

# Initiation à EMBOSS



D'après les notes de cours de Valérie LEDENT et Guy BOTTU  
ULB-VUB, Bruxelles, Belgique



Daniel Gautheret, 2004  
ESIL, Université de la Méditerranée

# European Molecular Biology Open Software Suite

- freeware package for sequence analysis, distributed under GNU licence
- EMBOSS development team first headed by Peter Rice at Sanger Centre (Hinxton), now by Alan Bleasby at HGMP-RC (Hinxton)
- contributions from EMBnet Nodes
- extra programs (Embassadors) distributed under their own licence
- current version 2.8 (nov '03)

# Features of EMBOSS

- Many programs, each specialized in one or a related set of functions
- EMBOSS programs run under UNIX with support for a complete command line
- Some programs are specialized in using data produced by another program
- Some programs produce graphic output, which can be sent directly to screen (X11, Tektronix) or written in file in a variety of formats (PS, HPGL, PNG,...)

# Features of EMBOSS

- Sequence analysis programs can operate on "private" sequences (stored in files) as well as on "public" sequences from databanks
- No proprietary sequence format, EMBOSS programs can read and write sequence files in any format (EMBL, GenBank, fastA, "raw",...)
- Other data used by programs (genetic codes, amino acid comparison tables, restriction sites,...) usually stored in data files. User can request usage of alternative data file

# EMBOSS configuration and data files

Variable d'environnement  
(export ou setenv)

- `$EMBOSS_DATA` 

directory with program data files

- `$EMBOSS_DATA/acd`

directory with ACD files

- `$EMBOSS_DATA/./doc/programs`

on-line help (xxx.txt for tfm and xxx.html)

- `$EMBOSS_DATA/./emboss.default`

file with databank definitions and other configuration settings

- `$PLPLOT_LIB`

directory with PLplot shared libraries

# EMBOSS download and information

- EMBOSS WWW site

<http://www.hgmp.mrc.ac.uk/Software/EMBOSS/>

- mailing lists :

[emboss@hgmp.mrc.ac.uk](mailto:emboss@hgmp.mrc.ac.uk) (for users)

[emboss-dev@hgmp.mrc.ac.uk](mailto:emboss-dev@hgmp.mrc.ac.uk) (for developers)

send bug reports to [emboss-bug@hgmp.mrc.ac.uk](mailto:emboss-bug@hgmp.mrc.ac.uk)

- citation :

Rice, Longden & Bleasby :TIBS 16, 276-277 (2000)

# The USA (*Uniform Sequence Address*) format

- ★ The Uniform Sequence Address, or USA, is a standard sequence naming used by all EMBOSS applications.
- ★ The format is not required. When reading in a sequence, EMBOSS will guess the sequence format by trying all known formats until one succeeds.
- ★ When writing out a sequence, EMBOSS will use **fasta** format by default. You can specify another format to use, for example:
  - ★ `"format::file"`
    - `fasta::toto.seq` (toto.seq est en format fasta)
    - `embl::toto.seq` (..ou en format embl)
  - ★ `"format::file:entry"`
    - `fasta::toto.seq:p345662` (séquence p345662 de toto.seq)
  - ★ `"dbname:entry"`
    - `embl:p345662` (séquence p345662 de la banque embl installée localement)
  - ★ `"@listfile"` (a file of file-names)
    - `@humangenome`

# La ligne de commande

The EMBOSS command syntax follows normal UNIX command conventions, (options start with a '-', for example: "program -format 2").

If in doubt, type:

**program -help** to get some help on the options  
or

**program -opt** to make the program prompt you for common options  
or

**tfm program** to get the full help on a program



# Interfaces d'EMBOSS

## ★ JAVA

- Jemboss
- S'installe par le Web sur votre machine

## ★ Interfaces WEB

- Pise
- W2H
- wEMBOSS

## ★ Interfaces GUI

- embossRunner (Mac)
- Staden
- CoLiMate
- Kaptain