

Asignatura	Datos del alumno	Fecha
Secuenciación y Ómicas de Próxima Generación	Apellidos: Tome Castro	18/11/24
	Nombre: Xose Manuel	

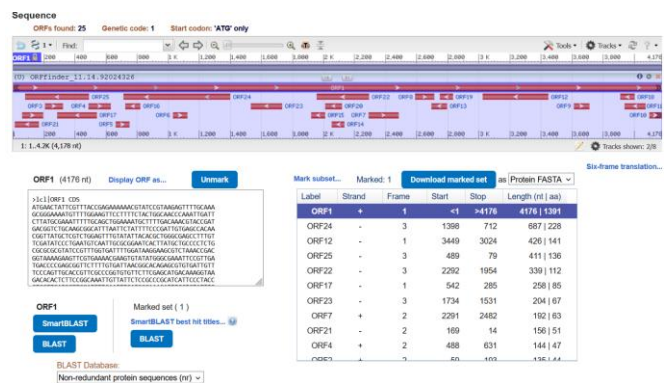
Análisis de secuencias usando herramientas disponibles en la red

La secuencia ORF1 proporcionada por el docente de la asignatura fue sometida a análisis bioinformáticos para determinar su relevancia funcional y filogenética. A continuación, se describen los métodos empleados y los hallazgos más destacados.

Identificación de ORFs con ORFfinder

Se utilizó ORFfinder para identificar los marcos de lectura abiertos (ORFs) en la secuencia problema. Inicialmente, se configuró para detectar ORFs con un mínimo de 75 nucleótidos, empleando el código genético estándar y permitiendo codones de iniciación alternativos (ATG y cualquier otro codón de inicio en ambos sentidos). Este análisis identificó 49 ORFs, aunque la mayoría eran fragmentos pequeños, posiblemente no funcionales.

Para refinar la búsqueda, se limitó a codones de inicio ATG, lo que redujo los ORFs detectados a 25. A pesar de esta mejora, muchos ORFs seguían siendo fragmentos. El ORF1 destacó como el más grande, abarcando gran parte de la secuencia y mostrando características que sugieren que podría codificar una proteína funcional, a pesar de que presenta un inicio y/o fin parcial.



Comparación y Anotación con BLAST

Se utilizó BLAST para comparar la secuencia de ORF1 contra bases de datos nucleotídicas y de proteínas:

1. **BlastN:** un resultado con Identidad del 100%, cobertura del

100% y un E-value de 0.0 de Neisseria gonorrhoeae.

2. **BlastX:** un resultado con secuencia fue traducida y comparada contra proteínas no redundantes (nr). El resultado mostró identidad del 100%, cobertura del 100%, y un E-value de 0.0 con la proteína DNA-directed RNA polymerase subunit beta.

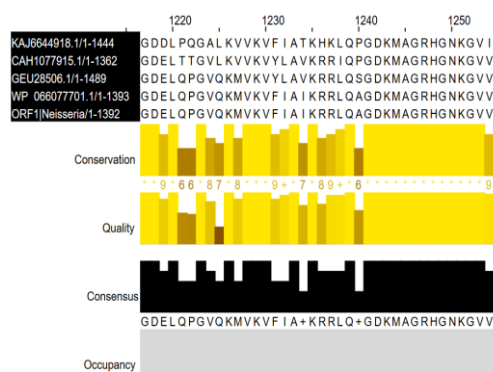
Asignatura	Datos del alumno	Fecha
Secuenciación y Ómicas de Próxima Generación	Apellidos: Tome Castro	18/11/24
	Nombre: Xose Manuel	

- Para evaluar homología en organismos diversos, se realizaron búsquedas excluyendo grupos específicos (*Neisseria* spp., bacterias y arqueas, y fungi/metazoos). De estas búsquedas, se seleccionaron cuatro proteínas con identidades del 94.47%-100%, cobertura entre 97%-100% y un E-value de 0.0.

Organismos	aa	Familia	Locus
<i>Neisseria Gonorrhoeae</i>	1392	Bacteria; Pseudomonadota; Betaproteobacteria	RPOB_NEIG2
<i>Bergeriella denitrificans</i>	1393	Bacteria; Pseudomonadati	WP_066077701
<i>Acinetobacter phage MD-2021a</i>	1362	Viruses; Duplodnaviria	CAH1077915
<i>Tanacetum cinerariifolium</i>	1489	Eukaryota; Viridiplantae	GEU28506
<i>Pseudolycoriella hygida</i>	1444	Eukaryota; Metazoa	KAJ6644918

Análisis de Conservación y Regiones Funcionales

Estos resultados destacan la conservación evolutiva de la proteína rpoB, reforzando su importancia funcional. Mediante alineamientos múltiples (**Clustal Omega**) y su visualización (**JalView**), se identificaron regiones conservadas y variables en las proteínas seleccionadas (ejemplo de una región conservada).

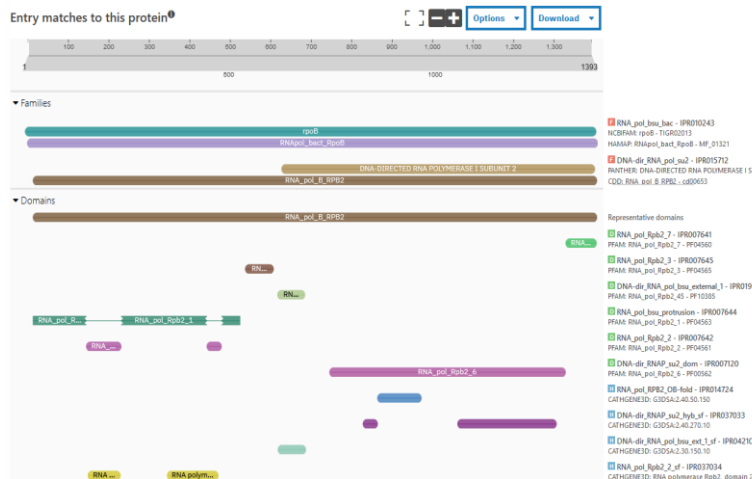


Mediante **Prosite**, entre las regiones conservadas, se destacaron (incluyendo MOTIFs), varios motivos funcionales esenciales para la actividad catalítica de la RNA polimerasa. Estas regiones posiblemente sean críticas para la interacción con RNA durante la transcripción.

- Regiones Conservadas: Presentan residuos cargados positivamente (Lys, Arg) e hidrofóbicos, típicos en dominios de interacción con ácidos nucleicos o lípidos como en posiciones 90-100, 410-420, o 1119-1420.

Asignatura	Datos del alumno	Fecha
Secuenciación y Ómicas de Próxima Generación	Apellidos: Tome Castro	18/11/24
	Nombre: Xose Manuel	

- Regiones Variable: con aminoácidos más polares, probablemente sean adaptaciones para interacciones en la superficie de la proteína, como en posiciones 200-220 y 800-840.



Se muestra a la derecha una imagen de InterProScan de los dominios de Tanacetum cinerariifolium, mostrando las regiones conservadas y posición de la secuencia.

Se muestran la conservación de regiones funcionales en

organismos tan diversos como bacterias, virus, plantas y metazoos destaca la importancia biológica de la proteína, reflejando una evolución bajo fuertes presiones selectivas que han mantenido características esenciales para la transcripción genética. Las regiones variables, por su parte, muestran adaptaciones específicas al entorno biológico, influyendo en interacciones celulares y estabilidad estructural.

1. ORF1 codifica la subunidad beta de la RNA polimerasa (rpoB), RNA_pol_B_RPB2 y DNA-Directed RNA Polymerase I subunit 2, esencial en Neisseria gonorrhoeae y los otros organismos.
2. Hay varias regiones conservadas críticas para su función catalítica, “GDKMAGRHHGNGKV”, o “KREEADELYE—I/OKKKK” entre otras.
3. La amplia distribución de la familia rpoB evidencia su relevancia evolutiva.
4. Las regiones variables indican posibles adaptaciones a contextos específicos.
5. Los dominios identificados por **InterProScan** se corresponden con las zonas conservadas detectadas en el alineamiento múltiple.