todo el conjunto de proteínas recuperado según las categorías que se mencionan más adelante. De acuerdo con los grupos funcionales (Fig. A) que poseen los sideróforos y que permiten

su unión al hierro, se les incluyeron en las principales familias de proteínas relacionadas con sideróforos. Las familias representadas fueron las proteínas relacionadas a la síntesis de catecolato, hidroxamato, carboxilato, combinados: hidroxamato-catecolato, carboxilato-hidroxamato, fenolato-carboxilato, salicilato-malonato-cisteína, citrato-carboxilato y catecol-carboxilato. Además, la mayor parte de las proteínas relacionadas con síntesis y transporte de sideróforos fueron incluidos en la categoría *otro* por no saber específicamente a qué tipo estructural de sideróforo sintetizan o están relacionadas, fueron los más abundantes; también se pueden observar enriquecidas las familias de proteínas relacionadas con los catecolatos, los hidroxamatos y los sideróforos mixtos: hidroxamato-catecolato e hidroxamato-carboxilato (Figura 1).

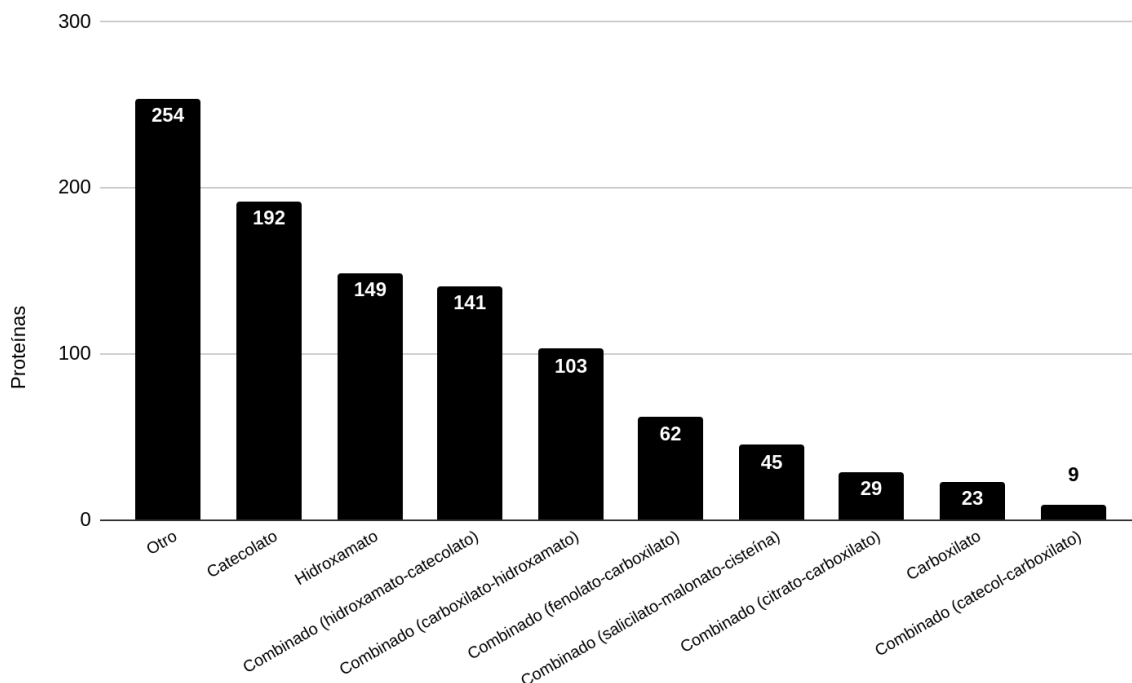


Figura 1. Abundancia de familias de proteínas de síntesis y transporte de sideróforos, clasificadas respecto a la estructura química presentada del sideróforo, presentes en la base de datos construida a partir de SEED *subsystems*.

El tipo de sideróforo hace referencia esencialmente a la procedencia filogenética de las proteínas relacionadas con los sideróforos, por lo que se consideraron 17 diferentes tipos de sideróforos que son: la pioverdina, enterobactina, aerobactina, ferricromo, pioquelina, yersiniabactina, bacilibactina, acromobactina, salmoquelina, alcaligina, vibrioferrina, antraquelina, estafilobactina, petrobactina, deferoxamina E, vulnibactina y vibriobactina (Figura 2). La categoría mejor representada es *sideróforo*, ya que su anotación no

proporcionó información más detallada, en cambio, podemos resaltar a la pioverdina, la enterobactina, aerobactina y ferricromo como los grupos más abundantes.

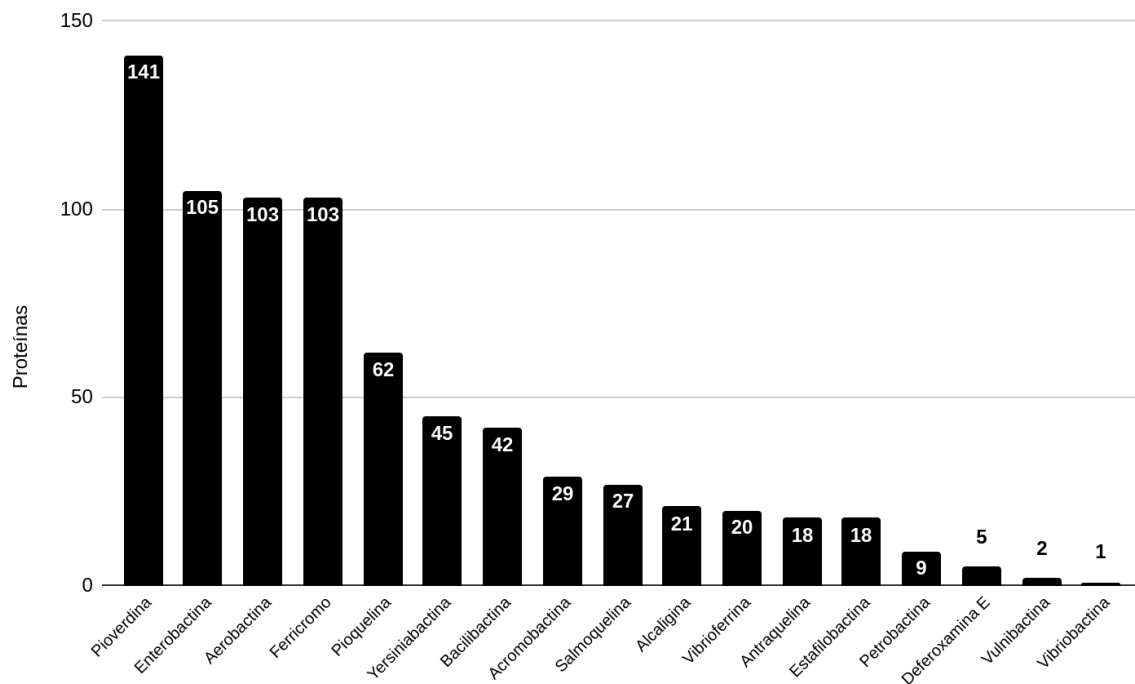


Figura 2. Los diferentes tipos de proteínas relacionadas con transporte y síntesis de sideróforos identificados por procedencia filogenética, en la base de datos construida a partir de SEED *subsystems*.

Se tomaron en cuenta las dos principales rutas de biosíntesis de sideróforos, por lo que se incluyó a la familia multienzimática de sintetetas de péptidos no ribosomales (NRPS) y la familia de sintetetas independientes de péptidos no ribosomales (NIS) (Figura 3). Del total de proteínas asociadas a sideróforos, se organizaron de tal manera que se pudieran distinguir únicamente los elementos asociados a síntesis de proteínas, por lo que se pudieron diferenciar 418 proteínas que en su mayoría siguen la ruta de biosíntesis que es dependiente de sinteteta de péptidos no ribosomales. Ambas rutas principales de síntesis se encuentran bien representadas, sin embargo, existe una tercera categoría que hace referencia a aquellas proteínas que no fue posible asignarles esta característica.

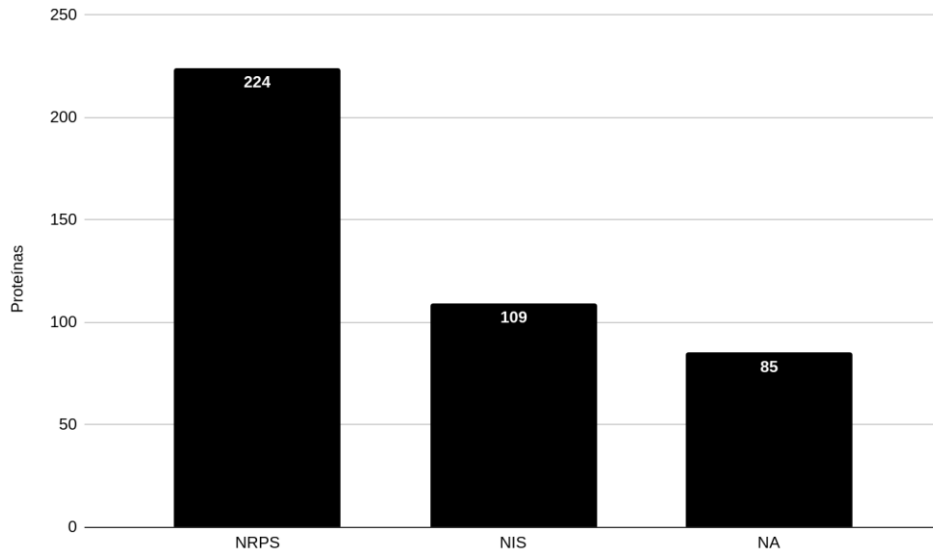


Figura 3. Principales rutas de biosíntesis de sideróforos representada en la base de datos construida a partir de SEED *subsystems*.

La asignación funcional hace referencia a la función biológica general de una proteína. Con base en la literatura disponible se establecieron cinco categorías que son síntesis, transporte, receptores, reacciones de tipo redox y regulación transcripcional (Figura 4). Las proteínas de biosíntesis de sideróforos se encuentran con mayor abundancia, seguidas por las proteínas encargadas del transporte y las proteínas receptoras que permiten la interacción con el complejo ferri-sideróforo.

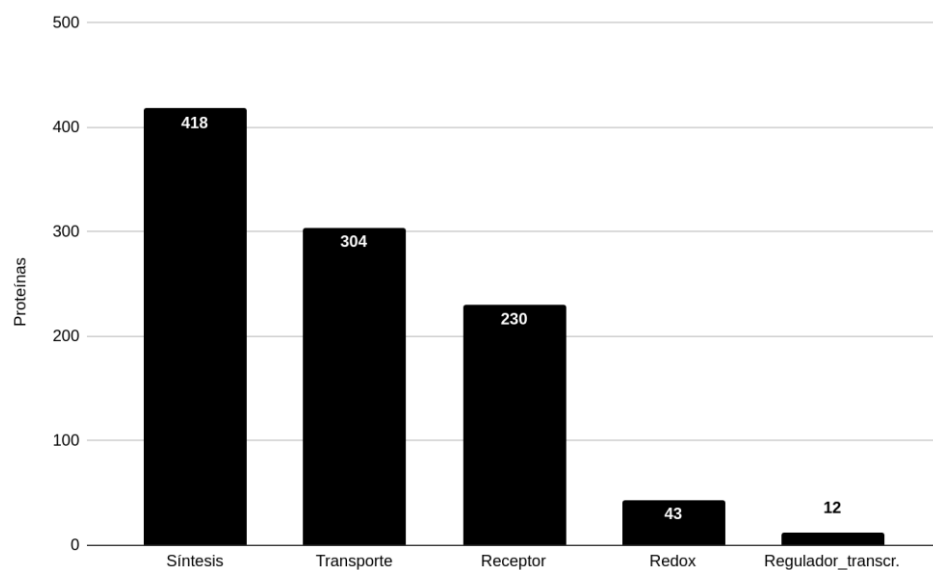


Figura 4. Principales funciones biológicas de las proteínas presentes en la base de datos construida a partir de SEED *subsystems*.

La anotación está definida por la base de datos SEED *subsystems* y hace referencia a las proteínas que forman parte de una misma ruta metabólica. Las anotaciones proporcionadas por el servidor SEED son implementadas a partir de métodos de alto rendimiento y asignan funciones a secuencias de proteínas o regiones de secuencias de DNA genómico (Aziz et al., 2012)□. La base de datos no redundante construida en este trabajo, considera 1,007 familias de proteínas asociadas a la adquisición de hierro a través de sideróforos.

Distribución filogenética de los sideróforos en Bacteria

La clasificación taxonómica se determinó mediante un análisis bioinformático de las proteínas recuperadas contra la base de datos no redundante del NCBI. La asignación taxonómica a nivel de phyla se logró para 994 proteínas incluidas en los grupos Euryarchaeota (Archaea), FCB (Fibrobacteres, Chlorobi, Bacteroidetes), Proteobacteria y Terrabacteria; se encontraron 5, 12, 764 y 213 representantes respectivamente (Figura 5). La asignación a nivel de género y especie sólo se logró para 367 proteínas, éstas representan 61 géneros y 111 especies diferentes (Tabla 2). Se puede observar que el taxón mejor representado en la base de datos es Terrabacteria, así como el género *Pseudomonas* y la bacteria *Bacillus cereus*.

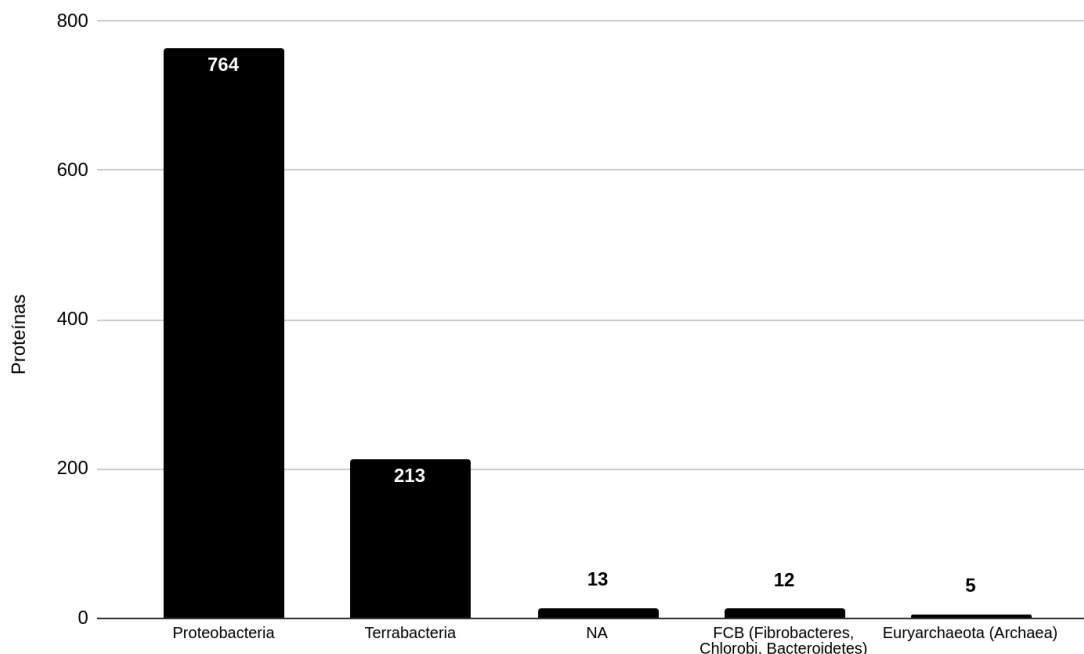


Figura 5. Asignación taxonómica a nivel de phylum para las proteínas encargadas de la adquisición de hierro presentes en la base de datos construida a partir de SEED

subsystems.

Tabla 2. Taxonomía a nivel de

Género	Frecuencia	Especie	Frecuencia
<i>Pseudomonas</i>	65	<i>Bacillus cereus</i> group	32
<i>Bacillus</i>	44	<i>Azotobacter vinelandii</i>	31
<i>Azotobacter</i>	31	<i>Yersinia pestis</i>	21
<i>Vibrio</i>	27	<i>Pseudomonas syringae</i> group genomsp. 1	15
<i>Burkholderia</i>	26	<i>Pseudomonas syringae</i> group genomsp. 2	14
<i>Yersinia</i>	26	<i>Pseudomonas putida</i>	12
<i>Streptomyces</i>	22	<i>Vibrio antiquarius</i>	11
<i>Acinetobacter</i>	10	<i>Acinetobacter baumannii</i>	10
<i>Escherichia</i>	7	<i>Vibrio alginolyticus</i>	9
<i>Stenotrophomonas</i>	7	<i>Burkholderia cenocepacia</i>	7
<i>Flavobacterium</i>	6	<i>Pseudomonas aeruginosa</i>	7
<i>Frankia</i>	6	<i>Pseudomonas fluorescens</i>	7
<i>Salinispora</i>	6	<i>Pseudomonas protegens</i>	7
<i>Rhizobium</i>	5	<i>Stenotrophomonas maltophilia</i>	7
<i>Sinorhizobium</i>	5	<i>Burkholderia thailandensis</i>	6
<i>Agrobacterium</i>	4	<i>Flavobacterium</i>	6
<i>Xanthomonas</i>	4	<i>Burkholderia pseudomallei</i>	5
<i>Campylobacter</i>	3	<i>Rhizobium leguminosarum</i>	5
<i>Cytophaga</i>	3	<i>Sinorhizobium meliloti</i>	5
<i>Desulfitobacterium</i>	3	<i>Streptomyces albidoflavus</i> group	5

géner
o y
especi
e de
las
proteí
nas
recopil
adas.

Con el objetivo de organizar las proteínas según su procedencia, se retomó la clasificación taxonómica y se complementó con una revisión bibliográfica para asignar el huésped en que fue inicialmente descrito cada sideróforo. La información taxonómica a nivel de género y especie fue la base para señalar el estado ante la tinción de Gram de la bacteria de la cual

procede cada proteína. Se reconocen 672 proteínas procedentes de organismos Gram negativos y 78 proteínas de bacterias Gram positivos, el resto no pudo ser reconocido (Figura 6).

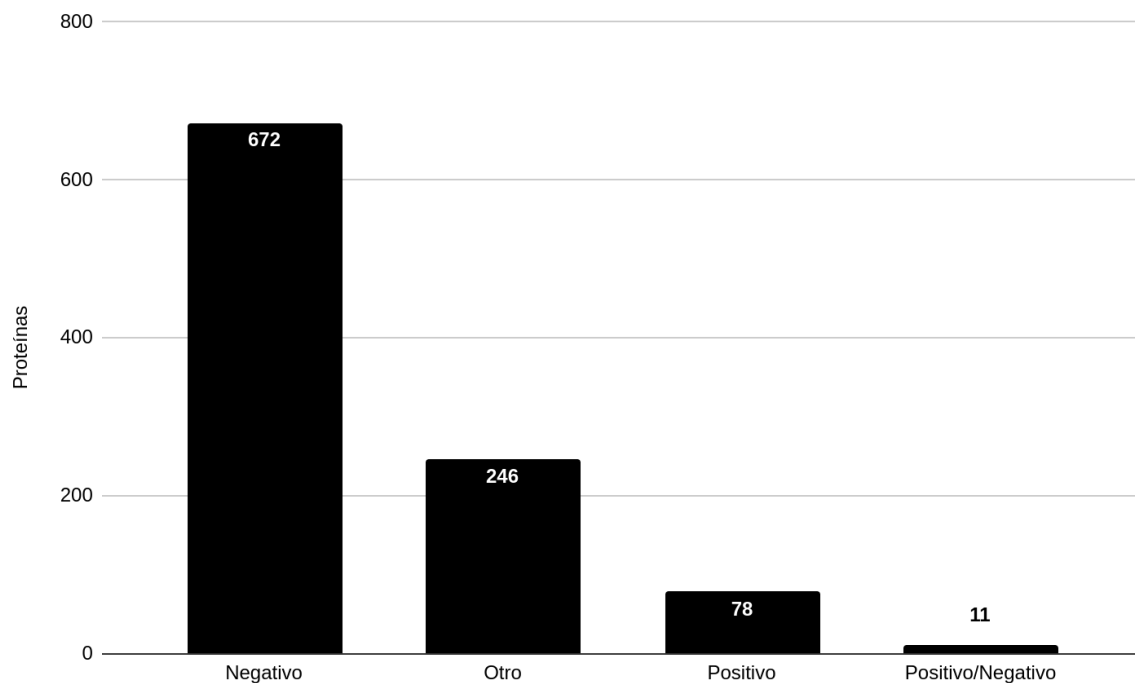


Figura 6. Tinción Gram asociada al huésped de la proteína presente en la base de datos construida a partir de SEED *subsystems*.

La construcción de la base de datos de proteínas asociadas únicamente a sideróforos se realizó mediante un método de agrupamiento de secuencias de proteínas (30% identidad; 70% longitud) distribuidas en 1,007 familias de proteínas. A pesar de ser 1,007 familias de proteínas, solamente hay 215 anotaciones redundantes.

Después de inspeccionar manualmente la anotación, se observó que estos elementos únicos recuperan 17 tipos de sideróforos y las respectivas proteínas responsables de su síntesis, transporte, regulación transcripcional y reacciones de reducción-oxidación (Figura 7). Es de resaltar, que existe un total de 39 proteínas que no pudieron ser asociadas a un tipo específico de sideróforo por su anotación. Además, se incluye una categoría en la que las proteínas involucradas pueden hacer uso mixto de enterobactina y vibriobactina.

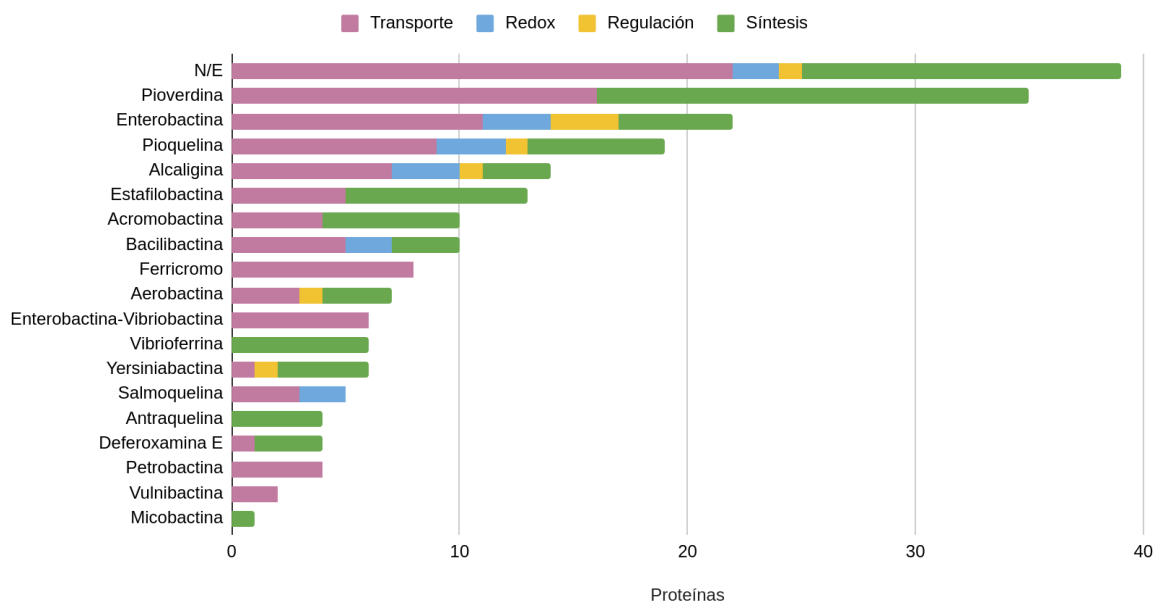


Figura 7. La anotación funcional propuesta por SEED permite distinguir 215 proteínas únicas involucradas específicamente en el aprovechamiento de hierro a través de 17 tipos de sideróforos. En diferentes colores se indica la proporción de proteínas destinadas a la síntesis, transporte, regulación transcripcional y reacciones de reducción-oxidación de cada sideróforo. La categoría N/E hace referencia a las proteínas que no pudieron ser clasificadas por tipo de sideróforo, siguiendo la anotación funcional.

Con el criterio de agrupamiento basado en la comparación entre secuencias, se esperaba que las anotaciones funcionales fueran similares dentro de cada familia; sin embargo, proteínas con la misma anotación funcional se hallaron distribuidas en diferentes familias, lo que puede hablar de paralelismo, convergencia o diversificación evolutiva a nivel de secuencia. Para evaluar estas posibilidades, fue necesario comparar un grupo de proteínas que compartieran la misma anotación.

Se realizó un alineamiento múltiple de 10 proteínas anotadas como receptoras de aerobactina lutA del phylum Proteobacteria y se compararon contra una proteína receptora para el mismo sideróforo descrita anteriormente en *V. parahaemolyticus* por Funahashi (et al., 2003). Las proteínas poseen diferentes longitudes, siendo de 87 residuos de aminoácidos la más corta y de 753 residuos la más larga.

Se observó que el valor de porcentaje de identidad (% PID, número de residuos idénticos por cada 100 pares de residuos de aminoácidos) de las secuencias no depende de la longitud de las mismas (Tabla 3), lo cual seguramente es consecuencia del proceso de

curación de SEED, donde las familias de proteínas han sido definidas de acuerdo a la estructura de dominio común y la similitud de secuencia de longitud completa (Meyer, Overbeek, & Rodriguez, 2009)□.

El alineamiento múltiple de proteínas muestra indicios del péptido señal típico de 25 residuos de aminoácidos del receptor férrico de aerobactina lutA, ubicado al inicio de las secuencias. También se puede observar la localización tentativa de la caja TonB, que es un motivo conservado con el que interactúa la proteína TonB y está cerca del extremo N-terminal, en las diferentes secuencias (Figura 8) (Funahashi et al., 2003; Noinaj, Guillier, Barnard, & Buchanan, 2010)□. La extensión de los huecos añadidos para lograr el alineamiento es muy amplia y los motivos con > 80% de identidad son pocos.

Tabla 3. Proteínas receptoras de aerobactina lutA. Se muestra el identificador (ID) y la longitud de cada proteína empleada en el MSA, así como el %PID contra la proteína de referencia lutA de *V. parahaemolyticus*. Las proteínas se muestran ordenadas de manera ascendente según el porcentaje de identidad, es así que la proteína 33334 tiene el menor %PID y la proteína 8755 posee el mayor valor de % PID.

ID proteína	Longitud (aa)	% PID
33334	696	24.01
79989	753	26.28
80436	214	34.72
34008	720	39.46
31848	726	42.8
34403	231	47.64
33566	322	48.77
79720	704	53.28
80437	87	57.47
8755	723	90.48

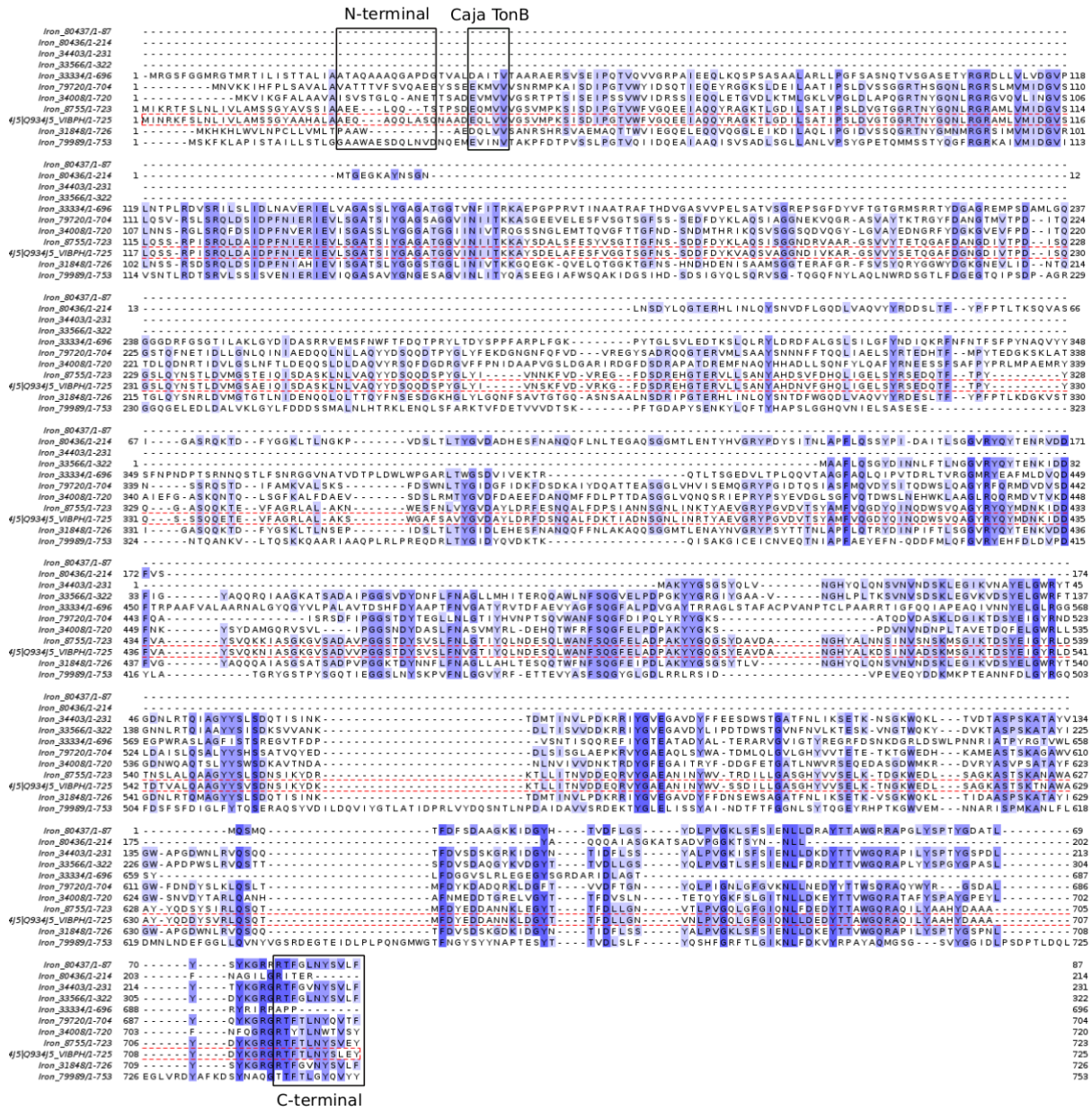


Figura 8. Alineamiento múltiple de proteínas receptoras de aerobactina *lutA*. Las secuencias comparadas están ordenadas de manera ascendente respecto a su longitud. La posición tentativa de la caja TonB, así como los residuos de los extremos N-terminal y C-terminal de estas proteínas receptoras de sideróforos se encuentran señalados por recuadros. La línea punteada roja indica la proteína de referencia *lutA*. El gradiente de colores de (4) indica el porcentaje de residuos de aminoácidos conservados por cada columna, partiendo del blanco (<=40%), gris (>40%), azul claro (>60%) y azul oscuro (>80%).

Búsqueda masiva de perfiles ocultos de Markov de proteínas relacionadas a transporte y biosíntesis de sideróforos, para construir perfiles filogenéticos en proteomas bacterianos

A partir de 70,171 secuencias de proteínas de referencia, organizadas en 1,007 familias de proteínas, se construyeron 995 pHMMs capaces de reconocer proteínas homólogas en 116 proteomas bacterianos. Las 12 familias restantes se descartaron por no reunir suficientes secuencias homólogas necesarias para la construcción de los respectivos pHMMs. Las proteínas descartadas se describen como módulos de péptidos no ribosomales para la síntesis de pioverdina, sintetetas de yersiniabactina, módulos de péptidos no ribosomales para la síntesis de pioquelina, una permeasa del complejo transportador de hidroxamato férrico y proteínas quelantes de hierro.

Los 995 pHMMs se implementaron en la búsqueda de proteínas homólogas en los proteomas propuestos por la filogenia de Ciccarelli (et al., 2006) y se consideraron cuatro diferentes condiciones experimentales (búsqueda bioinformática). En la primera condición se evaluó la capacidad de los pHMMs para detectar el máximo número de proteínas homólogas, por lo que no se definió ningún valor de corte en la búsqueda. Este primer ensayo validó la capacidad de los pHMMs para reconocer proteínas asociadas a la adquisición de hierro en los proteomas seleccionados. En la segunda condición experimental se definió un valor de selección a cada perfil para que fuera capaz de reconocer una proteína homóloga con la mínima probabilidad de reconocerla en una base de datos. Estos experimentos fueron la base de los ensayos de ajuste de los pHMMs y con ello fue necesario considerar la variabilidad de la capacidad de reconocer proteínas homólogas de cada perfil en una base de datos.

El primer ensayo de identificación de proteínas homólogas en los 116 proteomas bacterianos no consideró un valor de restricción de selección para los modelos, por lo que los perfiles tuvieron la capacidad de reconocer el máximo conjunto de proteínas potencialmente homólogas en los proteomas seleccionados. Se obtuvieron un total de 675,119 coincidencias en los proteomas que se distribuyen en las 995 familias de proteínas consideradas en este proyecto. La bacteria simbiótica de plantas *Bradyrhizobium japonicum* fue el microorganismo en el que más coincidencias se recolectaron con 20,863 *hits*, mientras que la bacteria endosimbionte *Buchnera aphidicola* presentó el menor valor con 348 *hits*.

El pHMM capaz de reconocer al componente de unión al ATP de los transportadores de tipo ABC para la pioverdina obtuvo 8,114 *hits*, siendo así el que presentó mayor número de coincidencias. La sintetasa de yersiniabactina con una sola coincidencia fue la de menor abundancia. A continuación, se muestra la distribución de las proteínas homólogas reclutadas bajo los parámetros anteriormente descritos (Figura 9).



Figura 9. Distribución de proteínas para la adquisición de hierro. En las filas se enlistan los nombres de las especies bacterianas, cuyos proteomas fueron analizados, las columnas corresponden a las familias de proteínas. La parte central corresponde a la frecuencia de las proteínas homólogas encontradas. A mayor abundancia, el gradiente del color se observa más oscuro. Para mayor resolución de la imagen, consultar el enlace:

<https://drive.google.com/file/d/1Cep2J6tl8vRLvqCZRyiJqHnkcH3qB2aP/view?usp=sharing>

El siguiente ensayo también se llevó a cabo con los 995 pHMMs contra los 116 proteomas bacterianos, a excepción de que se estableció un umbral de selección mínimo para cada modelo construido. La puntuación de selección mínima se estableció mediante el valor de *bit score* de cada pHMM y hace referencia a la mínima probabilidad de encontrar una secuencia homóloga (Eddy, 2018) □ .

Las 8,087 proteínas encontradas pertenecen a 713 familias, donde el microorganismo con mayor abundancia de proteínas encontradas fue la bacteria patógena de plantas *Pseudomonas syringae* con 836 coincidencias; por el contrario, los microorganismos con menor abundancia fueron *Aquifex aeolicus*, *Bifidobacterium longum*, *Haemophilus ducreyi*, *Helicobacter hepaticus*, *Helicobacter pylori*, *Onion yellows phytoplasma* y *Ureaplasma parvum* con sólo una coincidencia por proteoma. Los perfiles no encontraron coincidencias en un total de 21 proteomas, los cuales se enlistan a continuación: *Borrelia sp. CA690*, *Buchnera aphidicola*, *Candidatus Blochmannia floridanus*, *Chlamydia caviae*, *Chlamydia muridarum*, *Chlamydia trachomatis*, *Chlamydophila pneumoniae CWL029*, *Mycoplasma gallisepticum*, *Mycoplasma genitalium*, *Mycoplasma mobile*, *Mycoplasma mycoides*, *Mycoplasma penetrans*, *Mycoplasma pneumoniae*, *Mycoplasma pulmonis*, *Porphyromonas gingivalis*, *Rickettsia conorii*, *Rickettsia prowazekii*, *Synechococcus sp. WH 8102*, *Treponema pallidum*, *Wigglesworthia glossinidia* y *Wolbachia sp. wMel AMD*. Cabe destacar que el estilo de vida de la mayoría de los organismos cuyos proteomas en los que no se encontraron homólogos son patógenos y endosimbiontes de animales, a excepción de *Synechococcus sp. WH 8102* que es una cianobacteria marina no patógena.

La familia de proteínas con mayor número de aciertos fue el componente F de la sintetasa de bacilobactina con 612 *hits*. En cambio, fueron 245 familias que tuvieron sólo una coincidencia en todos los proteomas (Figura 10).

Los resultados obtenidos a partir de los valores mínimos de selección sugieren que son valores de corte muy estrictos, por ello se realizaron ajustes posteriores a los pHMMs para aumentar la posibilidad de identificar proteínas homólogas para la adquisición de hierro.



Figura 10. Distribución de proteínas para la adquisición de hierro. En las filas se encuentran acomodados los nombres de los organismos cuyos proteomas fueron analizados, mientras que las columnas corresponden a las familias de proteínas. La parte central corresponde a la frecuencia de las proteínas homólogas encontradas. A mayor abundancia, el gradiente del color se observa más oscuro. Para mayor resolución de la imagen, consultar el enlace: https://drive.google.com/file/d/1-95R12_up-F7ohvG3xSsVUg14SPaYdm_/view?usp=sharing

El ajuste de cada perfil se realizó considerando el 68% ($\mu - 1\sigma$) y el 95% ($\mu - 2\sigma$) de variabilidad de su capacidad de reconocer proteínas homólogas (*bit score*) en una base de datos y que es calculado por HMMER3, esto mejoró la capacidad de los pHMMs para identificar proteínas asociadas al proceso de quelar hierro en los diferentes proteomas.

Con el ensayo que implementó el 68% de variabilidad, se encontraron 16,203 proteínas homólogas incluidas en 751 familias. El proteoma de la bacteria patógena *Pseudomonas syringae* fue en el que se encontraron un mayor número de proteínas homólogas, con 1,180 aciertos; mientras que los patógenos *Mycoplasma genitalium*, *Mycoplasma gallisepticum*, *Chlamydia trachomatis*, *Chlamydia muridarum* y *Mycoplasma mycoides* tuvieron el mínimo de dos aciertos. Únicamente no se identificaron proteínas homólogas en el patógeno *Mycoplasma pulmonis*.

La familia de proteínas que incluye a la bomba de eflujo de tipo ABC del sideróforo pioquelina con 1,297 coincidencias, fue la familia de mayor abundancia; en tanto el mínimo valor lo presentaron 225 familias mostrando una sola coincidencia (Figura S3.1).

Cuando se consideró la diferencia de dos valores de desviación estándar de probabilidad de que un perfil reconozca una proteína homóloga, es decir, el 95% de variabilidad del *bit score* calculado por HMMER3, se encontraron 41,946 proteínas distribuidas en 772 familias. Con este ajuste de valores de selección de proteínas homólogas, fue posible encontrar homólogos en todos los proteomas analizados. El proteoma con mayor número de proteínas homólogas encontradas fue el del patógeno *Pseudomonas syringae*, en el cual se observaron 1,555 coincidencias; por el contrario, en el endosimbionte *Buchnera aphidicola* se encontró el mínimo valor de coincidencias con 18 *hits*.

La familia de proteínas mejor representada bajo los límites de este ensayo fue la bomba de eflujo de tipo ABC del sideróforo pioquelina con 6967 coincidencias, por el contrario, la frecuencia mínima de coincidencias fue de un acierto y la presentaron 209 familias de proteínas (Figura S3.2).

A manera de comparar y establecer las mejores condiciones de búsqueda, los resultados obtenidos en cada uno de los ensayos se recapitulan en la siguiente tabla:

Tabla 5. Implementación y ajuste de 997 pHMMs. La tabla compara el mínimo valor de *hits* o coincidencias (Mín. *hits*), el máximo valor de *hits* (Máx. *hits*), el total de *hits* observados y la cantidad de familias de proteínas en las que se distribuyen los homólogos encontradas por búsqueda. También se incluye la frecuencia de los perfiles o proteomas que no obtuvieron *hits* por cada evento experimental. Se hace distinción entre la información obtenida tomando como referencia los perfiles (n=995) y los proteomas (n=116).

	pHMMs (n=995)					Proteomas (n=116)			
	Familias de prot.	Sin hits	Mín. hits	Máx. hits	Total	Sin hits	Mín. hits	Máx. hits	Total
Sin valor de selección	995	0	2	8114	675119	0	348	20863	675119
Valor de selección mínimo	713	282	1	612	8087	21	1	836	8087
1 DS	751	244	1	1297	16203	1	2	1180	16203
2 DS	760	235	1	6967	41946	0	18	1555	41946

La estimación a partir de dos desviaciones estándar de la capacidad que tiene cada perfil para reconocer una proteína homóloga fue el método que pudo representar de mejor manera la variabilidad de las familias de proteínas analizadas en la totalidad de los proteomas, es que se eligieron las condiciones de este ensayo para los análisis definitivos de abundancia y distribución filogenética de las proteínas homólogas de síntesis y transporte de sideróforos.

A continuación, la matriz construida a partir de la diferencia de dos desviaciones estándar se tomó como base para realizar un análisis cualitativo. Para ello, las frecuencias absolutas de las proteínas encontradas en los diferentes proteomas se transformaron en datos binarios, de tal forma que los ceros indican ausencia de dicha familia y los uno denota presencia. En este análisis se pudo observar cómo es la distribución total de las familias de proteínas en los diferentes organismos (Figura 13).

Se identificaron dos familias de proteínas presentes de manera ubicua. Una es la ATPasa asociada a la bomba de eflujo de tipo ABC del sideróforo pioquelina, la otra corresponde a la ATPasa del complejo transportador ABC FhuC involucrado en la importación de hidroxamato férrico y que es responsable del acoplamiento de energía al sistema de

transporte.

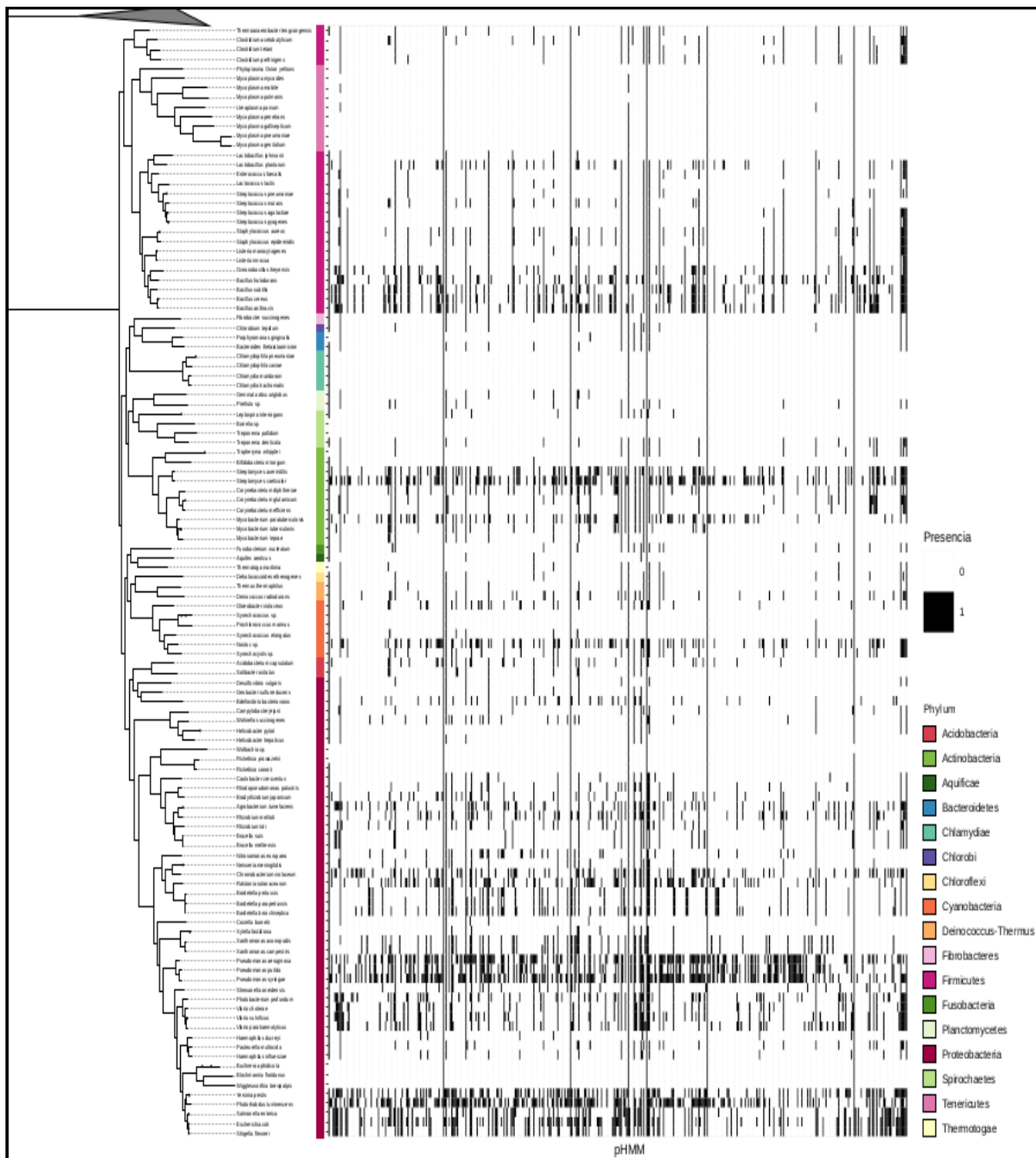


Figura 13. Distribución de las familias de proteínas responsables de quelar hierro mediante sideróforos en 116 proteomas representativos. En las filas están enlistados los proteomas ordenados filogenéticamente, las columnas corresponden a las familias de proteínas. La parte central corresponde a la presencia o ausencia de las proteínas homólogas encontradas, negro denota presencia y blanco hace referencia a ausencia. Para mayor resolución de la imagen, consultar el enlace: <https://drive.google.com/file/d/19xglQVQFH9E9iFi0BxE9zi-9aj46RnVk/view?usp=sharing>

Para conocer el porcentaje del proteoma que dedica cada organismo a los mecanismos de adquisición de hierro, se realizó una normalización de los datos considerando el número de proteínas homólogas recuperadas y el tamaño del proteoma de cada bacteria. Se observó que los organismos analizados dedican al menos 3% de sus proteomas para adquirir hierro del medio extracelular. Tal es el caso de la bacteria de agua dulce *Gemmata obscuriglobus*, que pertenece al phylum Planctomycetes y dedica 3.11% de su proteoma a estrategias para adquirir hierro. Por otra parte, son 66 organismos que destinan por lo menos 10% del proteoma a la biosíntesis y transporte de sideróforos para la adquisición de hierro. De acuerdo con los resultados, el organismo que asigna mayor porcentaje de su proteoma al consumo de hierro es *Pseudomonas syringae* con 30.56% (Figuras 14-16).

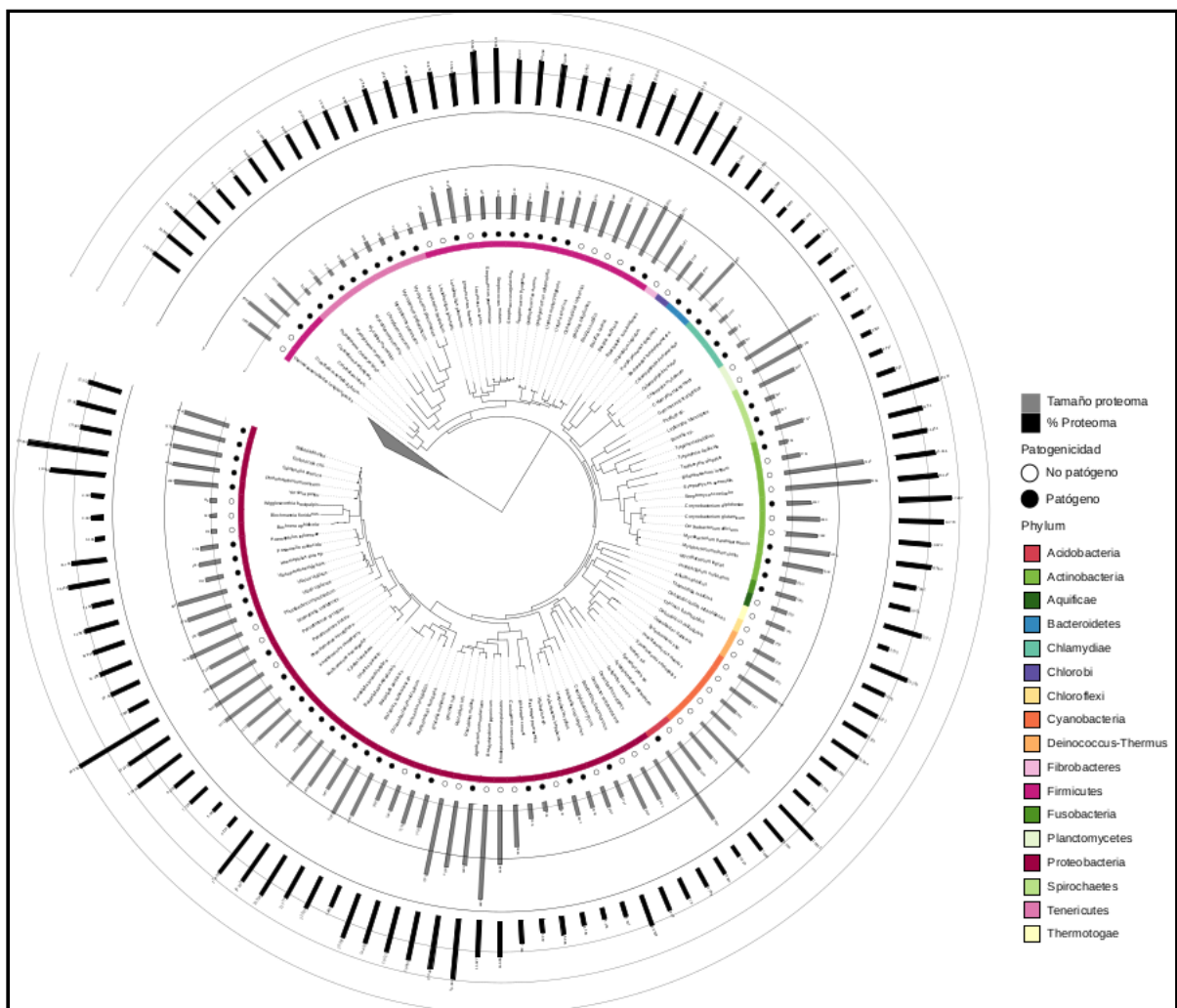


Figura 14. Distribución de las estrategias de biosíntesis y transporte de sideróforos. Los proteomas se encuentran ordenados filogenéticamente, con barras de colores se señala el phylum al que pertenecen. Con círculos se indica la patogenicidad de los organismos. El círculo interior con barras grises hace referencia al tamaño del proteoma de cada bacteria.

El círculo exterior con barras negras muestra el porcentaje de proteínas involucradas con la adquisición de hierro en cada organismo. Para mayor resolución de la imagen, consultar el enlace:

<https://drive.google.com/file/d/1sKG5YFLjcBbEC9b2PIhugMWsSPm9328H/view?usp=sharing>

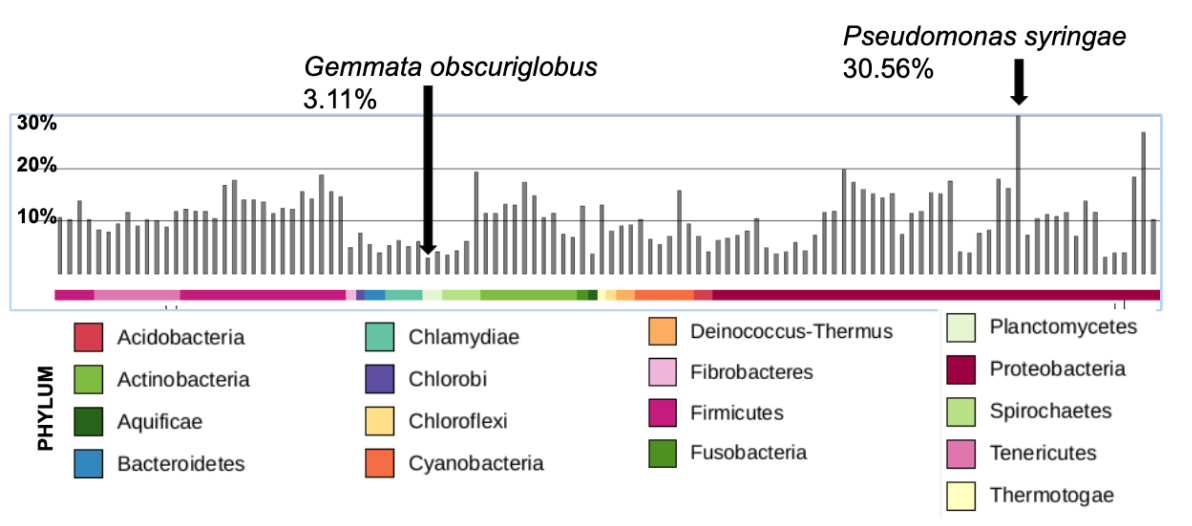


Figura 15. Proporción del proteoma destinado a la biosíntesis y transporte de sideróforos. Los proteomas se encuentran ordenados filogenéticamente y con barras de colores se señala el phylum al que pertenecen. Las barras grises corresponden al porcentaje de proteínas involucradas con la adquisición de hierro en cada organismo. Para mayor resolución de la imagen, consultar el enlace:

https://drive.google.com/file/d/18LANBblSxc0PXyKu-sQxfoUxxady_4_w/view?usp=sharing

De la distribución de las proteínas para adquirir hierro, se puede decir que *Pseudomonas syringae* y *Photorhabdus luminescens* se comportan como organismos especialistas de hierro, ya que sus proteomas destinan el mayor porcentaje de proteínas a este proceso y se ubican en la sección más delgada de la Figura 16. La sección más amplia de la trama del diagrama de violín está ligada a la proporción de los datos ubicados en esa área, de esta forma observamos que la mayoría de los organismos emplea cerca del 10% del total de proteínas para aprovechar el hierro del medio, independientemente del tamaño de su genoma.

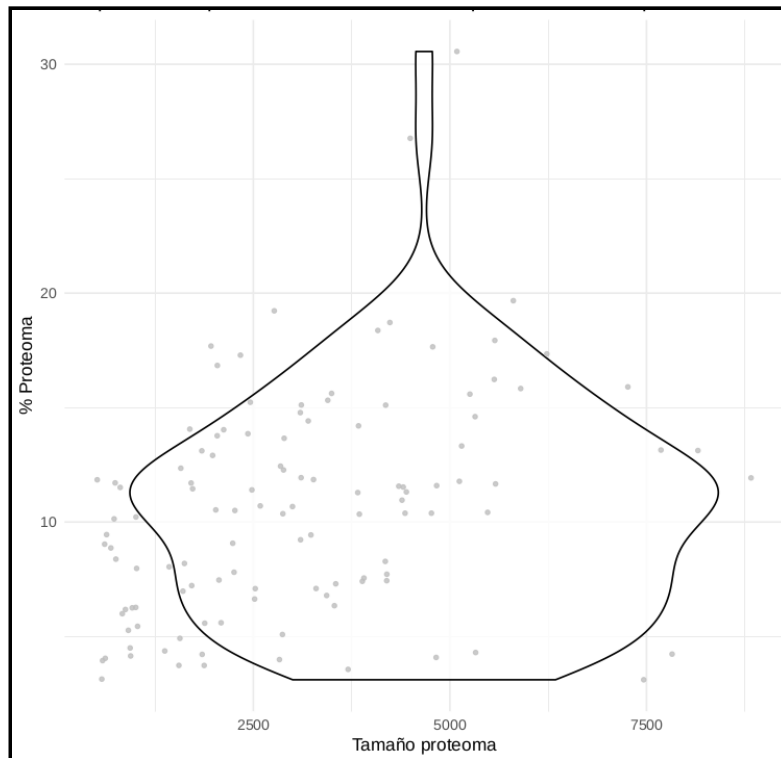


Figura 16. Distribución de las estrategias para adquirir hierro biodisponible por medio de sideróforos. La línea negra delimita la probabilidad de que un proteoma de cierto tamaño asocie un porcentaje de proteínas asociadas a la adquisición de hierro mediante sideróforos. Los puntos grises representan los proteomas analizados respecto al porcentaje de estrategias observadas.

La distribución de las proteínas homólogas relacionadas a sideróforos se analizó tomando en cuenta la filogenia propuesta por Ciccarelli, tomando en cuenta también la potencial patogenicidad de cada organismo porque es un tipo de interacción biológica de principal interés (Figuras 14) (Welch, 2015) □. Las proteínas vistas como estrategias que siguen los 17 phyla para aprovechar el hierro disponible mediante sideróforos son diferentes, sin embargo, se observan tendencias similares cuando las asociamos a la ecología de las bacterias. Se ha hecho énfasis en que los sideróforos son determinantes en las interacciones sociales porque la capacidad de producirlos es fundamental en el establecimiento de interacciones sinérgicas o antagónicas con otros microorganismos para competir por el hierro disponible del medio. Además, la capacidad de las bacterias para colonizar y sobrevivir en los huéspedes durante la infección depende de su aptitud para encontrar hierro e importarlo en cantidades adecuadas. Sin embargo, el huésped también necesita de este metal, lo que conlleva a la competencia entre el huésped y el patógeno por el acceso al hierro; por un lado las bacterias secretan sideróforos, mientras que los

huéspedes han desarrollado estrategias para restringir el acceso de iones metálicos esenciales a los patógenos (Kramer, Özkaya, & Kümmerli, 2020; Schalk & Cunrath, 2016)□.

De los proteomas incluidos en este estudio, podemos decir que los phyla Tenericutes, Bacteroidetes, Chlamydiae, Spirochaetes y Fusobacteria sólo incorporan ejemplos de bacterias patógenas. En cambio, los phyla Fibrobacteres, Chlorobi, Planctomycetes, Aquificae, Thermotogae, Chloroflexi, Deinococcus-Thermus y Cyanobacteria únicamente abarcan organismos no patógenos. En cambio, los phyla Firmicutes, Actinobacteria, Acidobacteria y Proteobacteria contienen ejemplos de bacterias consideradas como patógenas y no patógenas.

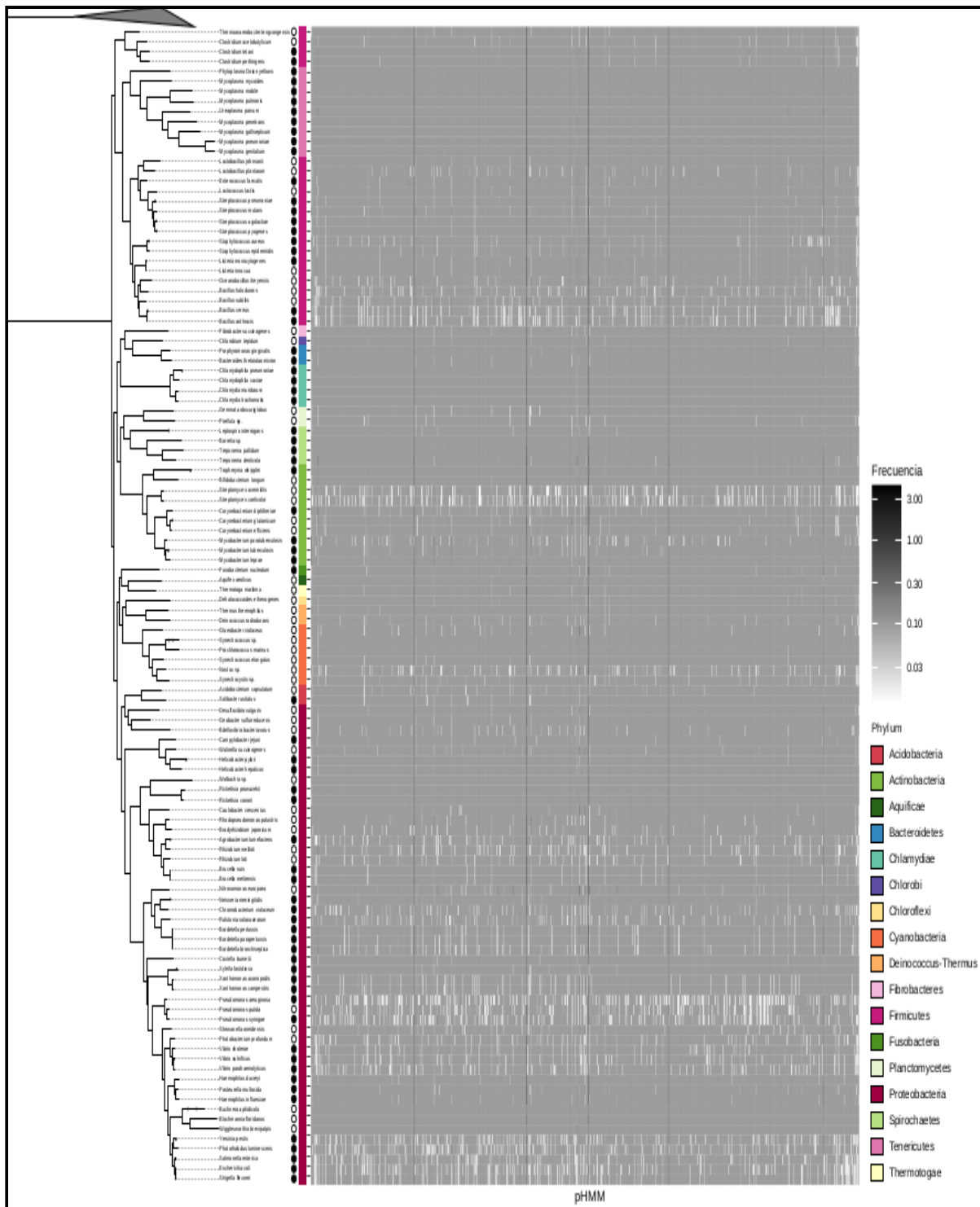


Figura 17. Perfil filogenético de proteínas para la adquisición de hierro normalizadas por el tamaño total de cada proteoma. En las filas están enlistados los proteomas ordenados filogenéticamente y las barras de colores señalan el phylum correspondiente. Con círculos se indica la patogenicidad de los organismos. En las columnas se ubican las familias de proteínas y la parte central ilustra la frecuencia de las proteínas homólogas encontradas. A mayor abundancia, el gradiente del color se observa más oscuro. Para mayor resolución de

la imagen, consultar el enlace:
<https://drive.google.com/file/d/1i9la1LEQqdGKzzOl3hfmkY7ZzviuztQa/view?usp=sharing>

La distribución de las proteínas de adquisición de hierro demostró que la principal función biológica que desempeñan es el transporte de sideróforos, seguida de las síntesis y la presencia de receptores específicos que interactúan con el complejo ferri-sideróforo. En menor medida se observan las proteínas asociadas a los procesos de oxidación-reducción y proteínas que regulan la transcripción de sideróforos (Figura 18).

El phylum Proteobacteria es el único conjunto de organismos donde se observó la presencia de proteínas que cubren todas las diferentes funciones biológicas asignadas e intervienen con la síntesis, transporte, respuesta al estrés oxidativo, regulación de la transcripción de sideróforos. Las bacterias *Buchnera aphidicola*, *Candidatus Blochmannia floridanus*, *Wigglesworthia glossinidia*, *Wolbachia sp.* y *Rickettsia prowazekii* destacan por presentar exclusivamente proteínas transportadoras. *Pseudomonas syringae* y *Photobacterium luminescens* destacan por ser las bacterias que requieren más proteínas para la síntesis de sideróforos; por otro lado, *Haemophilus ducreyi*, *Coxiella burnetii* y *Rickettsia conorii* sólo tienen proteínas de transporte y síntesis de sideróforos. Las proteobacterias fueron el único phylum con reguladores transcripcionales de sideróforos identificados, siendo *Pseudomonas aeruginosa* la bacteria que dedica mayor cantidad de proteínas a esta función. *Chromobacterium violaceum* fue la bacteria con más proteínas asociadas a procesos redox. Por último, las proteínas asociadas a receptores específicos de sideróforos se vieron enriquecidas en el proteoma de *Nitrosomonas europaea*.

Las bacterias del phylum Acidobacteria presentaron proteínas asociadas a las diferentes funciones biológicas de biosíntesis y transporte de sideróforos, con la excepción de reguladores transcripcionales. *Acidobacterium capsulatum* presenta mayor cantidad de proteínas de síntesis y receptores de sideróforos, en cambio, *Solibacter usitatus* destina más elementos que intervienen en procesos redox y transporte.

En cuanto al phylum Cyanobacteria, podemos decir que *Nostoc sp.* fue la bacteria que da preferencia a proteínas que sintetizan sideróforos; sin embargo, las bacterias *Synechococcus sp.*, *Prochlorococcus marinus* y *Synechococcus elongatus* dan preferencia exclusivamente a la síntesis y transporte de sideróforos. En cuanto a proteínas que regulan procesos redox sólo estuvieron presentes en *Gloeobacter violaceus* y *Nostoc sp.* La bacteria *Gloeobacter violaceus* también fue el organismo con mayor frecuencia de proteínas asociadas a receptores específicos de sideróforos.

En el phylum Deinococcus-Thermus observamos proteínas que cumplen con la síntesis, transporte y procesos redox. *Deinococcus radiodurans* da preferencia a las proteínas transportadoras, en comparación con *Thermus thermophilus* que concede más proteínas a la síntesis y procesos redox.

Los phyla Chloroflexi y Thermotogae presentan mayor número de proteínas transportadoras, seguidas por las asociadas a síntesis y procesos redox.

Igualmente, en *Aquifex aeolicus* que pertenece al phylum Aquificae, las proteínas transportadoras tienen un papel principal, seguida de las que cumplen con la síntesis, funciones redox y receptores de sideróforos.

En el proteoma de *Fusobacterium nucleatum* que pertenece al phylum Fusobacteria, observamos la preferencia de proteínas transportadoras, así como las proteínas de síntesis y receptoras.

En el phylum Actinobacteria observamos que *Tropheryma whipplei* únicamente presenta proteínas transportadoras; el resto de las bacterias sólo incluyen proteínas involucradas con el transporte, síntesis y procesos redox. *Mycobacterium avium* y *Mycobacterium tuberculosis* son los organismos con mayor número de proteínas de síntesis y proteínas redox.

En las bacterias *Borrelia sp.* y *Treponema pallidum* del phylum Spirochaetes, observamos la presencia exclusiva de proteínas transportadoras. *Treponema denticola* y *Leptospira interrogans* son las únicas bacterias de este phylum que presentan proteínas de síntesis. Además, el proteoma de *Treponema denticola* presenta una mínima proporción de proteínas redox y *Leptospira interrogans* incluye proteínas asociadas a receptores que reconocen específicamente sideróforos.

En el phylum Planctomycetes observamos que *Pirellula sp.* presenta proteínas asociadas al transporte y síntesis, en cambio, en el microorganismo *Gemmata obscuriglobus* también observamos proteínas redox.

El phylum Chlamydiae presentó únicamente proteínas relacionadas al transporte y síntesis de sideróforos, los cuatro organismos comparados presentaron estas proteínas en proporciones similares.

Los proteomas del phylum Bacteroidetes presentan mayor número de proteínas transportadoras y receptoras. A diferencia de *Porphyromonas gingivalis*, *Bacteroides thetaiotaomicron* presentó proteínas redox.

Los phyla Chlorobi y Fibrobacteres incluyen proteínas que transportan y sintetizan sideróforos, así como aquellas asociadas a receptores y procesos redox. La diferencia entre ellos es que *Chlorobium tepidum* destina mayor proporción de proteínas receptoras.

En el phylum Tenericutes observamos que todos los organismos exclusivamente contienen proteínas dedicadas al transporte de sideróforos.

Finalmente, en el phylum Firmicutes se observó la presencia de proteínas que transportan, sintetizan y llevan a cabo procesos redox. Las bacterias *Clostridium tetani* y *Clostridium perfringens* únicamente presentaron proteínas de transporte. *Bacillus subtilis* es la bacteria con más proteínas de síntesis. *Streptococcus agalactiae*, *Streptococcus pyogenes*, *Streptococcus pneumoniae*, *Enterococcus faecalis* y *Lactobacillus johnsonii* sólo destinan proteínas al transporte y la síntesis de sideróforos. El proteoma de *Oceanobacillus iheyensis* presentó la mayor frecuencia de proteínas redox.

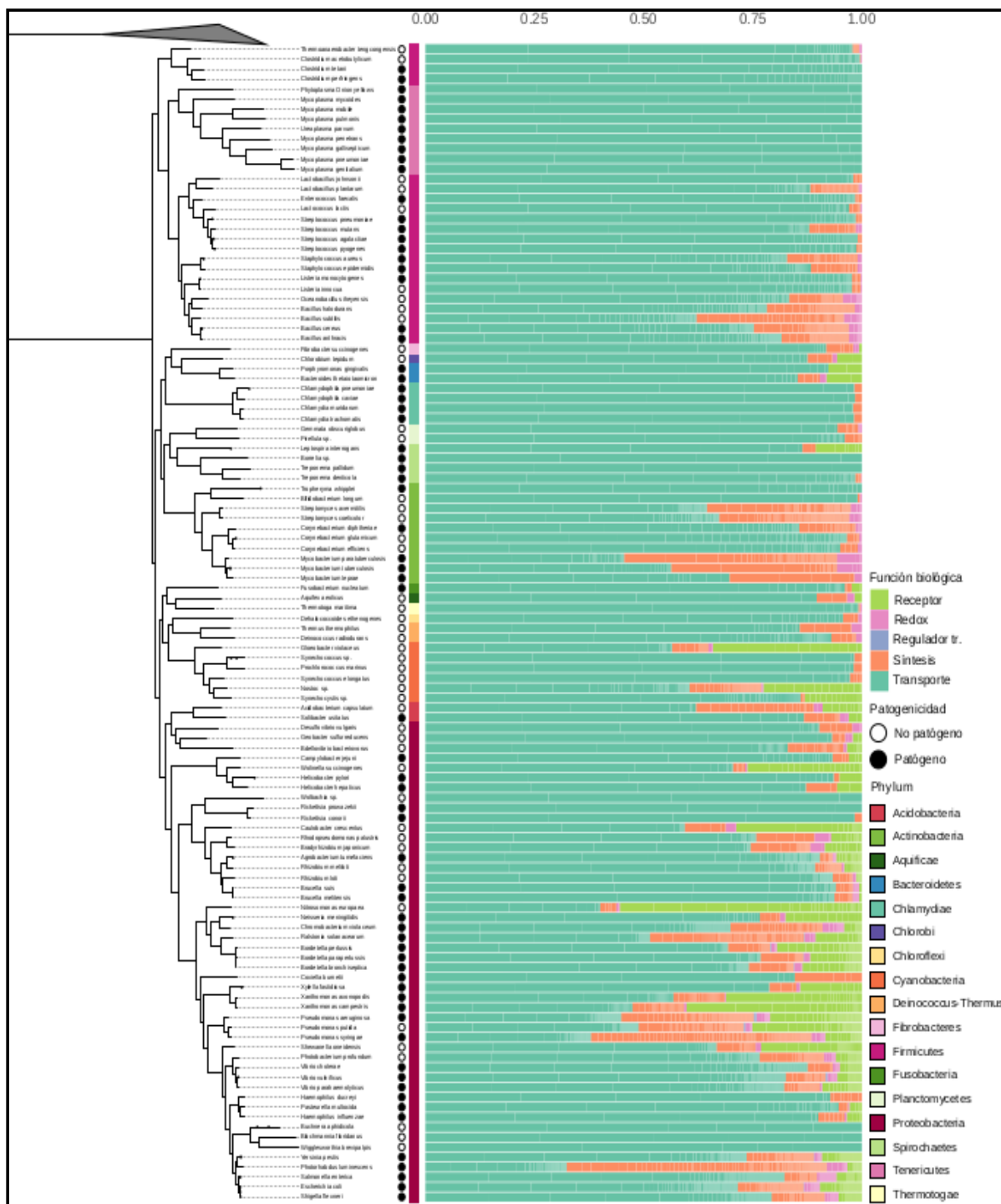


Figura 18. Frecuencia relativa de las proteínas asociadas a la síntesis, regulación transcripcional, receptores, transporte y procesos de reducción-oxidación de sideróforos. Las filas corresponden a 116 proteomas bacterianos ordenados filogenéticamente. Por cada organismo se observa la proporción de proteínas destinadas a cada función biológica. La barra de vertical de colores indica los phyla analizados. Con círculos se señala la patogenicidad de los organismos. Para mayor resolución de la imagen, consultar el enlace: <https://drive.google.com/file/d/1IS-G023XyEpBbDeT38y1o-lu-Nk727FZ/view?usp=sharing>

Discusión

El estudio exhaustivo sobre la adquisición de hierro en bacterias desarrollado previamente por otros grupos de trabajo nos ha permitido conocer inicialmente los aspectos mecanicistas de la síntesis, transporte y regulación de los sideróforos. Esto es porque la secreción de sideróforos es la forma más prevalente de las bacterias para acceder a este metal de transición, permiten su aprovechamiento a partir de reservas ambientales y facilitan su absorción a través de las membranas celulares (Kramer et al., 2020; Matzanke, 2011)□.

Existe un actual interés por explorar la base molecular de los sideróforos, así como sus consecuencias ecológicas y evolutivas. Por lo tanto, el presente trabajo tuvo como objetivo predecir y analizar la variedad de estrategias de adquisición de hierro mediante sideróforos en una selección de proteomas que comprenden la diversidad filogenética de las bacterias (Kramer et al., 2020)□.

El primer objetivo de este estudio fue identificar de manera extensiva a las proteínas asociadas específicamente con sideróforos y que son responsables de la síntesis de estos metabolitos secundarios, el transporte de complejos ferri-sideróforos, así como algunas proteínas encargadas de la protección contra estrés oxidativo.

Para recuperar la diversidad de secuencias de proteínas asociadas a estos procesos celulares se utilizó un enfoque basado en subsistemas, por lo tanto, se empleó la base de datos SEED. Esta base de datos fue desarrollada por la Fellowship for Interpretation of Genomes (FIG), la cual integra una gran cantidad de información genómica de procariontes no redundante, compilada de varias fuentes y organizada de manera modular (Rodríguez-Brito et al., 2006)□. El módulo de mayor jerarquía es el subsistema y está definido por el SEED como un conjunto de funciones relacionadas de productos genéticos (denominados roles funcionales); en otras palabras, un subsistema representa la relación funcional entre enzimas y que forman parte de una misma ruta metabólica (*MG-RAST A Technical Report and Manual for Version 3.3.6-rev1 Mathematics and Computer Science Division*, n.d.)□. Los subsistemas revelan información sobre cómo un particular proceso biológico es implementado en diferentes organismos. Por ello, el SEED *subsystems* contiene subsistemas que cubren la diversidad del metabolismo central e intermedio en muchos

organismos.

La ventaja de usar subsistemas es que permiten a los anotadores identificar proteínas homólogas en genomas de manera simultánea. Por esto, es importante establecer las características esenciales que definen a un subsistema. Durante su construcción, el anotador tiene acceso a una serie de recursos bioinformáticos que le permiten identificar variaciones sobre cómo se implementan los procesos biológicos particulares por parte de diversos organismos, permitiendo mejorar el proceso de anotación, así como reconocer proteínas que antes no se habían considerado y representar de manera más acertada el dinamismo de los procesos bioquímicos en la naturaleza. El SEED ha dado prioridad a la diversidad del metabolismo central de muchos organismos, sin embargo, el esfuerzo por complementarlo no ha finalizado (Overbeek et al., 2005)□.

La selección específica de proteínas a partir de SEED subsistemas permitió construir una base de datos con las principales estrategias empleadas por las bacterias para adquirir hierro del medio extracelular mediante sideróforos. En primera instancia, se pueden distinguir proteínas asociadas a los principales tipos de sideróforos, en función de su estructura química. Estos son los hidroxamatos, catecolatos, carboxilatos y mixtos. La base de datos creada contempla mayor información sobre las proteínas involucradas con los catecolatos, como la enterobactina, además confirma la abundancia de información disponible acerca de este tipo de sideróforos que son los que secretan de manera general las bacterias Gram negativas (Pi et al., 2012)□.

La importancia de la estructura y carga del complejo ferri-sideróforo recae en que influye en su movilidad y, por lo tanto, en su destino en el ambiente en el que se encuentra presente. Por ejemplo, el grupo catecolato libre y los complejos con el grupo α -hidroxicarboxílico tienden a ser fotoreactivos; en cambio, los complejos con catecolatos, el grupo α -hidroxicarboxílico libre, hidroxamatos libres y formando complejos con el hierro, no son fotoquímicamente activos. La degradación de los sideróforos por procesos fotoquímicos tiene como consecuencia la reducción de la constante de estabilidad del hierro, aumentando su biodisponibilidad (Harrington, Duckworth, & Haselwandter, 2015; Schalk & Guillon, 2013)□. Por otra parte, se ha demostrado que los complejos ferri-sideróforos son más resistentes a la degradación microbiana que los sideróforos libres (Winkelmann, 2007)□.

Los sideróforos se sintetizan mediante dos vías, NRPS y NIS. La síntesis a partir de dominios de sintetasas peptídicas no ribosomales (NRPS) es la vía más prevalente y consta de una línea de ensamblaje multienzimática, donde se constituye un péptido precursor a

partir de aminoácidos, grupos carboxi e hidroxiaácidos, para posteriormente ser modificado y formar el sideróforo final. Posteriormente, el sideróforo es liberado al medio extracelular, donde promueve la solubilización del hierro de fuentes biológicamente inaccesibles. Los sideróforos con mayor número de proteínas disponibles en la base de datos construida fueron la pioverdina y la enterobactina. Este tipo de sideróforos son sintetizados mediante enzimas NRPS, las cuales también están involucradas en la producción de muchos otros metabolitos peptídicos secundarios macrocíclicos, como los antibióticos (Batista, De Souza Santos, Ricci-Azevedo, & Da Silva Neto, 2019; Singh, Chaudhary, & Sareen, 2017) □.

Las proteínas relacionadas específicamente con sideróforos, se asociaron principalmente al phylum Proteobacteria. Las proteobacterias son una clase diversa de bacterias predominantemente Gram negativas, que incluye diversas especies de importancia médica (ej. *Neisseria*, *Yersinia*), tecnológica (ej. *Agrobacterium*, *Rhizobium*, *Escherichia*), entre otros (Timmermans, Paudel, & Ross, 2017) □. Si bien las especies clasificadas de bacterias oxidantes de hierro se encuentran en diferentes phyla dentro del dominio Bacteria, incluidas las Nitrospirae y Firmicutes, la mayoría están incluidas en las proteobacterias (Hedrich, Schlömann, & Barrie Johnson, 2011) □.

Las principales variaciones entre los organismos estudiados incluyen la potencial capacidad de síntesis, regulación, elementos de transporte del complejo ferri-sideróforo y funciones alternativas de los sideróforos como lo es la protección contra el estrés oxidativo. Estos principales aspectos se pueden sintetizar en 1,007 familias de proteínas asociadas específicamente a sideróforos y que representan la variabilidad de datos contenidos en SEED.

Las bacterias son capaces de utilizar diferentes grupos de genes biosintéticos que optimizan la biodisponibilidad del hierro, los cuales están regulados diferencialmente a nivel transcripcional en respuesta a las señales ambientales (McRose, Baars, Seyedsayamdost, & Morel, 2018; McRose, Seyedsayamdost, & Morel, 2018; Sharma, Sharma, & Verma, 2018) □. En el conjunto de proteínas específicas de sideróforos, la capacidad de síntesis fue la categoría más relevante. Una bacteria puede sintetizar más de un tipo de sideróforo con diferentes propiedades fisicoquímicas, las cuales determinarán su solubilidad y permanencia en el ambiente. Aunque la síntesis de estos metabolitos representa un costo metabólico alto para las células individuales, puede beneficiar a otros miembros de la comunidad bacteriana con la capacidad de reconocer y aprovechar los complejos ferri-sideróforos (Harrington et al., 2015) □.

La base de datos construida hace distinción entre transporte y receptores de sideróforos debido a que el transporte de estos agentes quelantes difiere entre las bacterias Gram positivas y negativas, esencialmente por la estructura celular. En las bacterias Gram negativas se requiere del reconocimiento específico del complejo ferri-sideróforo por un receptor embebido en la membrana externa que experimenta un cambio conformacional que favorece la translocación del complejo al periplasma. Este proceso es facilitado por el complejo TonB que proporciona la fuerza motriz de protones para la internalización del ión metálico (Cornelis, 2010; Kramer et al., 2020)□. En las bacterias Gram positivas no se requieren receptores de membrana externa y, al igual que en las bacterias Gram negativas, los transportadores de tipo ABC importan el complejo ferri-sideróforo (Kramer et al., 2020)□. En la frecuencia relativa de las proteínas que controlan el transporte de hierro por sideróforos (Figura 18) observamos que los Gram positivos, representados por los phyla Actinobacteria, Deinococcus-Thermus y Firmicutes están ausentes los receptores generalmente distintivos de las bacterias Gram negativas.

La liberación citoplasmática del hierro unido requiere de la actividad enzimática de reductasas y esterasas, que reducen el ión metálico y pueden degradar el sideróforo (Caza, Garénaux, Lépine, & Dozois, 2015; Neumann, Sassone-Corsi, Raffatellu, & Nolan, 2018)□.

Los sideróforos son agentes con alta afinidad que facilitan la biodisponibilidad del hierro, también son mediadores importantes de las interacciones entre los miembros de los conjuntos microbianos y sus huéspedes eucariotas (Kramer et al., 2020)□, por esto es posible observar la presencia de diferentes estrategias para adquirir hierro del medio extracelular en los distintos niveles taxonómicos del conjunto de proteomas analizados.

La mayor variabilidad y abundancia de estrategias de adquisición de hierro las encontramos en el phylum de las Proteobacterias (Figura 18), taxón en el que destacan las *Pseudomonas* y el sideróforo pioverdina. Estos son organismos ubicuos presentes en ambientes diversos como el agua o la rizósfera, con la capacidad de producir una gran variedad de factores de colonización y virulencia asociados a la absorción eficiente de hierro por la bacteria (Cornelis & Dingemans, 2013)□. Sin embargo, la disponibilidad de estudios está sesgada hacia el género *Pseudomonas* y la pioverdina, por ello se necesita de estudios adicionales que verifiquen que la información obtenida de este género es representativa de la biología general de los sideróforos, la filogenia bacteriana y los tipos de hábitat y huésped en los que están presentes (Kramer et al., 2020)□.

Se sabe que una parte sustancial del genoma, cerca del 10%, está involucrado con la

biosíntesis de diversos metabolitos secundarios, los cuales están altamente especializados y frecuentemente son tan diversos como lo son sus estructuras químicas y funciones. Los genes responsables de la producción de estos metabolitos incluyen genes transportadores especiales para la exportación y la absorción de sideróforos, además de los genes que codifican la maquinaria biosintética (Bruns et al., 2018) □. De la proporción de proteínas que participan en la adquisición de hierro en cada organismo, pudimos observar que la mayoría destina al menos el 10% de su proteoma.

El análisis comparativo de las estrategias de adquisición de hierro (Figura 15) también reveló que el organismo *Gemmata obscuriglobus* es el que tienen menor proporción de proteínas para cumplir con el requerimiento de hierro. De esta inusual bacteria, mediante un estudio de anotación del genoma se encontró que posiblemente no sean capaces de producir sideróforos y dependen de la disponibilidad de reductasas extracelulares para obtener sus suministro de hierro. Se han detectado en el duodeno humano, que es un ambiente rico en reductasas y hierro en su forma reducida (Aghnatiós & Drancourt, 2016) □. En comparación con este estudio, pudimos observar también la prioridad de transportar hierro mediante sideróforos, presentando también proteínas que le posibilitan reacciones redox; a diferencia de lo antes expuesto, también se observó la capacidad de síntesis de sideróforos.

Proteomas especializados en la producción de sideróforos

Existe un sesgo hacia los estudios sobre la biosíntesis de ciertos sideróforos porque es atractiva para el diseño de fármacos combinada con diversas aplicaciones terapéuticas, así como los mecanismos de transporte y liberación de hierro por ser objetivos interesantes con respecto al control de patógenos. A pesar de la información disponible, son reducidos los sistemas de sideróforos que han sido bien caracterizados. Por lo que en este estudio se planteó la posibilidad de analizar la biología general de los sideróforos, que trascienden la filogenia bacteriana y su ecología (Kramer et al., 2020; Miethke & Marahiel, 2007).

El análisis de abundancia de proteínas específicamente relacionadas con sideróforos, permitió reconocer a los organismos especializados en este tipo de estrategia para adquirir hierro. *Pseudomonas syringae* dedica 30% de su proteoma a mecanismos de síntesis, receptores, transportadores y proteínas con actividad redox que facilitan el aprovechamiento de este metal de transición. Esta bacteria es un patógeno de plantas que se encuentra principalmente en la superficie de las hojas y tallos. Pertenece a uno de los géneros Gram

negativos más importantes entre las bacterias aerobias cultivables y en el que se ha detectado alta abundancia de receptores de la membrana externa, lo que probablemente indica que la competencia por hierro es una presión de selección importante para este organismo (Varma & Chincholkar, 2007).

Las proteínas involucradas con el transporte del complejo al citoplasma a través de un transportador ABC afín, destacaron entre las estrategias mejor representadas en *Pseudomonas syringae*. Los transportadores de tipo ABC son proteínas que acoplan la hidrólisis de ATP para el intercambio de solutos a través de la membrana celular, además son importantes factores de virulencia porque permiten la absorción de nutrientes, intervienen en la secreción de toxinas y agentes antimicrobianos (Davidson & Chen, 2004; Reimann, 2012). Los transportadores ABC son importantes para agentes patógenos, como *P. syringae*, porque la colonización bacteriana depende de su capacidad para obtener nutrientes esenciales del hospedero. Además, se ha observado que la abundancia de transportadores ABC codificados en los genomas bacterianos se correlaciona con los nichos fisiológicos en los que pueden vivir (Garmory & Titball, 2004; Tanaka, Song, Mason, & Pinkett, 2018).

Una limitante clara es que la investigación actual sobre la producción de sideróforos está sesgada hacia el género *Pseudomonas* y la pioverdina, como el principal sideróforo que producen. Son necesarios sistemas de estudio adicionales que permitan verificar si los conocimientos adquiridos con este género de bacterias son representativos de la biología de los sideróforos (Kramer et al., 2020). Por ello, este trabajo pretendió explorar las estrategias potenciales que las bacterias pueden emplear para la absorción selectiva de hierro, con base en la diversidad de proteínas asociadas a todos los tipos de sideróforos reconocidos por SEED hasta el momento.

La capacidad de síntesis de sideróforos se observa con mayor frecuencia en la bacteria *Photobacterium luminescens* (Figura 18), el cual es un simbiote de nemátodos y un patógeno de insectos de amplio espectro. Al mantener este tipo de interacciones biológicas la adquisición de hierro parece ser de particular importancia. El análisis de su genoma reveló el mayor conjunto conocido de transportadores de hierro, grupos hemo, grupos hemina y sideróforos. También presenta una gran cantidad de genes similares a otras enterobacterias responsables de la biosíntesis de sideróforos, toxinas, factores de virulencia y proteínas de función desconocida. Lo interesante de esta bacteria es la cantidad de elementos genéticos móviles que sugieren que es un genoma con alto grado de plasticidad, sujeto a reordenamientos frecuentes y transferencia continua de genes (Ciche, Blackburn,

Carney, & Ensign, 2003; Duchaud et al., 2003; Watson, Millichap, Joyce, Reynolds, & Clarke, 2010) □. *Photorhabdus luminescens* es otro de los organismos que destacó como bacteria especializada en la producción de sideróforos, ya que dedica aproximadamente 26% de su proteoma a estrategias para el aprovechamiento de hierro, especialmente mediante la síntesis de sideróforos. Esta bacteria tiene un estilo de vida complejo que implica una interacción patógena con larvas de insectos y una interacción mutualista con el nemátodo bacterióforo *Heterorhabditis bacteriophora*. El nemátodo requiere de *P. luminescens* para su crecimiento, reproducción en larvas de insectos y supervivencia en condiciones artificiales (Ciche et al., 2003).

De manera general, la capacidad de síntesis de metabolitos para esta bacteria es la principal estrategia para el establecimiento de la colonización bacteriana, que causa la toxemia, la septicemia y con ello, la muerte del insecto. Las bacterias son letales para las larvas de insectos cuando se encuentran en la hemolinfa, pero no son patógenas cuando son ingeridas; por lo que resulta relevante su capacidad de síntesis de sideróforos de tipo catecolato para limitar la disponibilidad de hierro del huésped. La producción de sideróforos también es importante como determinante de interacciones biológicas, ya que pueden tener actividad antibiótica e inhibir la proliferación de bacterias y hongos competidores. Por lo tanto, el aprovechamiento de hierro es la base de la interacción simbiótica entre *P. luminescens* y *H. bacteriophora* (Ciche et al., 2003; Duchaud et al., 2003; Watson et al., 2010).

P. luminescens es reconocida como la bacteria con el mayor conjunto conocido de transportadores de hierro, asociado a grupos hemo, hemina y sideróforos. A diferencia de otros estudios bioinformáticos, no fue posible detectar proteínas implicadas en el transporte de hierro. Se han identificado al menos 9 receptores dependientes de TonB involucrados en la absorción de hierro y la virulencia de *P. luminescens*. Las proteínas del complejo TonB intervienen en la captación de transferrina, que es un hemóforo y también tiene la capacidad de quelar hierro unido a proteínas que contienen grupos hemo (Duchaud et al., 2003; Watson et al., 2010).

Los hemóforos aprovechan el grupo hemo de los organismos huéspedes por ser una fuente importante de hierro, son funcionalmente similares a los sideróforos y son secretados únicamente por bacterias Gram negativas (Wandersman & Delepelaire, 2004; Wandersman & Stojiljkovic, 2000). Aunque los hemóforos son parte del mecanismo general de adquisición de hierro para muchas bacterias, no fueron considerados en el desarrollo del presente

trabajo. Por lo que resulta congruente no haber identificado proteínas de transporte para la transferrina en *P. luminescens*.

Aunque *P. syringae* y *P. luminescens* se presentaron como organismos especialistas de la producción de proteínas asociadas a sideróforos, las estrategias que potencialmente pueden emplear y la diversidad de sideróforos con los que pueden interactuar son diferentes. Pese a que los principales quelantes de hierro sintetizados por *Pseudomonas* son la pioverdina y la pioquelina (Cornelis & Matthijs, 2007), en este trabajo también observamos el reclutamiento de proteínas asociadas a la síntesis y el aprovechamiento de acromobactina, aerobactina, alcaligina, antraquelina, bacilibactina, enterobactina y yersiniabactina.

En *P. luminescens* se ha estudiado en menor medida la diversidad de sideróforos que es capaz de sintetizar, como la fotobactina y elementos similares a la pioverdina, yersiniabactina y vibriobactina (Ciche et al., 2003; Duchaud et al., 2003; Heermann & Fuchs, 2008; Watson et al., 2010). De hecho, sólo se han descrito sistemas de secreción de sideróforos pocos microorganismos (Aznar & Dellagi, 2015). En la búsqueda de proteínas homólogas propuesta por este trabajo, se sugiere la presencia de proteínas asociadas a la síntesis de acromobactina, aerobactina, alcaligina, antraquelina, bacilibactina, deferoxamina E, enterobactina, pioquelina, pioverdina, estafilobactina, vibrioferrina y yersiniabactina.

Regulación transcripcional de sideróforos en los proteomas bacterianos

Para la utilización de sideróforos, las bacterias tienen que regular estrictamente las proteínas y los sistemas de transporte que permiten detectar la disponibilidad de este metal y regular los sistemas de adquisición. Las familias de proteínas encargadas de regular este proceso únicamente fueron observadas en el phylum Proteobacteria, como caso de mayor frecuencia se tiene a *Pseudomonas aeruginosa* (Figura 15).

P. aeruginosa es una bacteria patógena que tiene la capacidad de prosperar en muchos entornos diferentes, tanto fuera como dentro de un hospedero. Ante la limitación de hierro impuesto por el sistema inmune del huésped, pudo haberlo adquirido a través de una compleja red reguladora capaz de incorporar numerosos estímulos ambientales, asegurando la expresión oportuna de los sistemas individuales de absorción de hierro y los factores de virulencia. Se ha estimado que al menos el 10% de su genoma está dedicado a

estas redes reguladoras, en las que destacan los sistemas de dos componentes y que además se han relacionado con la virulencia, porque pueden contribuir a la supervivencia de las bacterias y a su capacidad de colonización (Francis, Stevenson, & Porter, 2017; Stover et al., 2000)□. Esencialmente, *P. aeruginosa* detecta los complejos ferri-pioverdina, ferri-pioquelina y ferri-xenosideróforo, que desencadenan una cascada de señalización según la naturaleza del sideróforo. Para detectar la presencia del complejo ferri-enterobactina, un sideróforo producido por *E. coli*, *P. aeruginosa* utiliza el sistema de dos componentes PfeS/PfeR (Miethke & Marahiel, 2007)□ y cuya presencia pudo observarse de manera enriquecida en la búsqueda de proteínas homólogas que intervienen con la regulación transcripcional de sideróforos en el proteoma de *P. aeruginosa*. Adicionalmente, se detectaron proteínas homólogas del regulador transcripcional PchR que es una proteína necesaria para la producción del receptor de pioquelina férrica FptA en respuesta a la limitación de hierro y también interviene en la formación de biopelículas, motilidad, citotoxicidad y resistencia a antibióticos (Heinrichs & Poole, 1996; Huang et al., 2019)□.

Además, el genoma de *P. aeruginosa* codifica más de 30 receptores de membrana externa dependientes de TonB, algunos de los cuales están implicados en la absorción de xenosideróforos. Asimismo, posee sistemas de absorción de hierro ferroso independientes de TonB, a través del sistema Feo (Lamont, Beare, Ochsner, Vasil, & Vasil, 2002; Minandri et al., 2016; Reinhart & Oglesby-Sherrouse, 2016)□.

Transporte de sideróforos en los proteomas bacterianos

La familia de transportadores ABC, que moviliza el hierro inorgánico a través de la membrana citoplasmática, es diversa y se ha reconocido que además de transportar metales permiten el paso de azúcares, aminoácidos, proteínas, metales, xenobioticos y otros compuestos. Esto complica la atribución de la especificidad del sustrato a los transportadores de tipo ABC (Hopkinson & Barbeau, 2012)□. Sin embargo, utilizamos con éxito los perfiles de modelos ocultos de Markov para encontrar las proteínas homólogas de los transportadores ABC. Los transportadores de sideróforos se observan de manera abundante en todos los proteomas (Figura 18), siendo los dominios de unión a ATP de estos complejos el componente encontrado de manera ubicua.

Receptores de sideróforos en los proteomas bacterianos

El análisis comparativo de las proteínas receptoras de sideróforos mostró que

Nitrosomonas europaea es el organismo que más da prioridad a este tipo de estrategia, además de las proteínas de transporte. *Nitrosomonas europaea* es un quimiolitoheterótrofo obligado, por lo que destina la mayor parte de su genoma al catabolismo de compuestos inorgánicos. La estrategia que sigue para adquirir hierro del ambiente es elevando la producción de receptores de sideróforos en condiciones limitantes de este metal (Vajrala, Sayavedra-Soto, Bottomley, & Arp, 2011) □. Es una bacteria oxidante que participa en el ciclo biogeoquímico del nitrógeno y los mecanismos para el transporte de hierro son esenciales para mantener los numerosos citocromos y otras proteínas de unión al grupo hemo involucradas en el metabolismo del amoníaco. *N. europaea* también tiene la capacidad de síntesis del sideróforo citrato (Chain et al., 2003) □. En el genoma de *Nitrosomonas europaea* se han identificado 42 receptores TBDT, lo cual podría indicar que esta bacteria puede utilizar un gran número de sideróforos (Schalk & Guillon, 2013) □. Nuestro análisis reconoció 35 familias de proteínas receptoras, asociadas a alcaligina, acromobactina, aerobactina y ferricromo. También observamos en la frecuencia relativa de proteínas de consumo de hierro (Figura 15) la presencia de proteínas redox y de síntesis.

Proteínas redox y adquisición de hierro

El hierro está asociado al estrés oxidativo y es una de las principales causas de la generación ROS a través de las reacciones de Fenton (S. C. Andrews et al., 2003). Por lo tanto, es de suponerse y se ha demostrado que la síntesis de sideróforos también puede estar regulada por el estrés oxidativo además de la disponibilidad de hierro y pueden conferir resistencia contra este tipo de estrés. Por ejemplo, la pioverdina puede absorber la radiación UV de forma extracelular y reducir la generación de ROS, además puede quelar el hierro ferroso que se libera de los clusters de Fe-S dañados, evitando así la producción de radicales hidroxilo dañinos en el interior de las células (Jin et al., 2018b; Kramer et al., 2020; Li et al., 2019) □. Otro ejemplo es la enterobactina, la cual es crucial para el crecimiento de *E. coli* en condiciones de estrés oxidativo, el cual en ocasiones puede ser causado por otro sideróforo como la pioquelina (Peralta et al., 2016) □. En la mayoría de los hábitats microbianos, el ión ferroso se oxida al ión férrico, de manera espontánea al reaccionar con el oxígeno molecular o enzimáticamente durante la asimilación y circulación en los organismos hospedadores (Miethke & Marahiel, 2007) □.

Las familias de proteínas asociadas a reacciones de oxidación/reducción que intervienen en la adquisición de hierro, revelaron estar en mayor frecuencia en organismos del género *Mycobacterium*, las cuales son patógenas y pueden estar expuestas a condiciones variables

y hostiles impuestas por el sistema inmunológico del huésped. Un componente importante de estas condiciones hostiles es el estrés reductor y oxidativo inducido por la respuesta inmune del huésped. Es esencial para estas bacterias estar monitoreando constantemente su microambiente. Son varios aspectos de su fisiología que están regulados por estos sistemas, incluida la virulencia, el metabolismo energético y la latencia. Por lo tanto, el éxito como patógenos puede atribuirse en parte a su capacidad para monitorear, responder y resistir este tipo de microambientes. Un ejemplo de las estrategias que emplea es sintetizar tiorinas de bajo peso molecular para ayudarlo a mantener un entorno reductor (Ashwani Kumar et al., 2011; Pacl, Reddy, Saini, Chinta, & Steyn, 2018)□.

El hierro es un cofactor obligado al menos para 40 enzimas diferentes codificadas en el genoma de *Mycobacterium tuberculosis*. Por su ecología, esta bacteria tiene que enfrentarse no sólo al ambiente reductor-oxidativo impuesto por el huésped, sino también a una limitación del ion metálico debido a su baja solubilidad a pH biológico y la retención de hierro por parte del huésped mamífero. Este patógeno expresa sideróforos específicos (micobactina, carboximicobactina) para quelar hierro insoluble y las proteínas del huésped transferrina, lactoferrina y ferritina. Por lo tanto, la captación de hierro mediada por sideróforos es esencial para la supervivencia de *M. tuberculosis* (De Voss, Rutter, Schroeder, & Barry, 1999; Sritharan, 2016)□.

Consecuencias ecológicas y estrategias de organismos con proteomas enriquecidos en la producción de sideróforos

Se ha planteado que la principal consecuencia física de la secreción de sideróforos es su difusión, lo cual compromete la aptitud de las bacterias porque muchas veces estas proteínas extracelulares son costosas de secretar y se pierden para el individuo productor, convirtiéndose en bienes públicos que benefician a otras bacterias en la comunidad. Por ello, los sideróforos son mucho más que moléculas portadoras de hierro y se consideran como importantes mediadores de interacciones entre bacterias y los huéspedes eucariotas que habitan (Dumas, Ross-Gillespie, & Kümmerli, 2013; Garcia-Garcera & Rocha, 2020b; Kramer et al., 2020; Kümmerli, Schiessl, Waldvogel, Mcneill, & Ackermann, 2014)□.

La capacidad de difundir de los sideróforos también permite que una célula productora no sólo reciba retroalimentación de sus propios sideróforos, sino que puede recibir una señal agregada que comprende moléculas secretadas por toda la comunidad. Este tipo de señalización basada en interacciones sociales complementa el control de la expresión de

los genes que dependen de la biodisponibilidad del hierro en el medio. Por lo tanto, la conmutación de sideróforos puede verse como una forma de toma de decisiones colectiva, mediante la cual surge un cambio coordinado a nivel de comunidad bacteriana como resultado de la retroalimentación positiva y negativa que operan entre los individuos (Dumas et al., 2013; Kramer et al., 2020) □.

Se ha discutido ampliamente que el desarrollo de este tipo de estrategias a lo largo de la evolución de las bacterias ha sido principalmente para adaptarse a las perturbaciones ambientales y nutricionales, lo cual generalmente tiene consecuencias en las interacciones de competencia y cooperación. Lo que ha dado razones para proponer que la producción de diferentes sideróforos podría aumentar la competitividad porque disminuye la probabilidad de que un competidor pueda explotar todo el repertorio de sideróforos de un productor (Jin et al., 2018b; Kramer et al., 2020) □.

Es así que la abundancia de proteínas asociadas a sideróforos de los proteomas bacterianos analizados señala que la diversidad de estrategias están impulsadas principalmente por la ecología de cada organismo, tal como fue señalado anteriormente para *P. syringae* y *P. luminescens*. Se pudo observar una tendencia a dedicar mayor porcentaje del genoma a la síntesis, receptores, transportadores y proteínas con actividad redox en organismos patógenos. Resulta crucial que las bacterias patógenas presenten diversas estrategias para aprovechar el hierro ante las limitante nutricionales impuestas por el huésped u otras bacterias durante las progresión de las infecciones (Aznar & Dellagi, 2015) □.

Otro aspecto importante determina las propiedades químicas y la asimilación de los sideróforos es la estructura del ambiente, así como la diversidad de la comunidad. Se ha reportado que los hábitats altamente estructurados, como el suelo o un hospedero, tienden a favorecer la cooperación y la síntesis de mayor cantidad de proteínas extracelulares como son los sideróforos altamente difusibles; mientras que en hábitats menos estructurados, como el mar, predominan los sideróforos con difusividad reducida (Kümmerli et al., 2014) □.

La asociación entre difusividad y compartimentalización del hábitat se explica con la capacidad de permanencia de un sideróforo cerca de la bacteria productora. La correlación entre la frecuencia de genes que codifican proteínas extracelulares con la estructura del hábitat puede explicar que las bacterias que están asociadas a un huésped presentan frecuencias altas de genes que codifican proteínas extracelulares (García-Garcera & Rocha, 2020b; Kramer et al., 2020; Kümmerli et al., 2014) □. Por ejemplo, las bacterias *P. syringae*,

P. luminescens y *T. denticola* presentan muchas proteínas asociadas a la secreción de sideróforos y están asociadas a un huésped, mientras que las bacterias acuáticas *G. obscuriglobus* y *A. aeolicus* pertenecen al grupo de organismos con menor proporción de estrategias de aprovechamiento de sideróforos.

Proteomas reducidos y captación de hierro

El tamaño del genoma de las bacterias oscila entre 0.6 y 10 Mb. En el extremo más bajo de este rango se encuentran los micoplasmas y bacterias relacionadas, con tamaños de genomas reportados de hasta 530 kb. Otros ejemplos destacados son *Rickettsia prowazekii* y otros patógenos relacionados incluidos en α -Proteobacterias, así como simbioses de insectos que pertenecen a γ -Proteobacterias, como *Buchnera aphidicola* (Moran, 2002)□. Es interesante que en los proteomas más pequeños como *Mycoplasma genitalium*, *Buchnera aphidicola*, *Blochmannia floridanus*, *Ureaplasma parvum* pudo observarse que la totalidad de las proteínas asociadas a sideróforos únicamente cumplen con el mecanismo de transporte. Además, la proporción de proteínas transportadoras se correlaciona mucho más estrechamente con el estilo de vida de las bacterias que con el tamaño del genoma (Charles et al., 2011)□. Cada uno de estos proteomas reducidos con interacciones patógenas o de endosimbiontes, refleja la dependencia a los metabolitos y nutrientes que puede proporcionar el huésped (McCutcheon & Moran, 2012)□. Lo anterior es evidente cuando se analiza la ausencia de proteínas que subyacen a la biosíntesis de sideróforos.

Los organismos con genomas pequeños suelen poseer un subconjunto de genes dedicados a la interacción con el huésped. Estos pueden ser genes necesarios para la proliferación en los tejidos del huésped (patógenos) o genes relacionados con la aptitud del huésped (mutualistas, simbioses). Este conjunto de genes depende esencialmente de la biología de la asociación en particular (McCutcheon & Moran, 2012)□.

Algunos simbioses con genoma reducido sustentan la fisiología de los huéspedes. Por ejemplo, la bacteria intracelular simbiótica obligada *Buchnera aphidicola*, se ha asociado durante más de 150 millones de años con pulgones (Moran, Munson, Baumann, & Ishikawa, 1993)□, proporcionándoles la capacidad de alimentarse exclusivamente del floema de las plantas. *Buchnera* se encuentra en el citosol de las células de estos insectos dentro de vesículas especializadas de tres membranas, llamadas bacteriocitos (Charles et al., 2011). La función de transporte de nutrientes en *Buchnera* se basa en una diversidad muy baja de

complejos de proteínas transportadores, en comparación con las bacterias de vida libre. El número de sistemas transportadores de *Buchnera* varía entre 12 y 34 (Ren, Chen, & Paulsen, 2007)□. En otros análisis bioinformáticos también se ha asegurado que el porcentaje de proteínas de transporte de *Buchnera aphidicola* corresponde al 3.4% (Charles et al., 2011)□. Nosotros estimamos el 3% correspondiendo a tres familias de proteínas transportadoras de tipo ABC. Se sugiere que la función de transporte podría estar asociada con el potencial adaptativo a la vida intracelular, la ecología del hospedero y la especialización de la interacción entre los microorganismos.

Conclusiones

Hemos recuperado todo el conjunto de datos de proteínas para el aprovechamiento de hierro específicamente mediante sideróforos, disponibles en SEED *subsystems*. Además, con revisión de la literatura y análisis bioinformáticos, la diversidad de proteínas de unión a hierro se organizó en 1,007 familias de proteínas, considerando su secuencia de aminoácidos, estructura química, proceso celular en el que intervienen (síntesis, transporte, respuesta al estrés oxidativo, regulación de la transcripción) y biología. Posteriormente, 995 familias de proteínas fueron la base de la búsqueda y reconocimiento de las principales estrategias de aprovechamiento de hierro específicamente vinculadas con sideróforos. En este trabajo se generó una base de datos de 995 pHMMs de proteínas relacionadas con sideróforos como una herramienta bioinformática para analizar proteomas. Validamos esta herramienta contra un conjunto representativo de 116 proteomas bacterianos y demostramos que son capaces de detectar de manera específica proteínas homólogas.

Se hizo un perfil genético para mapear la posibilidad de aprovechar hierro mediante estas proteínas y se reconoció en organismos representativos del dominio Bacteria. Las principales diferencias se observaron en proteínas asociadas a la síntesis, presencia de receptores, transportadores ABC, regulación transcripcional y respuestas al estrés oxidativo. Adicionalmente, fue posible reconocer la proporción del proteoma de cada organismo que potencialmente interviene en la adquisición de hierro y con ello, distinguir a las bacterias especialistas en la producción de sideróforos. Nuestro análisis partió del supuesto de que no habría diferencias en la abundancia de las proteínas relacionadas con sideróforos, sin embargo, encontró patrones que sugieren una distribución discreta de las proteínas y expansiones génicas en algunos organismos, sugiriendo especialización en comunidades de microorganismos para asimilar, captar y regular el hierro. Sugerimos que la diversidad de estrategias para utilizar el hierro extracelular mediante la secreción de sideróforos se correlaciona estrechamente con la biología y ecología de taxa particulares de bacterias.

Perspectivas

Las observaciones realizadas en este trabajo sugieren que los principales elementos involucrados en la síntesis y transporte de sideróforos son probablemente más diversos que los que se han podido describir anteriormente.

De acuerdo con que existen numerosas estrategias que involucran diferentes proteínas para la adquisición de hierro, será importante enriquecer este trabajo con una mayor cantidad de sideróforos, así como incluir la absorción altamente selectiva de hierro mediante hemóforos. También, deberán analizarse con mayor detalle las proteínas que intervienen en la regulación transcripcional, desacoplamiento del hierro en el interior de la célula, almacenamiento de hierro y reciclaje de sideróforos.

Convendría evaluar los resultados con los 116 proteomas representativos en los miles de genomas bacterianos disponibles en las bases de datos como GenBank para explorar a más detalle las hipótesis de la distribución taxonómica específica de las estrategias moleculares de producción, regulación y captación de hierro vía sideróforos.

Agregar un enfoque metagenómico comparativo podría proporcionar información importante para describir mejor la identidad de los sideróforos, su evolución y patrones ecológicos. Es evidente que se requiere de un esfuerzo multidisciplinario para comprender verdaderamente la dinámica de los sideróforos en el medio ambiente.

Material suplementario

<https://docs.google.com/document/d/1thJL1dkORLKyZJhifInjexK1ZA3v2psb4LB6p-ejlv/edit?usp=sharing>

Literatura consultada

- Aghnatos, R., & Drancourt, M. (2016). *Gemmata* species: *Planctomycetes* of medical interest. *Future Microbiology*, 11(5), 659–667. <https://doi.org/10.2217/fmb-2015-0001>
- Andrews, S. C., Robinson, A. K., & Rodríguez-Quñones, F. (2003). Bacterial iron homeostasis. *FEMS Microbiology Reviews*, 27(2–3), 215–237. [https://doi.org/10.1016/S0168-6445\(03\)00055-X](https://doi.org/10.1016/S0168-6445(03)00055-X)
- Andrews, S., Norton, I., Salunkhe, A. S., Goodluck, H., Aly, W. S. M., Mourad-Agha, H., & Cornelis, P. (2013). Control of Iron Metabolism in Bacteria. In *Metal Ions in Life Sciences* (Vol. 12, pp. 203–239). https://doi.org/10.1007/978-94-007-5561-1_7
- Apweiler, R. (2008). The Universal Protein resource (UniProt). *Nucleic Acids Research*, 36(SUPPL. 1), D190–D195. <https://doi.org/10.1093/nar/gkm895>
- Aziz, R. K., Devoid, S., Disz, T., Edwards, R. A., Henry, C. S., Olsen, G. J., ... Xia, F. (2012). SEED Servers: High-Performance Access to the SEED Genomes, Annotations, and Metabolic Models. *PLoS ONE*, 7(10), e48053. <https://doi.org/10.1371/journal.pone.0048053>
- Aznar, A., & Dellagi, A. (2015, June 1). New insights into the role of siderophores as triggers of plant immunity: What can we learn from animals? *Journal of Experimental Botany*. Oxford University Press. <https://doi.org/10.1093/jxb/erv155>
- Barker, D., & Pagel, M. (2005). Predicting Functional Gene Links from Phylogenetic-Statistical Analyses of Whole Genomes. *PLoS Computational Biology*, 1(1), e3. <https://doi.org/10.1371/journal.pcbi.0010003>
- Batista, B. B., De Souza Santos, R. E. R., Ricci-Azevedo, R., & Da Silva Neto, J. F. (2019). Production and uptake of distinct endogenous catecholate-type siderophores are required for iron acquisition and virulence in *Chromobacterium violaceum*. *Infection and Immunity*, 87(12). <https://doi.org/10.1128/IAI.00577-19>
- Baum, Eagon: An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. (n.d.). Retrieved October 9, 2020, from <https://projecteuclid.org/euclid.bams/1183528841>
- Baum, L. E., & Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, 37(6), 1554–1563. <https://doi.org/10.1214/aoms/1177699147>

- Braun, V., & Braun, M. (2002, October 2). Iron transport and signaling in *Escherichia coli*. *FEBS Letters*. John Wiley & Sons, Ltd. [https://doi.org/10.1016/S0014-5793\(02\)03185-X](https://doi.org/10.1016/S0014-5793(02)03185-X)
- Bruns, H., Crüsemann, M., Letzel, A. C., Alanjary, M., McInerney, J. O., Jensen, P. R., ... Ziemert, N. (2018). Function-related replacement of bacterial siderophore pathways. *ISME Journal*, *12*(2), 320–329. <https://doi.org/10.1038/ismej.2017.137>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. <https://doi.org/10.1186/1471-2105-10-421>
- Carroll, C. S., & Moore, M. M. (2018, July 4). Ironing out siderophore biosynthesis: a review of non-ribosomal peptide synthetase (NRPS)-independent siderophore synthetases. *Critical Reviews in Biochemistry and Molecular Biology*. Taylor and Francis Ltd. <https://doi.org/10.1080/10409238.2018.1476449>
- Cartron, M. L., Maddocks, S., Gillingham, P., Craven, C. J., & Andrews, S. C. (2006). Feo - Transport of ferrous iron into bacteria. In *BioMetals* (Vol. 19, pp. 143–157). Springer. <https://doi.org/10.1007/s10534-006-0003-2>
- Caza, M., Garénaux, A., Lépine, F., & Dozois, C. M. (2015). Catecholate siderophore esterases Fes, IroD and IroE are required for salmochelins secretion following utilization, but only IroD contributes to virulence of extra-intestinal pathogenic *Escherichia coli*. *Molecular Microbiology*, *97*(4), 717–732. <https://doi.org/10.1111/mmi.13059>
- Chain, P., Lamerdin, J., Larimer, F., Regala, W., Lao, V., Land, M., ... Arp, D. (2003). Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *Journal of Bacteriology*, *185*(9), 2759–2773. <https://doi.org/10.1128/JB.185.9.2759-2773.2003>
- Challis, G. L. (2005). A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *ChemBioChem*, *6*(4), 601–611. <https://doi.org/10.1002/cbic.200400283>
- Charles, H., Balmand, S., Lamelas, A., Cottret, L., Pérez-Brocal, V., Burdin, B., ... Rahbé, Y. (2011). A genomic reappraisal of symbiotic function in the aphid/buchnera symbiosis: Reduced transporter sets and variable membrane organisations. *PLoS ONE*, *6*(12). <https://doi.org/10.1371/journal.pone.0029096>
- Chatzou, M., Magis, C., Chang, J. M., Kemena, C., Bussotti, G., Erb, I., & Notredame, C. (2016, November 1). Multiple sequence alignment modeling: Methods and applications.

Briefings in Bioinformatics. Oxford University Press.
<https://doi.org/10.1093/BIB/BBV099>

- Choo, K. H., Tong, J. C., & Zhang, L. (2004, May 1). Recent applications of Hidden Markov Models in computational biology. *Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute*. Elsevier. [https://doi.org/10.1016/S1672-0229\(04\)02014-5](https://doi.org/10.1016/S1672-0229(04)02014-5)
- Ciche, T. A., Blackburn, M., Carney, J. R., & Ensign, J. C. (2003). Photobactin: A catechol siderophore produced by *Photobacterium luminescens*, an entomopathogen mutually associated with *Heterorhabditis bacteriophora* NC1 nematodes. *Applied and Environmental Microbiology*, 69(8), 4706–4713. <https://doi.org/10.1128/AEM.69.8.4706-4713.2003>
- Clarke, T., Tari, L., & Vogel, H. (2005). Structural Biology of Bacterial Iron Uptake Systems. *Current Topics in Medicinal Chemistry*, 1(1), 7–30. <https://doi.org/10.2174/1568026013395623>
- Colombo, C., Palumbo, G., He, J. Z., Pinton, R., & Cesco, S. (2014, November 30). Review on iron availability in soil: Interaction of Fe minerals, plants, and microbes. *Journal of Soils and Sediments*. Springer. <https://doi.org/10.1007/s11368-013-0814-z>
- Connorton, J. M., Balk, J., & Rodríguez-Celma, J. (2017). Iron homeostasis in plants-a brief overview. *Metallomics*, 9(7), 813–823. <https://doi.org/10.1039/c7mt00136c>
- Cornelis, P. (2010, May 30). Iron uptake and metabolism in pseudomonads. *Applied Microbiology and Biotechnology*. Springer. <https://doi.org/10.1007/s00253-010-2550-2>
- Cornelis, P., & Dingemans, J. (2013, November 14). *Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections. *Frontiers in Cellular and Infection Microbiology*. Frontiers. <https://doi.org/10.3389/fcimb.2013.00075>
- Cornelis, P., & Matthijs, S. (2007). *Pseudomonas* Siderophores and their Biological Significance. In *Microbial Siderophores* (pp. 193–203). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71160-5_9
- Crosa, J. H., & Walsh, C. T. (2002). Siderophore Genetics, 66(2), 1–27. <https://doi.org/10.1128/MMBR.66.2.223>
- Davidson, A. L., & Chen, J. (2004, June 9). ATP-binding cassette transporters in bacteria. *Annual Review of Biochemistry*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA . <https://doi.org/10.1146/annurev.biochem.73.011303.073626>

- De Voss, J. J., Rutter, K., Schroeder, B. G., & Barry, C. E. (1999, August 1). Iron acquisition and metabolism by mycobacteria. *Journal of Bacteriology*. American Society for Microbiology. <https://doi.org/10.1128/jb.181.15.4443-4451.1999>
- DeJongh, M., Formsma, K., Boillot, P., Gould, J., Rycenga, M., & Best, A. (2007). Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC Bioinformatics*, 8(1), 139. <https://doi.org/10.1186/1471-2105-8-139>
- Dey, G., & Meyer, T. (2015, August 26). Phylogenetic Profiling for Probing the Modular Architecture of the Human Genome. *Cell Systems*. Cell Press. <https://doi.org/10.1016/j.cels.2015.08.006>
- Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., ... Kunst, F. (2003). The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nature Biotechnology*, 21(11), 1307–1313. <https://doi.org/10.1038/nbt886>
- Dumas, Z., Ross-Gillespie, A., & Kümmerli, R. (2013). Switching between apparently redundant iron-uptake mechanisms benefits bacteria in changeable environments. *Proceedings of the Royal Society B: Biological Sciences*, 280(1764). <https://doi.org/10.1098/rspb.2013.1055>
- Eddy, S. R. (1995). *Multiple alignment using hidden Markov models*. Retrieved from www.aai.org
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*. Oxford University Press. <https://doi.org/10.1093/bioinformatics/14.9.755>
- Eddy, S. R. (2018). *HMMER User's Guide Biological sequence analysis using profile hidden Markov models*. Retrieved from <http://hmmer.org>
- Eddy, S. R. (2019). *HMMER User's Guide Biological sequence analysis using profile hidden Markov models*. Retrieved from <http://hmmer.org>
- Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5(1), 113. <https://doi.org/10.1186/1471-2105-5-113>
- Ezawa, K. (2016). General continuous-time Markov model of sequence evolution via insertions/deletions: Are alignment probabilities factorable? *BMC Bioinformatics*, 17(1), 304. <https://doi.org/10.1186/s12859-016-1105-7>
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39(SUPPL. 2), W29. <https://doi.org/10.1093/nar/gkr367>

- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ... Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. <https://doi.org/10.1093/nar/gkv1344>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Funahashi, T., Tanabe, T., Aso, H., Nakao, H., Fujii, Y., Okamoto, K., ... Yamamoto, S. (2003). An iron-regulated gene required for utilization of aerobactin as an exogenous siderophore in *Vibrio parahaemolyticus* Printed in Great Britain, 1217–1225. <https://doi.org/10.1099/mic.0.26066-0>
- Gabaldón, T., & Koonin, E. V. (2013, May 4). Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*. Nature Publishing Group. <https://doi.org/10.1038/nrg3456>
- Garcia-Garcera, M., & Rocha, E. P. C. (2020a). Community diversity and habitat structure shape the repertoire of extracellular proteins in bacteria. *Nature Communications*, *11*(1), 1–11. <https://doi.org/10.1038/s41467-020-14572-x>
- Garcia-Garcera, M., & Rocha, E. P. C. (2020b). Community diversity and habitat structure shape the repertoire of extracellular proteins in bacteria. *Nature Communications*, *11*(1), 1–11. <https://doi.org/10.1038/s41467-020-14572-x>
- Garmory, H. S., & Titball, R. W. (2004, December). ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. *Infection and Immunity*. American Society for Microbiology (ASM). <https://doi.org/10.1128/IAI.72.12.6757-6763.2004>
- Golonka, R., Yeoh, B. S., & Vijay-Kumar, M. (2019, April 1). The Iron Tug-of-War between Bacterial Siderophores and Innate Immunity. *Journal of Innate Immunity*. S. Karger AG. <https://doi.org/10.1159/000494627>
- Gu, S., Yang, T., Shao, Z., Wang, T., Cao, K., Jousset, A., ... Pommier, T. (2020). Siderophore-mediated interactions determine the disease suppressiveness of microbial consortia. *MSystems*, *5*(3). <https://doi.org/10.1128/mSystems.00811-19>
- Hannauer, M., Braud, A., Hoegy, F., Ronot, P., Boos, A., & Schalk, I. J. (2012). The PvdRT-OpmQ efflux pump controls the metal selectivity of the iron uptake pathway mediated by the siderophore pyoverdine in *Pseudomonas aeruginosa*. *Environmental Microbiology*, *14*(7), 1696–1708. <https://doi.org/10.1111/j.1462-2920.2011.02674.x>

- Hardison, R. C. (2003). Comparative Genomics. *PLoS Biology*, 1(2), e58. <https://doi.org/10.1371/journal.pbio.0000058>
- Harrington, J. M., Duckworth, O. W., & Haselwandter, K. (2015). The fate of siderophores: Antagonistic environmental interactions in exudate-mediated micronutrient uptake. *BioMetals*, 28(3), 461–472. <https://doi.org/10.1007/s10534-015-9821-4>
- Hayrapetyan, H., Siezen, R., Abee, T., & Groot, M. N. (2016). Comparative genomics of iron-transporting systems in *Bacillus cereus* strains and impact of iron sources on growth and biofilm formation. *Frontiers in Microbiology*, 7(JUN), 1–13. <https://doi.org/10.3389/fmicb.2016.00842>
- Hedrich, S., Schlömann, M., & Barrie Johnson, D. (2011, June). The iron-oxidizing proteobacteria. *Microbiology*. Microbiology (Reading). <https://doi.org/10.1099/mic.0.045344-0>
- Heermann, R., & Fuchs, T. M. (2008). Comparative analysis of the *Photorhabdus luminescens* and the *Yersinia enterocolitica* genomes: uncovering candidate genes involved in insect pathogenicity. *BMC Genomics*. BioMed Central. <https://doi.org/10.1186/1471-2164-9-40>
- Himpsl, S. D., Pearson, M. M., Arewng, C. J., Nusca, T. D., Sherman, D. H., & Mobley, H. L. T. (2010). Proteobactin and a yersiniabactin-related siderophore mediate iron acquisition in *Proteus mirabilis*. *Molecular Microbiology*, 78(1), 138–157. <https://doi.org/10.1111/j.1365-2958.2010.07317.x>
- Holden, V. I., & Bachman, M. A. (2015). Diverging roles of bacterial siderophores during infection. *Metallomics*, 7(6), 986–995. <https://doi.org/10.1039/c4mt00333k>
- Hopkinson, B. M., & Barbeau, K. A. (2012). Iron transporters in marine prokaryotic genomes and metagenomes. *Environmental Microbiology*, 14(1), 114–128. <https://doi.org/10.1111/j.1462-2920.2011.02539.x>
- Hsueh, K. L., Yu, L. K., Chen, Y. H., Cheng, Y. H., Hsieh, Y. C., Ke, S. chu, ... Huang, T. H. (2013). FeoC from *Klebsiella pneumoniae* contains a [4Fe-4S] cluster. *Journal of Bacteriology*, 195(20), 4726–4734. <https://doi.org/10.1128/JB.00687-13>
- Ilbert, M., & Bonnefoy, V. (2013, February). Insight into the evolution of the iron oxidation pathways. *Biochimica et Biophysica Acta - Bioenergetics*. Biochim Biophys Acta. <https://doi.org/10.1016/j.bbabi.2012.10.001>

- Jin, Z., Li, J., Ni, L., Zhang, R., Xia, A., & Jin, F. (2018a). Conditional privatization of a public siderophore enables *Pseudomonas aeruginosa* to resist cheater invasion. *Nature Communications*, 9(1), 1–11. <https://doi.org/10.1038/s41467-018-03791-y>
- Jin, Z., Li, J., Ni, L., Zhang, R., Xia, A., & Jin, F. (2018b). Conditional privatization of a public siderophore enables *Pseudomonas aeruginosa* to resist cheater invasion. *Nature Communications*, 9(1), 1–11. <https://doi.org/10.1038/s41467-018-03791-y>
- Jothi, R., Przytycka, T. M., & Aravind, L. (2007). Discovering functional linkages and uncharacterized cellular pathways using phylogenetic profile comparisons: A comprehensive assessment. *BMC Bioinformatics*, 8. <https://doi.org/10.1186/1471-2105-8-173>
- Kadi, N., & Challis, G. L. (2009). Chapter 17 Siderophore Biosynthesis. A Substrate Specificity Assay for Nonribosomal Peptide Synthetase-Independent Siderophore Synthetases Involving Trapping of Acyl-Adenylate Intermediates with Hydroxylamine. *Methods in Enzymology*. *Methods Enzymol.* [https://doi.org/10.1016/S0076-6879\(09\)04817-4](https://doi.org/10.1016/S0076-6879(09)04817-4)
- Kalantari, N., & Ghaffari, S. (2008). *EVALUATION OF TOXICITY OF HEAVY METALS FOR ESCHERICHIA COLI GROWTH*. *J. Environ. Health. Sci. Eng* (Vol. 5). Iranian Association of Environmental Health (IAEH). Retrieved from <https://tspace.library.utoronto.ca/handle/1807/62153>
- Kalantari, Narges. (2008). Evaluation of toxicity of iron, chromium and cadmium on *Bacillus cereus* growth. *Iranian Journal of Basic Medical Sciences*, 10(4), 222–228. <https://doi.org/10.22038/IJBMS.2008.5234>
- Kaplan, J., & Ward, D. M. (2013, August 5). The essential nature of iron usage and regulation. *Current Biology*. Cell Press. <https://doi.org/10.1016/j.cub.2013.05.033>
- Kawashima, T. (2018). Comparative and evolutionary genomics. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vol. 1–3, pp. 257–267). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20236-7>
- Köster, W. (2001). ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Research in Microbiology*, 152(3–4), 291–301. [https://doi.org/10.1016/S0923-2508\(01\)01200-1](https://doi.org/10.1016/S0923-2508(01)01200-1)
- Kraemer, S. M. (2004). Iron oxide dissolution and solubility in the presence of siderophores. *Aquatic Sciences*. <https://doi.org/10.1007/s00027-003-0690-5>

- Kramer, J., Özkaya, Ö., & Kümmerli, R. (2020). Bacterial siderophores in community and host interactions. *Nature Reviews Microbiology*, 18(3), 152–163. <https://doi.org/10.1038/s41579-019-0284-4>
- Kumar, Anoop, & Cowen, L. (2009). Augmented training of hidden Markov models to recognize remote homologs via simulated evolution. *Bioinformatics*, 25(13), 1602–1608. <https://doi.org/10.1093/bioinformatics/btp265>
- Kumar, Ashwani, Farhana, A., Guidry, L., Saini, V., Hondalus, M., & Steyn, A. J. C. (2011). Redox homeostasis in mycobacteria: the key to tuberculosis control? *Expert Reviews in Molecular Medicine*. Cambridge University Press. <https://doi.org/10.1017/s1462399411002079>
- Kümmerli, R., Schiessl, K. T., Waldvogel, T., McNeill, K., & Ackermann, M. (2014). Habitat structure and the evolution of diffusible siderophores in bacteria. *Ecology Letters*, 17(12), 1536–1544. <https://doi.org/10.1111/ele.12371>
- Lamb, A. L. (2015, May 30). Breaking a pathogen's iron will: Inhibiting siderophore production as an antimicrobial strategy. *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier. <https://doi.org/10.1016/j.bbapap.2015.05.001>
- Lamont, I. L., Beare, P. A., Ochsner, U., Vasil, A. I., & Vasil, M. L. (2002). Siderophore-mediated signaling regulates virulence factor production in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), 7072–7077. <https://doi.org/10.1073/pnas.092016999>
- Lemos, M. L., & Balado, M. (2020). Iron uptake mechanisms as key virulence factors in bacterial fish pathogens. *Journal of Applied Microbiology*, 129(1), 104–115. <https://doi.org/10.1111/jam.14595>
- Letunic, I., & Bork, P. (2011). Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*, 39(SUPPL. 2). <https://doi.org/10.1093/nar/gkr201>
- Li, C., Zhu, L., Pan, D., Li, S., Xiao, H., Zhang, Z., ... Long, M. (2019). Siderophore-mediated iron acquisition enhances resistance to oxidative and aromatic compound stress in *Cupriavidus necator* JMP134. *Applied and Environmental Microbiology*, 85(1). <https://doi.org/10.1128/AEM.01938-18>
- Lim, C. K., Hassan, K. A., Tetu, S. G., Loper, J. E., & Paulsen, I. T. (2012). The Effect of Iron Limitation on the Transcriptome and Proteome of *Pseudomonas fluorescens* Pf-5. *PLoS ONE*, 7(6), e39139. <https://doi.org/10.1371/journal.pone.0039139>

- Madera, M., & Gough, J. (2002, October 1). A comparison of profile hidden Markov model procedures for remote homology detection. *Nucleic Acids Research*. Oxford University Press. <https://doi.org/10.1093/nar/gkf544>
- Martin, D., Procter, J., Waterhouse, A., Shehata, S., Giang, N., Carstairs, M., ... Barton, G. (2019). *Jalview 2.11 Manual and Introductory Tutorial*.
- Martín, M. J., Herrero, J., Mateos, A., & Dopazo, J. (2003). Comparing bacterial genomes through conservation profiles. *Genome Research*, 13(5), 991–998. <https://doi.org/10.1101/gr.678303>
- Matzanke, B. F. (2011). Iron Transport: Siderophores. *Encyclopedia of Inorganic and Bioinorganic Chemistry*. <https://doi.org/10.1002/9781119951438.eibc0110>
- McCutcheon, J. P., & Moran, N. A. (2012, January 8). Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*. Nature Publishing Group. <https://doi.org/10.1038/nrmicro2670>
- McRose, D. L., Baars, O., Seyedsayamdost, M. R., & Morel, F. M. M. (2018). Quorum sensing and iron regulate a two-for-one siderophore gene cluster in *Vibrio harveyi*. *Proceedings of the National Academy of Sciences of the United States of America*, 115(29), 7581–7586. <https://doi.org/10.1073/pnas.1805791115>
- McRose, D. L., Seyedsayamdost, M. R., & Morel, F. M. M. (2018, October 1). Multiple siderophores: bug or feature? *Journal of Biological Inorganic Chemistry*. Springer Verlag. <https://doi.org/10.1007/s00775-018-1617-x>
- Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7(1), 1–9. <https://doi.org/10.1038/ncomms11257>
- Meyer, F., Overbeek, R., & Rodriguez, A. (2009). FIGfams: Yet another set of protein families. *Nucleic Acids Research*, 37(20), 6643–6654. <https://doi.org/10.1093/nar/gkp698>
- MG-RAST A Technical Report and Manual for Version 3.3.6-rev1 Mathematics and Computer Science Division*. (n.d.). Retrieved from www.anl.gov.
- Miethke, M., & Marahiel, M. A. (2007). Siderophore-Based Iron Acquisition and Pathogen Control. *Microbiology and Molecular Biology Reviews*, 71(3), 413–451. <https://doi.org/10.1128/mubr.00012-07>
- Minandri, F., Imperi, F., Frangipani, E., Bonchi, C., Visaggio, D., Facchini, M., ... Visca, P. (2016). Role of iron uptake systems in *Pseudomonas aeruginosa* virulence and airway

- infection. *Infection and Immunity*, 84(8), 2324–2335. <https://doi.org/10.1128/IAI.00098-16>
- Moran, N. A. (2002, March 8). Microbial minimalism: Genome reduction in bacterial pathogens. *Cell*. Cell Press. [https://doi.org/10.1016/S0092-8674\(02\)00665-7](https://doi.org/10.1016/S0092-8674(02)00665-7)
- Mosbahi, K., Wojnowska, M., Albalat, A., & Walker, D. (2018). Bacterial iron acquisition mediated by outer membrane translocation and cleavage of a host protein. *Proceedings of the National Academy of Sciences*, 115(26), 6840–6845. <https://doi.org/10.1073/pnas.1800672115>
- Nahas, M. El, Kassim, S., & Shikoun, N. (2012). Profile Hidden Markov Model for Detection and Prediction of Hepatitis C Virus Mutation. Retrieved from www.IJCSI.org
- Neumann, W., Sassone-Corsi, M., Raffatellu, M., & Nolan, E. M. (2018). Esterase-Catalyzed Siderophore Hydrolysis Activates an Enterobactin-Ciprofloxacin Conjugate and Confers Targeted Antibacterial Activity. *Journal of the American Chemical Society*, 140(15), 5193–5201. <https://doi.org/10.1021/jacs.8b01042>
- Noinaj, N., Guillier, M., Barnard, T. J., & Buchanan, S. K. (2010). TonB-Dependent Transporters: Regulation, Structure, and Function. *Annual Review of Microbiology*, 64(1), 43–60. <https://doi.org/10.1146/annurev.micro.112408.134247>
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., ... Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33(17), 5691–5702. <https://doi.org/10.1093/nar/gki866>
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., ... Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42(D1), D206. <https://doi.org/10.1093/nar/gkt1226>
- Pacl, H. T., Reddy, V. P., Saini, V., Chinta, K. C., & Steyn, A. J. C. (2018, July 1). Host-pathogen redox dynamics modulate *Mycobacterium tuberculosis* pathogenesis. *Pathogens and Disease*. Oxford University Press. <https://doi.org/10.1093/femspd/fty036>
- Pellegrini, M. (2012). Using phylogenetic profiles to predict functional relationships. *Methods in Molecular Biology*, 804, 167–177. https://doi.org/10.1007/978-1-61779-361-5_9
- Peralta, D. R., Adler, C., Corbalán, N. S., Paz García, E. C., Pomares, M. F., & Vincent, P. A. (2016). Enterobactin as Part of the Oxidative Stress Response Repertoire. *PLOS ONE*, 11(6), e0157799. <https://doi.org/10.1371/journal.pone.0157799>

- Pi, H., Jones, S. A., Mercer, L. E., Meador, J. P., Caughran, J. E., Jordan, L., ... Klebba, P. E. (2012). Role of Catecholate Siderophores in Gram-Negative Bacterial Colonization of the Mouse Gut. *PLoS ONE*, 7(11), e50020. <https://doi.org/10.1371/journal.pone.0050020>
- Pundir, S., Martin, M. J., & O'Donovan, C. (2017). UniProt protein knowledgebase. In *Methods in Molecular Biology* (Vol. 1558, pp. 41–55). Humana Press Inc. https://doi.org/10.1007/978-1-4939-6783-4_2
- Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2), 257–286. <https://doi.org/10.1109/5.18626>
- Raymond, K. N., Dertz, E. A., & Kim, S. S. (2003, April 1). Enterobactin: An archetype for microbial iron transport. *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences. <https://doi.org/10.1073/pnas.0630018100>
- Reimann, C. (2012). Inner-membrane transporters for the siderophores pyochelin in *Pseudomonas aeruginosa* and enantio-pyochelin in *Pseudomonas fluorescens* display different enantioselectivities. *Microbiology*, 158(5), 1317–1324. <https://doi.org/10.1099/mic.0.057430-0>
- Reinhart, A. A., & Oglesby-Sherrouse, A. G. (2016, December 14). Regulation of *Pseudomonas aeruginosa* virulence by distinct iron sources. *Genes*. MDPI AG. <https://doi.org/10.3390/genes7120126>
- Rigden, D. J., & Fernández, X. M. (2019). The 26th annual Nucleic Acids Research database issue and Molecular Biology Database Collection. *Nucleic Acids Research*, 47(D1), D1–D7. <https://doi.org/10.1093/nar/gky1267>
- Rodriguez-Brito, B., Rohwer, F., & Edwards, R. A. (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics*, 7(1), 162. <https://doi.org/10.1186/1471-2105-7-162>
- Schalk, I. J., & Guillon, L. (2013, May 27). Fate of ferrisiderophores after import across bacterial outer membranes: Different iron release strategies are observed in the cytoplasm or periplasm depending on the siderophore pathways. *Amino Acids*. Springer. <https://doi.org/10.1007/s00726-013-1468-2>
- Sharma, A., Sharma, D., & Verma, S. K. (2018). In silico study of Iron, zinc and copper binding proteins of *Pseudomonas syringae* pv. *lapsea*: Emphasis on secreted

metalloproteins. *Frontiers in Microbiology*, 9(JUL), 1838.
<https://doi.org/10.3389/fmicb.2018.01838>

- Singh, M., Chaudhary, S., & Sareen, D. (2017). Non-ribosomal peptide synthetases: Identifying the cryptic gene clusters and decoding the natural product. *Journal of Biosciences*, 42(1), 175–187. <https://doi.org/10.1007/s12038-017-9663-z>
- Sivashankari, S., & Shanmughavel, P. (2007). Comparative genomics - A perspective. *Bioinformatics*, 1(9), 376–378. <https://doi.org/10.6026/97320630001376>
- Sritharan, M. (2016). Iron homeostasis in *Mycobacterium tuberculosis*: Mechanistic insights into siderophore-mediated Iron uptake. *Journal of Bacteriology*. American Society for Microbiology. <https://doi.org/10.1128/JB.00359-16>
- Stojmirović, A., Gertz, E. M., Altschul, S. F., & Yu, Y. K. (2008). The effectiveness of position- and composition-specific gap costs for protein similarity searches. *Bioinformatics*, 24(13), 15–23. <https://doi.org/10.1093/bioinformatics/btn171>
- Tanaka, K. J., Song, S., Mason, K., & Pinkett, H. W. (2018, April 1). Selective substrate uptake: The role of ATP-binding cassette (ABC) importers in pathogenesis. *Biochimica et Biophysica Acta - Biomembranes*. Elsevier B.V. <https://doi.org/10.1016/j.bbamem.2017.08.011>
- Timmermans, M. L., Paudel, Y. P., & Ross, A. C. (2017, August 1). Investigating the biosynthesis of natural products from marine proteobacteria: A survey of molecules and strategies. *Marine Drugs*. MDPI AG. <https://doi.org/10.3390/md15080235>
- Toh, H., Hayashida, H., & Miyata, T. (1983). Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. *Nature*, 305(5937), 827–829. <https://doi.org/10.1038/305827a0>
- Vajrala, N., Sayavedra-Soto, L. A., Bottomley, P. J., & Arp, D. J. (2011). Role of a fur homolog in iron metabolism in *Nitrosomonas europaea*. *BMC Microbiology*, 11(1), 37. <https://doi.org/10.1186/1471-2180-11-37>
- Varma, A., & Chincholkar, S. B. (Eds.). (2007). *Microbial Siderophores* (Vol. 12). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-71160-5>
- Wandersman, C., & Delepelaire, P. (2004). Bacterial iron sources: From siderophores to hemophores. *Annual Review of Microbiology*. Annu Rev Microbiol. <https://doi.org/10.1146/annurev.micro.58.030603.123811>

- Wandersman, C., & Stojiljkovic, I. (2000, April 1). Bacterial heme sources: The role of heme, hemoprotein receptors and hemophores. *Current Opinion in Microbiology*. Current Biology Ltd. [https://doi.org/10.1016/S1369-5274\(00\)00078-3](https://doi.org/10.1016/S1369-5274(00)00078-3)
- Watson, R. J., Millichap, P., Joyce, S. A., Reynolds, S., & Clarke, D. J. (2010). The role of iron uptake in pathogenicity and symbiosis in *Photorhabdus luminescens* TT01. *BMC Microbiology*, 10(1), 177. <https://doi.org/10.1186/1471-2180-10-177>
- Welch, M. D. (2015). Why should cell biologists study microbial pathogens? *Molecular Biology of the Cell*, 26(24), 4295–4301. <https://doi.org/10.1091/mbc.E15-03-0144>
- What are profile hidden Markov models? | Pfam. (n.d.). Retrieved October 15, 2020, from <https://www.ebi.ac.uk/training-beta/online/courses/pfam-creating-protein-families/what-are-profile-hidden-markov-models-hmms/>
- Winkelmann, G. (2007). Ecology of siderophores with special reference to the fungi. In *BioMetals* (Vol. 20, pp. 379–392). Biometals. <https://doi.org/10.1007/s10534-006-9076-1>
- Wooldridge, K. G., Morrissey, J. A., & Williams, P. H. (1992). Transport of ferric-aerobactin into the periplasm and cytoplasm of *Escherichia coli* K12: Role of envelope-associated proteins and effect of endogenous siderophores. *Journal of General Microbiology*, 138(3), 597–603. <https://doi.org/10.1099/00221287-138-3-597>
- Yin, K., Guan, Y., Ma, R., Wei, L., Liu, B., Liu, X., ... Wang, Q. (2018). Critical role for a promoter discriminator in RpoS control of virulence in *Edwardsiella piscicida*. *PLoS Pathogens*, 14(8), e1007272. <https://doi.org/10.1371/journal.ppat.1007272>