

Cladismo: Ejercicios de Parsimonia para Proteínas

Elena Pérez-Urria Carril

Facultad de Biología. Universidad Complutense de Madrid

elenapuc@bio.ucm.es

Resumen: Diversas técnicas se utilizan para comparar características de las entidades vivas buscando analizar, en todo caso, la semejanza entre ellas. También son diversas las características que definen a los organismos: moleculares, metabólicas, celulares, ambientales, etológicas, etc. En función de la semejanza, de características semejantes, los organismos, y toda entidad de la jerarquía biológica, se ordenan y clasifican. Este trabajo plantea estudiar una característica de los organismos que es su metabolismo, en particular el metabolismo de porfirinas y clorofilas, haciendo uso de procedimientos cladísticos que, a diferencia de otros, permiten “rastrear” los cambios que experimentan los caracteres y que definen a los organismos. El resultado de todo análisis cladístico es una hipótesis evolutiva sobre relaciones de parentesco, es decir, una hipótesis filogenética.

Palabras clave: Cladismo. Parsimonia. Cladograma. Metabolismo. Enzimas. Hipótesis Evolutiva

INTRODUCCIÓN

El objetivo de establecer una clasificación natural de especies basada en sus relaciones filogenéticas se alcanza comparando algunas de sus características o atributos. Por ello, los métodos y técnicas utilizados para el estudio de la diversidad de especies son una aplicación del **método comparado**.

Los caracteres son rasgos que presentan las entidades biológicas, desde las moléculas hasta los ecosistemas, los cuales son recogidos por las ciencias descriptivas de todos los niveles de integración (genética, bioquímica, anatomía, morfología).

En un contexto matemático, los caracteres pueden ser de variación discontinua (caracteres discretos, cualitativos, absolutos como presencia/ausencia) o de variación continua (son los caracteres cuantitativos que comúnmente se refieren a dimensiones y cuya distribución en la población o en la especie se suele ajustar a una curva de Gauss), pudiendo convertir los continuos en discretos mediante el establecimiento de clases (por ejemplo, la edad: un carácter continuo que puede convertirse en discreto estableciendo clases de edad, de 10 a 20 años, de 20 a 30, etc.).

En el contexto biológico los caracteres sirven a la taxonomía siendo su acepción más clásica aquella que los define como rasgos que diferencian unas especies de otras. Estos caracteres han de ser susceptibles de ser comparados, homólogos y constantes, y pueden ser de muchos tipos: morfológicos, etológicos, fisiológicos, bioquímicos, geográficos, autoecológicos o moleculares, por ejemplo. El uso de estos últimos en Biología Sistemática representa un importante nexo entre Biología Molecular y Celular y Biología de Organismos y Sistemas.

Caracteres homólogos son estructuras o rasgos que tienen el mismo origen. La homología puede definirse como una comunidad de origen de rasgos. Naturalmente, hablar de homología es hablar de semejanza: estamos tratando el estudio de la diversidad biológica que se genera por evolución y entre las estirpes que constituyen esa diversidad existen diferencias y semejanzas. Nuestra atención se dirige a la semejanza (Fig.1).



Figura 1. Relación entre diversidad biológica y evolución.

Los caracteres morfológicos homólogos normalmente son muy semejantes a lo largo de la evolución de las estirpes emparentadas. Como ejemplo sirva la composición de la mano de todos los tetrápodos o la anatomía de las extremidades de los mamíferos, estructuras homólogas con distintas morfologías según sus diferentes funciones debidas a sus distintos modos de vida (terrestre, acuático, aéreo). Este último ejemplo sirve para intuir que la semejanza se debe a la posesión de un ancestro común. Con ello introducimos un nuevo término, **ancestría**.

También existe homología en el nivel molecular. La mayor parte de las moléculas que conforman los seres vivos se “inventan” pronto de manera que es posible seguir la evolución de moléculas homólogas desde los tiempos más remotos. Es el caso de la actina y la miosina, presentes en arqueobacterias, eubacterias y eucariontes. Estas proteínas (y los genes que las codifican) divergen a lo largo de la evolución entre unas estirpes y otras resultando más semejantes entre sí las de cada grupo natural. Dos genes o dos proteínas serán homólogas para un conjunto de estirpes si poseen el mismo origen, y comúnmente serán más semejantes cuánto más cercano sea el parentesco entre esas estirpes. En el nivel molecular, deben hacerse dos precisiones para la homología: caracteres homólogos “**ortólogos**” y “**parálogos**”, consecuencia de la duplicación de genes.

Pero no siempre la semejanza se debe a homología: puede deberse a **analogía** u **homoplasia** (construcción semejante). Analogía es la condición de semejanza presentada por estirpes no inmediatamente emparentadas pero con un modo de vida parecido. En consecuencia se definen caracteres análogos como aquellos que tienen la misma función pero origen distinto, habiendo diferentes tipos de analogía: convergencia, paralelismo o reversión al estado ancestral.

Podemos incorporar a nuestro pensamiento evolutivo la idea de que las estirpes más estrechamente emparentadas son las que comparten mayor número de caracteres y con ello introducimos un nuevo concepto, "cantidad de semejanza". Pero la cuestión de la analogía, que es un problema para la reconstrucción de la filogenia, conduce a un nuevo planteamiento: la necesidad de considerar no sólo la cantidad sino también la "**calidad**" de la semejanza. Por tanto, habrá que determinar homologías (debe tratarse de la misma estructura fundamental, de una misma posición respecto a una referencia y el mismo patrón de desarrollo embrionario, cuando se trate de caracteres morfológicos; proteínas sintetizadas por el mismo gen en un grupo de estirpes, etc.) y detectar homoplasias.

Se debe diferenciar entre **homologías** ancestrales y derivadas para llegar a la conclusión de que sólo las **derivadas** son indicadores fiables de **parentesco inmediato**. Por el momento podemos intuir para el término "derivado" el significado de "reciente" y considerar que las homologías ancestrales se denominan "plesiomorfías" y las derivadas "**sinapomorfías**" en términos cladistas que se ven a continuación. Para distinguir entre ambas se establece de varias formas la polaridad de los caracteres.

Respecto a los procedimientos para reconstruir la filogenia de las estirpes, éstos se enmarcan dentro de lo que denominamos **escuelas sistemáticas**: Escuela Evolutiva (Escuela de la Nueva Sistemática o Escuela de Simpson – Mayr), Escuela de la Taxonomía Numérica (Escuela del Feneticismo Numérico) y **Escuela Cladista**.

La Escuela Cladística (o de la Sistemática Filogenética) creada por W. Hennig (1913-1976) en la década de 1950 plantea una manera de realizar la reconstrucción filogenética basada en la "**calidad de la semejanza**" y no en la cantidad de semejanza. Según esto, **el parentesco inmediato no se deduce del hecho de compartir muchos caracteres (cantidad) sino de compartir alguno(s) que, además de homólogo(s), sea sinapomórfico (condición de calidad), es decir, exclusivo de los miembros de un grupo y de su ancestro común inmediato** (en el que apareció por primera vez el carácter). Estas sinapomorfías definen a los grupos. Por otra parte, se considera que las especies y los grupos mayores existen en la naturaleza como resultado de la evolución y se atribuye a la especie un papel protagonista como realidad evolutiva. Aporta un enorme poder analítico y precisa términos usuales de la teoría evolutiva: monofilia, parafilia, polifilia, grupos hermanos y estados de los caracteres.

El cladismo introduce un nuevo concepto: "**estados de los caracteres**". Ya antes se ha mencionado que los caracteres pueden ser derivados (avanzados) o ancestrales (primitivos). Ahora se habla de caracteres "plesiomórficos" y "apomórficos" que, lejos de

ser conceptos absolutos, se refieren siempre a un antecesor inmediato en un grupo determinado (**relativos a un nodo**). La terminología cladista incluye otros nuevos términos como “clado” para referirnos a un grupo y “cladograma” para referirnos al gráfico de ramificaciones dicotómicas (un árbol) que muestra las relaciones de parentesco. La condición de compartir un carácter plesiomórfico (debido a ancestros remotos) es un “simplesiomorfía” y los caracteres simplesiomórficos no garantizan la monofilia de un grupo, indican un ancestro común pero no inmediato y exclusivo. La condición de compartir un carácter apomórfico (una novedad evolutiva en el ancestro inmediato de un grupo en estudio) se denomina “sinapomorfía” y son los caracteres sinapomórficos los únicos que garantizan la monofilia de un grupo. El cladismo también precisa el concepto de semejanza: ésta puede ser debida a la posesión de rasgos homólogos o a homoplasia, pero aclara que, para cada nodo del cladograma, existen dos tipos de homología: la que se debe a ancestros remotos (“homología simplesiomórfica”) y la que se debe a ancestro inmediato (“**homología sinapomórfica**”) (Fig. 2).

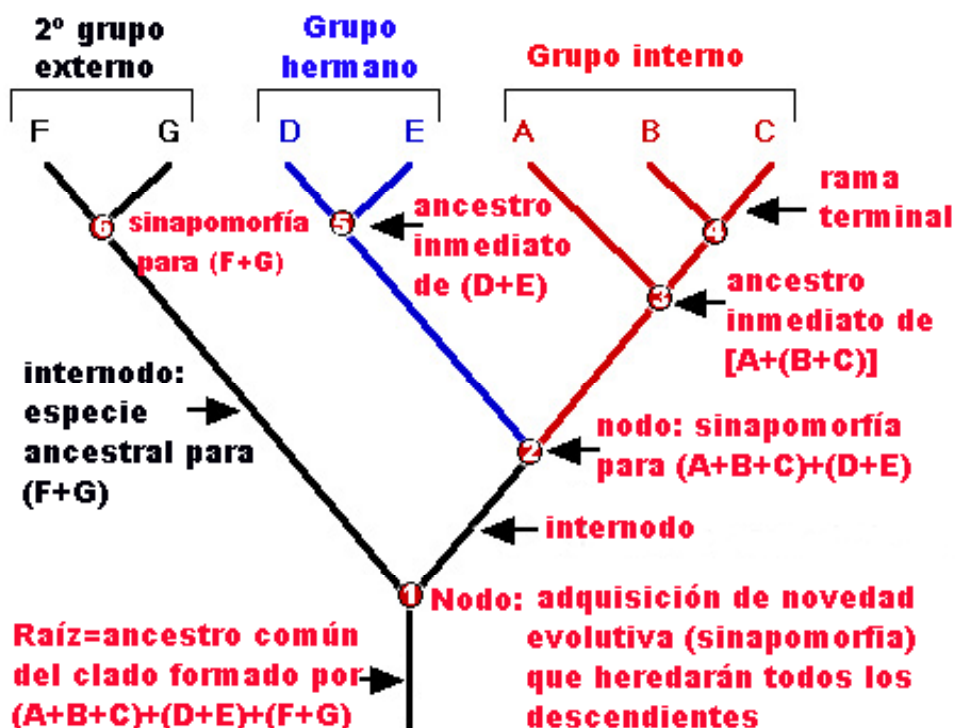


Figura 2. Conceptos del cladismo.

El análisis cladístico descubre relaciones entre grupos hermanos estudiando la distribución de sinapomorfías, es decir, caracteres derivados que representan homología y definen un grupo monofilético (un grupo natural). Esta distribución se representa en forma de cladograma, un diagrama de ramificaciones que muestran las relaciones entre taxones. Un **cladograma** es una **hipótesis evolutiva**; puede existir más de una hipótesis, es decir, más de un agrupamiento o más de una solución. Si hay soluciones **alternativas**, se escoge la **más simple o más parsimoniosa**, esto es, la que muestra menor número de cambios en la distribución de los caracteres. Este número de cambios es la **longitud** del cladograma.

Los datos requeridos para el análisis cladístico son taxones y caracteres. Los caracteres pueden ser morfológicos, anatómicos, moleculares, etc. Entre éstos últimos se consideran las secuencias de DNA, RNA o proteínas.

METABOLISMO DE PORFIRINAS

Las porfirinas son **tetrapirroles** a los que se une covalentemente un átomo metálico: con hierro se forman citocromos, peroxidasa, catalasa, mioglobina y hemoglobina; con cobre o níquel se forman moléculas para el transporte de electrones en bacterias metanógenas y acetógenas; con magnesio se forman clorofilas y bacterioclorofilas (Fig. 3).

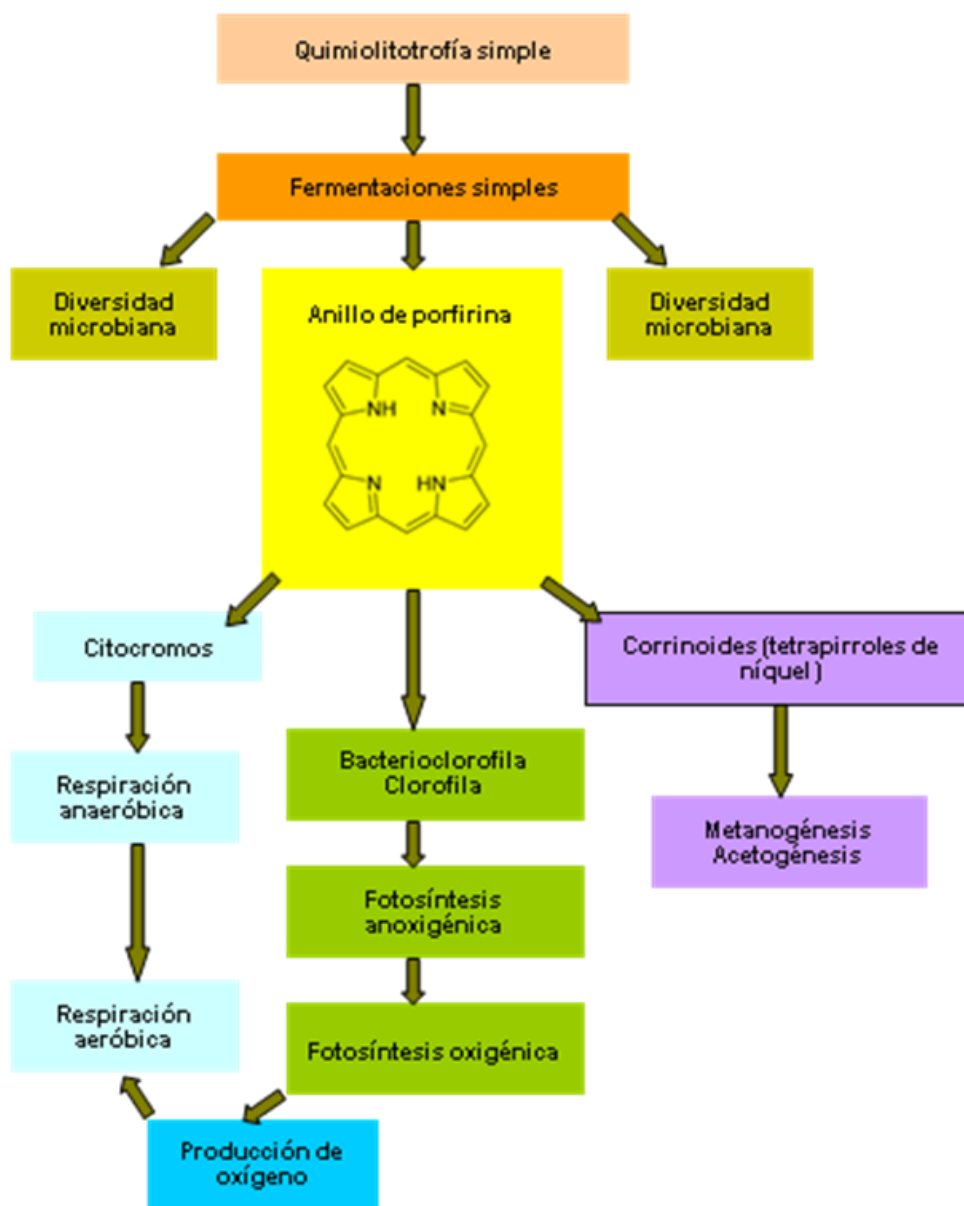


Figura 3. Tetrapirroles, formación de citocromos, clorofilas y corrinoides.

En la ruta de síntesis de porfirinas pueden distinguirse dos etapas:

- Síntesis de 5-aminolevulinato (ALA)** que difiere según los organismos. En eucariotas fotosintéticos y en todos los procariotas exceptuando el grupo de las α -proteobacterias, ALA se sintetiza a partir de glutamato que se une a glutamil-tRNA en reacción catalizada por glutamil-tRNA sintasa (GluRS), se reduce a glutamato-1-semialdehído en reacción catalizada por glutamil-tRNA reductasa y, finalmente, la glutamato-1-semialdehído 2,1 aminomutasa cataliza la transaminación que rinde ALA (Ruta C5). Por otra parte, todos los eucariotas no fotosintéticos (animales, hongos y apicomplexos) así como las α -proteobacterias forman ALA por condensación de succinil-CoA con glicina en una reacción catalizada por ALA sintasa (Ruta C4) (Fig. 4).

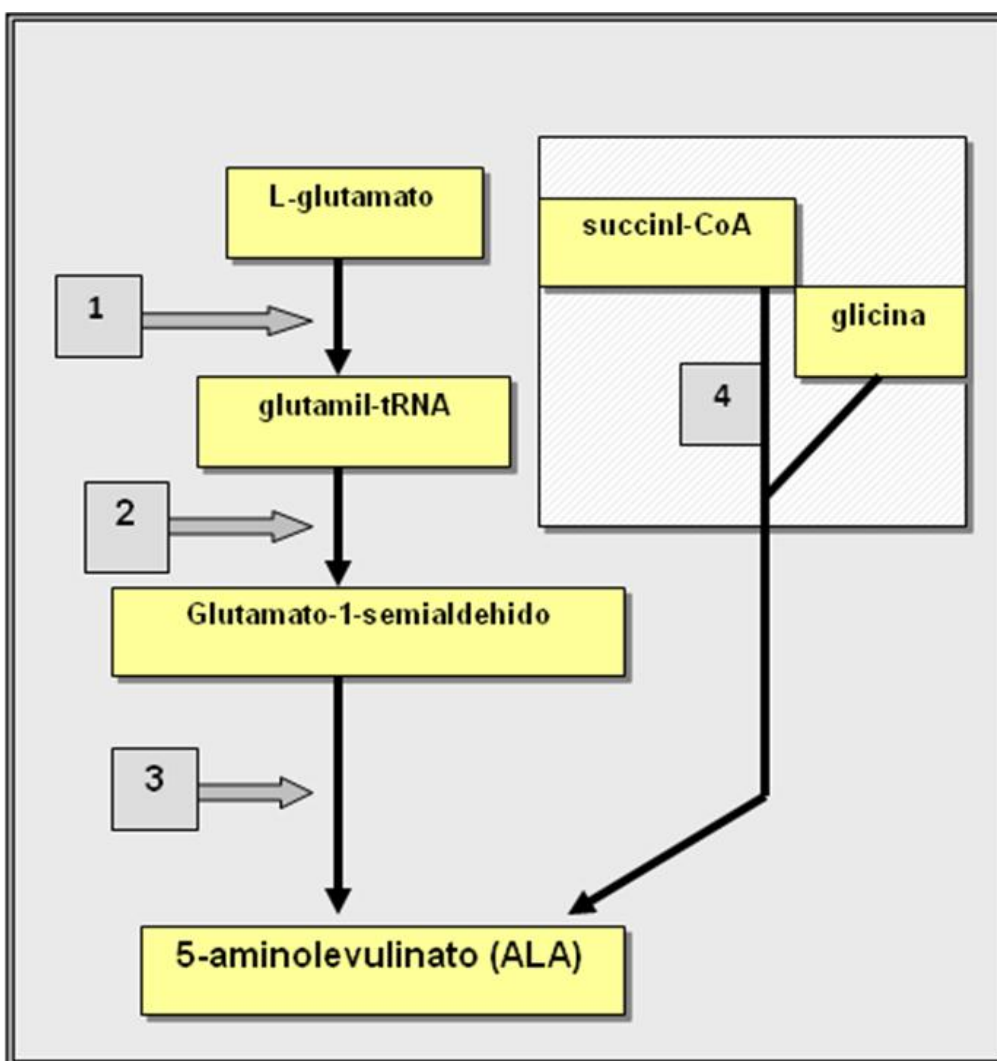


Figura 4. Rutas alternativas, C4 y C5, para la síntesis de 5-aminolevulinato. (Enzimas numeradas en Tabla 1).

- Formación de protoporfirina IX** a partir de 5-aminolevulinato, una cadena de seis reacciones, las mismas en todos los organismos (Fig. 5).

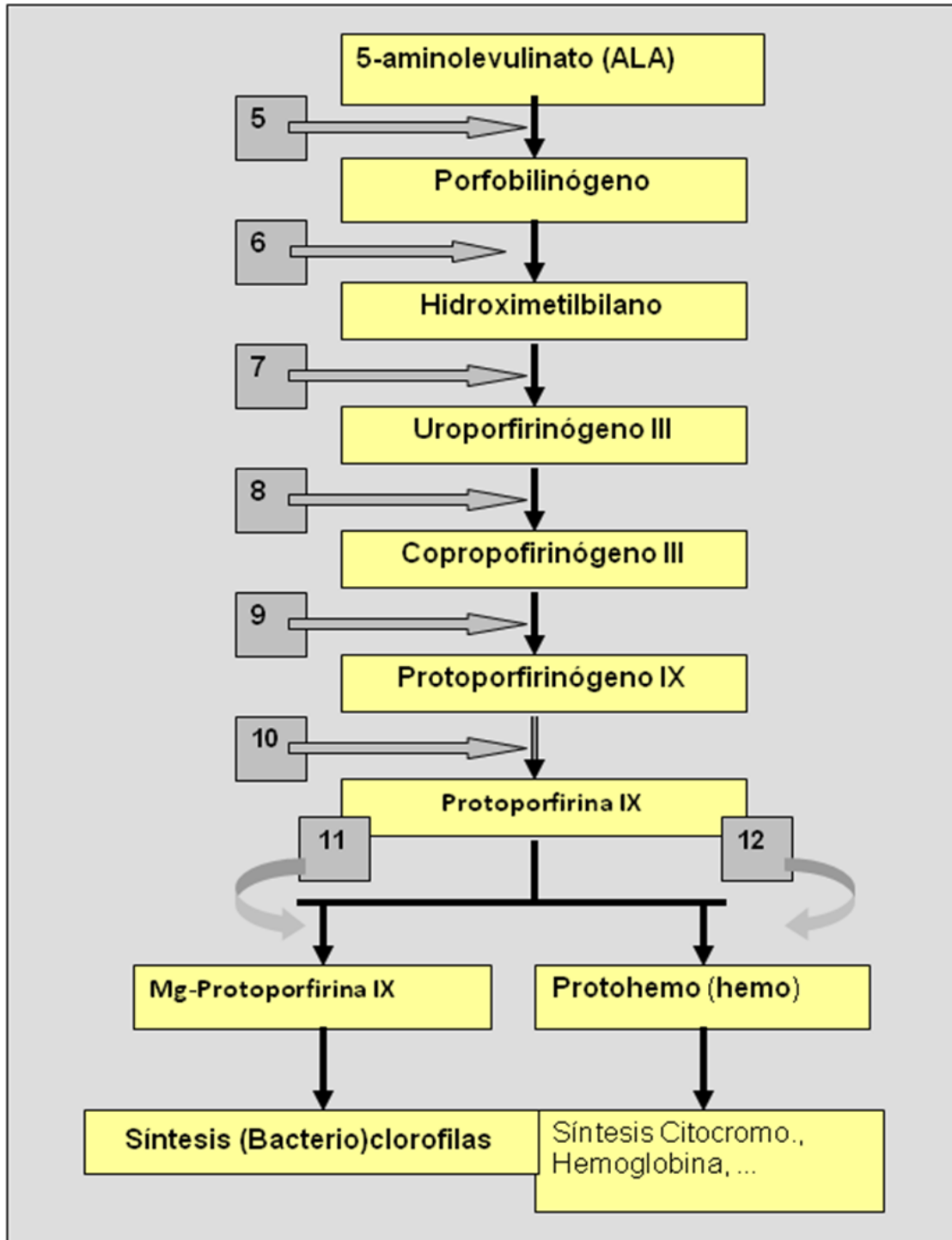


Figura 5. Ruta de síntesis de protoporfirina IX a partir de 5-aminolevulinato. (Enzimas numeradas en tabla 1).

Una perspectiva general del metabolismo de aminoácidos, cofactores y vitaminas en el que se ubica el metabolismo de porfirinas se encuentra en los siguientes enlaces http://www.sigmaaldrich.com/img/assets/4202/MetabolicPathways_6_17_04_.pdf <http://www.manet.illinois.edu/viewmap.php?release=2&map=map00860&mb=Metabolism%20of%20Cofactors&page=cofactors.php>

	ENZIMA	E.C.	GENES
1	glutamil-tRNA sintasa	6.1.1.17	gltX
2	glutamil-tRNA reductasa	1.2.1.70	hem A
3	glutamato-1-semialdehído 2,1 aminomutasa	5.4.3.8	hem L
4	5-aminolevulinato sintasa	2.3.1.37	hem A, ALAS, alas, HEM1
5	porfobilinógeno sintasa	4.2.1.24	hem B
6	porfobilinógeno deaminasa	2.5.1.61	hem C
7	uroporfirinógeno III sintasa	4.2.1.75	hem D
8	uroporfirinógeno III descarboxilasa	4.1.1.37	hem E
9	coproporfirinógeno III oxidasa	1.3.3.3	hem F
10	protoporfirinógeno IX oxidasa	1.3.3.4	hem Y, hem G
11	Mg-quelatasa	6.6.1.1	chl I, chl D, chlH
12	ferroquelatasa	4.99.1.1	hem H

Tabla 1. Enzimas de la ruta de síntesis de porfirinas (Ver figuras 4 y 5).

La mayoría de los organismos sintetizan tetrapirroles dado que son **esenciales** en las principales rutas de **conversión de energía**: respiración y fotosíntesis (Fig. 3). En términos evolutivos puede decirse que se trata de una ruta biosintética bien conservada. ALA es el precursor universal de todos los tetrapirroles existiendo 2 rutas alternativas para su biosíntesis (Fig. 4). Sin embargo, la ruta desde ALA hasta el primer tetrapirrol cíclico (Fig. 5), uroporfirinógeno III, es idéntica en todos los organismos.

A partir de uroporfirinógeno III (Fig. 5) surgen 3 ramas que conducen a la biosíntesis de productos finales específicos:

- a) **Corrinoides** (Vitamina B₁₂) que contiene **Co** en el tetrapirrol, y **Sirohemo** (el grupo prostético de nitrito y sulfito reductasas). Esta sería la primera ramificación.
- b) En segundo lugar, la ruta que conduce a las **clorofilas** que comienza con la quelación de **Mg** en el anillo tetrapirrólico.
- c) la tercera rama conduce a los distintos grupos **hemo** y comienza con la quelación de **Fe** en el tetrapirrol. De esta rama derivan también las **ficobilinas** y el cromóforo del fitocromo.

El **Uroporfirinógeno III** es el primer tetrapirrol cíclico y el último intermediario universal de todos los productos basados en tetrapirroles.

¿Cómo ha **evolucionado** esta ruta? Horowitz (1945) propuso que la ruta evolucionó hacia atrás, es decir, las últimas enzimas de la ruta aparecieron primero a medida que la disponibilidad en el medio de los productos finales fue disminuyendo gradualmente. Tiempo después, Granick (1965) postuló la evolución hacia delante de la ruta de los tetrapirroles proponiendo que los primeros intermediarios configuraron los actuales

ciclos hacia los distintos productos finales, con distintos grado de eficacia cada uno, y siendo la evolución la que fue perfeccionando estos productos finales.

Pero es difícil imaginar un medio prebiótico conteniendo productos tales como la vitamina B₁₂ que aparecen espontáneamente por síntesis orgánica prebiótica. Es lo que plantea la hipótesis de Horowitz.

Pero también es poco probable que los primeros intermediarios de la ruta de los tetrapirroles funcionaran como pigmentos/moléculas captadoras de luz o como transportadores de electrones en las células primitivas. Es lo que se deduce del modelo de Granick. Sería razonable pensar que la ruta de los tetrapirroles evolucionó según un modelo que combine ambas propuestas.

La uroporfirina es un compuesto relativamente simple que se puede concebir actuando en procesos de absorción de luz y de transporte de electrones. Pudo sintetizarse espontáneamente en la “sopa primitiva” por medio de reacciones químicas, teniendo en cuenta la relativa facilidad de síntesis no biológica de porfirinas. Modificaciones en la configuración de la uroporfirina pudieron dar lugar a los actuales productos finales, en un desarrollo evolutivo hacia delante. Por otra parte, las etapas que conducen a la formación de uroporfirina/uroporfirinógeno pudieron ir apareciendo gradualmente en “sentido inverso”: primero uroporfirina hasta que sus precursores fueron escasos en el medio o en el ambiente primitivo. Este proceso evolutivo podría explicar la ramificación hacia los productos finales y la de las primeras etapas, considerando que es más fácil imaginar las bifurcaciones de la ruta como etapas posteriores de la evolución, es decir, como especializaciones de un diseño original más simple.

ALA es el primer intermediario universal en la biosíntesis de tetrapirroles y se forma según dos rutas alternativas de manera que todos los organismos utilizan una u otra, a excepción de *Euglena gracilis* (fitoflagelado) que emplea ambas. La producción de ALA siempre implica un intermediario activado como sustrato y una o dos etapas enzimáticas que implican la transferencia de un grupo amino con la participación de piridoxal fosfato.

La ruta C5 implica tRNA como sustrato y es probablemente más ancestral que la ruta C4 que implica glicina y succinil-CoA.

Ruta C4: ocurre en animales, hongos y algunas bacterias (alfa-proteobacterias). C-4 se refiere a los 4 átomos de C que procedentes del succinato se incorporan al esqueleto de ALA. En esta ruta, utilizada por todos los eucariotas no fotosintéticos, la reacción que conduce a la formación de ALA ocurre en la mitocondria. Un organismo fotosintético modelo en el que se ha estudiado esta ruta es *Rhodobacter spheroides*, una alfa-proteobacteria capaz de producir hemo, bacterioclorofila y corrinas. Contiene, como en otros organismos, dos genes que codifican ALAS (E.C. 2.3.1.37) y que producen por tanto isoenzimas que pueden desempeñar distintas funciones metabólicas. Los mamíferos también tienen 2 genes ALAS, uno de los cuales se expresa en tejidos eritropoyéticos y el otro en todos los tejidos (Fig. 4).

Ruta C5: es muy significativa e importante para la formación de clorofilas. Es utilizada por plantas, algas y la mayoría de las bacterias. Esta vía de producción de ALA comienza con la activación del glutamato a α -carboxil a través de la unión a un t-RNA específico. La enzima que cataliza esta reacción, tRNA sintetasa, es aparentemente idéntica a la enzima que participa en la síntesis de proteínas incorporando glutamato. Glutamyl-tRNA es el sustrato de Glutamyl-tRNA reductasa (GR) en la reacción que genera el intermediario glutamato-1-semialdehído. Es probablemente la etapa limitante de esta ruta de síntesis de ALA (Fig. 4).

En el marco de la evolución es importante destacar la diferencia conceptual entre “evolución de fotosíntesis” y “evolución de organismos fotosintetizadores”: la primera implica un número limitado de genes y/o productos génicos mientras que la evolución de los organismos implica el genoma completo. Los trabajos de Woese (1987) sobre filogenia de la subunidad 16S de ARNr muestran la evolución de los organismos desde un ancestro común, estableciendo tres dominios: *Archaea* (arqueobacterias), *Bacteria* (eubacterias) y *Eucarya* (eucariontes). Aunque esta clasificación fue discutida por otros autores (Cavalier-Smith, 1992), lo cierto es que otros estudios filogenéticos realizados a la luz de nuevos y continuos datos de secuencias y genomas que recogen las bases de datos, ofrecen resultados consistentes y acordes con los tres dominios. Sin embargo, cuando se comparan filogenia basadas en genes de herencia vertical (16S ARNr) con filogenias de genes de fotosíntesis por ejemplo, aparecen incongruencias que demuestran que las primeras no necesariamente reflejan los segundos: evolución de organismos *versus* evolución de genes.

En la actualidad, las **bases de datos** recogen los resultados de los numerosísimos trabajos de secuenciación de genes y genomas de suerte que es notoria la riqueza de datos que esto supone. Sin embargo, esta misma abundancia de resultados hace necesaria su recapitulación e integración con el fin de generar información y nuevos conocimientos.

La Biología Sistemática realiza esta tarea descriptiva a nivel molecular siguiendo el método observacional: responde a la pregunta ¿qué moléculas conocemos de un aspecto particular del metabolismo? , recopilando datos procedentes de catálogos, museos, colecciones o, en el nivel molecular, bases de datos. Esta tarea descriptiva en el nivel molecular culmina con la ordenación de los elementos o datos recopilados en sistemas de referencia que permitan recuperar con facilidad la información.

Para el estudio comparativo de entidades de diversa naturaleza (moléculas, metabolismo, órganos, organismos, poblaciones, comportamientos, distribución geográfica, ecosistemas) se hace uso de diferentes técnicas, todas ellas destinadas a analizar la semejanza entre las estirpes en estudio. Entre los distintos procedimientos, el análisis cladístico aporta un grado de objetividad al estudio que no contemplan otras técnicas como los análisis de distancias o los de máxima probabilidad aunque aportan información útil, si no necesaria, en el contexto evolutivo.

Como se ha descrito anteriormente, el metabolismo de porfirinas y clorofilas es un proceso complejo que implica numerosas reacciones químicas catalizadas por enzimas. Su estructura (elementos y funciones) en la diversidad de organismos es producto de la evolución y sólo puede reconstruirse comparando los componentes.

Sirvan estos datos para mostrar una perspectiva general del metabolismo de porfirinas y clorofilas, el carácter universal de estas moléculas y la complejidad de su estudio desde la perspectiva evolutiva.

OBJETIVOS

El objetivo de este ejercicio es el **análisis cladístico con criterio de parsimonia** de un conjunto de secuencias de proteínas: subunidad grande de la enzima **rubisco (rbCL)**, **porfobilinógeno sintasa** y **glutamyl-tRNA sintasa**, presentes en un conjunto de taxones que será el grupo de estudio.

GRUPO DE ESTUDIO

El grupo de estudio está formado por representantes de los tres dominios considerando que se trata de moléculas universales (Tabla 2).

Registro	Especie	
ath	<i>Arabidopsis thaliana</i>	(Planta)
osa	<i>Oryza sativa</i>	(Planta)
ppp	<i>Physcomitrella patens</i>	(Musgo)
olu	<i>Ostrococcus lucimarinus</i>	(Clorofila)
neu	<i>Nitrosomonas europaea</i>	(β -Proteobacteria)
ret	<i>Rhizobium etli</i>	(α -Proteobacteria)
bbt	<i>Bradyrhizobium</i>	(α -Proteobacteria)
syn	<i>Synechocistis</i>	(Cianobacteria)
syw	<i>Synechococcus</i>	(Cianobacteria)
syc	<i>Synechococcus elongatus</i>	(Cianobacteria)
tel	<i>Thermosynechococcus</i>	(Cianobacteria)
gvi	<i>Gloeobacter violaceus</i>	(Cianobacteria)
ana	<i>Anabaena</i>	(Cianobacteria)
pma	<i>Prochlorococcus marinus</i>	(Cianobacteria)
cte	<i>Chlorobium tepidum</i>	(Clorobiacea)
mja	<i>Methanococcus jannaschii</i>	(Euryarchaea)
afu	<i>Archaeoglobus fulgidus</i>	(Euryarchaea)
nph	<i>Natronomonas pharaonis</i>	(Euryarchaea)
pab	<i>Pyrococcus abyssi</i>	(Euryarchaea)
rci	Uncultured methanogenic archaeon RC-I	(Euryarch.)



Tabla 2. Grupo de estudio e imágenes de algunos de los organismos.

ESCRUTINIO DE GENES Y OBTENCIÓN DE SECUENCIAS

Se trata de los **genes que codifican las enzimas** rbcL, porfobilinógeno sintasa y glutamil-tRNA sintasa en el grupo de estudio (organismos cuyos genomas han sido completamente secuenciados). La base de datos en la que se realiza la búsqueda será **GenomeNet** (<http://www.genome.ad.jp/>) (Fig. 6). Como se muestra a continuación, la búsqueda proporciona 111 registros de los que seleccionaremos los correspondientes al grupo de estudio.

Search for Database:

(Total 111 hits)

GENES - Search term: rbcL

[ath:ArthCp030](#)
rbcL; RuBisCO large subunit ; K01601 ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]

[pop:Poptr_cp030](#)
rbcL; ribulose bisphosphate carboxylase large chain precursor (EC:4.1.1.39); K01601 ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]

[osa:3131463](#)
rbcL, OrsajCp033; ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit ; K01601 ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]

[ppp:PhpapaCp031](#)
rbcL; RuBisCO large subunit ; K01601 ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]

[cre:ChreCp049](#)
rbcL; RuBisCO large subunit ; K01601 ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]

[olu:OSTLU_32608](#)
rbcL; predicted protein; K01601 ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]

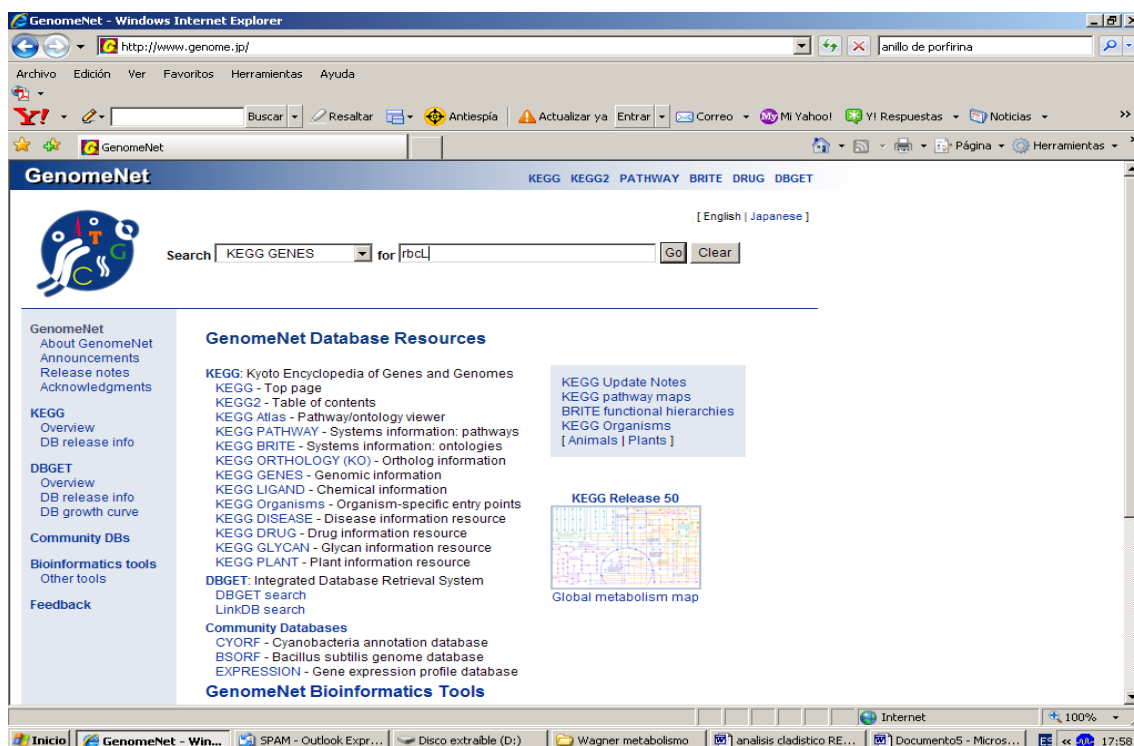


Figura 6. GenomeNet para la búsqueda de secuencias.

En la información contenida en cada registro figura la secuencia del gen y de la proteína que codifica (Fig. 7). Para realizar este ejercicio se selecciona la **secuencia de aminoácidos** (Fig. 8).

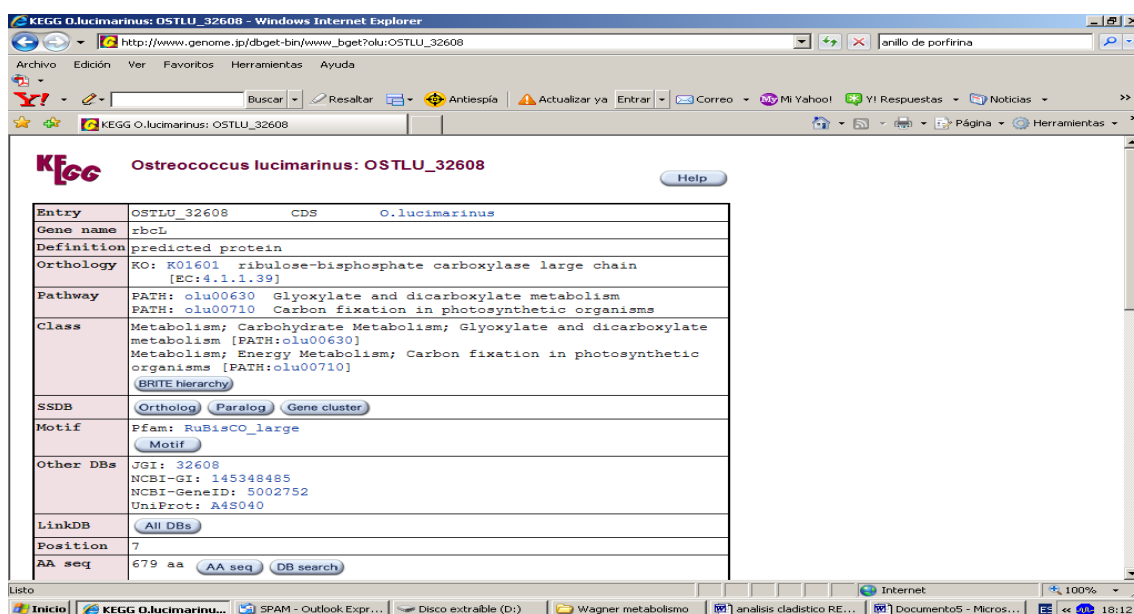


Figura 7. Información relativa al registro de *Ostreococcus lucimarinus*.

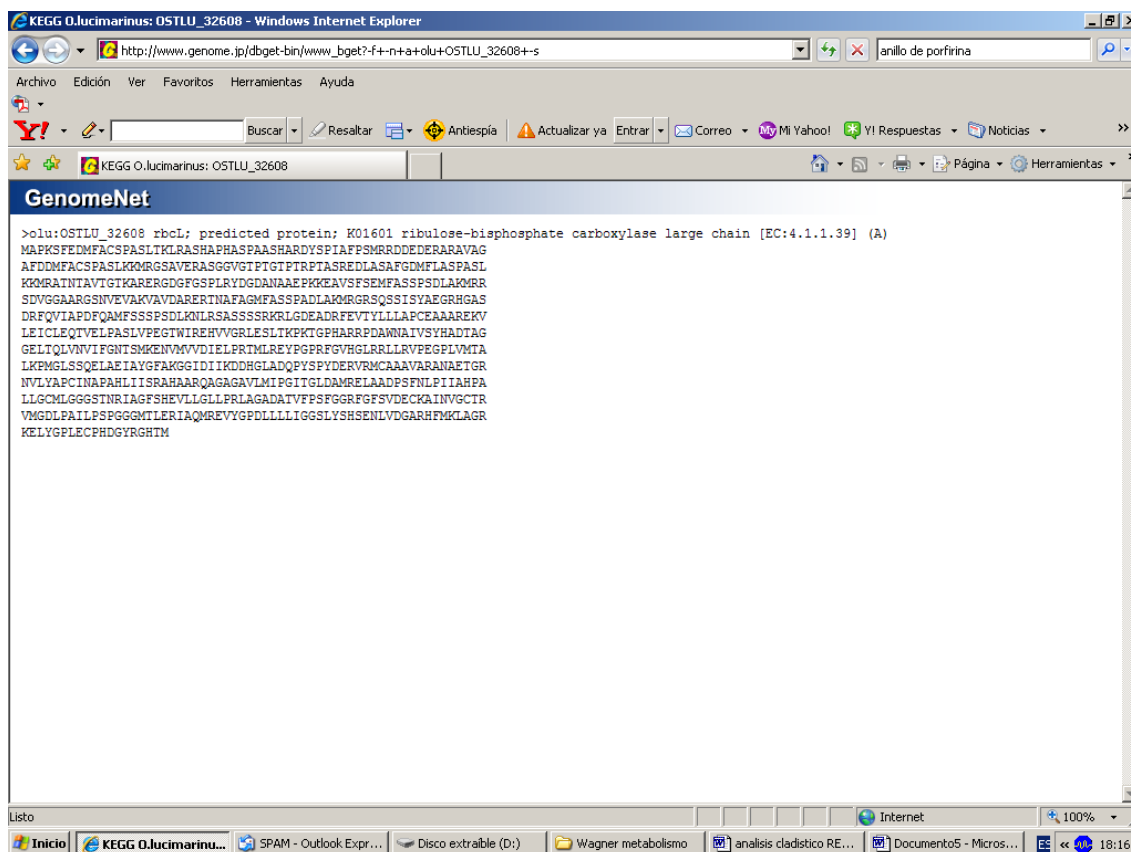


Figura 8. Secuencia de aminoácidos de la proteína codificada por el gen *rbcL* de *Ostreococcus lucimarinus*.

Otras bases a las que se puede acceder y obtener datos e información son las siguientes:

ExpASY Proteomics Server (<http://www.expasy.ch/>)

European Bioinformatics Institute (<http://www.ebi.ac.uk/>)

National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>)

Genbank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>).

ALINEAMIENTO DE SECUENCIAS

El conjunto de datos, caracteres y taxones, se ordena en una **matriz básica de datos**. Cuando se trata de caracteres moleculares, cada elemento de la secuencia, base de ácidos nucleicos o aminoácido de proteínas, constituye un carácter. Así, en las secuencias *rbcL* cada aminoácido es un carácter, de manera que la matriz básica de datos se construye con *n* secuencias pertenecientes a otros tantos taxones, y tantos caracteres como aminoácidos tiene la secuencia. Ahora bien, ocurre que la misma proteína (homóloga) en diferentes especies muestra variaciones en la secuencia (tanto en longitud como en los aminoácidos que la constituyen). Las diferencias en la propia secuencia son el objeto de análisis y su estudio permitirá agrupar y clasificar las especies a las que

pertenecen. Sin embargo, las diferencias en la longitud de la secuencia pueden “eliminarse” introduciendo huecos (“gaps”) en la propia secuencia. Los programas de **alineamiento de secuencias** como **Clustalw** (<http://align.genome.jp/>) realizan esta tarea buscando los alineamientos óptimos con el menor número de huecos, es decir, aquellos que maximizan la semejanza entre secuencias. El resultado de este alineamiento se puede escribir en distintos formatos. En el caso que nos ocupa, el formato de salida se denomina “phylip” y consiste en una matriz apropiada para los programas de reconstrucción filogenética que se utilizarán y que están contenidos en el paquete de programas de inferencia filogenético denominado **PHYLIP** (libre distribución).

La matriz básica de datos que contiene n secuencias (n taxones o especies) y m caracteres (m aminoácidos o posiciones) refleja en la primera línea estos números (esta configuración es obligada ya que los programas que se utilizarán sólo leen este formato). Existen diferentes formas (“formatos”) de escribir las secuencias y para esta tarea existen **programas de conversión de secuencias** (<http://www.bimas.cit.nih.gov/molbio/readseq/>).

A modo de ejemplo, las secuencias alineadas con ClustalW y la matriz de datos de rbcl es la siguiente (Fig. 9):

```

                10         20         30         40         50
24  752  .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
ath_ArthCp -----
osa_313146 -----
ppp_Ph papa -----
cre_ChreCp -----
tel_tll150 -----
ana_alr152 -----
gvi_glr215 -----
syn_slr000 -----
syc_syc013 -----
neu_NE1921 -----
bbt_BBta_2 -----
syw_SYNW17 -----
pma_Pro055 -----
sme_SMb201 -----
bbt_BBta_0 -----
nph_NP2770 -----
pab_PAB158 -----
afu_AF1638 -----
mja_MJ1235 -----
rci_RCIX22 -----
afu_AF1587 -----
cte_CT1772 -----
ret_RHE_PF -----
olu_OSTLU_ MAPKSEFEDMFACSPASLTKLRASHAPHASPAASHARDYSPIAFPSMRRDD
                60         70         80         90        100
ath_ArthCp -----
osa_313146 -----
ppp_Ph papa -----
cre_ChreCp -----
tel_tll150 -----

```

```

ana_alr152 -----
gvi_glr215 -----
syn_slr000 -----
syc_syc013 -----
neu_NE1921 -----
bbt_BBta_2 -----
syw_SYNW17 -----
pma_Pro055 -----
sme_SMb201 -----
bbt_BBta_0 -----
nph_NP2770 -----
pab_PAB158 -----
afu_AF1638 -----
mja_MJ1235 -----
rci_RCIX22 -----
afu_AF1587 -----
cte_CT1772 -----
ret_RHE_PF -----
olu_OSTLU_ EDERARAVAGAFDDMFACSPASLKKMRGSAVERASGGVGTPTGTPTRPTA

```

110 120 130 140 150

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

```

ath_ArthCp -----
osa_313146 -----
ppp_Phpapa -----
cre_ChreCp -----
tel_tll150 -----
ana_alr152 -----
gvi_glr215 -----
syn_slr000 -----
syc_syc013 -----
neu_NE1921 -----
bbt_BBta_2 -----
syw_SYNW17 -----
pma_Pro055 -----
sme_SMb201 -----
bbt_BBta_0 -----
nph_NP2770 -----
pab_PAB158 -----
afu_AF1638 -----
mja_MJ1235 -----
rci_RCIX22 -----
afu_AF1587 -----
cte_CT1772 -----
ret_RHE_PF -----
olu_OSTLU_ SREDLASAFGDMFLASPASLKKMRATNTAVTGTKARERGDGFGSPLRYDG

```

160 170 180 190 200

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

```

ath_ArthCp -----
osa_313146 -----
ppp_Phpapa -----
cre_ChreCp -----
tel_tll150 -----
ana_alr152 -----
gvi_glr215 -----
syn_slr000 -----
syc_syc013 -----
neu_NE1921 -----
bbt_BBta_2 -----

```



```

syw_SYNW17 -----
pma_Pro055 -----
sme_SMb201 -----
bbt_BBta_0 -----
nph_NP2770 -----
pab_PAB158 -----
afu_AF1638 -----
mja_MJ1235 -----
rci_RCIX22 -----
afu_AF1587 -----
cte_CT1772 -----
ret_RHE_PF -----
olu_OSTLU_  DANAAEPKKEAVSFSEMFASSPSDLAKMRRSDVGGAARGSNVEVAKVAVD

                210         220         230         240         250
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
ath_ArthCp -----MSPQT
osa_313146 -----MSPQT
ppp_Phpapa -----MSPRP
cre_ChreCp -----MVPQT
tel_tll150 -----MAYTQ
ana_alr152 -----MSYAQT
gvi_glr215 -----MSYT
syn_slr000 -----
syc_syc013 -----MP
neu_NE1921 -----
bbt_BBta_2 -----
syw_SYNW17 -----
pma_Pro055 -----
sme_SMb201 -----MNADAKT
bbt_BBta_0 -----MNEALKSL
nph_NP2770 -----
pab_PAB158 -----
afu_AF1638 -----
mja_MJ1235 -----
rci_RCIX22 -----
afu_AF1587 -----
cte_CT1772 -----
ret_RHE_PF -----
olu_OSTLU_  ARERTNAFAGMFASSPADLAKMGRSQSSISYAEGRHGASDRFQVIAPDF

                260         270         280         290         300
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
ath_ArthCp  ETKASVGFKAGVKYKLT-YYTPEYETKDTDILAAFRVTPQPGVPPEEEAG
osa_313146  ETKASVGFKAGVKDYKLT-YYTPEYETKDTDILAAFRVTPQPGVPPEEEAG
ppp_Phpapa  EIKAGVGFKAGVKDYRLT-YYTPDYQTKDTDILAAFRMTPQPGVPAEECCG
cre_ChreCp  ETKAGAGFKAGVKDYRLT-YYTPDYVVRDTDILAAFRMTPQLGVPPEECCG
tel_tll150  SKSQKVGYQAGVKDYRLT-YYTPDYTPKTDILAAFRVTPQPGVPFEEAA
ana_alr152  KTQTKSGYKAGVQDYRLT-YYTPDYTPKTDILAAFRVTPQPGVPFEEAA
gvi_glr215  KTQAKAGYQAGVKDYRLT-YYTPDYTPKTDVLAAFRVTPQPGVPIEEAG
syn_slr000  MVQAKAGFKAGVQDYRLT-YYTPDYTPKTDLLACFRMTPQPGVPAEEAA
syc_syc013  KTQSAAGYKAGVKDYKLT-YYTPDYTPKTDLLAAFRFSPQPGVPADEAG
neu_NE1921  --MSAKTYNAGVKYRHT-YWEPHYNVQDTDILACFKIVPQPGVDREEAA
bbt_BBta_2  --MAEKSYQAGVKYRKT-YWTPDYVPLDTDLLAVFKIVAQAGVPREEAA
syw_SYNW17  ---MSKKYDAGVKYRDT-YWTPDYVPLDTDLLACFKCTGQEGVPKEEVA
pma_Pro055  ---MSKKYDAGVKYRDT-YWTPDYVPLDTDLLACFKCTGQEGVPPREEVA
sme_SMb201  EIKGRERYKAGVLKYAQMGYWNGDYEPKDTDLIALFRITPQDGVDPIEAA
bbt_BBta_0  TVTGKERYKSGVLEYKRMGYWEPDYEPKDTDVIALFRVTPQNGVDPIEAS
nph_NP2770  -----MEYADFLDESYEPSDDDLVCTFRLVPGEGISVADAA
pab_PAB158  -----MVSSMKVEWYLDFVDLNYEPGRDELIVEYYFEPN-GVSPEEEAA

```

```

afu_AF1638 -----MAEFEIYREYVDKSYEPQKDDIVAVFRITPAEGFTIEDAA
mja_MJ1235 -----MDYINLNYPNEGDLLSCMVIKGE---NLEKLA
rci_RCIX22 -----MAIHGDKDLIIQVPLTLINIMTMVVRTTYVVEADAPIAKVA
afu_AF1587 MQLGVRLRFQKFEYPEANPEALPEGIDPEEYIIGTYYSFPGKGMNPFEIT
cte_CT1772 -----MNAEDVKGFFASRESLDMEQYLVLDYYLESVGDIIETAL
ret_RHE_PF -----MITLTYRIETPGSVETMA
olu_OSTLU_ QAMFSSSSPDLKNLRSASSSSSRKRLGDEADRFVYTYLLAPCEAAAREKV

          310          320          330          340          350
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
ath_ArthCp AAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVPGEETQ-----
osa_313146 AAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVPGEDNQ-----
ppp_Ph papa AAVAAESSTGTWTTVWTDGLTSLDRYKGRCYAIEPVAGEENQ-----
cre_ChreCp AAVAAESSTGTWTTVWTDGLTSLDRYKGRCYDIEPVPGEDNQ-----
tel_tll150 AAVAAESSTGTWTTVWTDLLTDLDRYKGCCYDIEPLPGEDNQ-----
ana_alr152 AAVAAESSTGTWTTVWTDLLTDLDRYKGRCYDIEPVPGEDNQ-----
gvi_glr215 AAVAAESSTGTWTTVWTDGLTELDRYKGRCYDIEPVPGEDNQ-----
syn_slr000 AAVAAESSTGTWTTVWTDNLTDLDRYKGRCYDLEAVPNEDNQ-----
syc_syc013 AAIAAESSTGTWTTVWTDLLTDMDRYKGRCYHIEPVQGEENS-----
neu_NE1921 AAVAAESSTGTWTTVWTDLLTDLDYKGRSYRIEDVPGDDSS-----
bbt_BBta_2 AAVAAESSTGTWTTVWTDLLTDLDYKGRAYRIEVPVGDDNA-----
syw_SYNW17 AAVAAESSTGTWSTVWSELLTDLDYKGRCYRIEDVPGDKES-----
pma_Pro055 AAVAAESSTGTWSTVWSELLTDLEFYKGRCYRIEDVPGDKES-----
sme_SMb201 AAVAGESSTATWTVVWTDRLTACDQYRAKAYRVDVPVPGTPGQ-----
bbt_BBta_0 AAVAGESSTATWTVVWTDRLTAAEKYRAKCYRVDVPVNTPGS-----
nph_NP2770 ARVASESSNGTWAALSPESD---VRQYSALACDIGPEDEHG-----
pab_PAB158 GRIASESSIGTWTTLWKLPE---MAKRSMKVFYLEKHGEG-----
afu_AF1638 GAVAAESSTGTWTSLHPWYDEERVKGLSAKAYDFVDLGDGS-----
mja_MJ1235 NIIAGESSIGTWTKVQTMKS---DIYEKLRPKVYEIKEIGEENGY-----
rci_RCIX22 KEIAAEQSTGTWTEVAAEKEVHEKLGAVVSAEG-----
afu_AF1587 QVLALEQSTGTWLPVPGETPEVRRKHVAKVGVYIIPDYEIMV-----
cte_CT1772 AHEFCSEQSTAQWKRVGVEDDFRLVHAAKVIDYEVIEELEQLSYPVKHSET
ret_RHE_PF DKIASDQSTGTFFVPGETEELKSRVAARVLAIRPLENARHPTWPESAPD
olu_OSTLU_ LEICLEQTVELPASLVPEGTWIREHVVGRLLESLTKPKTGPHAR-----

          360          370          380          390          400
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
ath_ArthCp -----FIAYVAYPLDLFEEGSVTNMFTSIVGNVFG-FKALAAALRLEDLRI
osa_313146 -----YIAYVAYPLDLFEEGSVTNMFTSIVGNVFG-FKALRALRLEDLRI
ppp_Ph papa -----YIAYVAYPLDLFEEGSVTNLFTSIVGNVFG-FKALRALRLEDLRI
cre_ChreCp -----YIAYVAYPIDLFEEGSVTNMFTSIVGNVFG-FKALRALRLEDLRI
tel_tll150 -----FIAYIAYPLDLFEEGSVTNMLTSIVGNVFG-FKALKALRLEDLRI
ana_alr152 -----FIAYIAYPLDLFEEGSITNVLTSIVGNVFG-FKALRALRLEDIRF
gvi_glr215 -----WICYIAYPLDLFEEGSVTNVLTSIVGNVFG-FKALRALRLEDIRF
syn_slr000 -----YFAFIAYPLDLFEEGSVTNVLTSIVGNVFG-FKALRALRLEDIRF
syc_syc013 -----YFAFIAYPLDLFEEGSVTNILTSIVGNVFG-FKAIRSLRLEDIRF
neu_NE1921 -----FYAFIAYPIDLFEEGSVVNVLTSLVGNVFG-FKAVRSLRLEDVRF
bbt_BBta_2 -----FYAFIAYPIDLFEEGSITNVLTSIVGNVFG-FKAVRSLRLEDIRF
syw_SYNW17 -----FYAFIAYPLDLFEEGSITNVLTSIVGNVFG-FKALRHLRLEDIRF
pma_Pro055 -----FYAFIAYPLDLFEEGSITNVLTSIVGNVFG-FKALRHLRLEDIRF
sme_SMb201 -----YFCYVAYDLILFEEGSIANLTASIIIGNVFS-FKPLKAARLEDMRL
bbt_BBta_0 -----YFAYIAYDLDLFEPGSIANLSASIIIGNVFG-FKPLKALRLEDMRF
nph_NP2770 -----TQVTVAYPSGLFEDGSLPQILSCIAGNIMG-MKAVETIRLLDCEW
pab_PAB158 -----YIAKIAYPLTLFEEGSLVQLFSAIAGNVFG-MKALKNLRLLDLDFHP
afu_AF1638 -----SIVRIAYPSELFEPHNMPGLLASIAGNVFG-MKRVKGLRLEDLQL
mja_MJ1235 ---KVGLIKIAYPLYDFEINNMPGVLGAGIAGNIFG-MKIAKGLRILDFRF
rci_RCIX22 -----NTVVIDFVVEIFEPDNVPQILSVVAGNLFGLGGLKACRLMDVDF
afu_AF1587 --PQEVDRNFIVQIAFPWRNIGSKLSMLFSTVVGNI SMAPKLLDLRF
cte_CT1772 GKIHACRVTIAHPCNF GP-KIPNLLTAVCGEGTYFTPGVPVVKLMDIHF
ret_RHE_PF TLLHRADV DIAFPLEAIGTDLSALMTIAIGG--VYSIKGMTGIRIVDMKL
    
```

```

olu_OSTLU_  -----RPDAWNAIVSYHADTAGGELTQLVNVIFGNTSMKENVMVVDIEL
                410          420          430          440          450
                ....|....|....|....|....|....|....|....|....|....|....|
ath_ArthCp    PPAYTKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVY
osa_313146    PPTYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRACY
ppp_Phpapa    PPAYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVY
cre_ChreCp    PPAYVKTFVGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVY
tel_tll150    PVAYLKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVY
ana_alr152    PVAYIKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVY
gvi_glr215    PIALVKTYQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVY
syn_slr000    PVALIKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVY
syc_syc013    PVALVKTFQGPPHGIQVERDLLNKYGRPMLGCTIKPKLGLSAKNYGRAVY
neu_NE1921    PIAYVKTCGGPPNGIQVERDILNKYGRAYLGCTIKPKLGLSAKNYGRAVY
bbt_BBta_2    PLAYVKTCGGPPNGIQLERDRNLKYGRPLLGCTIKPKLGLSAKNYGRAVY
syw_SYNW17    PMAFIKSCYGPPNGIQVERDRMNKYGRPLLGCTIKPKLGLSGKNYGRVY
pma_Pro055    PMAFIKTCGGPPQGIVVERDRLNKYGRPLLGCTIKPKLGLSGKNYGRVY
sme_SMb201    PVAYVKTFRGPPTGIVVERERLDFKFKPLLGATTKPKLGLSGKNYGRVY
bbt_BBta_0    PVAYVKTFQGPATGIVVERERLDFKGRPLLGATVTKPKLGLSGRNYGRVY
nph_NP2770    PAVIARSFPGPQYGSDVRTELLDAGDRPPLATVPKPKVGLSTEEHVSVAE
pab_PAB158    PYEYLRHFFKGPQFGVKGIREFMGIKDRPLTATVPKPKMGWSVEEYAEIAY
afu_AF1638    PKSEFLKDFKGPSKGKEGVKKIFGVADRPIVGTVPKPKVGYSAEEVEKLAY
mja_MJ1235    PAEFVKAYKGPRFFGIEGVRETLKIKERPLLGTIVKPKVGLKTEEHAKVAY
rci_RCIX22    G-PLTKYYNGPEFGIEEVRRKILGVYDRPLVGTIIKPKVGLSPKRTAEVAE
afu_AF1587    PKEFVKGFFKGPKFFGIEGVRDVLGVKDRPLLNNMIKPDVYSPPDLGAKLAY
cte_CT1772    PDTYLADFEGPKFFGIEGLRDILNAHGRPIFFGVVKNIGLSPGEFAEIAY
ret_RHE_PF    PEAFRSAHPGPQFGIAGSRRLTGVEGRPIIGTIVKPALGLRPHETAELVG
olu_OSTLU_    PRTMLREYPGPRFGVHGLRLLRVPEGLVMTALKP-MGLSSQELAEIAY
                460          470          480          490          500
                ....|....|....|....|....|....|....|....|....|....|....|
ath_ArthCp    ECLRGGLDFTKDDENVNSQPFMRWRDRFLFCAEAIYKSQAETGEIKGHYL
osa_313146    ECLRGGLDFTKDDENVNSQPFMRWRDRFVFCAEAIYKSQAETGEIKGHYL
ppp_Phpapa    ECLRGGLDFTKDDENVNSQPFMRWRDRFLFCAEAIYKSQGETGEIKGHYL
cre_ChreCp    ECLRGGLDFTKDDENVNSQPFMRWRDRFLFVAEAIYKAQAETGEVKGHYL
tel_tll150    ECLRGGLDFTKDDEINSQPFQWRWRDRFLFVADAIHKAQAETGEIKGHYL
ana_alr152    ECLRGGLDFTKDDEINSAPFQWRWRDRFLFVADAITKAQAETGEIKGHYL
gvi_glr215    ECLRGGLDFTKDDEINSQPFMRWRDRFLFVQDAIVKSQAETGEIKGHYL
syn_slr000    ECLRGGLDFTKDDEINSQPFMRWRDRFLFVQEAIEKAQAETNEMKGHYL
syc_syc013    ECLRGGLDFTKDDEINSQPFQWRWRDRFLFVADAIHKSQAETGEIKGHYL
neu_NE1921    ECLRGGLDFTKDDENVNSQPFMRWRQRFDFVMEAIHKAERETGERKGHYL
bbt_BBta_2    ECLRGGLDFTKDDEINSQPFMRWQHRFEFVMEAVHKATSETGERKGHYL
syw_SYNW17    ECLRGGLDFTKDDEINSQPFQWRQNRFFVAEAIKLSQETGERKGHYL
pma_Pro055    ECLRGGLDLTKDDEINSQPFQWRDRFFVAEAVKLAQOETGEVKGHYL
sme_SMb201    EGLKGGLDFMKDDEINSQPFMHWRDRYLCMEAVNHASAVTGEVKGHYL
bbt_BBta_0    EALKGGLDFTKDDENTNSQPFMHWRDRFLCMEAVNKAQAATGEVKGTYL
nph_NP2770    SAWRGGVDLLKDDENLTDQTFNPFEQRVADSFAADRLEEETGERKDYLV
pab_PAB158    ELWSGGIDLLKDDENFTSFPFNRFEERVKKLYRVRDRVEAETGETKEYLI
afu_AF1638    ELLSGGMDYIKDDENLTSPAYCRFEERAERIMKVIEKVEAETGEKKSWF
mja_MJ1235    EAWVGGVDLVKDDENLTSQEFNKFEDRIYKTLEMRDKAEETGERKAYMP
rci_RCIX22    QAALGGLDLIKDDETLTDQKFCPLEERLTMVMDRLHKVEDRIGKPCFYAV
afu_AF1587    EVARGGVDIIKDDELLANPEFNRIEERVPKFMEAIDRADEEKGEKTYAV
cte_CT1772    QSWLGGLDIAKDDEMLADVTWSSIEERAHLGKARRKAEAETGEPKIYLA
ret_RHE_PF    ELIGSGVDFIKDDEKLMSPAYSPLKERVAAIMPRILDHEQKTGKKVMYAF
olu_OSTLU_    GFAKGGIDIIKDDEHGLADQPYSPYDERVRMCAAAVARANAETGRNVLYAP
                510          520          530          540          550
                ....|....|....|....|....|....|....|....|....|....|....|
ath_ArthCp    NATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLSHYCRDNGLL
osa_313146    NATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLAHYCRDNGLL

```

```

ppp_Ph papa      N A T A G T C E E M I K R A Q F A R E L G M P I V M H D Y L T G G F T A N T S L A H Y C R D N G L L
cre_ChreCp      N A T A G T C E E M M K R A V C A K E L G V P I I M H D Y L T G G F T A N T S L A I Y C R D N G L L
tel_tll150      N V T A P T C E E M L K R A E F A K E L E M P I I M H D F L T A G F T A N T T L S K W C R D N G M L
ana_alr152      N V T A P T C E E M L K R A E Y A K E L K Q P I I M H D Y L T A G F T A N T T L A R W C R D N G V L
gvi_glr215      N C T A G T C E E M M E R A E F A K E L K T P I I M H D Y L T G G F T A N T T L A K W C R R N G I L
syn_slr000      N V T A G T C E E M M K R A E F A K E I G T P I I M H D F F T G G F T A N T T L A R W C R D N G I L
syc_syc013      N V T A P T C E E M M K R A E F A K E L G M P I I M H D F L T A G F T A N T T L A K W C R D N G V L
neu_NE1921      N V T A P T P E E M F K R A E Y A K E L K A P I I M H D Y I A G G F C A N T G L A N W C R D N G I L
bbt_BBta_2      N V T A P T P E E M Y K R A E F A K S L G A P I I M H D F L T A G F T A N T G L A N W C R E N G M L
syw_SYNW17      N V T A N T P E E M Y E R A E F A K E L G M P I I M H D F I T G G F T A N T G L S K W C R K N G M L
pma_Pro055      N C T A T T P E E M Y E R A E F A K E L D M P I I M H D Y I T G G F T A N T G L A N W C R K N G M L
sme_SMb201      N I T A G T M E E M Y R R A E F A K E L G S V I V M V D L V I G - W T A I Q S I S E W C R Q N D M I
bbt_BBta_0      N V T A A T M E D M Y E R A E F A K E L G S V I I M I D L V I G - Y T A I Q S M A K W A R R N D M I
nph_NP2770      N I T A - E T D E M V R R A E F V D D H G G S F V M V D I I T C G W S G L Q T V R R R T E D L D L A
pab_PAB158      N I T G - F V N V M E K R A E L V A N E G G Q Y V M I D I V V A G W S A L Q Y M R E V T E D L G L A
afu_AF1638      N I T A - D V R E M E R R L K L V A E L G N P H V M V D V V I T G W G A L E Y I R D L A E D Y D L A
mja_MJ1235      N I T A - P Y R E M I R R A E I A E D A G S E Y V M I D V V V C G F S A V Q S F R E - - E D F K F I
rci_RCIX22      N V T C G A D M I V E R - A E R A V E L G A N M V M V D I L T A G F S A V Q A L T D - - E K I G V P
afu_AF1587      N V T A D L P E V L E N - A E R A I E L G A N C L L V N Y L A T G F P V L R A L A E D E S I K V P I
cte_CT1772      N I T D E V D S L M E K H D V A V R N - G A N A L L I N A L P V G L S A V R M L S N - - - Y T Q V P
ret_RHE_PF      G I S H A D P D E M M R N H D I V A A A G G N C A V V N I N S I G F G G M S F L R K - - - R S S L V
olu_OSTLU_      C I N A P A H L I I S R A H A A R Q A G A G A V L M I P G I T G L D A M R E L A A D P S F N L P I I

                560          570          580          590          600
...|...|...|...|...|...|...|...|...|...|...|
ath_ArthCp      L H I H R A - - - M H A V I D R Q K N H G M H F R V L A K A L R L S G G D H I H A G T V V G - K L E
osa_313146     L H I H R A - - - M H A V I D R Q K N H G M H F R V L A K A L R M S G G D H I H A G T V V G - K L E
ppp_Ph papa      L H I H R A - - - M H A V L D R Q K N H G M H F R V L A K A L R L S G G D H I H S G T V V G - K L E
cre_ChreCp      L H I H R A - - - M H A V I D R Q N H G I H F R V L A K A L R M S G G D H L H S G T V V G - K L E
tel_tll150      L H I H R A - - - M H A V M D R Q K N H G I H F R V L A K C L R M S G G D H I H T G T V V G - K L E
ana_alr152      L H I H R A - - - M H A V I D R Q K N H G I H F R V L A K A L R L S G G D H I H T G T V V G - K L E
gvi_glr215      L H I H R A - - - M H A V I D R Q K N H G I H F R V L A K C L R L S G G D H I H T G T V V G - K L E
syn_slr000      L H I H R A - - - M H A V V D R Q K N H G I H F R V L A K C L R L S G G D H L H S G T V V G - K L E
syc_syc013      L H I H R A - - - M H A V I D R Q N H G I H F R V L A K C L R L S G G D H L H S G T V V G - K L E
neu_NE1921      L H I H R A - - - M H A V I D R N P H H G I H F R V L A K M L R L S G G D H L H S G T V V G - K L E
bbt_BBta_2      L H I H R A - - - M H A V L D R N P M H G I H F R V L T K C L R L S G G D H L H S G T V V G - K L E
syw_SYNW17      L H I H R A - - - M H A V I D R H P K H G I H F R V L A K C L R L S G G D Q L H T G T V V G - K L E
pma_Pro055      L H I H R A - - - M H A V I D R H P K H G I H F R V L A K C L R L S G G D Q L H T G T V V G - K L E
sme_SMb201      L H M H R A - - - G H G T Y T R Q K N H G I S F R V I A K W L R L A G V D H L H A G T A V G - K L E
bbt_BBta_0      L H L H R A - - - G H G T Y T R Q K S H G V S F R V I A K W M R L A G V D H I H A G T V V G - K L E
nph_NP2770      I H A H R A - - - M H A A F D R L P Q H G V S M R C L A Q F A R L A G V D H I H T G T A G - - - - -
pab_PAB158      I H A H R A - - - M H A A F T R N P K H G I T M F A L A K A A R M I G V D Q I H T G T A V G - - - -
afu_AF1638      I H G H R A - - - M H A A F T R N A K H G I S M F V L A K L Y R I I G I D Q L H I G T A G A G K L E
mja_MJ1235      I H A H R A - - - M H A A M T R S R D F G I S M L A L A K I Y R L L G V D Q L H I G T V V G K M E G
rci_RCIX22      I H I H R T - - - M H G A L T R G - K Y G I A M P V I S K L T R M C G G T N L H T G T Y A G - - - -
afu_AF1587      M A H M D V - - - A G A Y Y V S P I S G V R S T L I L G K L P R L A G A D I V V Y P A P Y G - - - -
cte_CT1772      L I G H F P - - - F I A S F S R M E K Y G I H S K V M T K L Q R L A G L D A V I M P G F G D - - - -
ret_RHE_PF      L H A H R N - - - G W D V L T R D P G A G M D F K V Y Q Q F W R L L G V D Q F Q I N G I R I - - - -
olu_OSTLU_      A H P A L L G C M L G G G S T N R I A G F S H E V L L G L L P R L A G A D A T V F P S F G G - - - -

                610          620          630          640          650
...|...|...|...|...|...|...|...|...|...|...|
ath_ArthCp      G D R E S T L G F V D L L R D D Y V E K D R S R G I F F T Q D W V S L P G V L P V A S G G I H V W H
osa_313146     G E R E M T L G F V D L L R D D F I E K D R A R G I F F T Q D W V S M P G V I P V A S G G I H V W H
ppp_Ph papa      G E R Q V T L G F V D L L R D D Y I E K D R S R G I Y F T Q D W V S L P G V L P V A S G G I H V W H
cre_ChreCp      G E R E V T L G F V D L M R D D Y V E K D R S R G I Y F T Q D W C S M P G V M P V A S G G I H V W H
tel_tll150      G D K A V T L G F V D L L R E N Y I E Q D R S R G I Y F T Q D W A S M P G V M A V A S G G I H V W H
ana_alr152      G E R G I T M G F V D L L R E N Y V E Q D K S R G I Y F T Q D W A S L P G V M A V A S G G I H V W H
gvi_glr215      G E R A S T M G F V D L L R E E H V E R D L S R G I Y F T Q D W A S M P G V M A V A S G G I H V W H
syn_slr000      G E R G I T M G F V D L M R E D Y V E E D R S R G I F F T Q D Y A S M P G T M P V A S G G I H V W H
    
```

```

syc_syc013      GDKASTLGFVDLMREDHIEADRSRGVFFFTQDWASMPGVLPVASGGIHVWH
neu_NE1921      GDREATLGWIDIMRDSFIKEDRSRGIMFDQDWGSMGVVPVASGGIHVWH
bbt_BBta_2      GDREATIGWVDLMREPFVPENRARGIFFDQDWGAMPGVMPVASGGIHVWH
syw_SYNW17      GDRQTTLGYIDQLRESFVPEDRSRGNFFDQDWGSMGVFAVASGGIHVWH
pma_Pro055      GDRQTTLGYIDNLRESFVPEDRTRGNFFDQDWGSMGVFAVASGGIHVWH
sme_SMb201      GDEPPTVQGYYNVCREMKNEVDLPRGLFFEQDWADLKKVMPVASGGIHAGQ
bbt_BBta_0      GDEPNTTRGYDICREDHNPMALEYGLFFEQHWASLNKLMVPVASGGIHAGQ
nph_NP2770      -----L GKLENETAGINEWLRSDLHGHS DVLVPVASSGGLHPGI
pab_PAB158      -----K MAGDYEEIKKINDFLLSKWEHIRPVFPVASSGGLHPGL
afu_AF1638      GQKWDTVQNARIFSEVEYTPDEGDAFHLSQNFHHIKPAMPVSSGGLHPGN
mja_MJ1235      GEK--EVKAIKDEIVYDKVEADNENKFFNQDWFDIKPVFPVSSGGVHPRL
rci_RCIX22      -----K MERNVCEIDASRDILRKPWAGYKRVWVPVSSGGYLPQK
afu_AF1587      -----K APMMEEKYIEVAKQHRYPFYHIKPCFPMPSSGGIAPIM
cte_CT1772      -----R MMTPEEEVLENVIECTKPMGRIPCLPVPGGSDSALT
ret_RHE_PF      -----K YWEPDESEFVSSFKAVSTPLFDAADCPLPVAGSGQWGGQ
olu_OSTLU_      -----R FGFSSVDECKAINVGC TRVMGDLPAILPSPGGGMTLER

                660          670          680          690          700
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
ath_ArthCp      MPALTEIFG--D DSVLQFGGGT LGHPWGNAPGAVANRVALEACVQARNEG
osa_313146      MPALTEIFG--D DSVLQFGGGT LGHPWGNAPGAAANRVALEACVQARNEG
ppp_Phpapa      MPALTEIFG--D DSVLQFGGGT LGHPWGNAPGAVANRVALEACVQARNEG
cre_ChreCp      MPALVEIFG--D DACLQFGGGT LGHPWGNAPGAAANRVALEACTQARNEG
tel_tll150      MPALVDIFG--D DAVLQFGGGT LGHPWGNAPGATANRVALEACIQARNEG
ana_alr152      MPALVEIFG--D DSVLQFGGGT LGHPWGNAPGATANRVALEACVQARNEG
gvi_glr215      MPALLDIFG--D DAVLQFGGGT LGHPWGNAPGATANRVALEACVKARNEG
syn_slr000      MPALVEIFG--D DSCLQFGGGT LGHPWGNAPGATANRVALEACVQARNEG
syc_syc013      MPALVEIFG--D DSVLQFGGGT LGHPWGNAPGATANRVALEACVQARNEG
neu_NE1921      MPALVTIFG--D DACLQFGGGT LGHPWGNAAAGAAANRVALEACVEARNRG
bbt_BBta_2      MPALTAIFG--D DACFQFGGGT LGHPWGNAAAGAHANRVALEACVEARNQG
syw_SYNW17      MPALVAIFG--D DSVLQFGGGT THGHPWGSAAAGAAANRVALEACVKARNAG
pma_Pro055      MPALLAIFG--D DSCLQFGGGT THGHPWGSAAAGAAANRVALEACVKARNAG
sme_SMb201      MHQLLDLFG--D DVVLQFGGGT IGHPMGIQAGATANRVALEAMVLARNEG
bbt_BBta_0      MHQLLNLYG--D EDVVLQFGGGT IGHPLGIQAGATANRVALEAMILARNEG
nph_NP2770      VDQLLDALG--D TNVMVQAGGGI HGHDPDGEAGARALRAAVDAYADGES--
pab_PAB158      MPELIRLFG--D KDLVIQAGGGV MGHDPDGPRAKALRDAIDAAIEGLD--
afu_AF1638      LEPVIDALG--D KEIIVIQAGGGV LGHMPGAKAGAKAVRQALDAIISAIP--
mja_MJ1235      VPKIVEILG--D RDLIIQAGGGV HGHDPDGTTRAGAKAMRAAIEAIEGKS--
rci_RCIX22      VRENDCYG--D IDVILQAGGGI HGHPEGTTVGVKAMFQAVEAWQQQKT--
afu_AF1587      VPKLVNTLG--D KDFVVAAGGGI HAHDPDGPAAAGARAFRQIDAAMQGYTDL
cte_CT1772      LQTVYEKVG N-VDFGFV PGRGVFGHPMGPKAGAKSIRQAWEAIEQGIS--
ret_RHE_PF      APETYERTGR TIDL LLYLCGGGIVSHPGGPAAGVRAVQQAQAAVAGIP--
olu_OSTLU_      IAQMREVYG--D PDL LLLIGGSLYSHSEN LVDGARHFMKLAGRKE-----

                710          720          730          740          750
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
ath_ArthCp      RDLAVEGNEI IREACKWSP ELAAACEVWKEITFNFPPTIDKLDGQE-----
osa_313146      RDLAREGNEI IRSACKWSP ELAAACEIWKAIKFEFEPVDKLDSE-----
ppp_Phpapa      RDLAREGNEI IREAAKWSP ELAAACEVWKEIKFEFDTVDL-----
cre_ChreCp      RDLAREGGD VIRSACKWSP ELAAACEVWKEIKFEFDTIDKL-----
tel_tll150      RDLMREGGD IIREAARWSP ELAAACELWKEIKFEFEAQDTI-----
ana_alr152      RNLAREGND VIREAAKWSP ELAVACELWKEIKFEFEAMDTV-----
gvi_glr215      RDLMREAGD IIREAARWSP ELAAACELWKEIKFEYEAVDKL-----
syn_slr000      RNLAREGND VIREACRWSP ELAAACELWKEIKFEFEAMDTL-----
syc_syc013      RDLYREGGD ILREAGKWSPE LAAAALDWKEIKFEFETMDKL-----
neu_NE1921      VPIEKEGKAIL TEAAKHSPE LKIAMETWKEIKFEFDTVDKLDVAHK----
bbt_BBta_2      REVEREGRE ILLTEAAQHSPE LKIAMETWKEIKFEFDTVDKLDGTPMLRVV
syw_SYNW17      REIEKESRD ILM EAGKHSPE LAIALETWKEIKFEFDTVDKLDVQN-----
pma_Pro055      REIEKESRD ILM EAAKHSPE LAIALETWKEIKFEFDTVDKLDVQ-----
sme_SMb201      RDIAHEGPE IILRAAAKWC KPLEAALDIWGNISFNYPPTDTSDFVPSVTAA
    
```

```

bbt_BBta_0  RDYVHEGPEILAKAAATCTPLKQALDVWKNVTFNYDSTDTPTDFVPTAAVT
nph_NP2770  -----LDSRAESVPAALRTALDEWGTQNP-----
pab_PAB158  -----LEEKAKSSPELKKALDKWGYLKPK-----
afu_AF1638  -----LEEHAKQHPELQAALDKWGRVTPI-----
mja_MJ1235  -----LEEKAEVVAELKKALEYWK-----
rci_RCIX22  -----LEEYAKTHKELAGALKQWGPSQ-----
afu_AF1587  R-----KYAEENNLOELLKALQL-----
cte_CT1772  -----IETWAETHPELQAMVDQSLLKKQD-----
ret_RHE_PF  -----LEVYAKDHPELAASIAKFSDBGKA-----
olu_OSTLU_  -----LYGPLECPHDCYRGHTM-----

ath_ArthCp  --
osa_313146  --
ppp_Phpapa  --
cre_ChreCp  --
tel_t11150  --
ana_alr152  --
gvi_glr215  --
syn_slr000  --
syc_syc013  --
neu_NE1921  --
bbt_BBta_2  NA
syw_SYNW17  --
pma_Pro055  --
sme_SMb201  --
bbt_BBta_0  A-
nph_NP2770  --
pab_PAB158  --
afu_AF1638  --
mja_MJ1235  --
rci_RCIX22  --
afu_AF1587  --
cte_CT1772  --
ret_RHE_PF  --
olu_OSTLU_  --

```

Figura 9. Alineamiento de secuencias rbcL y matriz de datos.

Construida la matriz de datos, se accede al programa de reconstrucción filogenética seleccionado. En este ejercicio de análisis cladístico con criterio de parsimonia, se utilizará un programa de parsimonia para proteínas denominado **PROTPARS** (contenido en PHYLIP: Joe Felsenstein 1986-1995)

<http://evolution.genetics.washington.edu/phylip.html>

ANÁLISIS Y OBTENCIÓN DE RESULTADOS

El análisis comienza con el acceso al programa **Protpars** (Parsimonia para proteínas) pinchando el icono correspondiente (Fig. 10).

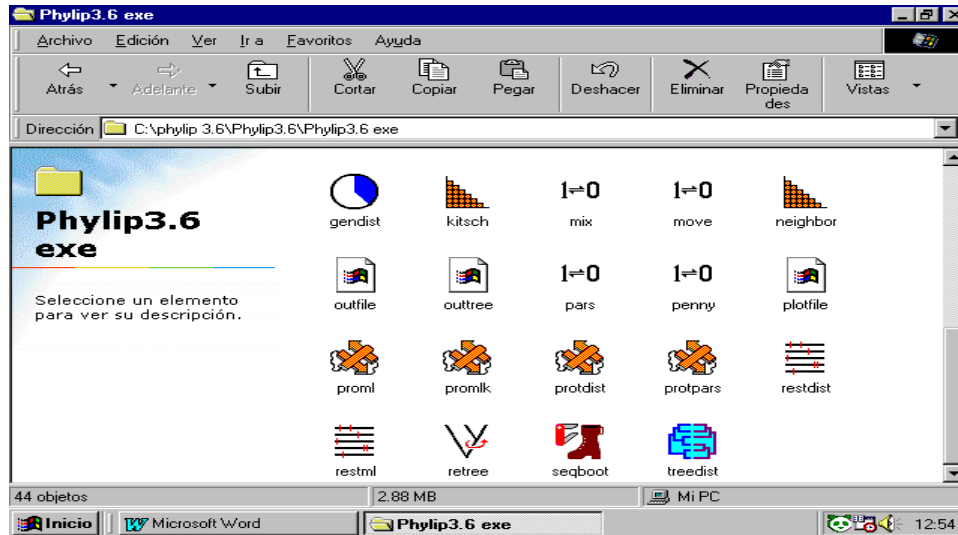


Figura 10. Programas contenidos en Phylip. Acceso al programa PROTPARS.

Cuando se accede al programa aparece una pantalla (Fig. 11) en la que se escribe la ruta de acceso al archivo que contiene la matriz de datos (guardada como texto “.txt”). Realizado este paso, que es la entrada de los datos al programa, aparece el menú de opciones.

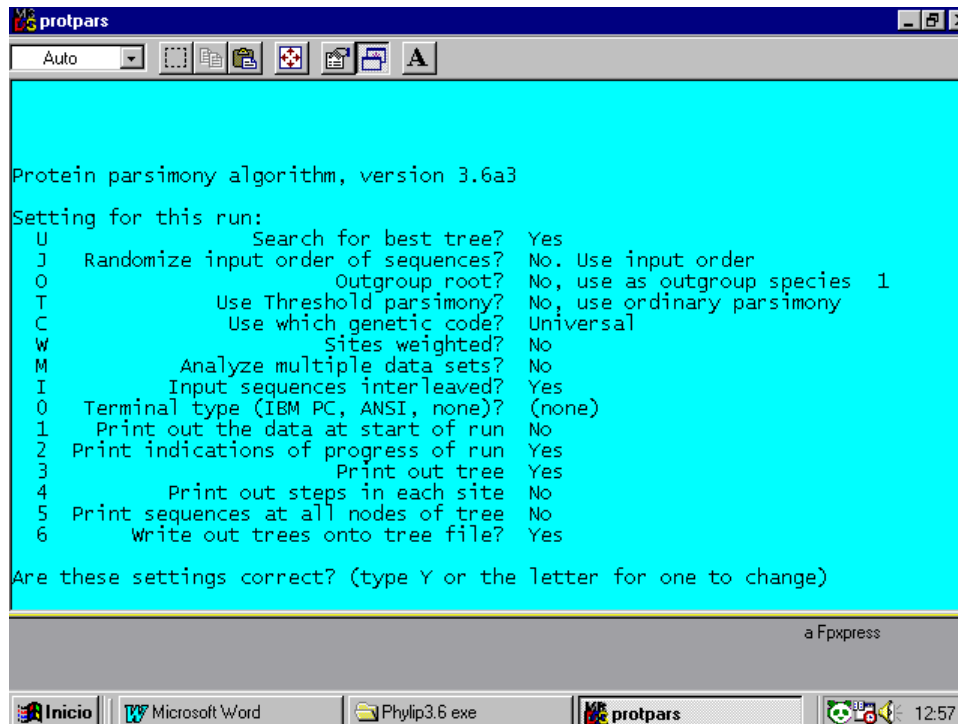
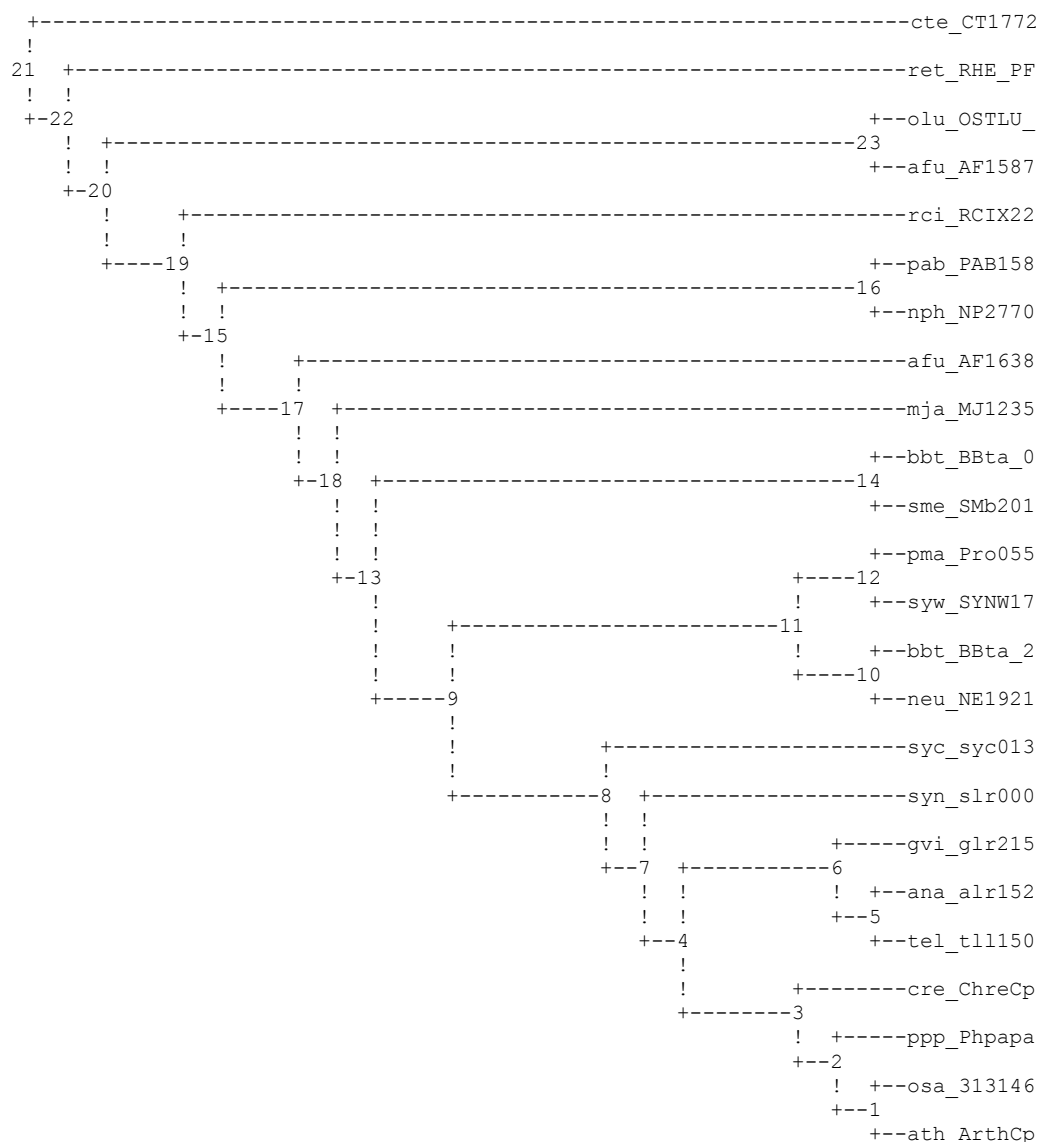


Figura 11. Programa PROTPARS: menú de opciones.

Las distintas opciones se pueden modificar tecleando la letra/número correspondiente. Una vez fijadas estas opciones se procede al desarrollo del programa. Finalizado éste, el resultado se graba en el archivo outfile (que se abre con WORDPAD) el cual contiene el **cladograma** resultante (Fig. 12), una tabla con el **número de cambios que experimenta cada carácter** (Tabla 3), y otra tabla que sirve para **situar** estos **cambios** en el árbol o cladograma (Tabla 4). Para el caso de las secuencias rbcl, el archivo outfile es el siguiente:

Protein parsimony algorithm, version 3.65

One most parsimonious tree found:



remember: (although rooted by outgroup) this is an unrooted tree!
requires a total of 5037.000

Figura 12. Cladograma de secuencias rbcl . Archivo outfile de Protpars de Phylip.

steps in each position:

	0	1	2	3	4	5	6	7	8	9
*	-----									
0!		3	3	3	3	3	3	3	3	3
10!	3	3	3	3	3	3	3	3	3	3
20!	3	3	3	3	3	3	3	3	3	3
30!	3	3	3	3	3	3	3	3	3	3
40!	3	3	3	3	3	3	3	3	3	3
50!	3	3	3	3	3	3	3	3	3	3
60!	3	3	3	3	3	3	3	3	3	3
70!	3	3	3	3	3	3	3	3	3	3
80!	3	3	3	3	3	3	3	3	3	3
90!	3	3	3	3	3	3	3	3	3	3
100!	3	3	3	3	3	3	3	3	3	3
110!	3	3	3	3	3	3	3	3	3	3
120!	3	3	3	3	3	3	3	3	3	3
130!	3	3	3	3	3	3	3	3	3	3
140!	3	3	3	3	3	3	3	3	3	3
150!	3	3	3	3	3	3	3	3	3	3
160!	3	3	3	3	3	3	3	3	3	3
170!	3	3	3	3	3	3	3	3	3	3
180!	3	3	3	3	3	3	3	3	3	3
190!	3	3	3	3	3	3	3	3	3	3
200!	3	3	3	3	3	3	3	3	3	3
210!	3	3	3	3	3	3	3	3	3	3
220!	3	3	3	3	3	3	3	3	3	3
230!	3	3	3	3	3	3	3	3	3	3
240!	3	3	3	6	8	11	14	17	14	19
250!	17	17	16	16	16	18	12	11	12	16
260!	12	15	15	18	19	14	18	17	12	14
270!	9	13	14	15	13	8	19	12	16	11
280!	11	10	13	9	7	16	7	11	13	13
290!	11	7	19	8	9	13	17	9	6	12
300!	8	10	8	6	4	5	1	2	1	5
310!	3	4	4	8	10	3	10	9	9	18
320!	9	13	13	13	12	13	10	8	10	13
330!	12	10	15	10	13	18	14	15	14	12
340!	17	16	17	11	5	5	4	5	4	5
350!	5	5	5	7	9	11	19	12	10	13
360!	6	4	5	4	8	11	8	2	4	13
370!	9	6	10	9	10	12	6	5	9	6
380!	8	2	8	8	6	3	6	6	5	7
390!	10	10	13	8	3	3	7	0	11	9
400!	9	2	18	9	13	14	5	10	7	16
410!	0	0	8	9	0	4	13	4	6	2
420!	8	10	6	8	5	13	4	2	3	6
430!	7	7	6	6	9	0	0	7	5	2
440!	4	5	7	8	6	11	3	5	7	4
450!	6	5	11	8	9	1	0	6	0	6
460!	6	0	0	0	2	6	5	4	7	9
470!	9	5	13	6	7	6	5	0	6	13
480!	12	12	16	11	8	8	17	8	9	9
490!	11	5	2	2	4	14	7	9	5	8
500!	8	4	9	2	5	18	9	13	12	7
510!	3	15	9	10	6	7	12	5	11	9
520!	13	7	15	10	11	6	3	7	3	13
530!	10	10	12	5	10	8	4	7	7	13
540!	8	10	16	16	16	8	14	9	11	12
550!	8	5	4	10	4	5	6	3	3	3
560!	7	4	4	7	12	7	2	14	13	13

570!	7	2	6	9	7	8	6	5	6	4
580!	13	10	0	6	11	0	5	2	11	8
590!	4	11	4	3	8	6	7	6	4	5
600!	4	3	8	7	15	15	7	10	6	9
610!	9	6	14	15	10	10	17	17	14	14
620!	12	10	14	15	11	11	13	14	6	9
630!	7	9	10	14	13	12	9	9	10	12
640!	2	3	4	2	1	1	7	5	5	9
650!	10	6	8	10	8	14	12	14	10	0
660!	5	3	10	2	11	14	7	4	8	0
670!	1	1	6	10	3	0	1	11	2	7
680!	11	3	0	2	10	3	7	4	7	2
690!	7	5	2	13	13	14	7	11	11	9
700!	6	8	8	7	11	8	3	5	12	7
710!	9	5	10	12	11	8	12	13	8	8
720!	4	2	14	13	5	7	9	15	5	10
730!	17	14	12	13	16	5	10	10	11	3
740!	7	7	9	12	12	13	8	7	7	7
750!	7	6	3							

Tabla 3. Número de cambios que experimentan los caracteres.

From	To	Any Steps?	State at upper node			
(. means same as in the node below it on tree)						
root	21		-----	-----	-----	-----
21	cte_CT1772	no
21	22	no
22	ret_RHE_PF	no
22	20	no
20	23	no
23	olu_OSTLU_	yes	MAPKSFEDMF	ACSPASLTKL	RASHAPHASP	AASHARDYSP
23	afu_AF1587	no
20	19	no
19	rci_RCIX22	no
19	15	no
15	16	no
16	pab_PAB158	no
16	nph_NP2770	no
15	17	no
17	afu_AF1638	no
17	18	no
18	mja_MJ1235	no
18	13	no
13	14	no
14	bbt_BBta_0	no
14	sme_SMb201	no
13	9	no
9	11	no
11	12	no
12	pma_Pro055	no
12	syw_SYNW17	no
11	10	no
10	bbt_BBta_2	no
10	neu_NE1921	no
9	8	no
8	syc_syc013	no

8	7	no
7	syn_slr000	no
7	4	no
4	6	no
6	gvi_glr215	no
6	5	no
5	ana_alr152	no
5	tel_tll150	no
4	3	no
3	cre_ChreCp	no
3	2	no
2	ppp_Phpapa	no
2	1	no
1	osa_313146	no
1	ath_ArthCp	no
root	21		-----	-----	-----	-----
21	cte_CT1772	no
21	22	no
22	ret_RHE_PF	no
22	20	no
20	23	no
23	olu_OSTLU_	yes	IAFPSMRDD	EDERARAVAG	AFDDMFACSP	ASLKKMRGSA
23	afu_AF1587	no
20	19	no
19	rci_RCIX22	no
19	15	no
15	16	no
16	pab_PAB158	no
16	nph_NP2770	no
15	17	no
17	afu_AF1638	no
17	18	no
18	mja_MJ1235	no
18	13	no
13	14	no
14	bbt_BBta_0	no
14	sme_SMb201	no
13	9	no
9	11	no
11	12	no
12	pma_Pro055	no
12	syw_SYNW17	no
11	10	no
10	bbt_BBta_2	no
10	neu_NE1921	no
9	8	no
8	syc_syc013	no
8	7	no
7	syn_slr000	no
7	4	no
4	6	no
6	gvi_glr215	no
6	5	no
5	ana_alr152	no
5	tel_tll150	no
4	3	no
3	cre_ChreCp	no
3	2	no
2	ppp_Phpapa	no

2	1	no
1	osa_313146	no
1	ath_ArthCp	no
root	21		-----	-----	-----	-----
21	cte_CT1772	no
21	22	no
22	ret_RHE_PF	no
22	20	no
20	23	no
23	olu_OSTLU_	yes	VERASGGVGT	PTGTPTRPTA	SREDLASAFG	DMFLASPASL
23	afu_AF1587	no
20	19	no
19	rci_RCIX22	no
19	15	no
15	16	no
16	pab_PAB158	no
16	nph_NP2770	no
15	17	no
17	afu_AF1638	no
17	18	no
18	mja_MJ1235	no
18	13	no
13	14	no
14	bbt_BBta_0	no
14	sme_SMb201	no
13	9	no
9	11	no
11	12	no
12	pma_Pro055	no
12	syw_SYNW17	no
11	10	no
10	bbt_BBta_2	no
10	neu_NE1921	no
9	8	no
8	syc_syc013	no
8	7	no
7	syn_slr000	no
7	4	no
4	6	no
6	gvi_glr215	no
6	5	no
5	ana_alr152	no
5	tel_tll150	no
4	3	no
3	cre_ChreCp	no
3	2	no
2	ppp_Ph papa	no
2	1	no
1	osa_313146	no
1	ath_ArthCp	no
root	21		-----	-----	-----	-----
21	cte_CT1772	no
21	22	no
22	ret_RHE_PF	no
22	20	no
20	23	no
23	olu_OSTLU_	yes	KKMRATNTAV	TGTKARERGD	GFGSPLRYDG	DANAAEPKKE
23	afu_AF1587	no

20	19	no
19	rci_RCIX22	no
19	15	no
15	16	no
16	pab_PAB158	no
16	nph_NP2770	no
15	17	no
17	afu_AF1638	no
17	18	no
18	mja_MJ1235	no
18	13	no
13	14	no
14	bbt_BBta_0	no
14	sme_SMb201	no
13	9	no
9	11	no
11	12	no
12	pma_Pro055	no
12	syw_SYNW17	no
11	10	no
10	bbt_BBta_2	no
10	neu_NE1921	no
9	8	no
8	syc_syc013	no
8	7	no
7	syn_slr000	no
7	4	no
4	6	no
6	gvi_glr215	no
6	5	no
5	ana_alr152	no
5	tel_tll150	no
4	3	no
3	cre_ChreCp	no
3	2	no
2	ppp_Ph papa	no
2	1	no
1	osa_313146	no
1	ath_ArthCp	no
root	21		-----	-----	-----	-----
21	cte_CT1772	no
21	22	no
22	ret_RHE_PF	no
22	20	no
20	23	no
23	olu_OSTLU_	yes	AVSFSEMFAS	SPSDLAKMRR	SDVGGAARGS	NVEVAKVAVD
23	afu_AF1587	no
20	19	no
19	rci_RCIX22	no
19	15	no
15	16	no
16	pab_PAB158	no
16	nph_NP2770	no
15	17	no
17	afu_AF1638	no
17	18	no
18	mja_MJ1235	no
18	13	no
13	14	no

14	bbt_BBta_0	no
14	sme_SMb201	no
13	9	no
9	11	no
11	12	no
12	pma_Pro055	no
12	syw_SYNW17	no
11	10	no
10	bbt_BBta_2	no
10	neu_NE1921	no
9	8	no
8	syc_syc013	no
8	7	no
7	syn_slr000	no
7	4	no
4	6	no
6	gvi_glr215	no
6	5	no
5	ana_alr152	no
5	tel_tll150	no
4	3	no
3	cre_ChreCp	no
3	2	no
2	ppp_Ph papa	no
2	1	no
1	osa_313146	no
1	ath_ArthCp	no
root	21		-----	-----	-----	-----
21	cte_CT1772	no
21	22	no
22	ret_RHE_PF	no
22	20	no
20	23	no
23	olu_OSTLU_	yes	ARERTNAFAG	MFASSPADLA	KMRGRSQSSI	SYAEGRHGAS
23	afu_AF1587	no
20	19	no
19	rci_RCIX22	no
19	15	no
15	16	no
16	pab_PAB158	no
16	nph_NP2770	no
15	17	no
17	afu_AF1638	no
17	18	no
18	mja_MJ1235	no
18	13	no
13	14	no
14	bbt_BBta_0	no
14	sme_SMb201	no
13	9	no
9	11	no
11	12	no
12	pma_Pro055	no
12	syw_SYNW17	no
11	10	no
10	bbt_BBta_2	no
10	neu_NE1921	no
9	8	no
8	syc_syc013	no

8	7	no
7	syn_slr000	no
7	4	no
4	6	no
6	gvi_glr215	no
6	5	no
5	ana_alr152	no
5	tel_tll150	no
4	3	no
3	cre_ChreCp	no
3	2	no
2	ppp_Ph papa	no
2	1	no
1	osa_313146	no
1	ath_ArthCp	no
root	21		-----	-----	--??ED???	????E?LD?
21	cte_CT1772	maybeMNA..VKG	FFASR.S..M
21	22	yes???.???	?????.?..?
22	ret_RHE_PF	yes-----	-----
22	20	yes	??KH?..???	?L???.?..?
20	23	yes	???????????	?.?.?.A???	?..?.?D???
23	olu_OSTLU_	yes	DRFQVIAPDF	QAMF...P.D	LKNLRS.SSS	SRKRLG.EAD
23	afu_AF1587	yes	MQLGVRLRFQ	KFEYP..NPE	A.PEGI.PEE
20	19	yesV..?..?..?	????..?..I..
19	rci_RCIX22	yes	MAI.GDKDLI	IQVPLTL.NI
19	15	yesVE.YV?F	?D.SY.P..D
15	16	yes	??.....D.S..
16	pab_PAB158	yesMVS	.M..W.L..	V.LN...GR.
16	nph_NP2770	yes	----M..A..	L.E.....
15	17	yesK.M.Y	..?..K?.
17	afu_AF1638	yes	MAEF.I.RE.	V.K....QK.
17	18	yes	?..???.?..	.NLN.....?
18	mja_MJ1235	yes	-----D.	I....R.NEG
18	13	yes	?.K??ERYKA	G..EY.H.?	W.PD....DT
13	14	yes??A????	??G?.....	..L...?G.
14	bbt_BBta_0	yes	..MNE.LK.L	TVT.K....SR...	.E.....
14	sme_SMb201	yes	..MN.DAKT	EI..R.....	..K.AQ...	..G.....
13	9	yesA?.....?..T.	.T...?.....
9	11	yes	?.?.?K?..?R...V..?..
11	12	yesMS.K.D.D...L..
12	pma_Pro055	yes	-.....-
12	syw_SYNW17	yes	-.....-
11	10	yesM.?.?
10	bbt_BBta_2	yes	-..E.S.Q.K.-L..
10	neu_NE1921	yes	-..S.A.T.N.-	.E.H.NVQ..
9	8	yes	KTQ..AG...	..D..L...	Y....T....
8	syc_syc013	yesMP	..SA.....K.-
8	7	yes?..?R...
7	syn_slr000	yes	MV..K..F..	..Q.....-
7	4	yes??QT
4	6	yesK..Y?
6	gvi_glr215	yes-M.Y.Q.-
6	5	yes???
5	ana_alr152	yesM.YA..	..T.S..K.	..Q.....-
5	tel_tll150	yesMAYTQ	SKSQ.V..Q.-
4	3	yesM.P..	E.K.G..F..??
3	cre_ChreCp	yesV...-VVR..
3	2	yesS...V...ET...
2	ppp_Ph papa	yesRP	.I.....-Q....

2	1	yesS.....K.....	...E.....
1	osa_313146	yes-
1	ath_ArthCp	yesE.....-
root	21		?????L?Y??	E??G??ET??	????SEQST?
21	cte_CT1772	maybe	EQYLV.D.YL	.SV.DI..AL	AHFC.....A
21	22	yes	??V.T.T...	.?..?..VA	.?IA.....G
22	ret_RHE_PF	yes	--MI....RI	.TP.SV..M	DK...D....
22	20	yes	??..V..???	.?G??..??	?E.....
20	23	yes	??...YSL??	P?..??EI.	?...L.....
23	olu_OSTLU_	yes	RFE...L.LA	.CEAAAR.KV	L..C...TVE
23	afu_AF1587	yes	YIIG..YMSF	.K.MNPF..T	QVL.....
20	19	yes	?.....	.?..?.....T?..A...
19	rci_RCIX22	yes	MTM.RT.YYV	.ADAPIAK..	K...A.....
19	15	yes	??..?FRF?P	...V...EA.	???.S...
15	16	yes	.L.....?	??..S.....	.R..S...?
16	pab_PAB158	yes	E.I.EYY.E.	N-...P....	G.I.....I.
16	nph_NP2770	yes	D..CT..LV.	GE.I.VAD..	A.V.....N.
15	17	yes	D..A?...I??	.?..A.....
17	afu_AF1638	yes	.I..V...T.	AE.FTI.D..	GAV.....
17	18	yes	.LL.S.....	.?..?L....	?...?.....
18	mja_MJ1235	yes	...SCMV.KG	---N..KL.	NEI.G...I.
18	13	yesT.	Q?...P....	AAV.....
13	14	yes	..I.L.....	.D..D.I...G...A
14	bbt_BBta_0	yes	.V.....V..	.N.....S
14	sme_SMb201	no
13	9	maybe?..P.....A.....
9	11	yesC.K..?R.....
11	12	yesC.G	.E.....V.S...SE..
12	pma_Pro055	no
12	syw_SYNW17	yesK....
11	10	yesV..	.A.....
10	bbt_BBta_2	yesV....A
10	neu_NE1921	yes	.I.....P	.P..D.....
9	8	yes?.....	.P...?..?
8	syc_syc013	yes	...A..FS.AD..G	..I.....
8	7	yesM..?
7	syn_slr000	yesC.....A...AN.
7	4	yes	.I..A.....??
4	6	yesV..?
6	gvi_glr215	yes	.V.....I...GG.
6	5	maybeF...AL.
5	ana_alr152	no
5	tel_tll150	no
4	3	maybe?..?GG.
3	cre_ChreCp	yesL...P..C.
3	2	no
2	ppp_Ph papa	maybeA..C.
2	1	yesV..P..A.
1	osa_313146	no
1	ath_ArthCp	no
root	21		??????A???	??????E?EQ	???????????
21	cte_CT1772	maybe	FRLVHA.KVI	DYEVIE.L..	LSYPVKHSET
21	22	yes	E.R.RV.?V	V?V?R...?	.??????????
22	ret_RHE_PF	yes	.LKS...AR.	LAIRPL.NAR	HPTWPE.APD
22	20	yes	.?..?..????B.?	???.?.....
20	23	yes	.?..?..V???	.?..?K?..?	???.?.....
23	olu_OSTLU_	yes	WI.EH..GRL	ESLT.PKTGP	HAR-----
23	afu_AF1587	yes	.V.RKH.AK.	.G.YEIPDYE	IMV-----

20	19	yes	.?...G.???	A....?..???	?.....??..?.
19	rci_RCIX22	yes	VHEKL.AHV.	SAEG-----	-----	-----NT.V.
19	15	yes	.?..?..???	.?.I..?.D.?	??.....??..?.
15	16	yes	?..?..??S..?.....	G?.....?.
16	pab_PAB158	yes	---MAKR.MA	KVFYLEKHGE	.-----	-----YIAK.
16	nph_NP2770	yes	---V.QY.AL	.CD.GPE..H	.-----	-----TQ.TV
15	17	yes	.D.VK.?.AK	.?..?..?G.E	?.....?.....
17	afu_AF1638	yes	.E...GLS..	.YDFVDL.DG	S-----	-----SI.R.
17	18	yes	?..?D.*.?P.	?.....??..?
18	mja_MJ1235	yes	-.IYEKLR..	VYE.KEI...	NGY-----	---KVGLIK.
18	13	yes	T..?.YR?RC	Y?.?PVP...	.S.....YFA?.
13	14	yes	.A?....AK.	.R.D....TP	G.....Y.
14	bbt_BBta_0	yes	..AE.....N..	.-----	-----
14	sme_SMb201	yes	..C.Q....AQ-----	-----C.V
13	9	yes	..L.?.KG..	..IE.....F.
9	11	yes?.....	.R..?..D.FY..
11	12	yesF.....D...K	E.....
12	pma_Pro055	yes	...E.....-----	-----
12	syw_SYNW17	yes-----	-----
11	10	yesY....SD
10	bbt_BBta_2	yesAP.....	NA-----	-----
10	neu_NE1921	yesD.....	S-----	-----
9	8	yesR.....	.H.....	N.....
8	syc_syc013	yes	..M.....K.Q...	.-----	-----
8	7	yesD.....D	.Q.....
7	syn_slr000	yesL.A..N..	.-----	-----
7	4	yesI.Y.
4	6	yes?
6	gvi_glr215	yes	.E.....-----	-----W.C..
6	5	yesF.....
5	ana_alr152	yes-----	-----
5	tel_tll150	yesC.L....	.-----	-----
4	3	yes	.S.....V
3	cre_ChreCp	yes-----	-----
3	2	yes?..?
2	ppp_Phpapa	yesA....A..E	.-----	-----
2	1	yesH.....
1	osa_313146	yesV..D	.-----	-----
1	ath_ArthCp	yesP..E	T.-----	-----F.....
root	21		A?P???FG??	?I???LTA??	G???Y???GV	????L?DI??
21	cte_CT1772	maybe	.H.HCN..P-	K.PNL...VC	.EGT.FTP..	PVVK.M..HF
21	22	yes	.?.LE?...D	...?.....G	.????.?R..V..Q.
22	ret_RHE_PF	yes	.F...AI.T.	LSALMTI.I.	--V.SIK.M	TGIRI..MKL
22	20	yes???	N?..?..???	.?V?G?....	?..?.....
20	23	yes	S?A.V...???	..G???.?LV	S?...N.SM?	??..?.....
23	olu_OSTLU_	yes	WN.I.SYHA.	TA.GE.TQ..	NVIF..T..K	ENVMV...EL
23	afu_AF1587	yes	FIVQIA.PWR	.I.SK.SM.F	.T.V..I..A	PKLK.L.LRF
20	19	yes?..E..	.?P??S..A	.N.F..?..	K??R...?..
19	rci_RCIX22	yes	DF.V.I..P.	.V.QI..VV.	..L..-LG.L	.AC..M.VDF
19	15	yes	.Y...L..?.?..GI.MK..	..L....?..
15	16	yes?..?G	SL.Q.....A.	.?..L.F.?
16	pab_PAB158	yesT...E.	..V.LF.A..-...L	.N.....HP
16	nph_NP2770	yes	...SG...D.I..C..	..IM.-....	ETI....CEW
15	17	yes?G?.A?..R.	.?..?.....
17	afu_AF1638	yes	...S....PH	.M..L..S..-....	.G...E.L.L
17	18	yesD...?.?.....RF
18	mja_MJ1235	yesYD..IN	.M..V..G..	..I..-..IA	.G..IL.F..
18	13	yes?G	S.?N...S.VF.PL	.A...E.I..
13	14	yes	..D.....	.IA.LS...IM..

14	bbt_BBta_0	yesP.-
14	sme_SMb201	yes	...I...E.T...	...S-....	..A.....L
13	9	yesE.	???.V.T...A.	R?.....
9	11	yesL.	?.....
11	12	yesIT.....H.....
12	pma_Pro055	yes-
12	syw_SYNW17	yes-
11	10	yes	...I.....	.VV.....V	.S.....
10	bbt_BBta_2	yes-
10	neu_NE1921	yesT-V..
9	8	yesVT.....
8	syc_syc013	yesI.....-...I	.S.....
8	7	yesA.....
7	syn_slr000	yesL.-
7	4	yes
4	6	yes
6	gvi_glr215	yesL.-
6	5	yes
5	ana_alr152	yesI.....-
5	tel_tll150	yesM.....-	K.....L.I
4	3	yesMF.....L.I
3	cre_ChreCp	yes	...I.....-
3	2	yes
2	ppp_Ph papa	yesL.....-
2	1	yes
1	osa_313146	yes-
1	ath_ArthCp	yes-	A.....
root	21		P???L?D??G	P?FGIEGLR?	?L???GRPI?	??VVKPB?GL
21	cte_CT1772	maybe	.DTY.A.FE.	.K.....D	I.NAH....F	FG....NI..
21	22	maybe	.EAF.....?	.GVQ.....	?T.....
22	ret_RHE_PF	yesRSAHP.	.Q...A.S.R	LT..E....I	G.I...AL..
22	20	yesK...?.?..D..L?	...?...V..
20	23	yes	.K....?...	?.....	?.....	?.....?...
23	olu_OSTLU_	yes	.RTM.REYP.	.R..VH.L.R	L.R.PEG..V	M.AL..-M..
23	afu_AF1587	yes	..E.V.GFK.	.K.....V.D	V...K....L	NNMI..D.YS
20	19	yes	.?....?.K.	?.....V.?*.....	G.....K...
19	rci_RCIX22	yes	G-PLT.YYN.	.E....E..K	I...Y....V	..II.....
19	15	yes	.?....F..	?.....?.....
15	16	yes	???.R?...	.Q.....I.E	?.....?	A..P.....
16	pab_PAB158	yes	.YFY..H...	...VK....	FM.IK....TM.W
16	nph_NP2770	yes	.AVIA.S.P.	..Y.SDVRT.	L.DAG...PL
15	17	maybeD...	??.....	?.....
17	afu_AF1638	yes	.KS.....	.SK.K...KK	IF..A...IV	...P.....Y
17	18	yes	.?.V.A...E	??.I?...L	..I?.....
18	mja_MJ1235	yes	.AE....Y..	.RF.....	T.K.KE....	...V.....
18	13	yes	.V...T.?.	.P?...VE..	R.BKYG....	.?T?...L..
13	14	yes	...Y.....	.T..V....	..D.F.....	.A.....
14	bbt_BBta_0	yesQ.	.A.....V.....
14	sme_SMb201	yesR.K...	...T.....
13	9	yesN..Q...D	..N.....	.C.I.....
9	11	yes	.?....C?
11	12	yes	.M..I.....
12	pma_Pro055	yesG.	..Q..V....
12	syw_SYNW17	yesS.Y.M.....
11	10	yes	...Y....G.
10	bbt_BBta_2	yes	.L.....	...L....
10	neu_NE1921	yes	.I.....	I.....AY.
9	8	yes	...?....Q.	..H.....
8	syc_syc013	yes	...L.....	L.....M.

8	7	yes?	K.....
7	syn_slr000	yes	...LI.....T.....
7	4	maybe	...?
4	6	no
6	gvi_glr215	yes	.I.LV..Y..V.....	.I.....
6	5	maybe	...Y?.....
5	ana_alr152	maybe	...I.....
5	tel_tll150	maybe	...L.....
4	3	yes	.P.Y.....
3	cre_ChreCp	yes	...V...V.G.
3	2	yes	...S.....
2	ppp_Ph papa	no
2	1	no
1	osa_313146	yes	..T.....
1	ath_ArthCp	yes	...T.....
root	21		SP?E?AE?AY	????GG?DI?	KDDE?L?B??	?S??EER?A?
21	cte_CT1772	maybe	..G.F..I..	QSWL..L..A	...M.ADV T	W.SI...A.H
21	22	maybe?.....	E.?R..V..I?.	?P....V.?
22	ret_RHE_PF	yes	R.H.T..LVG	.LIGS...F.	...K.MSPA	Y..LK...A
22	20	yes	..?..?.....	..??.....?.	?.....?
20	23	yes?.....	?A.....	...?A..?P.
23	olu_OSTLU_	yes	.SQ.L..I..	GF.K..I...	...HG..DQP	Y..YD...RM
23	afu_AF1587	yes	P.DLG.KL..	.V.....	...L..NPE	FNRI.....K
20	19	yes	..E....V..L.?.T.Q.	F.....
19	rci_RCIX22	yes	..KRT....E	QAAL..L...	...T..D.K	.C.L...LTM
19	15	yes	.A..?.....	..?.....	...N..?.	...?F....?K
15	16	yesW.....L?	.N.....
16	pab_PAB158	yes	.V..Y..I..	.L.S..I...	...F.SFP	.R.....K.
16	nph_NP2770	yes	.T..HVS..E	SA.....D.T	.P..Q..AD
15	17	yesK.....	?.....?	...S.A	.R.....
17	afu_AF1638	yes	...VE.L..	.LLS..M.Y.P.	YC....AER
17	18	yes	?.....	?...D.??.
18	mja_MJ1235	yes	KT..H.....	.AWV...LVE	.NK....IY.
18	13	yes	..KNYGR.V.	?L...L.FT	...IN..P	?..WR..F??
13	14	yes	.G.....	?..K.....MH....LY
14	bbt_BBta_0	yes	..R.....	.A.....	...T.....
14	sme_SMb201	yesG.....MY..
13	9	maybe?	.C.....F
9	11	yesE.
11	12	yes	.G....V..Q.....
12	pma_Pro055	yesL.
12	syw_SYNW17	yesQN.....
11	10	yesA..M...H....
10	bbt_BBta_2	yesQ.....
10	neu_NE1921	yesV.....	...Q..D.
9	8	maybeA..L.
8	syc_syc013	yesQ.....
8	7	maybeM.....
7	syn_slr000	no
7	4	no
4	6	no
6	gvi_glr215	no
6	5	yesQ.....
5	ana_alr152	yesA.
5	tel_tll150	no
4	3	yesV.....
3	cre_ChreCp	no
3	2	no
2	ppp_Ph papa	no

2	1	no
1	osa_313146	yesC.V.
1	ath_ArthCp	no
root	21		?????KAEA	ETG????Y??	NITD?????M	E??D??V???
21	cte_CT1772	maybe	LGKARR....	...EPKI.LA	...EVDSL.	.KH.VA.RN-
21	22	yes	.M?.I.....?...AVA..?V.	.R?..?..???
22	ret_RHE_PF	yes	I.PR.LDH.Q	K..KKVM..F	G.SH.DPDE.	M.NHDI.AAA
22	20	yes	..?..?.....?...	...A..?..?	..??E??..??
20	23	yes	F..A.DR.D.K?L...H?..?	..?..?..?..
23	olu_OSTLU_	yes	CAA.VA..N.	...RNV...P	C.N.PA.LII	S.AHAARQAG
23	afu_AF1587	yes	..E.....E	.K.E.T....	.V..DLPE.L	.N-A.RAIEL
20	19	yes	...?..V....R??...?AD?..	...?A.R.?EL
19	rci_RCIX22	yes	V.DRLH...D	RI.KPCF...	.V.CG..MIV	..-...AV..
19	15	yes	...?.D....	...E.K?..?V???	..R..L.A..
15	16	yes	.F?.R.R...?..L.?..V.D?
16	pab_PAB158	yes	LYRV.....T.E..I	...G-P.NV.	.K.....NE
16	nph_NP2770	yes	S.AA...L.ED...-ETDE.	V....F.D.H
15	17	yes	I.E.....?..?A..?E.
17	afu_AF1638	yes	..KV.E....K.SWFA-D.R..	...LK.V...
17	18	yesA..?..??..	?...?AE..
18	mja_MJ1235	yes	TL.MR....EA.MP-PYR..	I....I..DA
18	13	yes	?..A..?..??.GHYL	?...TME..	??...F.K..
13	14	yes	C...VN..??	?...V.....	Y.....
14	bbt_BBta_0	yesQ.	A.....T..	.V.....D.	.E.....
14	sme_SMb201	yesH.S.	V.....	.I..G.....	.R.....
13	9	yes	V?.....V..?.R...	.K.....
9	11	yes??.P...	?.....
11	12	yes	.A...KL..QT.....	YE.....
12	pma_Pro055	yes	...V...Q.	...V.....	.C.....
12	syw_SYNW17	yesSE.R.....N.....
11	10	maybe	.M...H..??R.....P.....
10	bbt_BBta_2	yes	...V...TS	Y.....S.
10	neu_NE1921	yesER	F...Y....
9	8	yes	.A.....Q.I.....C...	M.....
8	syc_syc013	yes	..D..H.S..P.....
8	7	yesG.....
7	syn_slr000	yes	.Q...E....	..N.M.....I
7	4	no
4	6	yes	..D.....
6	gvi_glr215	yes	.Q...V.S..C.....	.E.....
6	5	yesP.....	L.....
5	ana_alr152	yesT....Y....
5	tel_tll150	maybeH....
4	3	yesY....A.....	...?.....
3	cre_ChreCp	yesV.....VC...
3	2	yes	C.....S..	I.....R..
2	ppp_Ph papa	yesGQ.....
2	1	maybeV.....
1	osa_313146	maybe
1	ath_ArthCp	maybe
root	21		G?N????N?L	?VG??A?R?L	??---?????	L??H??---?
21	cte_CT1772	maybe	.A.ALLI.A.	P...LS.V.M.	SN...YTQVP	.IG.FP...F
21	22	maybe	...C..V.?	...F....?	?.....?L.	.HA.?...V
22	ret_RHE_PF	yes	.G..AV..IN	SI..GGMSF.	RK...RSS.V	...RN...G
22	20	yes?	TA.....???	??..??N???	?.....
20	23	yesL...??VL...	A.???.?..I	?..P???.
23	olu_OSTLU_	yes	AGAV.MIPGI	.GLDAMRELA	.DP.F.LPI.	A..ALLGCML
23	afu_AF1587	yes	.A...L..Y.	AT..P..RA.	.EDESIVP.	MAHMDV...A

20	19	yesVM.D..S..Q?.E...??	I...R....M
19	rci_RCIX22	yes	.A.M....I.V.A.	TD.-.KIGVP	..I..T....
19	15	yes	.GPY.....V	?...?....S?	R.L?.D???.AA....
15	16	yesI.	..W..L...	.?.T..L...
16	pab_PAB158	yes	..Q...I...	V.....YM	.EV....G..
16	nph_NP2770	yes	..SF.....I	TC...G..TV	.RR....D..
15	17	yes	.S.....V.	I.....?	...?..?
17	afu_AF1638	yes	.N.H.....	.T.WG.LEYI	.D.A..YD..	..G.....
17	18	yes	..?..?..?	.G..?..?	.E??.....I
18	mja_MJ1235	yes	..E...I...	VC.F..V..F	..--..FKF.
18	13	yes	...I?..???T.I...	A?W?R?N.M.	L.?
13	14	yes	..V.....L.	..??.....??.D..?.G
14	bbt_BBta_0	yesI.I..V	..-Y.....M	.K.A.R....	..L.....
14	sme_SMb201	yesV.V..I	V.-W.....I	SE.C.Q....	..M.....
13	9	yes	.?P.I.H.F?	T..F..NT?L	.K.C...G.L	..I.....
9	11	yes?..?G.	.?.....
11	12	yes	.M.....IK....
12	pma_Pro055	yes	D.....Y.N.....
12	syw_SYNW17	yes	SK.....
11	10	yes	.A.....N.....
10	bbt_BBta_2	yesL	.A.....E....
10	neu_NE1921	yes	K.....YI	A...C....D..I.
9	8	maybeLT.D..?
8	syc_syc013	yes	.M.....	.A.....V....
8	7	maybe?.....
7	syn_slr000	yes	.T.....FR.....I.
7	4	yesY.
4	6	yes	?.....
6	gvi_glr215	yes	KT.....K...R..I.
6	5	yes	.?.....	.A.....
5	ana_alr152	yes	KQ.....R.....V.
5	tel_tll150	yes	EM.....F.	SK.....M.
4	3	yes	.?.....S.	.?Y.....L.
3	cre_ChreCp	maybe	.V.....I.....
3	2	yes	...V.....H.....
2	ppp_Ph papa	maybe	.M.....
2	1	maybe	.V.....
1	osa_313146	no
1	ath_ArthCp	yes	S.....
root	21		?A??R????	GI?SKV??KL	?RL?G?D?V?	?PGFG?----
21	cte_CT1772	maybe	I.SFS.MEKY	..H...MT..	Q..A.L.A.I	M....D....
21	22	maybe	R.?..T..P..L...V...L	I.....
22	ret_RHE_PF	yes	WDVL..D.GA	.MDF..YQQF	W..L...QFQ	.N.IRI....
22	20	yes?..?	..?TL..?	?.....	..A?.G....
20	23	yes	G.??S?I?G	V??..?.G..	P..A.A.?V	F.....
23	olu_OSTLU_	yes	.GGS.NR.A.	FSHEVL..L.AT.	..SF.....
23	afu_AF1587	yes	..YV.P.S.	.RS..I....I..	Y..PY....
20	19	yes	H.A...?.K?	..?M.....	?...?.?LH	?GT?.....
19	rci_RCIX22	yes	.G.L..G-.Y	..A.P.IS..	T.MC.GTN..	T..YA....
19	15	yes	...F..?.H	..S....A..	...?..?..	...?.....
15	16	yes?..?	A.....I.	T..A....
16	pab_PAB158	yesN...	..T.FA...A	..MI...Q..V....
16	nph_NP2770	yesD.L.Q.	.V..RC..QF	..A...H..-....
15	17	yes	?.....KLE
17	afu_AF1638	yesNA..F.....	Y.II.I.Q..	I..A.AG...
17	18	yes	...?.?RN.?.....	..VV.....
18	mja_MJ1235	yes	...M..S.DFA...I	Y..L...Q..	I.....KMEG
18	13	yes?..?	...FR....W	L..?.H..	?.....
13	14	yes	.GTY..QK..I...	..A.....	A.....

14	bbt_BBta_0	yesS.	.V.....	M.....	I.....
14	sme_SMb201	yesA.....
13	9	yes	..VID.....	..H.....	C...S.G....	?.....
9	11	yesHP?
11	12	yesK.Q..	T.....
12	pma_Pro055	maybe
12	syw_SYNW17	maybe
11	10	yesN.?	S.....
10	bbt_BBta_2	yes	...L...M.T.
10	neu_NE1921	yesH.M
9	8	maybeQ...	S.....
8	syc_syc013	maybe
8	7	yesK..
7	syn_slr000	yes	...V.....
7	4	maybe?
4	6	yesI.	T.....
6	gvi_glr215	maybe
6	5	no
5	ana_alr152	yesA
5	tel_tll150	yes	...M.....M.....
4	3	yesA
3	cre_ChreCp	yesR..M....L.
3	2	yesM.....I.
2	ppp_Phpapa	yes	...L.....
2	1	yes	A.....
1	osa_313146	yesM.....
1	ath_ArthCp	maybe
root	21		-----	--?M???EEE	V???????T?	P???IKP?LP	
21	cte_CT1772	maybeR.MTP...	..LENVIEC.K	..MGR...C..	
21	22	maybeE.....	..AI?V...R	..LD?.....	
22	ret_RHE_PF	yesKYW.PD.SF	..SSFKA.VS.P	..LF.AADCP..	
22	20	yes?..???.	??..?..???	...?..?..	
20	23	yesV?..?	??..?..???	...D.....	
23	olu_OSTLU_	yesRFGFSVD.	CK..N.GCT.	VMG.LPAI..	
23	afu_AF1587	yesKAPMME.K	YIEVAKQHRY	..FYH...CF.	
20	19	yes??.	??..?..??LR.	..?.....V..	
19	rci_RCIX22	yesK..RNVC.	ID.SRDI..K	..WAGY.R.W.	
19	15	yes?.?D?..	..?..?..?	B.....	
15	16	yes	I??IN?...	...R.....	
16	pab_PAB158	yesKMAG.Y..	..KK..DF.L.	KWEH....F.	
16	nph_NP2770	yesLGKLEN.D	TAG..EW...	DLHGHS...D	
15	17	yes	G?K..TV???	R???.V..?	D..???.SQ	...D....?	
17	afu_AF1638	yes	..Q.WD..QNA	..IFS..EYTP	..EGDAFH...	NFHH...AM.	
17	18	yes	..E.....???	..?..D.N..	..????IFFN.	DW?.....	
18	mja_MJ1235	yesE.KAI	..DEIVY.KV.	ADNENK....	..F.....F.	
18	13	yes	..DR?T.LGY.	D.?R..?..	..?A?G...B.	..A??R...	
13	14	yes	..P?..?..Y	..C.....V	..L...L..E.	...L.K.M.	
14	bbt_BBta_0	yes	...N..R...	..I...H.PM	A.EY.....	H..S.N.L..	
14	sme_SMb201	yes	...P.VQ...	NV...MK.E.	..PR.....	...D.....	
13	9	yes	...?.....	..?M...??.	..RSR.....	...SMPG...	
9	11	yes	...?.....SF.?D.	..G.....?	
11	12	yes	...Q.....I	..?L...VP.N....FA	
12	pma_Pro055	yesN.....	..T.....	
12	syw_SYNW17	maybeQ.....	
11	10	yes	...EA...W.?	
10	bbt_BBta_2	yesI..V	..L...P.VP.	N.A.....	...A...M.	
10	neu_NE1921	yesI	..I..D..IK.M...V.	
9	8	yes	...A?...FV	..L.....E.T.	
8	syc_syc013	yes	..K.S.....HI.AV....L.	

8	7	yes	.E..I.?...YV..M.
7	syn_slr000	yes	...G..M...Y.....T..
7	4	yes	?.....?Y
4	6	yesL.....?A
6	gvi_glr215	yes	...S.M...EH..R	.L.....
6	5	yesN...Q
5	ana_alr152	yes	...G..M...K.....L.....
5	tel_tll150	yes	.DK.V.L...I..
4	3	yes	...E?.L...D...K?
3	cre_ChreCp	yes	...V.....	..M.....C.....
3	2	yesL.....?V.?.L.
2	ppp_Ph papa	yes	...QV.....I..L.....
2	1	yes	...?......F..
1	osa_313146	yes	...M.....FI..	..A.....	...M...I.
1	ath_ArthCp	yes	.D..S.....V..L.....
root	21		VPGG??SA??	?????YE??G?	-?D??F??G?	G??GHP?GP?
21	cte_CT1772	maybeSD..LT	LQTV..KV.N	.V.FG.VP.R	.VF...M..K
21	22	yesG?..??	VP?.....?	.I.LV.L..G	.I.....?..?
22	ret_RHE_PF	yes	.A.S.QWGGQ	A.ET..RT.R	T...LY.C..	..VS..G..A
22	20	yes	..?.L.P..	...?????..	.?.L.A..	..?.?..?
20	23	yes	?...??..???.....	..??.....A
23	olu_OSTLU_	yes	S.G..MTLER	IAQMREVV-	.P..L..I..	SLYS.SENLV
23	afu_AF1587	yes	M.S..IA.IM	..KLVNTL-	.K.F.VA...	..HA..D...
20	19	yes	.SS...Y..?D?..Q...	..?.....TT
19	rci_RCIX22	yesQK	.RENLCY.-	.I.VI.....	..H...E...
19	15	yesH.G?	..??..??..	.K...I....	.?.....?..?
15	16	yes	.A.....?	..L.....D...
16	pab_PAB158	yesL	M.E.IRLF.-VM.....PR
16	nph_NP2770	yesI	.DQ.L.AL.-	.TNVMV....	..IH.....E
15	17	yes????
17	afu_AF1638	yesN	LEPVI.AL.-	..EI...V..	..VL...M.AK
17	18	yes?..????	.?.....?
18	mja_MJ1235	yesV..RL	..KIVEIL.-	.R..I..A..	..VH...D..R
18	13	yes	.A...I.A.	M..L.....	.?..L.F..	..TL...?..??
13	14	yesQ	.HQ.LD....	..V.....	..I...L.IQ
14	bbt_BBta_0	yesNYL.-	.E.....
14	sme_SMb201	yesLF.-	.D.....M...
13	9	yesVWH	..A.V?IF..	.D.S.....W.NA
9	11	yesA...?
11	12	yesH.....S.
12	pma_Pro055	yesL.....-C.....
12	syw_SYNW17	yesV.....
11	10	yesAC
10	bbt_BBta_2	yesT.....-F.....
10	neu_NE1921	yesT.....-
9	8	yesE.....
8	syc_syc013	yes
8	7	yes
7	syn_slr000	yesC.....
7	4	yes?
4	6	yes?
6	gvi_glr215	yesLD..-	..A.....
6	5	yes
5	ana_alr152	yesE...-	..S.....
5	tel_tll150	yesD...-	..A.....
4	3	yes
3	cre_ChreCp	yesAC
3	2	yesT.....-	..S.....
2	ppp_Ph papa	yes

2	1	yes
1	osa_313146	yes	-
1	ath_ArthCp	yes	-
root	21		AGA??RQAW	EA??QGI?--	-----	?E??A?THPE
21	cte_CT1772	maybe	...KSI....	..IE...S..	I.TW.E....
21	22	maybe	...A.....	..?....?	L.A*.K....
22	ret_RHE_PF	yes	..VR.VQ...	Q.AVA..P..VY..D...
22	20	yesV	?...??????
20	23	yes	...R.F....	?...???	*?..E?..?
23	olu_OSTLU_	yes	D...H.MKLA	GRKE----	..L	YGPL.CP.DG
23	afu_AF1587	yesI	D.AM..YTDL	R.....	KY.E.NNLQ.
20	19	yes	...K.?.?	...??..?E.....
19	rci_RCIX22	yes	V.V..MF...	E.WQ.QKT..Y...K.
19	15	yes?	...??..??..??P.
15	16	yesL.??	D...E.??SL..
16	pab_PAB158	yesD.I	..AI..LD..K...S..
16	nph_NP2770	yes	...R...A..	..YAD.ES..DSR.E.V.A
15	17	yesL	..I..??...
17	afu_AF1638	yesV.Q.	D..ISAIP..H..QH..
17	18	maybe?	E.....??..L..
18	mja_MJ1235	yes	...M.A.I	...IEGKS..K.EEVA.
18	13	yes	...T.N.V..	..?.ARNEG	RDIA?EG??I	?..A...S..
13	14	yesM.L.....	...H..PE.CTP
14	bbt_BBta_0	yesI.....	..YV.....	.AK..AT...
14	sme_SMb201	yesV.....RA...W.K.
13	9	yesCV.....	...R..R?	?.....
9	11	yes	...A.....?	?..E?	?...H...
11	12	yesK...A.	.E..K.S.D.	.M.....
12	pma_Pro055	maybe
12	syw_SYNW17	yesG.....
11	10	yesE...Q.	.P.....?	.T.....
10	bbt_BBta_2	yes	...H.....V.R...E.	...Q....
10	neu_NE1921	yesR.	V...K..KA.
9	8	yes	P.....	...Q.....	..L...GD.	.R..G.W...
8	syc_syc013	yesY.....
8	7	yes	I...??...
7	syn_slr000	yesN...N.V	...CR....
7	4	no
4	6	yes?	...A....
6	gvi_glr215	yesK.....	..M..A..	...R....
6	5	no
5	ana_alr152	yesN.A...N.V	...K....
5	tel_tll150	yesI.....	..M.....	...R....
4	3	yes	...A.....K....
3	cre_ChreCp	yesT.....V	..S.C....
3	2	yes	...?.....NE.
2	ppp_Ph papa	yes	...V.....A....
2	1	maybeC....
1	osa_313146	yes	...A.....S.....
1	ath_ArthCp	yes	...V.....V....
root	21		L?A???QS??	?K??-----	-----	--
21	cte_CT1772	maybe	.Q.MVD..LL	K.QD.....
21	22	yes	...?....?	?..??
22	ret_RHE_PF	yes	.A.SIAKFSD	G.GA.....
22	20	yes	..?A??...??	?.....
20	23	yes	..??...??	..?
23	olu_OSTLU_	yes	YRGHTM----	-----
23	afu_AF1587	yes	.LK.LQL---	-----

20	19	yesL..WG? ?.....
19	rci_RCIX22	yes	.AG..K...P SQ--.....
19	15	yes	.?A..EK..? ?.P?.....
15	16	yes	.?T..D...? ?..K.....
16	pab_PAB158	yes	.KK.....Y L.....
16	nph_NP2770	yes	.R....E..T QN.R.....
15	17	maybe	.?.....? ??.....
17	afu_AF1638	yes	.Q.....R VT.I.....
17	18	yes	.?....?.K? ?.?.....
18	mja_MJ1235	yes	.KK...Y..- ----.....
18	13	yes	.?....I..? ?.F.?D??D? ??.....
13	14	yesD...N .T.NY.?T.T ?DFVP?????
14	bbt_BBta_0	yes	.KQ....V... V.....S... P.....TAAVT A.
14	sme_SMb201	yes	.E.....G. IS...TP... S.....VTAA
13	9	yes	.?.....E IK.EF.TM.K L.....
9	11	yes	..I...T... ..V... .DV?.....
11	12	yes	.A..... ..Q.....
12	pma_Pro055	no
12	syw_SYNW17	yesN.....
11	10	yes	.K..M.....???. ..
10	bbt_BBta_2	yesV... ..TGPMLRVV NA
10	neu_NE1921	yesAHK.....
9	8	yes	.A....L... ..E.... ?.....
8	syc_syc013	yesD.... ..-.....
8	7	yesC..... ..?.....
7	syn_slr000	yesA..T .-.....
7	4	yes?.....
4	6	yesA.....
6	gvi_glr215	yesY..V.. .-.....
6	5	yes?..T ?.....
5	ana_alr152	yes	..V.....M.. V-.....
5	tel_tll150	yesQ.. I-.....
4	3	yesV... ..?T?... .D.....
3	cre_ChreCp	yesD.I.. .-.....
3	2	maybe?.....
2	ppp_Ph papa	yesD.V.T .-.....
2	1	yes?..... ?.....
1	osa_313146	yesI..A ..EPV..
1	ath_ArthCp	yesT.N.P.I.. ..GQE.....

Tabla 4. Información sobre los cambios de los caracteres en el cladograma de la figura 12.

RESULTADOS

Cladograma de secuencias *rbcl* (Fig. 13).

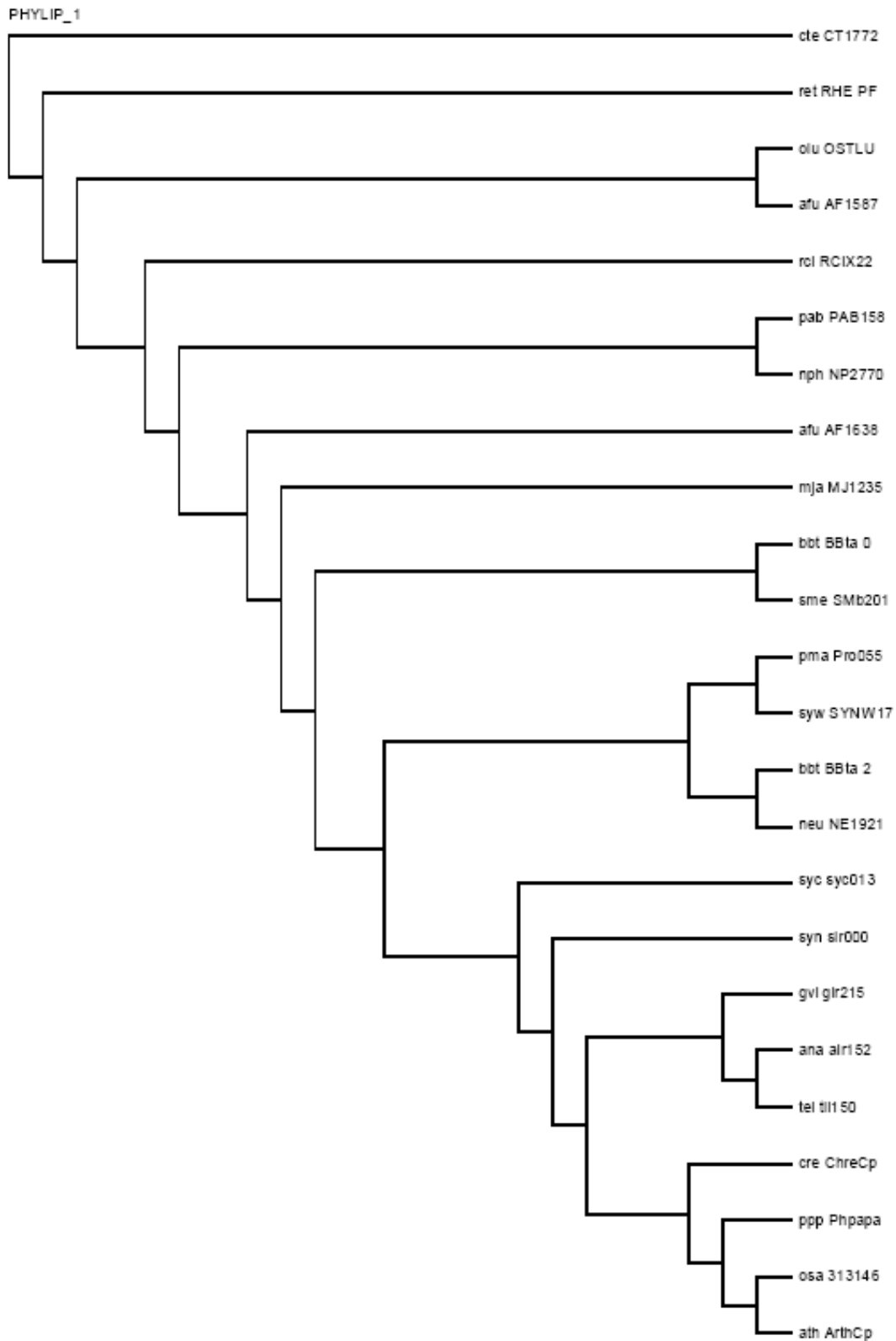


Figura 13. Cladograma de secuencias *rbcl*. Longitud 5037 cambios.

Cladograma de secuencias porfobilinógeno sintasa (Fig. 14)

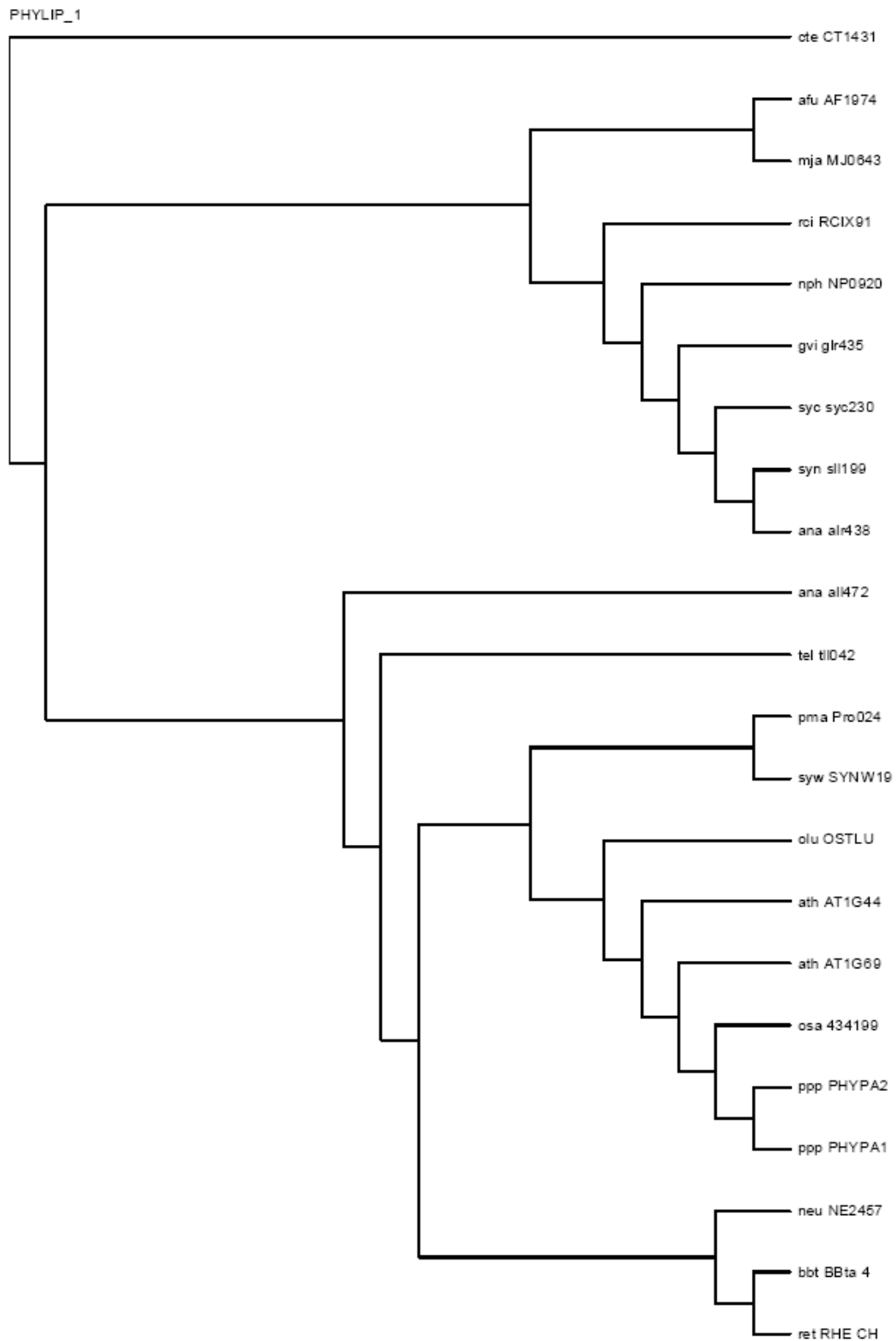


Figura 14. Cladograma de secuencias de porfobilinógeno sintasa.

Cladograma de secuencias glutamil-tRNA sintasa (Fig. 15).

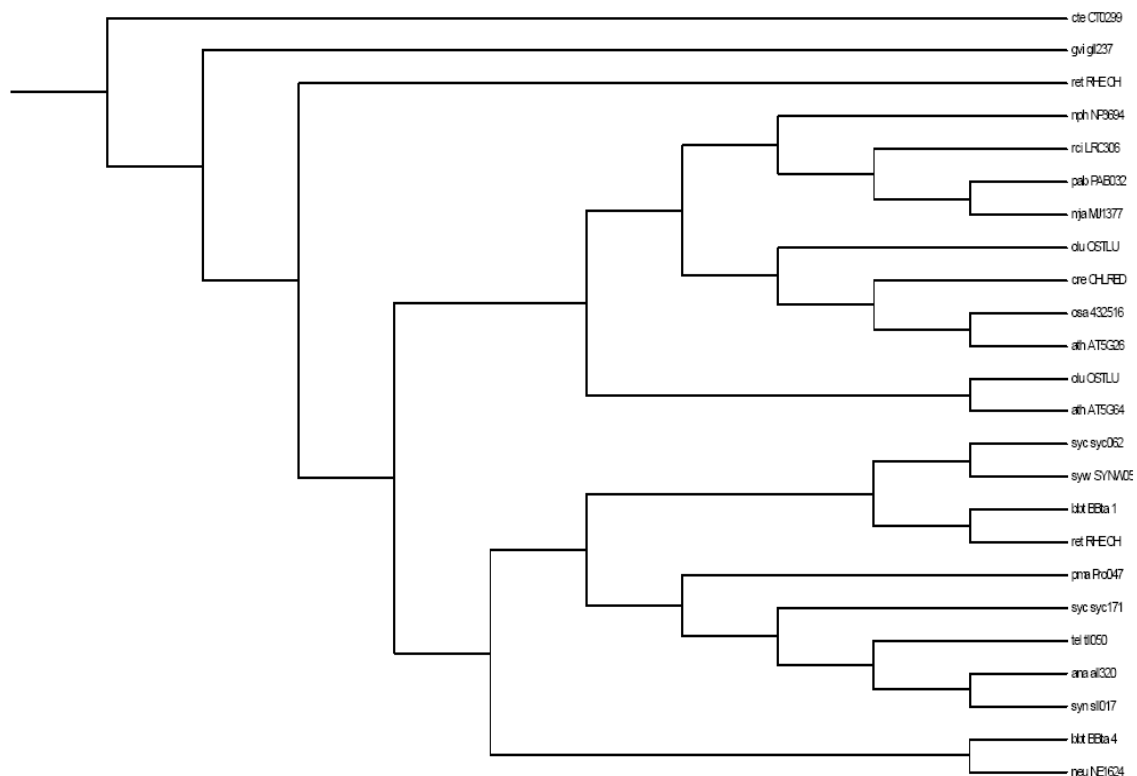


Figura 15. Cladograma de secuencias de glutamil-tRNA sintasa.

CUESTIONES Y EJERCICIOS

1. ¿Qué es el cladismo?
2. ¿Qué son grupos hermanos?
3. ¿Qué es un cladograma?
4. ¿Qué es una matriz básica de datos?
5. ¿Qué tipo de análisis realiza el programa PROTPARS? ¿Qué resultados proporciona?
6. Realizar análisis cladístico de secuencias porfobilinógeno sintasa y glutamil-tRNA sintasa del grupo de estudio.
7. Analizar los cladogramas de rbcl, porfobilinógeno sintasa y glutamil-tRNA sintasa, tanto análisis individual como comparativo.
8. A continuación tienes una serie de preguntas sobre los resultados obtenidos en los ejercicios que acabas de hacer:
 - a) ¿Qué clados se distinguen en el árbol más parsimoniosos de rbcl, porfobilinógeno sintasa y glutamil-tRNA sintasa?
 - b) ¿Son coherentes los grupos anteriores?, ¿se trata de grupos monofiléticos?
 - c) ¿Qué grupos presentan secuencias plesiotípicas?
 - d) ¿Qué explicación se puede dar a la presencia de rbcl plastidial de "ath" entre cianobacterias?

- e) ¿Qué explicaciones pueden darse a la semejanza de *rbcl* de arqueas (Archaea) y otras bacterias (proteobacterias y bacteria verde-azufrada) con la *rbcl* cianobacteriana?
9. ¿Qué indica la longitud del árbol?
11. ¿Cuál es la longitud de los cladogramas de las secuencias porfobilinógeno sintasa y glutamil-tRNA sintasa?. ¿Se obtienen más de un cladograma idénticos o de diferente longitud?
10. Interpretar los cladogramas resultantes. Puedes hacer uso de los siguientes elementos para la discusión:
- ¿Qué organismos/especies componen el grupo de estudio?
 - Características básicas de los organismos en estudio:
 - ✓ procariotas: archaea y bacterias.
 - ✓ eucariotas: plantas, algas.
 - **Archaea:** filogenéticamente más próximas a eucariontes que a bacterias; no contienen peptidoglicano ni celulosa en la pared (como bacterias) sino otros polisacáridos, glicoproteínas y proteínas; aerobias y anaerobias; autótrofas (productores primarios usando CO₂ o CH₄ como fuente de C y obteniendo energía de la oxidación de moléculas inorgánicas como el SH₂, H₂, CH₄; no fijan C a través de fotosíntesis, son quimiosintéticas ó quimiolitioautótrofos.
 - **Chlorobiaceae:** Bacterias verdes del Azufre, fotosíntesis anoxigénica; utilizan SH₂ ó S como donantes de e⁻; de zonas ricas en S; anaerobias de lagos y algunas fumarolas de profundidades oceánicas.
 - **Musgos:** plantas no vasculares; cloroplastos con *chl a* y *chl b*, carotenoides, almidón y celulosa, no poseen lignina.
 - **Algas:** fotosíntesis oxigénica.
 - **Cianobacterias:** fotosíntesis oxigénica.
 - **α-Proteobacterias (Bradyrhiz. Y Rhiz.):** bacterias púrpura, fotosintéticas (fotoautótrofas); bacterioclorofila y carotenos; en simbiosis o de vida libre.
 - **β-Proteobacterias (Nitrosomonas):** quimioautótrofas; quimiosintéticas transforman amonio en nitrito.

BIBLIOGRAFÍA

- Cavalier-Smith, T. 1992. Bacteria and Eucaryotes. *Nature*, 356: 570.
- Granick, S. 1965. Evolution of heme and chlorophyll. Pp: 67-88. En: Bryson, V. y H.J. Vogel (Eds.) *Evolving Genes and Proteins*. Academic Press, New York.
- Felsenstein, J. 2001. PHYLIP (phylogeny inference package) version 3.6a3. Department of Genetics. University of Washington, Seattle.
- Woese, C.R. 1987. Bacterial evolution. *Microbiol. Rev.*, 51: 221-271.

BIBLIOGRAFÍA DE CONSULTA

- Ávalos, A.; Costa, M.; Moreno, M. y Pérez-Urria, E. 2007. Evolución y Filogenia: Parte I. Editorial Complutense, Madrid.
- Ávalos, A.; Costa, M.; Moreno, M. y Pérez-Urria, E. 2009. Evolución y Filogenia: Parte II. Vicerrectorado de Desarrollo y Calidad de la Docencia, Universidad Complutense de Madrid.
- Blankenship, R.E. 2001. Molecular evidence for the evolution of photosynthesis. Trends in Plant Science, 6 (1): 4-6.
- Des Marais, D.J. 2000. When did photosynthesis emerge on Earth? Science, 289: 1703-1705.
- Doolittle, W.F. 1999. Phylogenetic classification and the universal tree. Science, 284: 2124-2128.
- Forey, P.L.; Humphries, C.J.; Kitching, I.L.; Scotland, R.W.; Siebert, D.J. y Williams, D.M. 1992. Cladistics: A Practical Course in Systematics. Clarendon Press, Oxford.
- Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press. Chicago.
- Lazcano, A. Y Miller, S.L. 1999. On the origin of metabolic pathways. J. Mol. Evol., 49: 424-431.
- Olson, J.M. 2001. Evolution of Photosynthesis 1970, re-examined thirty years later. Photosynth. Res., 68:95-112.
- Pérez-Urria, E. 2009. Evolución del Metabolismo: desde las Porfirinas hacia la Fotosíntesis. Sociedad Española de Biología Evolutiva, 4(1): 37-41.

RECURSOS ELECTRÓNICOS

Clustalw

<http://align.genome.jp/>

European Bioinformatics Institute

<http://www.ebi.ac.uk/>

ExPASy Proteomics Server

<http://www.expasy.ch/>

Genbank

<http://www.ncbi.nlm.nih.gov/Genbank/index.html>

GenomeNet

<http://www.genome.ad.jp/>

Manet

<http://www.manet.illinois.edu/viewmap.php?release=2&map=map00860&mb=Metabolism%20of%20Cofactors&page=cofactors.php>

Metabolic Pathways

http://www.sigmaaldrich.com/img/assets/4202/MetabolicPathways_6_17_04_.pdf

Nacional Center for Biotechnology Information (NCBI)

<http://www.ncbi.nlm.nih.gov/>

PHYLP: Joe Felsenstein 1986-1995)

<http://evolution.genetics.washington.edu/phylip.html>

ReadSeq

<http://www-bimas.cit.nih.gov/molbio/readseq/>

TreeView

<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>

Recibido: 29 mayo 2009.

Aceptado: 6 julio 2009.