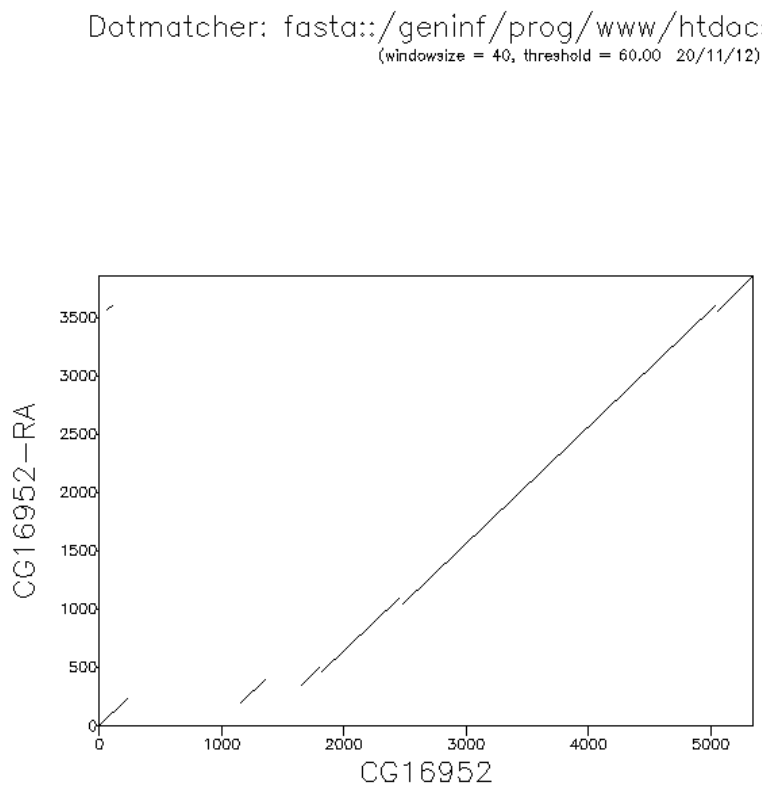
Bioinformatique des Séquences
EM7BBSAM

TD Alignement de séquences

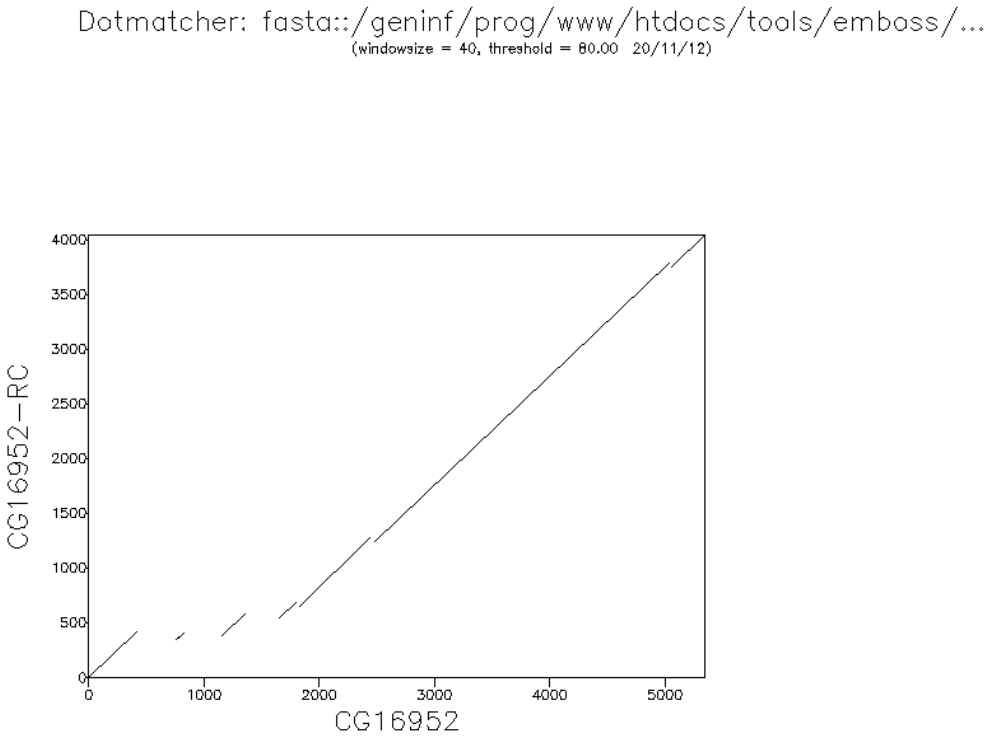
Exercice 5 : Comparer différents ARNm d'un même gène

Nous allons maintenant tenter d'identifier des phénomènes de transcrits alternatifs.
Récupérer la séquence génomique CG16952 et deux transcrits CG16952-RA et CG16952-RC d'un gène de *Drosophila melanogaster*.

a. Dotplot.
Comparer la séquence génomique du gène à ses deux transcrits.
- Comment apparaissent les exons dans le dotplot ?
- Combien y-a-t-il d'exons pour chaque ARNm ?
- Y-a-t-il une différence entre les deux ARNm ?
- Que peut-on dire sur la taille relative des exons et des introns pour ce gène ?



6 exons



7 exons

En 5' les introns sont plus longs qu'en 3'.

En 3' de très grands exons et petits introns

1 exon en plus sur RC vers 400-500 sur transcrit et 800 sur génomique

1er exon + long sur RC

b. Alignement global.

Utiliser le programme stretcher pour comparer les deux transcrits.
Quel est le score de l'alignement ?
Pouvez-vous identifier une variation d'épissage à partir de l'alignement ?
Pouvez-vous la décrire ?

Stretcher => 1 morceau en + sur RC entre 215 et 402

Score = 18506 (param par défaut)

Comparer maintenant la séquence génomique du géne avec le transcrit CG16952-RC.
Quel est le score de l'alignement ?
Correspond-t-il à ce qui était attendu ?

Pouvez-vous identifier le méme nombre d'exons qu'à partir du dotplot ?

Score (stretcher, param par défaut) = 14918

Score +petit qu'entre les transcrits car + d'indels

On ne voit que 6 exons ; on ne voit pas le tout petit vers 400 (RC) et 800 sur génomique

POURQUOI ??

Si on fait un alignement local avec matcher (7 matches alternatifs), GC=100 ; GE=50

	1	2	3	4	5	6	7
Gène	1-400	770-820	1180-	1670-	1850-	2510-	5080-

			1340	1790	2430	5020	5350
Transcrit RC	1-400	360-400	400-565	560-680	680-1260	1260-3770	3770-4045

En fait le 2 n'est pas un exon : c'est la fin de l'exon1 qui s'aligne aussi sur l'intron 1

```
Identity:       34/45 (75.6%)

             780        790        800        810
  gene TGTGTACATAAGAGTATATATGTGTGTATGTATTTTGTGGATTTA
       ::: ::::::    :::::::   : ::::::::: :::   ::: ::
    RC TGTATACATATTGGTATATACATATGTATGTATGTTGGAGATATA
        360        370        380        390        400
```

Dans le dotplot si on met T=80 et W=20, on ne le voit plus

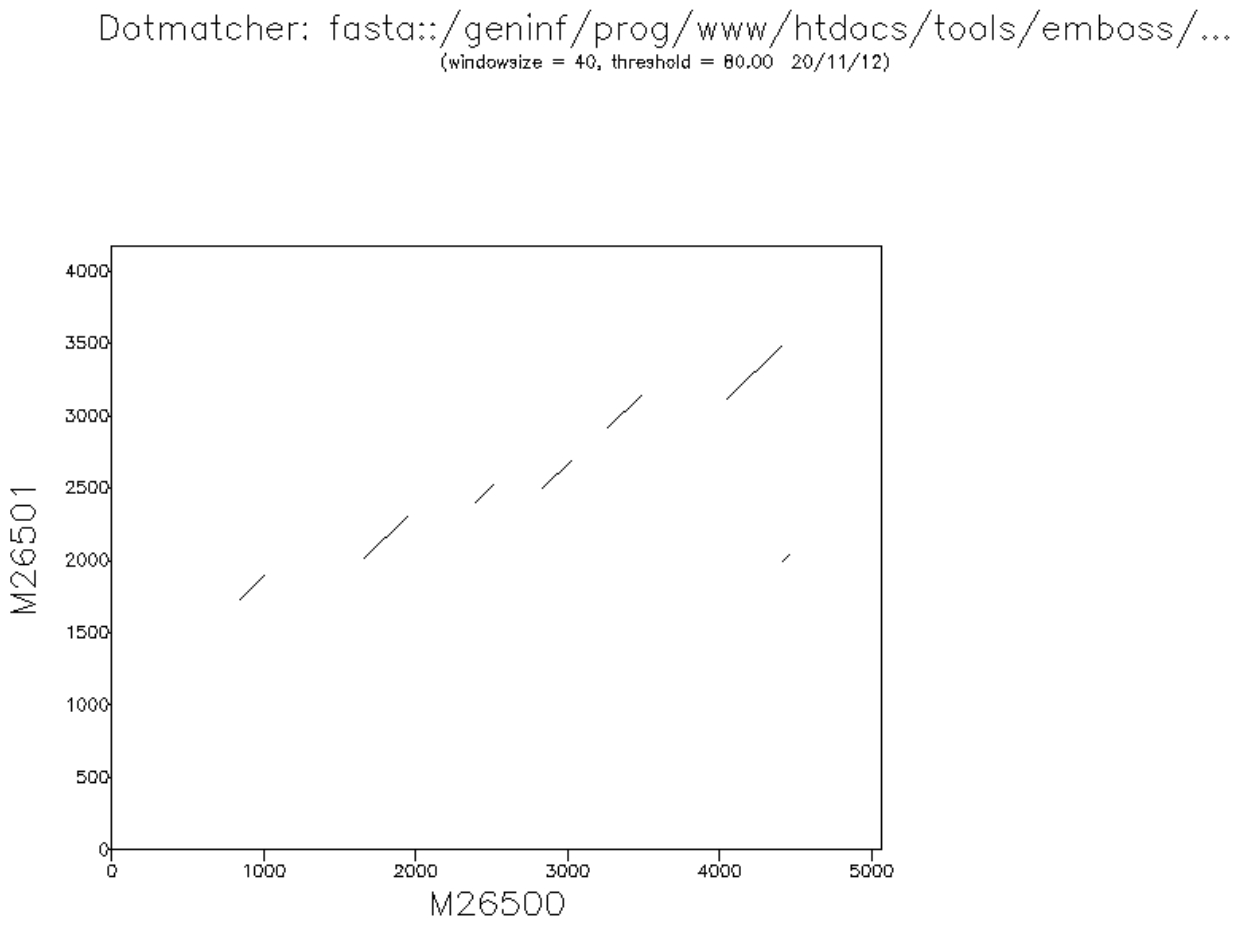
Concl : attention aux interprétations trop rapides !

Exercice 6 : Comparaison entre 2 gènes

1. Faire un DotPlot des séquences de M26500 et M26501
   Combien voyez-vous de diagonales ? Combien d'exons sur chaque gène ?
2. Faire alignement local et global sur ces séquences.
3. Aligner la séquences M26500 et son CDS avec needle.
4. Comparer la CDS de M26500 avec la séquence AK069426.
5. Chercher l'ORF de AK069426 et la comparer avec la protéine codée par M26500.
   Comparer les résultats de l'alignement local et global
   Changer la matrice de score

Starfish (P.ochraceus) muscle actin gene, complete cds
5,070 bp linear DNA
Accession: **M26500**.1

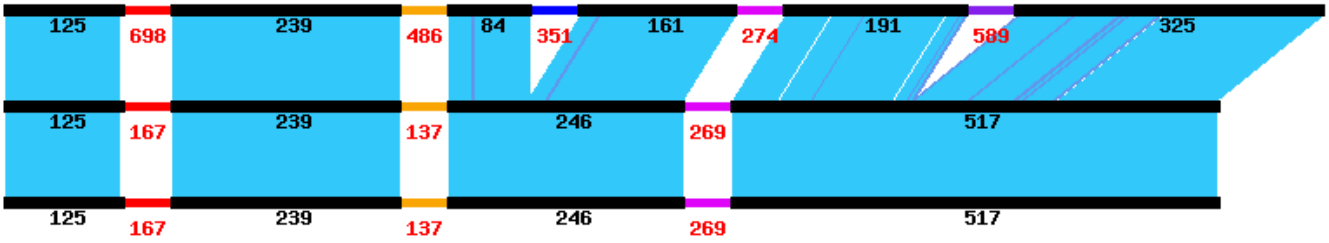
Starfish (P.ochraceus) cytoplasmic actin gene, complete cds
4,172 bp linear DNA
Accession: **M26501**.1



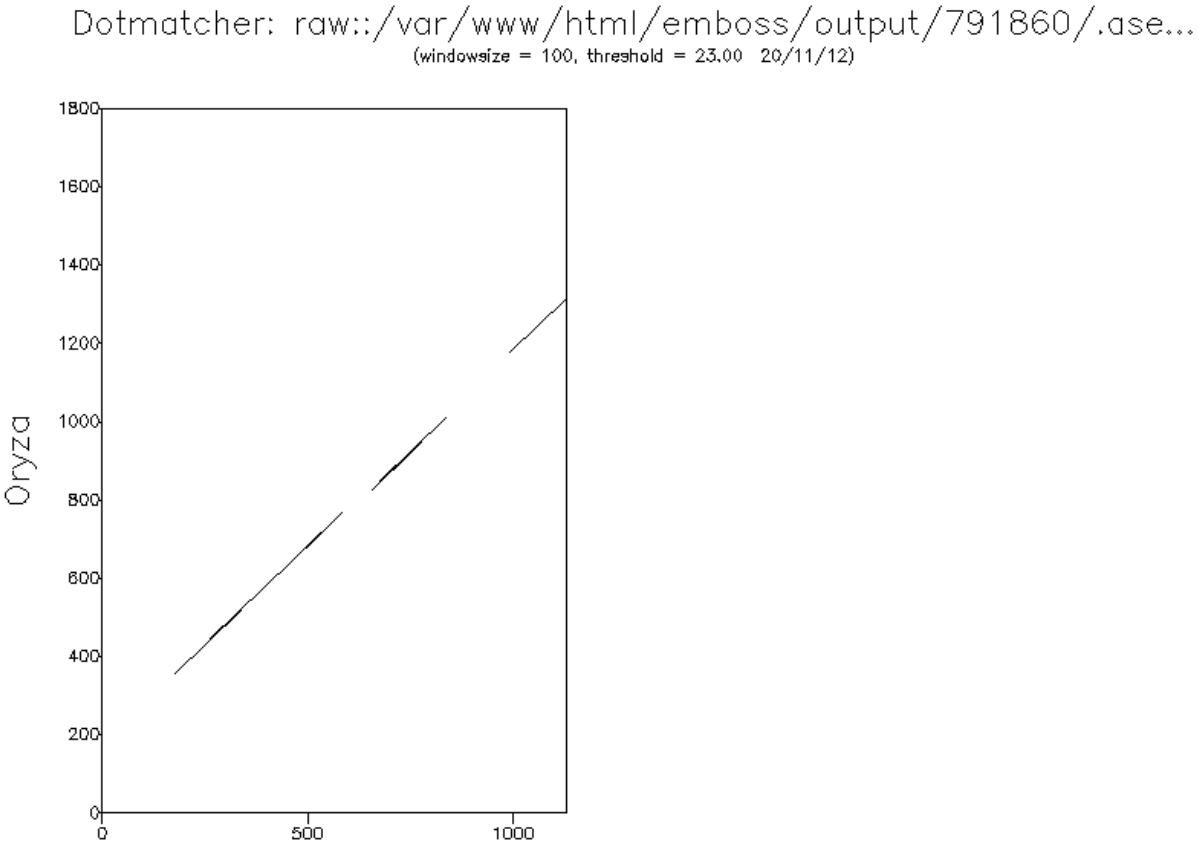
6 diagonales sur chaque gène => 6 exons sur M26500, sans doute 4 sur M26501 (3+4) et (4+5) sont joints. D'après l'annotation seuls 4 exons

Alignement local avec Water et GC=60

M26500	866..991	1689..1928	2414..2498	2849..3010	3284..3475	4064..4389
M26501	1756..1881	2048..2287	2424..2670		2939..3456	



4. Comparer la CDS de M26500 avec la séquence AK069426.



Alignement avec Water assez bon 45% id avec GC=30

5. Chercher l'ORF de AK069426 et la comparer avec la protéine codée par M26500.

```
>AK069426 ORF:234..1316 Frame +3
MEAVVVDAGSKLLKAGIALPDQSPSLVMPSKMKLEVEDGQMGDGAVVEEVVQPVVRGFVKDWDAMEDLLN
YVLYSNIGWEIGDEGQILFTEPLFTPKALREQLAQLMFEKFNVSGFYDSEQAVLSLYAVGRISGCTVDIG
HGKIDIAPVCEGAVQHIASKRFDIGGTDLTNLFAEELKKSNSSVNIDISDVERLKEQYACCAEDQMAFEA
IGSSCRPERHTLPDGQVITIEKERYIVGEALFQPHILGLEDYGIVHQLVTSVSNVTPEYHRQLLENTMLC
GGTASMTGFEDRFQREANLSASAIYPSLVKPPEYMPENLARYSAWLGGAILAKVVFPQNQHVTKGDYDET
GPSIVHKKCF*
```

Comparer les résultats de l'alignement local et global

Quasi pareil

Needle

```
Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 388
# Identity:     140/388 (36.1%)
# Similarity:   197/388 (50.8%)
# Gaps:          40/388 (10.3%)
# Score: 547.0
```

Water

```
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 381
# Identity:     140/381 (36.7%)
# Similarity:   196/381 (51.4%)
# Gaps:          35/381 ( 9.2%)
# Score: 547.0
```

Changer la matrice de score

Water

```
# Matrix: EBLOSUM30
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 388
# Identity:     146/388 (37.6%)
# Similarity:   221/388 (57.0%)
# Gaps:          49/388 (12.6%)
```