Sequence Lab

A short intro to bio-knoppix

Humberto Ortiz Zuazaga & Carlos Rodriguez

`humberto|carlos@hpcf.upr.edu`

Outline

- Why bio-knoppix?
- Booting bio-knoppix
- knoppix is linux is UNIX
- Tools on the CD
- EMBOSS
- Sequence file formats
- Errata

bio-knoppix

Carlos's custom version of KNOPPIX, a live-cd version of Debian GNU/linux. Debian is free software, software you are allowed to modify and redistribute.

Contact the HPCf for more information on bio-knoppix, or see www.debian.org and www.gnu.org for more information on debian and free software.

Booting bio-knoppix

Place the CD and reboot the computer. Some machines require changing the BIOS options to boot from the CD. As a last resort, there is also a floppy image on the CD that you may use to boot the PC.

When you log out of KNOPPIX, the CD ejects automatically.

UNIX

bio-knoppix has menus, icons, and a desktop. Most work will be done in a “shell” window. The icon has a picture of a sea shell.

- `ls` - LiSt files
- `cp` - CoPy files
- `man` - Read MANuals

Most unix tasks can also be accomplished via the menus and icons.

Tools in bio-knoppix

- EMBOSS - European Molecular Biology Open Software Suite
- Artemis - free genome browser and annotation tool.
- ImageJ - NIH sponsored image processing tool

EMBOSS

- European Molecular Biology Open Software Suite
- more than 100 free tools for molecular biology
- written with the express purpose of remaining non-commercial
- Online, free tutorials, examples, manuals,...

EMBOSS tools

- Sequence alignment
- Rapid database searching with sequence patterns
- Protein motif identification, including domain analysis
- Nucleotide sequence pattern analysis, for example to identify CpG islands or repeats.
- Codon usage analysis for small genomes
- Rapid identification of sequence patterns in large scale sequence sets.
- Presentation tools for publication
- more...

EMBOSS Introduction

The program `wossname` searches for EMBOSS programs by keyword:

```
$ wossname abi
Finds programs by keywords in their one-line documentation
SEARCH FOR 'ABI'
abiview          Reads ABI file and display the trace
```

The Fine Manual

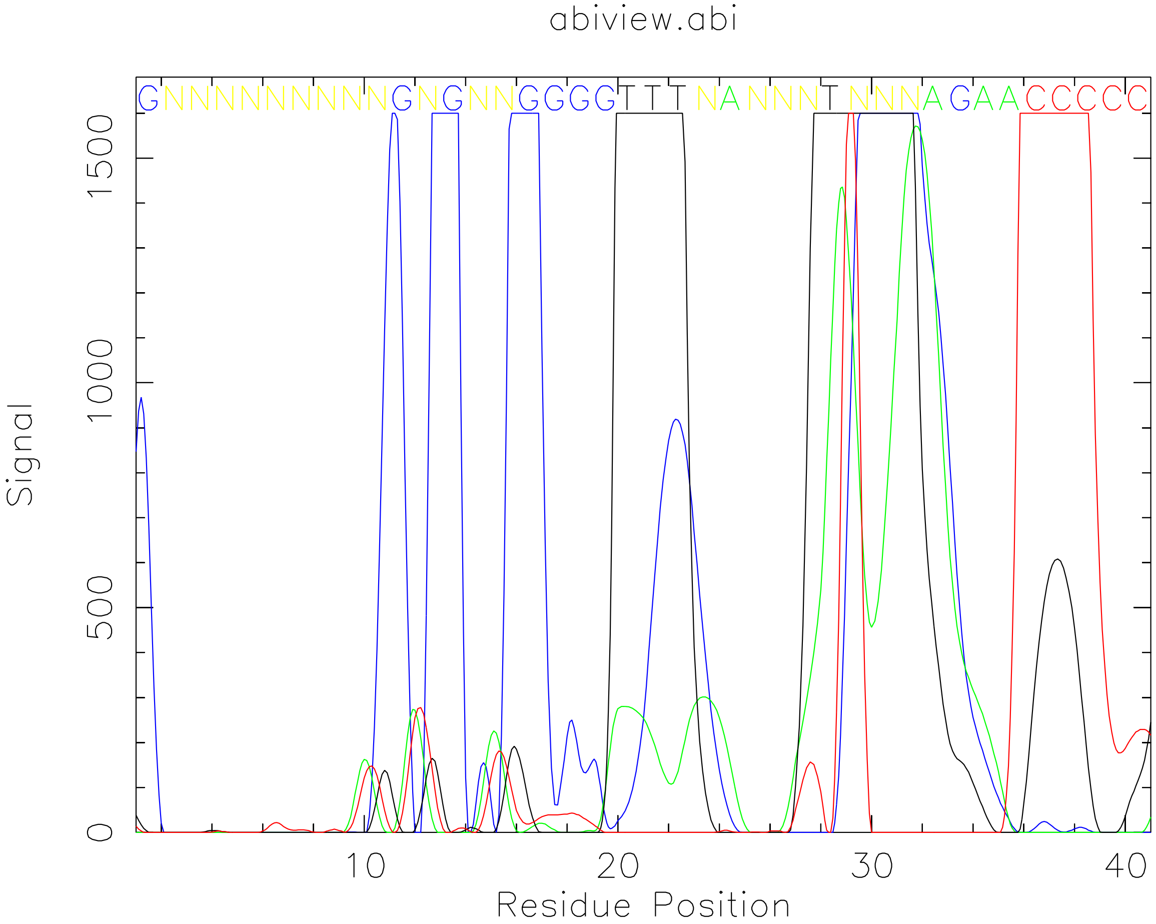
The program `tfm` reads a program manual:

```
$ tfm abiview
Displays a program's help documentation manual

                                  abiview

Function

   Reads ABI file and display the trace

Description

   abiview reads in an ABI sequence trace file and graphically displays
   the results.
⋮
```

Sequence formats

- Chromatograms .abi, .scf
- Sequence: Fasta, GenBank, GCG, ...
- Multiple sequence: clustal, MSF, nexus, phylip
- Phylogeny new jersey tree format
- Structure .pdb

Abiview


abiview.abi
GNNNNNNNNNGNGNNGGGGTTTNANNNTNNNAGAACCCCCC
Signal
0
500
1000
1500
10
20
30
40
Residue Position

IUPAC Nucleotide Codes

IUPAC Code	Meaning	Complement
A	A	T
C	C	G
G	G	C
T/U	T	A
M	A or C	K
R	A or G	Y
W	A or T	W
S	C or G	S
Y	C or T	R
K	G or T	M
V	A or C or G	B
H	A or C or T	D
D	A or G or T	H
B	C or G or T	V
N	G or A or T or C	N

Amino acid codes

A	Ala	alanine	P	Pro	proline
B	Asx	aspartate or asparagine	Q	Gln	glutamine
C	Cys	cysteine	R	Arg	arginine
D	Asp	aspartate	S	Ser	serine
E	Glu	glutamate	T	Thr	threonine
F	Phe	phenylalanine	U	Sec	selenocysteine
G	Gly	glycine	V	Val	valine
H	His	histidine	W	Trp	tryptophan
I	Ile	isoleucine	Y	Tyr	tyrosine
K	Lys	lysine	Z	Glx	glutamate or glutamine
L	Leu	leucine	X	Xaa	any
M	Met	methionine			
N	Asn	asparagine			

B, Z , and X are IUPAC ambiguity codes. Selenocysteine (U) is the 21st amino acid. See online resources.

Fasta format

```
>XLRHODOP L07770.1 Xenopus laevis rhodopsin mRNA, complete cds.
ggtagaacagcttcagttgggatcacaggcttctagggatcctttgggcaaaaaagaaac
acagaaggcattctttctatacaagaaaggactttatagagctgctaccatgaacggaac
agaaggtccaaatttttatgtccccatgtccaacaaaactggggtggtacgaagcccatt
⋮
```

GenBank flatfile format

```
LOCUS       XELRHODOP               1684 bp    mRNA    linear   VRT 15-FEB-1996
DEFINITION  Xenopus laevis rhodopsin mRNA, complete cds.
ACCESSION   L07770
VERSION     L07770.1  GI:214734
KEYWORDS    G protein-coupled receptor; phototransduction protein; retinal
            protein; rhodopsin; transmembrane protein.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1684)
  AUTHORS   Knox,B.E., Scalzetti,L.C., Batni,S. and Wang,J.Q.
  TITLE     Molecular cloning of the abundant rhodopsin and transducin from
            Xenopus laevis
  JOURNAL   Unpublished (1992)
REFERENCE   2  (bases 1 to 1684)
  AUTHORS   Batni,S., Scalzetti,L., Moody,S.A. and Knox,B.E.
  TITLE     Characterization of the Xenopus rhodopsin gene
  JOURNAL   J. Biol. Chem. 271 (6), 3179-3186 (1996)
  MEDLINE   96216396
   PUBMED   8621718
```

⋮

Sequence features

```
FEATURES             Location/Qualifiers
     source          1..1684
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /tissue_type="retina"
                     /dev_stage="adult"
                     /tissue_lib="lambda-ZAPII"
     CDS             110..1174
                     /note="gene accession number U23808"
                     /codon_start=1
                     /product="rhodopsin"
                     /protein_id="AAC42232.1"
                     /db_xref="GI:214735"
                     /translation="MNGTEGPNFYVPMSNKTGVVRSPFDYPQYYLAEPWQYSALAAYM
                     FLLILLGLPINFMTLFVTIQHKKLRTPLNYILLNLVFANHFMVLCGFTVTMYTSMHGY
                     FIFGQTGCYIEGFFATLGGEVALWSLVVLAVERYMVVCKPMANFRFGENHAIMGVAFT
                     WIMALSCAAPPLFGWSRYIPEGMQCSCGVDYYTLKPEVNNESFVIYMFIVHFTIPLIV
                     IFFCYGRLLCTVKEAAAQQQESATTQKAEKEVTRMVVIMVVFFLICWVPYAYVAFYIF
                     THQGSNFGPVFMTVPAFFAKSSAIYNPVIYIVLNKQFRNCLITTLCCGKNPFGDEDGS
                     SAATSKTEASSVSSSQVSPA"
```

⋮

Errata

1. When following the tutorial, add the letter t to all sequence database names (tembl, tsp, tgenbank, ...)
2. Before running `dottup` type:

```
export PLPLOT_LIB=/usr/local/share/EMBOSS
```