

EMBOSS

European Molecular Biology Open Software Suite



**Centro de Bioinformática
Instituto de Biotecnología
Universidad Nacional de Colombia**

...qué es EMBOSS?

- Suite de análisis especialmente desarrollada para biología molecular (EMBnet).
- Toolkit para aplicaciones o workflows.
- **Código Abierto.**
- Provee alrededor de 150 programas.
- Integra otros programas (EMBASSY).
- Soporta todas las plataformas UNIX más comunes.
- EMBOSS para MS/WINDOWS en desarrollo.

...150 programas?

- Alineamiento de Secuencias.
- Búsqueda de secuencias en bases de datos.
- Identificación de motivos en proteínas.
- Identificación de repeticiones.
- Análisis de uso codónico.
- Identificación de patrones.
- Herramientas de presentación.
- Visualización.
- etc....



The screenshot shows the EMBOSSt website interface. At the top center is the EMBOSSt logo in blue. Below it is a horizontal menu of navigation links: news, what?, apps, docs, interfaces, download, admin, developers, support, co-ordination, licence, and credits. On the right side, there is a SourceForge logo with the text 'hosted by' and 'net', and below it, 'latest stable release: 3.0.0'. A paragraph of text states: 'The EMBOSSt project gratefully acknowledges the support of CCP11.' Below this is a section titled 'Latest news about EMBOSSt' with a link to 'News page'. Underneath is a sub-section for 'Release 3.0.0' with the text: 'EMBOSSt 3.0.0 is available for download. As usual, the release date was 15th July 2005.' This is followed by a sub-section for 'FTP Server at Open Bio' with the text: 'The EMBOSSt FTP server at ftp.uk.embnet.org was hosted by the Rosalind Franklin Centre for Genomics Research. From 15th July 2005 the EMBOSSt FTP server has moved to a new home provided by the Open Bio Foundation at ftp://emboss.open-bio.org/pub/EMBOSSt/'

<http://emboss.sourceforge.net/>



... algo de historia.

- **EGCG** (1988) surgió como colaboración entre la **EMBnet** y otros institutos, utilizaba las librerías de **GCG**.
- Pensado para el Sanger Center.
- Problemas de licenciamiento: imposible distribuir más licencias académicas.
- Se inicia el desarrollo de **EMBOSS**.

Cómo luce EMBOSS?

Línea de comandos!

Muestra la versión de EMBOSS instalada

```
xterm
andipin@linux:~$ embossversion
Writes the current EMBOSS version number
3.0.0
andipin@linux:~$ wosname
Finds programs by keywords in their one-line documentation
Keyword to search for, or blank to list all programs: prettyplot
SEARCH FOR 'PRETTYPLOT'
prettyplot      Displays aligned sequences, with colouring and boxing
andipin@linux:~$ prettyplot
Displays aligned sequences, with colouring and boxing
Input sequence set: 
```

Lista los programas de la actual versión de EMBOSS

Ruta a Archivo

GUIs para EMBOSS

The image shows a screenshot of the Jemboss graphical user interface. The window title is "Jemboss". The menu bar includes "File", "Preferences", "Tools", and "Help". On the left, there is a vertical sidebar with a list of tool categories: ALIGNMENT, DISPLAY, EDIT, ENZYME KINETICS, FEATURE TABLES, HMM, INFORMATION, NUCLEIC, PHYLIP, PHYLOGENY, PROTEIN, and UTILS. Below this list is a "GoTo:" field containing "cpgplot" and a scrollable list of tool names, with "cpgplot" highlighted in blue. The main area of the window is titled "CPGPLOT" in red, with the subtitle "Plot CpG rich areas". Under the heading "input section", it says "Enter the sequence as:" followed by three radio buttons: "file / database entry" (selected), "paste", and "list of files". Below this is a text field containing "embl:hsfau" and a "Browse files..." button. There are two buttons: "Input Sequence Options" and "Reset". A red text box contains "LOAD SEQUENCE ATTRIBUTES". Under the heading "required section", there are four rows of configuration options, each with a text field and a label: "Window size" (value: 100, range: min:1 max:518 default:100), "Window shift increment" (value: 1, range: min:1 max:100 default:1), "Minimum length of an island" (value: 200, range: min:1 max:518 default:200), and "Minimum observed/expected" (value: 0.6, range: min:0. max:10. default:0.6). At the bottom, there is a status bar with a red and blue icon, the text "(No Current Jobs)", and a dropdown menu set to "interactive" with a right-pointing arrow icon.

Jemboss

File Preferences Tools Help

CPGPLOT
Plot CpG rich areas

input section

Enter the sequence as:
 file / database entry or paste or list of files

Sequence Filename
embl:hsfau

LOAD SEQUENCE ATTRIBUTES

required section

Window size
(min:1 max:518 default:100)

Window shift increment
(min:1 max:100 default:1)

Minimum length of an island
(min:1 max:518 default:200)

Minimum observed/expected
(min:0. max:10. default:0.6)

(No Current Jobs) interactive

- EMBOSS
- Alignment consensus
- Alignment dot
- Alignment multiple
- Alignment
- Display
- Edit
- Enzyme kinetics
- Feature tables
- Information
- Nucleic acid 2D
- Nucleic acid CPG
- Nucleic acid codon
- Nucleic acid composition
- Nucleic acid gene
- Nucleic acid motifs
- Nucleic acid primers
- Nucleic acid profiles
- Nucleic acid repeats
- Nucleic acid restriction
- Nucleic acid translation
- Nucleic acids mutation
- Nucleic acids transcription
- Phylogeny
- Protein 2D
- Protein 3D
- Protein composition
- Protein motifs
- Protein mutation
- Protein profiles
- Utils database
- EMBOSS-Launcher

```

File Edit Search Preferences Shell Macro Windows Help
Untitled line 10, col 43, 3069 bytes
CLUSTAL W(1.4) multiple sequence alignment
Dm_BABII_1  AMAEAIIFSVLKEG-LLSLSQAARKFDIPYPTFVLYAN
Dm_BAB_1___ TMAEAIIFSVLKEG-LLSLSQAARKYDIPYPTFVLYAN
Dm_Pfk_1___ ALDLAADAVIIEG-LSLQKAADRFDISKTVLWRRVR
Dm_Pfk_3___ KLAKAVHACLNEG-MSQNHAANLFEIPKSTLWRHLQ
Mm_E93Like  TLFEPTSWWMSCK-MSVSKQASTYGTGHTL EYKVK
EYKVK
EYKVK
EYKVK
EYKVK
EYKVK
EYKVK
YDKVR
HDRLK

```

Emma - Multiple alignment program - interface to ClustalW

Multiple alignment | Pairwise alignment | Advanced | Format

Gap opening penalty: 10.0

Gap extension penalty: 5.0

Substitution matrix: **blosum**

DNA substitution matrix: **Scoring table which describes the similarity of each amino acid to each other.**

Gap separation distance: 0

End gap separation

Residue specific penalties

Use 'Transition weight'

Reduce gap penalty for hydrophilic residues

Hydrophilic residues: GPSNDQEKR

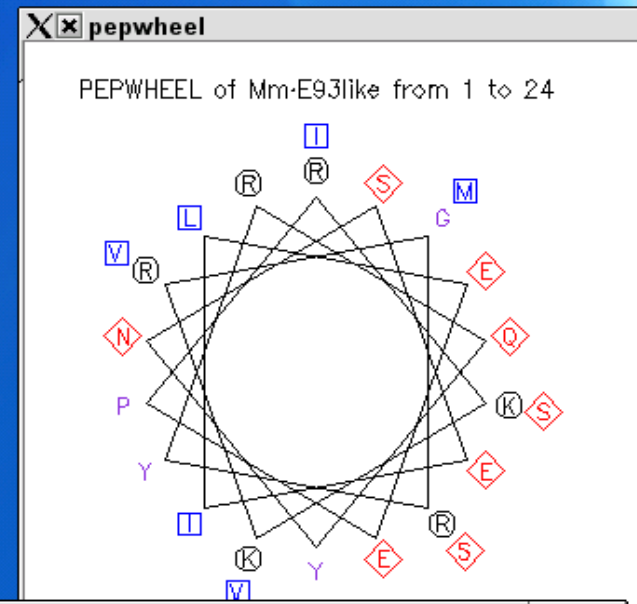
Delay alignment of distant sequences (%): 30

Multiple sequence file: /home/sgmd/Projekte/piefke/psq-domains.fasta

Write multiple alignment to: []

Write dendrogram file to: []

Run emma | Show | Help | Cancel



Pepwheel - Shows protein sequences as helices

Plot parameters

Steps: 18 | Turns: 5

Squares: ILVM | Diamonds: DENQST | Octagons: HKR

Amphipathic

From position 1 to 24

Formats

Input: auto

Graphic: Colour PS

Sequence file: /home/sgmd/Projekte/piefke/psq-domains.fasta

Write output to: []

Run pepwheel | Show plot | Help | Cancel

Showdb - List available databases

Name	Type	ID	Qry	All	Comm
embl	N	OK	-	-	EMBL
genbank	N	OK	-	-	Genbank
nr_protein	P	OK	OK	OK	NCBI
swall	P	OK	-	-	SWISS
swissprot	P	OK	OK	OK	SWISS



megamerger
meme
merger
msbar
mwcontam
mwfilter
needle
newcpgreport
newcpgseek
newseq
noreturn
notseq
nthseq
octanol
oddcomp
palindrome
pasteq
patmatdb
patmatmotifs
pepcoil
pepinfo
pepnet
pepstats
pepwheel
pepwindow
pepwindowall
plotcon
plotorf
polydot
preg
prettyplot
prettyseq
primersearch
profit
prophecy
prophet
pscan

PEPWHEEL

(Shows protein sequences as helices)

Fields with a coloured background are optional and can safely be ignored...



[[Hide optional fields](#)]

1. SET THE PARAMETERS FOR THE RUN (OR ACCEPT THE DEFAULTS...)

input section

Select an input sequence.

Use one of the following three fields: *(file must contain a protein sequence)*

1. To access a sequence from a database, enter the USA path here: *(dbname:entry)*

2. Or, upload a sequence file from your local computer here:

 Examiner...

3. Or enter the sequence data manually here:

output section

Plot the wheel? ▾

Number of steps (integer)

Number of turns (integer)

- ▶ ALIGNMENT
- ▶ DISPLAY
- ▶ EDIT
- ▶ ENZYME KINETICS
- ▶ FEATURE TABLES
- ▶ HMM
- ▶ INFORMATION
- ▶ MENUS
- ▶ NUCLEIC
- ▶ PHYLOGENY
- ▶ PROTEIN
- ▶ UTILS
- ▶ ALPHABETIC LIST OF PROGRAMS

David project

PROJECT MANAGEMENT

▶ New project subproject?

Rename
▶ proj.

Move proj. to

Delete
▶ proj.

PROJECT FILES

▶ New file

View Edit Copy Delete List G-E&G

.command
emma.html
error
mult1
nuclist
outseq.aln
prettyplot.2.png

View with

Upload Examinar...

FILE TYPE

PROJECT RESULTS

▶ Sort by name

Program Output	yy mm dd	hh mm ss	Copy
<i>showalign</i>	05.08.18	12.06.22	Files Delete
<i>prettyplot</i>	05.08.18	12.05.39	Files Delete
<i>plotorf</i>	05.07.16	21.17.50	Files Delete
<i>cons</i>	05.06.08	14.09.39	Files Delete
<i>infoalign</i>	05.06.08	14.07.14	Files Delete
<i>showalign</i>	05.06.08	14.03.27	Files Delete
<i>prettyplot</i>	05.06.08	14.03.02	Files Delete
<i>plotcon</i>	05.06.08	14.01.59	Files Delete
<i>infoalign</i>	05.06.08	14.01.16	Files Delete
<i>emma</i>	05.06.08	14.0.15	Files Delete
<i>emma</i>	05.06.08	13.58.33	Files Delete
<i>sixpack</i>	05.06.08	12.12.28	Files Delete

▶ Search for programs

by keywords :

and or

Marc Colet & Martin Sarachi
Version 1.2.1
7/10/09

Workflows con EMBOSS

The screenshot shows the Wildfire application window with the menu open. The 'MULTIPLE' menu item is selected, and a sub-menu is displayed with the following items: wildfire, emma, infection, biorcon, prettyplot, showalign, and tranalign. The console output at the bottom shows the following text:

```
Open project  
Opening project: myproj  
Creating subdirectories...Created. Done!
```

The screenshot shows the Wildfire application window with a workflow diagram displayed in the main area. The diagram consists of two yellow boxes: 'getorf 1' on the left and 'garnier 2' on the right, connected by a black arrow pointing from 'getorf 1' to 'garnier 2'. The console output at the bottom shows the following text:

```
struct  
-sequence seq01.orf -idc 0 -rformat tagseq -outfile seq01.  
struct  
[CmdDelete] dolt
```