

# CURSO: CC471

## Practica 2.

### 1. Introducción

PRACTICA : ALINEAMIENTO DE SECUENCIAS.

#### Objetivo general:

Conocer, Aplicar, y programar **algoritmos de alineamiento** de secuencias. Utilizar metodos manuales y computarizados para el analisis de secuencias y su alineamiento.

#### Objetivos particulares:

- Metodo de Dot – Plot
- Programacion de algoritmos de alineamiento
- Utilizacion de herramientas de alineamiento de secuencias.

#### RECURSOS INFORMÁTICOS

<http://www.ncbi.nlm.nih.gov/> (Nacional Center for Biotechnology Information – NCBI)

<http://www.ebi.ac.uk/> (European Bioinformatics Institute- EBI- EMBL)

<http://www.ddbj.nig.ac.jp/Welcome-e.html> (DNA Databank of Japan- DDBJ)

<http://cube.univie.ac.at/gepard>

#### DESARROLLO

##### I. Alineamiento de secuencias con el método de Dot – Plot

**Hacer un programa que dibuje un dotplot de dos secuencias de la misma longitud (en modo texto)**

1.1 Dadas dos secuencias de nucleótidos:

A = ACTGG

B = ACTAA

Realizar el **dotplot** correspondiente - haga un screenshot.

1.2 Dada una secuencia con repeticiones en tandem perfectas. Un **dotplot** obtenido a partir de la comparación de esta secuencia consigo misma:

- a. Muestra dos diagonales que se cruzan en el centro (X).
- b. Muestra dos diagonales paralelas
- c. Muestra dos diagonales invertidas que se cruzan en un vértice (V).
- d. Muestra una diagonal principal y varias diagonales más pequeñas correspondientes a las repeticiones .
- e. Muestra una diagonal principal y varias diagonales más pequeñas invertidas correspondientes a las repeticiones.

1.3 Dibuje el **dotplot** para las siguientes secuencias y deduzca (sin usar el algoritmo) cuál es el camino que tomaría un algoritmo de alineamiento global:

A = ACTGGTCAACTGGTCA

B = ACTAACTGGTCAATCA

1.4 Realizar un dotplot de la siguiente secuencia consigo misma. ¿Qué características tiene esta secuencia? Puede usar GEPARD (<http://cube.univie.ac.at/gepard>)

```
> SLIT_DROME (P24014)
MAAPSRTTLMPPPPFRLQLRLLILPILLLLRHDAVHAEPYSGGFGSSAVSSGGLGSVGIHIPGGGVG
VITEARCPRVCSTGLNVDCSHRGLTSVPRKISADVERLELQGNLTVIYETDFQRLTKLRMLQLT
DNQIHTIERNFSQDLVSLERLDISNNVITTVGRRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVEL
EILTLNNNNLTSLPHNIFGGLGRLRALRLSDNPFACDCHLSWLSRFLRSATRLAPYTRCQSPSQLK
GQNVADLHDQEFKCSGLTEHAPMECGAENSCPHPCRCADGIVDCREKSLTSVPVTLPPDDTTDVRLE
QNFITELPPKSFSSFRRLRRIDLNNNISRIAHDAISGLKQLTTLVLVYGNKIKDLPSGVFKGLGSL
RLLLLNANEISCIKDAFRDLHSLSLSLYDNNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRW
LADYLHKNPIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGT
TVDCTGRRLKEIPRDIPLHTTELLLNDELGRISDGLFGRPLHLVKLELKRNLGTGIEPNAFEGA
SHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNLASNPFCNCH
LAWFAECVRKKS LNNGAARCGAPSKVRDVQIKDLPHSEFKCSSENSEGLGDGYCPPSCTCTGTVV
ACSRNQLKEIPRGI PAETSELYLESNEIEQIHYERIRHLRSLTRLDSLNNQITILSNYTFANLTKL
STLIISYNKLQCLQRHALSGLNNLRVSVLHGNRISMLPEGSFEDLKS LTHIALGSNPLYCDCGLKW
FSDWIKLDYVEPGIARCAEPEQMMDKLILSTPSSSFVCRGRVRNDILAKCNACFEQPCQNQAQCVA
LPQREYQCLCQPGYHGKHCFMIDACYGNPCRN NATCTVLEEGRFSCQCAPGYTGARCETNIDDC
GEIKCQNNATCIDGVESYKCECQPGFSGEFCDTKIQFCSPEFNP CANGAKCMDHFTHYSCDCQAGF
HGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTS PCQNHECKH
GVC FQPN AQGSDYLCRCHPGYTGWCEYLT S ISFVHNNSFVELEPLRTRPEANVTIVFSSAEQNGI
LMYDGDQDAHLAVELFNGRIRVSYDVGNHPVSTMYSFEMVADGKYHAVELLA IKKNFTLRVDRGLAR
SI INEGSNDYKLTT PMFLGGLPVDPAQQAYKNWQIRNLTSFKGCMKEVWINHKLVD FGNAQRQOK
ITPGCALLEGEEQ QEEEDDEQDFMDETPHIKEEPVDPCLENKCRRGSRCPVNSNARDGYQCKCKHGQ
RGRYCDQGE GSTEPPTVTAASTCRKEQVREYYTENDCRSRQPLKYAKCVGGCGNQCCA AKIVRRRK
VRMVCSNNRKYIKNLDIVRKCGCTKKCY
```

1.5 Dadas las siguientes secuencias, realizar un dotplot. ¿Tienen algo en común? Realizar un alineamiento local de ambas secuencias.

```
>MS2_HUMAN (P78325)
MRGLGLWLLGAMMLPAIAPSRPWALMEQYEVVLPRLPGPRVRRALPSHLGLHPERSYVLGATGHNFTLHLRK
NRDLLGSGYTETETYTAANGSEVTEQPRGQDHCLYQGHVEGYPDSAASLSTCAGLRGFFQVGSDDLHLEPLDEGGE
GGRHAVYQAEHLLQTAGTCGVSDDSLGSLLGPRTAAVFRPRPGDSLPSRETRYVELYVVVDNAEFQMLGSEAAV
RHRVLEVNVHVDKLYQKLNFRVVLVGLIWN SQDRFHVSPDPSVTLENLLTWQARQRTRRHLHDNVQLITGVDF
TGTTVG FARVSAMCSHSSGAVNQDHSKNPVG VACTMAHEMGNLGMHDENVQGCRCQERFEAGRCIMAGSIGS
SFPRMFSDCSQAYLESFLERPQSVCLANAPDLSHLVGGPVCGNLFVERGEQCDGPPEDCRNRCCNSTTCQLAE
GAQCAHGTCCQECKVKPAGELCRPKKDMCDLEEFCDGRHPECPEDAFQENGTPCSGGYCYNGACPTLAQQCQAF
WGPGGQAAEESCF SYDILPGCKASRYRADMCGVLQCKGGQQPLGRAICIVDVCHALT TEDGTAYEPVPEGTRCG
PEKVCWKGRQCQDLHVYRSSNCSAQCHNHGVCNHNKQECHCHAGWAPPHCAKLLTEVHAASGSLPVLVVVVLVLLA
VVLVTLAGIIVYRKARSRI LSRNVAPKTTMGRSNPLFHQAASRVPAKGGA PAPS RGPQELVPTTHPGQPARHPA
SSVALKRPPPAPPVTVSSPPFPVPVYTRQAPKQVIKPTFAPPVPPVKPGAGAAPGPAEGAVGPKVALKPPIQR
KQGAGAPTAP
```

```
>ADAM_CROAD (P34179)
QQNL PQR YIELVVVADRRVFMKYNSDLNII RTRVHEIVNII NGFYRSLNIDVSLVNLEIWSGQDPLTIQSSSN
TLNSEGLWREKVLNKKKKDNAQLLTAIEFKCETLGKAYLNSMCNPRSSVGIVKDHSPINLLVAVTMAHELGHN
LGMEHDGKDCLRGASLCIMRPLTPGRSYEFSDDSMGYYQKFLNQYKQPQCILNKP
```

## 2. (Realizar Alineamientos)

2.1 Usando los parámetros por defecto para match (1), mismatch (-1) y gap (-2), calcule el score para cada uno de los siguientes alineamientos:

GG-A-TC	GGA--TC
GGAAATC	GGAAATC

¿Tiene sentido biológico el resultado obtenido?

Pruebe con un sistema de score no lineal (affine gap = open gap penalty + extend gap penalty \* n), en donde match (1), mismatch (-1), open gap penalty (-2) y extend gap penalty (-1).

### 2.2

Busque en alguna base de datos biológicas dos secuencias de proteínas que tengan la misma función, p.ej homologas a la proteína L36. P. Ej podría usar la base de datos **Protein** del sitio de **NCBI** para buscar 3 homologos. p. Ej (en los organismos Homo Sapiens, mus musculus, y otro organismo que usted elija)

Realice 2 alineamientos globales (Homo Sapiens - Mus Musculus y Homo Sapiens – el organismo que Usted escogió) utilizando el formato FASTA de cada una de las secuencias de las proteínas. Para esto puede utilizar el algoritmo BLASTP de NCBI

Grabe los resultados obtenidos y las secuencias en formato FASTA utilizadas  
¿Tienen sentido los resultados obtenidos?

## 3. (Codificar)

1. Modificar el programa que hizo para tomar las secuencias biológicas desde un archivo.
2. Modificar el programa que hizo para poder escribir el dotplot a un archivo de nombre "salida.txt".

**Cree un archivo con el nombre <Nombre-Apellido-Lab-Bio-02.zip> y subalo al sitio web del curso.**