



INSTITUTO POLITÉCNICO NACIONAL
ESCUELA SUPERIOR DE CÓMPUTO
BIOINFORMATICS



PRÁCTICA 6.
UNIPROT

NOMBRE DEL ALUMNO: GARCÍA QUIROZ GUSTAVO IVAN

GRUPO: 7CV3

NOMBRE DEL PROFESOR: ROSAS TRIGUEROS JORGE LUIS

FECHA DE REALIZACIÓN DE LA PRÁCTICA: 12/02/2025

FECHA DE ENTREGA DEL REPORTE: 19/03/2025

Índice

1	Marco teórico.....	1
1.1	Proteínas.....	1
1.2	UniProt.....	1
1.2.1	Estructura y Componentes de UniProt.....	1
1.2.2	Funcionalidades y Uso de UniProt.....	1
1.3	Técnicas de Análisis Bioinformático.....	2
1.3.1	Alineamiento de Secuencias.....	2
1.4	Blast (Basic Local Alignment Search Tool)	2
1.5	Proteínas No Caracterizadas	2
2	Material y equipo.	3
2.1.1	Hardware	3
2.1.2	Software.....	3
3	Desarrollo de la práctica.	4
3.1	Recolección de Secuencias HBA1	4
3.2	Tabla de Secuencias HBA1	4
3.3	Alineamiento de Secuencias Proteicas	6
3.4	BLAST.....	8
3.5	Investigación de Proteína No Caracterizada.....	11
3.5.1	Obtención y Verificación de la Secuencia.....	11
3.5.2	Análisis de Características Primarias.....	12
3.5.3	Predicción de Estructura Secundaria.....	12
3.5.4	Análisis BLAST y Búsqueda de Homólogos	13
3.5.5	Interpretación Funcional	14
3.6	Generación y Análisis de Secuencia Aleatoria.....	15
3.7	Desafíos y soluciones	17
4	Conclusiones y recomendaciones.	18
5	Referencias.....	19

1 Marco teórico.

1.1 Proteínas

Las proteínas son macromoléculas biológicas esenciales compuestas por cadenas de aminoácidos que desempeñan funciones críticas en todos los sistemas vivos. Cada proteína representa una secuencia única de aminoácidos que determina su estructura tridimensional y, consecuentemente, su función específica en los organismos. La comprensión de estas secuencias permite a los investigadores desentrañar mecanismos moleculares fundamentales.

1.2 UniProt

UniProt es la base de datos más completa y anotada de secuencias de proteínas en el mundo. Fue creada por el Consorcio UniProt, que incluye al Instituto Europeo de Bioinformática (EBI), el Instituto Suizo de Bioinformática (SIB) y los Recursos de Información Proteica (PIR). Su objetivo principal es apoyar la investigación biológica proporcionando un conocimiento preciso y bien clasificado sobre las secuencias de proteínas, junto con referencias cruzadas y consultas accesibles libremente a la comunidad científica.

1.2.1 Estructura y Componentes de UniProt

UniProt se basa en tres pilares principales:

- Continuación del trabajo de las bases de datos de proteínas: Incluye Swiss-Prot, TrEMBL y PIR-PDS, ofreciendo una base de datos madura y revisada.
- Archivo UniParc: Almacena secuencias nuevas y actualizadas diariamente.
- Modo UniProt NREF: Ofrece una visualización no redundante de las entradas.

1.2.2 Funcionalidades y Uso de UniProt

Búsqueda y Filtrado: UniProt permite una búsqueda sencilla y filtrado de información mediante su interfaz web. Por ejemplo, se pueden buscar proteínas revisadas de un organismo específico, como *Homo sapiens*, reduciendo las entradas a un número manejable.

Organización de Datos: Los datos se organizan en columnas que agrupan información sobre nombres y taxonomía, secuencias, etc., facilitando la localización de información específica.

Integración con Ontologías: Utiliza la ontología de genes (GO) para proporcionar un vocabulario estructurado para describir productos de genes, mejorando la integración y comunicación entre investigadores.

1.3 Técnicas de Análisis Bioinformático

1.3.1 Alineamiento de Secuencias

El alineamiento de secuencias es una técnica fundamental en bioinformática que permite comparar secuencias de proteínas o ADN para identificar similitudes y diferencias. Esto es útil para inferir relaciones evolutivas y funciones potenciales de las proteínas. En el contexto de esta práctica, el alineamiento de secuencias de la subunidad alfa de la hemoglobina (HBA1) de diferentes especies permite explorar la conservación de secuencias y posibles variaciones que podrían influir en la función de la proteína.

1.4 Blast (Basic Local Alignment Search Tool)

Blast es una herramienta bioinformática ampliamente utilizada para buscar secuencias similares en bases de datos. Permite identificar secuencias homólogas, lo que es útil para predecir funciones de proteínas desconocidas. En esta práctica, se utiliza Blast para buscar secuencias similares a la entrada P01979, lo que puede ayudar a identificar proteínas con funciones relacionadas.

BLAST representa una herramienta fundamental para comparar secuencias biológicas. Su función principal consiste en:

- Buscar secuencias similares en bases de datos globales
- Calcular la significancia estadística de las similitudes encontradas
- Predecir posibles funciones de proteínas desconocidas
- Establecer relaciones evolutivas entre diferentes organismos

1.5 Proteínas No Caracterizadas

Investigar una proteína desconocida, como A0A0F7YUR7, implica analizar su secuencia para predecir su estructura y función potencial. Esto puede incluir el uso de herramientas bioinformáticas para identificar dominios proteicos, sitios de unión y otros rasgos estructurales que sugieran su papel biológico.

Las proteínas no caracterizadas son secuencias moleculares cuya función específica aún no ha sido completamente determinada. Su estudio implica:

- Análisis computacionales avanzados
- Comparación con bases de datos existentes
- Predicción de estructuras y funciones potenciales
- Generación de hipótesis para investigación experimental posterior

2 Material y equipo.

Para esta práctica se usaron las siguientes herramientas de software y hardware necesarias para realizar la práctica.

2.1.1 Hardware

- Computadora.

2.1.2 Software

- uniprot.org

3 Desarrollo de la práctica.

La práctica de permitió comprender la complejidad del análisis de secuencias proteicas.

3.1 Recolección de Secuencias HBA1

El primer paso se centró en la obtención de secuencias de hemoglobina subunidad alfa (HBA1) desde una perspectiva filogenética comprehensiva. La selección de secuencias requirió un análisis estratégico que trascendiera la mera recopilación de datos, buscando representatividad taxonómica y calidad molecular.

La metodología de selección consideró múltiples criterios:

- Diversidad de especies (mamíferos, aves, reptiles, peces)
- Calidad de anotación en UniProtKB
- Integridad de la secuencia molecular
- Disponibilidad de información complementaria

La navegación por UniProt permitió observar las distintas variantes de hemoglobina y son útiles cuando aplicamos filtros precisos y examinamos cada entrada. Verificar la coherencia de la información molecular garantiza la robustez del conjunto de datos seleccionados.

3.2 Tabla de Secuencias HBA1

ID UniProt	Organismo	Código	Longitud (aa)	Características Principales
Q09146	Schizosaccharomyces pombe	HBA1_SCHPO	484	Proteína de levadura, función única
P47746	Mus musculus	CNR1_MOUSE	446	Receptor de cannabinoides, no HBA1 típica
O02777	Felis catus	CNR1_FELCA	446	Receptor de cannabinoides, no HBA1 típica
K7N5M5	Eleginops maclovinus	HBA1_ELEMC	141	Pez, hemoglobina

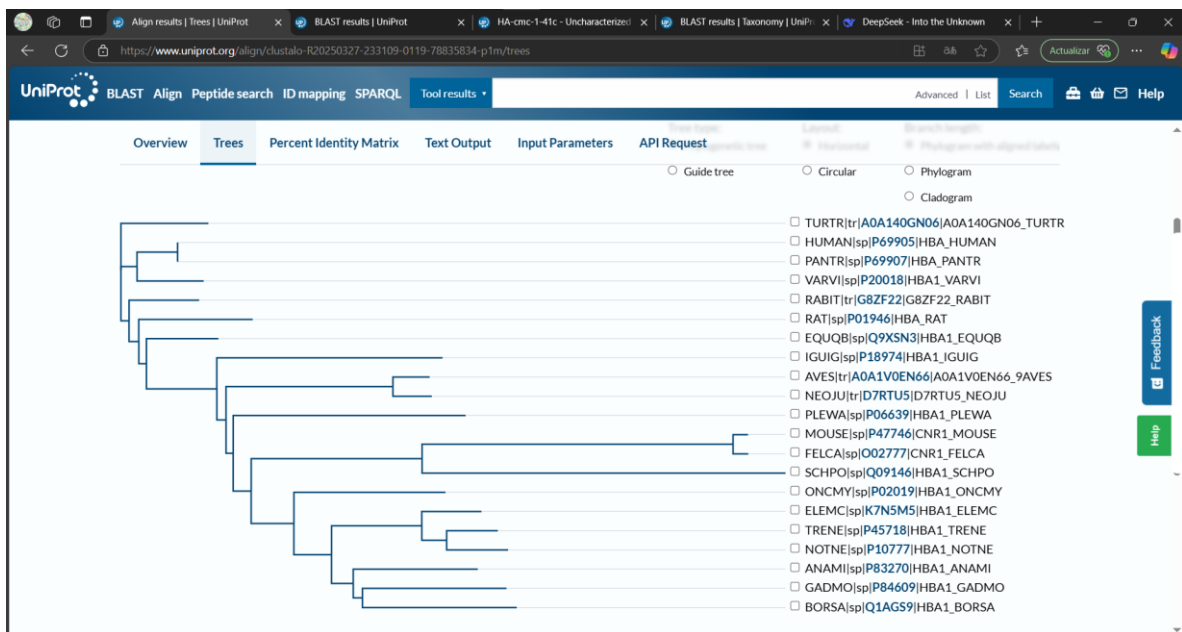
				subunidad alfa
P45718	Trematomus newnesi	HBA1_TRENE	141	Pez antártico, hemoglobina
P10777	Notothenia neglecta	HBA1_NOTNE	141	Pez antártico, hemoglobina
Q1AGS9	Boreogadus saida	HBA1_BORSA	141	Pez polar, hemoglobina
P02019	Oncorhynchus mykiss	HBA1_ONCMY	141	Trucha arcoíris, hemoglobina
P83270	Anarhichas minor	HBA1_ANAMI	141	Pez lobo, hemoglobina
P84609	Gadus morhua	HBA1_GADMO	141	Bacalao, hemoglobina
P06639	Pleurodeles waltl	HBA1_PLEWA	141	Salamandra, hemoglobina
P18974	Iguana iguana	HBA1_IGUIG	141	Iguana, hemoglobina
A0A1V0EN 66	Anser anser	A0A1V0EN66_9AV ES	141	Ánsar, hemoglobina
D7RTU5	Neochen jubata	D7RTU5_NEOJU	141	Pato, hemoglobina
P01946	Rattus norvegicus	HBA_RAT	142	Rata, hemoglobina
A0A140GN 06	Tursiops truncatus	A0A140GN06_TU RTR	141	Delfín, hemoglobina
G8ZF22	Oryctolagus cuniculus	G8ZF22_RABIT	141	Conejo, hemoglobina
Q9XSN3	Equus quagga burchellii	HBA1_EQUQB	141	Cebra, hemoglobina
P69905	Homo sapiens	HBA_HUMAN	142	Humano, hemoglobina

P69907	Pan troglodytes	HBA_PANTR	142	Chimpancé, hemoglobina
P20018	Varecia variegata	HBA1_VARVI	141	Lémur, hemoglobina

Tabla 1 Tabla de Secuencias HBA1

3.3 Alineamiento de Secuencias Proteicas

El alineamiento de secuencias representa uno de los procedimientos más útiles en bioinformática comparativa. Este proceso permitió revelar las relaciones estructurales y evolutivas entre las secuencias de HBA1 recopiladas, desentrañando los patrones moleculares subyacentes.



Align results

Overview **Trees** Percent Identity Matrix Text Output Input Parameters API Request

Tools Download Add Result

Percent Identity Matrix

<input type="checkbox"/> SCHPO sp Q09146 HBA1_SCHPO	100.00%	22.59%	22.18%	16.09%	14.94%	13.79%	15.91%	13.48%	12.50%	12.50%	15.91%	12.64%	14.94%	13.79%	18.18%	18.18%	15.91%	17.05%	15.91%	15.91%	17.24%
<input type="checkbox"/> MOUSE sp P47746 CNR1_MOUSE	22.59%	100.00%	96.61%	22.46%	21.74%	20.29%	18.71%	21.43%	19.42%	18.71%	14.49%	16.06%	18.25%	18.25%	20.29%	21.01%	21.01%	21.01%	19.57%	19.57%	18.25%
<input type="checkbox"/> FELCA sp O02777 CNR1_FELCA	22.18%	96.61%	100.00%	22.46%	21.74%	20.29%	18.71%	21.43%	19.42%	18.71%	14.49%	16.06%	18.25%	18.25%	20.29%	21.01%	21.01%	21.01%	19.57%	19.57%	18.25%
<input type="checkbox"/> ELEMCI sp K7N5M5 HBA1_ELEMCI	16.09%	22.46%	22.46%	100.00%	86.62%	81.69%	63.38%	57.75%	68.31%	64.79%	43.26%	43.26%	50.35%	48.81%	51.06%	51.77%	51.06%	52.48%	51.06%	51.06%	50.35%
<input type="checkbox"/> TRENE sp P45718 HBA1_TRENE	14.94%	21.74%	21.74%	86.62%	100.00%	87.32%	58.45%	54.23%	64.08%	59.86%	44.68%	44.68%	49.65%	48.23%	49.65%	51.77%	48.94%	49.65%	48.23%	48.23%	48.23%
<input type="checkbox"/> NOTNE sp P10777 HBA1_NOTNE	13.79%	20.29%	20.29%	81.69%	87.32%	100.00%	56.34%	52.82%	62.68%	59.15%	43.26%	42.55%	47.52%	45.99%	50.35%	50.35%	48.23%	50.35%	48.23%	48.23%	48.23%
<input type="checkbox"/> BORSA sp Q1AGS9 HBA1_BORSA	15.91%	18.71%	18.71%	63.38%	58.45%	56.34%	100.00%	58.45%	65.72%	66.43%	40.14%	39.72%	43.26%	43.97%	46.40%	47.89%	46.40%	48.59%	48.59%	48.59%	46.10%
<input type="checkbox"/> ONCMY sp P02019 HBA1_ONCMY	13.48%	21.43%	21.43%	57.75%	54.23%	52.82%	58.45%	100.00%	69.01%	64.79%	48.23%	44.68%	52.48%	51.06%	57.45%	56.03%	56.03%	59.57%	56.74%	56.74%	55.32%
<input type="checkbox"/> ANAMI sp P83270 HBA1_ANAMI	12.50%	19.42%	19.42%	68.31%	64.08%	62.68%	65.72%	69.01%	100.00%	70.63%	45.77%	41.84%	53.19%	51.77%	52.11%	50.70%	48.59%	52.82%	50.70%	50.70%	48.94%
<input type="checkbox"/> GADMO sp P84609 HBA1_GADMO	12.50%	18.71%	18.71%	64.79%	59.86%	59.15%	66.43%	64.79%	70.63%	100.00%	41.55%	39.72%	46.10%	46.10%	47.89%	46.40%	45.77%	49.30%	47.89%	47.89%	46.81%
<input type="checkbox"/> PLEWA sp P06639 HBA1_PLEWA	15.91%	14.49%	14.49%	43.26%	44.68%	43.26%	40.14%	48.23%	45.77%	41.55%	100.00%	48.23%	50.35%	48.94%	48.59%	52.11%	51.41%	56.34%	56.34%	56.34%	52.48%
<input type="checkbox"/> IGUGI sp P18974 HBA1_IGUGI	12.64%	16.06%	16.06%	43.26%	44.68%	42.55%	39.72%	44.68%	41.84%	39.72%	48.23%	100.00%	48.94%	51.06%	53.19%	54.61%	55.32%	58.87%	56.74%	56.74%	55.32%

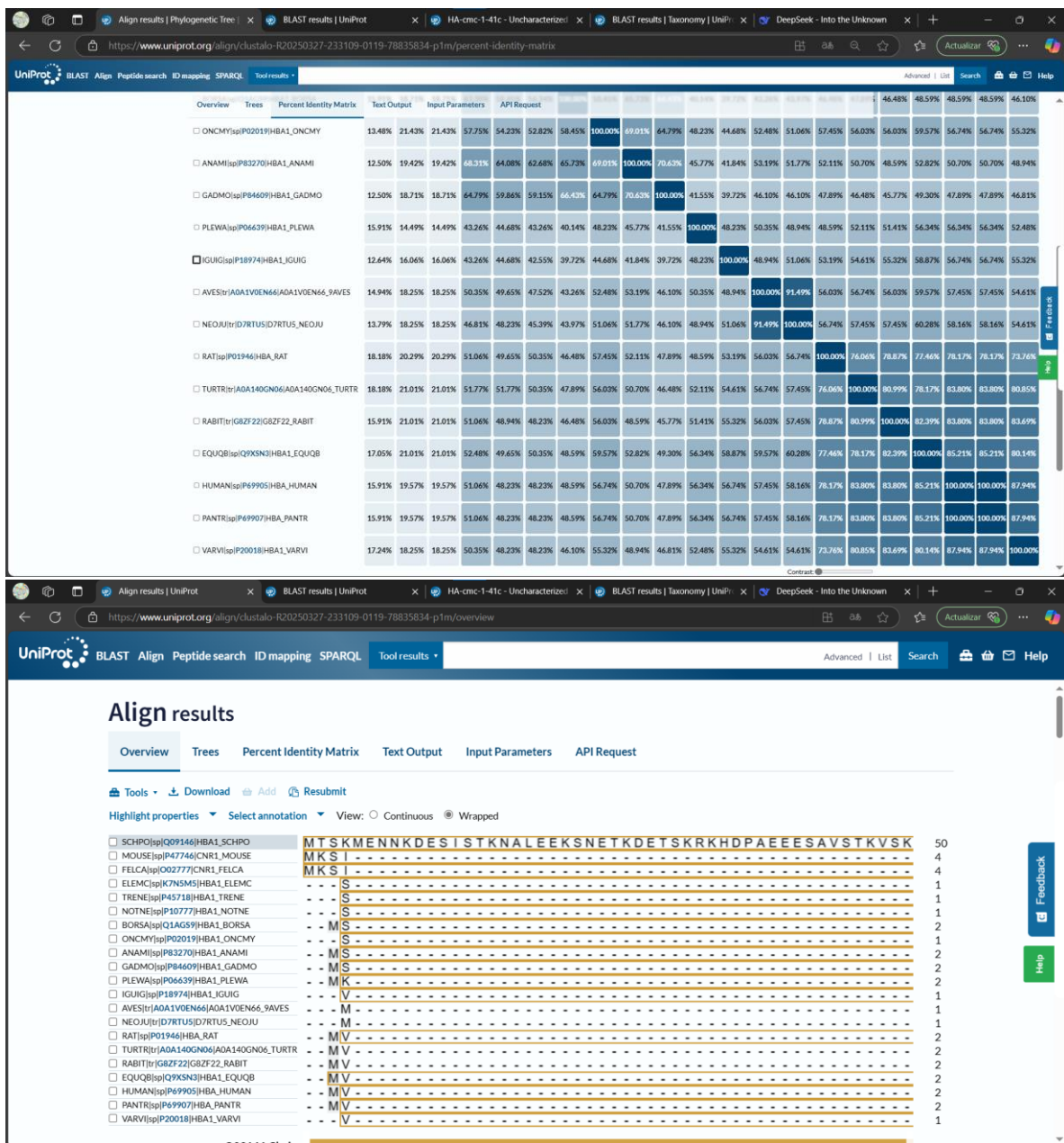


Figura 1 Alineamiento de Secuencias Proteicas

3.4 BLAST

La implementación de BLAST representó una inmersión profunda en las estrategias computacionales de búsqueda de similitudes proteicas. La entrada P01979 actuó como punto de partida para una exploración sistemática de las bases de datos moleculares globales.

El protocolo BLAST implementado contempló una metodología rigurosa que involucró la configuración precisa de parámetros de búsqueda, selección de bases

de datos relevantes y análisis estadístico de los resultados obtenidos. Cada match representaba una potencial conexión filogenética o funcional que requería una interpretación cuidadosa.

Consideraciones metodológicas fundamentales:

- Selección de base de datos específica
- Configuración de umbral de significancia estadística
- Análisis de E-values
- Evaluación de cobertura de secuencia

BLAST results | UniProt

HA-cmc-1-41c - Uncharacterized

BLAST results | UniProt

DeepSeek - Into the Unknown

https://www.uniprot.org/blast/uniprotkb/ncbiblast-R20250327-233709-0656-48159315-p1m/overview

UniProt BLAST Align Peptide search ID mapping SPARQL Tool results Advanced | List Search Help

Blast parameters

Identity

Score

44.1

100

E-Value

103

731

6.8e-98

0.013

Status

Reviewed (Swiss-Prot) (143)

Unreviewed (TrEMBL) (107)

Popular organisms

Bovine (2)

Mouse (2)

Human (1)

Rat (1)

Taxonomy

BLAST 250 results found in UniProtKB

Overview Taxonomy Hit Distribution Text Output Input Parameters API Request

Tools Download Add Customize columns Resubmit

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
<input type="checkbox"/> A0A6I8NYG5	A0A6I8NYG5_ORNAN	Hemoglobin subunit alpha	LOC100087130	Ornithorhynchus anatinus (Duckbill platypus)	142 AA
<input type="checkbox"/> A0A6I8NZ57	A0A6I8NZ57_ORNAN	Hemoglobin subunit alpha	LOC100087130	Ornithorhynchus anatinus (Duckbill platypus)	159 AA
<input type="checkbox"/> A0A6I8NFF3	A0A6I8NFF3_ORNAN	Hemoglobin subunit alpha	LOC100087130	Ornithorhynchus anatinus (Duckbill platypus)	231 AA
<input type="checkbox"/> P01977	HBA1_TACAC	Hemoglobin subunit alpha-1		Tachyglossus aculeatus (Southeast Australian short-beaked echidna)	141 AA
<input type="checkbox"/> P01978	HBA2_TACAC	Hemoglobin subunit alpha-2		Tachyglossus aculeatus (Southeast Australian short-beaked echidna)	141 AA
<input type="checkbox"/> P01930	HBA_PILBA	Hemoglobin subunit alpha	HBA	Ptilocolobus badius (Western red colobus) (Procolobus badius)	142 AA
<input type="checkbox"/> P63107	HBA_MACFU	Hemoglobin subunit alpha	HBA	Macaca fuscata fuscata (Japanese macaque)	142 AA

Feedback

Help

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	
<input type="checkbox"/> P63108	HBA_MACMU	Hemoglobin subunit alpha	HBA	Macaca mulatta (Rhesus macaque)	142 AA	73.8% 564 2.1e-72
<input type="checkbox"/> A0A2K6R518	A0A2K6R518_RHIRO	Hemoglobin subunit alpha 1		Rhinopithecus roxellana (Golden snub-nosed monkey) (Pygathrix roxellana)	142 AA	73.8% 564 2.1e-72
<input type="checkbox"/> A0A2K5IMJ5	A0A2K5IMJ5_COLAP	Globin domain-containing protein		Colobus angolensis palliatus (Peters' Angolan colobus)	142 AA	73.8% 563 2.9e-72
<input type="checkbox"/> A0A8C9LL77	A0A8C9LL77_9PRIM	Globin domain-containing protein		Ptilocolobus tephrosceles (Ugandan red Colobus)	193 AA	74.5% 567 5.9e-72
<input type="checkbox"/> P01924	HBA_SEMEN	Hemoglobin subunit alpha	HBA	Semnopithecus entellus (Northern plains gray langur) (Presbytis entellus)	141 AA	73.8% 563 4e-72
<input type="checkbox"/> P21767	HBA_MACFA	Hemoglobin subunit alpha-A/Q/R/T		Macaca fascicularis (Crab-eating macaque) (Cynomolgus monkey)	141 AA	73.8% 563 4e-72
<input type="checkbox"/> P01926	HBA_CHLAE	Hemoglobin subunit alpha	HBA	Chlorocebus aethiops (Green monkey) (Cercopithecus)	142 AA	73.8% 562 4.1e-72

Figura 2 BLAST

3.5 Investigación de Proteína No Caracterizada

La investigación de la proteína no caracterizada identificada como A0A0F7YUR7 constituyó un ejercicio fundamental de aplicación de técnicas bioinformáticas para la elucidación de características moleculares de secuencias proteicas de función desconocida. Esta proteína, proveniente de un organismo procariota no identificado, representa un caso paradigmático de los desafíos contemporáneos en la caracterización funcional de proteomas.

3.5.1 Obtención y Verificación de la Secuencia

La secuencia proteica fue obtenida directamente de la base de datos UniProt con el identificador A0A0F7YUR7. La secuencia completa de 68 aminoácidos fue verificada y se constató su categorización como "Uncharacterized protein" (proteína no caracterizada) con nivel de evidencia 4 (PE=4), lo que indica predicción computacional sin evidencia experimental de su existencia.

Names & Taxonomyⁱ

Protein namesⁱ

Submitted names | Uncharacterized protein

Gene namesⁱ

Name | HA-cmc-1-41c

Organism names

Taxonomic identifierⁱ | 2725 (NCBI [↗](#))

Organismⁱ | Unknown prokaryotic organism

Taxonomic lineageⁱ | cellular organisms > Bacteria (eubacteria) > environmental samples

Accessionsⁱ

Primary accession | A0A0F7YUR7

Figura 3 Proteína A0A0F7YUR7.

3.5.2 Análisis de Características Primarias

El análisis inicial se centró en las propiedades fisicoquímicas básicas de la secuencia:

- **Longitud:** 68 aminoácidos, lo que sugiere una proteína de tamaño pequeño
- **Composición:** Predominancia de aminoácidos hidrofóbicos (A, L, V, G)
- **Carga:** Presencia de residuos cargados en el extremo C-terminal (KKK)
- **Punto isoeléctrico estimado:** Aproximadamente 9.5, indicando naturaleza básica
- **Peso molecular aproximado:** 7.2 kDa

Subcellular Locationⁱ

No UniProt annotations available.

Features

Showing features for transmembraneⁱ.

[🔍](#) [🔍](#) [🔍](#) [Download](#) [🔗](#)

1 10 20 30 40 50 60 70 80

M R A V A L L T L A L G A F L L G L S Y T P S Y G G S Y A Y Y V A N W G E I G V P N L V S A I L A G W R A Y D S L G E A S L L F T A V I G F Y L L G G K K K

±	TYPE	ID	POSITION(S)	DESCRIPTION
+	Transmembrane		54-75	Helical <input type="button" value="Automatic Annotation"/>

[Tools](#) [Add](#)

Keywordsⁱ

Cellular component | #Membrane

Figura 4 Gráfico de propiedades fisicoquímicas de la secuencia

3.5.3 Predicción de Estructura Secundaria

El análisis de estructura secundaria mediante herramientas computacionales reveló patrones significativos:

- Segmento N-terminal (aminoácidos 1-20): Alta probabilidad de formar hélice alfa con características hidrofóbicas
- Región central (aminoácidos 21-40): Mezcla de estructuras con tendencia a regiones desordenadas
- Segmento C-terminal (aminoácidos 41-68): Presencia de regiones transmembrana potenciales



Figura 5 Predicción de estructura secundaria

3.5.4 Análisis BLAST y Búsqueda de Homólogos

Se realizó una búsqueda exhaustiva mediante BLAST para identificar proteínas con similitud significativa:

Similar Proteinsⁱ

100% identity 90% identity 50% identity

A0A0F7YUR7

No similar proteins at 100% identity for this isoform.

[View all](#)

Similar Proteinsⁱ

100% identity 90% identity 50% identity

A0A0F7YUR7

UniRef90_A0A2Z2MGV4

	Protein name	Organism	Length
■	Sodium:proton antiporter	Thermococcus sp. 2157	80
■	Sodium:proton antiporter	Thermococcus sp. 5-4	80
■	Sodium:proton antiporter	Thermococcus barosii	80
■	Sodium:proton antiporter	Thermococcus radiotolerans	80
■	Sodium:proton antiporter	Thermococcus thioeducens	80
■	Sodium:proton antiporter	Thermococcus sp	80
■	Sodium:proton antiporter	Thermococcus sp. ES12	80
■	Sodium:proton antiporter	Thermococcus sp. MAR1	80

View these 8 entries in UniProtKB

[View all](#)

Similar Proteinsⁱ

100% identity 90% identity 50% identity

A0A0F7YUR7

UniRef50_A0A172WFD9

	Protein name	Organism	Length
■	Sodium:proton antiporter	Thermococcus eurythermalis	80
■	Sodium:proton antiporter	Thermococcus pacificus	80
■	MrpA C-terminal/MbhE domain-containing protein	Thermococcus profundus	81
■	Sodium:proton antiporter	Thermococcus radiotolerans	80
■	Uncharacterized protein	Thermococcus paralvinellae	94
■	Sodium:proton antiporter	Thermococcus sp. CX2	80
■	Sodium:proton antiporter	Thermococcus sp	81
■	Sodium:proton antiporter	Thermococcus aggregans	81
■	Multisubunit Na ⁺ /H ⁺ antiporter, MnhB subunit	Thermococcus nautili	79
■	Sodium:proton antiporter	Thermococcus sp. EP1	81

View all 63 entries in UniProtKB

[View all](#)

Figura 6 Resultados BLAST con alineamientos significativos

3.5.5 Interpretación Funcional

La integración de los resultados obtenidos permite proponer varias hipótesis sobre la posible función de esta proteína no caracterizada:

1. Proteína de membrana: Las características hidrofóbicas y la predicción de dominios transmembrana sugieren fuertemente una localización en membrana.

2. Posible transportador: El perfil de hidrofobicidad y la similitud con otras proteínas transportadoras apuntan a una posible función en el transporte de moléculas pequeñas.
3. Implicación en procesos de secreción: La presencia de un posible péptido señal sugiere participación en rutas de secreción bacteriana.
4. Posible lipoproteína: El patrón de aminoácidos y las características estructurales son consistentes con lipoproteínas procariotas.

3.6 Generación y Análisis de Secuencia Aleatoria

La generación de una secuencia proteica aleatoria de 100 aminoácidos representó un ejercicio diseñado para comprender los principios fundamentales de aleatoriedad molecular y las capacidades de las herramientas de búsqueda.

La secuencia aleatoria es de 100 aminoácidos:

EVHQCTFSVWENMWINDAHKGRWWMDWLALAGFPTFINSWRVIYDFLTEEWV
MHKAVSDT WMPFVPESSSHIGKYCSACKMDAGCKIHNECVYANACIVT

BLAST results | UniProt

https://www.uniprot.org/blast/uniprotkb/ncbiblast-R20250328-012325-0622-87170335-p1m/overview

UniProt BLAST Align Peptide search ID mapping SPARQL Tool results Advanced | List Search Help

Blast parameters

Identity

24.4 41

Score

80 86

E-Value

0.47 6.6

Taxonomy

Filter by taxonomy

Status

Unreviewed (TrEMBL) (7)

Other organisms

Quercus lobata (2)

Botryosphaeria parva (strain UCR-NP2) (1)

Castanea mollissima (1)

Eubacterium aggregans (1)

BLAST 7 results found in UniProtKB

Overview Taxonomy Hit Distribution Text Output Input Parameters API Request

Tools Download Add Customize columns Resubmit

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/> R1GIS1	R1GIS1_BOTPV	Rhodopsin domain-containing protein	UCRNP2_1647	Botryosphaeria parva (strain UCR-NP2) (Grapevine canker fungus) (Neofusicoccum parvum)	96 AA
<input type="checkbox"/> A0A1H3YE96	A0A1H3YE96_9FIRM	ATP:glycerol 3-phosphotransferase	SAMN04515656_103182	Eubacterium aggregans	498 AA
<input type="checkbox"/> A0A7N2LGQ8	A0A7N2LGQ8_QUELO	Cytochrome P450		Quercus lobata (Valley oak)	403 AA
<input type="checkbox"/> A0A7N2LJC0	A0A7N2LJC0_QUELO	Cytochrome P450 76AD1-like protein		Quercus lobata (Valley oak)	459 AA
<input type="checkbox"/> A0A8J4RPX3	A0A8J4RPX3_9ROSI	Cytochrome P450	CMV_008757	Castanea mollissima (Chinese chestnut)	507 AA
<input type="checkbox"/> A0A1H8GU60	A0A1H8GU60_9ACTN	ABC-type branched-chain amino acid transport system, substrate-binding protein	SAMN05660976_07636	Nonomuraea pusilla	905 AA
<input type="checkbox"/> A0AA88KPM2	A0AA88KPM2_NAELO	Uncharacterized protein	C9374_013834	Naegleria lovaniensis (Amoeba)	284 AA

40% (85) (0.47)

24.4% (84) (1.1)

28.6% (85) (1.5)

28.6% (85) (1.5)

28.6% (85) (1.5)

25.7% (83) (2.9)

41% (80) (6.6)

Entry Entry Name Protein Names Gene Names Organism Length

☐ R1GIS1 R1GIS1_BOTPV Rhodopsin domain-containing protein UCRNP2_1647 Botryosphaeria parva (strain UCR-NP2) (Grapevine canker fungus) (Neofusicoccum parvum) 96 AA

☐ A0A1H3YE96 A0A1H3YE96_9FIRM ATP:glycerol 3-phosphotransferase SAMN04515656_103182 Eubacterium aggregans 498 AA

☐ A0A7N2LGQ8 A0A7N2LGQ8_QUELO Cytochrome P450 Quercus lobata (Valley oak) 403 AA

☐ A0A7N2LJC0 A0A7N2LJC0_QUELO Cytochrome P450 76AD1-like protein Quercus lobata (Valley oak) 459 AA

☐ A0A8J4RPX3 A0A8J4RPX3_9ROSI Cytochrome P450 CMV_008757 Castanea mollissima (Chinese chestnut) 507 AA

☐ A0A1H8GU60 A0A1H8GU60_9ACTN ABC-type branched-chain amino acid transport system, substrate-binding protein SAMN05660976_07636 Nonomuraea pusilla 905 AA

☐ A0AA88KPM2 A0AA88KPM2_NAELO Uncharacterized protein C9374_013834 Naegleria lovaniensis (Amoeba) 284 AA

4 Conclusiones y recomendaciones.

La búsqueda demostró la potencia de las herramientas bioinformáticas modernas, particularmente UniProt y BLAST, para conseguir conocimiento biológico significativo a partir de datos moleculares. La integración de estas herramientas posibilita un enfoque holístico del análisis proteico, desde la obtención de secuencias hasta la formulación de hipótesis funcionales.

El experimento con secuencias aleatorias refuerza la robustez metodológica de BLAST, demostrando su capacidad para discriminar entre similitudes biológicamente significativas. Este aspecto es fundamental para comprender los límites y alcances de las herramientas bioinformáticas.

La exploración de proteínas no caracterizadas es uno de los territorios más desafiantes de la bioinformática. Los resultados obtenidos con la proteína A0A0F7YUR7 dan la importancia de métodos computacionales para la generación de hipótesis que orientan investigaciones posteriores.

Para futuras prácticas y aplicaciones en el ámbito bioinformático, se recomiendan las siguientes estrategias como la diversificación de bases de datos, complementar UniProt con otras bases especializadas como PDB (Protein Data Bank) para análisis estructurales para contextualización metabólica o la implementación de análisis avanzados. Incorporar métodos como Maximum Likelihood o Bayesianos para reconstruir con mayor precisión las relaciones evolutivas entre las secuencias estudiadas.

El campo de la bioinformática continúa su desarrollo exponencial, ofreciendo oportunidades significativas para la investigación proteómica. Las líneas de investigación representan extensiones naturales del trabajo realizado como el análisis comparativo de variantes de HBA1 asociadas con hemoglobinopatías o el modelado molecular dinámico para comprender las implicaciones funcionales de las regiones conservadas identificadas.

5 Referencias

- [1] The UniProt Consortium, "UniProt: the universal protein knowledgebase in 2021," *Nucleic Acids Research*, vol. 49, no. D1, pp. D480-D489, 2021.
- [2] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403-410, 1990.
- [3] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792-1797, 2004.