

1. ALINEAMIENTO DE SECUENCIAS CON DOTPLOT

1.1 DotPlot correspondiente a la comparación de estas secuencias:

A = ACTGG

B = ACTAA

```
#Strings to compare
s1 <- unlist(strsplit(c('ACTGG'), ''))
s2 <- unlist(strsplit(c('ACTAA'), ''))

#To set up the matrix.
n1 <- length(s1)
n2 <- length(s2)

#Finding matches
a <- ifelse(s1=='A', 'X', '-')
c <- ifelse(s1=='C', 'X', '-')
g <- ifelse(s1=='G', 'X', '-')
t <- ifelse(s1=='T', 'X', '-')

x <- matrix (data=NA, nrow=n1, ncol=n2)

#s1 vs s2, filling dotplot.
colnames(x) <- s2
rownames(x) <- s1
for(i in 1:n2){
  if(s2[i]=='A'){
    x[,i] <- a
  } else if(s2[i]=='C'){
    x[,i] <- c
  } else if(s2[i]=='G'){
    x[,i] <- g
  } else {
    x[,i] <- t
  }
}
x
```

	A	C	T	A	A
A	X	-	-	X	X
C	-	X	-	-	-
T	-	-	X	-	-
G	-	-	-	-	-
G	-	-	-	-	-

1.2 Usando la secuencia **ACTGACTGACTG** que tiene **repeticiones tandem perfectas** y analizaremos su respectivo dotplot.

	A	C	T	G	A	C	T	G	A	C	T	G
A	X	-	-	-	X	-	-	-	X	-	-	-
C	-	X	-	-	-	X	-	-	-	X	-	-
T	-	-	X	-	-	-	X	-	-	-	X	-
G	-	-	-	X	-	-	-	X	-	-	-	X
A	X	-	-	-	X	-	-	-	X	-	-	-
C	-	X	-	-	-	X	-	-	-	X	-	-
T	-	-	X	-	-	-	X	-	-	-	X	-
G	-	-	-	X	-	-	-	X	-	-	-	X
A	X	-	-	-	X	-	-	-	X	-	-	-
C	-	X	-	-	-	X	-	-	-	X	-	-
T	-	-	X	-	-	-	X	-	-	-	X	-
G	-	-	-	X	-	-	-	X	-	-	-	X

Se muestra varias diagonales pequeñas paralelas a la principal que son causadas por las repeticiones directas.

1.3 DotPlot correspondiente a la comparación de estas secuencias:

A = ACTGGTCAACTGGTCA

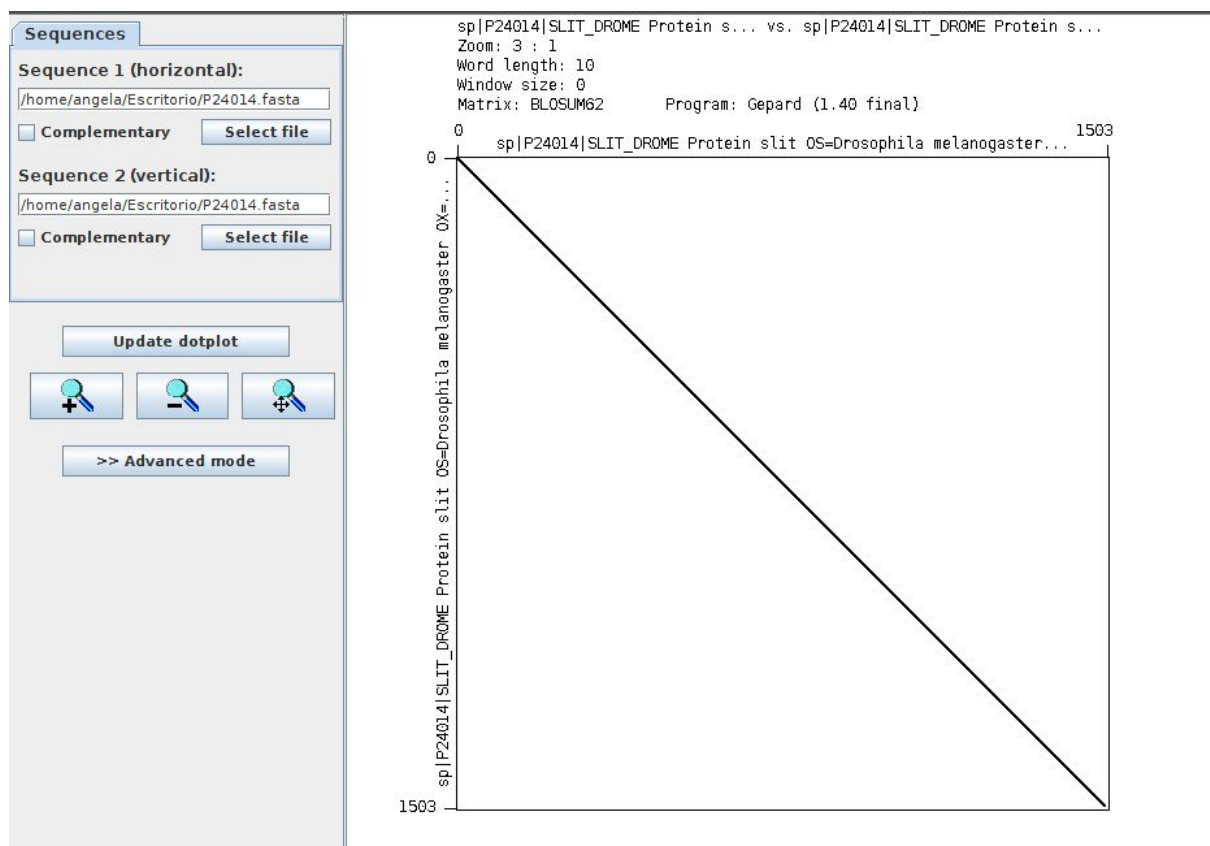
B = ACTAACTGGTCAATCA

	A	C	T	A	A	C	T	G	G	T	C	A	A	T	C	A
A	X	-	-	X	X	-	-	-	-	-	-	X	X	-	-	X
C	-	X	-	-	-	X	-	-	-	-	X	-	-	-	X	-
T	-	-	X	-	-	-	X	-	-	X	-	-	-	X	-	-
G	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-
T	-	-	X	-	-	-	X	-	-	X	-	-	-	X	-	-
C	-	X	-	-	-	X	-	-	-	-	X	-	-	-	X	-
A	X	-	-	X	X	-	-	-	-	-	-	X	X	-	-	X
A	X	-	-	X	X	-	-	-	-	-	-	X	X	-	-	X
C	-	X	-	-	-	X	-	-	-	-	X	-	-	-	X	-
T	-	-	X	-	-	-	X	-	-	X	-	-	-	X	-	-
G	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-
T	-	-	X	-	-	-	X	-	-	X	-	-	-	X	-	-
C	-	X	-	-	-	X	-	-	-	-	X	-	-	-	X	-
A	X	-	-	X	X	-	-	-	-	-	-	X	X	-	-	X

El alineamiento correspondiente usando gaps en las dos secuencias sería:

A	C	T	_	_	_	_	G	G	T	C	A	A	C	T	G	G	T	C	A
A	C	T	A	A	C	T	G	G	T	C	A	A	_	_	_	_	T	C	A

1.4 DotPlot realizado con Gepard a partir de la secuencia **P24014 SLIT_DROME**



1.5 DotPlot comparando las siguientes secuencias **P78325 MS2_HUMAN** y **P34179 ADAM_CROAD**

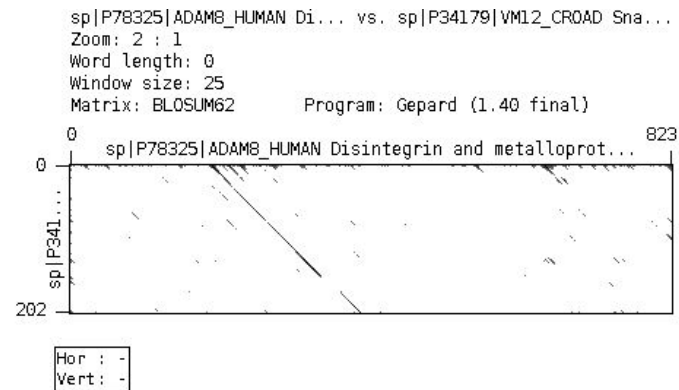
Sequences

Sequence 1 (horizontal):

☐ Complementary

Sequence 2 (vertical):

☐ Complementary



2. REALIZAR ALINEAMIENTOS

2.1 Con el **score** de MATCH(+1), MISMATCH(-1) Y GAP(-2) calcular score:

G	G	-	A	-	T	C
G	G	A	A	A	T	C

matches = 4

mismatches = 0

gaps = 2

$$\text{SCORE: } 4*(+1) + 0*(-1) + 2*(-2) = 0$$

G	G	A	-	-	T	C
G	G	A	A	A	T	C

matches = 5

mismatches = 0

gaps = 2

$$\text{SCORE: } 5*(+1) + 0*(-1) + 2*(-2) = 1$$

2.2 Usando el algoritmo BLAST de NCBI con las siguientes secuencias de la proteína L36 en las especies HomoSapiens, Mus Musculus y Chlamydia

L36 HOMOSAPIENS VS L36 MUS MUSCULUS

Blast 2 sequences

Job title: NP_378669.1 60S ribosomal protein L36 [Homo

RID9NM2N7SF114 (Expires on 03-28 09:22 am)

Query IDIcl|Query_10952

DescriptionNP_378669.1 60S ribosomal protein L36 [Homo sapiens]

Molecule typeamino acid

Query Length105

Subject ID2 subjects

DescriptionSee details

Molecule typeamino acid

Subject Length224

ProgramBLASTP 2.9.0+ Citation

Other reports: Search Summary Distance tree of results Multiple alignment MSA viewer

Graphic Summary

Distribution of the top 2 Blast Hits on 2 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

<40

40-50

50-80

80-200

>=200

Query

120406080100

Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Perc. ident	Accession
<input type="checkbox"/>	NP_061200.2 60S ribosomal protein L36 [Mus musculus]	209	209	100%	6e-77	100.00%	Query_10954
<input type="checkbox"/>	XP_006524693.1 PREDICTED: 60S ribosomal protein L36 isoform X1 [Mus musculus]	147	147	70%	6e-52	100.00%	Query_10955

Download Graphics

Ne

NP_061200.2 60S ribosomal protein L36 [Mus musculus]

Sequence ID: Query_10954 Length: 105 Number of Matches: 1

Range 1: 1 to 105 Graphics

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
209 bits(532)	6e-77	Compositional matrix adjust.	105/105(100%)	105/105(100%)	0/105(0%)

Query 1MALRYPMAVGLNKGHKVTKNVSKPRHSRRRGRLLKHTKFVRDMIREVCGFAPYERRAMEL60

Sbjct 1MALRYPMAVGLNKGHKVTKNVSKPRHSRRRGRLLKHTKFVRDMIREVCGFAPYERRAMEL60

Query 61LKVSKDKRALKFIKKRVGTHIRAKRKREELSNVLAAMRKAANKD105

Sbjct 61LKVSKDKRALKFIKKRVGTHIRAKRKREELSNVLAAMRKAANKD105

Download Graphics

Ne

XP_006524693.1 PREDICTED: 60S ribosomal protein L36 isoform X1 [Mus musculus]

Sequence ID: Query_10955 Length: 119 Number of Matches: 1

Range 1: 46 to 119 Graphics

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
147 bits(370)	6e-52	Compositional matrix adjust.	74/74(100%)	74/74(100%)	0/74(0%)

Query 32RLTKHTKFVRDMIREVCGFAPYERRAMELLKVSKDKRALKFIKKRVGTHIRAKRKREELS91

Sbjct 46RLTKHTKFVRDMIREVCGFAPYERRAMELLKVSKDKRALKFIKKRVGTHIRAKRKREELS105

Query 92NVLAAMRKAANKD105

Sbjct 106NVLAAMRKAANKD119

L36 HOMOSAPIENS VS L36 CHLAMYDIA

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [▶ Formatting options](#) [▶ Download](#) [You Tube](#) [How to read this page](#) [Blast report description](#)

Blast 2 sequences

Job title: NP_378669.1 60S ribosomal protein L36 [Homo]

RID	9NM6XCT9114 (Expires on 03-28 09:24 am)	Subject ID	Ic Query_18143
Query ID	Ic Query_18141	Description	AAA16204.1 L36 [Chlamydia muridarum]
Description	NP_378669.1 60S ribosomal protein L36 [Homo sapiens]	Molecule type	amino acid
Molecule type	amino acid	Subject Length	37
Query Length	105	Program	BLASTP 2.9.0+ ▶ Citation

❗ No significant similarity found. For reasons why, [click here](#)

Other reports: [▶ Search Summary](#)


L36 MUS MUSCULUS VS L36 CHLAMYDIA

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [▶ Formatting options](#) [▶ Download](#) [You Tube](#) [How to read this page](#) [Blast report description](#)

Blast 2 sequences

Job title: NP_061200.2 60S ribosomal protein L36 [Mus]

Results for: *1:Ic|Query_80620 NP_061200.2 60S ribosomal protein L36 [Mus musculus](105aa) ▼ 

RID	9NM8NPXC114 (Expires on 03-28 09:25 am)	Subject ID	Ic Query_80623
Query ID	Ic Query_80620	Description	AAA16204.1 L36 [Chlamydia muridarum]
Description	NP_061200.2 60S ribosomal protein L36 [Mus musculus]	Molecule type	amino acid
Molecule type	amino acid	Subject Length	37
Query Length	105	Program	BLASTP 2.9.0+ ▶ Citation

❗ No significant similarity found. For reasons why, [click here](#)

Other reports: [▶ Search Summary](#)