



MICB 405 Bioinformatics

Mini-Lab #4 – ClustalX

Dr. Joanne Fox

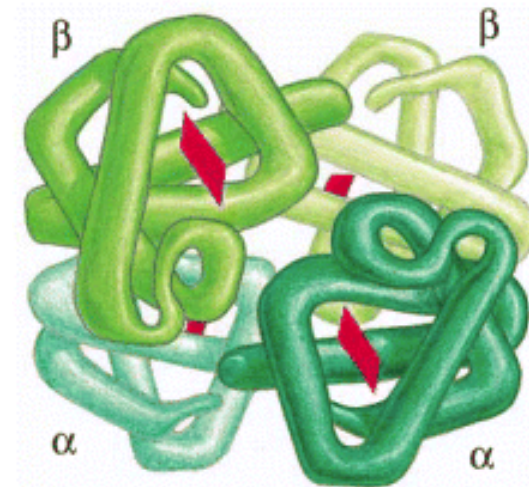
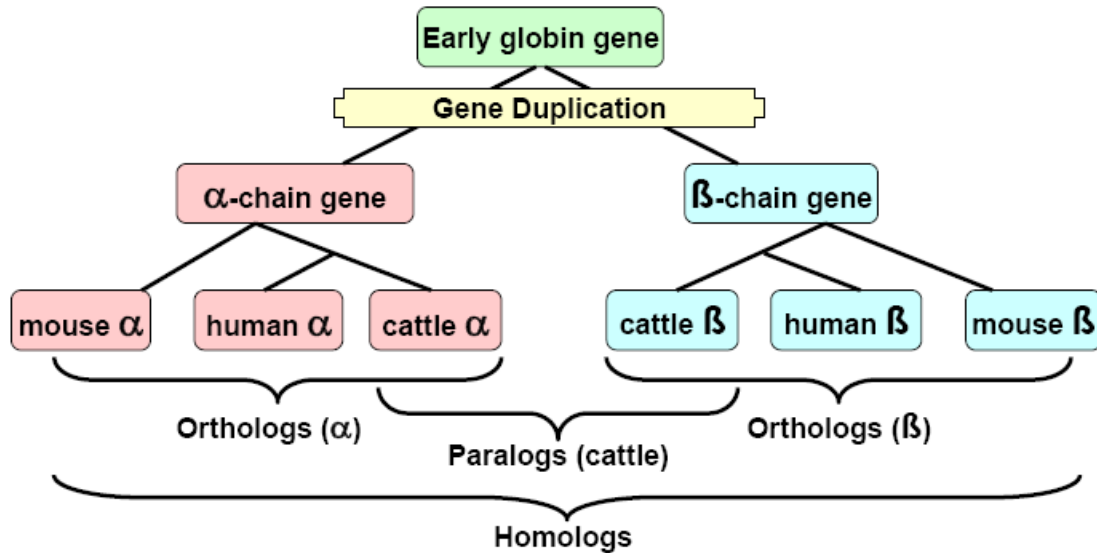
joanne@misl.ubc.ca

We gratefully acknowledge the funding for the development of these teaching materials provided by UBC students through the Teaching and Learning Enhancement Fund (TLEF) from 2005-2007.

Objectives

- By the end of today's tutorial:
 - You will download and install ClustalX on your own computer.
 - You will use ClustalX to generate a multiple sequence alignment for a set of globin sequences.
 - You will view a phylogenetic tree generated from this set of globin sequences.

The Globin Genes

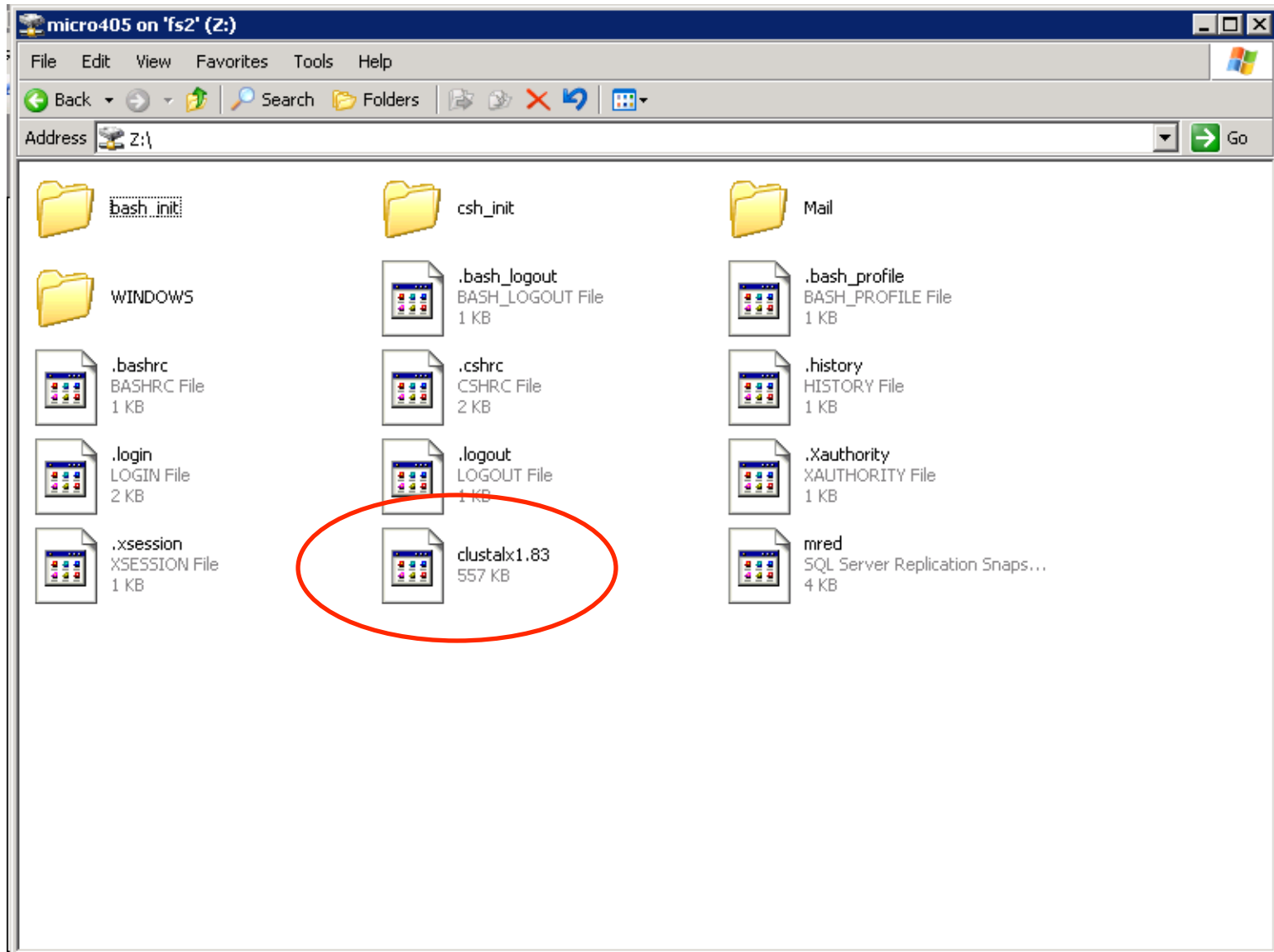


Hemoglobin is an abundant protein in red blood cells that contains two copies of α globin and two copies of β globin.

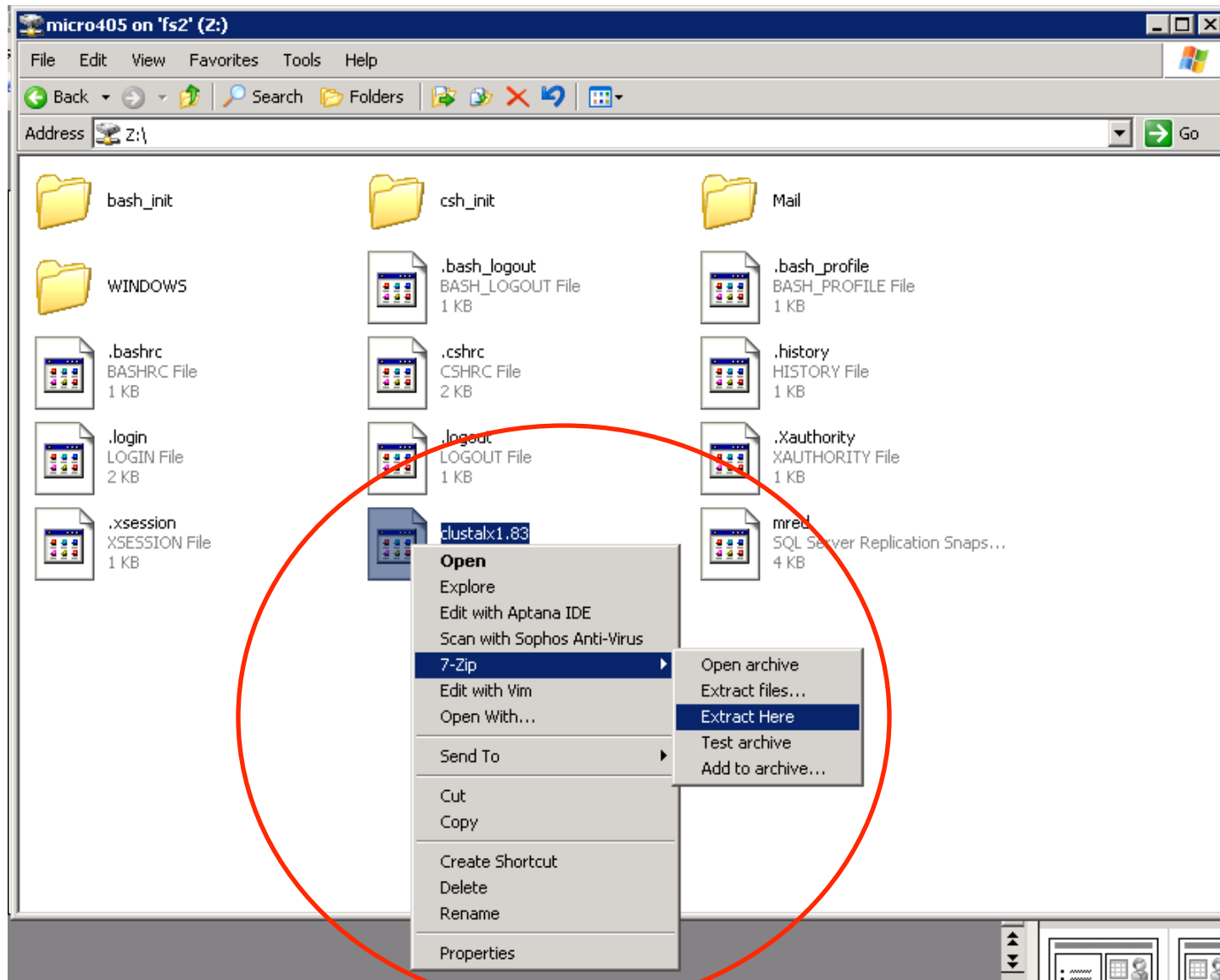
Download ClustalX v1.83

- Go to: <ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/>
- Scroll down list and choose:
 - clustalx1.83.zip
- Download
- Extract all files to a fold where you can find it. ie. Desktop

Find 'clustalx1.83' File



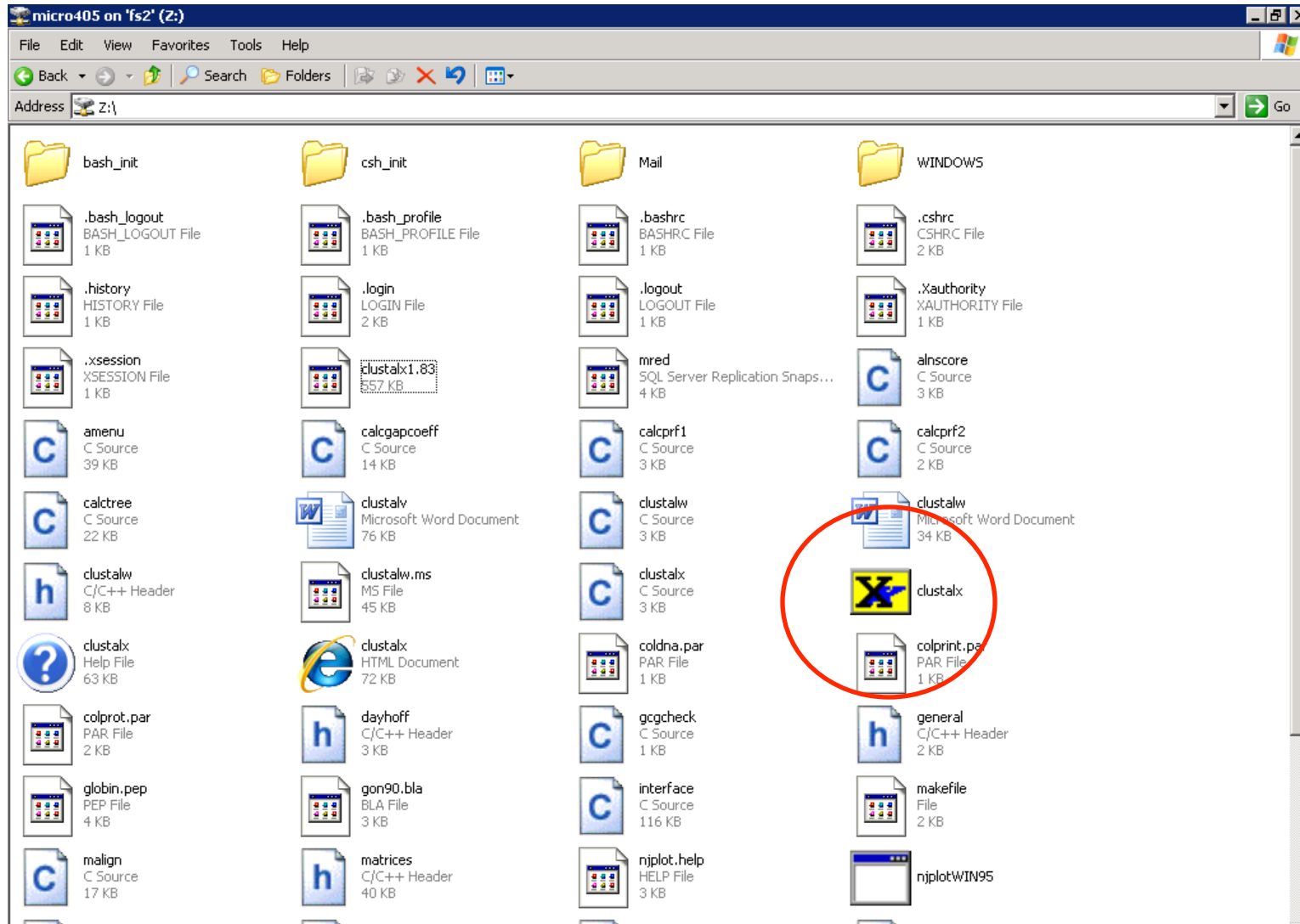
Right Click 'clustalx1.83' and 'Extract Here' with 7-Zip Option





Clustalx.exe

Double Click 'clustalx' File



Starting up ClustalX

File:

-Load sequences

Edit:

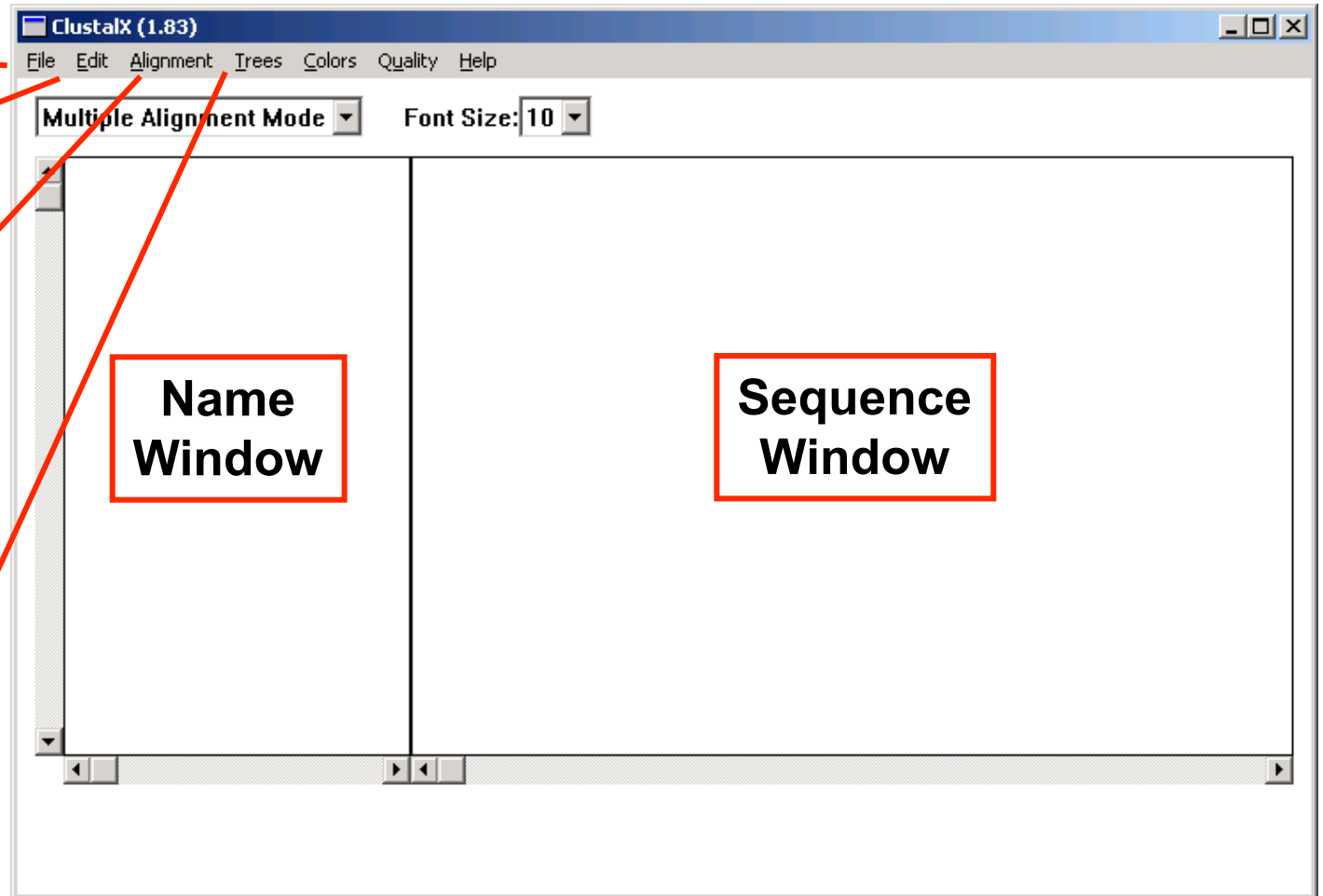
-Remove all gaps

Alignment:

-Do complete alignment
-Alignment parameters

Trees:

-Bootstrapped NJ
-Output format options




```

globin.pep - WordPad
File Edit View Insert Format Help
VHLTPEEKSA VTLWGVKVV DEVGGEALGR LLVVPWTQR FFESFGDLST
PDAVMGNPKV KAHGKVLGA FSDGLAHLDN LKGTFFATLSE LHCCKLHVDP
ENFRLLGNVL VCVLAHFFGK EFTPPVQAA Y QKVVAGVANA LAHKYH*
C:ID HBB_HUMAN STANDARD; PRT: 146 AA.
C:AC P02023;
C:DT 21-JUL-1986 (REL. 01, CREATED)
C:DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
C:DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
C:DE HEMOGLOBIN BETA CHAIN. . . .

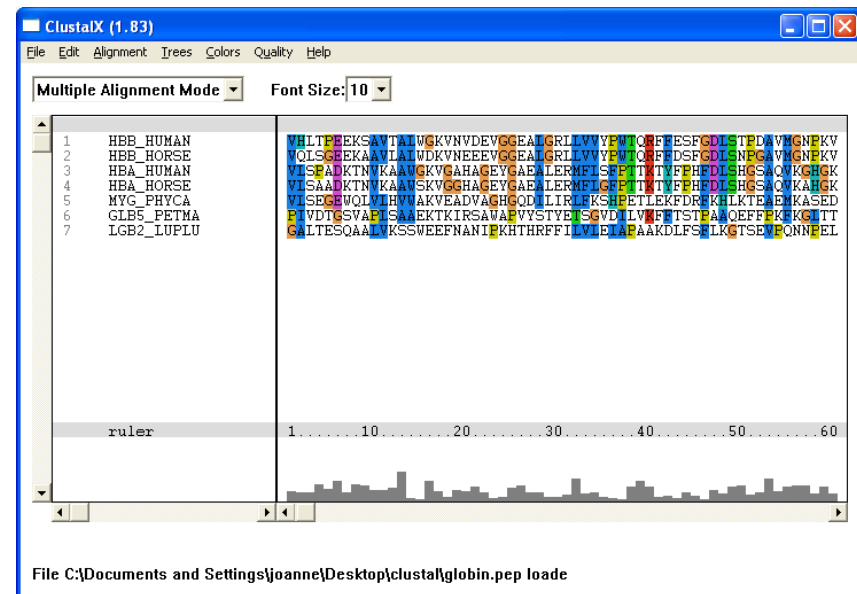
>P1:HBB_HORSE
Sw:Hbb_Horse => HBB_HORSE
VQLSGEEKAA VLALWVKVNE EEVGGGEALGR LLVVPWTQR FFDSFGDLSE
PGAVMGNPKV KAHGKVLHS FGEGVHHLDN LKGTFFAALSE LHCCKLHVDP
ENFRLLGNVL VVVLARHFGK DFTPELQASY QKVVAGVANA LAHKYH*
C:ID HBB_HORSE STANDARD; PRT: 146 AA.
C:AC P02062;
C:DT 21-JUL-1986 (REL. 01, CREATED)
C:DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
C:DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
C:DE HEMOGLOBIN BETA CHAIN. . . .

>P1:HBA_HUMAN
Sw:Hba_Human => HBA_HUMAN
VLSPADKTNV KAAWGKVGAAH AGEYGAEALE RMFLSFPTTK TYFPFHDLSE
GSAQVKHGK KVADALTNV AVHDDMPNAL SALSDLHAHK LRVDPVNFKL
LSHCLLVTLA AHLPAEFTPA VHASLDKFLA SVSTVLTSKY R*
C:ID HBA_HUMAN STANDARD; PRT: 141 AA.
C:AC P01922;
C:DT 21-JUL-1986 (REL. 01, CREATED)
C:DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
C:DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
C:DE HEMOGLOBIN ALPHA CHAIN. . . .

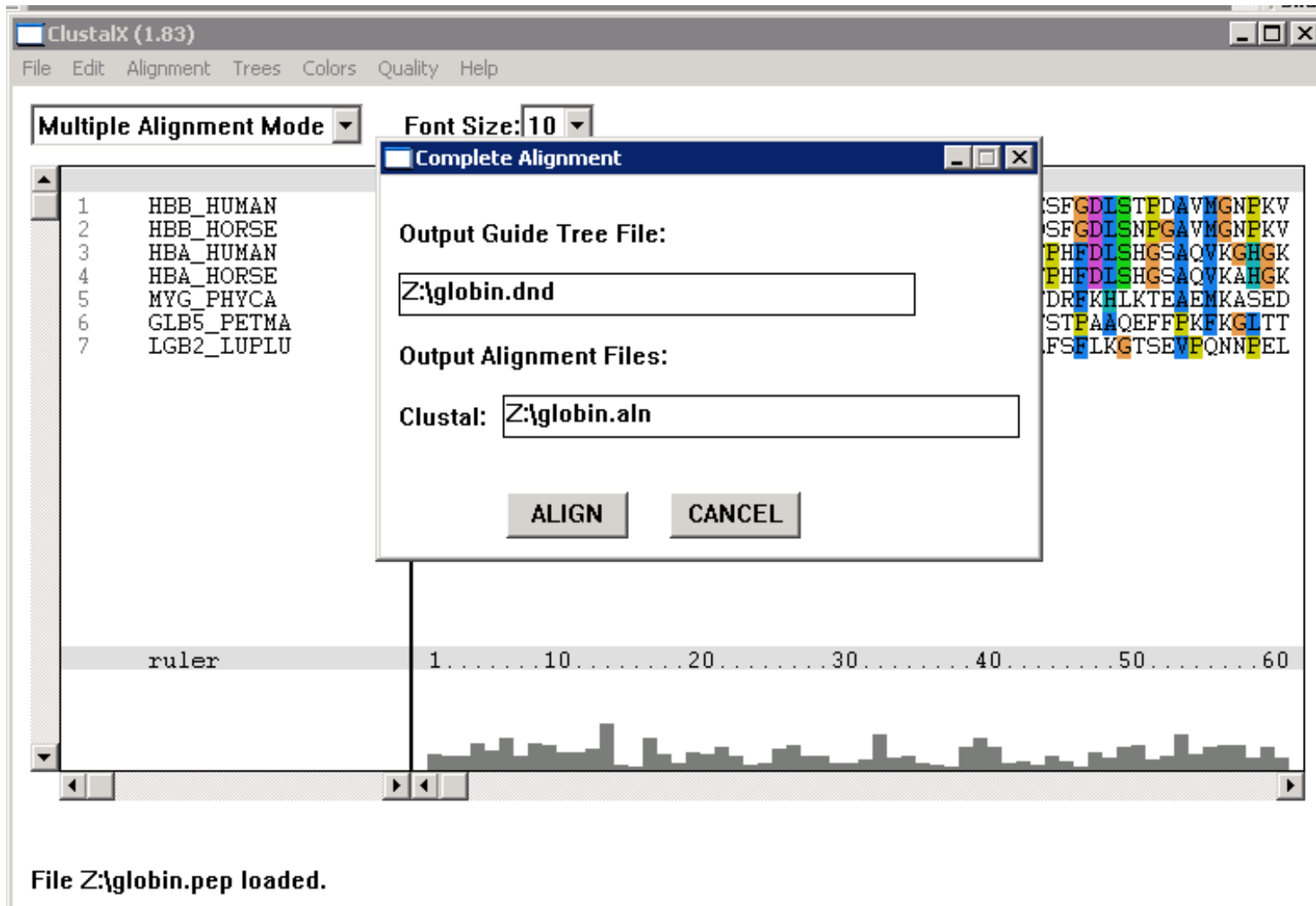
>P1:HBA_HORSE
Sw:Hba_Horse => HBA_HORSE
VLSAADKTNV KAAWSKVGGAH AGEYGAEALE RMFLGFPTTK TYFPFHDLSE
GSAQVKAHGK KVGDPALTLAV GHLDLDPAL SNLSDLHAHK LRVDPVNFKL
LSHCLLSTLA VHLPNDFTPA VHASLDKFLS SVSTVLTSKY R*
C:ID HBA_HORSE STANDARD; PRT: 141 AA.
C:AC P01958;
C:DT 21-JUL-1986 (REL. 01, CREATED)
C:DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
C:DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
C:DE HEMOGLOBIN ALPHA CHAINS (SLOW AND FAST). . . .

```

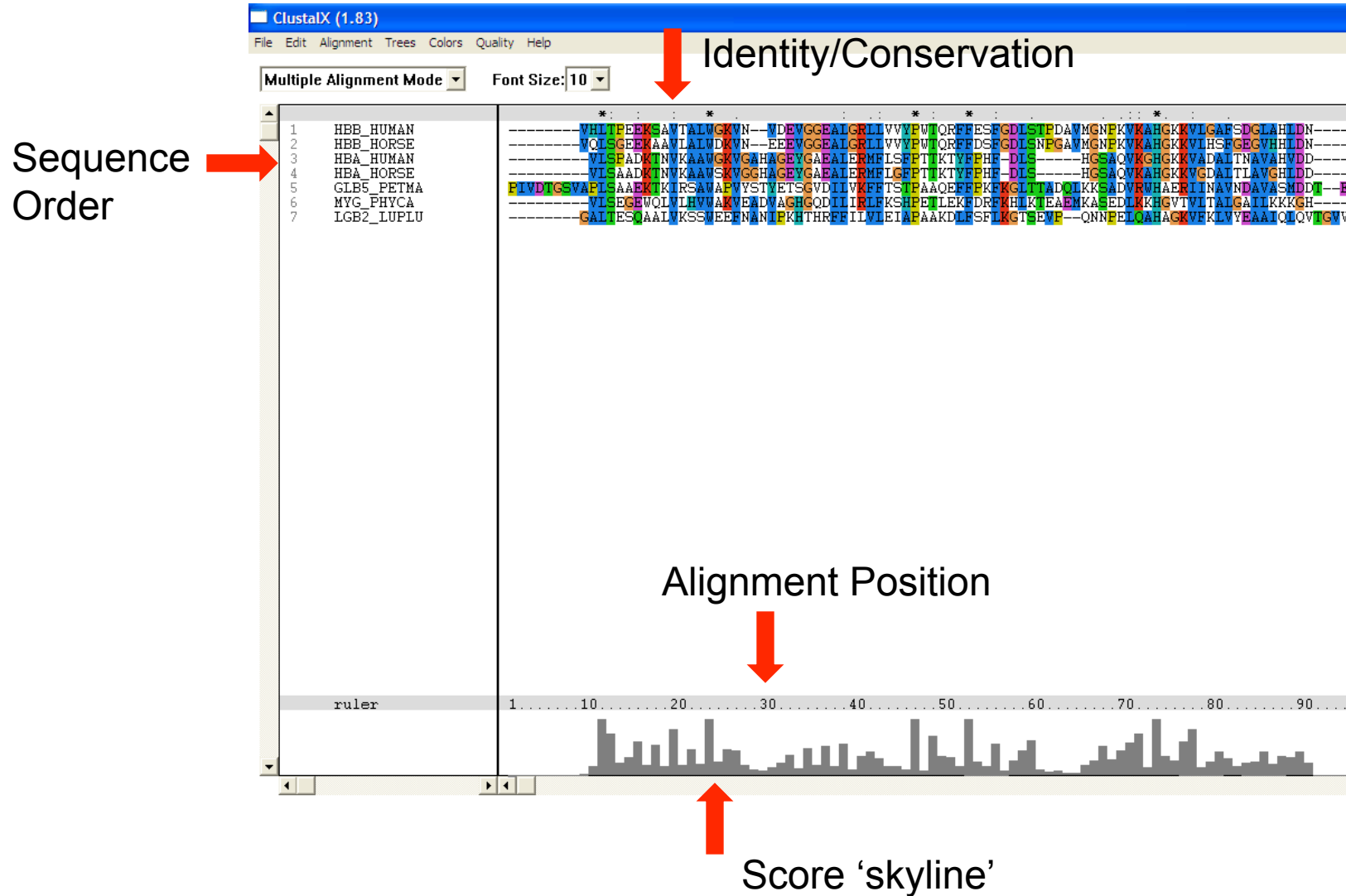
Load the sequences – globin.pep



Alignment > Do Complete Alignment



Examine Alignment



Alignment Menu

The screenshot shows the ClustalX (1.83) software interface. The 'Alignment' menu is open, displaying several options. Three red arrows point to the following items:

- Do Complete Alignment
- Alignment Parameters
- Output Format Options

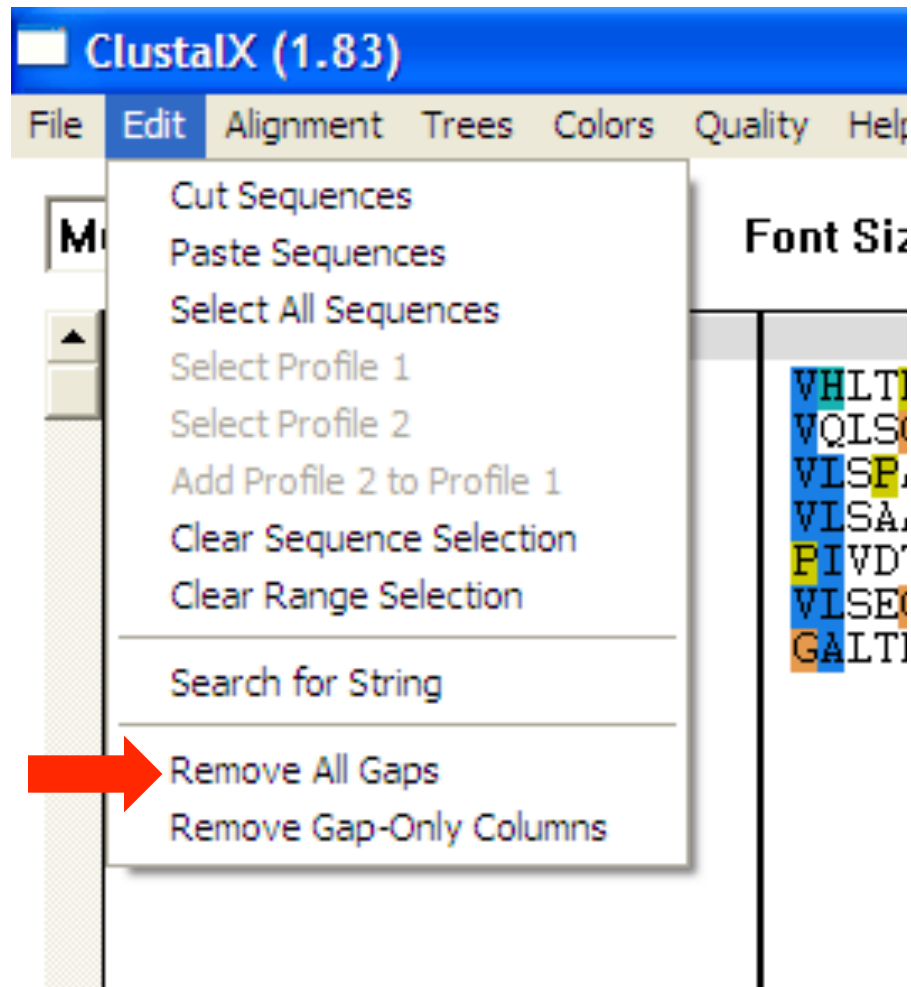
The 'Alignment Parameters' sub-menu is also visible, listing the following options:

- Reset New Gaps before Alignment
- Reset All Gaps before Alignment
- Pairwise Alignment Parameters
- Multiple Alignment Parameters
- Protein Gap Parameters
- Secondary Structure Parameters

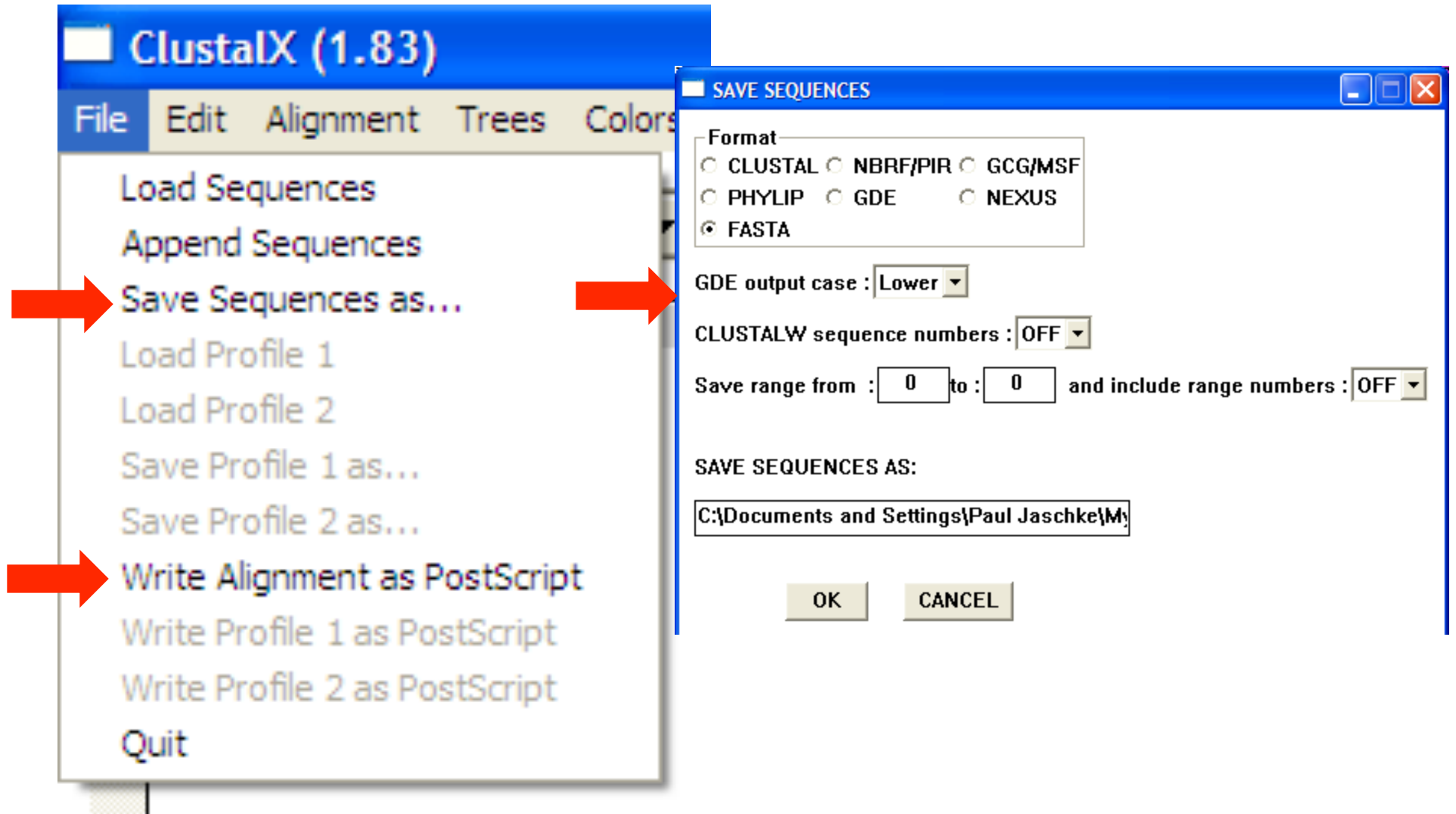
The background shows a sequence alignment window with a list of sequences (1-7) and a corresponding alignment view. The alignment view shows the following sequences:

```
1 EKSAVTALWGKVNVDDEVGGEALGRLLVVY
2 EKAAVLALWDKVNDEEVGGEALGRLLVVY
3 KTNVKAAWGKVGAHAGEYGAEALERMFLS
4 KTNVKAAWSKVGGHAGEYGAEALERMFLG
5 IQLVLIHVAWAKVEADVAGHGQDILIRLFKS
6 SVAFI SAAEKTKIRSAWAFVYSTYETISGV
7 DAALVKSSWEEFNANIPKHTRFFILVLE
```

Edit Menu



File Menu



Phylogenetic Tree Building

Trees Menu

The screenshot shows the ClustalX (1.83) interface. The 'Trees' menu is open, displaying options: 'Draw N-J Tree', 'Bootstrap N-J Tree', 'Exclude Positions with Gaps' (checked), 'Correct for Multiple Substitutions' (checked), 'Save Log File', and 'Output Format Options'. A red arrow points to the 'Multiple Alignment' label, and another red arrow points to the 'Exclude Positions with Gaps' option. The main window displays a sequence alignment for seven sequences: 1 (IPEEKSAVTALWGKVN--VDEVGGEALGRLLVVVFWIQRFHESHFGDLSI), 2 (SGEEKAAVLAALWVKVN--EEVGGGEALGRLLVVVFWIQRFHDSFGDLSN), 3 (SPADKTNVKAAGKVGGAHAGEYGAEALERMFISFPTTKTYFPHF-DLS-), 4 (SAADKTNVKAAWSKVGGAHAGEYGAEALERMFLGFPTTKTYFPHF-DLS-), 5 (SAAEKTKIRSAWAPVYSTVETSQVDIIVKFPTSIPAAQEFFPKFKGLT), 6 (VLSGEWQIVLHVWAKVEADVAGHGQDILIRLFKSHPEILEKEDRFKHLKI), and 7 (GALTESQAALVKSSWEFNANIPKHTHRFFILVLEIAPAAKDLFSFLKGTSE). A 'DRAW TREE' dialog box is open in the foreground, with the text 'SAVE PHYLP TREE AS : id Settings\joanne\Desktop\clustal\globin.ph' and 'OK' and 'CANCEL' buttons. At the bottom of the ClustalX window, a status bar reads 'CLUSTAL-Alignment file created []'.

Use NJPlot to visualize the trees

<http://pbil.univ-lyon1.fr/software/njplot.html>

Google: 'njplot'

NJplot

NEW: NJplot plots trees in PDF and PostScript formats (not for MacOS).

NEW: NJplot allows to open several tree windows.

NJplot is a tree drawing program able to draw any phylogenetic tree expressed in the [Newick](#) phylogenetic tree format (e.g., the format used by the PHYLIP package). NJplot roots the unrooted trees obtained from parsimony, distance or maximum likelihood tree-building methods.


A screen shot of the main window of njplot is available [here](#).

Use of NJPlot

Any rooting of the unrooted tree can be interactively specified using the mouse. NJplot also allows zooming, branch swapping, display of bootstrap scores and printing in the PDF format. NJplot can be used as a graphical extension of any package of phylogenetic programs which employs the standard tree format for storing trees (i.e., with most such packages).

Download NJplot

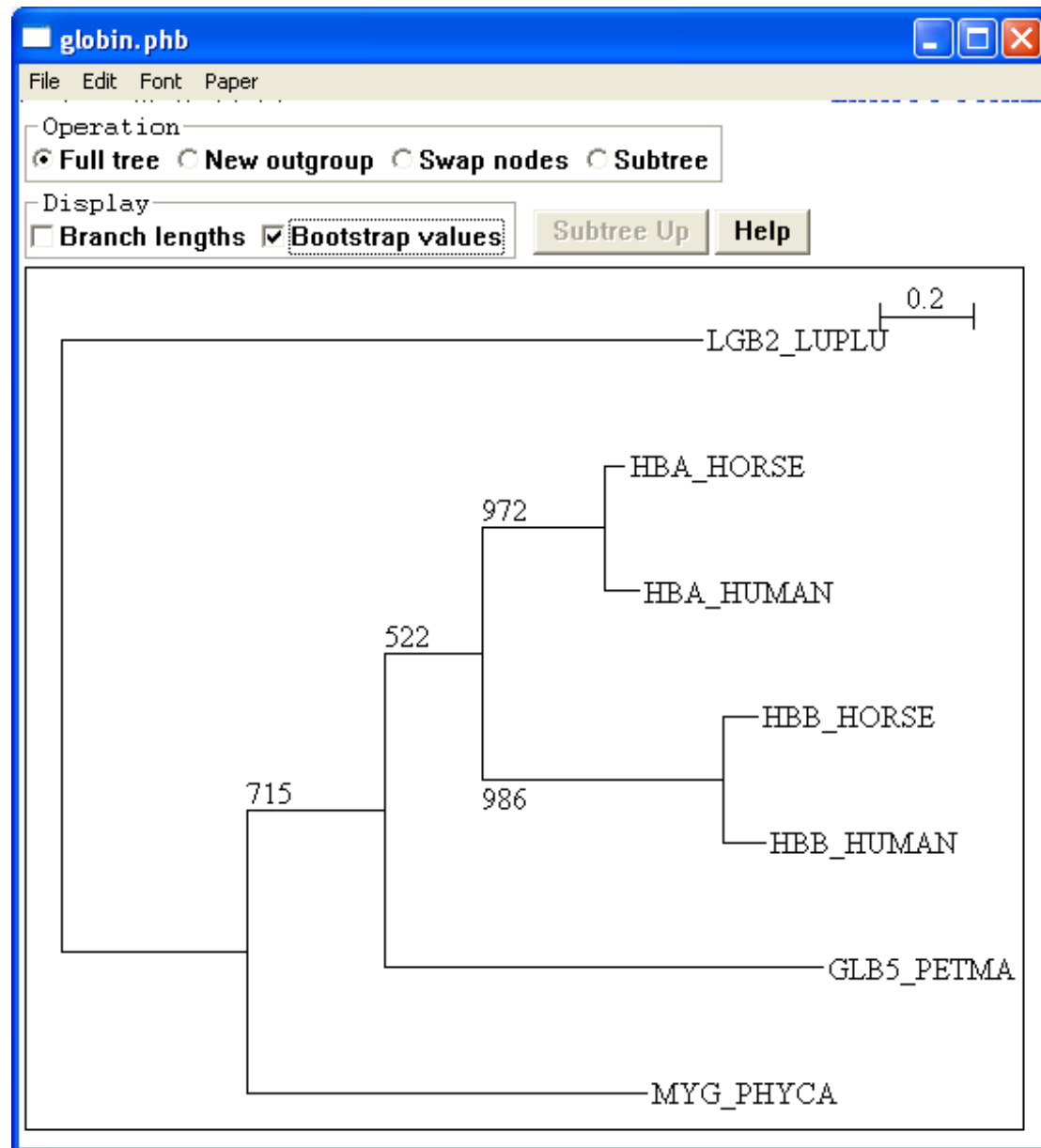
Executables and full [source code](#) can be downloaded through our [FTP server](#). You may also use the following table to directly access the version corresponding to your computer.

 MacOS X	 Linux on PC in case of font problem	 MS Windows	 Sun-Solaris	 HP Alpha
 MacOS (8,9)				

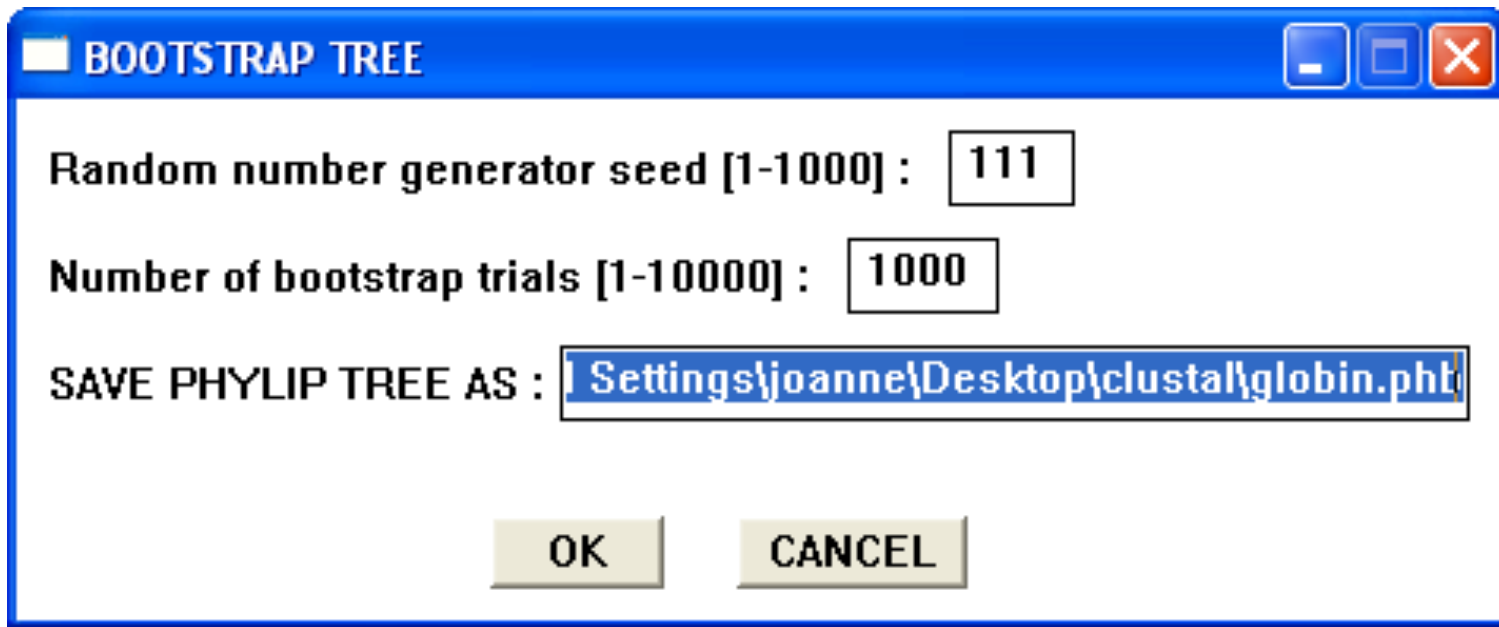
Extract njplotWIN95.exe and run njplot.exe



NJPlot



Trees > Bootstrap N-J Tree



1. Open tree file (.ph)
2. Open bootstrap file (.phb)
3. Can now display tree with both distances and/or bootstrap values.
4. Can save tree as **.pdf** for later use in poster.

Links

- ClustalX

<http://bips.u-strasbg.fr/en/Documentation/ClustalX/>

- The web version: ClustalW

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

- NJ-Plot

<http://pbil.univ-lyon1.fr/software/njplot.html>

- Tree View (software for tree visualization)

<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>