

#### Marco teórico.

UniProt significa "Universal Protein Resource" y es, como su nombre lo indica, un recurso central y gratuito de información sobre proteínas. Se ha convertido en la base de datos de proteínas más importante del mundo. UniProt colabora con el Instituto Europeo de Bioinformática (EMBL-EBI), el Instituto Suizo de Bioinformática SIB y el Protein Information Resource (PIR). En los tres institutos, más de 100 personas participan en diferentes tareas, como curación de bases de datos, desarrollo de software y soporte. [1]

En UniProt se encontrará información sobre:

- La secuencia de aminoácidos de la proteína, que son los bloques de construcción básicos de las proteínas.
- La función de la proteína, o qué hace en el cuerpo.
- La estructura de la proteína, o su forma tridimensional.
- Cualquier modificación que pueda tener la proteína.

La información de UniProt proviene de dos fuentes principales:

- Proyectos de secuenciación del genoma, que determinan el orden de los nucleótidos en el ADN de un organismo.
- Publicaciones en revistas científicas, donde los investigadores describen sus descubrimientos sobre proteínas específicas.

---

#### Material y equipo.

Para esta práctica se utilizaron los siguientes materiales y recursos:

- Un equipo con acceso a internet.
- La plataforma UniProt.

---

#### Desarrollo.

1. Recopilación de secuencias de HBA1 de al menos 50 especies diferentes.  
En este paso se seleccionaron y añadieron 50 secuencias de la proteína HBA1 a la canasta de UniProt (Figura 1-7).



Figura 1. Búsqueda de secuencias de HBA1.

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/> A0A2J8IND9	A0A2J8IND9_PANTR	HBA1 isoform 3 [...]	CK820_G0054685, CK820_G0054686	Pan troglodytes (Chimpanzee)	110 AA
<input checked="" type="checkbox"/> A0A2J8INE6	A0A2J8INE6_PANTR	HBA1 isoform 1 [...]	CK820_G0054685, CK820_G0054686	Pan troglodytes (Chimpanzee)	142 AA
<input type="checkbox"/> Q02777	CNR1_FELCA	Cannabinoid receptor 1 [...]	CNR1	Felis catus (Cat) (Felis silvestris catus)	472 AA
<input checked="" type="checkbox"/> Q5IS73	CNR1_PANTR	Cannabinoid receptor 1 [...]	CNR1	Pan troglodytes (Chimpanzee)	472 AA
<input type="checkbox"/> Q71SP5	CNR1_MACMU	Cannabinoid receptor 1 [...]	CNR1	Macaca mulatta (Rhesus macaque)	472 AA
<input type="checkbox"/> Q9PUI7	CNR1_TARGR	Cannabinoid receptor 1 [...]	CNR1	Taricha granulosa (Roughskin newt)	473 AA

Figura 2. Selección de las primeras secuencias de HBA1.

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/> A0A8V0ZAH8	A0A8V0ZAH8_CHICK	Hemoglobin subunit alpha-A [...]	HBA1	Gallus gallus (Chicken)	168 AA
<input type="checkbox"/> A0A8V0ZB98	A0A8V0ZB98_CHICK	Hemoglobin subunit alpha-A [...]	HBA1	Gallus gallus (Chicken)	203 AA
<input checked="" type="checkbox"/> A9L8T9	A9L8T9_PAPAN	Hemoglobin, alpha 1 (Predicted)	HBA1	Papio anubis (Olive baboon)	142 AA
<input type="checkbox"/> C1BG68	C1BG68_ONCMY	Hemoglobin subunit alpha-1	HBA1	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	143 AA
<input checked="" type="checkbox"/> C1BGX1	C1BGX1_ONCMY	Hemoglobin subunit alpha-1	HBA1	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	95 AA
<input type="checkbox"/> C1BIB0	C1BIB0_ONCMY	Hemoglobin subunit alpha-1	HBA1	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	143 AA
<input type="checkbox"/> C3KID2	C3KID2_ANOFI	Hemoglobin subunit alpha-1	HBA1	Anoplopoma fimbria (Sablefish)	143 AA

Figura 3. Selección de mas secuencias de HBA1.

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/> E1B2D1	E1B2D1_HUMAN	Hemoglobin alpha-1 globin chain variant	HBA1	Homo sapiens (Human)	100 AA
<input type="checkbox"/> E9M4D4	E9M4D4_HUMAN	Hemoglobin alpha-1 globin chain	HBA1	Homo sapiens (Human)	100 AA
<input checked="" type="checkbox"/> F7BMQ3	F7BMQ3_MACMU	Globin family profile domain-containing protein	HBA1	Macaca mulatta (Rhesus macaque)	141 AA
<input type="checkbox"/> G8ZF10	G8ZF10_RABIT	Alpha-globin 1	HBA1	Oryctolagus cuniculus (Rabbit)	142 AA
<input type="checkbox"/> G8ZF12	G8ZF12_RABIT	Alpha-globin 1	HBA1	Oryctolagus cuniculus (Rabbit)	142 AA
<input type="checkbox"/> G8ZF16	G8ZF16_RABIT	Alpha-globin 1	HBA1	Oryctolagus cuniculus (Rabbit)	142 AA
<input checked="" type="checkbox"/> G8ZF18	G8ZF18_RABIT	Alpha-globin 1	HBA1	Oryctolagus cuniculus (Rabbit)	142 AA
<input type="checkbox"/> G8ZF19	G8ZF19_RABIT	Alpha-globin 1	HBA1	Oryctolagus cuniculus (Rabbit)	142 AA
<input type="checkbox"/> G8ZF22	G8ZF22_RABIT	Alpha-globin 1	HBA1	Oryctolagus cuniculus (Rabbit)	142 AA
<input type="checkbox"/> G8ZF24	G8ZF24_RABIT	Alpha-globin 1	HBA1	Oryctolagus cuniculus (Rabbit)	142 AA
<input type="checkbox"/> G8ZF25	G8ZF25_RABIT	Alpha-globin 1	HBA1	Oryctolagus cuniculus (Rabbit)	142 AA

Figura 4. Selección de las ultimas secuencias de HBA1.

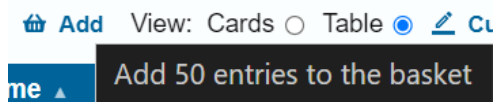


Figura 5. Adición de las 50 secuencias a la canasta de UniProt.

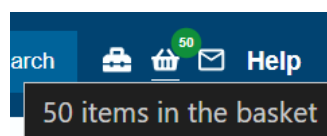


Figura 6. Cantidad de secuencias dentro de la canasta una vez fueron agregadas.

My Basket			
UniProtKB (50) UniRef UniParc			
BLAST Align Map IDs Download Remove View: Cards Table Customize columns			
Entry	Entry Name	Organism	
<input type="checkbox"/> P02019	HBA1_ONCMY	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	
<input type="checkbox"/> P47746	CNR1_MOUSE	Mus musculus (Mouse)	
<input type="checkbox"/> B8XJS4	B8XJS4_GADMO	Gadus morhua (Atlantic cod)	
<input type="checkbox"/> P0CH25	HBA1_CAPHI	Capra hircus (Goat)	
<input type="checkbox"/> B3EX31	B3EX31_SORAR	Sorex araneus (Eurasian common shrew) (European shrew)	
<input type="checkbox"/> P18974	HBA1_IGUIG	Iguana iguana (Common iguana)	
<input type="checkbox"/> P29624	HBA1_NOTAN	Notothenia angustata (Rockcod)	
<input type="checkbox"/> V9GZG8	V9GZG8_HUMAN	Homo sapiens (Human)	
<input type="checkbox"/> P21554	CNR1_HUMAN	Homo sapiens (Human)	
			Clear all (50) Full view

Figura 7. Visualización de la canasta de UniProt.

## 2. Alineamiento de Secuencias.

En esta sección se seleccionaron las 50 secuencias dentro de la canasta y se aplicó la función de alineamiento proporcionada por UniProt (Figura 8-10).

My Basket			
UniProtKB (50) UniRef UniParc			
BLAST Align Map IDs Download Remove View: Cards Table Customize columns 50 rows sele			
Entry	Entry Name	Organism	
<input checked="" type="checkbox"/> P02019	HBA1_ONCMY	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	
<input checked="" type="checkbox"/> P47746	CNR1_MOUSE	Mus musculus (Mouse)	
<input checked="" type="checkbox"/> B8XJS4	B8XJS4_GADMO	Gadus morhua (Atlantic cod)	
<input checked="" type="checkbox"/> P0CH25	HBA1_CAPHI	Capra hircus (Goat)	
<input checked="" type="checkbox"/> B3EX31	B3EX31_SORAR	Sorex araneus (Eurasian common shrew) (European shrew)	
<input checked="" type="checkbox"/> P18974	HBA1_IGUIG	Iguana iguana (Common iguana)	
<input checked="" type="checkbox"/> P29624	HBA1_NOTAN	Notothenia angustata (Rockcod)	
<input checked="" type="checkbox"/> V9GZG8	V9GZG8_HUMAN	Homo sapiens (Human)	
			Remove (50) Full view

Figura 8. Selección de las 50 secuencias dentro de la canasta de UniProt.

# Align

Find a protein sequence by UniProt ID (e.g. P05067 or A4\_HUMAN or UPI0000000001) to align with the [Clustal Omega program](#) . You can also paste a list of IDs.

UniProt IDs

OR

Enter multiple protein or nucleotide sequences, separated by a FASTA header. You may also [load from a text file](#).

```
KKVADALTNAAHVHMDPNALSALSDLHAHLKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
AVHASLDKFLASVSTVLTSKYR
>tr|Q86YQ4|Q86YQ4_HUMAN Alpha-1 globin (Fragment) OS=Homo sapiens OX=9606 GN=HBA1 PE=3 SV=1
QVKGHGKGVADALTNAAHVHMDPNALSALSDLHAHLKLRVDPVNFKLLSHCLLVTLAAHL
PAEFTPAVHASLDKFLASVSTVLTSKYR
>tr|A0A0U2VE96|A0A0U2VE96_BIZLO Hemoglobin alpha D subunit OS=Biziura lobata OX=45648 GN=HBA1 PE=3 SV=1
MLTADKKLIVQLWEKVAGHQEEFGNEALQRMFVTPQTKTYPHFDLSPGSDQIRGHGK
KVAAALGSAVKSLDNLSQLSELNLHAYNLRVDPANFKLLAQCFQVLAHLGDKDYSPE
MHAADFDFLSAVAVALAEKYR
```

Figura 9. Ventana de la función del alineamiento de secuencias.

# Tool results

Your tool analysis results from the last 7 days are listed below. If you have tools jobs running, you can navigate away to other pages and you will be notified once the job is completed.

Job type	Name	Created	Status
ALIGN	sp P02019 HBA1_ONCMY +49	2024-05-11 02:46	Queued <div></div>
clustalo-R20240511-094557-0614-29144803-p1m			

Figura 10. Vista de la alineación en proceso.

Una vez se realizó el alineamiento de las secuencias de las 50 especies se muestra una matriz con las secuencias resaltando los aminoácidos que se logran alinear de forma vertical, añadiendo algunos guiones para favorecer el alineamiento (Figura 11 y 12).

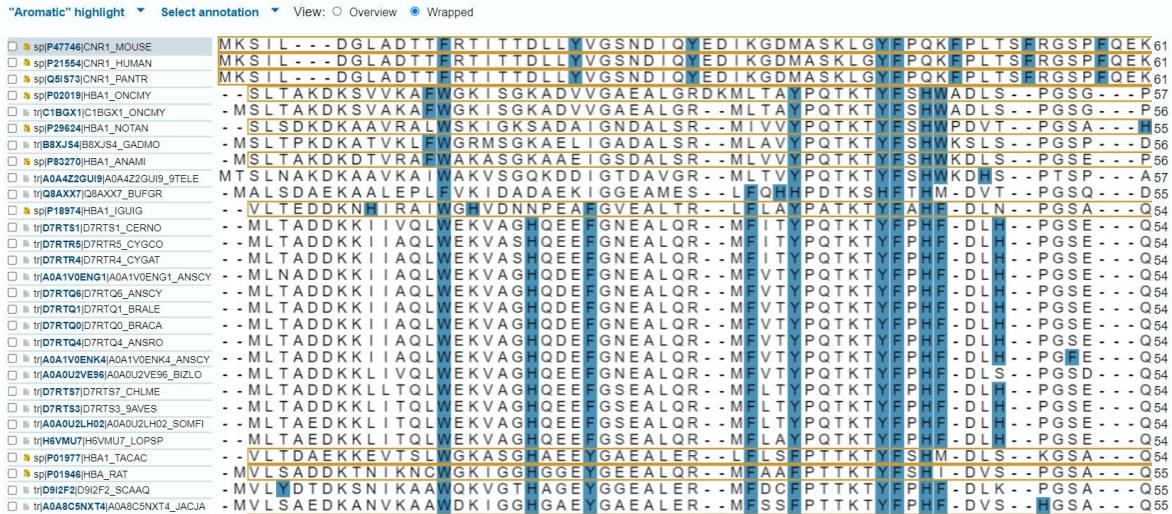


Figura 11. Matriz de alineamiento de secuencias aromáticas.

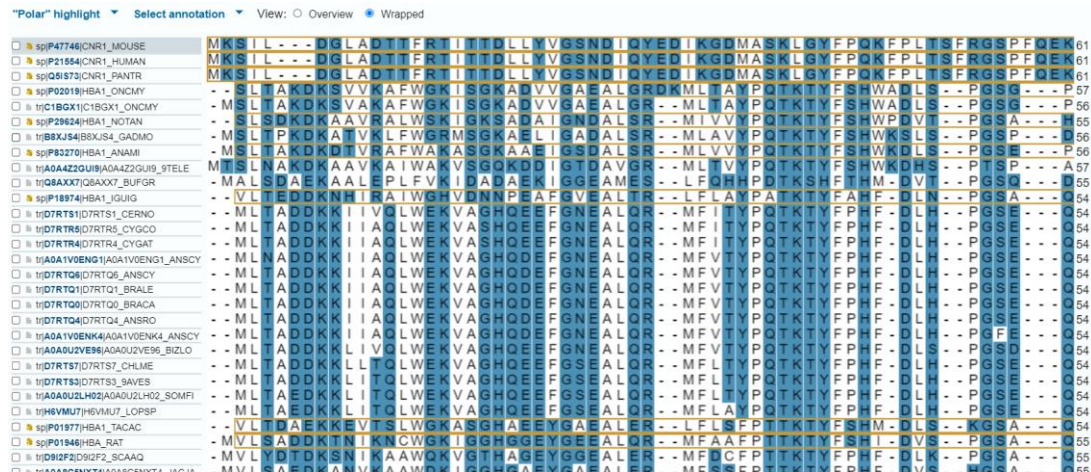


Figura 12. Matriz de alineamiento de secuencias polares.

De igual forma, se obtuvo el árbol de secuencias después de aplicar el alineamiento en el cual se muestra la similitud entre las distintas especies y el ancestro que comparten como se muestra en la figura 13.

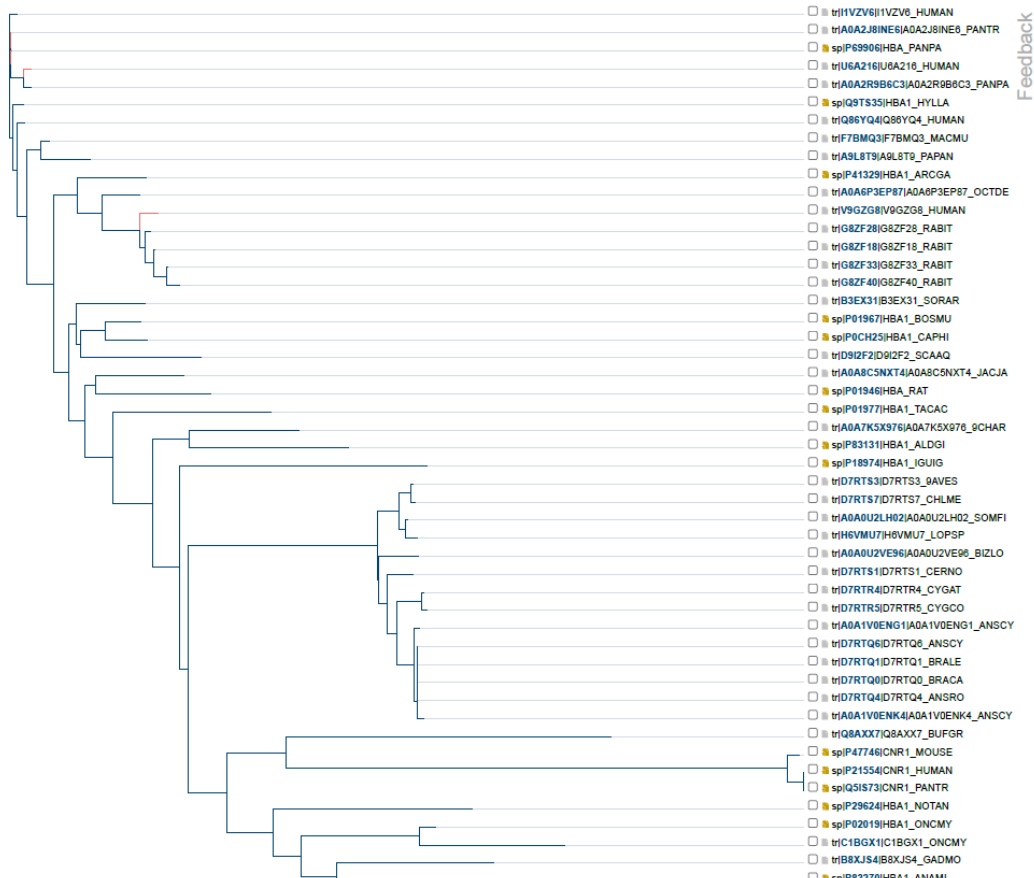


Figura 13. Árbol de secuencias de las 50 especies.



Por ultimo se visualizo la matriz de similitud entre las secuencias obtenidas (figura 14).

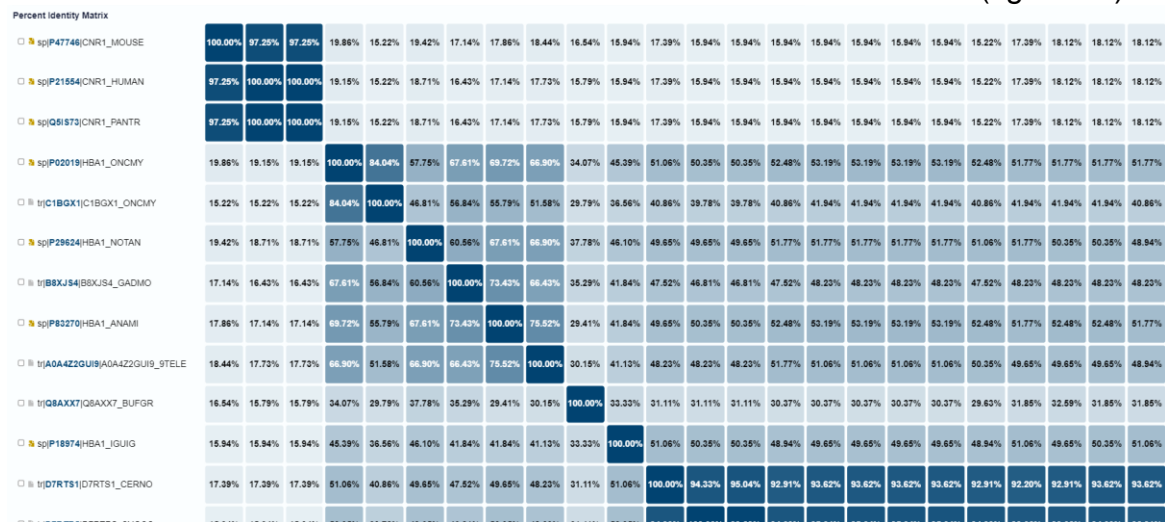


Figura 14. Matriz de similitud entre las especies.

3. “BLAST” usando la entrada P01979.

Dentro de UniProt se ejecutó la función “BLAST” utilizando como entrada el ID P01979 el cual corresponde a la especie “Ornitorrinco” como se aprecia en la figura 15 y 16.

BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4\_HUMAN or UPI0000000001).

UniProt IDs

OR

Enter one or more sequences (5 max). You may also [load from a text file](#).

```
>sp|P01979|HBA_ORNAN Hemoglobin subunit alpha OS=Ornithorhynchus anatinus OX=9258 GN=HBA PE=1 SV=1
MLTDAEKKEV TALWGKAAGH GEEYGAEAL RLFQAFPTTK TYFSHFDLSH GSAQIKAHGK
KVADALSTAA GHFDDMDSDAL SALSDLHAHK LRVDPVNFKL LAHCILVLA RHCPGEFTPS
AHAAMDKFLS KVATVLTSKY R
```

Figura 15. Ventana de la función BLAST con la entrada P01979.

Job type	Name	Created	Status	
BLAST	sp P01979 HBA_ORNAN	2024-05-11 10:41	Running	
ncbiblast-R20240511-174122-0623-39878320-p1m		Target database: UniProtKB reference proteomes + Swiss-Prot		

Figura 16. Ejecución del “BLAST” con la secuencia P01979.

Posteriormente se visualiza la similitud entre la secuencia con las mas parecidas a la misma (Figura 17).

## BLAST 250 results found in UniProtKB

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	
<input type="checkbox"/> P01979	HBA_ORNAN	Hemoglobin subunit alpha [...]	HBA	Ornithorhynchus anatinus (Duckbill platypus)	141 AA	100%  (231) (6.4e-80)
<input type="checkbox"/> A0A6I8NYG5	A0A6I8NYG5_ORNAN	Hemoglobin subunit alpha	LOC100087130	Ornithorhynchus anatinus (Duckbill platypus)	142 AA	100%  (231) (6.7e-80)
<input type="checkbox"/> A0A6I8NZ57	A0A6I8NZ57_ORNAN	Hemoglobin subunit alpha	LOC100087130	Ornithorhynchus anatinus (Duckbill platypus)	159 AA	100%  (231) (1.2e-97)
<input type="checkbox"/> A0A6I8NFF3	A0A6I8NFF3_ORNAN	Hemoglobin subunit alpha	LOC100087130	Ornithorhynchus anatinus (Duckbill platypus)	231 AA	100%  (231) (1.4e-80)
<input type="checkbox"/> P01977	HBA1_TACAC	Hemoglobin subunit alpha.1 [...]		Tachyglossus aculeatus aculeatus (Southeast Australian short-beaked echidna)	141 AA	88.7%  (253) (6.1e-80)
<input type="checkbox"/> P01978	HBA2_TACAC	Hemoglobin subunit alpha.2 [...]		Tachyglossus aculeatus aculeatus (Southeast Australian short-beaked echidna)	141 AA	84.4%  (253) (6.7e-80)
<input type="checkbox"/> P01930	HBA_PILBA	Hemoglobin subunit alpha [...]	HBA	Ptilocolobus badius (Western red colobus) (Procolobus badius)	142 AA	74.5%  (224) (2e-72)
<input type="checkbox"/> P63107	HBA_MACFU	Hemoglobin subunit alpha [...]	HBA	Macaca fuscata fuscata (Japanese macaque)	142 AA	73.8%  (224) (2e-72)
<input type="checkbox"/> P63108	HBA_MACMU	Hemoglobin	HBA	Macaca mulatta (Rhesus)	142 AA	73.8%  (224) (2e-72)

Figura 17. Similitud de la entrada P01979 con otras secuencias.

Como se pudo apreciar en la figura 17 la secuencia de un Ornitorrinco es mas similar a la de un Echidna tipo Tachyglossidae en un 88.7%.

- Obtención de una secuencia de proteínas no caracterizada de un organismo desconocido e investigación de esta.

Proteína: Cisteína/Cisteína ácido sulfínico descarboxilasa.

Secuencia:

```
>sp|A0A2I9|CCSD_UNKP Cysteine/Cysteine sulfinic acid decarboxylase OS=Unknown
prokaryotic organism OX=2725 GN=undec1A PE=1 SV=1
MITPLTLATLSKNPILVDFDPEDGRWNSHVDLGLWSDLYLIAPATANTIGKMAAGIADN
LLTSYLSARCPVFIAPAMDVDMLMHPATQRNLGILKSSGNHIIIEPGSGELASGLTGKGR
MAEPEEIVREVISFFSKKKITEKPLNGRRVFINAGPTIEPIDPVRFISNYSSGRMGIALA
DAAAAMGAEVTLVLGPVTLRPSSQDINVIDVRSAAEMKEASVEAFRECDIILAAVADF
TPLTTSDKKIKRGSGEMVINLRPTEDIAAELGKMKKKNQLLVGFALETDDDEITNASSKLK
RKNLDMIVLNSLKDPGAGFGHETNRITIIDKSNNIDKFELKTKGEVAADIIRKILTLVH
```

Función:

Cataliza la descarboxilación de L-cisteína a cisteamina y de 3-sulfino-L-alanina (ácido cisteína sulfínico) a hipotaurina (Figura 18).

## A0A2I9 · CCSD\_UNKP

**Protein**<sup>i</sup> | Cysteine/Cysteine sulfinic acid decarboxylase

**Gene**<sup>i</sup> | undec1A

**Status**<sup>i</sup> | UniProtKB reviewed (Swiss-Prot)

**Organism**<sup>i</sup> | Unknown prokaryotic organism

**Amino acids** | 359 ([go to sequence](#))

**Protein existence**<sup>i</sup> | Evidence at protein level

**Annotation score**<sup>i</sup> | (4/5)

Figura 18. Información de la proteína A0A2I9.

Se aplicó un BLAST con la entrada A0A2I9 y se observó una semejanza del 65.7% con una “*Sanguibacteroides justesenii*” la cual es una bacteria recientemente descubierta que se ha aislado de hemocultivos en una prueba de laboratorio para detectar bacterias en el torrente sanguíneo. No se sabe mucho sobre esta bacteria, pero se considera un agente del grupo de riesgo 2. Esto significa que se cree que tiene un riesgo moderado de causar enfermedades en humanos y animales. Se necesitan más investigaciones para comprender mejor *Sanguibacteroides justesenii*, incluido su potencial en la causa de enfermedades y cómo se propaga (Figura 19). [2].

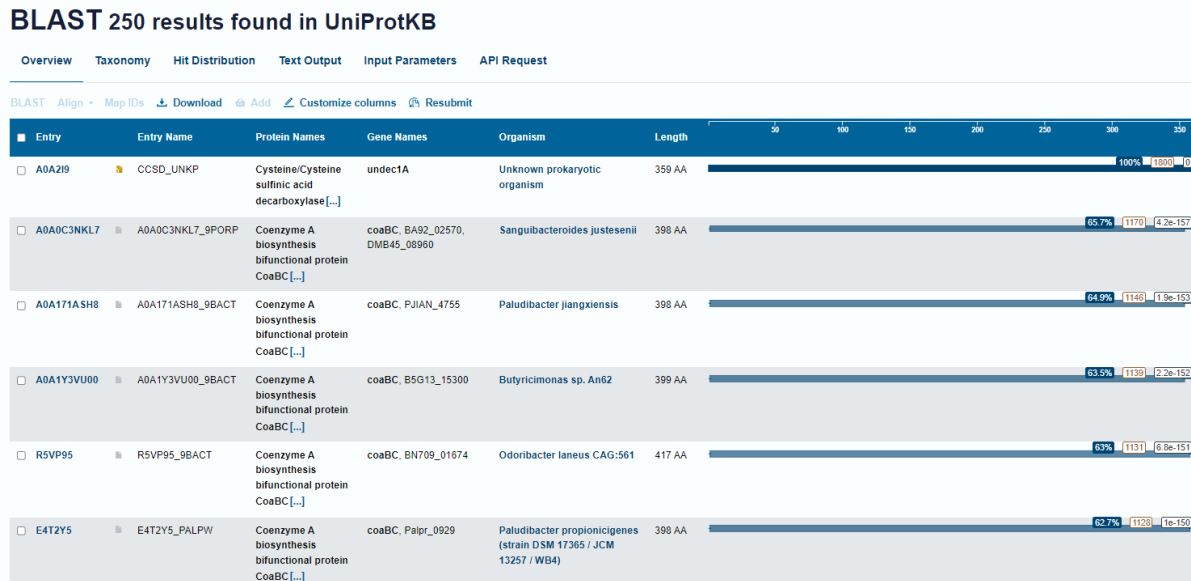


Figura 19. Vista de la similitud de la proteína A0A2I9.

## Conclusiones y recomendaciones.

UniProt al igual que PDB es una plataforma para el estudio y la visualización de proteínas y secuencias de aminoácidos que implementa algunas herramientas para el alineamiento de secuencias y la similitud entre las mismas.

Conocer estas herramientas es crucial para el estudio y la comprensión de las proteínas tanto las que ya se han estudiado profundamente como las que aún no se estudian en su totalidad y gracias a UniProt, los científicos de todo el mundo tienen acceso a un entorno para acceder a información completa y confiable sobre las proteínas lo cual es fundamental para la investigación en biología molecular, genética, medicina y muchos otros campos.

## Referencias.

[1] **UniProt: the universal protein knowledgebase in 2021:**

<https://academic.oup.com/nar/article/45/D1/D158/2605721>

[2] Sydenham, TV, Hasman, H, & Justesen, US. (2015). Draft Genome Sequences of *Sanguibacteroides justesenii*, gen. nov., sp. nov., Strains OUH 308042T (= ATCC BAA-2681T) and OUH 334697 (= ATCC BAA-2682), Isolated from Blood Cultures from Two Different Patients. *Genome Announc*, 3(1), e00005-15.

<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1547597>