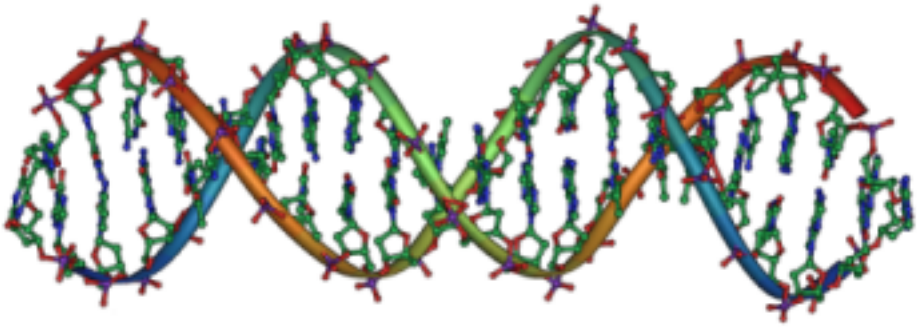
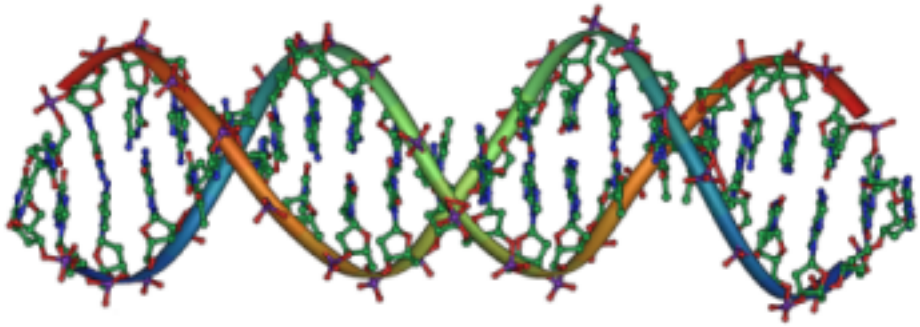
Bio Users Guide



6.2 Edition

Bio Users Guide :



6.2 Edition
Published May 07 2015

Table of Contents

List of Tables

Preface

Bio-Informatics is the use of techniques from applied mathematics, informatics, statistics, and computer science to solve biological problems. Major research efforts in the field include sequence alignment, gene finding, genome assembly, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, and the modeling of evolution.

To address the requirements of these efforts, a wide spectrum of bio-informatics tools are available. These tools, while powerful, are packaged according to the individual tastes of the developers.

The Bio-informatics Roll is a collection of some of the most common bio-informatics tools that are being used by the community today. This roll is being developed in an attempt to standardize and ease packaging and installation of these tools.

Chapter 1. Overview

Table 1-1. Summary

Name	bio
Version	6.2
Maintained By	Rocks Group
Architecture	i386, x86_64
Compatible with Rocks®	6.2

The bio roll has the following requirements of other rolls. Compatability with all known rolls is assured, and all known conflicts are listed. There is no assurance of compatiblity with third-party rolls.

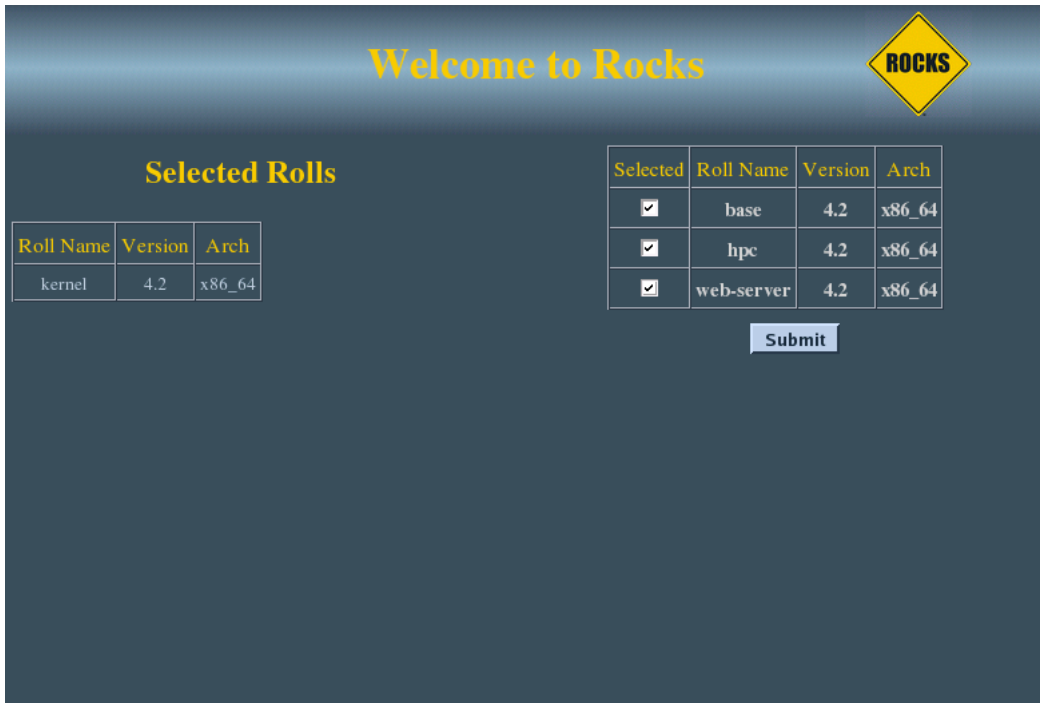
Table 1-2. Compatibility

Requires	Conflicts
Base	
HPC	
Kernel	
OS	
Web Server	

Chapter 2. Installing

2.1. On a New Server

The bio roll should be installed during the initial installation of your server (or cluster). This procedure is documented in section 3.2 of the Rocks® usersguide. You should select the bio roll from the list of available rolls when you see a screen that is similar to the one below.



2.2. On an Existing Server

The bio Roll may also be added onto an existing server (or frontend). For sake of discussion, assume that you have an iso image of the roll called `bio.iso`. The following procedure will install the Roll, and after the server reboots the Roll should be fully installed and configured.

```
$ su - root
# rocks add roll bio.iso
# rocks enable roll bio
# cd /export/rocks/install
# rocks create distro
# rocks run roll bio | bash
# init 6
```

Chapter 3. Using

3.1. List of packages present in the Bio Roll

The Bio Roll contains a suite of Bio-informatics applications, most commonly in use by the bio-informatics community. The list of applications is as follows:

- HMMER - http://hmmer.janelia.org/
- NCBI BLAST - From National Center for Biotechnology Information - www.ncbi.nlm.nih.gov/BLAST/[2]
- MpiBLAST - From Los Alamos National Laboratory - http://mpiblast.lanl.gov/
- biopython - www.biopython.org
- ClustalW - From the European BioInformatics Institute - http://www.ebi.ac.uk/clustalw/
- MrBayes - From School of Computational Science at the Florida State University - http://mrbayes.csit.fsu.edu/
- T_Coffee - From Information Genomique et Structurale at Centre National de la Recherche Scientifique - The T-Coffee Home Page[7]
- Emboss - From European Molecular Biology Institute - http://emboss.sourceforge.net/
- Phylip - From the Dept. of Biology at the University of Washington - http://evolution.genetics.washington.edu/phylip.html
- fasta - From the University of Virginia - http://fasta.bioch.virginia.edu/
- Glimmer - From Center for Bioinformatics and Computational Biology at the University of Maryland - http://www.cbcb.umd.edu/software/glimmer/
- TIGR Assembler - From the J. Craig Venter Institute - http://www.jcvi.org/cms/research/software/
- All the perl utilities mentioned below are from CPAN
- perl-bioperl
- perl-bioperl-ext
- perl-bioperl-run
- perl-bioperl-db

All the packages that appear below are dependencies and are already present in the base and OS Rolls. They are installed automatically during system installation.

foundation-python	flex	readline-devel
foundation-python-extras	xorg-x11-devel	gd
ReportLab	readline	gd-devel

3.2. HMMER

3.2.1. About

HMMER is an implementation of profile HMM methods for sensitive database searches using multiple sequence

alignments as queries.

The version of HMMER that is distributed with this version of Rocks was obtained from here[11]. The version as of code freeze is v2.3.2 and is distributed under the GNU General Public License v2.0.

3.2.2. Usage

HMMER is setup in the /opt/bio/hmmer directory. The HMMER execution environment is setup automatically by the login scripts. The environment contains HMMER_DB variable which points to the directory containing the hmmer databases. By default, this is set to $HOME/bio/hmmer/db/.

HMMER has many modes of execution. For a description of all the executables that come with HMMER, please refer to the current HMMER online userguide[12]. This guide is also available on your rocks installation at /opt/bio/hmmer/Userguide.pdf

There is also a tutorial available on your cluster at /opt/bio/hmmer/tutorial/ . The description of how to use the tutorial is given in the Userguide.pdf file.

3.3. NCBI BLAST

3.3.1. About

BLAST, or Basic Local Alignment Search Tool, is a collection of tools that are used to search for and find regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases, and calculates the statistical significance of the matches. This software suite has been released free to the public by the National Centre for Biotechnology Information.

3.3.2. Usage

BLAST can be used for protein-protein comparisons or nucleotide-nucleotide comparisons. Before an example of the usage is presented, we must first define some environmental variables.

- $BLASTDB - This is the variable which points to the Blast Database. This is set to $HOME/bio/ncbi/db/. This directory should contain the databases that you would want to search. BLAST, by default, checks this location and the current working directory for the presence of the databases. This variable is set during login by system login scripts , and may be changed by the user to point to her preferred location in her startup scripts.
- $BLASTMAT - This variable points to the location where the BLAST scoring matricies are present. It is set to /opt/bio/ncbi/data. Again, they may be changed to point to a desired location on a per-user basis.

BLAST requires the presence of 2 datasets. One dataset is the input sequence that you want to search for, and the other dataset is the database that you want to search against.

Use the following procedure to run blast

- Download a BLAST database that you want to run the comparison against. The databases can be obtained from the NCBI ftp site at ftp://ftp.ncbi.nlm.nih.gov/blast/db/.

The databases available on the site mentioned above are pre-formatted.

It is recommended that the blast databases be stored at the $BLASTDB location.

Visit ftp://ftp.ncbi.nlm.nih.gov/blast/db/ in your browser to see a list of available preformatted databases.

Download one of these on to your cluster using wget.

```
[nostromo@xxx ~]$ wget -q ftp://ftp.ncbi.nlm.nih.gov/blast/db/nt.08.tar.gz
[nostromo@xxx ~]$ gunzip -c nt.08.tar.gz | ( cd $BLASTDB/ && tar -xf -)
```

- The above method downloads a formatted database, and untars it into $BLASTDB.

  Unformatted databases can be obtained in FASTA format at ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/[15].

  Visit ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/[16] in your web browser

If you've downloaded the databases from ftp://ftp.ncbi.nlm.nih.gov/blast/db/, then DO NOT run formatdb.

Run the formatdb command to format the database to the BLAST format. For this example, we'll use the Drosophila Melanogaster (fruitfly) nucleotide database

```
[nostromo@xxx ~]$ cd $BLASTDB
[nostromo@xxx ~]$ wget -q ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/drosoph.nt.gz
[nostromo@xxx ~]$ gunzip drosoph.nt.gz
[nostromo@xxx ~]$ formatdb -p F -V T -i drosoph.nt
[nostromo@xxx ~]$ ls drosoph.nt*
drosoph.nt  drosoph.nt.nhr  drosoph.nt.nin  drosoph.nt.nsq
[nostromo@xxx ~]$ cd $HOME
```

- After the database is formatted, create a test input file.

```
[nostromo@xxx ~]$ cat > test.txt
>Test
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGC
TTCTGAACTGGTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAA
TATAGGCATAGCGCACAGACAGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACC
ATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAG
CCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGTAACGAGGTAACAACCATGCGAGTGTTGAA
GTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTTGCCGATATTCTGGAAAGCAATGCC
AGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGGTGGCGATGATTG
AAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCCGAACTTTT
```

- Run the blastall program on the test input against the formatted database.

```
[nostromo@xxx ~]$ blastall --help
```

gives a list of all the options that you can use to run the blastall program.

```
[nostromo@xxx ~]$  blastall -d drosoph.nt -p blastn -i test.txt
BLASTN 2.2.18 [Mar-02-2008]


Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.
```

```
Query= Test
         (560 letters)

Database: drosoph.nt
           1170 sequences; 122,655,632 total letters

Searching.....................................................done



                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

gi|10729531|gb|AE002936.2|AE002936 Drosophila melanogaster genom...    36   0.86
gi|10728232|gb|AE003493.2|AE003493 Drosophila melanogaster genom...    36   0.86
gi|10726497|gb|AE003698.2|AE003698 Drosophila melanogaster genom...    36   0.86
gi|10726398|gb|AE003681.2|AE003681 Drosophila melanogaster genom...    36   0.86
gi|10729308|gb|AE002665.2|AE002665 Drosophila melanogaster genom...    34   3.4
gi|10729264|gb|AE002615.2|AE002615 Drosophila melanogaster genom...    34   3.4
gi|7298233|gb|AE003648.1|AE003648 Drosophila melanogaster genomi...    34   3.4
gi|7297628|gb|AE003628.1|AE003628 Drosophila melanogaster genomi...    34   3.4
gi|10728546|gb|AE003447.2|AE003447 Drosophila melanogaster genom...    34   3.4
gi|7290819|gb|AE003441.1|AE003441 Drosophila melanogaster genomi...    34   3.4
gi|10728461|gb|AE003431.2|AE003431 Drosophila melanogaster genom...    34   3.4
gi|10728241|gb|AE003495.2|AE003495 Drosophila melanogaster genom...    34   3.4
gi|7292554|gb|AE003484.1|AE003484 Drosophila melanogaster genomi...    34   3.4
gi|10727872|gb|AE003525.2|AE003525 Drosophila melanogaster genom...    34   3.4
gi|10727399|gb|AE003587.2|AE003587 Drosophila melanogaster genom...    34   3.4
gi|10727114|gb|AE003673.2|AE003673 Drosophila melanogaster genom...    34   3.4
gi|10726705|gb|AE003740.2|AE003740 Drosophila melanogaster genom...    34   3.4
```

The above example shows how to search for the test input in a drosophila nucleotide database, and a snippet of the output file.

3.3.3. Running Blast with SGE

This section gives a very simple example of running BLAST through the provided batch system SGE.

- Create a simple submission script called blast_sge.sh containing the following -

```
#!/bin/bash
#
#$ -cwd
#$ -S /bin/bash
#$ -j y

export BLASTDB=$HOME/bio/ncbi/db/
export BLASTMAT=/opt/bio/ncbi/data/

/opt/bio/ncbi/bin/blastall -d drosoph.nt \
        -p blastn -i $HOME/test.txt      \
        -o $HOME/result.txt
```

- Run

```
[nostromo@xxx ~]$ qsub blast_sge.sh
Your job 10 ("blast_sge.sh") has been submitted
```

- The output of the Blast job is similar to the one given above and will be stored in $HOME/result.txt

3.3.4. Further Information

For further information about BLAST and its usage, please refer to the following sources

- THE NCBI Blast website - http://www.ncbi.nlm.nih.gov/BLAST/[17]
- BLAST Help page on your cluster BLAST Help Page[18]
- BLAST Program selection Guide - http://www.ncbi.nlm.nih.gov/blast/BLAST_guide.pdf[19]

3.4. ClustalW

3.4.1. About

ClustalW is a multiple sequence alignment program. The version included with this distribution is v2.0.12.

3.4.2. Using ClustalW

ClustalW can be run at the command line as

```
[nostromo@xxx ~]$ clustalw2


 ***********************************************************
 ******** CLUSTAL 2.0.12 Multiple Sequence Alignments  ********
 ***********************************************************


     1. Sequence Input From Disc
     2. Multiple Alignments
     3. Profile / Structure Alignments
     4. Phylogenetic trees

     S. Execute a system command
     H. HELP
     X. EXIT (leave program)


Your choice:
```

Choosing the option 'H' brings up the help on clustalW.

3.4.3. Further Information

Further information on the usage of ClustalW can be obtained from clustalw.doc(MS Word Document) available at /opt/bio/clustalw/doc/clustalw.doc on the frontend of your cluster.

3.5. EMBOSS

3.5.1. About

EMBOSS is the European Molecular Biology Open Software Suite, a set of tools that are used for sequence analysis by the Molecular Biology community (EMBnet).

The version of EMBOSS included with this version of Rocks is 6.1.0

3.5.2. Further Information

Information about using EMBOSS is available at http://emboss.sourceforge.net/. You may also register at their mailing list here[21].

3.6. Glimmer

3.6.1. About

Glimmer is a system for finding genes in microbial DNA, especially the genomes of bacteria, archaea, and viruses. Glimmer was developed at the Centre for BioInformatics and Computational Biology. The version that is distributed with Rocks is Glimmer v3.02.

3.6.2. Using Glimmer

Glimmer is installed at /opt/bio/glimmer/. Glimmer is run in 2 stages.

- Glimmer is trained on a particular training set of similar species to recognize genes
- Glimmer is then run on an input DNA sequence to find genes

3.6.3. Further Information

Further information about the usage of Glimmer can be found in the release notes of the software, available here[22]. This file is also available on the frontend of your cluster at /opt/bio/glimmer/glim302notes.pdf

3.7. Fasta

3.7.1. About Fasta

FASTA is a program used to search in large Protein or DNA sequence data banks. It was developed at the University of Virginia by William R. Pearson, and D.J. Lippman.

3.7.2. Usage

FASTA is installed in /opt/bio/fasta/. FASTA is run in a similar manner to NCBI Blast.

- First create a test query file

```
[nostromo@xxx ~]$ cat > test.txt
>Test
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGC
TTCTGAACTGGTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAA
TATAGGCATAGCGCACAGACAGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACC
ATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAG
CCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGTAACGAGGTAACAACCATGCGAGTGTTGAA
GTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTTGCCGATATTCTGGAAAGCAATGCC
AGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGGTGGCGATGATTG
AAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCCGAACTTTT
```

- The next step is to search for this against a database sequence. For this, we can download a DNA or protein sequence database or use the ones that are provided by the program. For this example, we will use the ones present along with the fasta program in /opt/bio/fasta/.

```
[nostromo@xxx ~]$ fasta35
# fasta35
FASTA searches a protein or DNA sequence data bank
 version 35.04 Oct. 7, 2008
Please cite:
 W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

 test sequence file name: test.txt
 library file name: drosoph.nt
 ktup? (1 to 6) [6]
Query: test.txt
  1>>>Test - 560 nt
Library: drosoph.nt
.......  Done!
 122655632 residues in  1170 sequences

       opt      E()
< 20     0     0:
  22     0     0:           one = represents 3 library sequences
  24     0     0:
  26     0     0:
  28     0     0:
  30     3     2:*
  32    12     9:==*=
  34    37    23:=======*=====
  36    59    48:===============*====
  38    90    79:==========================*===
```

```
  40   110   110:==================================*
  42   133   135:===========================================*
  44   147   149:================================================*
  46   151   151:=================================================*
  48   129   145:============================================     *
  50   131   132:===========================================*
  52   102   116:==================================    *
  54    92    99:================================ *
  56    80    83:============================*
  58    68    68:======================*
  60    43    55:===============   *
  62    44    44:==============*
  64    42    35:===========*==
  66    30    28:=========*
  68    25    22:=======*=
  70    20    17:=====*=
  72    18    13:====*=
  74    14    10:===*=
  76     7     8:==*
  78     7     6:=*=
  80     9     5:=*=
  82     3     4:=*
  84     0     3:*
  86     0     2:*
  88     2     2:*          inset = represents 1 library sequences
  90     1     1:*
  92     0     1:*         :*
  94     0     1:*         :*
  96     2     1:*         :*=
  98     0     0:          *
 100     0     0:          *
 102     0     0:          *
 104     0     0:          *
 106     0     0:          *
 108     0     0:          *
 110     0     0:          *
 112     0     0:          *
 114     0     0:          *
 116     0     0:          *
 118     0     0:          *
>120     0     0:          *
122902592 residues in  1611 sequences
Statistics:  Expectation_n fit: rho(ln(x))= 7.6751+/-0.00204; mu= 6.7759+/- 0.231
 mean_var=233.8700+/-93.821, 0's: 0 Z-trim: 0  B-trim: 0 in 0/53
 Lambda= 0.083866
 Kolmogorov-Smirnov  statistic: 0.0247 (N=27) at  38
Algorithm: FASTA (3.5 Sept 2006) [optimized]
Parameters: +5/-4 matrix (5:-4) ktup: 6
 join: 52, opt: 37, open/ext: -12/-4, width:  16
 Scan time: 10.680
Enter filename for results []:  How many scores would you like to see? [20]

The best scores are:                                      opt bits E(1611)
gi|10727961|gb|AE003541.2|AE003541 Drosophila  (265536) [r]  171 36.0       1
gi|10728546|gb|AE003447.2|AE003447 Drosophila  (304085) [f]  171 36.0       1
gi|7290382|gb|AE003426.1|AE003426 Drosophila m (300193) [f]  159 34.5     2.8
gi|7290880|gb|AE003443.1|AE003443 Drosophila m (302357) [f]  157 34.3     3.3
```

```
gi|10727731|gb|AE003838.2|AE003838 Drosophila  (263411) [r]  149 33.3     6.4
gi|7291133|gb|AE003450.1|AE003450 Drosophila m (300732) [f]  148 33.2     6.9
gi|7300931|gb|AE003741.1|AE003741 Drosophila m (233313) [r]  151 33.2     7.1
gi|10726402|gb|AE003682.2|AE003682 Drosophila  (224400) [f]  147 33.1     7.5
gi|10728339|gb|AE003512.2|AE003512 Drosophila  (301457) [f]  147 33.1     7.5
gi|10728273|gb|AE003500.2|AE003500 Drosophila  (327446) [f]  145 32.8     8.9
gi|10726452|gb|AE003691.2|AE003691 Drosophila  (226773) [f]  145 32.8     8.9
gi|10727164|gb|AE003603.2|AE003603 Drosophila  (294914) [r]  144 32.6      10
gi|7290252|gb|AE003423.1|AE003423 Drosophila m (291976) [r]  144 32.6      10
gi|10727489|gb|AE003803.2|AE003803 Drosophila  (282567) [r]  143 32.6      10
gi|10727489|gb|AE003803.2|AE003803 Drosophila  (282567) [r]  143 32.5      11
gi|10727339|gb|AE003577.2|AE003577 Drosophila  (267662) [f]  142 32.3      13
gi|7292734|gb|AE003488.1|AE003488 Drosophila m (302797) [f]  140 32.2      13
gi|7298684|gb|AE003667.1|AE003667 Drosophila m (263704) [r]  139 31.9      17
gi|10727995|gb|AE003546.2|AE003546 Drosophila  (281602) [f]  137 31.9      17
gi|10728551|gb|AE003448.2|AE003448 Drosophila  (310364) [f]  137 31.9      18
 More scores? [0]
 Display alignments also? (y/n) [n]


560 residues in 1 query   sequences
122655632 residues in 1170 library sequences
 Scomplib [35.04]
 start: Wed Dec 10 19:45:41 2008 done: Wed Dec 10 19:46:04 2008
 Total Scan time: 10.680 Total Display time:  0.000

Function used was FASTA [version 35.04 Oct. 7, 2008]
```

3.7.3. Further Information

Further information about the usage of fasta can be obtained from /opt/bio/fasta/fasta3x.doc present on the frontend of your installation.

More information is also available at the FASTA home page[23].

For support, you are encouraged to join the FASTA mailing list at
http://list.mail.virginia.edu/mailman/listinfo/fasta_list

3.8. MrBayes

3.8.1. About

MrBayes is a program used for bayesian inference of phylogeny. MrBayes is cowritten by John Huelsenbeck and Fredrik Ronquist.

The version of MrBayes included with this version of Rocks is MPI enabled, and can be used in either parallel or serial modes of execution.

3.8.2. Usage

MrBayes uses the NEXUS file format for input. To use MrBayes in interactive mode, just type mb at the command line

```
[nostromo@xxx mrbayes]$ mb
MrBayes v3.1.2

                      (Bayesian Analysis of Phylogeny)

                            (Parallel version)
                         (1 processors available)

                                     by

                    Fredrik Ronquist and John P. Huelsenbeck

                          School of Computational Science
                              Florida State University
                               ronquist@csit.fsu.edu

                       Section of Ecology, Behavior and Evolution
                             Division of Biological Sciences
                          University of California, San Diego
                                johnh@biomail.ucsd.edu

                  Distributed under the GNU General Public License

                   Type "help" or "help <command>" for information
                         on the commands that are available.


MrBayes >
```

To use MrBayes in the parallel version, you'll need to use it in non-interactive mode. It can be invoked as shown.

```
[nostromo@xxx ~]$ /opt/openmpi/bin/mpirun -np 4 /opt/bio/mrbayes/mb /opt/bio/mrbayes/primates.nex
[nostromo@xxx ~]$ cat log.txt
                                  MrBayes v3.1.2

                      (Bayesian Analysis of Phylogeny)

                            (Parallel version)
                         (4 processors available)

                                     by

                    John P. Huelsenbeck and Fredrik Ronquist

                       Section of Ecology, Behavior and Evolution
                             Division of Biological Sciences
                          University of California, San Diego
                                johnh@biomail.ucsd.edu

                          School of Computational Science
                              Florida State University
                               ronquist@csit.fsu.edu
```

```
                Distributed under the GNU General Public License

                 Type "help" or "help <command>" for information
                       on the commands that are available.



   Executing file "/opt/bio/mrbayes/primates.nex"
   UNIX line termination
   Longest line length = 915
   Parsing file
   Expecting NEXUS formatted file
   Reading data block
      Allocated matrix
      Matrix has 12 taxa and 898 characters
      Data is Dna
      Data matrix is not interleaved
      Gaps coded as -
      Setting default partition (does not divide up characters).
      Taxon  1 -> Tarsius_syrichta
      Taxon  2 -> Lemur_catta
      Taxon  3 -> Homo_sapiens
      Taxon  4 -> Pan
      Taxon  5 -> Gorilla
      Taxon  6 -> Pongo
      Taxon  7 -> Hylobates
      Taxon  8 -> Macaca_fuscata
      Taxon  9 -> M_mulatta
      Taxon 10 -> M_fascicularis
      Taxon 11 -> M_sylvanus
      Taxon 12 -> Saimiri_sciureus
      Setting output file names to "/opt/bio/mrbayes/primates.nex.run<i>.<p/t>"
      Successfully read matrix
   Exiting data block
   Reached end of file

   Tasks completed, exiting program because mode is noninteractive
   To return control to the command line after completion of file processing,
   set mode to interactive with 'mb -i <filename>' (i is for interactive)
   or use 'set mode=interactive'

[nostromo@xxx ~]$
```

3.8.3. Further Information

A wealth of information about MrBayes is available at the MrBayes Home Page[24] and at the MrBayes Wiki[25]

3.9. Phylip

3.9.1. About

Phylip - Phylogeny Inference Package - is a package of programs for inferring phylogenies or evolutionary trees. The version distributed with Rocks is v3.68.

3.9.2. Further Information

Furtherinformation about Phylip is available at the Phylip home page[26].

3.10. T_Coffee

3.10.1. About

T_Coffee is a multiple sequence alignment package. The version included with this distribution of Rocks is v8.14

3.10.2. Usage

T-coffee is used for standard alignments and alignment combinations. It is installed at /opt/bio/tcoffee/ on the Rocks distribution. To use T-Coffee, just type t_coffee at the command line for a list of all possible parameters that can be used. T-coffee recognizes formats such as fasta, clustalw, blast, etc. Example input files are available at /opt/bio/tcoffee/example/

A simple sequence alignment example is shown below about. It is run against a sample fasta file present in the example directory. Parts of the output are deleted for the sake of brevity. Where missing, output is substituted by ellipses (.....)

```
[nostromo@xxx ~]$ t_coffee /opt/bio/tcoffee/example/sample_aln2.fasta

PROGRAM: T-COFFEE (Version_8.14)
-full_log       S       [0]
-run_name       S       [0]
-mem_mode       S       [0]     mem
-extend         D       [1]     1
-extend_mode    S       [0]     very_fast_triplet
-max_n_pair     D       [0]     10
-seq_name_for_quadruplet        S       [0]     all
-compact        S       [0]     default
-clean          S       [0]     no
-do_self        FL      [0]     0
-do_normalise   D       [0]     1000
-template_file  S       [0]
-template_mode  S       [0]
-remove_template_file   D       [0]     0
-profile_template_file  S       [0]
-in             S       [0]
-seq            S       [1]     /opt/bio/tcoffee/example/sample_aln2.fasta
-aln            S       [0]
```

```
-method_limits  S       [0]
-method         S       [0]
-lib            S       [0]
-profile        S       [0]
-profile1       S       [0]
-profile2       S       [0]
-pdb            S       [0]
-relax_lib      D       [0]     1
-filter_lib     D       [0]     0
-shrink_lib     D       [0]     0
-out_lib        W_F     [0]     no
-out_lib_mode   S       [0]     primary
-lib_only       D       [0]     0
-outseqweight   W_F     [0]     no
-dpa            FL      [0]     0
-seq_source     S       [0]     ANY
-cosmetic_penalty       D       [0]     0
-gapopen        D       [0]     0
-gapext         D       [0]     0
-fgapopen       D       [0]     0
-fgapext        D       [0]     0
-nomatch        D       [0]     0
-newtree        W_F     [0]     default
-tree           W_F     [0]     NO
-usetree        R_F     [0]
-tree_mode      S       [0]     nj
-distance_matrix_mode   S       [0]     ktup
-distance_matrix_sim_mode       S       [0]     idmat_sim1
-quicktree      FL      [0]     0
-outfile        W_F     [0]     default
-maximise       FL      [1]     1
-output         S       [0]     aln     html
-infile         R_F     [0]
-matrix         S       [0]     default
-tg_mode        D       [0]     1
-profile_mode   S       [0]     cw_profile_profile
-profile_comparison     S       [0]     profile
-dp_mode        S       [0]     linked_pair_wise
-ktuple         D       [0]     1
-ndiag          D       [0]     0
-diag_threshold D       [0]     0
-diag_mode      D       [0]     0
-sim_matrix     S       [0]     vasiliky
-transform      S       [0]
-outorder       S       [0]     input
-inorder        S       [0]     aligned
-seqnos         S       [0]     off
-case           S       [0]     keep
-cpu            D       [0]     0
-maxnseq        D       [0]     1000
-maxlen         D       [0]     -1
-weight         S       [0]     default
-seq_weight     S       [0]     t_coffee
-align          FL      [1]     1
-mocca          FL      [0]     0
-domain         FL      [0]     0
-start          D       [0]     0
```

```
-len            D       [0]     0
-scale          D       [0]     0
-mocca_interactive      FL      [0]     0
-method_evaluate_mode   S       [0]     default
-evaluate_mode  S       [0]     t_coffee_fast
-get_type       FL      [0]     0
-clean_aln      D       [0]     0
-clean_threshold        D       [1]     1
-clean_iteration        D       [1]     1
-clean_evaluate_mode    S       [0]     t_coffee_fast
-extend_matrix  FL      [0]     0
-prot_min_sim   D       [0]     0
-prot_max_sim   D       [90]    90
-prot_min_cov   D       [0]     0
-pdb_min_sim    D       [35]    35
-pdb_max_sim    D       [100]   100
-pdb_min_cov    D       [50]    50
-pdb_blast_server       W_F     [0]     EBI
-blast          W_F     [0]
-blast_server   W_F     [0]     EBI
-pdb_db         W_F     [0]     pdb
-protein_db     W_F     [0]     uniprot
-method_log     W_F     [0]     no
-struc_to_use   S       [0]
-cache          W_F     [0]     use
-align_pdb_param_file   W_F     [0]     no
-align_pdb_hasch_mode   W_F     [0]     hasch_ca_trace_bubble
-external_aligner       S       [0]     NO
-msa_mode       S       [0]     tree
-one2all        S       [0]
-subset2all     S       [0]
-lalign_n_top   D       [0]     10
-iterate        D       [0]     0
-trim           D       [0]     0
-split          D       [0]     0
-trimfile       S       [0]     default
-split          D       [0]     0
-split_nseq_thres       D       [0]     0
-split_score_thres      D       [0]     0
-check_pdb_status       D       [0]     0
-clean_seq_name D       [0]     0
-seq_to_keep    S       [0]
-dpa_master_aln S       [0]
-dpa_maxnseq    D       [0]     0
-dpa_min_score1 D       [0]
-dpa_min_score2 D       [0]
-dpa_keep_tmpfile       FL      [0]     0
-dpa_debug      D       [0]     0
-multi_core     S       [0]     templates_jobs_relax_msa
-n_core         D       [0]     0
-lib_list       S       [0]
-prune_lib_mode S       [0]     5
-tip            S       [0]     one
-rna_lib        S       [0]
-no_warning     D       [0]     0
-run_local_script       D       [0]     0
-plugins        S       [0]     default
```

```
-proxy          S       [0]     unset
-email          S       [0]
-clean_overaln  D       [0]     0
-overaln_param  S       [0]
-overaln_mode   S       [0]
-overaln_model  S       [0]
-overaln_threshold      D       [0]     0
-overaln_target D       [0]     0
-overaln_P1     D       [0]     0
-overaln_P2     D       [0]     0
-overaln_P3     D       [0]     0
-overaln_P4     D       [0]     0
-exon_boundaries        S       [0]

INPUT FILES
        Input File (S) /opt/bio/tcoffee/example/sample_aln2.fasta  Format clustal_aln
        Input File (M) proba_pair

INPUT SEQUENCES: 6 SEQUENCES  [PROTEIN]
  Input File /opt/bio/tcoffee/example/sample_aln2.fasta Seq 1cms   Length  175 type PROTEIN Struc
  Input File /opt/bio/tcoffee/example/sample_aln2.fasta Seq 4pep   Length  174 type PROTEIN Struc
  Input File /opt/bio/tcoffee/example/sample_aln2.fasta Seq 4ape   Length  178 type PROTEIN Struc
  Input File /opt/bio/tcoffee/example/sample_aln2.fasta Seq 3app   Length  174 type PROTEIN Struc
  Input File /opt/bio/tcoffee/example/sample_aln2.fasta Seq 2apr   Length  178 type PROTEIN Struc
  Input File /opt/bio/tcoffee/example/sample_aln2.fasta Seq 1cms_1 Length  148 type PROTEIN Struc

COMPUTE PAIRWISE SIMILARITY [dp_mode: ] [distance_matrix_mode: ktup][Similarity Measure: idmat_si

        Seq: 1cms
        Seq: 1cms_1
        Seq: 2apr
        Seq: 3app
        Seq: 4ape
        Seq: 4pep

READ/MAKE LIBRARIES:[2]

        proba_pair [method]

        Multi Core Mode: 2 processors [subset]

                [Submit   Job][TOT=    8][100 %][ELAPSED TIME:    0 sec.]
MANUAL PENALTIES: gapopen=0 gapext=0

        Library Total Size: [6175]

Library Relaxation: Multi_proc [2]
                [Submit   Job][TOT= 3087][100 %][ELAPSED TIME:    0 sec.]
Total Relaxation: [6175]--->[5092] Entries

        #### File Type=     WEIGHT Format=  tc_weight Name= no | NOT PRODUCED [WARNING:T-COFFEE:V


WEIGHTED MODE:t_coffee

          1cms 1.00
        1cms_1 1.10
```

```
          2apr 1.00
          3app 0.96
          4ape 0.95
          4pep 0.99


MAKE GUIDE TREE
        [MODE=nj][DONE]

PROGRESSIVE_ALIGNMENT [Tree Based]

        Group    8: [Group    5 (   1 seq)] with [Group    4 (   1 seq)]-->[Score=  83][Len=  179
        Group    7: [Group    6 (   1 seq)] with [Group    1 (   1 seq)]-->[Score=  92][Len=  176
        Group    9: [Group    8 (   2 seq)] with [Group    3 (   1 seq)]-->[Score=  74][Len=  186
        Group   10: [Group    9 (   3 seq)] with [Group    7 (   2 seq)]-->[Score=  77][Len=  186
        Group   11: [Group    2 (   1 seq)] with [Group   10 (   5 seq)]-->[Score=  24][Len=  209

CLUSTAL FORMAT for T-COFFEE Version_8.14 [http://www.tcoffee.org] [MODE:  ], CPU=0.15 sec, SCORE=

1cms            GE---VASVPLTNY------LDSQYFGKIYLGTPPQEFTVLFDTGSSDFWVPSIYCKSNA
4pep            -----IGDEPLENY------LDTEYFGTIGIGTPAQDFTVIFDTGSSNLWVPSVYCSSLA
4ape            S-TGSATTTPID-S------LDDAYITPVQIGTPAQTLNLDFDTGSSDLWVFSSETTASE
3app            AASGVATNTPTA--------NDEEYITPVTIGG--TTLNLNFDTGSADLWVFSTELPASQ
2apr            AG---VGTVPMTDY-----GNDIEYYGQVTIGTPGKKFNLDFDTGSSDLWIASTLCTN-C
1cms_1          Y-TGSLHWVPVTVQQYWQFTVDSVTISGVVVACEG-GCQAILDTGTSKLVGPSSD-----
                         *           *      : :.         :***::.:   *

1cms            CKNHQRFDPRKSSTFQ-NLGKPLSIHYGTGS-MQGILGYDTVTVSNIVDIQQTVGLSTQE
4pep            CSDHNQFNPDDSSTFE-ATSQELSITYGTGS-MTGILGYDTVQVGGISDTNQIFGLSETE
4ape            VDGQTIYTPSKSTTAKLLSGATWSISYGDGSSSSGDVYTDTVSVGGLTVTGQAVESAKKV
3app            QSGHSVYNPSATG-KE-LSGYTWSISYGDGSSASGNVFTDSVTVGGVTAHGQAVQAAQQI
2apr            GSGQTKYDPNQSSTYQ-ADGRTWSISYGDGSSASGILAKDNVNLGGLLIKGQTIELAKRE
1cms_1          ----------------------------------------------ILNIQQAIGATQNQ
                                                             :    * .  :

1cms            PGDVFTYAEFD--------GILGMAYPSLASEY-------SIPVFDNM-MNRHLVA----
4pep            PGSFLYYAPFD--------GILGLAYPSISASG-------ATPVFDNL-WDQGLVS----
4ape            SSSFTEDSTID--------GLLGLAFSTLNTVSPTQ----QKTFFDNA---KASLD----
3app            SAQFQQDTNND--------GLLGLAFSSINTVQPQS----QTTFFDTV---KSSLA----
2apr            AASFAS-GPND--------GLLGLGFDTITTVRG------VKTPMDNL-ISQGLIS----
1cms_1          YGEFDIDCDNLSYMPTVVFEINGKMYPLTPSAYTSQDQGFCTSGFQSENHSQKWILGDVF
                 ...               : *  :    :           . ::.    :  :

1cms            -QDLFSVYMDRN-G-QESMLTLGAIDPSY
4pep            -QDLFSVYLSSN-DDSGSVVLLGGIDSSY
4ape            -SPVFTADLGY---HAPGTYNFGFIDTTA
3app            -QPLFAVALKH---QQPGVYDFGFIDSSK
2apr            -RPIFGVYLGKAKNGGGGEYIFGGYDSTK
1cms_1          IREYYSVFDR--------ANNLVGLAKAI
                    : .             :     :



OUTPUT RESULTS
        #### File Type= GUIDE_TREE Format=     newick Name= sample_aln2.dnd
```

```
        #### File Type=        MSA Format=        aln Name= sample_aln2.aln
        #### File Type=        MSA Format=       html Name= sample_aln2.html


# TIP :See The Full Documentation on www.tcoffee.org
# TIP  1:  Get the most accurate protein alignments with: t_coffee <yourseq> -special_mode accura
# TIP  4:  -special_mode=expresso to fetch your structures automatically

# Command Line: t_coffee /opt/bio/tcoffee/example/sample_aln2.fasta  [PROGRAM:T-COFFEE]
# T-COFFEE Memory Usage: Current= 11.819 Mb, Max= 13.181 Mb
# T-COFFEE CPU Usage: 160 millisec
# Results Produced with T-COFFEE (Version_8.14)
# T-COFFEE is available from http://www.tcoffee.org
```

3.10.3. Further Information

Further information about t_coffee is available at -

- The T-coffee home page[27]
- On your cluster head node at /opt/bio/tcoffee/doc/
- T-Coffee Documentation[28]

3.11. TIGR Assembler v2

3.11.1. About

The TIGR Assembler is a tool to assemble large shotgun sequencing projects. The version included with this distribution of Rocks is v2

3.11.2. Usage

TIGR is used for assembling large shotgun DNA sequences. It is installed at /opt/bio/tigr on the Rocks Distribution. To use TIGR, just type TIGR_Assembler at the command line for a list of all possible parameters that can be used.

3.11.3. Further Information

Further information is available at the JCVI TIGR Assembler page[29]

3.12. MPI-Blast

3.12.1. About

MPI-Blast is a program from LANL[30] which parallelizes the NCBI Blast algorithms using Message Passing Interface library. The version of MPI-Blast included with Rocks is v1.5.0-pio patched and compiled against NCBI Blast 2.2.19.

3.12.2. Usage

MPI-Blast is used in a similar manner to NCBI-Blast. MPI-Blast uses the same variables that are available for NCBI-Blast.

There are 3 steps to running MPI-Blast.

- Download a FASTA database to $BLASTDB. For this example we will download the ecoli nucleotide database.

```
[nostromo@xxx ~]$ cd $BLASTDB
[nostromo@xxx ~]$ wget ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/ecoli.nt.gz
--17:06:23--  ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/ecoli.nt.gz
           => 'ecoli.nt.gz'
Resolving ftp.ncbi.nlm.nih.gov... 165.112.7.10
Connecting to ftp.ncbi.nlm.nih.gov|165.112.7.10|:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done.    ==> PWD ... done.
==> TYPE I ... done.  ==> CWD /blast/db/FASTA ... done.
==> PASV ... done.    ==> RETR ecoli.nt.gz ... done.
Length: 1,438,199 (1.4M) (unauthoritative)

100%[======================================================>] 1,438,199
610.14K/s

17:06:27 (607.91 KB/s) - 'ecoli.nt.gz' saved [1438199]
```

- Format the database using mpiformatdb as follows. A good rule is to format the database to atleast 4 processors, as follows.

```
[nostromo@xxx ~]$ gunzip ecoli.nt.gz
[nostromo@xxx ~]$ ls
ecoli.nt
[nostromo@xxx ~]$ mpiformatdb --nfrags=4 -i ecoli.nt -pF --quiet
Reading input file
Done, read 58882 lines
Reordering 400 sequence entries
Breaking ecoli.nt into 4 fragments
Executing: formatdb -p F -i /tmp/reorderncq8B1 -N 4 -n /home/nostromo/bio/ncbi/db/ecoli.nt -o T
Removed /tmp/reorderncq8B1
Created 4 fragments.
[nostromo@xxx ~]$ ls
ecoli.nt          ecoli.nt.000.nsq  ecoli.nt.001.nsq  ecoli.nt.002.nsq  ecoli.nt.003.nsq
ecoli.nt.000.nhr  ecoli.nt.001.nhr  ecoli.nt.002.nhr  ecoli.nt.003.nhr  ecoli.nt.mbf
ecoli.nt.000.nin  ecoli.nt.001.nin  ecoli.nt.002.nin  ecoli.nt.003.nin  ecoli.nt.nal
ecoli.nt.000.nnd  ecoli.nt.001.nnd  ecoli.nt.002.nnd  ecoli.nt.003.nnd  formatdb.log
ecoli.nt.000.nni  ecoli.nt.001.nni  ecoli.nt.002.nni  ecoli.nt.003.nni
ecoli.nt.000.nsd  ecoli.nt.001.nsd  ecoli.nt.002.nsd  ecoli.nt.003.nsd
```

```
ecoli.nt.000.nsi  ecoli.nt.001.nsi  ecoli.nt.002.nsi  ecoli.nt.003.nsi
```

- Now create a test sequence file and run mpiblast on the sequence against the formatted database.

```
[nostromo@xxx ~]$ cat > test.txt
>Test
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGC
TTCTGAACTGGTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAA
TATAGGCATAGCGCACAGACAGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACC
ATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAG
CCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGTAACGAGGTAACAACCATGCGAGTGTTGAA
GTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTTGCCGATATTCTGGAAAGCAATGCC
AGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGGTGGCGATGATTG
AAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCCGAACTTTT

[nostromo@xxx mpiblast]$ /opt/openmpi/bin/mpirun -np 4 /opt/bio/mpiblast/bin/mpiblast -d ecoli.
```

After mpirun terminates, result.txt contains the result of your computation.

3.12.3. Running MPI Blast and SGE

This section gives a brief overview of running MPI Blast with SGE

- Create a simple SGE submission scripts called mpiblast_sge.sh with the following contents

```
#!/bin/bash

#$ -cwd
#$ -j y
#$ -S /bin/bash

export MPI_DIR=/opt/openmpi/
export BLASTDB=$HOME/bio/ncbi/db/
export BLASTMAT=/opt/bio/ncbi/data/

$MPI_DIR/bin/mpirun -np $NSLOTS \
 /opt/bio/mpiblast/bin/mpiblast \
 -d ecoli.nt -i $HOME/test.txt \
 -p blastn -o $HOME/result.txt
```

- Run

```
[nostromo@xxx ~]$ qsub -pe orte 4 mpiblast_sge.sh
Your job 11 ("mpiblast_sge.sh") has been submitted
```

- The results of your computation will be present in $HOME/result.txt

Please note that an MPI blast job requires atleast 3 processors to run. The argument for mpirun specifying the number of processors should be factor of the number of pieces the blast database was divided into. If you're running on a cluster with 2 processors, SGE, by default, will not schedule a job which requires more than 2 slots to run.

3.12.4. Further Information

Further information about using mpiblast can be found at the MPI-Blast home page[31].

For support, please join the mpiblast mailing list[32]

3.13. GROMACS

3.13.1. About

GROMACS - Groningen MAchine for Chemical Simulation - is a software suite meant for molecular dynamics simulation.

The version of GROMACS included with the distribution is version 4.0.5. It is available at http://www.gromacs.org under the GNU General Public Licence v2.0.

3.13.2. Usage

GROMACS is setup in /opt/bio/gromacs directory. The version included in this distribution is compiled with mpi support. OpenMPI v1.3.3 is used as the MPI library.

To get more help on using GROMACS, please refer to the following resources:

- GROMACS Home Page[33]
- GROMACS Documentation[34]
- GROMACS Online Reference Manual[35]
- GROMACS FAQ[36]
- Tutorials available on your machines at /opt/bio/gromacs/share/tutor

3.14. Bioperl

3.14.1. About

Bioperl is a set of perl modules for Bio-informatics computation.

3.14.2. Usage

Bioperl modules can be used to supplement already existing applications such as t_coffee, clustalw, and blast. For information on how to use the library, please refer to the API Docs[37].

3.14.3. Further Information

Further information about bioperl is available at the Bioperl home page[38]

3.15. Biopython

3.15.1. About

Biopython is a set of python modules for Bio-informatics computation.

3.15.2. Usage

Biopython modules can be used to supplement already existing applications such as blast. For information on how to use the library, please refer to the biopython documentation[39].

3.15.3. Further Information

Further information about biopython is available at the Biopython home page[40]

Notes

1. http://hmmer.janelia.org/
2. http://www.ncbi.nlm.nih.gov/BLAST/
3. http://mpiblast.lanl.gov/
4. www.biopython.org
5. http://www.ebi.ac.uk/clustalw/
6. http://mrbayes.csit.fsu.edu/
7. http://www.tcoffee.org/Projects_home_page/t_coffee_home_page.html
8. http://emboss.sourceforge.net/
9. http://evolution.genetics.washington.edu/phylip.html
10. http://fasta.bioch.virginia.edu/
11. http://hmmer.janelia.org/#download
12. ftp://selab.janelia.org/pub/software/hmmer/CURRENT/Userguide.pdf
13. ftp://ftp.ncbi.nlm.nih.gov/blast/db/
14. ftp://ftp.ncbi.nlm.nih.gov/blast/db/
15. ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/
16. ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/
17. http://www.ncbi.nlm.nih.gov/BLAST/
18. /blast/docs/

19. http://www.ncbi.nlm.nih.gov/blast/BLAST_guide.pdf
20. http://emboss.sourceforge.net/
21. http://emboss.sourceforge.net/support/
22. http://www.cbcb.umd.edu/software/glimmer/glim302notes.pdf
23. http://fasta.bioch.virginia.edu/
24. http://mrbayes.csit.fsu.edu/index.php
25. http://mrbayes.scs.fsu.edu/wiki/index.php/Manual
26. http://evolution.genetics.washington.edu/phylip.html
27. http://www.tcoffee.org/Projects_home_page/t_coffee_home_page.html
28. http://www.tcoffee.org/Documentation/t_coffee/t_coffee_technical.htm
29. http://www.jcvi.org/cms/publications/listing/abstract/article/tigr-assembler-a-new-tool-for-assembling-large-shotgun-sequencing-projects/
30. http://www.lanl.gov/
31. http://mpiblast.lanl.gov/
32. http://mpiblast.lanl.gov/Support.Lists.html
33. http://www.gromacs.org/
34. http://www.gromacs.org/gromacs/documentation/documentation.html
35. /gromacs/online.html
36. /gromacs/gmxfaq.html
37. http://doc.bioperl.org/
38. http://www.bioperl.org/
39. http://biopython.org/DIST/docs/tutorial/Tutorial.html
40. http://www.biopython.org/

Appendix A. Rocks® Copyright

Appendix B. Third Party Copyrights and Licenses

This section enumerates the licenses from all the third party software components of this Roll. A "best effort" attempt has been made to insure the complete and current licenses are listed. In the case of errors or ommisions please contact the maintainer of this Roll. For more information on the licenses of any components please consult with the original author(s) or see the Rocks CVS repository[1].

B.1. Biopython

```
Permission to use, copy, modify, and distribute this software and its
documentation with or without modifications and for any purpose and
without fee is hereby granted, provided that any copyright notices
appear in all copies and that both those copyright notices and this
permission notice appear in supporting documentation, and that the
names of the contributors or copyright holders not be used in
advertising or publicity pertaining to distribution of the software
without specific prior permission.

THE CONTRIBUTORS AND COPYRIGHT HOLDERS OF THIS SOFTWARE DISCLAIM ALL
WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED
WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL THE
CONTRIBUTORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY SPECIAL, INDIRECT
OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS
OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE
OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE
OR PERFORMANCE OF THIS SOFTWARE.
```

B.2. Clustal W

```
Date:29 November 2007

The copyright for ClustalW and ClustalX is held by Des Higgins, Julie Thompson and Toby Gibson.
The binaries and source code are made available and can be distributed subject to the following c
Users are free to redistribute ClustalW or ClustalX in it’s unmodified form as long as it is not
for commercial gain.
Anyone wishing to redistribute Clustal commercially should contact Toby Gibson at gibson@embl.de
If users make changes/have ideas that they believe would be useful to the broader research commun
they can send their suggestions to the clustal development team at clustalw@ucd.ie where they wil
be considered for inclusion in future releases.
```

B.3. EMBOSS

GNU GENERAL PUBLIC LICENSE

Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc.
59 Temple Place - Suite 330, Boston, MA 02111-1307, USA

Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom
to share and change it. By contrast, the GNU General Public License is
intended to guarantee your freedom to share and change free
software--to make sure the software is free for all its users. This
General Public License applies to most of the Free Software
Foundation's software and to any other program whose authors commit to
using it. (Some other Free Software Foundation software is covered by
the GNU Library General Public License instead.) You can apply it to
your programs, too.

When we speak of free software, we are referring to freedom, not
price. Our General Public Licenses are designed to make sure that you
have the freedom to distribute copies of free software (and charge for
this service if you wish), that you receive source code or can get it
if you want it, that you can change the software or use pieces of it
in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid
anyone to deny you these rights or to ask you to surrender the
rights. These restrictions translate to certain responsibilities for
you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether
gratis or for a fee, you must give the recipients all the rights that
you have. You must make sure that they, too, receive or can get the
source code. And you must show them these terms so they know their
rights.

We protect your rights with two steps: (1) copyright the software, and
(2) offer you this license which gives you legal permission to copy,
distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain
that everyone understands that there is no warranty for this free
software. If the software is modified by someone else and passed on,
we want its recipients to know that what they have is not the
original, so that any problems introduced by others will not reflect
on the original authors' reputations.

Finally, any free program is threatened constantly by software
patents. We wish to avoid the danger that redistributors of a free
program will individually obtain patent licenses, in effect making the

program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone’s free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program’s source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.

c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you

provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the

operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made

generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A

```
FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF
SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH
DAMAGES.
```

B.4. FASTA

```
Copyright 1988, 1991, 1992, 1993, 1994 1995, by William
R. Pearson and the University of Virginia.  All rights
reserved. The FASTA program and documentation may not be sold or
incorporated into a commercial product, in whole or in part,
without written consent of William R. Pearson and the University
of Virginia.  For further information regarding permission for
use or reproduction, please contact:

 David Hudson
 Assistant Provost for Research
 University of Virginia
 P.O. Box 400301
 Charlottesville, VA  22906-9025

 (434) 924-3606

Code in the smith_waterman_sse2.c and smith_waterman_sse2.h files
is copyright (c) 2006 by Michael Farrar.

This program may not be sold or incorporated into a commercial
product, in whole or in part, without written consent of Michael
Farrar.  For further information regarding permission for use or
reproduction, please contact: Michael Farrar at
farrar.michael@gmail.com.
```

B.5. Glimmer

```
Preamble

The intent of this document is to state the conditions under which a
Package may be copied, such that the Copyright Holder maintains some
semblance of artistic control over the development of the package, while
giving the users of the package the right to use and distribute the
Package in a more-or-less customary fashion, plus the right to make
reasonable modifications.

Definitions:

    * "Package" refers to the collection of files distributed by the
      Copyright Holder, and derivatives of that collection of files
      created through textual modification.
    * "Standard Version" refers to such a Package if it has not been
```

```
  modified, or has been modified in accordance with the wishes of
  the Copyright Holder.
* "Copyright Holder" is whoever is named in the copyright or
  copyrights for the package.
* "You" is you, if you're thinking about copying or distributing this Package.
* "Reasonable copying fee" is whatever you can justify on the basis
  of media cost, duplication charges, time of people involved, and
  so on. (You will not be required to justify it to the Copyright
  Holder, but only to the computing community at large as a market
  that must bear the fee.)
* "Freely Available" means that no fee is charged for the item
  itself, though there may be fees involved in handling the item. It
  also means that recipients of the item may redistribute it under
  the same conditions they received it.

1.  You may make and give away verbatim copies of the source form of the
    Standard Version of this Package without restriction, provided that
    you duplicate all of the original copyright notices and associated
    disclaimers.

2.  You may apply bug fixes, portability fixes and other modifications
    derived from the Public Domain or from the Copyright Holder. A
    Package modified in such a way shall still be considered the Standard
    Version.

3.  You may otherwise modify your copy of this Package in any way,
    provided that you insert a prominent notice in each changed file
    stating how and when you changed that file, and provided that you do
    at least ONE of the following:

    a) place your modifications in the Public Domain or otherwise make
    them Freely Available, such as by posting said modifications to
    Usenet or an equivalent medium, or placing the modifications on a
    major archive site such as ftp.uu.net, or by allowing the Copyright
    Holder to include your modifications in the Standard Version of the
    Package.

    b) use the modified Package only within your corporation or
    organization.

    c) rename any non-standard executables so the names do not conflict
    with standard executables, which must also be provided, and provide a
    separate manual page for each non-standard executable that clearly
    documents how it differs from the Standard Version.

    d) make other distribution arrangements with the Copyright Holder.

4.  You may distribute the programs of this Package in object code or
    executable form, provided that you do at least ONE of the following:

    a) distribute a Standard Version of the executables and library
    files, together with instructions (in the manual page or equivalent)
    on where to get the Standard Version.

    b) accompany the distribution with the machine-readable source of
    the Package with your modifications.
```

c) accompany any non-standard executables with their corresponding Standard Version executables, giving the non-standard executables non-standard names, and clearly documenting the differences in manual pages (or equivalent), together with instructions on where to get the Standard Version.

d) make other distribution arrangements with the Copyright Holder.

5. You may charge a reasonable copying fee for any distribution of this Package. You may charge any fee you choose for support of this Package. You may not charge a fee for this Package itself. However, you may distribute this Package in aggregate with other (possibly commercial) programs as part of a larger (possibly commercial) software distribution provided that you do not advertise this Package as a product of your own.

6. The scripts and library files supplied as input to or produced as output from the programs of this Package do not automatically fall under the copyright of this Package, but belong to whomever generated them, and may be sold commercially, and may be aggregated with this Package.

7. C or perl subroutines supplied by you and linked into this Package shall not be considered part of this Package.

8. The name of the Copyright Holder may not be used to endorse or promote products derived from this software without specific prior written permission.

9. THIS PACKAGE IS PROVIDED "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTIBILITY AND FITNESS FOR A PARTICULAR PURPOSE.

B.6. GROMACS

GNU GENERAL PUBLIC LICENSE
Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc.
59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to

using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

GNU GENERAL PUBLIC LICENSE
TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not

covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

    a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

    b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.

    c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program
with the Program (or with a work based on the Program) on a volume of
a storage or distribution medium does not bring the other work under
the scope of this License.

  3. You may copy and distribute the Program (or a work based on it,
under Section 2) in object code or executable form under the terms of
Sections 1 and 2 above provided that you also do one of the following:

    a) Accompany it with the complete corresponding machine-readable
    source code, which must be distributed under the terms of Sections
    1 and 2 above on a medium customarily used for software interchange; or,

    b) Accompany it with a written offer, valid for at least three
    years, to give any third party, for a charge no more than your
    cost of physically performing source distribution, a complete
    machine-readable copy of the corresponding source code, to be
    distributed under the terms of Sections 1 and 2 above on a medium
    customarily used for software interchange; or,

    c) Accompany it with the information you received as to the offer
    to distribute corresponding source code. (This alternative is
    allowed only for noncommercial distribution and only if you
    received the program in object code or executable form with such
    an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for
making modifications to it. For an executable work, complete source
code means all the source code for all modules it contains, plus any
associated interface definition files, plus the scripts used to
control compilation and installation of the executable. However, as a
special exception, the source code distributed need not include
anything that is normally distributed (in either source or binary
form) with the major components (compiler, kernel, and so on) of the
operating system on which the executable runs, unless that component
itself accompanies the executable.

If distribution of executable or object code is made by offering
access to copy from a designated place, then offering equivalent
access to copy the source code from the same place counts as
distribution of the source code, even though third parties are not
compelled to copy the source along with the object code.

  4. You may not copy, modify, sublicense, or distribute the Program
except as expressly provided under this License. Any attempt
otherwise to copy, modify, sublicense or distribute the Program is
void, and will automatically terminate your rights under this License.
However, parties who have received copies, or rights, from you under
this License will not have their licenses terminated so long as such
parties remain in full compliance.

  5. You are not required to accept this License, since you have not
signed it. However, nothing else grants you permission to modify or
distribute the Program or its derivative works. These actions are
prohibited by law if you do not accept this License. Therefore, by
modifying or distributing the Program (or any work based on the

Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the program's name and a brief idea of what it does.>
Copyright (C) 19yy <name of author>

This program is free software; you can redistribute it and/or modify

```
    it under the terms of the GNU General Public License as published by
    the Free Software Foundation; either version 2 of the License, or
    (at your option) any later version.

    This program is distributed in the hope that it will be useful,
    but WITHOUT ANY WARRANTY; without even the implied warranty of
    MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
    GNU General Public License for more details.

    You should have received a copy of the GNU General Public License
    along with this program; if not, write to the Free Software
    Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA  02111-1307  USA


Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this
when it starts in an interactive mode:

    Gnomovision version 69, Copyright (C) 19yy name of author
    Gnomovision comes with ABSOLUTELY NO WARRANTY; for details type `show w'.
    This is free software, and you are welcome to redistribute it
    under certain conditions; type `show c' for details.

The hypothetical commands `show w' and `show c' should show the appropriate
parts of the General Public License.  Of course, the commands you use may
be called something other than `show w' and `show c'; they could even be
mouse-clicks or menu items--whatever suits your program.

You should also get your employer (if you work as a programmer) or your
school, if any, to sign a "copyright disclaimer" for the program, if
necessary.  Here is a sample; alter the names:

  Yoyodyne, Inc., hereby disclaims all copyright interest in the program
  `Gnomovision' (which makes passes at compilers) written by James Hacker.

  <signature of Ty Coon>, 1 April 1989
  Ty Coon, President of Vice

This General Public License does not permit incorporating your program into
proprietary programs.  If your program is a subroutine library, you may
consider it more useful to permit linking proprietary applications with the
library.  If this is what you want to do, use the GNU Library General
Public License instead of this License.
```

B.7. HMMER

```
                    GNU GENERAL PUBLIC LICENSE
                       Version 2, June 1991

 Copyright (C) 1989, 1991 Free Software Foundation, Inc.
                          675 Mass Ave, Cambridge, MA 02139, USA
```

Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

GNU GENERAL PUBLIC LICENSE
TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

```
  0. This License applies to any program or other work which contains
a notice placed by the copyright holder saying it may be distributed
under the terms of this General Public License.  The "Program", below,
refers to any such program or work, and a "work based on the Program"
means either the Program or any derivative work under copyright law:
that is to say, a work containing the Program or a portion of it,
either verbatim or with modifications and/or translated into another
language.  (Hereinafter, translation is included without limitation in
the term "modification".)  Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not
covered by this License; they are outside its scope.  The act of
running the Program is not restricted, and the output from the Program
is covered only if its contents constitute a work based on the
Program (independent of having been made by running the Program).
Whether that is true depends on what the Program does.

  1. You may copy and distribute verbatim copies of the Program's
source code as you receive it, in any medium, provided that you
conspicuously and appropriately publish on each copy an appropriate
copyright notice and disclaimer of warranty; keep intact all the
notices that refer to this License and to the absence of any warranty;
and give any other recipients of the Program a copy of this License
along with the Program.

You may charge a fee for the physical act of transferring a copy, and
you may at your option offer warranty protection in exchange for a fee.

  2. You may modify your copy or copies of the Program or any portion
of it, thus forming a work based on the Program, and copy and
distribute such modifications or work under the terms of Section 1
above, provided that you also meet all of these conditions:

    a) You must cause the modified files to carry prominent notices
    stating that you changed the files and the date of any change.

    b) You must cause any work that you distribute or publish, that in
    whole or in part contains or is derived from the Program or any
    part thereof, to be licensed as a whole at no charge to all third
    parties under the terms of this License.

    c) If the modified program normally reads commands interactively
    when run, you must cause it, when started running for such
    interactive use in the most ordinary way, to print or display an
    announcement including an appropriate copyright notice and a
    notice that there is no warranty (or else, saying that you provide
    a warranty) and that users may redistribute the program under
    these conditions, and telling the user how to view a copy of this
    License.  (Exception: if the Program itself is interactive but
    does not normally print such an announcement, your work based on
    the Program is not required to print an announcement.)


These requirements apply to the modified work as a whole.  If
identifiable sections of that work are not derived from the Program,
and can be reasonably considered independent and separate works in
themselves, then this License, and its terms, do not apply to those
```

sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt

otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in

certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS

Appendix: How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it

free software which everyone can redistribute and change under these terms.

  To do so, attach the following notices to the program. It is safest
to attach them to the start of each source file to most effectively
convey the exclusion of warranty; and each file should have at least
the "copyright" line and a pointer to where the full notice is found.

    <one line to give the program's name and a brief idea of what it does.>
    Copyright (C) 19yy <name of author>

    This program is free software; you can redistribute it and/or modify
    it under the terms of the GNU General Public License as published by
    the Free Software Foundation; either version 2 of the License, or
    (at your option) any later version.

    This program is distributed in the hope that it will be useful,
    but WITHOUT ANY WARRANTY; without even the implied warranty of
    MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
    GNU General Public License for more details.

    You should have received a copy of the GNU General Public License
    along with this program; if not, write to the Free Software
    Foundation, Inc., 675 Mass Ave, Cambridge, MA 02139, USA.

Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this
when it starts in an interactive mode:

    Gnomovision version 69, Copyright (C) 19yy name of author
    Gnomovision comes with ABSOLUTELY NO WARRANTY; for details type `show w'.
    This is free software, and you are welcome to redistribute it
    under certain conditions; type `show c' for details.

The hypothetical commands `show w' and `show c' should show the appropriate
parts of the General Public License. Of course, the commands you use may
be called something other than `show w' and `show c'; they could even be
mouse-clicks or menu items--whatever suits your program.

You should also get your employer (if you work as a programmer) or your
school, if any, to sign a "copyright disclaimer" for the program, if
necessary. Here is a sample; alter the names:

  Yoyodyne, Inc., hereby disclaims all copyright interest in the program
  `Gnomovision' (which makes passes at compilers) written by James Hacker.

  <signature of Ty Coon>, 1 April 1989
  Ty Coon, President of Vice

This General Public License does not permit incorporating your program into
proprietary programs. If your program is a subroutine library, you may
consider it more useful to permit linking proprietary applications with the
library. If this is what you want to do, use the GNU Library General
Public License instead of this License.

B.8. mpiBlast

```
BLAST - open source parallel BLAST

Source code and other files distributed with mpiBLAST are
copyright (c) 2002-2005 by Aaron Darling of Los Alamos National
Laboratory for the Regents of the University of California.
Numerous authors have contributed to mpiBLAST and are listed in
the included AUTHORS file.

License:

Copyright 2002-2005. The Regents of the University of California.

This material was produced under U.S. Government contract
W-7405-ENG-36 for Los Alamos National Laboratory, which is operated
by the University of California for the U.S. Department of Energy.
The Government is granted for itself and other acting on its behalf
a paid-up, non-exclusive, irrevocable worldwide license in the
material to reproduce, prepare derivative works, and perform
publicly and display publicly. Beginning five (5) years after
November 6, 2002, subject to additional five-year worldwide
renewals, the Government is granted for itself and others acting on
its behalf a paid-up, nonexclusive, irrevocable worldwide license
in this material to reproduce, prepare derivative works, distribute
copies to the public, perform publicly and display publicly, and to
permit others to do so. NEITHER THE UNITED STATES NOR THE UNITED
STATES DEPARTMENT OF ENERGY, NOR THE UNIVERSITY OF CALIFORNIA, NOR
ANY OF THEIR EMPLOYEES, MAKES ANY WARRANTY, EXPRESS OR IMPLIED, OR
ASSUMES ANY LEGAL LIABLITY OR RESPONSIBILITY FOR THE ACCURACY,
COMPLETENESS, OR USEFULNESS OF ANY INFORMATION, APPARATUS, PRODUCT,
OR PROCESS DISCLOSED, OR REPRESENTS THAT ITS USE WOULD NOT INFRINGE
PRIVATELY OWNED RIGHTS.

Additionally, this program is free software; you can distribute it
and/or modify it under the terms of the GNU General Public License
as published by the Free Software Foundation; either version 2 of
the License, or any later version. Accordingly, this program is
distributed in the hope that it will be useful, but WITHOUT ANY
WARRANTY; without even the implied warranty of MERCHANTABILITY or
FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public
License for more details. You should have received a copy of the
GNU General Public License along with this program; if not, write
to the Free Software Foundation, Inc., 59 Temple Place - Suite 330,
Boston, MA 02111-1307, USA.

-------------------------------------------------------------

You may contact the primary author (Aaron E. Darling) via email
at darling@cs.wisc.edu, or via postal mail at 425 Henry Mall,
Madison WI, 53706.


See the file 'COPYING' for the details of the GNU General Public License

Some portions of mpiBLAST are derivitive works of the NCBI Toolbox.
```

In particular, src/blast_hooks.c and src/mpiblast_formatdb.c are derivitive
of blastall.c and formatdb.c from the NCBI Toolbox. As such, they remain
under the NCBI public domain license given at the beginning of those files.

B.9. Mr.Bayes

GNU GENERAL PUBLIC LICENSE
Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc.
59 Temple Place, Suite 330, Boston, MA 02111-1307 USA
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so

that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

GNU GENERAL PUBLIC LICENSE
TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.

c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to

control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING

WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR
REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES,
INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING
OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED
TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY
YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER
PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE
POSSIBILITY OF SUCH DAMAGES.

B.10. NCBI

PUBLIC DOMAIN NOTICE
National Center for Biotechnology Information

This software/database is a "United States Government Work" under the terms of the United States Copyright Act. It was written as part of the author's official duties as a United States Government employee and thus cannot be copyrighted. This software/database is freely available to the public for use. The National Library of Medicine and the U.S. Government have not placed any restriction on its use or reproduction.

Although all reasonable efforts have been taken to ensure the accuracy and reliability of the software and data, the NLM and the U.S. Government do not and cannot warrant the performance or results that may be obtained by using this software or data. The NLM and the U.S. Government disclaim all warranties, express or implied, including warranties of performance, merchantability or fitness for any particular purpose.

Please cite the author in any work or product based on this material.

B.11. Perl Modules

The "Artistic License"

Preamble

The intent of this document is to state the conditions under which a Package may be copied, such that the Copyright Holder maintains some semblance of artistic control over the development of the package, while giving the users of the package the right to use and distribute the Package in a more-or-less customary fashion, plus the right to make reasonable modifications.

Definitions:

"Package" refers to the collection of files distributed by the Copyright Holder, and derivatives of that collection of files

```
        created through textual modification.

        "Standard Version" refers to such a Package if it has not been
        modified, or has been modified in accordance with the wishes
        of the Copyright Holder as specified below.

        "Copyright Holder" is whoever is named in the copyright or
        copyrights for the package.

        "You" is you, if you're thinking about copying or distributing
        this Package.

        "Reasonable copying fee" is whatever you can justify on the
        basis of media cost, duplication charges, time of people involved,
        and so on.  (You will not be required to justify it to the
        Copyright Holder, but only to the computing community at large
        as a market that must bear the fee.)

        "Freely Available" means that no fee is charged for the item
        itself, though there may be fees involved in handling the item.
        It also means that recipients of the item may redistribute it
        under the same conditions they received it.

1. You may make and give away verbatim copies of the source form of the
Standard Version of this Package without restriction, provided that you
duplicate all of the original copyright notices and associated disclaimers.

2. You may apply bug fixes, portability fixes and other modifications
derived from the Public Domain or from the Copyright Holder.  A Package
modified in such a way shall still be considered the Standard Version.

3. You may otherwise modify your copy of this Package in any way, provided
that you insert a prominent notice in each changed file stating how and
when you changed that file, and provided that you do at least ONE of the
following:

    a) place your modifications in the Public Domain or otherwise make them
    Freely Available, such as by posting said modifications to Usenet or
    an equivalent medium, or placing the modifications on a major archive
    site such as uunet.uu.net, or by allowing the Copyright Holder to include
    your modifications in the Standard Version of the Package.

    b) use the modified Package only within your corporation or organization.

    c) rename any non-standard executables so the names do not conflict
    with standard executables, which must also be provided, and provide
    a separate manual page for each non-standard executable that clearly
    documents how it differs from the Standard Version.

    d) make other distribution arrangements with the Copyright Holder.

4. You may distribute the programs of this Package in object code or
executable form, provided that you do at least ONE of the following:

    a) distribute a Standard Version of the executables and library files,
    together with instructions (in the manual page or equivalent) on where
    to get the Standard Version.
```

b) accompany the distribution with the machine-readable source of the Package with your modifications.

c) give non-standard executables non-standard names, and clearly document the differences in manual pages (or equivalent), together with instructions on where to get the Standard Version.

d) make other distribution arrangements with the Copyright Holder.

5. You may charge a reasonable copying fee for any distribution of this Package. You may charge any fee you choose for support of this Package. You may not charge a fee for this Package itself. However, you may distribute this Package in aggregate with other (possibly commercial) programs as part of a larger (possibly commercial) software distribution provided that you do not advertise this Package as a product of your own. You may embed this Package's interpreter within an executable of yours (by linking); this shall be construed as a mere form of aggregation, provided that the complete Standard Version of the interpreter is so embedded.

6. The scripts and library files supplied as input to or produced as output from the programs of this Package do not automatically fall under the copyright of this Package, but belong to whoever generated them, and may be sold commercially, and may be aggregated with this Package. If such scripts or library files are aggregated with this Package via the so-called "undump" or "unexec" methods of producing a binary executable image, then distribution of such an image shall neither be construed as a distribution of this Package nor shall it fall under the restrictions of Paragraphs 3 and 4, provided that you do not represent such an executable image as a Standard Version of this Package.

7. C subroutines (or comparably compiled subroutines in other languages) supplied by you and linked into this Package in order to emulate subroutines and variables of the language defined by this Package shall not be considered part of this Package, but are the equivalent of input as in Paragraph 6, provided these subroutines do not change the language in any way that would cause it to fail the regression tests for the language.

8. Aggregation of this Package with a commercial distribution is always permitted provided that the use of this Package is embedded; that is, when no overt attempt is made to make this Package's interfaces visible to the end user of the commercial distribution. Such use shall not be construed as a distribution of this Package.

9. The name of the Copyright Holder may not be used to endorse or promote products derived from this software without specific prior written permission.

10. THIS PACKAGE IS PROVIDED "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTIBILITY AND FITNESS FOR A PARTICULAR PURPOSE.

The End

B.12. Phylip

```
The following copyright notice is intended to cover all source code, all
documentation, and all executable programs of the PHYLIP package.

Copyright 1980-2004. University of Washington and Joseph Felsenstein. All
rights reserved. Permission is granted to reproduce, perform, and modify these
programs and documentation files. Permission is granted to distribute or
provide access to these programs provided that this copyright notice is not
removed, the programs are not integrated with or called by any product or
service that generates revenue, and that your distribution of these
documentation files and programs are free. Any modified versions of these
materials that are distributed or accessible shall indicate that they are
based on these program. Institutions of higher education are granted
permission to distribute this material to their students and staff for a fee
to recover distribution costs. Permission requests for any other distribution
of this program should be directed to  license @ u.washington.edu
```

B.13. T_Coffee

```
  ACADEMIC LICENCE AGREEMENT

  © Centre National de la Recherche Scientifique (CNRS) and Cedric Notredame  ( Tue Feb 26 16:54:
      GNU GENERAL PUBLIC LICENSE
         Version 2, June 1991

 Copyright (C) 1989, 1991 Free Software Foundation, Inc.
                       59 Temple Place, Suite 330, Boston, MA  02111-1307  USA
 Everyone is permitted to copy and distribute verbatim copies
 of this license document, but changing it is not allowed.

       Preamble

  The licenses for most software are designed to take away your
freedom to share and change it.  By contrast, the GNU General Public
License is intended to guarantee your freedom to share and change free
software--to make sure the software is free for all its users.  This
General Public License applies to most of the Free Software
Foundation's software and to any other program whose authors commit to
using it.  (Some other Free Software Foundation software is covered by
the GNU Library General Public License instead.)  You can apply it to
your programs, too.

  When we speak of free software, we are referring to freedom, not
price.  Our General Public Licenses are designed to make sure that you
have the freedom to distribute copies of free software (and charge for
this service if you wish), that you receive source code or can get it
if you want it, that you can change the software or use pieces of it
in new free programs; and that you know you can do these things.

  To protect your rights, we need to make restrictions that forbid
anyone to deny you these rights or to ask you to surrender the rights.
```

These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

GNU GENERAL PUBLIC LICENSE
TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and
you may at your option offer warranty protection in exchange for a fee.

  2. You may modify your copy or copies of the Program or any portion
of it, thus forming a work based on the Program, and copy and
distribute such modifications or work under the terms of Section 1
above, provided that you also meet all of these conditions:

    a) You must cause the modified files to carry prominent notices
    stating that you changed the files and the date of any change.

    b) You must cause any work that you distribute or publish, that in
    whole or in part contains or is derived from the Program or any
    part thereof, to be licensed as a whole at no charge to all third
    parties under the terms of this License.

    c) If the modified program normally reads commands interactively
    when run, you must cause it, when started running for such
    interactive use in the most ordinary way, to print or display an
    announcement including an appropriate copyright notice and a
    notice that there is no warranty (or else, saying that you provide
    a warranty) and that users may redistribute the program under
    these conditions, and telling the user how to view a copy of this
    License. (Exception: if the Program itself is interactive but
    does not normally print such an announcement, your work based on
    the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If
identifiable sections of that work are not derived from the Program,
and can be reasonably considered independent and separate works in
themselves, then this License, and its terms, do not apply to those
sections when you distribute them as separate works. But when you
distribute the same sections as part of a whole which is a work based
on the Program, the distribution of the whole must be on the terms of
this License, whose permissions for other licensees extend to the
entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest
your rights to work written entirely by you; rather, the intent is to
exercise the right to control the distribution of derivative or
collective works based on the Program.

In addition, mere aggregation of another work not based on the Program
with the Program (or with a work based on the Program) on a volume of
a storage or distribution medium does not bring the other work under
the scope of this License.

  3. You may copy and distribute the Program (or a work based on it,
under Section 2) in object code or executable form under the terms of
Sections 1 and 2 above provided that you also do one of the following:

    a) Accompany it with the complete corresponding machine-readable
    source code, which must be distributed under the terms of Sections
    1 and 2 above on a medium customarily used for software interchange; or,

    b) Accompany it with a written offer, valid for at least three
    years, to give any third party, for a charge no more than your

cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not

excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

```
<one line to give the program's name and a brief idea of what it does.>
Copyright (C) <year>  <name of author>

This program is free software; you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation; either version 2 of the License, or
(at your option) any later version.

This program is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
GNU General Public License for more details.

You should have received a copy of the GNU General Public License
along with this program; if not, write to the Free Software
Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA  02111-1307  USA
```

Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

```
    Gnomovision version 69, Copyright (C) year name of author
    Gnomovision comes with ABSOLUTELY NO WARRANTY; for details type `show w'.
    This is free software, and you are welcome to redistribute it
    under certain conditions; type `show c' for details.

The hypothetical commands `show w' and `show c' should show the appropriate
parts of the General Public License.  Of course, the commands you use may
be called something other than `show w' and `show c'; they could even be
mouse-clicks or menu items--whatever suits your program.

You should also get your employer (if you work as a programmer) or your
school, if any, to sign a "copyright disclaimer" for the program, if
necessary.  Here is a sample; alter the names:

  Yoyodyne, Inc., hereby disclaims all copyright interest in the program
  `Gnomovision' (which makes passes at compilers) written by James Hacker.

  <signature of Ty Coon>, 1 April 1989
  Ty Coon, President of Vice

This General Public License does not permit incorporating your program into
proprietary programs.  If your program is a subroutine library, you may
consider it more useful to permit linking proprietary applications with the
library.  If this is what you want to do, use the GNU Library General
Public License instead of this License.
```

B.14. TIGR Assembler

```
Copyright (c) 2003, The Institute for Genomic Research (TIGR), Rockville,
Maryland, U.S.A.  All rights reserved.

The Artistic License

Preamble

The intent of this document is to state the conditions under which a
Package may be copied, such that the Copyright Holder maintains some
semblance of artistic control over the development of the package,
while giving the users of the package the right to use and distribute
the Package in a more-or-less customary fashion, plus the right to
make reasonable modifications.

Definitions:
    * "Package" refers to the collection of files distributed by the
      Copyright Holder, and derivatives of that collection of files
      created through textual modification.
    * "Standard Version" refers to such a Package if it has not been
      modified, or has been modified in accordance with the wishes of
      the Copyright Holder.
    * "Copyright Holder" is whoever is named in the copyright or
      copyrights for the package.
    * "You" is you, if you're thinking about copying or distributing
```

```
      this Package.
    * "Reasonable copying fee" is whatever you can justify on the
      basis of media cost, duplication charges, time of people
      involved, and so on. (You will not be required to justify it to
      the Copyright Holder, but only to the computing community at
      large as a market that must bear the fee.)
    * "Freely Available" means that no fee is charged for the item
      itself, though there may be fees involved in handling the
      item. It also means that recipients of the item may redistribute
      it under the same conditions they received it.

1. You may make and give away verbatim copies of the source form of
   the Standard Version of this Package without restriction, provided
   that you duplicate all of the original copyright notices and
   associated disclaimers.

2. You may apply bug fixes, portability fixes and other modifications
   derived from the Public Domain or from the Copyright Holder. A
   Package modified in such a way shall still be considered the
   Standard Version.

3. You may otherwise modify your copy of this Package in any way,
   provided that you insert a prominent notice in each changed file
   stating how and when you changed that file, and provided that you
   do at least ONE of the following:

    a) place your modifications in the Public Domain or otherwise make
    them Freely Available, such as by posting said modifications to
    Usenet or an equivalent medium, or placing the modifications on a
    major archive site such as ftp.uu.net, or by allowing the
    Copyright Holder to include your modifications in the Standard
    Version of the Package.

    b) use the modified Package only within your corporation or
    organization.

    c) rename any non-standard executables so the names do not
    conflict with standard executables, which must also be provided,
    and provide a separate manual page for each non-standard
    executable that clearly documents how it differs from the Standard
    Version.

    d) make other distribution arrangements with the Copyright Holder.

4. You may distribute the programs of this Package in object code or
   executable form, provided that you do at least ONE of the
   following:

    a) distribute a Standard Version of the executables and library
    files, together with instructions (in the manual page or
    equivalent) on where to get the Standard Version.

    b) accompany the distribution with the machine-readable source of
    the Package with your modifications.

    c) accompany any non-standard executables with their corresponding
    Standard Version executables, giving the non-standard executables
```

```
    non-standard names, and clearly documenting the differences in
    manual pages (or equivalent), together with instructions on where
    to get the Standard Version.

    d) make other distribution arrangements with the Copyright Holder.

5. You may charge a reasonable copying fee for any distribution of
   this Package. You may charge any fee you choose for support of this
   Package. You may not charge a fee for this Package itself. However,
   you may distribute this Package in aggregate with other (possibly
   commercial) programs as part of a larger (possibly commercial)
   software distribution provided that you do not advertise this
   Package as a product of your own.

6. The scripts and library files supplied as input to or produced as
   output from the programs of this Package do not automatically fall
   under the copyright of this Package, but belong to whomever
   generated them, and may be sold commercially, and may be aggregated
   with this Package.

7. C or perl subroutines supplied by you and linked into this Package
   shall not be considered part of this Package.

8. The name of the Copyright Holder may not be used to endorse or
   promote products derived from this software without specific prior
   written permission.

9. THIS PACKAGE IS PROVIDED "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED
   WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES
   OF MERCHANTIBILITY AND FITNESS FOR A PARTICULAR PURPOSE.

The End
This license is approved by the Open Source Initiative
(www.opensource.org) for certifying software as OSI Certified Open
Source.
```

Notes

1. http://cvs.rocksclusters.org