**Taller 1-Herramientas Computacionales para el Análisis de Bases de Datos Biológicas (Bioinformática)**. **Profesor Pedro A. Moreno. – 2025\_I**

**EISC, Facultad de Ingeniería, Universidad del Valle**

Trabaje en grupos de dos estudiantes, con las 12 secuencias “anónimas” de ADN dadas en el Anexo. Responda las siguientes preguntas y presente los resultados en formato artículo científico (título, autores, créditos institucionales, resumen, introducción, metodología, resultados, discusión, conclusiones, referencias, tablas y figuras). Entrega el taller en formato pdf para el día 15 de noviembre a las 10pm, vía mail.

1) Determine a qué organismo pertenecen las secuencias (Utilice el algoritmo BLAST del NCBI) y en qué cromosomas mapean cada una de estas con sus coordenadas en pb.

2) Mediante alineamiento múltiple con el algoritmo CLUSTAL, identifique los motivos de secuencias conservadas que tienen y su potencial función como gen en las regiones de: control (promotor), codificante y terminal, utilizando el programa BioEdit 7.2 y LOGO u otro programa disponible. Discuta

3) Prediga si estas secuencias tienen un potencial promotor, utilizando el programa BANANAde EMBOSS. Discuta

4) Determine las secuencias de las proteínas que estos potenciales genes podrían codificar utilizando ORFfinder del NCBI. Discuta

5) Determine qué tipo de estructuras secundarías tienen estas proteínas, utilizando el programa JPred 4. Analice su estructura molecular y discuta.

6) Busque si existe en el PDB un modelo en 3D de estas proteínas. Analice su estructura molecular a diferentes niveles de representación y su utilidad práctica.

7) Tome una de las secuencias de ADN, tradúzcala a proteína y busque con ésta en el NCBI (con el algoritmo BLAST) otras moléculas similares en otros organismos similares o diferentes. Escoja las secuencias de 10 organismos (las que quieras). Elabore árboles filogenéticos. Utilice el programa MEGA o MEGA X) y utilice los cuatro algoritmos básicos: UPGMA, NJ, MP, MV (o ML). Incline su conclusión por alguno o algunos algoritmos. Utilice como secuencia Outgroup aquella que tenga la menor similaridad entre las 10 que escogió. Discuta los resultados.

8) Utilice la plataforma STRING para determinar el grafo de interacciones de las 12 proteínas. Discuta.

9) Finalmente, discuta cómo estas proteínas podrían participar en el proceso del movimiento del espermatozoide del organismo en cuestión que ha determinado en el punto 1).

Nota: Utilice algoritmos, herramientas sugeridas u otras herramientas disponibles en la red para responder a cada una de las preguntas.

**ANEXO**

>Seq1

gaaatgtacatcattgtttgctatagatctaggacagagataagagagaaatttgctata

aattagcaggttgatttagggggatccggatttgttttcacacactacacagcaGTcGcC ATGGCCCAATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTCTTC AATGCACCATACGACGACAAGCACACGTATCACATCAAAGTGATTAACTCCTCGGCTCGC CGTATTGGATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCATGT GGAGTTTTGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCCTTC GGTCAGGAGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCCGGATGGA GCTGCCAAGCAATTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGAAC CTCCCAATTGAGTACAACccataagaccatctcctctgaatcctctaataaatggagatt ggaggccactattgacttcatcaatttgaaaaaaaatgtttgccttgagagtaagaggga ttttattcccccaccctttgaatgcctggagaagcacaatttcgtacatcccacatgctt

gattgaactttg

>Seq2

gaaatgtacatcttagggattactatagatctaggacagagataagagagaaatttgcta taaattaattatttgggccagggggccggtgcggagttttcacactctacaaagaaGTCG CCATGGCCCAATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTCT TCAATGCACCATACGACGACAAGCACACGTATCACATCAAAGTGATTAACTCCTCGGCTC GCCGTATTGGATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCAT GTGGAGTTTTGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCCT TCGGTCAGGAGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCCGGATG GAGCTGCCAAGCAGCTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGA ACCTCCCAGCTGAGTACAACccataagacaagttaccctctgaatcctctaataaatTGA AAATTTAAAAAAAGAATTAAGTGAAAGAATACGAAAACTATTTATCAAATTTAAAATTCA ACGAAAAACAAAGCAAAGCAGATCTTTTAGTTTTTAATGCTCCAAATGAACTCttcttat taaaaagattgaactttg

>Seq3

gaaatgttgtttgctatagatctaggacagagataagagagaaatttgctataaattggt

tttcccgttttggcgccaggaccgagagccggctctcttctacaaagaaGTCGCCATGGC CCAATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTCTTCAATGC

ACCATACGACGACAAGCACACGTATCACATCAAAGTGATACTGTCCTCGGCTCGCCGTAT TGGATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCATGTGGAGT TTTGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCCTTCGGTCA GGAGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCCGGATGGAGCTGC CAAGCAGCTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGAACCTCCC AGCTGAGTACAACccataagaccatctcctctgaatcctctaataaatggtatacaccca cccattaccaggagggattatttgaaaaaaaatgtttgccttgagttacctaggggcctt

agttcaacttaactgattttgaatgcctggagaacaggcacaatttcggtacacccagct

tgattgaactttg

>Seq4

gaaatgtacatcacgttgtttgctatagatctaggacagagataagagagaaatttgcta taaaggcggggtggcgcgctttaccacattcggctttggctttcgctcgctcgctaaGTC GCCATGGCCCAATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTC TTCAATGCACCATACGACGACAAGCACACGTATCACATCAAAGTGATACTGTCCTCGGCT CGCCGTATTGGATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCA TGTGGAGTTTTGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCC TTCGGTCAGGAGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCCGGAT GGAGCTGCCAAGCAGCTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAG AACCTCCCAGCTGAGTACAACaattccatttgggtgtttaggtttttaaaataaataggt tacccaaaaggggacctcaattgggcttgtttgaaaaaaaatgtttgccttgagagtaag aaggccttagttcaacttttctgattttgaatgcctggaggagctttattttatttacac

ccacatgcttgattgaactttg

>Seq5

gaaatgtacatcacgttgtttgctatagatctaggacagagataagagagaaatttgcta taaattgagcaggttttgccataatcttcactgtgagttttcacactctacaaagaaGTC

GCCATGGCCCAATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTC TTCAATGCACCATACGGGCATGGGGGGGCACACGTATCACATCAAAGTGATACTGTCCTC GGCTCGCCGTATTGGATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCC

ACCATGTGGAGTTTTGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTT TGCCTTCGGTCAGGAGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCC GGATGGAGCTGCCAAGCAGCTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCG CAAGAACCTCCCAATTGAGTACAACccgagcgaccaggggatttaggccacccccaataa atcgttctctgattatgtgtttttgttcatcaatttgaaaaaaaatgtttgccttgagag

taagaaggccttagttcaacttttctgattttgaatgcctggagaagcacaatttcggta

cacccacatgcttgattgaactttg

>Seq6

gaaatgtcgttgtttgctatagatctaggacagagataagagagaaatttgctataaagg cccttccccaccccgcgcgtacgggcccctatccccgttaagggcgcttaaGTCGCCATG GCCCAATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTCTTCAAT GCACCATACGACGACAAGCACACGTATCACATCAAAGTGATACTGTCCTCGGCTCGCCGT ATTGGATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCATGTGGA GTTTTGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCCTTCGGT CAGGAGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCCGGATGGAGCT GCCAAGCAATTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGAACCTC CCAATTGAGTACAACccataagaccatctcctctgaatcctctaataaatcgttctctga ttatgtgtttttgttcatcaatttgaaaaaaaatgtttgccttgagagtaagaaggcctt

agttcaacttttctgattttgaatgcctggagaagcacaatttcggtacacccacatgct

tgattgaactttg

>Seq7

gaaatgtacatcacgttgttctaggacagagataagagagaaatttgctataaaccctct ggccacccttcttaaacccggcggagagcggtgtgttctacaaagaaGTCGCCATGGCCC AATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTCTTCAATGCAC CATACGACGACAAGCACACGTATCACATCAAAGTGATTAACTCCTCGGCTCGCCGTATTG GATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCATGTGGAGTTT TGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCCTTCGGTCAGG AGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCCGGATGGAGCTGCCA

AGCAATTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGAACCTCCCAA TTGAGTACAACccataagaccatctcctctgaatcctctaataaatcgttctctgattat gtgtttttgttcatcaatttgaattcgttctttatatgtctgtctaagaagaacacaatc

aacactacctgattttgaatgcctggagaagcacaatttcggtacacccacatgcttgat tgaactttg

>Seq8

gaaatgtacatcacgttgtttgctatagatctaggacagagataagagagaaatttgcta taaaattagggcggttcgggcttatcgggggccctcttctatctagctagctgatcgtag

ctCATGGCCCAATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTC TTCAATGCACCATACGACGACAAGCACACGTATCACATCAAAGTGATTAACTCCTCGGCT CGCCGTATTGGATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCA TGTGGAGTTTTGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCC TTCGGTCAGGAGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCCGGAT GGAGCTGCCAAGCAATTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAG AACCTCCCAATTGAGTACAACggaggattcttttttcccttgggtcttataataaatcgt tctctgattatgtgtttttgttcatcaatttgaaaaaaaatgtttgccttgagagtaaga

aggccttagttcattagggcttttatttcccgttcttcgaagcacaatgtatttatggcc

cacatgcttgattgaactttg

>Seq9

gaaatgtacatcacgttgtttgctatagatagagataagagagaaatttgctataaagga tttctctttcttcactgtgagttttcacactctacaaagaaGTCGCCATGGCCCAATCCG TCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTCTTCAATGCACCATACG ACGACAAGCACACGTATCACATCAAAGTGATTAACTCCTCGGCTCGCCGTATTGGATACG GTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCATGTGGAGTTTTGGACC CAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCCTTCGGTCAGGAGGATA CAATTTATTGACCGTATCACCGTTGAGTGGACCAATACCCCGGATGGAGCTGCCAAGCAA TTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGAACCTCCCAATTGAG TACAACccataagaccatctcctctgaatcctctaataaatcgttctctgattatgtgtt

tttgttcatcaatttgaaaaaaaatgtttgccttgagagtaagaaggccttagttcaact

tttctgattttgaatgcctggagaagcacaatttcggtacacccacatgcttgattgaac

tttg

>Seq10

gaaatgtacatcacgtctatagatctaggacagagataagagagaaatttgctataaatt gagcaggttttgccataaggttaaatctctttccccacaaagaaGTCGCCATGGCCCAAT CCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTCTTCAATGCACCAT ACGACGACAAGCACACGTATCACATCAAAGTGATACTGTCCTCGGCTCGCCGTATTGGAT ACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCATGTGGAGTTTTGG ACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCCTTCGGTCAGGAGG ATACAAACAACGACCGTATCACCGTTCCCGGGACCAATACCCCGGATGGAGCTGCCAAGC AATTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGAACCTCCCAATTG AGTACAACaaggccgtgtggcttttgggctggcgaggaataaatcgttctctgattatgt gtttttgttcatcaatttgaaaaaaaatgtttgccttgagagaaccgtttcttttgttca

acttttctgattttgaatgcctggagaagcacaatttcgtacatcccacatgcttgataa

ttcggggg

>Seq11

gaaatgtacatcacgtagatctaggacagagataagagagaaatttgctataaagggctt attattttttcccccgcggggcgcgcgcgcttttcGCCATGGCCCAATCCGTCCCACCAG GAGACATCCAGACCCAGCCGGGTACGAAGATCGTCTTCAATGCACCATACGACGACAAGC ACACGTATCACATCAAAGTGATTAACTCCTCGGCTCGCCGTATTGGATACGGTATCAAGA CCACCAATATGAAGAGACTTGGAGTTGATCCACCATGTGGAGTTTTGGACCCAAAGGAGG CAGTTCTTCTGGCAGTTTCCTGTGATGCCTTTGCCTTCGGTCAGGAGGATACAAACAACG ACCGTATCACCGTTGAGTGGACCAATACCCCGGATGGAGCTGCCAAGCAGCTCCGCCGTG AATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGAACCTCCCAGCTGAGTACAACTTTT CTCGGATATCGCTtatatcttctcgaataaatcgttctctgattatgtgtttttgttcat caatttgaaaaaaaatgtttgccttgagagtaagaaggccttagttcaacttttctgatt

ccggtttccggggttttcacaatttcgtacatcccacatgcttgattgaactttg

>Seq12

gaattatcaatgtacatgctatagatctaggacagagataagagagaaatttgctataaa ttgagcaggttttgccataatcttcactgtgagttttcacactctacaaagaaGTCGCCA TGGCCCAATCCGTCCCACCAGGAGACATCCAGACCCAGCCGGGTACGAAGATCGTCTTCA AGCTACCATACGACGACAAGCACACGTATCACATCAAAGTGATACTGTCCTCGGCTCGCC GTATTGGATACGGTATCAAGACCACCAATATGAAGAGACTTGGAGTTGATCCACCATGTG GAGTTTTGGACCCAAAGGAGGCAGTTCTTCTGGCAGTTTCCTGTGAGCTCTTTGCCTTCG GTCAGGAGGATACAAACAACGACCGTATCACCGTTGAGTGGACCAATACCCCGGATGGAG CTGCCAAGCAGCTCCGCCGTGAATGGTTCCAAGGAGACGGTATGGTTCGTCGCAAGAACC TCCCAGCTGAGTACAACattatgggacgggattaggagaaggggaaataaaaaaggtctt tcttgcgcgcttttgttcatttgaaaaaaaatgtttgccttgagagtaagaaggccttag

ttcaacttttctgattttgaatgcctggagaagcacaatttcgtacatgcttgattgaac

tttg