



Construcción y edición de árboles filogenéticos

Día 2

Contenido del curso

1. Conceptos clave
2. Preparación de datos y herramientas
3. Reconstrucción de árboles filogenéticos
4. Edición e interpretación de árboles filogenéticos



Día 2: Preparación de datos y herramientas

18:05 - 18:20

- Repaso

18:20 - 18:50 *

- Enfoque y planteamiento del proyecto

19:00 - 19:20

- Exploración de bases de datos de secuencias
- Descarga de secuencias

19:20 - 19:50

- Instalación de herramientas bioinformáticas
- Registro de usuario en las herramientas online

19:50 - 20:00

- Dudas y comentarios finales



Repaso



Diagrama de flujo para reconstruir un árbol filogenético

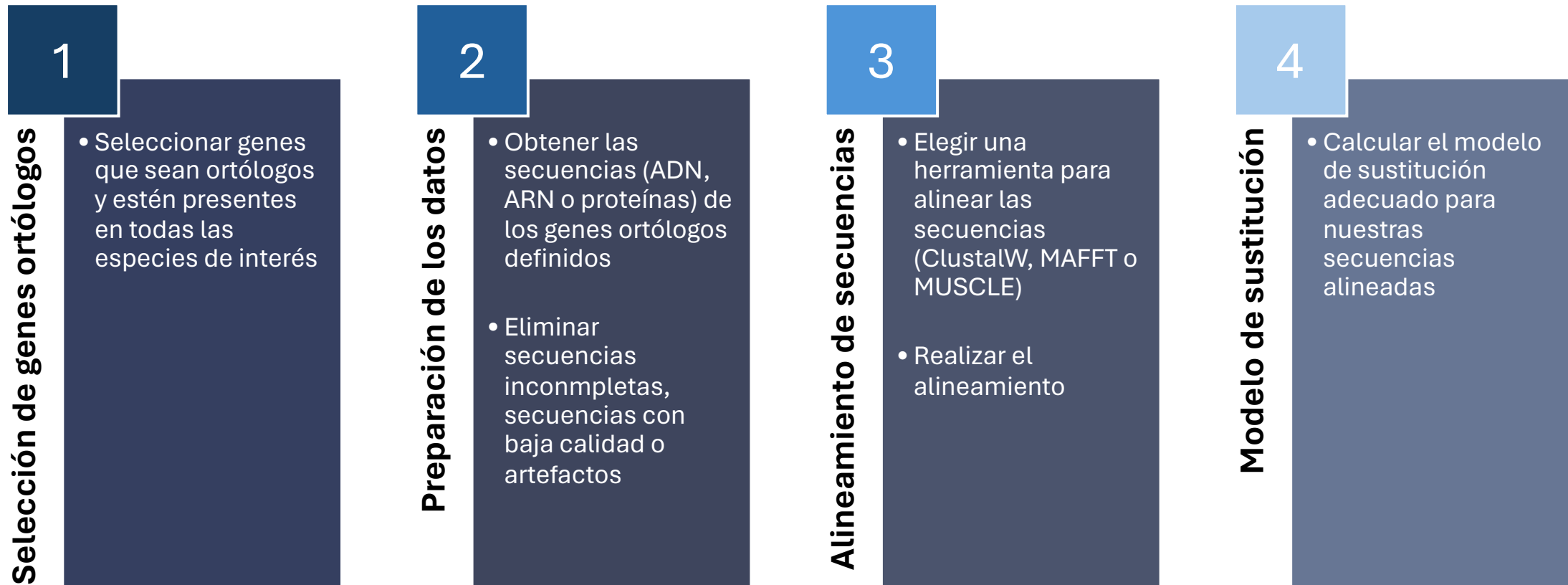


Diagrama de flujo para reconstruir un árbol filogenético

5

Reconstrucción filogenética

- Elegir uno de los métodos de reconstrucción filogenética (Maximum likelihood, Neighbor joining, parsimonia)
- Ajustar los parámetros (modelo de sustitución, bootstrap, particiones) y ejecutar el análisis

6

Evaluación del árbol

- Realizar análisis de bootstrap para evaluar la robustez del árbol obtenido
- Comparar el árbol con reportes previos o con conocimiento biológico previo

7

Visualización e interpretación

- Visualizar el árbol filogenético con herramientas especializadas
- Analizar y discutir la topología del árbol

8

Refinamiento e iteración

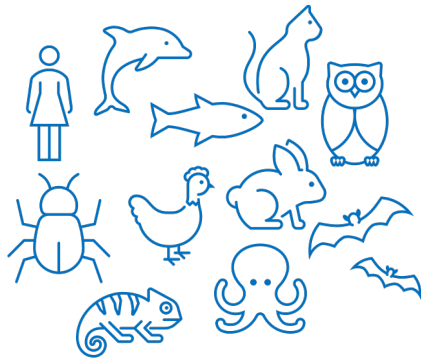
- Realizar ajustes en los parámetros, el modelo de sustitución o los datos según sea necesario

¿Qué tipo de árbol filogenético
vamos a construir en este curso?

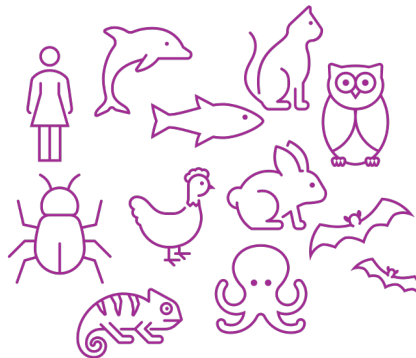
Enfoque del proyecto

Plantear una pregunta real que podamos resolver mediante la reconstrucción de un árbol filogenético, el cual nos permita comprender la evolución biológica de una **población** o de un **gen**

Árbol de genes



Árbol de especies



Tipos de árboles filogenéticos

Árbol de genes

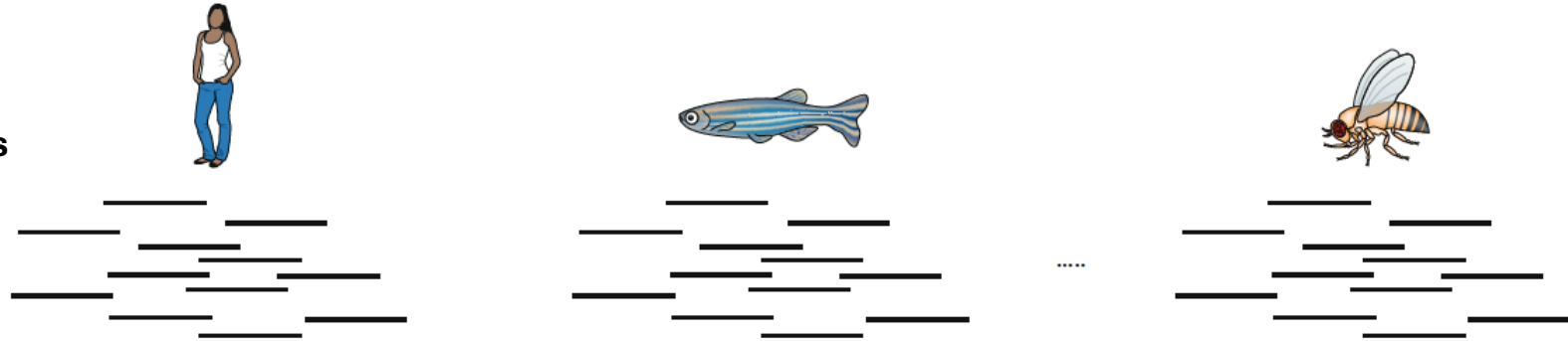
El árbol filogenético o genealógico de las secuencias de un solo gen o región genómica

Árbol de especies

Un árbol filogenético para un conjunto de especies que unifica los árboles de genes en uno solo

Proceso general y simplificado para construir un árbol de especies

Ensamblado de secuencias



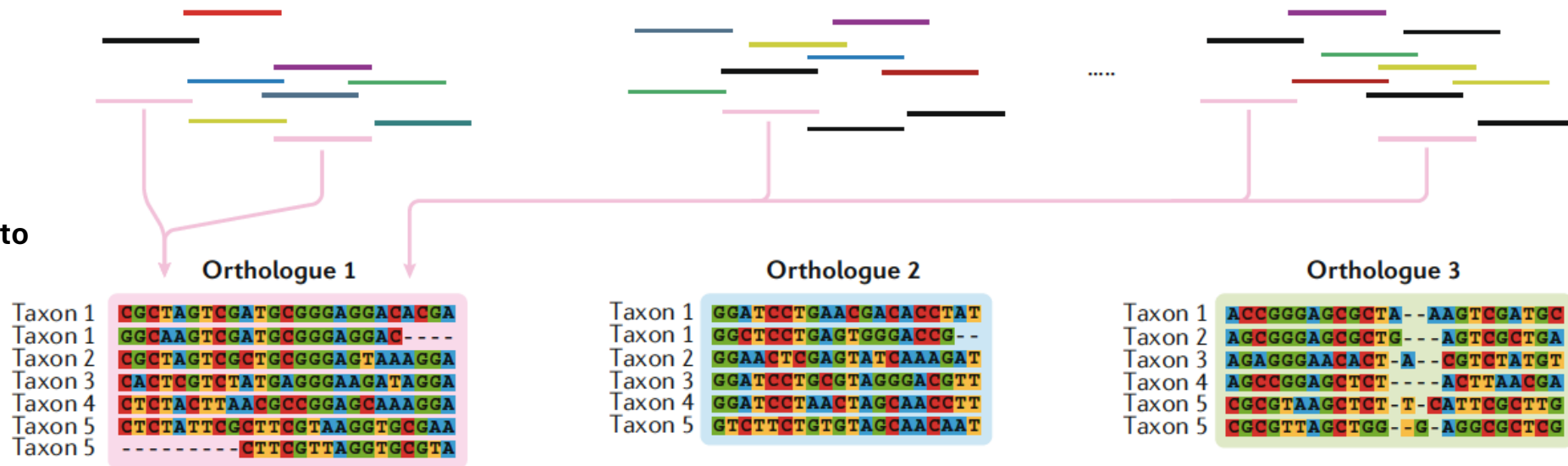
Identificación de secuencias contaminantes



Proceso general y simplificado para construir un árbol de especies

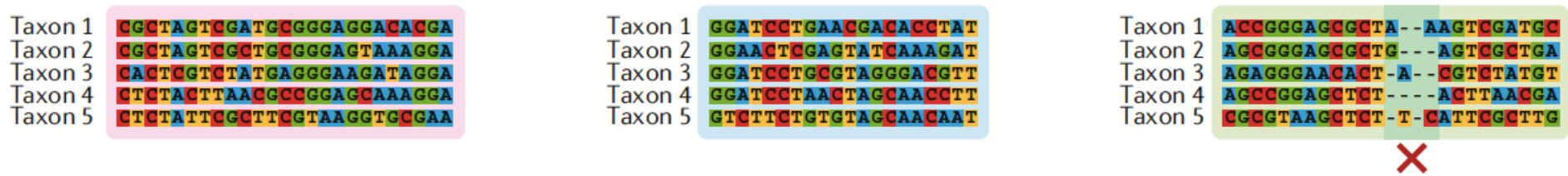
Predicción de ortólogos

Alineamiento

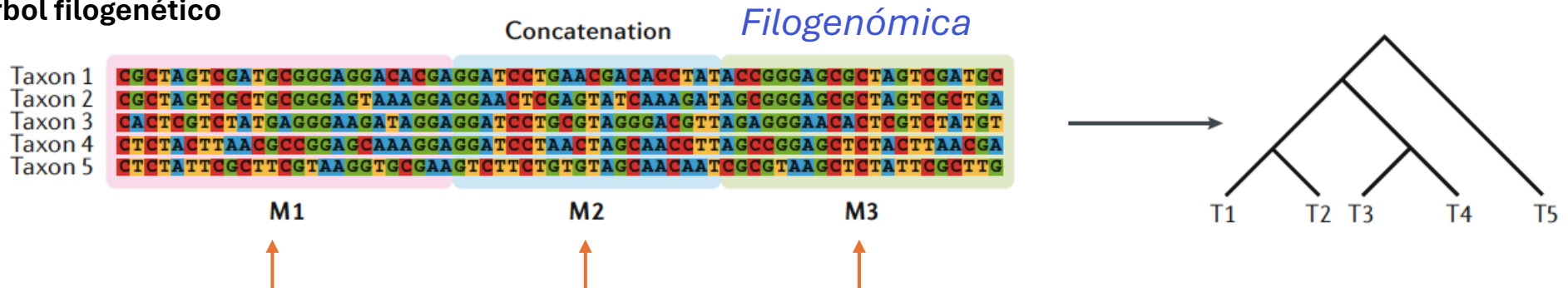


Proceso general y simplificado para construir un árbol de especies

Filtrado de sitios o secuencias



Inferencia del árbol filogenético



¡Modelo de sustitución específico para cada gen!

Heterogeneidad genealógica de los genes

Cada gen cuenta su propia historia evolutiva, derivada de los múltiples procesos biológicos que realizan.

Tratar de representar un árbol de especies usando un solo gen no nos asegura capturar la verdadera historia evolutiva entre las especies.

A este fenómeno se le conoce como:

- Incomplete lineage sorting (ILS)
- Deep coalescence
- Gene-tree-species-tree incongruence



Multispecies coalescent (MSC) model

- Lidia con el fenómeno de ILS
- Bajo el modelo MSC, los árboles de genes (topologías y longitudes de las ramas) pueden ser diferentes entre los genes o regiones genómicas debido al proceso de coalescencia en las especies ancestrales: tienen una distribución estadística determinada por el árbol de especies y por parámetros como los tiempos de divergencia entre especies y los tamaños de población.

El término "coalescente" se refiere a cómo las líneas genéticas de diferentes individuos en una población se retroceden en el tiempo hasta encontrar un ancestro común.

Multispecies coalescent (MSC) model

Métodos de dos pasos

1. Construir árboles de genes por loci
2. Usa los árboles de genes como datos para el árbol de especies



Accurate Species Tree Estimator (ASTER*)

Métodos full likelihood

Calculan la verosimilitud de los alineamientos y, por lo tanto, tienen en cuenta las incertidumbres en los árboles de genes.



Otros retos en la reconstrucción de árboles filogenéticos de especies

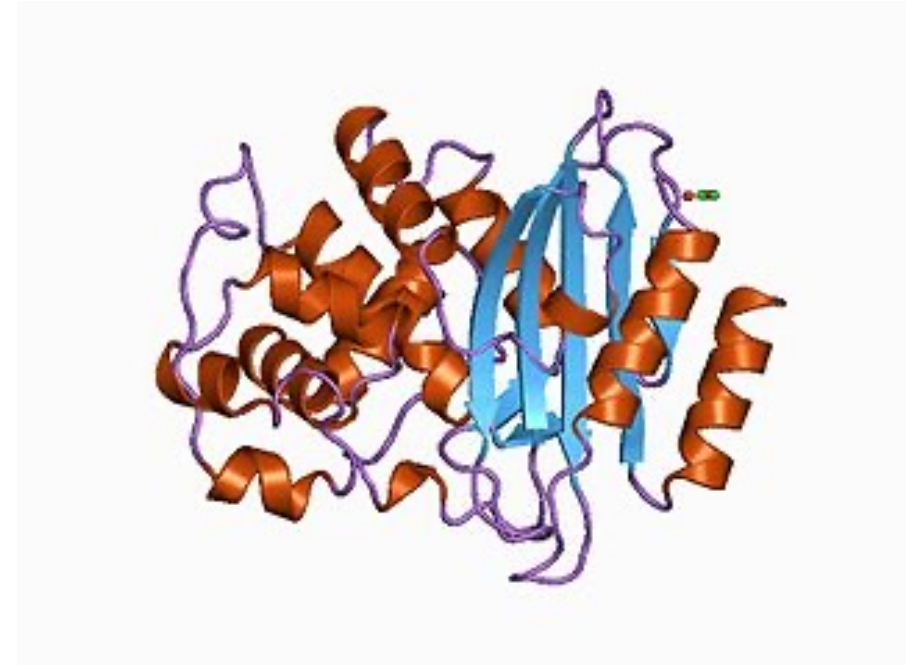
- **Muestreo sesgado de especies**
- **Homoplasia** (ocurre cuando caracteres similares evolucionan de forma independiente en diferentes linajes)
- **Convergencia evolutiva** (similitudes entre especies que no comparten un ancestro común)
- **Especies híbridas**
- **Falta de datos genéticos**



Planteamiento del proyecto

Árbol de genes: **Beta-lactamasas**

- Es una proteína presente en algunas bacterias que se encarga de degradar los anillos beta-lactámicos, que forman parte de varios antibióticos incluyendo las penicilinas, cefalosporinas, monobactámicos y carbapenémicos



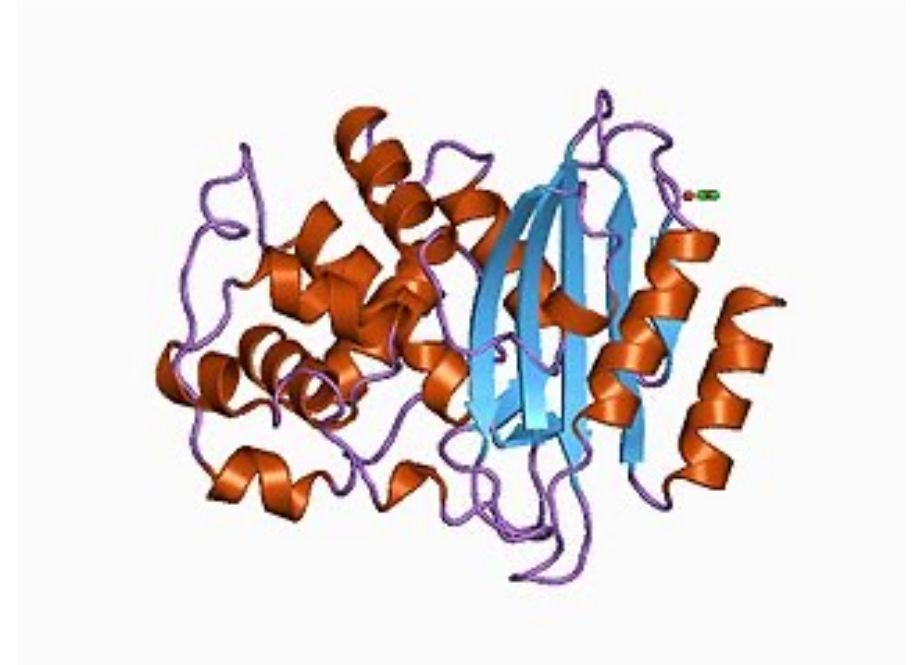
Árbol de genes: **Beta-lactamasas**

Pregunta:

¿Cómo han evolucionado las beta-lactamasas en bacterias?

Objetivo:

Crear un árbol filogenético del gen de las beta-lactamasas considerando únicamente genes ortólogos en distintas bacterias



Exploración de bases de datos de secuencias y descarga de secuencias

Principales bases de datos de secuencias genéticas y proteicas

Secuencias:

- [NCBI protein](#)
- [NCBI nucleotide](#)
- [KEGG](#)

Grupos de secuencias (ortólogos):

- [UniProt](#)
- [EnsemblGenomes](#)
- [OrthoMCL](#)





UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt](#)”

UniProt invites you to participate in a survey on the use and value of UniProt. [Please click on the link provided to complete the survey](#) .

Proteins
UniProt Knowledgebase

Reviewed (Swiss-Prot)
571,282

Unreviewed (TrEMBL)
248,234,451

Species
Proteomes

Protein sets for species with sequenced genomes from across the tree of life

Protein Clusters
UniRef

Clusters of protein sequences at 100%, 90% & 50% identity


Sequence Archive
UniParc


Non-redundant archive of publicly available protein sequences seen across different databases


Feedback



1. <https://www.uniprot.org>
2. Protein Clusters
3. Seleccionar el nombre de una proteína
4. Filtrar por categorías
5. Alinear las secuencias
6. Descargar el alineamiento


[BLAST/BLAT](#) | [VEP](#) | [Tools](#) | [BioMart](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

 Search all species...

 **Human** (GRCh38.p14) ▼

Location: 13:32,315,086-32,400,268 Gene: **BRCA2**

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- Sequence
 - Secondary Structure
- Comparative Genomics
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Ontologies
 - GO: Molecular function
 - GO: Biological process
 - GO: Cellular component
- Phenotypes
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- Gene expression
- Pathway
- Molecular interactions
- Regulation
- External references
- Supporting evidence
- ID History
 - Gene history

Gene: BRCA2 ENSG00000139618

Description
 BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:[HGNC:1101](#)]

Gene Synonyms
 BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11

Location
[Chromosome 13: 32,315,086-32,400,268](#) forward strand.
 GRCh38:CM000675.2

About this gene
 This gene has 15 transcripts ([splice variants](#)), [172 orthologues](#) and is associated with [123 phenotypes](#).

Transcripts
[Show transcript table](#)

Summary ⓘ

Name
[BRCA2](#) (HGNC Symbol)

MANE
 This gene contains MANE Select [ENST00000380152](#), [ENSP00000369497](#)

UniProtKB
 This gene has proteins that correspond to the following UniProtKB identifiers: [P51587](#)

RefSeq
 This Ensembl/Gencode gene contains transcript(s) for which we have [selected identical RefSeq transcript\(s\)](#). If there are other RefSeq transcripts will be in the [External references](#) table

CCDS
 This gene is a member of the Human CCDS set: [CCDS9344.1](#)

LRG
[LRG_293](#) provides a stable genomic reference framework for describing sequence variants for this gene

Ensembl version
 ENSG00000139618.18

Other assemblies
 This gene maps to [32,889,223-32,974,405](#) in GRCh37 coordinates.
 View this locus in the GRCh37 archive: [ENSG00000139618](#)

Gene type
 Protein coding

Annotation method
 Annotation for this gene includes both automatic annotation from Ensembl and Havana manual annotation

Annotation Attributes
 overlapping locus [Definitions](#)

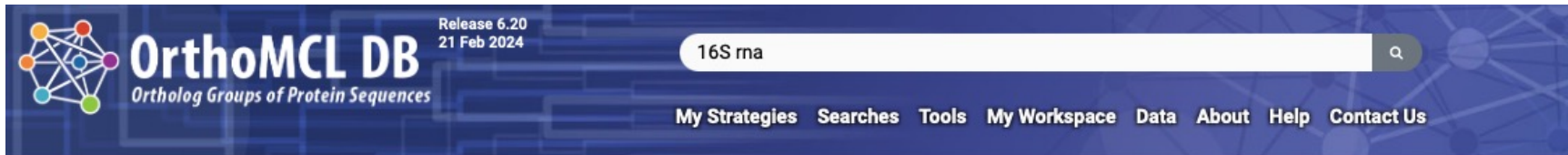
[Configure this page](#)

[Custom tracks](#)

[Export data](#)

1. <https://www.ensembl.org/index.html>
2. Seleccionar una especie y un gen
3. Ir a sección **Orthologues** del panel izquierdo
4. Filtrar la tabla
5. Descargar la lista de genes ortólogos

Búsqueda por nombre del gen o proteína



All results matching 16S rna


☒ Hide zero counts

Filter results

Genome
Protein Sequences 226,664

Orthology
Ortholog Groups 9,708

Filter fields
Select a result filter above



1 - 20 of 236,372



Protein Sequence - bovi|BaOVIS_009450 16S ribosomal protein RNA methyltransferase

Ortholog group: OG6_118254
PFam Domains: RNA methyltransferase (PF04452)
Taxon Name: Babesia ovis Selcuk

► Fields matched: PFam Domains; Product

Protein Sequence - bova-old|BOVATA_000600 16S rRNA methyltransferase

Ortholog group: OG6_118254
PFam Domains: RNA methyltransferase (PF04452)
Taxon Name: Babesia ovata strain Miyake (old build 2017-12-16)

► Fields matched: PFam Domains; Product

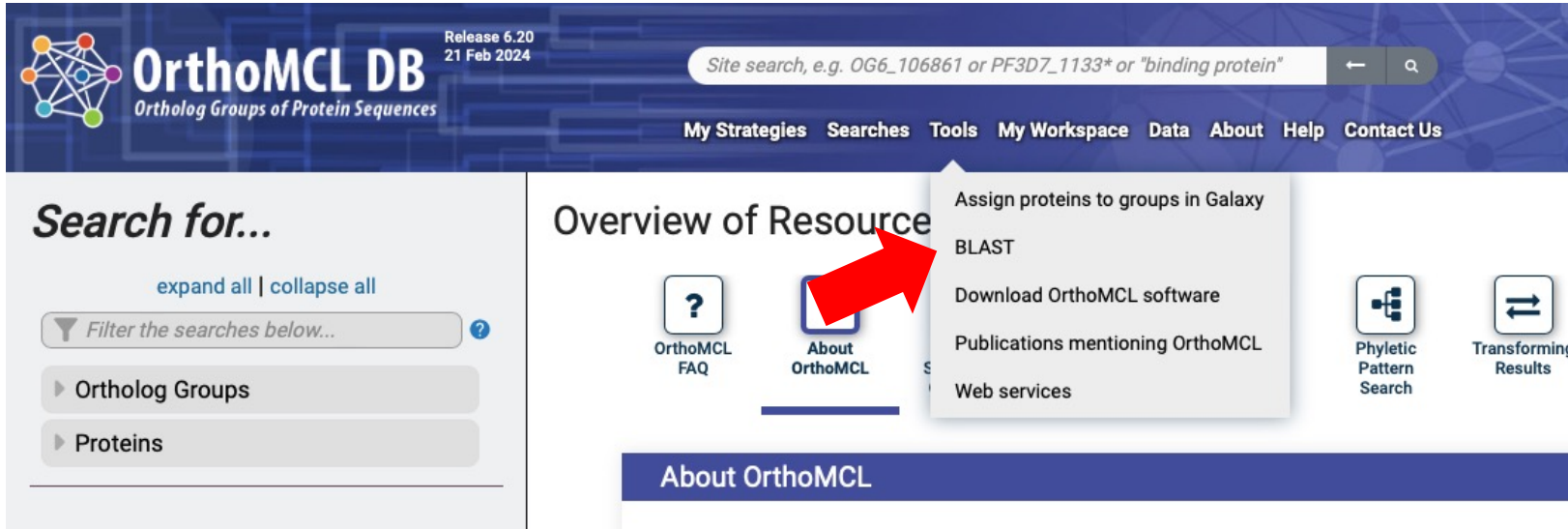
Protein Sequence - nfow-old|NF0087150 16s rna methyltransferase

Ortholog group: OG6_106261
PFam Domains: RNA methyltransferase (PF04452)
Taxon Name: Naegleria fowleri ATCC 30863 (old build 2014-11-05)

► Fields matched: PFam Domains; Product

1. <https://orthomcl.org/orthomcl/app>
2. Buscar el nombre de una proteína
3. Seleccionar la opción de Orthologues
4. Seleccionar el nombre de una proteína
5. Filtrar por categorías
6. Alinear las secuencias
7. Descargar el alineamiento

Búsqueda por secuencia - BLAST



OrthoMCL DB
Ortholog Groups of Protein Sequences

Release 6.20
21 Feb 2024

Site search, e.g. OG6_106861 or PF3D7_1133* or "binding protein"

My Strategies Searches Tools My Workspace Data About Help Contact Us

Search for...

expand all | collapse all

Filter the searches below...

- Ortholog Groups
- Proteins

Overview of Resources

- Assign proteins to groups in Galaxy
- BLAST**
- Download OrthoMCL software
- Publications mentioning OrthoMCL
- Web services

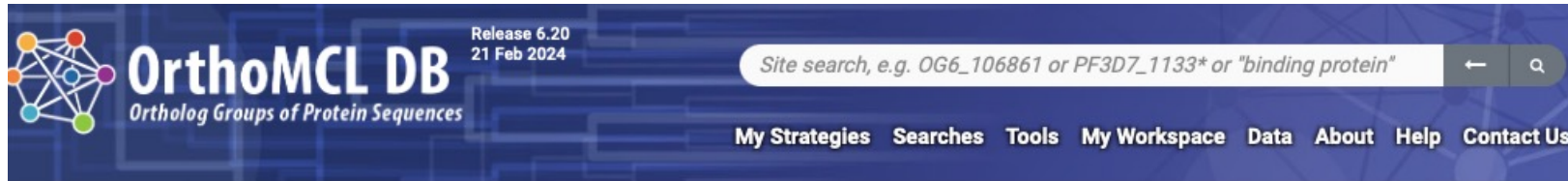
OrthoMCL FAQ About OrthoMCL

Phyletic Pattern Search Transforming Results

About OrthoMCL

1. <https://orthomcl.org/orthomcl/app>
2. Seleccionar BLAST en la pestaña "Tools"

Búsqueda por secuencia - BLAST



Identify Proteins based on BLAST

Configure Search Learn More

Reset values to default

BLAST Program

☒ blastp
☐ blastx

Input Sequence

MAKSHLEIAKQLKHNSIELKEYFEDLYSWQNDIKKKEEKEK

1. <https://orthomcl.org/orthomcl/app>
2. Seleccionar BLAST en la pestaña “Tools”
3. Colocar la secuencia proteica (sin espacios, sin identificador) en el recuadro indicado y procesar
4. En la pestaña “Protein”, dar click en el botón de “Download” para descargar la tabla y/o las secuencias en formato FASTA

Instalación de herramientas bioinformáticas

Dirígete el siguiente link, donde encontrarás información para la instalación de las herramientas necesarias:

<https://github.com/paumayell/Phylogenetic-tree-reconstruction>

Dudas y comentarios finales

Contacto

Dra. Paulina M. Mejía-Ponce



paumayell@gmail.com