

Introducción a la Bioinformática

Práctica 2: Alineamiento múltiple e Identificación y búsqueda de Motivos.

El alineamiento múltiple es una de las técnicas bioinformáticas más usadas, ya que por medio de ella podemos realizar diversos análisis, que van desde la filogenia hasta la búsqueda de motivos.

En esta oportunidad estudiaremos diferentes algoritmos para alineamiento múltiple, y la utilización de esta técnica en la búsqueda de motivos.

Alineamiento múltiple

El primer paso para realizar un alineamiento múltiple es contar con las secuencias necesarias para llevar a cabo el análisis. En este caso utilizaremos 5 secuencias de lipocalinas distantemente relacionadas.

Obtenga las siguientes secuencias de proteínas, puede ser en un archivo FASTA múltiple o de manera separada:

NP_006735.1 (Retinol Binding Protein 4, human)

BAB25881.1 (Retinol Binding Protein 4, mouse)

NP_001638.1 (human apolipoprotein D)

MUP4_MOUSE (Major urinary protein 4, mouse)

gi|732003 (E. Coli outer membrane lipocalin)

Si tienen dificultad en la recuperación de estas secuencias, consulte el taller número 1: BLAST y recuperación de secuencias.

Inicialmente vamos a realizar un alineamiento múltiple utilizando la implementación de **CLUSTALW** realizada por el **EBI** (European Institute of Bioinformatics).

Visite el sitio web de CLUSTALW en el EBI, en la siguiente dirección:

<http://www.ebi.ac.uk/clustalw/index.html>

Después de unos segundos aparecerá el formulario de envío de secuencias de

CLUSTALW:

ClustalW
Submission Form

ClustalW is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms. [New users, please read the FAQ.](#)

>> [Download Software](#)

YOUR EMAIL	ALIGNMENT TITLE	RESULTS	ALIGNMENT	CPU MODE
<input type="text"/>	<input type="text" value="Sequence"/>	<input type="text" value="interactive"/>	<input type="text" value="full"/>	<input type="text" value="single"/>
KTUP (WORD SIZE)	WINDOW LENGTH	SCORE TYPE	TOPDIAG	PAIRGAP
<input type="text" value="def"/>	<input type="text" value="def"/>	<input type="text" value="percent"/>	<input type="text" value="def"/>	<input type="text" value="def"/>
MATRIX	GAP OPEN	END GAPS	GAP EXTENSION	GAP DISTANCES
<input type="text" value="def"/>	<input type="text" value="def"/>	<input type="text" value="def"/>	<input type="text" value="def"/>	<input type="text" value="def"/>

OUTPUT		PHYLOGENETIC TREE		
OUTPUT FORMAT	OUTPUT ORDER	TREE TYPE	CORRECT DIST.	IGNORE GAPS
<input type="text" value="aln w/numbers"/>	<input type="text" value="aligned"/>	<input type="text" value="none"/>	<input type="text" value="off"/>	<input type="text" value="off"/>

Enter or Paste a Set of Sequences in any supported format: Help

```

>gi|5803139|ref|NP_006735.1| RBP4 gene product
[Homo sapiens]
MEWVWALLLLAANAARDCRVSSFRVKEINFDKARFSGTWYMAKGD
ATAKGRVRLLIHNDVCA DMVGTFTDTEPAKFKMKYGVASF LQKGI
LDGTCADSYSFVSRDPNGLPPEAQKIVRQEQEELCLARQYRLIVHI
>gi|12843160|dbj|BAB25881.1| unnamed protein
product [Mus musculus]
MEWVWALVLLAALGGGSAERDCRVSSFRVKEINFDKARFSGLWYALAK
MSATAKGRVRLLSHWEVCA DMVGTFTDTEPAKFKMKYGVASF LQR
QNLGTCADSYSFVSRDPNGLSPETRRIVRQEQEELCLERQYRWLE
                
```

Upload a file:

Introduzca las secuencias descargadas anteriormente. Revise cuidadosamente cada uno de los parámetros que ofrece CLUSTALW. ¿Esta seguro-a de conocer el propósito de cada uno de ellos? En esta oportunidad correremos el programa con los parámetros por defecto.

Presione el botón **“Run”**, aparecerá una ventana similar a esta:

Your job is currently running...
...please be patient

The results of your job will appear in this browser window.

Your job output:
<http://www.ebi.ac.uk/cgi-bin/clustalw/result?tool=clustalw&jobid=clustalw-20060706-17144376&poll=yes>

Please Note the Following:

- You may press Shift+Refresh or Reload on your browser at any time to check if results are ready. Should this window go blank please press the Shift+Refresh or Reload button on your browser.
- You may bookmark this page to view your results later if you wish.
Netscape users: Use Bookmark - Add Bookmark or CTRL-D | Alt-K to bookmark this page.
IE users: Click -> [BookMark](#) to bookmark this page.
- Results are stored for 24 hours. Some big files will be deleted after ca. 15 minutes.

Básicamente, debemos ser pacientes y esperar por los resultados. El tiempo que le toma al programa obtener los resultados depende de la carga del servidor en ese momento, así que el tiempo de espera debe ser directamente proporcional al número de personas que estén haciendo uso de este sistema al mismo tiempo.

Después de unos segundos o (minutos) aparecerá la página de resultados de CLUSTALW, la cual se encuentra dividida en 4 secciones.

1. Resumen de resultados.

Results of search	
Number of sequences	5
Alignment score	1459
Sequence format	Pearson
Sequence type	aa
ClustalW version	1.83
JaView	<input type="button" value="Start JaView"/>
Output file	clustalw-20060706-17144376.output
Alignment file	clustalw-20060706-17144376.aln
Guide tree file	clustalw-20060706-17144376.dnd
Your input file	clustalw-20060706-17144376.input
<input type="button" value="SUBMIT ANOTHER JOB"/>	

2. Tabla de puntajes, esta tabla muestra los puntajes de los alineamientos pareados realizados por el algoritmo.

Scores Table

Seqs. Name	Len(aa)	Seqs. Name	Len(aa)	Score
1 gi 5803139 ref NP_006735.1	199	2 gi 12843160 dbj BAE25881.1	201	84
1 gi 5803139 ref NP_006735.1	199	3 gi 4502163 ref NP_001638.1	189	14
1 gi 5803139 ref NP_006735.1	199	4 gi 127528 sp P11590 HMP4_MOUSE	178	8
1 gi 5803139 ref NP_006735.1	199	5 gi 732003 sp P39281 BLC_ECOLI	177	12
2 gi 12843160 dbj BAE25881.1	201	3 gi 4502163 ref NP_001638.1	189	17
2 gi 12843160 dbj BAE25881.1	201	4 gi 127528 sp P11590 HMP4_MOUSE	178	5
2 gi 12843160 dbj BAE25881.1	201	5 gi 732003 sp P39281 BLC_ECOLI	177	19
3 gi 4502163 ref NP_001638.1	189	4 gi 127528 sp P11590 HMP4_MOUSE	178	5
3 gi 4502163 ref NP_001638.1	189	5 gi 732003 sp P39281 BLC_ECOLI	177	27
4 gi 127528 sp P11590 HMP4_MOUSE	178	5 gi 732003 sp P39281 BLC_ECOLI	177	7

3. Alineamiento, este es el alineamiento múltiple en sí.

Alignment

Seqs. Name	Len(aa)	Seqs. Name	Len(aa)	Score
gi 5803139 ref NP_006735.1	199	gi 12843160 dbj BAE25881.1	201	84
gi 5803139 ref NP_006735.1	199	gi 4502163 ref NP_001638.1	189	14
gi 5803139 ref NP_006735.1	199	gi 127528 sp P11590 HMP4_MOUSE	178	8
gi 5803139 ref NP_006735.1	199	gi 732003 sp P39281 BLC_ECOLI	177	12
gi 12843160 dbj BAE25881.1	201	gi 4502163 ref NP_001638.1	189	17
gi 12843160 dbj BAE25881.1	201	gi 127528 sp P11590 HMP4_MOUSE	178	5
gi 12843160 dbj BAE25881.1	201	gi 732003 sp P39281 BLC_ECOLI	177	19
gi 4502163 ref NP_001638.1	189	gi 127528 sp P11590 HMP4_MOUSE	178	5
gi 4502163 ref NP_001638.1	189	gi 732003 sp P39281 BLC_ECOLI	177	27
gi 127528 sp P11590 HMP4_MOUSE	178	gi 732003 sp P39281 BLC_ECOLI	177	7

4. Árbol guía, árbol creado por CLUSTAL como guía para realizar el proceso de alineamiento múltiple. Este árbol no debe considerarse un árbol filogenético.

Guide Tree

Seqs. Name	Len(aa)	Seqs. Name	Len(aa)	Score
gi 5803139 ref NP_006735.1	199	gi 12843160 dbj BAE25881.1	201	84
gi 5803139 ref NP_006735.1	199	gi 4502163 ref NP_001638.1	189	14
gi 5803139 ref NP_006735.1	199	gi 127528 sp P11590 HMP4_MOUSE	178	8
gi 5803139 ref NP_006735.1	199	gi 732003 sp P39281 BLC_ECOLI	177	12

Cladogram



Estrategia de CLUSTALW para el alineamiento múltiple (AM).



CLUSTALW lleva a cabo una estrategia denominada “progresiva” para realizar el AM. Básicamente podemos dividir esta estrategia en 3 pasos.

Primer paso: Alineamientos Globales pareados

CLUSTALW realiza un alineamiento global pareado (mediante el algoritmo Needleman-Wunsch) de cada una de las secuencias incluidas en el alineamiento, de tal manera que para un alineamiento de **5** secuencias, el número de alineamientos a realizar será de **10**.

Segundo paso: Cálculo del árbol guía.

Se caála un árbol guía a partir del puntaje (o distancia) de los alineamiento pareados realizados en el primer paso. Este árbol puede ser calculado ya sea por cualquiera de los siguientes métodos: UPGMA o neighbor-joining.

El árbol describe entonces la relación existente entre las secuencias que van a ser alineadas. No se debe considerar este como un árbol filogenético, sino como guía para ir añadiendo las secuencias al alineamiento múltiple.

Tercer paso: Creación del alineamiento múltiple.

El alineamiento de las secuencias se lleva a cabo en el orden determinado por el árbol guía. El algoritmo selecciona primero las dos secuencias más relacionadas y crea un alineamiento pareado de estas, y de manera progresiva va sumando una secuencia al resultado de dicho alineamiento.

Compare los resultados de la tabla de puntajes con el árbol guía. ¿Cuáles son las dos secuencias con las que se inicia el alineamiento múltiple y cuál es la última que es utilizada en este?

Una herramienta muy interesante que implementa el EBI es **JalView**, una aplicación que nos permite visualizar nuestros alineamientos y realizar operaciones sobre estos. Al ser JalView una aplicación desarrollada en el lenguaje de programación JAVA, es necesario contar con el JRE de Sun instalado en nuestros equipos para que este programa pueda funcionar¹.

En esta práctica usaremos no JalView sino que utilizaremos **BioEdit** como herramienta de visualización y análisis . Si requiere más información acerca de JalView consulte su sitio web:

<http://www.jalview.org>

¹ El JRE se puede obtener gratuitamente del sitio web de SUN: <http://www.sun.com>

Guarde en su computador los archivos disponibles en el cuadro de resultados, especialmente aquellos con extensión “.aln” y “.dnd”. Estos archivos serán los que utilizemos en BioEdit.

Results of search	
Number of sequences	5
Alignment score	1459
Sequence format	Pearson
Sequence type	aa
ClustalW version	1.83
JaView	<input type="button" value="Start JaView"/>
Output file	c:\ustalw-20060706-17144376.output
Alignment file	c:\ustalw-20060706-17144376.aln
Guide tree file	c:\ustalw-20060706-17144376.dnd
Your input file	c:\ustalw-20060706-17144376.input
<input type="button" value="SUBMIT ANOTHER JOB"/>	

Bioedit es un programa gratuito para edición de alineamientos y análisis de secuencias que funciona únicamente sobre ambiente MS/Windows. Es, sin lugar a duda, uno de los programas más conocidos para edición de secuencias para dicho sistema operativo.

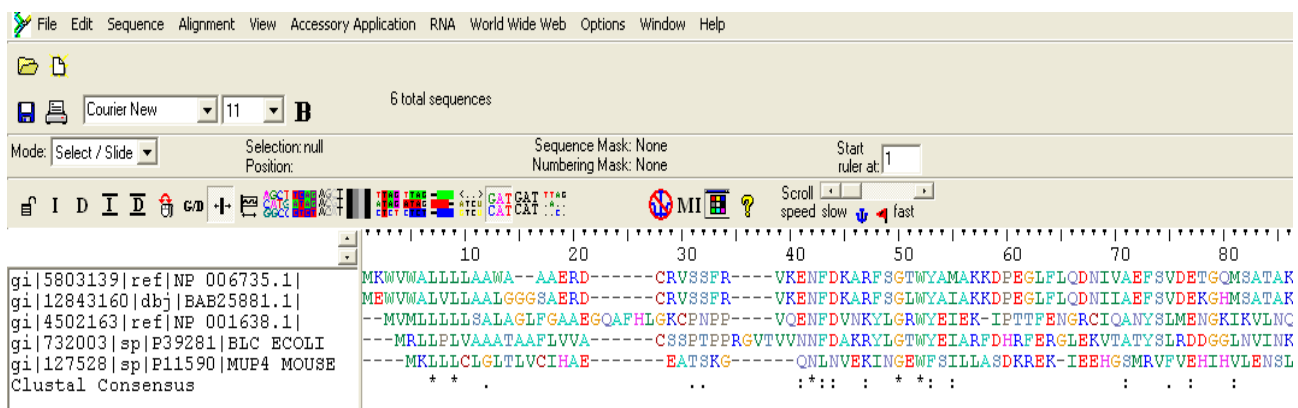
BioEdit cuenta con varias herramientas que van desde la creación de alineamientos hasta la anotación de plásmidos. En esta oportunidad utilizaremos este programa para manipular parcialmente nuestro alineamiento recién creado.

Si tiene interés en conocer a profundidad BioEdit y todas sus posibilidades puede consultar su sitio web: <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

Abra con BioEdit el archivo “.aln”, que ha guardado en su computador.

BioEdit, así como todas las aplicaciones para MS/Windows tienen un comportamiento similar, por lo tanto para abrir un archivo debe dirigirse al menú **Archivo** (file) y escoger la opción **Abrir** (open).

Inmediatamente aparecerá una ventana similar a la siguiente:



Para nuestro siguiente ejemplo utilizaremos de nuevo 5 secuencias de aminoácidos, pero esta vez muy relacionadas. De esta manera podremos evaluar y conocer la forma en que se ve un alineamiento de este tipo.

Recupere las secuencias con estos identificadores en formato FASTA:

P18902 A39486 NP_006735.1 Q00724 P04916

T-COFFEE

CLUSTALW se ha convertido con el paso del tiempo en el programa por defecto para alineamiento múltiple de secuencias. Sin embargo, desde hace ya varios años existen nuevos programas que han probado ser de alguna manera más eficientes y eficaces que éste. Entre estos programas podemos contar con **MUSCLE** y **T-COFFEE**. De hecho, los autores de MUSCLE aseguran (algunas publicaciones parecen confirmarlo) que dicho programa logra mejores alineamientos que los realizados con T-COFFEE o CLUSTALW.

En esta oportunidad trabajaremos con T-COFFEE, especialmente porque posee una característica que aunque no sea utilizada esta vez le hace muy interesante, y es el hecho de que nos permite combinar alineamientos hechos con diferentes métodos. Así por ejemplo, si tenemos un alineamiento hecho con CLUSTALW y otro con MUSCLE, T-COFFEE es capaz de combinar estos dos alineamientos y generar uno nuevo.

Ingrese las secuencias que ha recuperado en el formulario de envío de secuencias de T-COFFEE, disponible en la siguiente dirección:

<http://www.ebi.ac.uk/t-coffee/>

Presione el botón **Run** y espere un momento.

EMAIL	RESULTS	RUN NAME	MATRIX
<input type="text"/>	interactive ▾	Sequence	none ▾

Enter or Paste a Set of Sequences in any supported format:		Help
<div style="border: 1px solid black; height: 100px;"></div>		
Upload a file:	<input type="text"/> Examinar...	Run Reset

El comportamiento de esta aplicación es similar al de CLUSTALW, es decir, gracias a que estamos realizando nuestros análisis desde un mismo sitio, el EBI, las interfaces gráficas de los programas son consistentes y nuestra manera de interactuar con ellos muy similar.

alineamiento de la manera que lo hacemos por ejemplo con el valor E en BLAST. Sin embargo existen algunas “reglas” que pueden ser de ayuda al momento de interpretarlo, por ejemplo, generalmente podemos interpretar un bloque de alrededor de 30 residuos, con dos o tres identidades y alrededor de 7 sustituciones conservativas, como un indicio de algo importante.

Identificación de Motivos

Los alineamientos múltiples son utilizados para obtener diversos tipos de información. Un uso muy común de estos es la búsqueda de motivos.

Un motivo es un patrón de DNA o proteínas, al que se le podría asociar una función, es decir que tiene una significancia biológica.

Es común encontrar varios patrones asociados a un motivo, uno que es el más probable y los otros que son las posibles variaciones sobre este (más adelante veremos en detalle este aspecto).

En esta oportunidad usaremos el programa **MEME** para identificación de motivos y **MAST** para la búsqueda de estos motivos en otras secuencias.

Visite el sitio web de MEME en la siguiente dirección:

<http://meme.sdsc.edu/meme/meme.html>

Menu

- [Submit A Job](#)
- [Resources](#)
- [Alternate Servers](#)
- [Other Tools](#)



MEME
Multiple Em for Motif Elicitation

Version 3.5.3

Use this form to submit DNA or protein sequences to MEME. MEME will analyze your sequences for similarities among them and produce a description (**motif**) for each pattern it discovers. Your results will be sent to you by e-mail.

Data Submission Form

Required

Your **e-mail address**:

Re-enter e-mail address:

Please enter the **sequences** which you believe share one or more motifs. The sequences may contain no more than **60,000 characters** total in any of a large number of **formats**.
Enter the **name of a file** containing the sequences here:
 Examinar...
or
The **actual sequences** here (Sample Input Sequences):

How do you think the occurrences of a single motif are **distributed** among the sequences?

One per sequence
 Zero or one per sequence
 Any number of repetitions

MEME will find the optimum **width** of each motif within the limits you specify here:

Minimum width (≥ 2)
 Maximum width (≤ 300)
 Maximum **number of motifs** to find

Los resultados de MEME nos son enviados por correo, para eso debemos dar nuestra dirección de correo y confirmarla.

En la parte izquierda del formulario aparecen algunas opciones, la primera de ellas nos permite establecer el número de motivos que esperamos encontrar en nuestras secuencias, el segundo nos permite establecer la longitud del motivo.

Ingrese los datos que pide el formulario, copie y pegue las secuencias en formato FASTA utilizadas para el primer alineamiento múltiple realizado anteriormente, en la casilla inferior izquierda.

Utilizaremos los parámetros por defecto, así que presione el botón “**Start Search**” en la parte inferior del formulario.

Inmediatamente aparecerá una nueva página con un resumen de los datos ingresados.

Your MEME search results will be sent to: **ampinzonv@unal.edu.co**
 If you do not receive a confirming email message, there could be an error in your email address.

- E-mail address: **ampinzonv@unal.edu.co**
- Sequence file: **pasted_sequences**
- Description:
- Distribution of motif occurrences: **Zero or one per sequence**
- Number of different motifs: **3**
- Minimum number of sites:
- Maximum number of sites: **5**
- Minimum motif width: **6**
- Maximum motif width: **50**
- Statistics on your dataset:

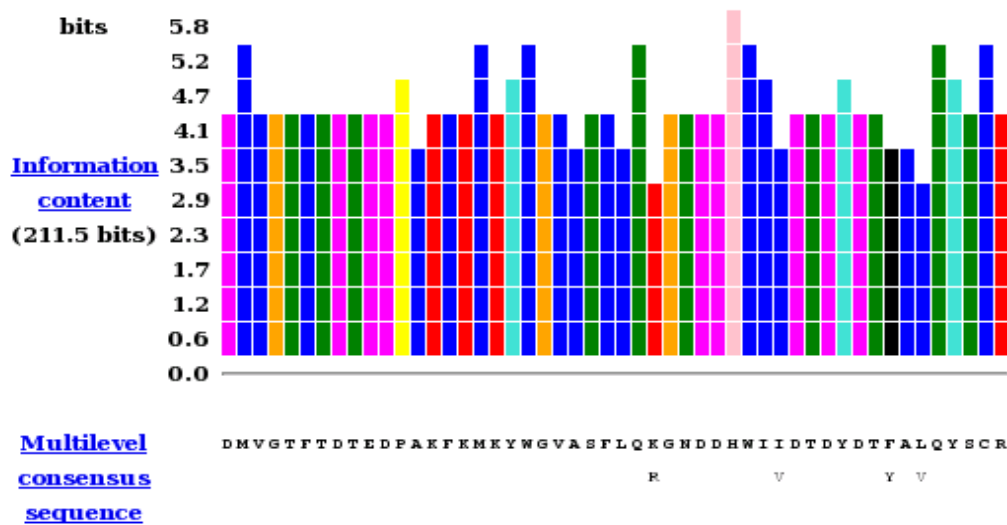
type of sequence	protein
number of sequences	5
shortest sequence (residues)	177
longest sequence (residues)	201
average sequence length (residues)	188.8
total dataset size (residues)	944

Dependiendo del número y longitud de nuestras secuencias los resultados de MEME pueden tardar en llegar. Lo primero que recibiremos es un correo de **confirmación** de MEME, en el cual nos dicen que la identificación está corriendo. Un tiempo después llegarán los **resultados** de MEME conjuntamente con los resultados de un trabajo enviado a MAST que el programa hace por defecto, por ahora podemos obviar este último.

🏠 ☆ meme	» MAST job 24853 results: (Use web browser to view results) - MAST job 24853 results: (Use web browser to view results)	🕒 12:42
🏠 ☆ meme	» MEME job 24853 results: (Use web browser to view results) - MEME job 24853 results: (Use web browser to view results)	🕒 12:42
🏠 ☆ meme	» MEME job 24853 confirmation: - MEME job 24853 confirmation: Your MEME search request 24853 is being	🕒 12:42

Los resultados llegan como archivo adjunto en formato HTML.

La segunda gráfica que encontramos es el “**Diagrama de contenido de información**”:



Este diagrama nos da una idea de cuales posiciones son más conservadas en el motivo, lo cual se mide en **bits** (para nuestro propósito es suficiente saber que la altura de la columna es directamente proporcional a su nivel de conservación). Los colores corresponden al tipo de residuos con mayor prevalencia en dicha posición (ej, el color rojo hace referencia a que en dicha posición se encuentra mayoritariamente aminoácidos cargados positivamente).

En la parte inferior observamos la secuencia consenso de dicho motivo. Esta es una manera de recordar mejor nuestro motivo, es en realidad una representación de este. Dicho de otra manera, **un motivo es en última instancia una matriz de probabilidades de aparición para cada residuo**, la secuencia consenso nos ayuda a visualizar este motivo de mejor manera.

Esta representación del motivo se le puede llamar “**patrón**”, y es común describirlo en una notación estandar denominada “**Patrones Prosite**”.

Si requiere más información al respecto de la notación para patrones, puede consultar el sitio web del manual de Prosite :

<http://ca.expasy.org/prosite/prosuser.html>

La sección “**Sites**” muestra las secuencias en las cuales fue encontrado dicho motivo, con sus respectivas regiones flanqueantes (10 residuos corriente arriba y 10 residuos corriente abajo).

NAME	START	P-VALUE	SITES	
gi 12843160 dbj BAB25881	90	1.13e-64	PLAHHHDDCA	LVQLDSTCAD
gi 5803139 ref NP_006735	88	1.58e-64	PLAHHHDDCA	LVQLDSTCAD

Búsqueda de Motivos

Una vez que hemos **identificado** los posibles motivos en nuestras secuencias, puede ser deseable ver si dicho motivo se encuentra presente en otras secuencias.

Para lograr este objetivo utilizamos la herramienta MAST, la cual dado un motivo realiza una búsqueda de este en otras secuencias.

Presione el botón "**Mast**" que se encuentra en la parte superior derecha, al inicio de la página de resultados de MEME.

Esto le llevará a la página principal de MAST.

Menu

- Submit A Job
- Resources
- Alternate Servers
- Other Tools

MAST
Motif Alignment & Search Tool
Version 3.5.3

Use this form to submit motifs to **MAST** to be used in searching a sequence database. Your data will be processed at the **San Diego Supercomputer Center** and the results will be sent to you via e-mail.

Data Submission Form

Required

Your **e-mail address**:

Re-enter **e-mail address**:

Your **motif file**: **MEME results on pasted sequences**
Sample DNA motif, right click and "save target as" to download

Sequence database to search--select **one** of the following:
A **MAST database**:

or

Your **FASTA** sequence file (1,000,000 sequence characters maximum):
Sample DNA sequence, right click and "save target as" to download.

Optional

Scale motif display threshold by sequence length; **recommended** for **nucleotide** searches

Text output format

DNA-ONLY OPTIONS

Search nucleotide database with protein motifs

Treatment of **reverse complement** strands:

Use individual sequence composition in *E*- and *p*-value calculation

Please send comments and questions to: meme@nbcrc.net

MEME Introduction FAQ Forum Release Notes Papers Download License

Ingrese de nuevo su dirección de correo, seleccione la base de datos **swissprot** del menú desplegable de la derecha llamado "Sequence Database", esta será la base de datos en la que se buscará el motivo.

Opcionalmente es posible proporcionar nuestra propia base de datos (en formato FASTA) mediante la opción "**examinar**" ubicada inmediatamente abajo.

Sequence database to search--select **one** of the following:
A MAST database:

or

Your **FASTA** sequence file (1,000,000 sequence characters maximum):
Sample DNA sequence, right click and "save target as" to download.

Es posible dar una breve descripción de nuestra búsqueda en la casilla “**Description**”, lo que allí escribamos será el asunto del correo que MAST nos hará llegar con los resultados.

Presione el botón “**Start Search**”, el comportamiento de llegada de resultados será similar al de MEME. Después de unos minutos llegará a un correo con un archivo HTML adjunto.

Tal vez se esté preguntando por qué razón MAST no le ha pedido que ingrese un motivo. La razón para esto es que al presionar el botón MAST desde la hoja de resultados de MEME, el sistema se ha encargado directamente de hacerle llegar el motivo a MAST. Si usamos MAST directamente desde el sitio web, nos veríamos obligados a introducir un motivo.

Es posible guardar el archivo del motivo desde la hoja de resultados de MEME, presionando el botón “**view PSSM1**” (para el motivo 1).

Los resultados de MAST son más sencillos de interpretar que los de MEME, básicamente lo que encontrará es una lista de secuencias en las cuales se ha encontrado el motivo que buscamos.

Links	Sequence Name	Description	E-value	Length
EDA?	sp Q00724 RETBP_MOUSE	Plasma retinol-binding pr...	4e-147	201
EDA?	sp P04916 RETBP_RAT	Plasma retinol-binding pr...	5e-146	201
EDA?	sp P02753 RETBP_HUMAN	Plasma retinol-binding pr...	8e-146	201
EDA?	sp Q28369 RETBP_HORSE	Plasma retinol-binding pr...	2.3e-137	201
EDA?	sp P27485 RETBP_PIG	Plasma retinol-binding pr...	3.4e-137	201
EDA?	sp P06912 RETBP_RABIT	Plasma retinol-binding pr...	7.7e-137	201
EDA?	sp P18902 RETBP_BOVIN	Plasma retinol-binding pr...	3.3e-133	183
EDA?	sp P41263 RETBP_CHICK	Plasma retinol-binding pr...	2.8e-124	196
EDA?	sp P06172 RETBP_XENLA	Plasma retinol-binding pr...	3e-78	197
EDA?	sp P24774 RETB1_ONCMY	Plasma retinol-binding pr...	5.2e-76	176
EDA?	sp P24775 RETB2_ONCMY	Plasma retinol-binding pr...	1.7e-75	176
EDA?	sp P08938 PURP_CHICK	Purpurin precursor	3.1e-54	196
EDA?	sp P51910 APOD_MOUSE	Apolipoprotein D precurso...	6.5e-18	189
EDA?	sp P23593 APOD_RAT	Apolipoprotein D precurso...	5.1e-17	189

La primera columna consta de 4 enlaces:

1. **E**, nos lleva a la página de la entrada de dicha proteína en el NCBI.
2. **D**, nos lleva al diagrama de motivos para dicha secuencia, en la misma página. Donde se nos muestra la ocurrencia de dicho motivo dentro de la secuencia.
3. **A**, nos lleva al *alineamiento* del motivo con dicha secuencia en la misma página.
4. **?**, nos lleva a la ayuda acerca de estos enlaces.

Guía elaborada por Andrés M. Pinzón V., del **Centro de Bioinformática** del Instituto de Biotecnología en la Universidad Nacional de Colombia y está distribuida bajo licencia:



Bogotá Colombia - Julio de 2006.

Cualquier sugerencia o inquietud dirigirla a:

ampinzonv@unal.edu.co ó andrespinzon@gmail.com