

Ana Sofia Santos Tedim Sousa Pedrosa

08/11/2025

Análisis de una secuencia usando herramientas disponibles en la red

Para realizar esta actividad se emplearon las siguientes herramientas de análisis disponibles en la red: i) **ORFfinder**(1) (NCBI) para detección de ORFs usando "*ATG and alternative initiation codons*" para buscar las ORFs. Los demás parámetros se han mantenido estándar. Se usó también **ORFfinder** para la traducción del gen a preteína; ii) se usó **BLASTn/BLASTp**(2) (NCBI) para búsqueda de similitud; iii) **COBALT**(3) y **Clustal Omega**(4) para el alineamiento múltiple; iv) **PROSITE**(5) para detección de motivos.

Los resultados obtenidos para la búsqueda de ORF pueden ser observados en la Figura 1. El ORF correcto se localizó en la **hebra positiva, frame +1**, corresponde a una proteína larga con 1392aa (rectángulo naranja en la Figura 1).



3. Análisis funcional de la secuencia

El mismo ORFfinder traduce a la secuencia de ADN a proteína (rectángulo en rojo Figura 1). Se usó blastn/blastp para identificar la secuencia de nucleótidos y la de proteína, respectivamente. La secuencia problema es idéntica a la subunidad β de la RNA polimerasa (*rpoB*) de *Neisseria gonorrhoeae*, comprando tanto en nucleótidos como en aminoácidos (Figura 2 y 3).

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Neisseria gonorrhoeae strain 5035 chromosome complete genome	Neisseria gonorrhoeae	7718	7718	100%	0.0	100.00%	2223133	CP104546.2
Neisseria gonorrhoeae isolate G97587 genome assembly chromosome I	Neisseria gonorrhoeae	7718	7718	100%	0.0	100.00%	2174841	LS999565.1

Figura 2. Resultados *blastn* para la secuencia problema.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=DNA-directed RNA polymerase subunit beta; Short=RNAP subunit beta; AltName: Full=RNA polyme... Neisseria gonorr...	Neisseria gonorrhoeae	2846	2846	100%	0.0	100.00%	1392	B4RQW2.1
RecName: Full=DNA-directed RNA polymerase subunit beta; Short=RNAP subunit beta; AltName: Full=RNA polyme... Neisseria gonorr...	Neisseria gonorrhoeae	2843	2843	100%	0.0	99.93%	1392	Q5F5R5.1
RecName: Full=DNA-directed RNA polymerase subunit beta; Short=RNAP subunit beta; AltName: Full=RNA polyme... Neisseria mening...	Neisseria meningitidis	2837	2837	100%	0.0	99.57%	1392	A1KR66.2

Figura 3. Resultados *blastp* para la secuencia problema.

Fueron el porcentaje de identidad y la cobertura fueron de 100%, bien como el e-value fue de 0.0, que permitieron identificar la secuencia como el gen *rpoB* *N. gonorrhoeae*, que codifica *pa*. Además, el valor de e-value obtenido de 0.0, indica que la probabilidad de que el alineamiento se deba al azar extremadamente baja, lo que suele ocurrir cuando hay 100% de identidad en las secuencias.

4. Comparación evolutiva

Para buscar RpoB en otros microorganismos se usó blastp con la base de datos RefSeq para los géneros/especies: *Escherichia coli*, *Enterococcus faecium* y *Salmonella* spp. Se seleccionó una secuencia proteica para cada microorganismo (en anexo) y procedió al alineamiento múltiple usando COBALT que permite una mejor visualización de las secuencias de Clustal Omega para la identificación de regiones conservadas y variables como se puede ver en la Figura 4. Las regiones en rojo observadas en la Figura 4 son las regiones conservadas y las regiones en gris son las regiones variables entre las 4 secuencias de proteínas usadas en el alineamiento.

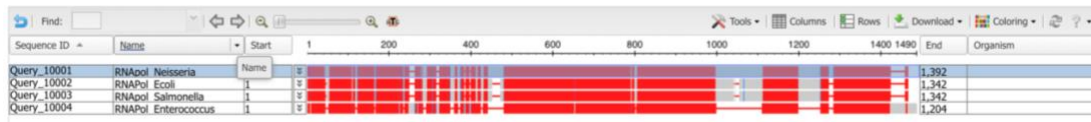


Figura 4. Resultados obtenidos en COBALT para las 4 secuencias comparadas.

Sin embargo, clustal omega da más información relativa al porcentaje de identidad entre las 4 secuencias (ver Figura S1 en anexo).

5. Análisis estructural

Se detectó un único dominio proteico conservado en RpoB de *N. gonorrhoeae* (Figura 5). Como indica PROSITE el dominio detectado es “RNA_POL_BETA, PS01166; RNA polymerases beta chain signature”, que se encuentra entre los aminoácidos 1104 y 1116. Confirmando que el gen problema codifica para RpoB.

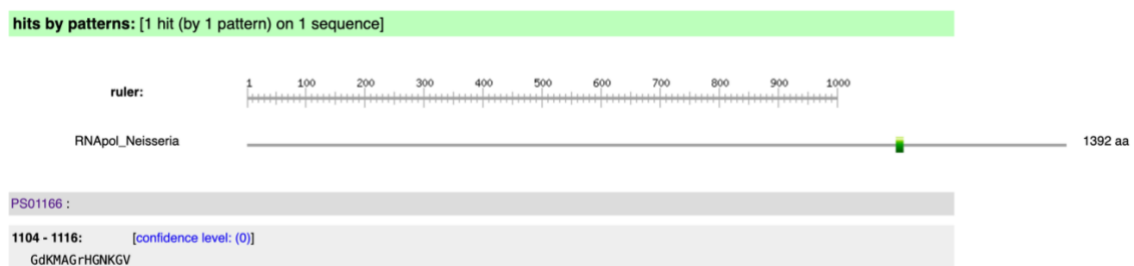


Figura 5. Resultado de dominios proteicos obtenido usando PROSITE.

Para verificar si este dominio se encontraba en alguna de las regiones conservadas de la proteína se buscó en el alineamiento obtenido con COBALT (Figura S2, en anexo) se los aminoácidos comprendidos entre 1104 e 1116 se encontraban en una región marcada a roja que indica conservación, lo que se ha verificado, indicando que el motivo encontrado es un motivo conservado.

6. Conclusiones:

La secuencia analizada codifica para la subunidad beta (RpoB) de la ARN polimerasa de *N. gonorrhoeae*. El alineamiento múltiple con ortólogos de *E. coli*, *Salmonella* spp y *Enterococcus* spp revela una alta conservación global, especialmente en el motivo funcional PS01166, lo que concuerda con el papel esencial y altamente conservado de la proteína RpoB en bacterias.

La combinación de herramientas web empleadas (ORFfinder, BLAST, COBALT, Clustal Omega y PROSITE) permitió realizar un flujo de análisis reproducible y bien documentado, siguiendo buenas prácticas de bioinformática básica.

7. Referencias:

1. ORFfinder Home - NCBI [Internet]. [cited 2025 Nov 6]. Available from: <https://www.ncbi.nlm.nih.gov/orffinder/>
2. BLAST: Basic Local Alignment Search Tool [Internet]. [cited 2025 Nov 9]. Available from: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
3. COBALT: - Cobalt RID (seqs) [Internet]. [cited 2025 Nov 7]. Available from: <https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi>
4. Clustal Omega < Job Dispatcher < EMBL-EBI [Internet]. [cited 2025 Nov 7]. Available from: <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>
5. ScanProsite [Internet]. [cited 2025 Nov 8]. Available from: <https://prosite.expasy.org/scanprosite/>

Anexo:

Secuencias usadas para hacer la comparación múltiple (en formato FASTA):

```
>WP_001314338.1 MULTISPECIES: DNA-directed RNA polymerase subunit beta [Escherichia]
MVYSYTEKKRIRKDFGKRPQVLDVPYLLSIQLDSFQKFIEQDPEGQYGLEAAFRSVFPIQSYSGNSELQY
VSYRLGEPVFDVQECQIRGVYTSAPLRVKLRVLIYEREAEPTVKDIKEQEVYMGELPLMTDNGTFVING
TERVIVSQLHRS PGVFFDSKGTTHSSGKVLNARIIPYRGSWLDFEFDPKDNLFVRIDRRRKLPAIIIL
RALNYTTEQILDLFFKEKVFIEIRDNLQOMELVPERLRGETASFIEANGKVYVEKGRRITARHIRQLEKD
DVKLIIEVPVEYIAGKVVAKYIDESTGELICAA NMELSLDLLAKLSQSGHKRIETLFTNDLDHGPYISET
LRVDPNTDRLSALVEIYRMMRPGEPTTREA AESLFENLFFSEDRYDLSAVGRMKFNRSLLREEIEGSGIL
SKDDIIDVMKKLIDIRNGKGEVDDIDHLGNRRIRSVGEMAENQFRVGLVVRVERAVKERLSLGDLDTLMPQ
DMINAKPISAAVKEFFGSSQLSQFMDQNNPLSEITHKRRISALGPGGLTRERAGFEVRDVHPTHYGRVCP
IETPEGPNIGLINSLSVYAQTNEYGFLETPYRKVTDGVVTDEIHYLSAIEEGNYVIAQANSNLDEEGHFV
EDLVTCRSKGESSLSRQDVYMDVSTQQVVS VGASLIPFLEHDDANRALMGANMQRQAVPTLRADKPLV
GTGMERAVAVDSGVTAVAKRGGVQYVDASRIVIKVNEDEMYPG EAGIDIYNLT KYTRSNQNTCINQMPC
VSLGEPVERGDVLDAGPSTDLGELALGQNM RVAFMPWNGYNFEDSILVSE RVVQEDRFTTIIHQELACVS
RDTKLGP EETADIPNVGEAALS KLDES GIVYIGA E V TGGDILVGKVT PKGETQLTPEEKL LRAIFGEKA
SDVKDSSLRVPNGVSGTVIDVQVFTRDGVEKDKRALEIEEMQLKQAKKDLSEELQILEAGLFSRIRAVLV
AGGVEAEKLDKLPDRDWLELGLTDEEKQNQLEQLAEQYDELKHEFEKKLEAKRRKITQGDDLAPGV LKIV
KVYLAVKRRIQPGDKMAGRHGNKGVISKINPIEDMPYDENGTPVDIVLNLGVP SRMNIGQILETHLGMA
AKGIGDKINAMLKQQQEVAKLREFIQRAYDLGADV RQKVDLSTFSDEEVMRLAENLRKGMPIATPVFDGA
KEAEIKELLKLGDLPTSGQITLFDGRTGEQFERPVTVGMYMLKLNHLVDDKMHARSTGSYSLVTTQQPLG
GKAQFGGQRFGE MEVWALEAYGAAYTLQ EMLTVKSDDVNGRTKMYKNIVDGNHQMPEPGMPESFNVLLKEI
RSLGINIELEDE

>NP_463022.1 RNA polymerase, beta subunit [Salmonella enterica subsp. enterica
serovar Typhimurium str. LT2]
MVYSYTEKKRIRKDFGKRPQVLDVPYLLSIQLDSFQKFIEQDPEGQYGLEAAFRSVFPIQSYSGNSELQY
VSYRLGEPVFDVQECQIRGVYTSAPLRVKLRVLIYEREAEPTVKDIKEQEVYMGELPLMTDNGTFVING
TERVIVSQLHRS PGVFFDSKGTTHSSGKVLNARIIPYRGSWLDFEFDPKDNLFVRIDRRRKLPAIIIL
RALNYTTEQILDLFFKEKVFIEIRDNLQOMELVPERLRGETASFIEANGKVYVEKGRRITARHIRQLEKD
DIKHIEVPVEYIAGKVVS KDYVDESTGELICAA NMELSLDLLAKLSQSGHKRIETLFTNDLDHGPYISET
VRVDPNTDRLSALVEIYRMMRPGEPTTREA AESLFENLFFSEDRYDLSAVGRMKFNRSLLRDEIEGSGIL
SKDDIIDVMKKLIDIRNGKGEVDDIDHLGNRRIRSVGEMAENQFRVGLVVRVERAVKERLSLGDLDTLMPQ
DMINAKPISAAVKEFFGSSQLSQFMDQNNPLSEITHKRRISALGPGGLTRERAGFEVRDVHPTHYGRVCP
IETPEGPNIGLINSLSVYAQTNEYGFLETPYRRVVDGVVTDEIHYLSAIEEGNYVIAQANSNLDEEGHFV
EDLVTCRSKGESSLSRQDVYMDVSTQQVVS VGASLIPFLEHDDANRALMGANMQRQAVPTLRADKPLV
GTGMERAVAVDSGVTAVAKRGGVQYVDASRIVIKVNEDEMYPG EAGIDIYNLT KYTRSNQNTCINQMPC
VSLGEPVERGDVLDAGPSTDLGELALGQNM RVAFMPWNGYNFEDSILVSE RVVQEDRFTTIIHQELACVS
RDTKLGP EETADIPNVGEAALS KLDES GIVYIGA E V TGGDILVGKVT PKGETQLTPEEKL LRAIFGEKA
SDVKDSSLRVPNGVSGTVIDVQVFTRDGVEKDKRALEIEEMQLKQAKKDLSEELQILEAGLFSRIRAVLV
SSGVEAEKLDKLPDRDWLELGLTDEEKQNQLEQLAEQYDELKHEFEKKLEAKRRKITQGDDLAPGV LKIV
KVYLAVKRRIQPGDKMAGRHGNKGVISKINPIEDMPYDENGTPVDIVLNLGVP SRMNIGQILETHLGMA
AKGIGDKINAMLKQQQEVAKLREFIQRAYDLGADV RQKVDLSTFSDEEVLRLAENLRKGMPIATPVFDGA
KEAEIKELLKLGDLPTSGQITLFDGRTGEQFERPVTVGMYMLKLNHLVDDKMHARSTGSYSLVTTQQPLG
GKAQFGGQRFGE MEVWALEAYGAAYTLQ EMLTVKSDDVNGRTKMYKNIVDGNHQMPEPGMPESFNVLLKEI
RSLGINIELEDE

>WP_002389470.1 MULTISPECIES: DNA-directed RNA polymerase subunit beta [Enterococcus]
MAGHVVKYKGKRRERSFARISEVLELPNLIEIQTDSYQWFLDEGLREMFEDILPIDDFQGNLSLEFVDYE
LKEPKYTVEEARAH DANYSAPLHVTLRLTNRETGEIKSQEVFFGDFPLMTEMGTFIINGAERVIVSQLVR
SPGVYFHGKVDKNGKEGFGSTVIPN RGAWLEMETDAKDISYVRIDRTRKIPLTVLVRALGFGSDDTIFEI
FGDSESLRNTIEKDLHKNASDSRTEEGLKDIYERLRPGEPKTADSSRSLLTARFFDPKRYDLANVG RYKV
NKKLDLKRLLNLT LAETLVDPETGEIIVEKGTVLTHQIMETLGEYIDNGLNSVTYYPSED AVVTEPMTI
QVIQVLSPKDPERIVNVIGNGYPDDSVKTVRPADIVASMSYFFNLMEDIGNVDDIDHLGNRRIRSVGELL
QNQFRIGLARMERVVRERMSIQDTETLTPQQLINIRPVVASIKEFFGSSQLSQFMDQTNPLGELTHKRR
SALGPGGLTRDRAGYEV RDVHYSHYGRMCPIETPEGPNIGLINSLSYAKVNKFGFIETPYRRVD RATGR
VTDQVDYLTADIEDHYI V AQANSLLNEDGTFANDVVMARLQSENLEVAVDKVDYMDVSPKQVAVATACI
PFLENDSSNRALMGANMQRQAVPLIQPRSPWVG TGM EYKSAHDSGAALLCKHDGVVEFVDAKEIRVRDN
GALDKYMVTKFRSNSGTSYNQRP I VHLGEKVEKGD TLADGPSMEEGEMALGQNVLVAFMTWEGYNYEDA
IIMSRRLVKDDVYTSVHIEEYSEARDTKLGPEEITREIPNVGEDALKDLDEMGIIRIGA EVQDGDLLVG
KVTPKGVTELS AEERLLHAIFGEKAREVRDTSLRVPHGGGGIVHDVKIFTREAGDELSPGVNMLVRVYIV
QKRKIH EGD KMAGRHGNKGVS RIMPEEDMPFLPDGTPVDIMLNLGVP SRMNIGQVLELHLGMAARQLG
IHVATPVFDGATDEDVWETVREAGMASDAKTVLYDGRTEGPF DNRI SVGVMYMIKLAHMVDDKLHARSIG
PYSLVTTQQPLGGKAQFGGQRFGE MEVWALEAYGAAYTLQ EILTYKSDDVVG RVKTYE AIVKG EPIPKPGV
PESFRVLVKELQSLGLDMRVL D IEEAEIELRDMDDDDDLITVDALTKFAEQQSAKQLEKEAESVVKEEA
QDVVQEIETAE DRD
```

```
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: RNAPol_Enterococcus 100.00 56.09 56.51 56.32
2: RNAPol_Neisseria    56.09 100.00 66.24 65.80
3: RNAPol_Ecoli        56.51 66.24 100.00 98.73
4: RNAPol_Salmonella    56.32 65.80 98.73 100.00
```

Figura S1. Matriz de identidad obtenida usando clustal omega.

Query_10001	1048	LAKQLELIKLSLQQKREEADELYEIKKKKL	TGDELPQGVQKMKVFI	AIKRRRLQAGDKMAGRHGNKGVVSRILPVEDMP	1127
Query_10002	1007	KQNQLEQLAEQYDELKHEFEKKLEAKRRKIT	QGDDLPAGVLKIVKVYLAVKRRIQPGDKMAGRHGNKGVISKINPIEDMP		1086
Query_10003	1007	KQNQLEQLAEQYDELKHEFEKKLEAKRRKIT	QGDDLPAGVLKIVKVYLAVKRRIQPGDKMAGRHGNKGVISKINPIEDMP		1086
Query_10004	893	-----	AGDELSPGVNMLVRVYIVQKRKI	HEGDKMAGRHGNKGVVSRIMPEEDMP	941

Figura S2. Representación del alineamiento de COBALT en la región donde se localiza el motivo conservado.