



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ESTUDIO EVOLUTIVO DE REDES METABÓLICAS
ANCESTRALES: METABOLISMO DE COENZIMAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

PRESENTA:

INGRID MIRANDA PÉREZ

DIRECTOR DE TESIS:

ARTURO CARLOS II BECERRA BRACHO



CIUDAD UNIVERSITARIA, Cd. Mx., 2023



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.

Datos del jurado

1. Datos del estudiante

Apellido paterno: Miranda
Apellido materno: Pérez
Nombre(s): Ingrid
Teléfono: 5549225262
Universidad: Universidad Nacional
Autónoma de México
Facultad o escuela: Facultad
de Ciencias
Carrera: Biología
Número de cuenta: 417086068

2. Datos del tutor

Grado: Dr.
Nombre(s): Arturo Carlos II
Apellido paterno: Becerra
Apellido materno: Bracho

3. Datos del sinodal 1

Grado: Dr.
Nombre(s): Pablo
Apellido paterno: Padilla
Apellido materno: Longoria

4. Datos del sinodal 2

Grado: Dr.
Nombre(s): Ricardo
Apellido paterno: Hernández
Apellido materno: Morales

5. Datos del sinodal 3

Grado: M. en C.
Nombre(s): Fabiola
Apellido paterno: Ramírez
Apellido materno: Corona

6. Datos del sinodal 4

Grado: Dr.
Nombre(s): Israel
Apellido paterno: Muñoz
Apellido materno: Velasco

7. Datos del trabajo escrito

Título: Estudio evolutivo de redes metabólicas ancestrales: Metabolismo de coenzimas
Número de páginas: 51
Año: 2023

«It's the questions we can't answer that teach us the most. They teach us how to think. If you give a man an answer, all he gains is a little fact. But give him a question and he'll look for his own answers.»

Patrick Rothfuss

*A Joaquín,
el compañero de vida que llegó inesperadamente.*

Agradecimientos

Repetidamente se veía muy lejos el fin de este capítulo de mi vida. Y como siempre en la vida, hubo gente que llegó a acompañarme durante periodos pequeños y se fueron, a las que les agradezco las experiencias compartidas. Pero quiero agradecer especialmente a aquellas que llegaron y permanecieron...

Cronológicamente, a mi familia pequeña (Don Víctor, Doña Juanita, Montserrat y Víctor) por darme un hogar en cualquier rincón donde quiera que se encuentren. Juanita, espero que duermas más tranquila porque por fin lo logramos. A mis familias gigantes y maravillosas, encabezadas por mis pilares mayores (Doña Carlotita, el dúo dinámico de Doña Olivia y Don Pioquinto), a todos los Miranda Barrera y a todos los Pérez Soto, que han formado gran parte de mi carácter y me han dado infinidad de risas y enseñanzas. Nunca dejaré de apreciar la buena suerte que tuve de crecer con ustedes, ni de recordar que coincidimos gracias a todos aquellos que nos antecedieron. Aunque no lo verbalice a menudo, les quiero infinitamente.

A mi familia elegida... Sé que querían un párrafo individual pero es importante para mí que los tres sepan que todo lo que me han enseñado, cuidado y querido también se ve reflejado en este pequeño texto. Espero ser como ustedes cuando crezca y ojalá nunca me falten en mi vida. José María, sigo asombrándome de cómo la vida poco a poco llevó a encontrarnos y de toda la luz que has traído a mi vida porque por más prendas negras que vistas nunca serán suficientes para ocultar tu luz interna. Paloma, el primer día que te conocí nunca imaginé lo importante que te volverías para mí. He aprendido tantas cosas observando cómo cuidas de ti misma y lo valiente que eres por enfrentar la vida cada día. Carlos, es increíble ver cómo hemos crecido juntos desde que nuestro mayor problema eran aquellas interminables tareas de Orgánica y vernos ahora donde la vida misma es lo que pesa. Les agradezco por todas las risas, los secretos, las pláticas banales y profundas, los recordatorios de autocuidado y los sueños compartidos.

A los macacos del Laboratorio de Origen de la Vida les agradezco por todo lo que he aprendido de ustedes y ayudar en mi formación científica. Con mención especial para los becerritos, Abelardo e Hilda, que son mi recordatorio diario de la científica que quiero llegar ser, por ser mi respiro de vida y por hacer más ligero el quehacer diario para cambiar la academia. Si ustedes, no hubiera terminado esto.

Y finalmente, a mis sinodales les agradezco por la paciencia que me tuvieron, los grandes consejos que me dieron y el tiempo que me dedicaron. Arturo, momentáneamente puedes respirar.

Índice general

Agradecimientos	IV
Resumen	1
1. Introducción	2
1.1. Ensamblando metabolismos	2
1.2. Evolución de las coenzimas	5
1.3. Metanogénesis	7
1.4. Coenzima B	11
1.5. Del árbol a la red	13
2. Materiales y procedimiento	16
3. Resultados y discusión	19
3.1. Caso de la homocitrato sintasa (EC 2.3.3.14)	20
4. Conclusiones	25
Apéndice A. Caso de la homoaconitasa(EC 4.2.1.114)	26
A.1. Subunidad grande (AksD)	26
A.2. Subunidad pequeña (AksE)	28
Apéndice B. Caso de la homoisocitrato deshidrogenasa (EC 1.1.1.87)	30
Bibliografía	30

Resumen

Actualmente, el metabolismo comprende una red compuesta de una amplia diversidad de metabolitos, enzimas, reacciones, cofactores inorgánicos y coenzimas. Conocer su historia evolutiva temprana ha sido de gran interés desde mediados del siglo pasado. Una de las rutas metabólicas de interés en el contexto de la evolución temprana de la vida es la metanogénesis. Algo particular de este metabolismo son las coenzimas que están asociadas a las enzimas biosintéticas, las cuales son exclusivas de los organismos metanógenos. Estas coenzimas tienen un papel clave en la generación de metano, sobretudo, las coenzimas M, B y F430 que participan en la reacción final (la reducción del metil-coenzima M por la coenzima B) de la metanogénesis.

En este trabajo se indagó sobre el papel evolutivo de la coenzima B. Para esto, se realizaron redes de similitud de secuencias y se crearon filogenias de las enzimas que participan en la biosíntesis de la coenzima B. Nuestros resultados sugieren que este módulo enzimático parece estar filogenéticamente relacionado con enzimas relacionadas a las vías biosintéticas de lisina, leucina e isoleucina. En conjunto con datos de distribución filogenética, se concluye que la biosíntesis de la coenzima B se derivó de las rutas de aminoácidos antes mencionadas. Estos resultados apoyan la idea de que la metanogénesis no es un metabolismo tan antiguo como algunos autores han sugerido. Comprender los detalles de las biosíntesis de las coenzimas es fundamental para conocer la historia evolutiva del metabolismo metanogénico.

1 Introducción

Los organismos tienen la capacidad de realizar un conjunto acoplado de reacciones químicas (bioquímicas) que les permiten asimilar nutrientes del medio al que se le denomina metabolismo, y se caracteriza por ser una gran red de vías interconectadas responsables de sintetizar (anabolismo) y degradar (catabolismo) biomoléculas. Esta red biológica está compuesta por una extraordinaria diversidad de enzimas, cofactores y metabolitos, y puede estar interconectada a través de diferentes interacciones como las reacciones químicas que se llevan a cabo, las similitudes e interacciones que comparten las enzimas, entre otras. La red que vemos actualmente es producto de miles de millones de años de evolución que han dado forma a la topología de la red y a las propiedades de sus componentes proteínicos. Este papel clave que tiene el metabolismo destaca aún más el interés por saber el origen de las rutas metabólicas que lo integran, y que, desde hace décadas ha sido la pregunta principal de numerosos trabajos [1–6].

1.1. Ensamblando metabolismos

Fueron Aleksandr I. Oparin [7] y John B. S. Haldane [8] quienes desde las publicaciones de sus obras comenzaron una de las controversias que hasta la fecha no se ha podido llegar a un consenso sobre la naturaleza metabólica de las primeras formas de vida. En su obra publicada en 1938, Oparin sugirió un ambiente altamente reductor donde había compuestos de carbono y hierro de origen geológico los cuales reaccionaban con el vapor de agua para formar hidrocarburos. Éstos últimos al oxidarse producirían compuestos como alcoholes, cetonas, aldehídos, etc., que luego reaccionarían con el amoníaco (NH_3) para formar aminas, amidas y sales de amonio. Entre los compuestos proteínicos y otras moléculas resultantes se formaría una solución diluida de compuestos orgánicos que permitiría la formación de sistemas coloidales a partir de los cuales evolucionarían las primeras formas de vida heterótrofas. Por otra parte,

Haldane en 1929 argumentó que el origen de la vida había sido precedido por la síntesis de compuestos orgánicos; también sugirió que la ausencia de oxígeno en una atmósfera primitiva rica en dióxido de carbono (CO_2) dando lugar a una "sopa caliente diluida". A partir de las diferencias entre estas hipótesis surgió un debate histórico sobre la naturaleza metabólica de los primeros seres vivos.

A lo largo del tiempo, se han generado diversas hipótesis sobre el origen y evolución de las rutas metabólicas, dentro de las cuales destacan cuatro modelos principales: el retrógrado [9,10], el progresivo [11], el denominado *patchwork* [12,13]; y el semienzimático [4] **figura 1.1**. Aunque cabe aclarar que el origen de todas las rutas metabólicas pudo no ser de la misma forma y que es posible explicar el surgimiento de algunas de ellas por la combinación de las diferentes hipótesis. Además, todas éstas hipótesis suponen que las rutas metabólicas actuales se derivan de las rutas prebióticas.

Inicialmente, el modelo retrógrado propuesto por Horowitz en 1945 [9]; **figura 1.1-a** sugiere que las vías se fueron ensamblando hacia atrás (de su dirección de acción actual), utilizando intermediarios que se encontraban disponibles en el ambiente prebiótico; por lo que, es un modelo válido sólo si los compuestos prebióticos están disponibles. Más tarde, Horowitz [10] modificó su propuesta al argumentar que la agrupación de genes que codifican las enzimas de una determinada ruta podría explicarse como el resultado de duplicaciones tempranas en tándem de un gen ancestral. Es decir, si la biosíntesis contemporánea del compuesto 'A' requiere la transformación secuencial de los precursores 'D', 'C' y 'B' a través de las enzimas correspondientes, el producto final 'A' de una determinada ruta metabólica fue el primer compuesto utilizado por los organismos heterótrofos primordiales. Esta hipótesis ha recibido numerosas críticas entre las que se encuentran el hecho de que no explica el origen de mecanismos reguladores de las vías catabólicas, ni la existencia de rutas metabólicas con reacciones muy diferentes, y que, si las enzimas que catalizan pasos sucesivos en una determinada vía metabólica son el resultado de una serie de eventos de duplicación de genes, entonces deberían compartir similitudes estructurales [4]. Para pocos casos ha sido una explicación válida.

Granick [11] propuso el modelo progresivo en 1957. Éste se planteó para explicar el desarrollo de las vías biosintéticas en las que los compuestos prebióticos no tienen ningún papel. Según este modelo, las vías se habrían ensamblado en sentido progresivo,

desde precursores simples hasta productos más complejos. Por lo tanto, los genes más antiguos a lo largo del tiempo estarían representados por aquellos que catalizan los primeros pasos de las vías contemporáneas. En la **figura 1.1-b**, la complejidad que se añade al precursor inicial y a los productos intermedios de la vía se ejemplifica con reacciones de “ornamentación” como podría ser la alquilación. Algo necesario para que este modelo pueda ocurrir es que cada uno de los intermediarios sea útil para el organismo, ya que la aparición de múltiples genes simultáneamente en secuencia resultaría demasiado improbable [4]. Granick sugirió que este escenario podría funcionar con el grupo heme y la clorofila.

Con la hipótesis de patchwork, Yčas [13] y Jensen [12] sugirieron que las rutas metabólicas se pudieron haber ensamblado por el reclutamiento de enzimas primitivas con baja especificidad dotadas de un amplio rango de sustratos químicamente relacionados. Como lo ilustra la **figura 1.1-c**, los genes ancestrales que codifican enzimas promiscuas (E1) podrían haber ampliado las capacidades metabólicas de las células primordiales mediante la duplicación génica y la posterior divergencia. Un posible destino que de un evento de duplicación génica es la subfuncionalización (E2) [14, 15], en la que las actividades catalíticas del gen ancestral se dividen entre los parálogos. En nuestro ejemplo ilustrado, tras la divergencia de los genes duplicados, una de las reacciones ancestrales es asumida por uno de los parálogos. Un ejemplo en el que se toma como una explicación válida es la biosíntesis a partir de policétidos y los ciclos de la urea y del ácido tricarbóxico [16–18].

Finalmente, Lazcano y Miller [4] propusieron el modelo semienzimático. En este modelo suponen que: (i) un conjunto de compuestos presentes en el océano primitivo, los cuales deberían de ser estables para que sus concentraciones aumentaran, tal es el caso de la adenina y guanina; (ii) debido a la fuga de las rutas existentes dentro de las células, varios compuestos no tan estables también deberían estar disponibles y ser utilizados rápidamente antes de su degradación como oxaloacetato o α -cetoglutarato; (iii) asimismo, los tipos de enzimas existentes estaban disponibles después de la duplicación de genes, y eran poco específicos; y finalmente, (iv) las enzimas iniciales surgieron a partir de reacciones no-enzimáticas seguidas de la adquisición de la enzima. De esta forma, aunque la mayoría de las reacciones son mediadas por enzimas, también hay algunas reacciones que se producen de forma espontánea o cambiando las condiciones

de la reacción (en caso de no haber enzima). Este es un modelo que está respaldado por numerosos ejemplos [4], y a su vez, apoyan la idea de que las rutas metabólicas que antecedieron a las actuales pudieron existir con un menor número de enzimas, mediante una actividad semi-enzimática, y por la presencia de enzimas generalistas como lo que ocurre en el ensamblaje tipo *patchwork* [1].

Como podemos ver, el origen y evolución de las vías metabólicas representó un paso esencial en la evolución molecular y celular. Los modelos previamente descritos son muestras de aportaciones fundamentales para entender la gran pregunta “¿cómo se originó y evolucionó la red metabólica que observamos hoy en día en los organismos?”. Aunque se tiene un amplio conocimiento para intentar resolver esta incógnita, es una pregunta sumamente compleja que se presta para abordarla desde diferentes ángulos.

1.2. Evolución de las coenzimas

Los organismos contemporáneos llevan a cabo la combinación del metabolismo (una red metabólica que conecta, en términos muy generales, reacciones anabólicas y catabólicas) y la replicación (el proceso de hacer más copias de sí mismo), los cuales son indispensables para que los organismos sobrevivan y puedan heredar su material genético [20]. La relación tan estrecha entre ambos procesos biológicos ha llevado a la comunidad científica a enfrentarse a uno (de muchos) de los problemas tipo huevo-gallina o de causalidad circular (donde se necesita B para que exista A, y viceversa) en el caso de la investigación sobre el origen y evolución temprana de la vida: ¿qué fue primero? ¿la replicación o el metabolismo?, si bien no es materia de debate para este trabajo, este ejemplo es ilustrativo de escenarios similares que se presentan al momento de preguntarnos sobre el origen de otros componentes biológicos presentes en el metabolismo celular como lo son las enzimas y las coenzimas.

En general, los cofactores son actores químicos esenciales del metabolismo y usualmente se dividen en inorgánicos y orgánicos. El término “cofactor” usualmente es asociado a moléculas químicas e inorgánicas, e. g. cationes metálicos divalentes como el magnesio (Mg^{+2}), el manganeso (Mn^{+2}) y el hierro (Fe^{+2} o Fe^{+3}). Las coenzimas consisten de pequeñas moléculas orgánicas no proteicas que se unen específicamente a las

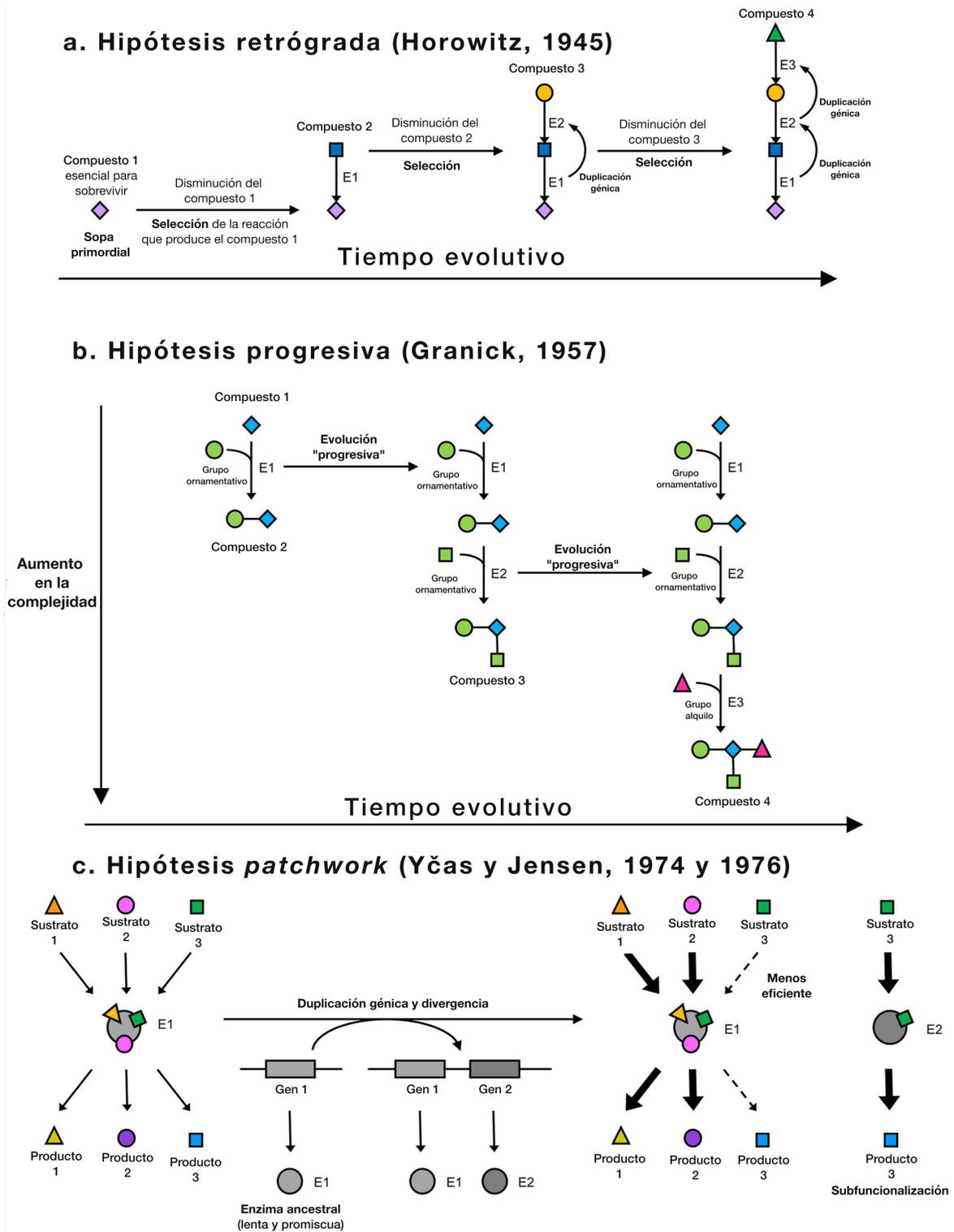


Figura 1.1: Representaciones gráficas de los principales modelos sobre el origen y evolución de las vías metabólicas. En a) podemos observar la hipótesis retrógrada. En b) se muestra la hipótesis progresiva. Y finalmente en c), observamos el resumen gráfico de la hipótesis conocida como *patchwork*. En este esquema no se incluye el modelo semienzimático. Abreviaturas: E1, enzima 1; E2, enzima 2. Modificada de Scossa y Fernie [19].

enzimas y pueden actuar como grupo prostético o como co-sustratos, *e. g.* moléculas como el flavín-adenín dinucleótido (FAD) y el nicotín-adenín dinucleótido fosfato (NADP⁺). Participan activamente promoviendo oxidaciones y reducciones específicas del sitio, reacciones de transferencia de grupos como la acilación, la fosforilación y la metilación [21, 22]. Esta relación es tan íntima e importante al momento de ocurrir la reacción que sin la presencia de la(s) coenzima(s) o cofactor(es) inorgánico(s) correspondientes, la enzima encargada de la transformación química disminuye drásticamente o deja de llevar a cabo la reacción en cuestión. Actualmente observamos que para la biosíntesis de los cofactores orgánicos se requieren de enzimas que les puedan producir.

El estudio evolutivo de estas moléculas inició a mitad del siglo XX [23–27]. A pesar de que durante muchos años las publicaciones con este enfoque fueron escasas, recientemente se han vuelto más frecuentes los trabajos relacionados con su evolución temprana [28–32]. Uno de los hallazgos más interesantes es que las coenzimas FAD, NADP⁺ y la coenzima A (CoA) contienen una parte de monofosfato de adenosina (*asa* de AMP) como elemento estructural [30]. Este hecho se ha utilizado como indicador de que el RNA y las coenzimas, o análogos más simples derivados de ellos, pudieron haber coexistido en la Tierra prebiótica [33]. Además, existen organismos cuyas enzimas dependen de varias coenzimas que no existen en otros organismos, este es el caso de los metanógenos [34–36].

1.3. Metanogénesis

Los metanógenos actuales son microorganismos capaces de producir metano (CH₄) como producto principal de su metabolismo energético. Estos seres generadores de metano comparten varias características: todos los organismos metanógenos conocidos hasta la fecha son parte del dominio Archaea [37]; son anaerobios obligados; son tan sensibles al O₂ que la presencia de trazas de este gas provoca la inhibición del proceso metanogénico temporal [38, 39]; por lo tanto, estos organismos están restringidos a entornos estrictamente anóxicos como humedales, tractos intestinales, y ventilas hidrotermales [40, 41]; son productores obligados de metano y obtienen la mayor parte de su energía para el crecimiento de la producción de CH₄, a pesar de que el rendimiento de ATP sea muy bajo (menos de un equivalente de ATP por cada molécula de metano generada). Además, muchas arqueas metanógenas también son autótrofas, siendo el

CO₂ su única fuente de carbono [34, 42–44].

Geológicamente, se tiene evidencia fósil, así como datos de fechamiento molecular que indican que la metanogénesis se originó tempranamente en la historia de la Tierra en algún punto de principios o mediados del eón Arcaico [45–48], hace aproximadamente unos 3,500 millones de años. También se ha sugerido que ha tenido implicaciones importantes en la evolución del clima de la Tierra [49, 50].

Las primeras reacciones de la metanogénesis y su similitud con la rama metilo de la vía Wood–Ljungdahl (MBWL) ha llevado a varios investigadores a sugerir que los metanógenos pueden representar una de las primeras formas de vida; incluso han propuesto que tanto la metanogénesis como la acetogénesis fueron metabolismos presentes en el Último Ancestro Común (LCA, por su acrónimo en inglés) de todos los seres vivos, y lo ubican como un evento cercano al origen de la vida [51–54]. Sin embargo, un análisis más detallado de esta ruta autotrófica de fijación de carbono nos permite observar que sólo la rama carbonilo (CBWL) está compartida en los dominios Archaea y Bacteria, ya que la rama metilo es diferente tanto en el número de enzimas involucradas, que no son homólogas entre ellas, así como en el número de reacciones metabólicas [52]. Otra gran diferencia radica en el uso de coenzimas, las cuales son específicas de cada una de las variantes de estas rutas metabólicas [42]. Estas diferencias, en conjunto con datos de distribución filogenética, llevó a Muñoz-Velasco y colaboradores a sugerir que la metanogénesis puede ser antigua, mas no ancestral como otros afirman [55].

Hasta hace poco menos de una década, se aceptaba que existían tres formas clásicas de producir CH₄, según el tipo de sustrato y donador de electrones que usan [34, 43, 56]. Siguiendo esta visión tradicional, la mayoría de los organismos metanógenos pueden producir CH₄ mediante la reducción del CO₂ con H₂ o formiato como donadores de electrones (hidrogenotrófica), algunos pueden dismutar compuestos C-1, como metanol o metilamina, que contienen grupos metilo (metilotrófica), y la metanogénesis acetoclástica, en la que el metano se produce a partir de la escisión del acetato [57]. Esta visión tradicional de la producción de metano ha cambiado desde que se han encontrado organismos metanógenos que usan sustratos como alcanos de cadenas más largas [58]. Garcia y colaboradores profundizan en ésta visión ampliada de la metanogénesis y metabolismos relacionados al CH₄ (ver referencia [37]).

Sin importar cómo estas arqueas generen CH_4 , la metil-coenzima M reductasa (Mcr) cataliza el paso final de la formación de metano por lo que es la enzima clave común a todas las variantes metanogénicas. En esta última reacción (**figura 1.2**), la metil-coenzima M ($\text{CH}_3\text{-S-CoM}$) es reducida por el tiol, coenzima B (CoB-SH), para formar el heterodisulfuro (CoM-S-CoB). Esta enzima presenta un grupo prostético tetrapirrólico único, la coenzima F430, que contiene níquel. Aunque la Mcr se considera una enzima clave para la metanogénesis, también puede catalizar la reacción de retroceso (metanotrofia u oxidación anaerobia del metano) si se dispone de un aceptor de electrones con un potencial redox suficientemente positivo [59]. Además, todas las arqueas metanógenas poseen un complejo heterodisulfuro reductasa (MvhADG-HdrABC) con un centro Fe-S que permite reciclar el heterodisulfuro de vuelta a CoB-SH y coenzima M (CoM-SH). Este es uno de los principales pasos de conservación de energía en la ruta, aunque el mecanismo varía dependiendo del tipo de organismo [44, 60].

En cuanto a su distribución filogenética, no ha habido un consenso a lo largo de los años. Tradicionalmente, la metanogénesis fue relacionada únicamente con miembros del *phylum* Euryarchaeota. En 2005, se realizaron análisis filogenéticos basados en secuencias de RNA ribosomal y enzimas clave para la metanogénesis (*e. g.* la Mcr) que agruparon a los metanógenos secuenciados en dos clases y seis órdenes, Clase I (*Methanobacteriales*, *Methanococcales*, y *Methanopyrales*) y Clase II (*Methanosarcinales*, *Methanomicrobiales*, y *Methanocellales*) [61].

Durante la última década, y gracias al uso de la metagenómica, se ha teorizado la presencia de metanogénesis en linajes diferentes y novedosos (en su mayoría por la detección de la Mcr o Mrc-like). En 2015, Evans y colaboradores secuenciaron muestras ambientales y propusieron un nuevo *phylum* para Archaea, “*Candidatus Bathyarchaeota*”, dentro del cual se encontraron miembros con enzimas claves para generar CH_4 [62]. Durante el 2016 [63], se analizaron muestras metagenómicas cuyos análisis filogenéticos llevaron a proponer un nuevo *phylum* “*Ca. Verstraetearchaeota*” dentro del cual también encontraron organismos con enzimas claves de la metanogénesis metilotrófica (específicamente tienen la capacidad de metabolizar metanol, metanotiol y metilaminas). La lista de linajes teóricamente metanógenos sigue creciendo: “*Ca. Methanohydrogenales*”/ “*Ca. Nezhaarchaeota*” [64, 65] y “*Ca. Methanomixophus*” en

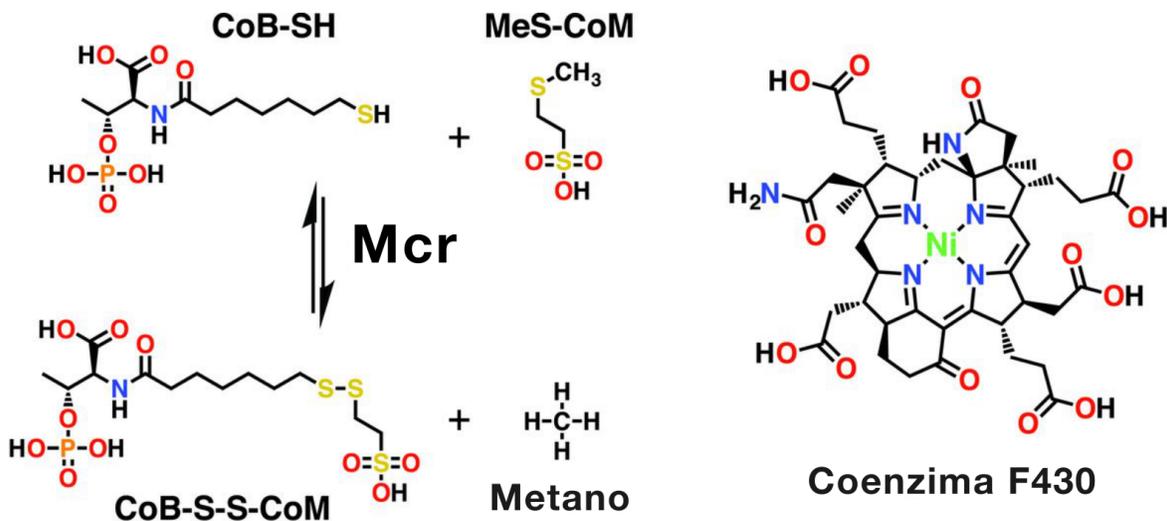


Figura 1.2: Reacción final en cualquier variante de la metanogénesis. Abreviaturas: coenzima B (CoB-SH); metil-coenzima M (MeS-CoM); metil-coenzima M reductasa (Mcr); heterodisulfuro mixto (CoB-S-S-CoM). Modificada de Zheng et al. [76].

Archaeoglobales [65, 66], Methanomassiliicoccales [67], “*Ca. Nuwarchaeales*” [68, 69], “*Ca. Methanofastidiosa*” [70], “*Ca. Methanomethyliales*” [63], “*Ca. Methanodesulfokores*” [68, 71]. Recientemente se descubrió que la alcanotrofia puede estar acoplada a la metanogénesis en “*Ca. Methanoliparia*” [68, 72–74], estos organismos pueden usar alcanos de cadena larga (más de 13 carbonos), al igual que alquilciclohexanos y alquilbencenos. Finalmente, cabe destacar que hay casos especiales en los que ciertos géneros, como *Methanosarcina*, tienen la capacidad metabólica para utilizar una amplia variedad de sustratos como metilaminas, metanol, H_2 y acetato [75].

Dada la importancia de esta ruta, es de interés particular para este trabajo dar luz a la investigación sobre la evolución de la metanogénesis en el contexto de las coenzimas que se necesitan para llevar a cabo las reacciones químicas. A pesar de la diversidad de las coenzimas en la metanogénesis [31, 35, 42], como ya se mencionó anteriormente, la metanogénesis en cualquiera de sus formas necesita en su reacción final de la Mcr y de ciertas coenzimas clave, la coenzima M, B y F430. Sin la presencia de alguna de ellas se interrumpe la biosíntesis de metano [42]. Este trabajo está orientado a evidenciar el papel evolutivo que tienen las coenzimas clave de la metanogénesis, en específico de la coenzima B.

1.4. Coenzima B

El 7-mercaptoheptanoiltreoninafosfato, coenzima B (CoB), es una coenzima que participa en la metanogénesis. Como su nombre lo dice, tiene dos componentes estructurales importantes: el 7-mercaptoheptanoato y la L-treonina. Como lo ilustra la **figura 1.3**, la biosíntesis de la coenzima B sigue repetidamente una secuencia de transformaciones en la elongación de la cadena de α -cetoácidos, involucrando tres reacciones básicas: una condensación aldólica entre un α -cetoácido y acetyl-CoA para formar un homólogo de citrato catalizada por la homocitrato sintasa (AksA, EC 2.3.3.14, [77]), una isomerización del homólogo de citrato a un homólogo de isocitrato catalizada por la homoaconitasa (AksDE, EC 4.2.1.114, [78, 79]), y una descarboxilación oxidativa del homólogo de isocitrato a un α -cetoácido que contiene un grupo metileno adicional, reacción que es catalizada por la homoisocitrato deshidrogenasa dependiente de NAD (AksF, EC 1.1.1.87, [80]). Este ciclo ocurre tres veces, aumentando el número de metileno de dos en el 2-oxoglutarato a tres en el 2-oxoadipato, cuatro en el 2-oxopimelato, y finalmente, cinco en el 2-oxosuberato [81, 82].

El producto del proceso de elongación de la cadena de α -cetoácidos, el 2-oxosuberato, se convierte en un tiol por descarboxilación y sulfonación de forma análoga a la conversión del 3-sulfopiruvato en coenzima M. La forma tiol, 7-mercaptoheptanoato, se acopla entonces a la treonina, formando 7-mercaptoheptanoiltreonina, que es fosforilada para formar la coenzima B [36]. Sin embargo, para estos últimos cuatro pasos no se tienen caracterizadas enzimas que participen en la ruta.

A pesar del desconocimiento de las últimas enzimas que participan en la formación de la CoB, ha resultado de interés el comportamiento químico que tiene la serie de reacciones conocidas por la similitud que comparten con ciertas reacciones del ciclo de ácido cítrico y de la biosíntesis de lisina, leucina e isoleucina [78, 80, 81]. Este comportamiento ha hecho que se estudie más a fondo denominándolo módulo de elongación C1 o módulo C1 (**figura 1.4**). Este módulo actúa de la siguiente forma (**figura 1.4-a**): La acetyl-coenzima A (CoA) dona dos nuevos carbonos (en azul), pero el sustrato α -cetoácido pierde un carbono por descarboxilación, lo que resulta en una ganancia neta de un grupo metileno. La oxidación y la descarboxilación suelen estar catalizadas por una sola enzima, pero esta última también puede producirse de forma no enzimática,

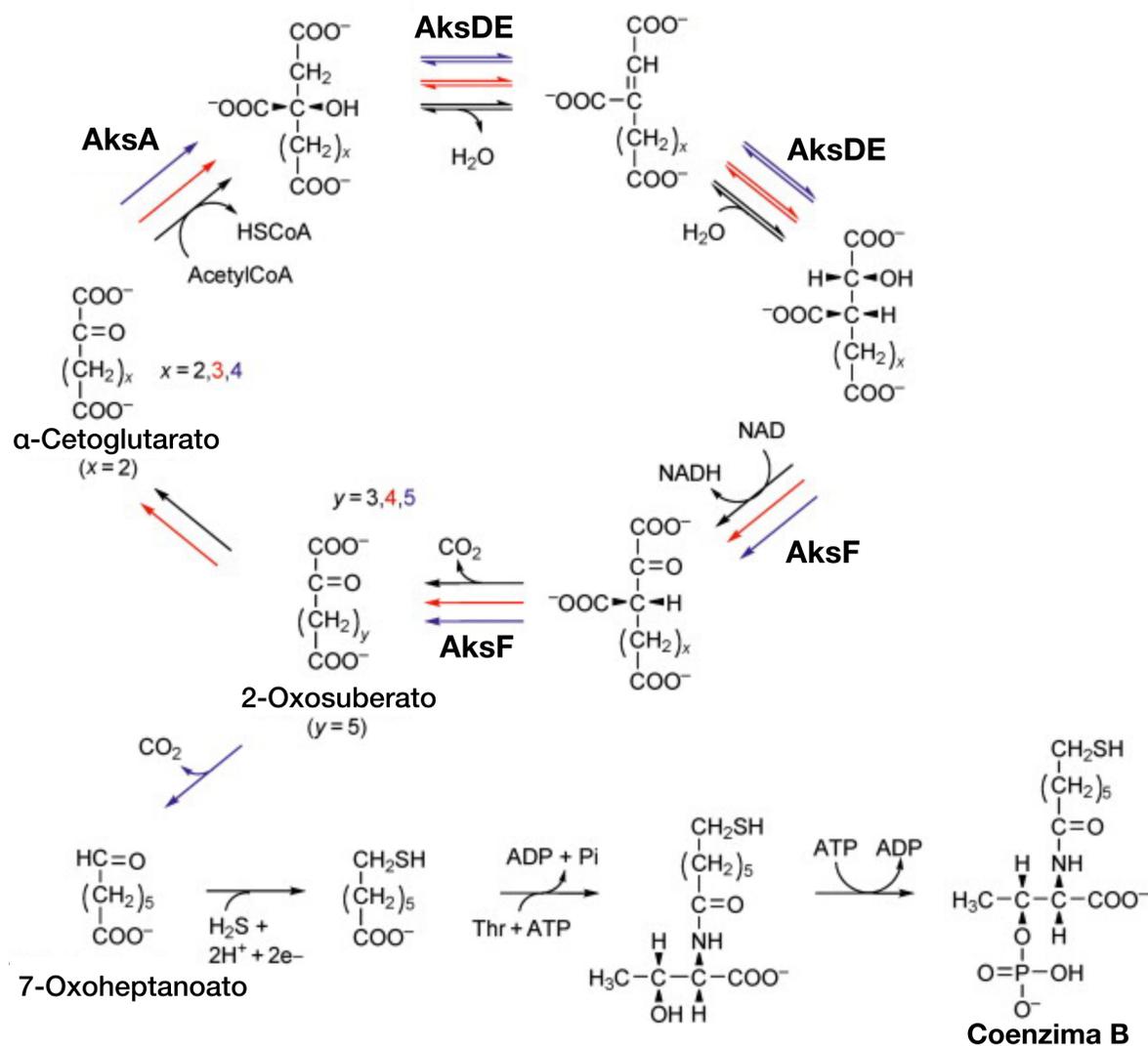


Figura 1.3: Ruta de biosíntesis de la coenzima B. Las reacciones implicadas en la biosíntesis del 2-oxosuberato son el resultado de tres iteraciones de la vía de elongación de 2-oxoácidos que comienza con el α -cetoglutarato. Las flechas negras indican el primer ciclo, mientras que las rojas y azules indican el segundo y el tercero, respectivamente. Imagen modificada de Grochowski y White [35].

por lo que se muestra por separado. Noda-Garcia y colaboradores [5] han sugerido que este fenómeno es debido a que es un módulo de reacción (cascada de dos o más reacciones que son energética y químicamente viables en su conjunto) que constituye un componente evolutivo elemental del metabolismo. Puesto que además de la similitud química, también está bien documentado que las enzimas que actúan como mediadoras del módulo C1, en vías que ahora no están relacionadas, comparten ancestros comunes (**figura 1.4-b**; [51, 68]). Entonces, ¿cómo surgió la biosíntesis de CoB? ¿conocer la historia evolutiva de la CoB nos podría decir algo sobre la evolución de la metanogénesis? En este estudio se pondrá a prueba la hipótesis de que las vías de biosíntesis de coenzimas metanógenas derivaron de rutas preexistentes, por lo tanto, son antiguas pero no ancestrales.

Dada la complejidad de las preguntas y los retos que otros autores han enfrentado al intentar responder alguna de éstas preguntas [5,31,51,54,55,83] usando procedimientos de inferencia filogenéticos o químicos teóricos, el trabajo presente intenta abordar y explicar la evolución temprana de este metabolismo desde una aplicación de teoría de grafos. Esto podría llegar a aportar un enfoque más integrativo para explicar la evolución de la metanogénesis y su papel en la evolución temprana de la vida.

1.5. Del árbol a la red

En 2009, Atkinson y Babbitt describieron a las redes de similitud de secuencias o SSNs [84]. Una SSN es una representación gráfica conformada por nodos, donde cada nodo representa una secuencia, y aristas que los conectan representando similitud de secuencias significativa. Dentro de sus ventajas, destaca que permiten el análisis de conjuntos de datos masivos y son más fiables a medida que aumenta el tamaño de los conjuntos de datos, ya que el mayor número de relaciones por pares resueltas en las redes conduce a una colocación más precisa de las secuencias entre posibles homólogos [84, 85].

Se han usado bastante en el estudio de súperfamilias de proteínas puesto que facilitan la visualización y análisis de las relaciones de similitud entre sus miembros [84, 86–88]. Esto ha permitido la evaluación simultánea de información ortogonal,

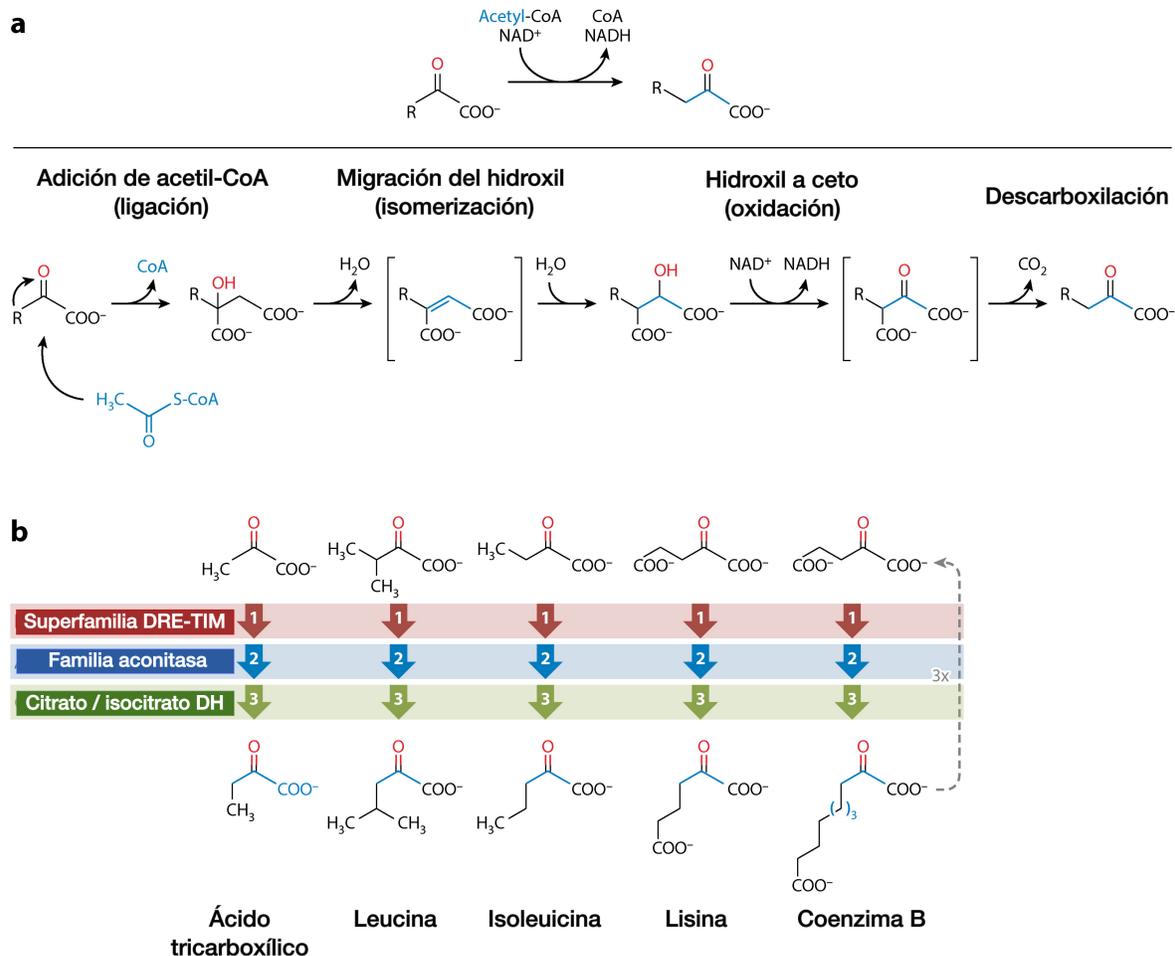


Figura 1.4: (a) El módulo de reacción C1 comienza con un α -cetoácido y termina con un α -cetoácido más largo en un carbono. Los corchetes indican intermediarios que pueden no existir en ciertas especies; por ejemplo, en el ciclo del ácido tricarbóxico (TCA) y la biosíntesis de la lisina. (b) El módulo C1 forma parte de al menos seis vías (de izquierda a derecha): una vía central (el ciclo TCA), tres vías centrales (biosíntesis de leucina, isoleucina y lisina) y dos vías especializadas (biosíntesis de coenzima B, un cofactor redox en metanógenos, y de glucosinolato, un producto natural de plantas). En las dos últimas vías, el módulo C1 se utiliza de forma recursiva (línea discontinua). La primera enzima cataliza la ligación acetil-CoA y pertenece a la superfamilia DRE-TIM. El segundo paso, la isomerización, es catalizado por miembros de la superfamilia aconitasa (a veces por dos enzimas consecutivas). El tercer paso, la descarboxilación oxidativa, está catalizado por miembros de la superfamilia citrato/isocitrato deshidrogenasa (DH). Modificada de Noda-García et al [5].

por ejemplo, la diversidad funcional, cuando se mapean en el contexto de la similitud de secuencias lo que facilita la caracterización de conjuntos de datos mucho más grandes en comparación con los análisis convencionales de alineamiento de secuencias múltiples (MSA) y de árboles filogenéticos. Las conexiones de similitud que proporcionan las SSNs permiten formular hipótesis sobre cómo se relacionan los subgrupos de miembros dentro de una súperfamilia, que normalmente comparten una identidad de secuencia muy baja. En conjunto con otros enfoques, como diferentes tipos de algoritmos de agrupamiento, pueden revelar homologías remotas entre conjuntos de secuencias distantes dentro de la súperfamilia que anteriormente habían permanecido desconocidas [84, 86, 89].

2 Materiales y procedimiento

Se trabajó con las secuencias de las tres enzimas caracterizadas que participan en la biosíntesis de CoB (**Tabla 2.1**). Un resumen del procedimiento que se realizó para cada secuencia *query* se ilustra en la **figura 2.1**.

Tabla 2.1: Información resumida de las enzimas (secuencias) utilizadas como entrada.

Código	Enzima	EC	Reacción	# de aa	UniProtKB
AksA	Homocitrato sintasa metanógena	2.3.3.14 / 2.3.3.-	2-oxoglutarato + acetil-CoA + H ₂ O = (2R)-homocitrato + CoA + H ⁺	406	Q57926
AksD	Homoaconitasa metanógena, subunidad larga		(2R)-homocitrato = (2R,3S)-homoisocitrato	420	Q58409
AksE	Homoaconitasa metanógena, subunidad chica	4.2.1.114	2-oxoglutarato => 2-oxoadipato => 2-oxopimelato => 2-oxosuberato	170	Q58667
AksF	Homoisocitrato deshidrogenasa metanógena	1.1.1.87 / 1.1.1.-	(2R,3S)-homoisocitrato + NAD ⁺ = 2-oxoadipato + CO ₂ + NADH	347	Q58991

La primera columna se refiere al código de la proteína. La segunda columna se refiere a la nomenclatura que reciben las enzimas de acuerdo a las reglas de la International Union of Biochemistry Biochemistry and Molecular Biology. La tercera columna (EC) se refiere a número asignado según el esquema de clasificación numérica para las enzimas de la Enzyme Commission, con base en las reacciones químicas que catalizan. La cuarta columna (Reacción) se refiere a la reacción que catalizan las enzimas. La quinta columna (# de aa) se refiere a la longitud de aminoácidos que constituyen a las enzimas en cuestión.

Finalmente, la sexta columna (UniProtKB) se refiere al identificador de las secuencias que tienen asignados en la base de datos de UniProtKB.

Primero, se localizó la ruta metabólica de interés en la base de datos pública MetaCyc [90] disponible a través de <https://metacyc.org/META/NEW-IMAGE?type=PATHWAY&object=P241-PWY>. Después, se obtuvo la secuencia de aminoácidos de la enzima de interés de la base de datos UniProt [91], proveniente de la arquea metanógena hipertermofílica marina *Methanocaldococcus jannaschii* (Clase Methanococci, [92]) de-

bido a que es el organismo de referencia de la ruta metabólica en la base de datos de MetaCyc.

Para la construcción de redes, se utilizó la colección de herramientas del sitio web Enzyme Function Initiative ó EFI [93] a través de <https://efi.igb.illinois.edu/>. Estas herramientas utilizan redes de similitud de secuencias (SSN) para visualizar y analizar el espacio secuencia-función en familias de proteínas, así como para transferir información entre las herramientas. A través de la herramienta EFI-EST (EFI- Enzyme Similarity Tool; disponible en <https://efi.igb.illinois.edu/efi-est/>) se generó la red de similitud de secuencia o SSN [84]. EFI-EST ofrece cuatro opciones para generar SSNs que difieren según el tipo de datos de entrada. Para este análisis, elegimos la opción A ("Sequence BLAST": Secuencia única). Esta secuencia se utiliza como entrada para una búsqueda BLAST en la base de datos UniProt; las secuencias recuperadas se utilizan para generar la SSN. De tal modo que, utilizando las secuencias de entrada recolectadas se generaron cuatro SSNs, una por cada secuencia (*Taxonomy categories: Archaea; UniProt BLAST query e-value = 1x10-5; sequence DB: UniProt: 2022-01 / InterPro: 88; SSN edge e-value = 1x10-5; SSN threshold = 35*).

Posteriormente, esta SSN se visualizó y analizó en Cytoscape v3.9.1 ([94]; disponible en <https://cytoscape.org/>). Además, se utilizó la aplicación clusterMaker [95] para realizar una agrupación de transitividad (TransClust, [96]). Este algoritmo hace uso de una propiedad de las redes llamada transitividad. Un grafo se denomina transitivo si, para cada triplete de nodos, se cumple la siguiente condición: siempre que veamos una arista del nodo A al B y una arista del B al C, necesitamos también una arista del nodo A al C. Este algoritmo se propuso como una aproximación para cualquier tipo de conjunto de datos en el que se disponga de una función de similitud por pares como para la detección de superfamilias de proteínas, detección de homología de proteínas con valores incompletos y agrupaciones de datos de expresión génica [97–100].

La sección de filogenias inicia a partir de la obtención de la lista de identificadores que EFI-EST ofrece, en conjunto con la red, después se obtuvieron las secuencias de cada nodo a través de UniProt [91]. Con estas secuencias se realizó un MSA con ayuda de MAFFT [101, 102] y TrimAl [103], para filtrar partes pobremente alineadas y con gaps grandes. Las opciones de MAFFT fue *-auto* y para trimAl se usó la opción

-automated1 para generar los alineamientos. Las filogenias de máxima verosimilitud fueron construidas usando IQ-Tree v1.6.12 [104]. Para la selección del modelo evolutivo se obtuvo agregando el comando *-m TEST*, que eligió como ideal al modelo *LG+R9*. También se obtuvieron valores de soporte (*bootstrap*) de mil repeticiones agregando el comando *-bb 1000* mediante el algoritmo Ultrafast Bootstrap [105, 106]. Finalmente, también se realizó la prueba de razón de verosimilitud aproximada tipo SH [107] de mil replicas con el comando *-alrt 1000*.

La visualización, anotación y generación de imágenes de las filogenias se realizó con el software iTOL v6.6 [108] disponible en línea a través de <https://itol.embl.de/>. La anotación de las filogenias se realizó de manera automática gracias a la información y datos que obtuvimos con la red. Una vez anotadas las filogenias se lograron generar las imágenes desde el portal web de iTOL con la opción *Export*.

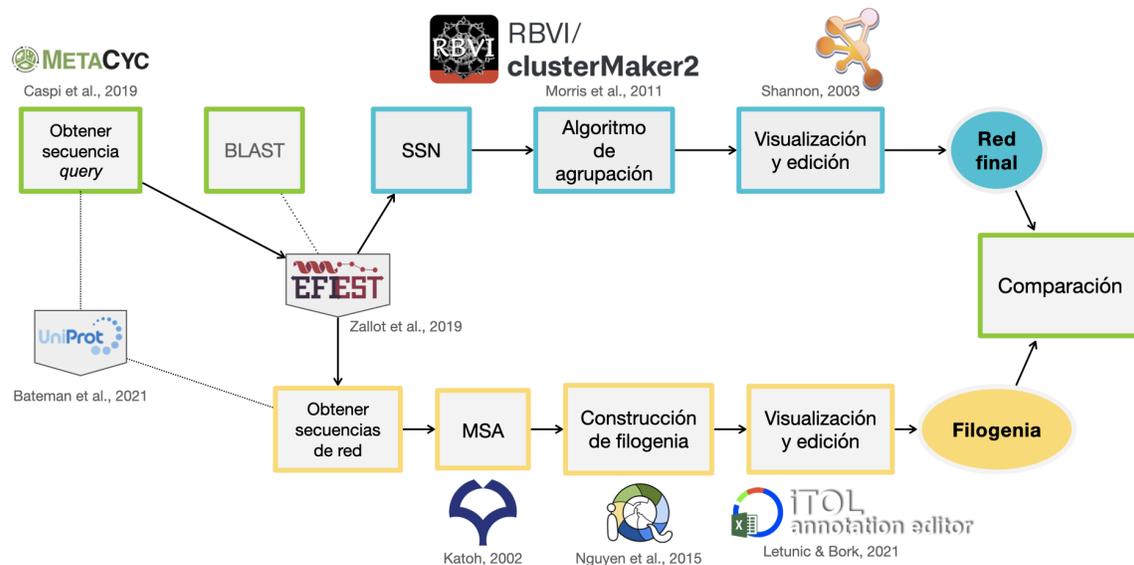


Figura 2.1: Resumen de los análisis que se realizaron para cada una de las secuencias de entrada.

3 Resultados y discusión

A continuación, se muestran los resultados obtenidos de las enzimas biosintéticas de la coenzima B. Para fines prácticos se detallará el caso de la homocitrato sintasa, y los resultados de la homoaconitasa y sus subunidades, y de la homoisocitrato deshidrogenasa se muestran en los anexos correspondientes (**apéndices A y B**).

Debido a que hay muchos términos compartidos entre la filogenia molecular y el modelo de redes, cuando se hable específicamente de la red se usarán los términos nodo (para cada vértice) y *cluster* (para cada grupo de nodos resultado del algoritmo de agrupación). Mientras que cuando se refiera específicamente a la filogenia, para la cual, se usarán términos como OTU (*Operational Taxonomic Unit*, para cada nodo terminal) y clados (para cada grupo que comparte un ancestro hipotético).

En las SSNs, los nodos representan dos tipos de información: la forma que representa el número EC de la secuencia y los diferentes colores de relleno que varían según el *phylum*. Las aristas, además de representar relaciones de homología, tienen su longitud en proporción al valor de identidad que comparten. Con respecto a los *clusters*, todas las redes siguen un orden descendente de acuerdo al número de nodos que los integran.

Con respecto a las filogenias, los colores de las OTUs están asignados según el *phylum*. Además, las filogenias están adornadas con tres caracteres: (i) el número de *cluster* al que pertenece en la red, (ii) el número EC, y (iii) si el organismo al que le pertenece la secuencia tiene algún reporte experimental o predicción de la presencia de algún tipo de metanogénesis. Los números EC que se muestran corresponden a las clasificaciones más frecuentes que se encontraron para cada secuencia. La presencia de alguna variante de metanogénesis está limitada a las variantes clásicas: hidrogenotrófica (cuadrado verde), metilotrófica (círculo azul) y acetoclástica (triángulo morado); y la presencia experimental está señalada con la figura rellena y la predicción está señalada

sólo con el contorno de la figura. Además, el color de las ramas está asignado según el rango de valores de bootstrap como se indica en la leyenda. Estas filogenias no tienen raíz, lo que muestra las relaciones relativas entre las OTUs.

3.1. Caso de la homocitrato sintasa (EC 2.3.3.14)

La **figura 3.1** muestra la red de similitud de secuencias agrupada relacionada con la primera enzima de la biosíntesis de la coenzima B. Originalmente, resultaron 977 nodos enlazados a través de 476,776 aristas, y posterior a la agrupación la red resultó conformada por los 977 nodos previos enlazados por medio de 63,290 aristas, agrupados entre 25 grupos y 7 *singletons*. Con relación a los *phyla* encontrados, de estos 977 nodos: 5 pertenecen a *Ca. Altiarchaeota*, 74 a *Ca. Bathyarchaeota*, 1 a *Ca. Helarchaeota*, 2 a *Ca. Hydrothermarchaeota*, 6 a *Ca. Korarchaeota*, 1 a *Ca. Nezharchaeota*, 32 a *Ca. Thermoplasmata*, 1 a *Ca. Thorarchaeota*, 26 a *Ca. Verstraetearchaeota*, 72 a *Crenarchaeota*, 620 a *Euryarchaeota*, 120 a *Thaumarchaeota* y 17 nodos sin taxonomía asignada. Sin embargo, se encontraron nodos anotados con diferentes números EC aunque todas las secuencias corresponden a actividades de la sub-subclase de las aciltransferasas que convierten los grupos acilo en grupos alquilo: 8 de 2.3.3.-, 365 a 2.3.3.13 (2-isopropilmalato sintasa), 41 a 2.3.3.14 (homocitrato sintasa), 78 a 2.3.3.21 (*(R)*-citramalato sintasa), y 485 de EC desconocido (NA en la figura).

Los *clusters* más grandes conjuntan a nodos pertenecientes a diferentes EC, *e. g.* el grupo A.2 con nodos de EC variados (2.3.3.13, 2.3.3.14, 2.3.3.21 y desconocido). Los grupos en donde se encontraron secuencias clasificadas con el EC metanogénico están sombreados con amarillo (A.2, A.9 y A.19) y en azul donde se encuentra la secuencia de entrada (A.3). La mayoría de los nodos de estos *clusters* metanogénicos son secuencias que pertenecen a *Euryarchaeota*. El *cluster* A.19 es el único que se compone de homocitrato sintasas que pertenecen a organismos de los *phyla* *Crenarchaeota*, *Ca. Verstraetearchaeota* y *Ca. Bathyarchaeota*. Este último resultado apoya la predicción de que estos *phyla* son capaces de realizar metanogénesis que otros grupos de investigación realizaron con base en análisis bioinformáticos [63, 64].

Por otro lado, la filogenia de la homocitrato sintasa (**figura 3.1-ii**; disponible electrónicamente a través de <https://itol.embl.de/export/18921720297266191663249080>)

muestra que en gran medida los *clusters* mantienen su integridad en los clados que se formaron; *e. g.* el *cluster* A.1 se mantiene casi igual excepto por la adición del singleton A.31. Es importante enfatizar que esta filogenia también revela que los grupos de las secuencias con el número EC metanogénico están agrupadas en clados monofiléticos diferentes (A.2, A.3, A.9 y A.19). Esto implicaría que además de ser muy similares en su estructura primaria (de acuerdo con la información de la SSN), esta similitud puede explicarse con ayuda de la filogenia al compartir ancestros comunes hipotéticos.

De los últimos clados mencionados, el clado de las OTUs A.9 está más alejado de los otros e incluye OTUs que pertenecen a los *clusters* A.15 y A. 28. El clado que incluye a todos los miembros del *cluster* A.3 también incluye a nodos de los *clusters* A.22 y A.30; es un clado que incluye a secuencias clasificadas como 2-isopropilmalato sintasa, homocitrato sintasa y (*R*)-citramalato sintasa de organismos del *phylum* Euryarchaeota. Este comportamiento también se muestra en el clado que incluye a las OTUs de los *clusters* A.2 y A.19. Esto resulta ser interesante debido a que respalda lo que visualizamos en la red: nodos con diferente actividad enzimática que resultan no ser tan diferentes. Otra interpretación a este hecho que sea la misma enzima que debido a una baja especificidad sea capaz de utilizar varios sustratos.

La presencia de secuencias de enzimas como la 2-isopropilmalato sintasa o (*R*)-citramalato sintasa, las cuales participan en la biosíntesis de leucina e isoleucina [109] respectivamente, es esperada puesto que previamente se reportó que podía existir una relación evolutiva entre estas dos rutas. En los casos de la homoaconitasa y la homoisocitrato deshidrogenasa los resultados son similares. Para la homoaconitasa, la representación de secuencias de enzimas implicadas en el metabolismo de éstos aminoácidos (*e. g.*, la 3-isopropilmalato deshidratasa que es una enzima implicada en la biosíntesis de leucina) supera en gran medida a aquellas anotadas con la actividad metanógena. En el caso de la homoisocitrato deshidrogenasa también sucede lo mismo con la presencia de secuencias de la enzima 3-isopropilmalato deshidrogenasa (también involucrada en la biosíntesis de leucina). Las filogenias correspondientes a cada una de éstas enzimas reafirman su cercanía evolutiva.

En conjunto con las distribuciones taxonómicas de la homocitrato sintasa, la homoaconitasa y la homoisocitrato sintasa indican que son caracteres derivados de la

3-isopropilmalato sintasa, 3-isopropilmalato deshidratasa y 3-isopropilmalato deshidrogenasa cuyas distribuciones no se restringen sólo a organismos metanógenos.

Sin embargo, cabe destacar que la presencia de la homocitrato sintasa, o de las demás enzimas analizadas, puede estar sobreestimada tomando en cuenta sólo el número EC puesto que la vía de biosíntesis de lisina se compone también de una homocitrato sintasa con el mismo número EC pero cuya distribución taxonómica es más amplia e incluye no sólo a arqueas metanógenas sino también a algunas bacterias y hongos [77]. Este podría ser el caso del *cluster* A.19 en el que encontramos homocitrato sintasas del *phylum* Crenarchaeota. A pesar de esto, no cambia nuestras conclusiones, sino que refleja el comportamiento evolutivo que permite el surgimiento de nuevas rutas metabólicas a través del “reciclaje” de bloques moleculares preexistentes como lo pueden explicar las diferentes hipótesis de surgimiento de rutas metabólicas. Por lo tanto, en términos relativos, los datos sugieren que la biosíntesis de coenzima B surgió después de las rutas de biosíntesis de los aminoácidos lisina, leucina e isoleucina en la historia evolutiva de la vida.

Comprender los detalles de las biosíntesis de las coenzimas también aporta conocimientos sobre los orígenes químicos del metabolismo metanogénico. Nuestros resultados apoyan la idea de que la metanogénesis no es un metabolismo tan antiguo [55] como han postulado por algunos grupos de investigación [51–54]. Sin embargo, aún deben ser extrapolados con cautela puesto que este estudio está limitado a una de las coenzimas clave de la metanogénesis y sus enzimas caracterizadas. Se sugiere realizar un estudio que incluya, al menos, a las otras coenzimas clave para poder tener un mejor soporte a las ideas sobre la antigüedad de la metanogénesis.

En general, se observa que las SSNs obtenidas recapitulan mucha de la información presente en las filogenias. Es evidente que las redes incorporan una serie de ventajas prácticas que mejoran las técnicas actuales para relacionar secuencias y, por último, la visualización de las redes de similitud permite percibir tendencias a partir del contexto de la similitud de secuencias. Esta utilidad de las redes resalta más cuando se quiere analizar conjuntos de datos masivos. Sin embargo, las redes de similitud de secuencias, por su naturaleza, no proporcionan una polaridad de la evolución, sino que representan las relaciones simultáneas de todas las proteínas que cumplen un determinado valor de

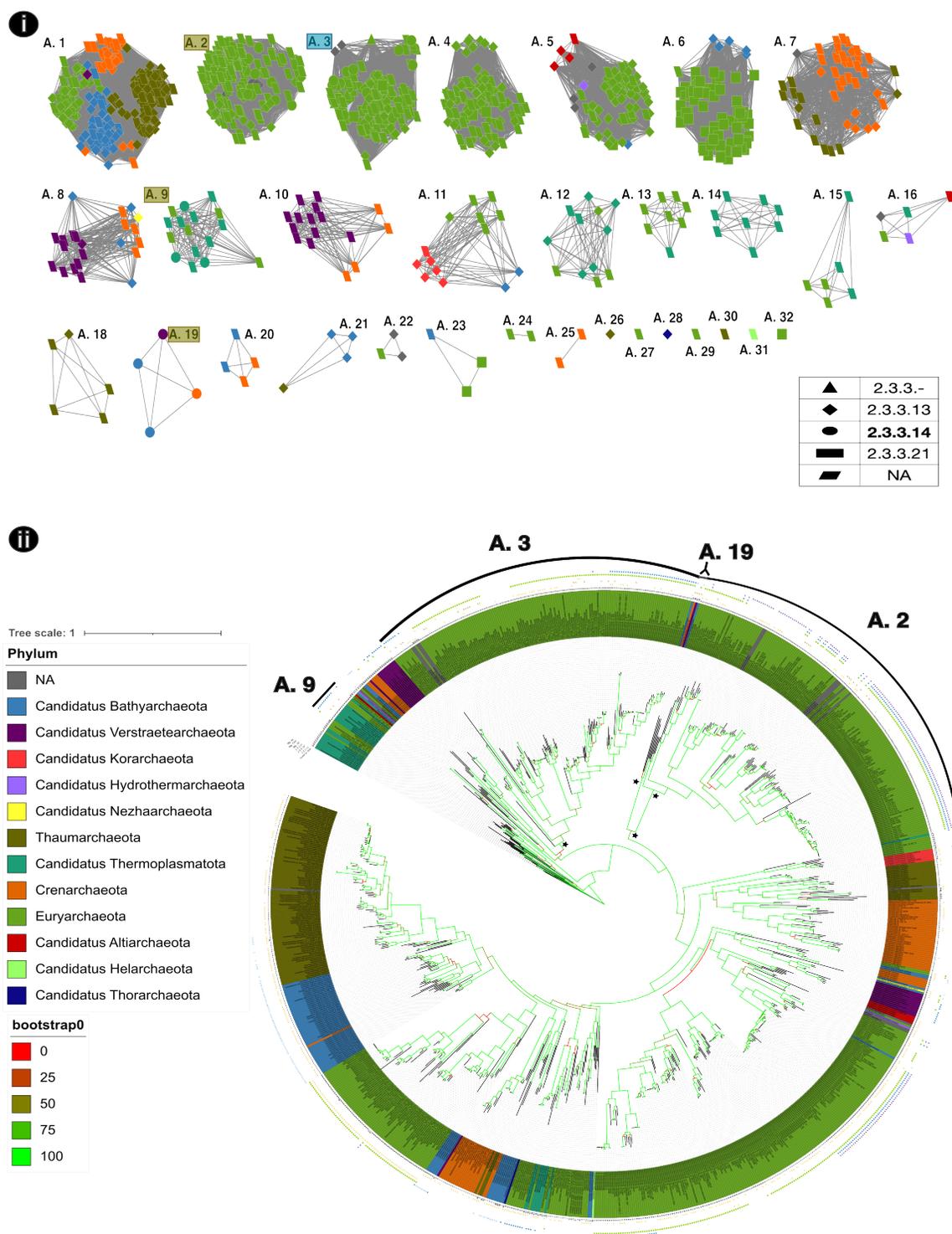


Figura 3.1: SSN y filogenia de la homocitrato sintasa metanógena. **i**) En la SSN, los nodos tienen dos atributos representados: el *phylum* al que pertenecen (los colores están señalados en la leyenda inferior) y la clasificación EC a la que pertenece la secuencia (denotada según la figura de la leyenda de la parte derecha). **ii**) En la filogenia, los colores de las OTUs están dados por la misma leyenda que en la SSN; además incluye el número de clúster al que pertenece en la red, el tipo de actividad enzimática que realiza, si y el organismo al que pertenece tiene la capacidad metabólica de realizar metanogénesis. “bootstrap0” se refiere al nombre del conjunto de datos.

corte de BLAST. Por esto, y para fines de este trabajo, fue fundamental el análisis filogenético al permitirnos dar polaridad evolutiva a la red. El potencial de usar estas redes aún está sin explotar puesto que se desconocen en gran medida el significado biológico y evolutivo de sus propiedades topológicas (en comparación con modelos clásicos como las redes aleatorias, de mundo pequeño y sin escala) pero iniciar usándolas con estos enfoques cambia la perspectiva y aporta mucha comprensión sobre el fenómeno como lo muestra el trabajo presente.

4 Conclusiones

Como muestran los resultados de este trabajo, las enzimas involucradas en la biosíntesis de coenzima B están evolutivamente relacionadas con las rutas biosintéticas de los aminoácidos lisina, leucina e isoleucina. Dada la distribución y filogenias de las enzimas que catalizan estas rutas podemos dar una temporalidad relativa al surgimiento de la biosíntesis de esta coenzima que participa en la metanogénesis, *i. e.*, la ruta de biosíntesis de coenzima B es una vía derivada de las vías metabólicas preexistentes, como pueden ser las vías de biosíntesis de lisina, leucina e isoleucina. Se recomienda hacer un estudio más detallado donde se analicen las demás coenzimas esenciales que participan en esta reacción final de la generación de metano, la coenzima M y F430, lo que podría permitirnos dar una explicación más robusta sobre la evolución temprana de la metanogénesis.

Esta investigación también resalta la importancia de conocer otros aspectos centrales de las vías metabólicas, además de metabolitos y enzimas principales, como el papel de las coenzimas y algunos otros cofactores necesarios lo que permitirá mejorar nuestras explicaciones sobre la evolución del metabolismo.

Además, se refuerza la necesidad de encontrar nuevas técnicas que permitan un análisis rápido y exhaustivo de grandes conjuntos de datos, especialmente para acceder al contexto biológicamente útil que puede extraerse de esta información. Una técnica útil es el uso de redes de similitud de secuencias que permiten visualizar fácilmente las relaciones entre secuencias de interés.

A Caso de la homoaconitasa(EC 4.2.1.114)

A.1. Subunidad grande (AksD)

La **figura A.1-i** muestra la red de similitud de secuencias agrupada. Inicialmente la red resultó compuesta por 800 nodos enlazados a través de 319,600 aristas. Posterior a la agrupación, el número de aristas se redujo a 77,982; se formaron 21 grupos y 2 *singletons*.

Con relación a los grupos taxonómicos, se encontraron 11 phyla; estos 800 nodos pertenecen: 11 a *Ca. Altiarchaeota*, 126 a *Ca. Bathyarchaeota*, 1 a *Ca. Geothermarchaeota*, 1 a *Ca. Odinarchaeota*, 35 a *Ca. Thermoplasmatota*, 1 a *Ca. Thorarchaeota*, 31 a *Ca. Verstratearchaeota*, 2 a *Ca. Woesearchaeota*, 13 a *Crenarchaeota*, 564 a *Euryarchaeota*, 5 a *Thaumarchaeota* y 10 nodos sin taxonomía asignada. Se encontraron nodos anotados con diferentes números EC pero todas coinciden con ser hidrolisas: 7 de 4.2.1.114 que corresponde a la homoaconitasa metanógena, 1 a 4.2.1.31 que agrupa a las maleato hidratatasas, 785 a 4.2.1.33 que clasifican a la actividad de las 3-isopropilmalato deshidratatasas, 2 a 4.2.1.35 que serían (*R*)-2-metilmalato deshidratatasas y 5 de EC desconocido (NA en la figura).

En la **figura A.1- ii**, disponible también en el siguiente hipervínculo, se muestra la filogenia de la subunidad grande de la homoaconitasa metanógena en la que se representa que los clados que agrupan a las secuencias de la actividad metanógena son monofiléticos e incluyen a secuencias de 3-isopropilmalato sintasas del mismo organismo.

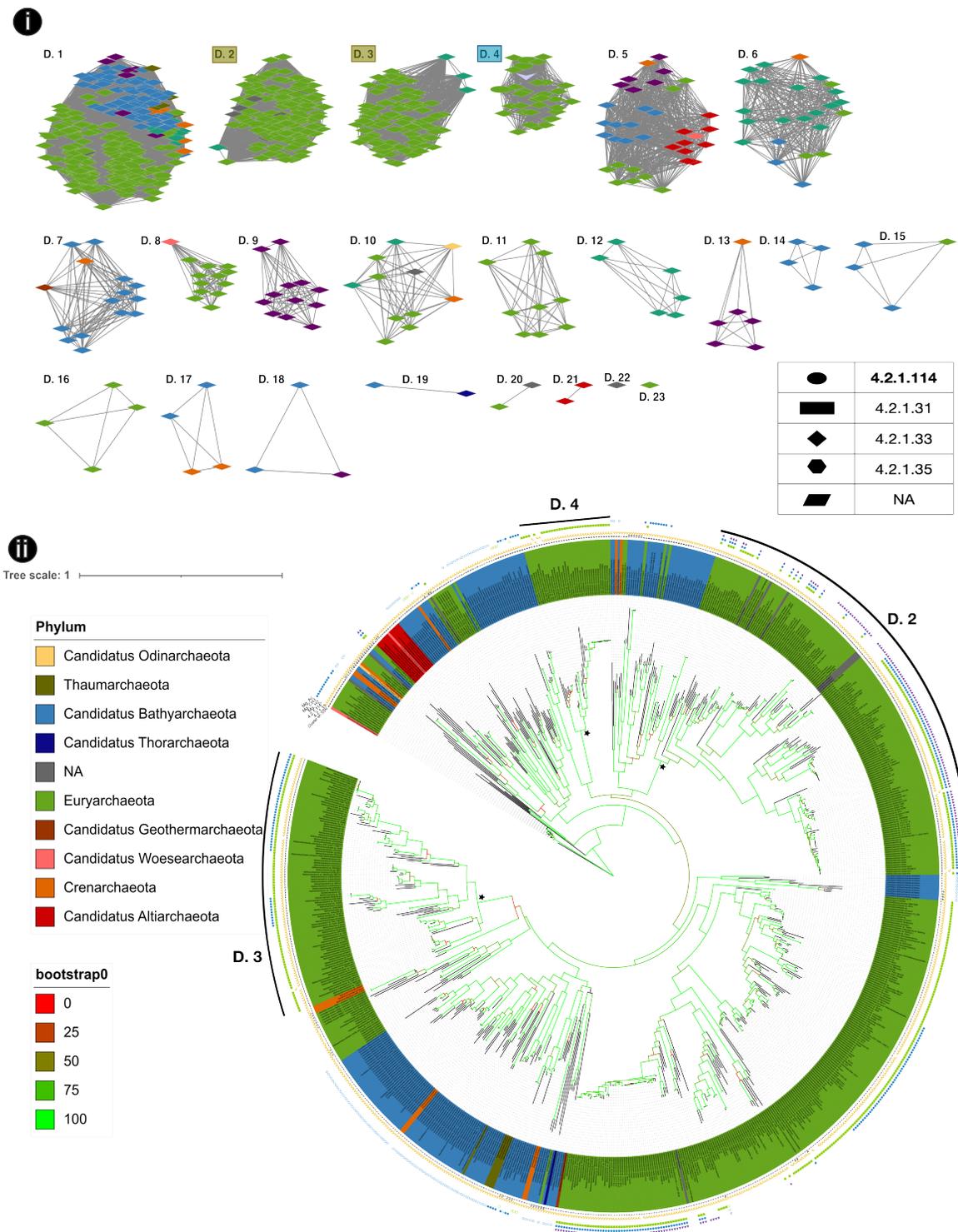


Figura A.1: SSN y filogenia de la subunidad grande de la homoaconitasa metanógena. **i)** En la SSN, los nodos tienen dos atributos representados: el phylum al que pertenecen (los colores están señalados en la leyenda inferior) y la clasificación EC a la que pertenece la secuencia (denotada según las formas diferentes del nodo mostradas en la leyenda de la parte derecha). **ii)** En la filogenia, los colores de las OTUs están dados por la misma leyenda que en la SSN; además incluye el número de clúster al que pertenece en la red, el tipo de actividad enzimática que realiza, si y el organismo al que pertenece tiene la capacidad metabólica de realizar metanogénesis. “bootstrap0” se refiere al nombre del conjunto de datos.

A.2. Subunidad pequeña (AksE)

La **figura A.2-i** muestra la red de similitud de secuencias agrupada. Previa a la agrupación, resultaron 274 nodos enlazados a través de 37,231 aristas. Posterior a la agrupación, la red (Figura 8-i) resultó conformada por los 274 nodos enlazados por medio de 5,281 aristas, agrupados entre 23 grupos y 4 *singletons*.

Con relación a los grupos taxonómicos, se encontraron 11 phyla; estos 274 nodos pertenecen: 4 a *Ca. Altiarchaeota*, 18 a *Ca. Bathyarchaeota*, 2 a *Ca. Geothermarchaeota*, 1 a *Ca. Korarchaeota*, 2 a *Ca. Nezhaarchaeota*, 29 a *Ca. Thermoplasmatota*, 12 a *Ca. Verstratearchaeota*, 6 a *Crenarchaeota*, 188 a *Euryarchaeota*, 7 a *Thaumarchaeota* y 5 nodos sin taxonomía asignada. Por otro lado, se encontraron nodos anotados con diferentes números EC: 131 de 4.2.1.114 (homoaconitasa), 1 a 4.2.1.31 (maleato hidrataza), 137 a 4.2.1.33 (3-isopropilmalato deshidratasa), 2 a 4.2.1.35 (*(R)*-2-matilmalato deshidratasa) y 3 de EC desconocido (NA en la figura). Cabe resaltar que el número EC de la homoaconitasa metanógena es el 4.2.1.114.

La filogenia de la subunidad pequeña de la homoaconitasa metanógena (**figura A.2-ii**, disponible también en <https://itol.embl.de/export/18921720297301511663545075>). Al igual que el caso de la subunidad grande, los clados que agrupan a las secuencias de la actividad metanógena son monofiléticos e incluyen a secuencias de 3-isopropilmalato sintasas del mismo organismo.

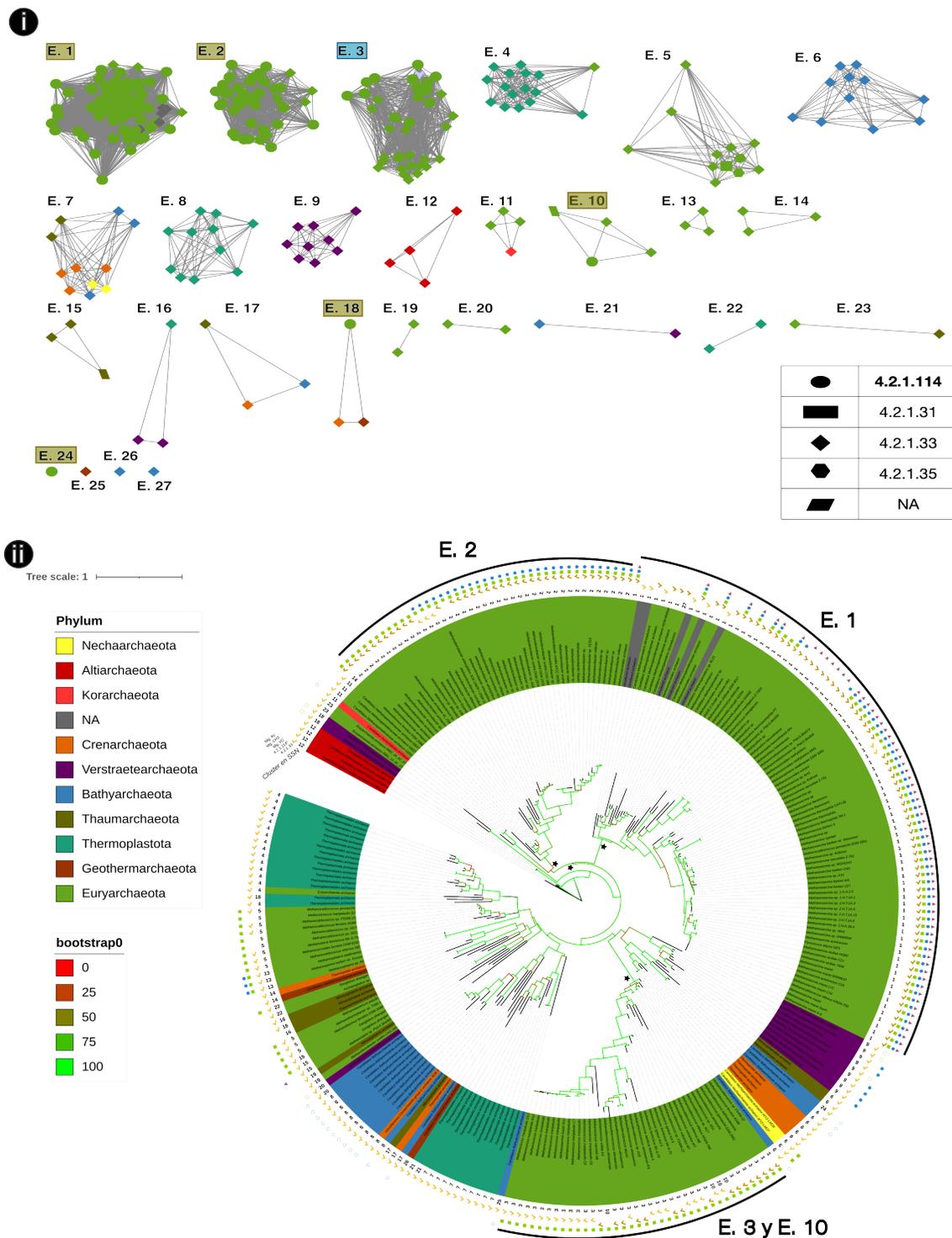


Figura A.2: SSN y filogenia de la subunidad pequeña de la homoacnitas metanógena. **i**) En la SSN, los nodos tienen dos atributos representados: el phylum al que pertenecen (los colores están señalados en la leyenda inferior) y la clasificación EC a la que pertenece la secuencia (denotada según las formas diferentes del nodo mostradas en la leyenda de la parte derecha). **ii**) En la filogenia, los colores de las OTUs están dados por la misma leyenda que en la SSN; además incluye el número de clúster al que pertenece en la red, el tipo de actividad enzimática que realiza, si y el organismo al que pertenece tiene la capacidad metabólica de realizar metanogénesis. “bootstrap0” se refiere al nombre del conjunto de datos.

B Caso de la homoisocitrato deshidrogenasa (EC 1.1.1.87)

Previo a la agrupación, resultaron 535 nodos enlazados a través de 142,845 aristas. Posterior a la agrupación, la red (**figura B.1-i**) resultó conformada por 535 nodos enlazados por medio de 32,795 aristas, agrupados entre 16 grupos y 4 *singletons*.

Con relación a los phyla encontrados, de estos 535 nodos: 1 pertenecen a *Ca.* Altiarchaeota, 54 a *Ca.* Bathyarchaeota, 1 a *Ca.* Geothermarchaeota, 1 a *Ca.* Korarchaeota, 2 a *Ca.* Micrarchaeota, 2 a *Ca.* Nezharchaeota, 13 a *Ca.* Thermoplasmatota, 1 a *Ca.* Undinarchaeota, 17 a *Ca.* Verstratearchaeota, 3 a *Ca.* Woesearchaeota, 10 a Crenarchaeota, 385 a Euryarchaeota, 36 a Thaumarchaeota y 9 nodos sin taxonomía asignada. Por otro lado, se encontraron nodos anotados con diferentes números EC, todas son alcohol deshidrogenasas (oxidorreductasas actuando sobre grupos HC-OH como donadores y con NAD⁺ o NADP⁺ como aceptores): 3 de 1.1.1.-, 34 a 1.1.1.41 (isocitrato deshidrogenasa dependiente de NAD⁺), 1 a 1.1.1.42 (isocitrato deshidrogenasa dependiente de NADP⁺), 8 a 1.1.1.83 (D-malato deshidrogenasa), 86 a 1.1.1.85 (3-isopropilmalato deshidrogenasa), 53 a 1.1.1.87 (homoisocitrato deshidrogenasa), y 350 de EC desconocido (NA en la figura). El número EC de la homoisocitrato deshidrogenasa metanógena es el 1.1.1.87.

La filogenia de la homoisocitrato deshidrogenasa metanógena (**figura B.1-ii**, disponible también en <https://itol.embl.de/export/18921720297301691663545087>).

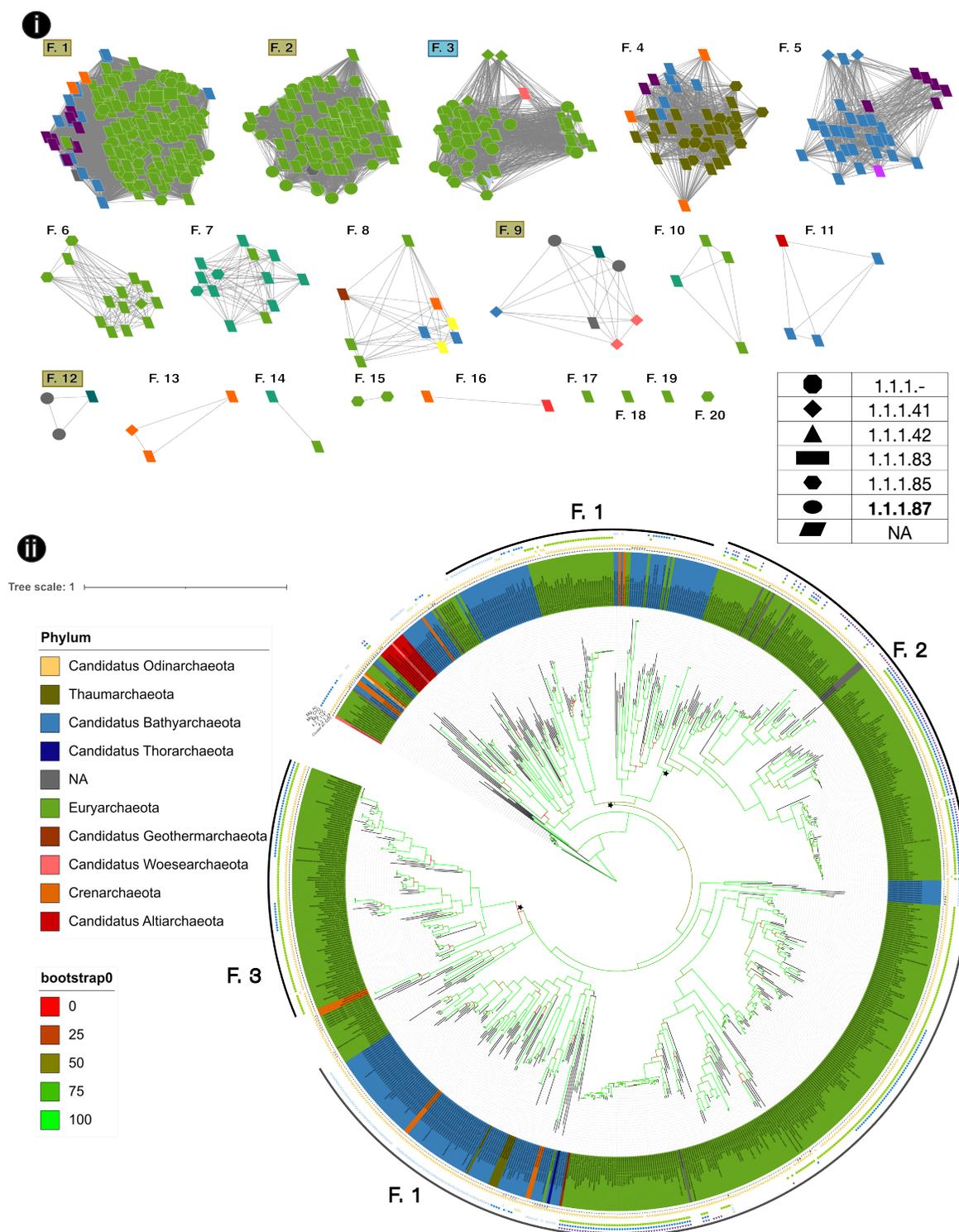


Figura B.1: SSN y filogenia de la homoisocitrato deshidrogenasa dependiente de NAD metanógena. **i)** En la SSN, los nodos tienen dos atributos representados: el phylum al que pertenecen (los colores están señalados en la leyenda inferior) y la clasificación EC a la que pertenece la secuencia (denotada según las formas diferentes del nodo mostradas en la leyenda de la parte derecha). **ii)** En la filogenia, los colores de las OTUs están dados por la misma leyenda que en la SSN; además incluye el número de clúster al que pertenece en la red, el tipo de actividad enzimática que realiza, si y el organismo al que pertenece tiene la capacidad metabólica de realizar metanogénesis. “bootstrap0” se refiere al nombre del conjunto de datos.

Bibliografía

- [1] Becerra A. The Semi-Enzymatic Origin of Metabolic Pathways: Inferring a Very Early Stage of the Evolution of Life. *Journal of Molecular Evolution*. 2021 4;89:183–188. Available from: <http://link.springer.com/10.1007/s00239-021-09994-0>.
- [2] Caetano-Anollés G, Yafremava LS, Gee H, Caetano-Anollés D, Kim HS, Mittenhal JE. The origin and evolution of modern metabolism. *The International Journal of Biochemistry and Cell Biology*. 2009 2;41:285–297. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1357272508003373>.
- [3] Fani R. The Origin and Evolution of Metabolic Pathways: Why and How did Primordial Cells Construct Metabolic Routes? *Evolution: Education and Outreach*. 2012 9;5:367–381. Available from: <https://evolution-outreach.biomedcentral.com/articles/10.1007/s12052-012-0439-5>.
- [4] Lazcano A, Miller SL. On the Origin of Metabolic Pathways. *Journal of Molecular Evolution*. 1999;49:424–431.
- [5] Noda-Garcia L, Liebermeister W, Tawfik DS. Metabolite–Enzyme Coevolution: From Single Enzymes to Metabolic Pathways and Networks. *Annual Review of Biochemistry*. 2018 6;87:187–216. Available from: <https://www.annualreviews.org/doi/10.1146/annurev-biochem-062917-012023>.
- [6] Wachtershauser G. Evolution of the first metabolic cycles. *Proceedings of the National Academy of Sciences*. 1990 1;87:200–204. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.87.1.200>.
- [7] Oparin AI. *The origin of life*. MacMillan; 1938.

- [8] Haldane JBS. The origin of life. *Rationalist Annual*. 1929;148:3–10. Available from: <https://cir.nii.ac.jp/crid/1570009751001377664.bib?lang=en>.
- [9] Horowitz NH. On the Evolution of Biochemical Syntheses. *Proceedings of the National Academy of Sciences*. 1945 6;31:153–157. Available from: <https://www.pnas.org/doi/abs/10.1073/pnas.31.6.153>.
- [10] Horowitz NH. In: *The Evolution of Biochemical Syntheses — Retrospect and Prospect*. Elsevier; 1965. p. 15–23. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9781483227344500097>.
- [11] Granick S. Speculations on the origins and evolution of photosynthesis. *Annals of the New York Academy of Sciences*. 1957 8;69:292–308. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/13479007>.
- [12] Jensen RA. Enzyme recruitment in evolution of new function. *Ann Rev MicrobioL*. 1976;30:409–434. Available from: www.annualreviews.org/aronline.
- [13] Yčas M. On earlier states of the biochemical system. *Journal of Theoretical Biology*. 1974;44:145–160.
- [14] Lewis EB. Pseudoallelism and gene evolution. *Cold Spring Harbor Symposia on Quantitative Biology*. 1951 1;16:159–174. Available from: <http://symposium.cshlp.org/cgi/doi/10.1101/SQB.1951.016.01.014>.
- [15] Ohno S. *Evolution by Gene Duplication*. Springer Berlin Heidelberg; 1970. Available from: <http://link.springer.com/10.1007/978-3-642-86659-3>.
- [16] Meléndez-Hevia E, Waddell TG, Cascante M. The puzzle of the Krebs citric acid cycle: Assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *Journal of Molecular Evolution*. 1996 9;43:293–303. Available from: <https://link.springer.com/article/10.1007/BF02338838><http://link.springer.com/10.1007/BF02338838>.
- [17] Peretó J. In: *Origin and evolution of metabolisms*. Cambridge University Press; 2011. p. 270–288. Available from: https://www.cambridge.org/core/product/identifier/CB09780511933875A031/type/book_part.

- [18] Takiguchi M, Matsubasa T, Amaya Y, Mori M. Evolutionary aspects of urea cycle enzyme genes. *BioEssays*. 1989;10:163–166.
- [19] Scossa F, Fernie AR. The evolution of metabolism: How to test evolutionary hypotheses at the genomic level. *Computational and Structural Biotechnology Journal*. 2020 1;18:482–500. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2001037019303654>.
- [20] Wagner A. *Arrival of the fittest: Solving evolution's greatest puzzle*. 1st ed. Current; 2014. Includes bibliographical references (pages 225-280) and index.
- [21] Begley TP. Cofactor biosynthesis: an organic chemist's treasure trove. *Natural Product Reports*. 2006;23:15. Available from: <http://xlink.rsc.org/?DOI=b207131m>.
- [22] Webb ME, Marquet A, Mendel RR, Rébeillé F, Smith AG. Elucidating biosynthetic pathways for vitamins and cofactors. *Natural Product Reports*. 2007;24:988. Available from: <http://xlink.rsc.org/?DOI=b703105j>.
- [23] Eakin RE. An approach to the evolution of metabolism. *Proceedings of the National Academy of Sciences*. 1963 3;49:360–366. Available from: <https://www.pnas.org/doi/abs/10.1073/pnas.49.3.360>.
- [24] Goldman AD, Kacar B. Cofactors are Remnants of Life's Origin and Early Evolution. *Journal of Molecular Evolution*. 2021;89:127–133. Available from: <https://doi.org/10.1007/s00239-020-09988-4>.
- [25] Handler P. *Evolution of the coenzymes*. MacMillan; 1963. p. 149–157.
- [26] Hartman H. Speculations on the origin and evolution of metabolism. *Journal of Molecular Evolution*. 1975 12;4:359–370. Available from: <http://link.springer.com/10.1007/BF01732537>.
- [27] White HB. Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution*. 1976;7:101–104. Available from: <https://link.springer.com/article/10.1007/BF01732468>.
- [28] Jadhav VR, Yarus M. Coenzymes as coribozymes. *Biochimie*. 2002 9;84:877–888. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0300908402014049>.

- [29] Kirschning A. The coenzyme/protein pair and the molecular evolution of life. *Natural Product Reports*. 2021;38:993–1010. Available from: <http://xlink.rsc.org/?DOI=DONP00037J>.
- [30] Kirschning A. Coenzymes and Their Role in the Evolution of Life. *Angewandte Chemie International Edition*. 2021 3;60:6242–6269. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/anie.201914786>.
- [31] Kirschning A. On the evolution of coenzyme biosynthesis. *Natural Product Reports*. 2022;39:2175–2199. Available from: <http://xlink.rsc.org/?DOI=D2NP00037G>.
- [32] Sharov AA. Coenzyme world model of the origin of life. *Biosystems*. 2016 6;144:8–17. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0303264716300107>.
- [33] Hernández-Morales R, Becerra A, Lazcano A. Alarmones as Vestiges of a Bygone RNA World. *Journal of Molecular Evolution*. 2019 1;87:37–51. Available from: <http://link.springer.com/10.1007/s00239-018-9883-3><https://doi.org/10.1007/s00239-018-9883-3>.
- [34] Deppenmeier U. The unique biochemistry of methanogenesis. *Progress in Nucleic Acid Research and Molecular Biology*. 2002;71:223–283.
- [35] Grochowski LL, White RH. In: *Biosynthesis of the Methanogenic Coenzymes*. Elsevier; 2010. p. 711–748. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780080453828001477>.
- [36] White RH. Biosynthesis of the methanogenic cofactors. *Vitamins and Hormones*. 2001 1;61:299–337.
- [37] Garcia PS, Gribaldo S, Borrel G. Diversity and Evolution of Methane-Related Pathways in Archaea. *Annual Review of Microbiology*. 2022 9;76:727–755. Available from: <https://www.annualreviews.org/doi/10.1146/annurev-micro-041020-024935>.
- [38] Ma K, Lu Y. Regulation of microbial methane production and oxidation by intermittent drainage in rice field soil. *FEMS Microbiology Ecology*. 2011 3;75:446–

456. Available from: <https://academic.oup.com/femsec/article-lookup/doi/10.1111/j.1574-6941.2010.01018.x>.
- [39] Yuan Y, Conrad R, Lu Y. Transcriptional response of methanogen *mcrA* genes to oxygen exposure of rice field soil. *Environmental Microbiology Reports*. 2011 6;3:320–328.
- [40] Liu Y, Whitman WB. Metabolic, Phylogenetic, and Ecological Diversity of the Methanogenic Archaea. *Annals of the New York Academy of Sciences*. 2008 3;1125:171–189. Available from: <http://doi.wiley.com/10.1196/annals.1419.019>.
- [41] Thomas CM, Quéméner EDL, Gribaldo S, Borrel G. Factors shaping the abundance and diversity of the gut archaeome across the animal kingdom. *Nature Communications*. 2022 6;13:3358. Available from: <https://www.nature.com/articles/s41467-022-31038-4>.
- [42] Graham DE, White RH. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. *Natural Product Reports*. 2002 3;19:133–147. Available from: <http://xlink.rsc.org/?DOI=b103714p>.
- [43] Liu Y, Beer LL, Whitman WB. Methanogens: A window into ancient sulfur metabolism. *Trends in Microbiology*. 2012 5;20:251–258.
- [44] Lyu Z, Shao N, Akinyemi T, Whitman WB. Methanogenesis. *Current Biology*. 2018 7;28:R727–R732. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0960982218306237>.
- [45] Battistuzzi FU, Feijao A, Hedges SB. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol*. 2004;4. Available from: <http://www.biomedcentral.com/1471-2148/4/44>.
- [46] Ueno Y, Yamada K, Yoshida N, Maruyama S, Isozaki Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature*. 2006 3;440:516–519. Available from: <http://www.nature.com/articles/nature04584>.

- [47] Wolfe JM, Fournier GP. Horizontal gene transfer constrains the timing of methanogen evolution. *Nature Ecology and Evolution*. 2018 5;2:897–903. Available from: <http://www.nature.com/articles/s41559-018-0513-7>.
- [48] Wordsworth R, Pierrehumbert R. Hydrogen-nitrogen greenhouse warming in earth's early atmosphere. *Science*. 2013 1;339:64–67. Available from: www.sciencemag.org.
- [49] Kasting JF, Siefert JL. Life and the evolution of Earth's atmosphere. *Science*. 2002 5;296:1066–1068. Available from: <https://www.science.org/doi/10.1126/science.1071184>.
- [50] Zhao Z, Shen B, Zhu JM, Lang X, Wu G, Tan D, et al. Active methanogenesis during the melting of Marinoan snowball Earth. *Nature communications*. 2021 12;12:955. Available from: <https://doi.org/10.1038/s41467-021-21114-6><http://www.nature.com/articles/s41467-021-21114-6><http://www.ncbi.nlm.nih.gov/pubmed/33574253><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7878791>.
- [51] Borrel G, Adam PS, Gribaldo S. Methanogenesis and the Wood–Ljungdahl Pathway: An Ancient, Versatile, and Fragile Association. *Genome Biology and Evolution*. 2016 6;8:1706–1711. Available from: <https://academic.oup.com/gbe/article/8/6/1706/2574027>.
- [52] Martin W, Russell MJ. On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2006 10;362:1887–1926. Available from: <https://royalsocietypublishing.org/doi/10.1098/rstb.2006.1881>.
- [53] Sousa FL, Thiergart T, Landan G, Nelson-Sathi S, Pereira IAC, Allen JF, et al. Early bioenergetic evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2013 7;368. Available from: <https://royalsocietypublishing.org/doi/10.1098/rstb.2013.0088>.
- [54] Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, et al. The physiology and habitat of the last universal common ancestor. *Nature Microbiology*. 2016;1:1–8. Available from: <http://dx.doi.org/10.1038/nmicrobiol.2016.116>.

- [55] Muñoz-Velasco I, García-Ferris C, Hernandez-Morales R, Lazcano A, Peretó J, Becerra A. Methanogenesis on Early Stages of Life: Ancient but Not Primordial. *Origins of Life and Evolution of Biospheres*. 2018;48:407–420. Available from: <https://doi.org/10.1007/s11084-018-9570-9>.
- [56] Blaut M. Metabolism of methanogens. *Antonie van Leeuwenhoek*. 1994;66:187–208. Available from: <http://link.springer.com/10.1007/BF00871639>.
- [57] Ferry JG. How to Make a Living by Exhaling Methane Methanogen: a microorganism that produces methane Anaerobe: a microorganism with energy-yielding metabolism independent of O₂ proliferating in anaerobic O₂-free environments. *Annu Rev Microbiol*. 2010;64:453–473. Available from: www.annualreviews.org.
- [58] Lemaire ON, Wagner T. A Structural View of Alkyl-Coenzyme M Reductases, the First Step of Alkane Anaerobic Oxidation Catalyzed by Archaea. *Biochemistry*. 2022 5;61:805–821. Available from: <https://pubs.acs.org/doi/10.1021/acs.biochem.2c00135>.
- [59] Thauer RK. Anaerobic oxidation of methane with sulfate: on the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO₂. *Current Opinion in Microbiology*. 2011 6;14:292–299. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1369527411000439>.
- [60] Evans PN, Boyd JA, Leu AO, Woodcroft BJ, Parks DH, Hugenholtz P, et al. An evolving view of methane metabolism in the Archaea. *Nature Reviews Microbiology*. 2019 4;17:219–232. Available from: <http://www.nature.com/articles/s41579-018-0136-7>.
- [61] Baptiste E, Brochier C, Boucher Y. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea (Vancouver, BC)*. 2005 5;1:353–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15876569><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2685549>.
- [62] Evans PN, Parks DH, Chadwick GL, Robbins SJ, Orphan VJ, Golding SD, et al. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science*. 2015 10;350:434–438. Available from: <https://www.science.org>.

- [63] Vanwonterghem I, Evans PN, Parks DH, Jensen PD, Woodcroft BJ, Hugenholtz P, et al. Methylo-trophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nature Microbiology*. 2016 12;1:16170. Available from: <http://www.nature.com/articles/nmicrobiol2016170>.
- [64] Berghuis BA, Yu FB, Schulz F, Blainey PC, Woyke T, Quake SR. Hydroge-notrophic methanogenesis in archaeal phylum Verstraetearchaeota reveals the shared ancestry of all methanogens. *Proceedings of the National Academy of Sciences of the United States of America*. 2019 3;116:5037–5044. Available from: <https://www.pnas.org/doi/abs/10.1073/pnas.1815631116>.
- [65] Wang Y, Wegener G, Hou J, Wang F, Xiao X. Expanding anaerobic alkane metabolism in the domain of Archaea. *Nature Microbiology*. 2019 4;4:595–602. Available from: <http://www.nature.com/articles/s41564-019-0364-2>.
- [66] Liu YF, Chen J, Zaramela LS, Wang LY, Mbadinga SM, Hou ZW, et al. Geno-mic and Transcriptomic Evidence Supports Methane Metabolism in Archaeoglobi. *mSystems*. 2020 4;5. Available from: <https://journals.asm.org/doi/10.1128/mSystems.00651-19>.
- [67] Borrel G, O’Toole PW, Harris HMB, Peyret P, Brugère JF, Gribaldo S. Phy-logenomic Data Support a Seventh Order of Methylo-trophic Methanogens and Provide Insights into the Evolution of Methanogenesis. *Genome Biology and Evolution*. 2013 10;5:1769–1780. Available from: <https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evt128>.
- [68] Borrel G, Adam PS, McKay LJ, Chen LX, Sierra-García IN, Sieber CMK, et al. Wide diversity of methane and short-chain alkane me-tabolisms in uncultured archaea. *Nature microbiology*. 2019 4;4:603–613. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30833729><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6453112>.
- [69] Wang Y, Wegener G, Williams TA, Xie R, Hou J, Tian C, et al. A methylo-trophic origin of methanogenesis and early divergence of anaerobic multicarbon alkane metabolism. *Science Advances*. 2021 7;7. Available from: <https://www.science.org/doi/10.1126/sciadv.abj1453>.

- [70] Nobu MK, Narihiro T, Kuroda K, Mei R, Liu WT. Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. *The ISME Journal*. 2016 10;10:2478–2487. Available from: <http://www.nature.com/articles/ismej201633>.
- [71] McKay LJ, Dlakić M, Fields MW, Delmont TO, Eren AM, Jay ZJ, et al. Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nature Microbiology*. 2019 4;4:614–622. Available from: <http://www.nature.com/articles/s41564-019-0362-4>.
- [72] Laso-Pérez R, Hahn C, van Vliet DM, Tegetmeyer HE, Schubotz F, Smit NT, et al. Anaerobic Degradation of Non-Methane Alkanes by “Candidatus Methanoliparia” in Hydrocarbon Seeps of the Gulf of Mexico. *mBio*. 2019 8;10. Available from: <https://journals.asm.org/doi/10.1128/mBio.01814-19>.
- [73] Wegener G, Laso-Pérez R, Orphan VJ, Boetius A. Anaerobic Degradation of Alkanes by Marine Archaea. *Annual Review of Microbiology*. 2022 9;76:553–577. Available from: <https://www.annualreviews.org/doi/10.1146/annurev-micro-111021-045911>.
- [74] Zhou Z, jing Zhang C, fei Liu P, Fu L, Laso-Pérez R, Yang L, et al. Non-syntrophic methanogenic hydrocarbon degradation by an archaeal species. *Nature*. 2022 1;601:257–262. Available from: <https://www.nature.com/articles/s41586-021-04235-2>.
- [75] Hippe H, Caspari D, Fiebig K, Gottschalk G. Utilization of trimethylamine and other N-methyl compounds for growth and methane formation by *Methanosarcina barkeri*. *Proceedings of the National Academy of Sciences*. 1979 1;76:494–498. Available from: <https://pnas.org/doi/full/10.1073/pnas.76.1.494>.
- [76] Zheng K, Ngo PD, Owens VL, peng Yang X, Mansoorabadi SO. The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science*. 2016 10;354:339–342. Available from: <https://www.sciencemag.org/lookup/doi/10.1126/science.aag2947>.
- [77] Okada T, Tomita T, Wulandari AP, Kuzuyama T, Nishiyama M. Mechanism of Substrate Recognition and Insight into Feedback Inhibition of Homo-

- citrate Synthase from *Thermus thermophilus*. *Journal of Biological Chemistry*. 2010;285:4195–4205. Available from: <http://www.rcsb.org/>.
- [78] Drevland RM, Jia Y, Palmer DRJ, Graham DE. Methanogen Homoaconitase Catalyzes Both Hydrolyase Reactions in Coenzyme B Biosynthesis. *Journal of Biological Chemistry*. 2008 10;283:28888–28896. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0021925820641187>.
- [79] Jeyakanthan J, Drevland RM, Gayathri DR, Velmurugan D, Shinkai A, Kuramitsu S, et al. Substrate Specificity Determinants of the Methanogen Homoaconitase Enzyme: Structure and Function of the Small Subunit. *Biochemistry*. 2010 3;49:2687–2696. Available from: <https://pubs.acs.org/doi/10.1021/bi901766z>.
- [80] Howell DM, Graupner M, Xu H, White RH. Identification of Enzymes Homologous to Isocitrate Dehydrogenase That Are Involved in Coenzyme B and Leucine Biosynthesis in Methanoarchaea. *Journal of Bacteriology*. 2000 9;182:5013–5016. Available from: <https://journals.asm.org/doi/10.1128/JB.182.17.5013-5016.2000>.
- [81] Howell DM, Harich K, Xu H, White RH. α -Keto Acid Chain Elongation Reactions Involved in the Biosynthesis of Coenzyme B (7-Mercaptoheptanoyl Threonine Phosphate) in Methanogenic Archaea. *Biochemistry*. 1998 7;37:10108–10117. Available from: <https://pubs.acs.org/doi/10.1021/bi980662p>.
- [82] Solow B, White RH. Biosynthesis of the Peptide Bond in the Coenzyme N-(7-Mercaptoheptanoyl)-l-threonine Phosphate. *Archives of Biochemistry and Biophysics*. 1997 9;345:299–304. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0003986197902721>.
- [83] Martin WF, Sousa FL. Early Microbial Evolution: The Age of Anaerobes. *Cold Spring Harbor Perspectives in Biology*. 2016 2;8:a018127. Available from: <http://cshperspectives.cshlp.org/lookup/doi/10.1101/cshperspect.a018127>.
- [84] Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. *PLoS ONE*. 2009 2;4:e4345. Available from: <https://dx.plos.org/10.1371/journal.pone.0004345>.

- [85] Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*. 2004 12;20:3702–3704. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bth444>.
- [86] Copp JN, Anderson DW, Akiva E, Babbitt PC, Tokuriki N. In: Exploring the sequence, function, and evolutionary space of protein superfamilies using sequence similarity networks and phylogenetic reconstructions. vol. 620. Academic Press; 2019. p. 315–347. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0076687919300965>.
- [87] González JM. Visualizing the superfamily of metallo- β -lactamases through sequence similarity network neighborhood connectivity analysis. *Heliyon*. 2021 1;7:e05867. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2405844020327092>.
- [88] Uberto R, Moomaw EW. Protein Similarity Networks Reveal Relationships among Sequence, Structure, and Function within the Cupin Superfamily. *PLoS ONE*. 2013 9;8:e74477. Available from: <https://dx.plos.org/10.1371/journal.pone.0074477>.
- [89] Copp JN, Akiva E, Babbitt PC, Tokuriki N. Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. *Biochemistry*. 2018 8;57:4651–4662. Available from: <https://pubs.acs.org/doi/10.1021/acs.biochem.8b00473>.
- [90] Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Research*. 2019;48:445–453. Available from: <https://academic.oup.com/nar/article/48/D1/D445/5581728>.
- [91] Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*. 2021 1;49:D480–D489. Available from: <https://academic.oup.com/nar/article/49/D1/D480/6006196>.
- [92] Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, et al. Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*.

- chii. *Science*. 1996 8;273:1058–1073. Doi: 10.1126/science.273.5278.1058. Available from: <https://www.science.org/doi/10.1126/science.273.5278.1058>.
- [93] Zallot R, Oberg N, Gerlt JA. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry*. 2019 10;58:4169–4182. Available from: <https://pubs.acs.org/doi/10.1021/acs.biochem.9b00735>.
- [94] Shannon P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*. 2003 11;13:2498–2504. Available from: <http://www.genome.org/cgi/doi/10.1101/gr.1239303>.
- [95] Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, et al. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*. 2011 12;12:436. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-436>.
- [96] Wittkop T, Emig D, Truss A, Albrecht M, Böcker S, Baumbach J. Comprehensive cluster analysis with Transitivity Clustering. *Nature Protocols*. 2011. Available from: <http://transclust.cebitec.uni-bielefeld.de/>.
- [97] Röttger R, Kreutzer C, Vu TD, Wittkop T, Baumbach J. Online Transitivity Clustering of Biological Data with Missing Values. *DROPS-IDN/3718*. 2012:57–68. Available from: <http://transclust.mmci.uni-saarland.de/>.
- [98] Wittkop T, Rahmann S, Röttger R, Böcker S, Baumbach J. Extension and Robustness of Transitivity Clustering for Protein–Protein Interaction Network Analysis. *Internet Mathematics*. 2011 11;7:255–273. Available from: <http://www.internetmathematicsjournal.com/article/1510>.
- [99] Wittkop T, Emig D, Lange S, Rahmann S, Albrecht M, Morris JH, et al. Partitioning biological data with transitivity clustering. *Nature Methods*. 2010 6;7:419–420. Available from: <http://www.nature.com/articles/nmeth0610-419>.
- [100] Wiwie C, Baumbach J, Röttger R. Comparing the performance of biomedical clustering methods. *Nature Methods*. 2015 11;12:1033–1038. Available from: <http://www.nature.com/articles/nmeth.3583>.

- [101] Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. 2002 7;30:3059–3066. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkf436>.
- [102] Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*. 2013 4;30:772–780. Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst010>.
- [103] Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *BIOINFORMATICS APPLICATIONS NOTE*. 2009;25:1972–1973. Available from: <https://academic.oup.com/bioinformatics/article/25/15/1972/213148>.
- [104] Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*. 2015 1;32:268–274. Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu300>.
- [105] Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*. 2013 5;30:1188–1195. Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst024>.
- [106] Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS, Rosenberg MS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol*. 2017;35:518–522. Available from: <http://www.iqtree.org>.
- [107] Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*. 2010 05;59(3):307–321. Available from: <https://doi.org/10.1093/sysbio/syq010>.
- [108] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*. 2021;49. Available from: <https://doi.org/10.1093/nar/gkab301>.

- [109] Yoshida A, Kosono S, Nishiyama M. Characterization of two 2-isopropylmalate synthase homologs from *Thermus thermophilus* HB27. *Biochemical and biophysical research communications*. 2018 6;501:465–470. Available from: <https://pubmed.ncbi.nlm.nih.gov/29738774/>.