Datos del estudiante

**Nombre y apellidos** Ana Sofia Santos Tedim Sousa Pedrosa

**Fecha de entrega** 08/11/2025

Análisis de una secuencia usando herramientas disponibles en la red

1. **Introducción:**

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 preoteina; 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.

1. **Identificación de marcos de lectura abiertos (ORF)**

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).

**Interfaz de usuario gráfica, Aplicación, Tabla

El contenido generado por IA puede ser incorrecto.**

**Figura 1. Resultados obtenidos para la secuencia problema usando ORFfinder.**

1. **Análisis funcional de la secuencia**

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

Interfaz de usuario gráfica, Aplicación, Tabla

El contenido generado por IA puede ser incorrecto.

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

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

**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 *rpo*B *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.

1. **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.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

**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).

1. **Análisis estructural**

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

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

**Figura 5. Resultado de dominios proteicos obtenido usando PROSITE.**

Para verificar si este domino 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.

1. **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.

1. **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

VSYRLGEPVFDVQECQIRGVTYSAPLRVKLRLVIYEREAPEGTVKDIKEQEVYMGEIPLMTDNGTFVING

TERVIVSQLHRSPGVFFDSDKGKTHSSGKVLYNARIIPYRGSWLDFEFDPKDNLFVRIDRRRKLPATIIL

RALNYTTEQILDLFFEKVIFEIRDNKLQMELVPERLRGETASFDIEANGKVYVEKGRRITARHIRQLEKD

DVKLIEVPVEYIAGKVVAKDYIDESTGELICAANMELSLDLLAKLSQSGHKRIETLFTNDLDHGPYISET

LRVDPTNDRLSALVEIYRMMRPGEPPTREAAESLFENLFFSEDRYDLSAVGRMKFNRSLLREEIEGSGIL

SKDDIIDVMKKLIDIRNGKGEVDDIDHLGNRRIRSVGEMAENQFRVGLVRVERAVKERLSLGDLDTLMPQ

DMINAKPISAAVKEFFGSSQLSQFMDQNNPLSEITHKRRISALGPGGLTRERAGFEVRDVHPTHYGRVCP

IETPEGPNIGLINSLSVYAQTNEYGFLETPYRKVTDGVVTDEIHYLSAIEEGNYVIAQANSNLDEEGHFV

EDLVTCRSKGESSLFSRDQVDYMDVSTQQVVSVGASLIPFLEHDDANRALMGANMQRQAVPTLRADKPLV

GTGMERAVAVDSGVTAVAKRGGVVQYVDASRIVIKVNEDEMYPGEAGIDIYNLTKYTRSNQNTCINQMPC

VSLGEPVERGDVLADGPSTDLGELALGQNMRVAFMPWNGYNFEDSILVSERVVQEDRFTTIHIQELACVS

RDTKLGPEEITADIPNVGEAALSKLDESGIVYIGAEVTGGDILVGKVTPKGETQLTPEEKLLRAIFGEKA

SDVKDSSLRVPNGVSGTVIDVQVFTRDGVEKDKRALEIEEMQLKQAKKDLSEELQILEAGLFSRIRAVLV

AGGVEAEKLDKLPRDRWLELGLTDEEKQNQLEQLAEQYDELKHEFEKKLEAKRRKITQGDDLAPGVLKIV

KVYLAVKRRIQPGDKMAGRHGNKGVISKINPIEDMPYDENGTPVDIVLNPLGVPSRMNIGQILETHLGMA

AKGIGDKINAMLKQQQEVAKLREFIQRAYDLGADVRQKVDLSTFSDEEVMRLAENLRKGMPIATPVFDGA

KEAEIKELLKLGDLPTSGQITLYDGRTGEQFERPVTVGYMYMLKLNHLVDDKMHARSTGSYSLVTQQPLG

GKAQFGGQRFGEMEVWALEAYGAAYTLQEMLTVKSDDVNGRTKMYKNIVDGNHQMEPGMPESFNVLLKEI

RSLGINIELEDE

>NP\_463022.1 RNA polymerase, beta subunit [Salmonella enterica subsp. enterica serovar Typhimurium str. LT2]

MVYSYTEKKRIRKDFGKRPQVLDVPYLLSIQLDSFQKFIEQDPEGQYGLEAAFRSVFPIQSYSGNSELQY

VSYRLGEPVFDVQECQIRGVTYSAPLRVKLRLVIYEREAPEGTVKDIKEQEVYMGEIPLMTDNGTFVING

TERVIVSQLHRSPGVFFDSDKGKTHSSGKVLYNARIIPYRGSWLDFEFDPKDNLFVRIDRRRKLPATIIL

RALNYTTEQILDLFFEKVVFEIRDNKLQMELIPERLRGETASFDIEANGKVYVEKGRRITARHIRQLEKD

DIKHIEVPVEYIAGKVVSKDYVDESTGELICAANMELSLDLLAKLSQSGHKRIETLFTNDLDHGPYISET

VRVDPTNDRLSALVEIYRMMRPGEPPTREAAESLFENLFFSEDRYDLSAVGRMKFNRSLLRDEIEGSGIL

SKDDIIDVMKKLIDIRNGKGEVDDIDHLGNRRIRSVGEMAENQFRVGLVRVERAVKERLSLGDLDTLMPQ

DMINAKPISAAVKEFFGSSQLSQFMDQNNPLSEITHKRRISALGPGGLTRERAGFEVRDVHPTHYGRVCP

IETPEGPNIGLINSLSVYAQTNEYGFLETPYRRVVDGVVTDEIHYLSAIEEGNYVIAQANSNLDDEGHFV

EDLVTCRSKGESSLFSRDQVDYMDVSTQQVVSVGASLIPFLEHDDANRALMGANMQRQAVPTLRADKPLV

GTGMERAVAVDSGVTAVAKRGGTVQYVDASRIVIKVNEDEMYPGEAGIDIYNLTKYTRSNQNTCINQMPC

VSLGEPVERGDVLADGPSTDLGELALGQNMRVAFMPWNGYNFEDSILVSERVVQEDRFTTIHIQELACVS

RDTKLGPEEITADIPNVGEAALSKLDESGIVYIGAEVTGGDILVGKVTPKGETQLTPEEKLLRAIFGEKA

SDVKDSSLRVPNGVSGTVIDVQVFTRDGVEKDKRALEIEEMQLKQAKKDLSEELQILEAGLFSRIRAVLV

SSGVEAEKLDKLPRDRWLELGLTDEEKQNQLEQLAEQYDELKHEFEKKLEAKRRKITQGDDLAPGVLKIV

KVYLAVKRRIQPGDKMAGRHGNKGVISKINPIEDMPYDENGTPVDIVLNPLGVPSRMNIGQILETHLGMA

AKGIGDKINAMLKQQQEVAKLREFIQRAYDLGADVRQKVDLSTFSDDEVLRLAENLRKGMPIATPVFDGA

KEAEIKELLKLGDLPTSGQITLFDGRTGEQFERPVTVGYMYMLKLNHLVDDKMHARSTGSYSLVTQQPLG

GKAQFGGQRFGEMEVWALEAYGAAYTLQEMLTVKSDDVNGRTKMYKNIVDGNHQMEPGMPESFNVLLKEI

RSLGINIELEDE

>WP\_002389470.1 MULTISPECIES: DNA-directed RNA polymerase subunit beta [Enterococcus]

MAGHVVKYGKHRERRSFARISEVLELPNLIEIQTDSYQWFLDEGLREMFEDILPIDDFQGNLSLEFVDYE

LKEPKYTVEEARAHDANYSAPLHVTLRLTNRETGEIKSQEVFFGDFPLMTEMGTFIINGAERVIVSQLVR

SPGVYFHGKVDKNGKEGFGSTVIPNRGAWLEMETDAKDISYVRIDRTRKIPLTVLVRALGFGSDDTIFEI

FGDSESLRNTIEKDLHKNASDSRTEEGLKDIYERLRPGEPKTADSSRSLLTARFFDPKRYDLANVGRYKV

NKKLDLKTRLLNLTLAETLVDPETGEIIVEKGTVLTHQIMETLGEYIDNGLNSVTYYPSEDAVVTEPMTI

QVIQVLSPKDPERIVNVIGNGYPDDSVKTVRPADIVASMSYFFNLMEDIGNVDDIDHLGNRRIRSVGELL

QNQFRIGLARMERVVRERMSIQDTETLTPQQLINIRPVVASIKEFFGSSQLSQFMDQTNPLGELTHKRRL

SALGPGGLTRDRAGYEVRDVHYSHYGRMCPIETPEGPNIGLINSLSSYAKVNKFGFIETPYRRVDRATGR

VTDQVDYLTADIEDHYIVAQANSLLNEDGTFANDVVMARLQSENLEVAVDKVDYMDVSPKQVVAVATACI

PFLENDDSNRALMGANMQRQAVPLIQPRSPWVGTGMEYKSAHDSGAALLCKHDGVVEFVDAKEIRVRRDN

GALDKYMVTKFRRSNSGTSYNQRPIVHLGEKVEKGDTLADGPSMEEGEMALGQNVLVAFMTWEGYNYEDA

IIMSRRLVKDDVYTSVHIEEYESEARDTKLGPEEITREIPNVGEDALKDLDEMGIIRIGAEVQDGDLLVG

KVTPKGVTELSAEERLLHAIFGEKAREVRDTSLRVPHGGGGIVHDVKIFTREAGDELSPGVNMLVRVYIV

QKRKIHEGDKMAGRHGNKGVVSRIMPEEDMPFLPDGTPVDIMLNPLGVPSRMNIGQVLELHLGMAARQLG

IHVATPVFDGATDEDVWETVREAGMASDAKTVLYDGRTGEPFDNRISVGVMYMIKLAHMVDDKLHARSIG

PYSLVTQQPLGGKAQFGGQRFGEMEVWALEAYGAAYTLQEILTYKSDDVVGRVKTYEAIVKGEPIPKPGV

PESFRVLVKELQSLGLDMRVLDIEEAEIELRDMDDDDDDLITVDALTKFAEQQSAKQLEKEAESVVKEEA

QDVVQEIETAEDRD

Interfaz de usuario gráfica, Texto, Aplicación, Word

El contenido generado por IA puede ser incorrecto.

**Figura S1. Matriz de identidad obtenida usando clustal omega.**

**Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.**

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