ISSN online: 2282-5665

Consiglio Nazionale delle Ricerche

CERIS ISTITUTO DI RICERCA SULL'IMPRESA E LO SVILUPPO

Dicembre

2013

Rapporto tecnico N.49

STEP BY STEP INSTALLATION GUIDE OF A VIRTUAL SERVER FOR BIO APPLICATIONS FROM SCRATCH

Simona Abbà, Giancarlo Birello, Ivano Fucile, Stefano Ghignone, Valter Giovanetti, Anna Perin

Consiglio Nazionale delle Ricerche

Istituto di Ricerche sull'Impresa e Lo Sviluppo

RAPPORTO TECNICO CNR-CERIS
Anno 8, N° 49; Dicembre 2013

Direttore Responsabile
Secondo Rolfo

Direzione e Redazione
CNR-Ceris
Istituto di Ricerca sull'Impresa e lo Sviluppo
Via Real Collegio, 30
10024 Moncalieri (Torino), Italy
Tel. +39 011 6824.911
Fax +39 011 6824.966
segreteria@ceris.cnr.it
http://www.ceris.cnr.it

Sede di Roma
Via dei Taurini, 19
00185 Roma, Italy
Tel. 06 49937810
Fax 06 49937884

Sede di Milano
Via Bassini, 15
20121 Milano, Italy
tel. 02 23699501
Fax 02 23699530

Segreteria di redazione
Enrico Viarisio
e.viarisio@ceris.cnr.it

STEP BY STEP INSTALLATION GUIDE OF A VIRTUAL SERVER FOR BIO APPLICATIONS FROM SCRATCH

Giancarlo Birello, Ivano Fucile Valter Giovanetti*
(CNR-Ceris, IT Office)

Anna Perin
(CNR-Ceris, Library)

Simona Abbà
(CNR-IVV)

Stefano Ghignone
(CNR-IPP)

CNR-Ceris, IT Office
Strada delle Cacce, 73
10135 Torino – Italy
Phone: +39 0113977 533/4/5

CNR-Ceris, Library
Via Real Collegio, 30
10024 Moncalieri (Torino) – Italy
Phone: +39 01168249 28

CNR-IVV
Strada delle Cacce, 73
I-10135 Torino - Italy
Phone: +39 0113977 925

CNR-IPP, Sez. Torino
c/o Dip. Scienze della Vita e Biologia dei Sistemi
V.le P.A. Mattioli, 25
I-10125 Turin - Italy
Phone: +39 0116502 927

* Corresponding author: g.birello@ceris.cnr.it

ABSTRACT: CNR IVV and CNR IPP Institutes have been dealing with NGS DATA (Next-gen sequencing) for years, but, until 2012, no CNR-based high-performance server was available. Powerful multi-core servers, clusters and virtual cloud-based machines represent are essential to address the computational demands of NGS. CNR CERIS Institute built up a server for such bio-applications and the joint effort amongst, IVV, IPP and CERIS put the basis for a new bioinformatics facility in the CNR Turin Research Area.

KEY WORDS: open-source, NGS data, bioinformatics, virtual cloud-based machines

Table of Contents

1 Introduction

Nucleic acid sequencing is a method for determining the exact order of nucleotides present in a given DNA or RNA molecule. Since 2005, the next-generation sequencing (also 'Next-gen sequencing' or NGS) technologies have been revolutionizing life sciences with unimaginable speed, enabling impressive scientific achievements and novel biological applications.

The term NGS refers to technologies that have enabled the massively parallel sequencing of DNA on the gigabase scale in a single day, through a complex integration of chemistry, biology, optical sensors, advanced imaging and computer hardware and software. Providing low-cost and high-throughput sequencing, NGS platforms have largely replaced earlier capillary sequencing methods and have created significant enthusiasm for researchers, who can move quickly from a project idea to full data sets in a matter of hours or days.

The progress in genome sequencing technology is so rapid that the cost efficiency of NGS has far outpaced the growth of compute power predicted by Moore's law[1] Moore's law says that computer processors double in complexity approximately every 18 months, whereas, in an extreme case, technology improvements for a next-generation platform have led to a yield doubling time of as little as 5 months[2].

Given the huge amounts of data generated by NGS platforms, these technologies are posing new major statistical and bioinformatics challenges, particularly for genome/transcriptome assemblies. In particular, some NGS technologies produce billions of short sequences ('reads' ≤ 100 bp) that make the assembly of full-length genomes computationally demanding. The bioinformatics community and the computer power have to keep pace with this flood of raw sequence data, otherwise computational analysis will be the rate-limiting factor in genomics research.

Memory and CPUs are not the only concerns in high-throughput sequencing: the hard disk capacity is another key factor, because the storage and the analysis (generating a huge amount of temporary data files) of each data set is estimated to require 500-600 GB of disk space[3], depending on the sequenced organism and the NGS platform used. It is simply unthinkable to use desktop computers for such a task, so powerful multi-core servers, clusters and virtual cloud-based machines represent so far the solution to address the computational demands of NGS.

Two CNR Institutes, IVV and IPP, working the Turin area have been dealing with NGS data for years, but, until 2012, no CNR-based high-performance server was available. The virtual server for bio-applications that we are going to describe in the next pages is the result of the joint effort amongst CERIS, IVV and IPP to put the basis for a new bioinformatics facility in the Turin area that we hope to improve and expand in the next years.

2 Hypervisor and base system

The main server acting as hypervisor ("*Towanda*") is equipped with 64 cores and 256 GB RAM due to application requirements, memory will be expanded to 512 GB RAM next year.

Virtualization framework is KVM (Kernel- based Virtual Machine) installed over an Ubuntu Linux operative system. Full hypervisor installation guide can be found in our technical report "*KVM: an open-source framework for virtualization* ", RT44 Ceris-CNR.

Virtual machine disks are LVM based for a flexible storage management (see details in RT 44). *Towanda* is equipped with 1 TB RAID-1 logical drive and 3.7 TB RAID-5 logical drive, both on local server storage. In addition a 1 TB iSCSI partition is available to virtual machines for backup and temporarily storage, the partition is located on two-nodes HA cluster (see technical reports RT37 and RT41 Ceris-CNR).

Applications are hosted on a virtual machine ("*Rachael*") based on Ubuntu server 12.04 LTS. We made 240GB RAM and 60 cores available to *Rachael* virtual machine, while remaining resources are reserved for hypervisor base functions.

Storage for virtual machine was divided in 4 main partitions mounted on root ("/"), home, db and storage. The whole 3.7 TB logical volume was assigned to home directory, 15 GB for root (base system), 90 GB for db directory and 700 GB for storage directory (SAN).

Operative system was installed from ISO image of standard distribution with default values. Next paragraph starts from a fresh installation with all system packages updated to last version available.

Note: why *Towanda*, do you remember *Fried green tomatoes* movie? We love Evelyn mood (http://en.wikiquote.org/wiki/Fried_Green_Tomatoes)

Note: why *Rachael*, do you remember *Blade Ranner* movie? We like the Replicant that is led to believe that she is human (http://en.wikipedia.org/wiki/List_of_Blade_Runner_characters)

3 Applications

3.1 Qiime

"*QIIME (canonically pronounced "chime") stands for Quantitative Insights Into Microbial Ecology. QIIME is an open source software package for comparison and analysis of microbial communities, primarily based on high-throughput amplicon sequencing data (such as SSU rRNA) generated on a variety of platforms, but also supporting analysis of other types of data (such as shotgun metagenomic data).*"

Dependencies for all Qiime versions

```
# apt-get install python-dev libncurses5-dev libssl-dev libzmq-dev libgsl0-dev
  openjdk-6-jdk libxml2 libxslt1.1 libxslt1-dev ant git subversion build-essential
  zlib1g-dev libpng12-dev libfreetype6-dev mpich2 libreadline-dev gfortran unzip
  libmysqlclient18 libmysqlclient-dev ghc sqlite3 libsqlite3-dev
```

Useful package

```
# apt-get install python-biopython
```

We report installation steps from Qiime 1.5.0 for history, while it's possible to start with Qiime 1.7.0 skipping previous versions.

We followed Qiime GitHub page (https://github.com/qiime/qiime-deploy) for all versions.

You may install more than one version of QIIME on your system.

3.1.1 Qiime 1.5.0

```
# mkdir /usr/share/qiime_software
# cd /root/
# git clone git://github.com/qiime/qiime-deploy.git
# git clone git://github.com/qiime/qiime-deploy-conf.git
# cd qiime-deploy
```

Run deployment script

```
# python qiime-deploy.py /usr/share/qiime_software/ -f /root/qiime-deploy-
  conf/qiime-1.5.0/qiime.conf --force-remove-failed-dirs
```

result

```
...
DEPLOYMENT SUMMARY
Packages deployed successfully:
data-lanemask, rtax, clearcut, cdbtools, raxml, ampliconnoise, chimeraslayer,
  vienna, infernal, data-core, fasttree, pplacer, python, parsinsert, blast, muscle,
  rdpclassifier, drisee, cdhit, uclust, pyzmq, SQLAlchemy, tornado, cytoscape,
  setuptools, numpy, MySQL-python, mpi4py, biom-format, mothur, pycogent, pynast,
  pprospector, sphinx, ipython, matplotlib, qiime, r, gg_otus
Packages skipped (assumed successful):
Packages failed to deploy:
```

Installation check with errors

```
$ print_qiime_config.py -t
...
======================================================================
FAIL: test_denoiser_supported_version (__main__.Qiime_config)
denoiser aligner is ready to use
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/usr/share/qiime_software/qiime-1.5.0-release/bin/print_qiime_config.py", line 536, in
  test_denoiser_supported_version
    "which components of QIIME you plan to use.")
AssertionError: Denoiser flowgram aligner not found or not executable.This may or may not be a problem depending
  on which components of QIIME you plan to use.
======================================================================
FAIL: test_mothur_supported_version (__main__.Qiime_config)
mothur is in path and version is supported
----------------------------------------------------------------------
Traceback (most recent call last):
```

```
  File "/usr/share/qiime_software/qiime-1.5.0-release/bin/print_qiime_config.py", line 523, in
   test_mothur_supported_version
    % ('.'.join(map(str,acceptable_version)), version_string))
AssertionError: Unsupported mothur version. 1.25.0 is required, but running mothur: error while loading shared
   libraries: libreadline.so.5: cannot open shared object file: No such file or directory
.======================================================================
FAIL: test_python_supported_version (__main__.Qiime_config)
python is in path and version is supported
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/usr/share/qiime_software/qiime-1.5.0-release/bin/print_qiime_config.py", line 338, in
   test_python_supported_version
    version_string))
AssertionError: Unsupported python version. Must be >= 2.7.1 and <= 2.7.2 , but running 2.7.3.
======================================================================
FAIL: test_rtax_supported_version (__main__.Qiime_config)
rtax is in path and version is supported
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/usr/share/qiime_software/qiime-1.5.0-release/bin/print_qiime_config.py", line 620, in
   test_rtax_supported_version
    % ('.'.join(map(str,acceptable_version)), version_string))
AssertionError: Unsupported rtax version. (0, 981).(0, 981) is required, but running 0.982.
======================================================================
FAIL: test_usearch_supported_version (__main__.Qiime_config)
usearch is in path and version is supported
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/usr/share/qiime_software/qiime-1.5.0-release/bin/print_qiime_config.py", line 668, in
   test_usearch_supported_version
    "which components of QIIME you plan to use.")
AssertionError: usearch not found. This may or may not be a problem depending on which components of QIIME you
   plan to use.
----------------------------------------------------------------------
Ran 34 tests in 0.724s
FAILED (failures=5)
```

Fix Qiime installation errors.

- USEARCH

Copy usearch folder from Qiime virtual image

```
# scp -r user@qiime_virtualbox:/usr/share/qiime_software/usearch-5.2.32-release
  /usr/share/qiime_software/
```

Then add path in activate.sh

```
# nano -w /usr/share/qiime_software/activate.sh
```

```
export PATH=...:/usr/share/qiime_software/usearch-5.2.32-release/.:$PATH
```

- RTAX

Copy rtax-0.981 folder from Qiime virtual image

```
# scp -r user@qiime_virtualbox:/usr/share/qiime_software/rtax-0.981-release
  /usr/share/qiime_software/
```

Then change path from rtax-0.982-release to rtax-0.981-release in activate.sh

```
# nano -w /usr/share/qiime_software/activate.sh
```

```
export PATH=...:/usr/share/qiime_software/rtax-0.981-release/.:$PATH
```

- MOTHUR

Install missed package

```
# apt-get install libreadline5
```

Replace mothur folder with the mothur folder on Qiime virtual image

```
# mv mothur-1.25.0-release mothur-1.25.0-release.ORI
# scp -r user@qiime_virtualbox:/usr/share/qiime_software/mothur-1.25.0-release
  /usr/share/qiime_software/
```

- DENOISER

Compile missed bin and install it

```
# cd /usr/share/qiime_software/qiime-1.5.0-
  release/qiime/support_files/denoiser/FlowgramAlignment
# make
# make install
# mkdir /usr/share/qiime_software/qiime-1.5.0-
  release/lib/qiime/support_files/denoiser/bin
# cp /usr/share/qiime_software/qiime-1.5.0-
  release/qiime/support_files/denoiser/bin/FlowgramAli_4frame
  /usr/share/qiime_software/qiime-1.5.0-release/lib/qiime/support_files/denoiser/bin/
```

- PYTHON

Edit these files and change min_unacceptable_version from (2,7,2) to (2,7,3)

```
/usr/share/qiime_software/qiime-1.5.0-release/bin/print_qiime_config.py
/usr/share/qiime_software/qiime-1.5.0-release/build/scripts-
  2.7/print_qiime_config.py
/usr/share/qiime_software/qiime-1.5.0-release/scripts/print_qiime_config.py
```

Finally check Qiime installation

```
$ print_qiime_config.py -t
```

```
System information
==================
         Platform:    linux2
   Python version:    2.7.3 (default, Aug  1 2012, 05:25:23)  [GCC 4.6.3]
Python executable:    /usr/share/qiime_software/python-2.7.1-release/bin/python

Dependency versions
===================
                     PyCogent version:    1.5.1
                       NumPy version:    1.5.1
                  matplotlib version:    1.1.0
                 biom-format version:    0.9.3
               QIIME library version:    1.5.0
                QIIME script version:    1.5.0
       PyNAST version (if installed):    1.1
RDP Classifier version (if installed):    rdp_classifier-2.2.jar
...
Ran 34 tests in 0.675s
OK
```

3.1.2 Qiime 1.6.0

```
# mkdir /usr/share/qiime-1.6.0
# cd /root/
# rm -R qiime-deploy
# rm -R qiime-deploy-conf
# git clone git://github.com/qiime/qiime-deploy.git
# git clone git://github.com/qiime/qiime-deploy-conf.git
# cd qiime-deploy
# python qiime-deploy.py /usr/share/qiime-1.6.0/ -f /root/qiime-deploy-conf/qiime-
  1.6.0/qiime.conf --force-remove-failed-dirs
```

Install `usearch` (=5.2.236) downloaded from http://www.drive5.com/usearch/download.html

```
# mkdir /usr/share/qiime-1.6.0/usearch-5.2.236-release/
# cd /usr/share/qiime-1.6.0/usearch-5.2.236-release
# wget /urltousearch/usearch
# chmod +x usearch
```

Add path to activate script

```
# nano -w /usr/share/qiime-1.6.0/activate.sh
```

```
export PATH=...:/usr/share/qiime-1.6.0/usearch-5.2.236-release/.:$PATH
```

Check Qiime installation

```
$ source /usr/share/qiime-1.6.0/activate.sh
$ print_qiime_config.py -t
```

```
System information
==================
         Platform:      linux2
   Python version:      2.7.3 (default, Jan 10 2013, 13:17:26)  [GCC 4.6.3]
Python executable:      /usr/share/qiime-1.6.0/python-2.7.3-release/bin/python

Dependency versions
===================
                    PyCogent version:  1.5.3
                       NumPy version:  1.5.1
                  matplotlib version:  1.1.0
                biom-format version:  1.1.1
               QIIME library version:  1.6.0
                QIIME script version:  1.6.0
       PyNAST version (if installed):  1.2
RDP Classifier version (if installed):  rdp_classifier-2.2.jar
         Java version (if installed):  1.6.0_24
...
Ran 35 tests in 0.859s
OK
```

3.1.3 Qiime 1.7.0

```
# cd /root/
# git clone git://github.com/qiime/qiime-deploy.git
# git clone git://github.com/qiime/qiime-deploy-conf.git
# cd qiime-deploy
# python qiime-deploy.py /usr/share/qiime-1.7.0/ -f /root/qiime-deploy-conf/qiime-
  1.7.0/qiime.conf --force-remove-failed-dirs
```

Install `usearch` (=5.2.236) downloaded from http://www.drive5.com/usearch/download.html

```
# mkdir /usr/share/qiime-1.7.0/usearch-5.2.236-release/
# cd /usr/share/qiime-1.7.0/usearch-5.2.236-release
# wget /urltousearch/usearch
# chmod +x usearch
```

Add path to activate script

```
# nano -w /usr/share/qiime-1.7.0/activate.sh                                    [5.1.1]
```

```
export PATH=...:/usr/share/qiime-1.7.0/usearch-5.2.236-release/.:$PATH
```

Install `pplacer` (=1.1) downloaded from http://matsen.fhcrc.org/pplacer/

```
# wget http://matsen.fhcrc.org/pplacer/builds/pplacer-v1.1-Linux.tar.gz
# tar -xvzf pplacer-v1.1-Linux.tar.gz
# mv pplacer-v1.1.alpha14-Linux-3.2.0 /usr/share/qiime-1.7.0/
```

Add path to activate script

```
# nano -w /usr/share/qiime-1.7.0/activate.sh
```

```
export PATH=...:/usr/share/qiime-1.7.0/pplacer-v1.1.alpha14-Linux-
  3.2.0/.:/usr/share/qiime-1.7.0/pplacer-v1.1.alpha14-Linux-3.2.0/scripts/.:$PATH
```

Check Qiime installation

```
$ source /usr/share/qiime-1.7.0/activate.sh
$ print_qiime_config.py -t
```

```
System information
==================
         Platform:      linux2
   Python version:      2.7.3 (default, Jun  6 2013, 13:11:23)  [GCC 4.6.3]
Python executable:      /usr/share/qiime-1.7.0/python-2.7.3-release/bin/python
```

```
Dependency versions
===================
                    PyCogent version:  1.5.3
                       NumPy version:  1.5.1
                    matplotlib version:  1.1.0
                  biom-format version:  1.1.2
                 QIIME library version:  1.7.0
                  QIIME script version:  1.7.0
         PyNAST version (if installed):  1.2
RDP Classifier version (if installed):  rdp_classifier-2.2.jar
          Java version (if installed):  1.7.0_21
...
Ran 35 tests in 0.763s
OK
```

3.2 Trinity

"*Trinity, developed at the Broad Institute and the Hebrew University of Jerusalem, represents a novel method for the efficient and robust de novo reconstruction of transcriptomes from RNA-seq data.*"

Install extra packages: bowtie, ncurses and PerlIO layer to gzip/gunzip

```
# apt-get install bowtie libncurses-dev libperlio-gzip-perl
```

Download and compile Trinity

```
# cd /root
# wget
  http://garr.dl.sourceforge.net/project/trinityrnaseq/trinityrnaseq_r20131110.tar.gz
# tar xzf trinityrnaseq_r20131110.tar.gz
# cd trinityrnaseq_r20131110/
# make
```

Remove if present previous version

```
# mv /usr/share/trinity /usr/share/trinityOLD
```

Install

```
# cd ..
# mv trinityrnaseq_r20131110 /usr/share/trinity
# chown -R root:root /usr/share/trinity
# chmod -R 755 /usr/share/trinity
```

Download and install RSEM

```
# cd /root/
# wget  http://deweylab.biostat.wisc.edu/rsem/src/rsem-1.2.7.tar.gz
# tar -xvzf rsem-1.2.7.tar.gz
# cd rsem-1.2.7/
# make
# cd ..
# mv rsem-1.2.7 /usr/local/bin/
# chown -R root:root /usr/local/bin/rsem-1.2.7
# cd /usr/local/bin/
# ln -s rsem-1.2.7 rsem
```

User script to activate Trinity and RSEM

```
nano -w /usr/local/bin/TrinityON.sh
```

```
export
  PATH=/usr/share/trinity:/usr/share/trinity/util:/usr/share/trinity/util/RSEM_util:
  /usr/local/bin/rsem:$PATH
export TRINITY_HOME=/usr/share/trinity
ulimit -s unlimited
```

Check installation as user.

Logon and activate Trinity

```
$ source /usr/local/bin/TrinityON.sh
```

Copy sample directory

```
$ cp -R /usr/share/trinity/sample_data ~/
```

Edit test script and remove script path

```
$ cd sample_data/test_Trinity_Assembly/
$ nano -w runMe.sh
```

```
#!/bin/bash -ve

if [ -e reads.left.fq.gz ] && ! [ -e reads.left.fq ]
then
    gunzip -c reads.left.fq.gz > reads.left.fq
fi
if [ -e reads.right.fq.gz ] && ! [ -e reads.right.fq ]
then
    gunzip -c reads.right.fq.gz > reads.right.fq
fi

################################################################
##  Run Trinity to Generate Transcriptome Assemblies ##
################################################################

## use jellyfish
Trinity.pl --seqType fq --JM 2G --left reads.left.fq.gz --right reads.right.fq.gz --SS_lib_type
  RF --CPU 4 --no_cleanup
##### Done Running Trinity #####

if [ ! $* ]; then
    exit 0
fi
sleep 2

#########################################################
## align reads back to the transcripts using Bowtie ##
#########################################################

sleep 2
alignReads.pl --left reads.left.fq --right reads.right.fq --target trinity_out_dir/Trinity.fasta
  --aligner bowtie --seqType fq --SS_lib_type RF
##### Done aligning reads #######
sleep 2

##########################################
# use RSEM to estimate read abundance  ####
##########################################

sleep 2
run_RSEM.pl --transcripts trinity_out_dir/Trinity.fasta --name_sorted_bam
  bowtie_out/bowtie_out.nameSorted.sam.+.sam.PropMapPairsForRSEM.bam --paired
###### Done running RSEM ########
```

Run test

```
$ ./runMe.sh
```

result

```
...
##### Done Running Trinity #####
```

Clean test directory

```
$ ./cleanme.pl
```

3.3 Blast+ and Blast legacy

"*The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.*"

3.3.1 Database

We'll use `/db` directory mounted as separated partition.

Copy db update script from Blast+ bin directory.

```
# cp /usr/share/ncbi-blast-2.2.27+/bin/update_blastdb.pl /db/
```

Script help

```
# cd /db
# ./update_blastdb.pl -help
```

List database names

```
# cd /db
# ./update_blastdb.pl --showall
```

Download/update database (i.e. swissprot)

```
# cd /db
# ./update_blastdb.pl --decompress --verbose swissprot
```

3.3.2 Blast+

Download and install

```
# wget /fromsomeserver/ncbi-blast-2.2.28+-x64-linux.tar.gz
# tar xzf ncbi-blast-2.2.28+-x64-linux.tar.gz
# mv ncbi-blast-2.2.28+ /usr/share/
# chown -R root:root /usr/share/ncbi-blast-2.2.28+
# chmod -R 755 /usr/share/ncbi-blast-2.2.28+
```

User script to activate Blast+

```
nano -w /usr/local/bin/Blast+ON.sh
```

```
export PATH=/usr/share/ncbi-blast-2.2.28+/bin:$PATH
export BLASTDB=/db
```

Check installation

```
$ source /usr/local/bin/Blast+ON.sh
$ blastp -db swissprot -query testseq.fasta
```

3.3.3 Blast legacy

Download and install

```
# wget /fromsomeserver/blast-2.2.26-x64-linux.tar.gz
# tar xzf blast-2.2.26-x64-linux.tar.gz
# mv /root/blast-2.2.26 /usr/share/
# chown -R root:root /usr/share/blast-2.2.26
# chmod -R 755 /usr/share/blast-2.2.26
```

User script to activate Blast legacy

```
nano -w /usr/local/bin/BlastON.sh
```

```
export PATH=/usr/share/blast-2.2.26/bin:$PATH
export BLASTMAT=/usr/share/blast-2.2.26/data
export BLASTDB=/db
```

Check installation

```
$ source /usr/local/bin/BlastON.sh
$ blastall -p blastp -i testseq.fasta -d swissprot
```

3.3.4 Schedule db update

Create schedule script in `/db` directory

```
#!/bin/bash
cd /db
date > "update.log"

#verify lock db
echo "verifica lock" >> "update.log"
DIRS=$(find "/home/" -mindepth 1 -maxdepth 1 -type d)
for user in $DIRS
do
        filelock=$user"/dblock"
        if [ -e $filelock ] && [ -f $filelock ]
        then
                echo $filelock" -> NO update" >> "update.log"
                exit
        fi
done
echo "SI Update" >> "update.log"
./update_blastdb.pl --decompress swissprot 16SMicrobial refseq_protein >> "update.log"
echo "DONE" >> "update.log"
exit
```

Schedule cron job

```
# crontab -e

    50 12 * * * /db/condupdatedb.sh > /dev/null 2>&1
```

Note: *to avoid update while blast is running, user must create a file named "`dblock`" in his home directory. At the end of the process the file must be removed.*

3.4 Bioperl

"*Welcome to BioPerl, a community effort to produce Perl code which is useful in biology.*"

From packages:

```
# apt-get install bioperl
```

3.5 Mira

"*MIRA is a whole genome shotgun and EST sequence assembler for Sanger, 454, Solexa (Illumina), IonTorrent data and PacBio (the later at the moment only CCS and error-corrected CLR reads).*"

Dependencies for all Mira versions

```
# apt-get install libboost-doc libboost.*1.48-dev libboost.*1.48.0 flex libgoogle-
  perftools-dev expat libbz2-dev libtcmalloc-minimal0-dbg
```

We report installation steps from Mira 3.9.9 for history, while it's possible to start with Mira 4.0rc4 skipping previous versions.

3.5.1 Install Mira from source (3.9.9)

Download

```
# wget http://downloads.sourceforge.net/project/mira-
  assembler/MIRA/development/mira-3.9.9.tar.bz2?r=&ts=1357647818&use_mirror=garr
# tar jxvf mira-3.9.9.tar.bz2
```

BUG: remove `-GE:uti` line from `parameters.C` source

```
# cd mira-3.9.9/
# nano -w src/mira/parameters.C
```

remove line

```
"\n\t-GE:uti=yes"
```

Compile

```
# ./configure
# make
# make install
```

Check installation

```
$ mira -v
3.9.9 ()
$ convert_project -v
3.9.9 ()
```

3.5.2 Update Mira from source (3.9.10/3.9.11/3.9.18)

Download

```
# wget http://garr.dl.sourceforge.net/project/mira-assembler/MIRA/development/mira-
  3.9.10.tar.bz2
# tar jxvf mira-3.9.10.tar.bz2
```

Compile

```
# cd mira-3.9.10/
# ./configure
# make
# make install
```

Check installation

```
$ mira -v
3.9.10 ()
$ convert_project -v
3.9.10 ()
```

3.5.3 Update Mira from binaries (4.0rc4)

Download

```
# wget http://garr.dl.sourceforge.net/project/mira-
  assembler/MIRA/stable/mira_4.0rc4_linux-gnu_x86_64_static.tar .bz2
# tar jxvf mira_4.0rc4_linux-gnu_x86_64_static.tar.bz2
```

Install

```
# mv mira_4.0rc4_linux-gnu_x86_64_static /usr/local/
# cd /usr/local/
# mkdir OLDMira-3.9.18
# mv bin/mira* OLDMira-3.9.18/
```

```
# mv bin/convert_project OLDMira-3.9.18/
# chown -R root:root mira_4.0rc4_linux-gnu_x86_64_static
# ln -s /usr/local/mira_4.0rc4_linux-gnu_x86_64_static/bin/* /usr/local/bin/
```

Check installation

```
$ mira -h
```

```
mira            MIRALIB version 4.0rc4
Author:         Bastien Chevreux (bach@chevreux.org)
Purpose:        assemble sequencing data.

To (un-)subscribe the MIRA mailing lists, see:
        http://www.chevreux.org/mira_mailinglists.html

After subscribing, mail general questions to the MIRA talk mailing list:
        mira_talk@freelists.org


To report bugs or ask for features, please use the SourceForge ticketing
system at:
        http://sourceforge.net/p/mira-assembler/tickets/
This ensures that requests do not get lost.


Compiled by: bach
Mon Oct 14 17:32:36 CEST 2013
On: Linux vk10464 2.6.32-41-generic #94-Ubuntu SMP Fri Jul 6 18:00:34 UTC 2012 x
  86_64 GNU/Linux
Compiled in boundtracking mode.
Compiled in bugtracking mode.
Compiled with ENABLE64 activated.
Runtime settings (sorry, for debug):
        Size of size_t  : 8
        Size of uint32  : 4
        Size of uint32_t: 4
        Size of uint64  : 8
        Size of uint64_t: 8
Current system: Linux rachaelx 3.2.0-29-generic #46-Ubuntu SMP Fri Jul 27 17:03:
  23 UTC 2012 x86_64 x86_64 x86_64 GNU/Linux

Usage:
mira [options] manifest_file [manifest_file ...]

Options:
  -c / --cwd=           directory       Change working directory
  -r / --resume                         Resume an interrupted assembly
  -h / --help                           Print short help and exit
  -v / --version                        Print version and exit
```

3.5.4 Mira 3rd party tools

Download

```
# wget http://garr.dl.sourceforge.net/project/mira-
  assembler/MIRA/development/mira_3rdparty_06-07-2012.tar.bz2
# tar jxvf mira_3rdparty_06-07-2012.tar.bz2
```

Install

```
# mv 3rdparty /usr/local/bin/
# cd /usr/local/bin/
# chown -R root:root 3rdparty
# chmod -R 755 3rdparty
# chmod -R 644 3rdparty/README.txt
# chmod -R 644 3rdparty/midi_screen.fasta
# ln -s 3rdparty/454pairedEnd2caf.pl 454pairedEnd2caf.pl
# ln -s 3rdparty/bin_fasta_on_mid_primers.pl bin_fasta_on_mid_primers.pl
# ln -s 3rdparty/caf2aceMiraConsed.pl caf2aceMiraConsed.pl
# ln -s 3rdparty/lucy2xml.pl lucy2xml.pl
# ln -s 3rdparty/qual2ball qual2ball
# ln -s 3rdparty/readpair2caf.pl readpair2caf.pl
# ln -s 3rdparty/sff_extract sff_extract
```

3.6 SMALT and SSAHA2

"*SMALT efficiently aligns DNA sequencing reads with a reference genome.*"

"*SSAHA2 (Sequence Search and Alignment by Hashing Algorithm) is a pairwise sequence alignment program designed for the efficient mapping of sequencing reads onto genomic reference sequences.*"

Download

```
# wget ftp://ftp.sanger.ac.uk/pub4/resources/software/smalt/smalt-0.7.2.tgz
# wget
  ftp://ftp.sanger.ac.uk/pub4/resources/software/ssaha2/ssaha2_v2.5.5_x86_64.tgz
# tar -zxvf smalt-0.7.2.tgz
# tar -zvxf ssaha2_v2.5.5_x86_64.tgz
```

Install

```
# mv smalt-0.7.2 /usr/local/bin/
# mv ssaha2_v2.5.5_x86_64 /usr/local/bin/
# cd /usr/local/bin/
# ln -s smalt-0.7.2/smalt_x86_64 smalt
# ln -s ssaha2_v2.5.5_x86_64/ssaha2 ssaha2
# ln -s ssaha2_v2.5.5_x86_64/ssaha2Build ssaha2Build
# ln -s ssaha2_v2.5.5_x86_64/ssahaSNP ssahaSNP
# chown -R root:root smalt-0.7.2
# chown -R root:root ssaha2_v2.5.5_x86_64
```

3.7 Tophat

"*TopHat is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to mammalian-sized genomes using the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results to identify splice junctions between exons.*"

We installed version 1.4.1 from binaries.

Download

```
# wget http://tophat.cbcb.umd.edu/downloads/tophat-1.4.1.Linux_x86_64.tar.gz
# tar xvzf tophat-1.4.1.Linux_x86_64.tar.gz
```

Install

```
# mv tophat-1.4.1.Linux_x86_64 /usr/local/bin/
# ln -s /usr/local/bin/tophat-1.4.1.Linux_x86_64/tophat /usr/local/bin/
```

Check installation

```
$ wget http://tophat.cbcb.umd.edu/downloads/test_data.tar.gz
$ tar zxvf test_data.tar.gz
$ cd test_data/
$ tophat -r 20 test_ref reads_1.fq reads_2.fq
```

3.8 R

3.8.1 Environment

"*R is a free software environment for statistical computing and graphics.*"

Add to /etc/apt/sources.list this line:

```
deb http://cran.mirror.garr.it/mirrors/CRAN/bin/linux/ubuntu precise/
```

then

```
sudo apt-key adv --keyserver keyserver.ubuntu.com --recv-keys E084DAB9
sudo apt-get update
sudo apt-get install r-base
```

Enter R at prompt:

```
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
```

3.8.2 Rstudio

"*RStudio IDE is a powerful and productive user interface for R.*"

Install dependencies

```
# apt-get install gdebi-core libapparmor1
# apt-get install libapache2-mod-proxy-html libxml2-dev
```

Download and install

```
# wget http://download2.rstudio.org/rstudio-server-0.97.551-amd64.deb
# gdebi rstudio-server-0.97.551-amd64.deb
```

Configure reverse-proxy

```
# a2enmod proxy
# a2enmod proxy_http
# nano -w /etc/apache2/sites-enabled/000-default
```

```
…
  <Proxy *>
    Allow from localhost
  </Proxy>

ProxyPass        /rstudio/ http://localhost:8383/
ProxyPassReverse /rstudio/ http://localhost:8383/
RedirectMatch permanent ^/rstudio$ /rstudio/
```

```
# service apache2 restart
```

Limit connection from localhost

```
nano -w /etc/rstudio/rserver.conf
```

```
  www-address=127.0.0.1
```

```
# rstudio-server stop
# rstudio-server start
```

3.8.3 Bioconductor

"*Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data.*"

Additional packages

```
# apt-get install libcurl4-openssl-dev
```

Login as root and at R prompt:

```
source("http://bioconductor.org/biocLite.R")
biocLite()
biocLite(c("GenomicFeatures", "AnnotationDbi"))
```

3.9 Pandaseq

"*PANDASEQ is a program to align Illumina reads, optionally with PCR primers embedded in the sequence, and reconstruct an overlapping sequence.*"

Install dependencies

```
# apt-get install libtool
```

Compile and install

```
# cd /root/
# git clone git://github.com/neufeld/pandaseq.git
# cd pandaseq
# ./autogen.sh
# ./configure
# make
# make install
# ldconfig
```

3.10 WGS-assembler

"*Celera Assembler is a de novo whole-genome shotgun (WGS) DNA sequence assembler. It reconstructs long sequences of genomic DNA from fragmentary data produced by whole-genome shotgun sequencing.*"

Download and install

```
# wget http://garr.dl.sourceforge.net/project/wgs-assembler/wgs-assembler/wgs-
  7.0/wgs-7.0-PacBio-Linux-amd64.tar.bz2
# bzip2 -dc wgs-7.0-PacBio-Linux-amd64.tar.bz2  |  tar  -xf  -
# mv wgs-7.0 /usr/local/bin/
# chown -R root:root /usr/local/bin/wgs-7.0
```

Patch correctPacBio

```
# wget http://www.cbcb.umd.edu/software/PBcR/data/correctPacBio
# mv /usr/local/bin/wgs-7.0/Linux-amd64/bin/correctPacBio /usr/local/bin/wgs-
  7.0/Linux-amd64/bin/correctPacBio.OLD
# mv correctPacBio /usr/local/bin/wgs-7.0/Linux-amd64/bin/
# chmod 755 /usr/local/bin/wgs-7.0/Linux-amd64/bin/correctPacBio
```

Check installation

```
$ /usr/local/bin/wgs-7.0/Linux-amd64/bin/pacBioToCA
```

```
usage: /usr/local/bin/wgs-7.0/Linux-amd64/bin/pacBioToCA [options] -s spec.file -fastq fastqfile
  <frg>
  -length                 Minimum length to keep.
  -partitions             Number of partitions for consensus
  -sge                    Submit consensus jobs to the grid
  -sgeCorrection          Parameters for the correction step for the grid. This should match the
  threads specified below, for example by using -pe threaded
  -l libraryname          Name of the library; freeformat text.
  -t threads              Number of threads to use for correction.
```

3.11 Amos

"*AMOS is a collection of tools and class interfaces for the assembly of DNA reads. The package includes a robust infrastructure, modular assembly pipelines, and tools for overlapping, consensus generation, contigging, and assembly manipulation.*"

Install dependencies

```
# apt-get install libqt4-dev
```

Download, compile and install

```
# cd /root/
# wget http://garr.dl.sourceforge.net/project/amos/amos/3.1.0/amos-3.1.0.tar.gz
# tar -xvzf amos-3.1.0.tar.gz
# cd amos-3.1.0/
# nano -w src/Align/find-tandem.cc
```

```
    #include <getopt.h>
```

```
# ./configure --prefix=/usr/local/AMOS
# make
# make check
# make install
# ln -s /usr/local/AMOS/bin/* /usr/local/bin/
```

Check installation

```
$ Minimo -h
```

```
Minimo is a de novo assembler based on the AMOS infrastructure. Minimo uses a
conservative overlap-layout-consensus algorithm to avoid mis-assemblies and
can be applied to short reads. The input is a FASTA file and there are options
to control the stringency of the assembly and the processing of the quality
scores. By default, the results are in the AMOS format and written to the
directory where the input FASTA file is located.

Usage:
    Minimo FASTA_IN [options]

Options:
    -D QUAL_IN=<file>   Input quality score file (in Phred format)
    -D GOOD_QUAL=<n>    Quality score to set for bases within the clear
                          range if no quality file was given (default: 30)
    -D BAD_QUAL=<n>     Quality score to set for bases outside clear range
                          if no quality file was given (default: 10). If your
                          sequences are trimmed, try the same value as GOOD_QUAL.
    -D MIN_LEN=<n>      Minimum contig overlap length (at least 20 bp,
                          default: 35)
    -D MIN_IDENT=<d>    Minimum contig overlap identity percentage (between 0
                          and 100 %, default: 98)
    -D ALN_WIGGLE=<d>   Alignment wiggle value (from 2 for short reads to 15 for
                          long reads, default: 2)
```

```
    -D FASTA_EXP=<n>    Export results in FASTA format (0:no 1:yes, default: 0)
    -D ACE_EXP=<n>      Export results in ACE format (0:no 1:yes, default: 0)
    -D OUT_PREFIX=<s>   Prefix to use for the output file path and name

Minimo v1.6. Copyright Florent Angly 2010. Under the GPL v3 open-source license.
```

3.12 MUMmer

"*MUMmer is a system for rapidly aligning entire genomes, whether in complete or draft form.*"

Install dependencies

```
# apt-get install csh
```

Download and install

```
# cd /root/
# wget http://garr.dl.sourceforge.net/project/mummer/mummer/3.23/MUMmer3.23.tar.gz
# tar -xvzf MUMmer3.23.tar.gz
# mv MUMmer3.23 /usr/local/
# cd /usr/local/MUMmer3.23/
# make check
# make install
# ln -s /usr/local/MUMmer3.23/nucmer /usr/local/bin/
# ln -s /usr/local/MUMmer3.23/show-coords /usr/local/bin/
# ln -s /usr/local/MUMmer3.23/delta-filter /usr/local/bin/
# chown -R root:root /usr/local/MUMmer3.23
```

Check installation

```
$ nucmer -h
```

```
USAGE: nucmer  [options]  <Reference>  <Query>

DESCRIPTION:
  nucmer generates nucleotide alignments between two mutli-FASTA input
  files. The out.delta output file lists the distance between insertions
  and deletions that produce maximal scoring alignments between each
  sequence. The show-* utilities know how to read this format.

MANDATORY:
  Reference       Set the input reference multi-FASTA filename
  Query           Set the input query multi-FASTA filename

OPTIONS:
  --mum           Use anchor matches that are unique in both the reference
                  and query
  --mumcand       Same as --mumreference
  --mumreference  Use anchor matches that are unique in in the reference
                  but not necessarily unique in the query (default behavior)
  --maxmatch      Use all anchor matches regardless of their uniqueness

  -b|breaklen     Set the distance an alignment extension will attempt to
                  extend poor scoring regions before giving up (default 200)
  --[no]banded    Enforce absolute banding of dynamic programming matrix
                  based on diagdiff parameter EXPERIMENTAL (default no)
  -c|mincluster   Sets the minimum length of a cluster of matches (default 65)
  --[no]delta     Toggle the creation of the delta file (default --delta)
  --depend        Print the dependency information and exit
  -D|diagdiff     Set the maximum diagonal difference between two adjacent
                  anchors in a cluster (default 5)
  -d|diagfactor   Set the maximum diagonal difference between two adjacent
                  anchors in a cluster as a differential fraction of the gap
                  length (default 0.12)
  --[no]extend    Toggle the cluster extension step (default --extend)
  -f
  --forward       Use only the forward strand of the Query sequences
  -g|maxgap       Set the maximum gap between two adjacent matches in a
                  cluster (default 90)
  -h
  --help          Display help information and exit
  -l|minmatch     Set the minimum length of a single match (default 20)
  -o
  --coords        Automatically generate the original NUCmer1.1 coords
                  output file using the 'show-coords' program
  --[no]optimize  Toggle alignment score optimization, i.e. if an alignment
                  extension reaches the end of a sequence, it will backtrack
                  to optimize the alignment score instead of terminating the
```

```
                  alignment at the end of the sequence (default --optimize)
    -p|prefix     Set the prefix of the output files (default "out")
    -r
    --reverse     Use only the reverse complement of the Query sequences
    --[no]simplify  Simplify alignments by removing shadowed clusters. Turn
                  this option off if aligning a sequence to itself to look
                  for repeats (default --simplify)
    -V
    --version     Display the version information and exit
```

3.13 BWA

"*BWA is a software package for mapping low-divergent sequences against a large reference genome, such as the human genome.*"

Download, compile and install

```
# wget http://garr.dl.sourceforge.net/project/bio-bwa/bwa-0.7.5a.tar.bz2
# bzip2 -dc bwa-0.7.5a.tar.bz2 | tar -xf  -
# mv bwa-0.7.5a /usr/local/
# cd /usr/local/bwa-0.7.5a/
# make
# ln -s /usr/local/bwa-0.7.5a/bwa /usr/local/bin/
# chown -R root:root /usr/local/bwa-0.7.5a
```

Check installation

```
$ bwa
```

```
Program: bwa (alignment via Burrows-Wheeler transformation)
Version: 0.7.5a-r405
Contact: Heng Li <lh3@sanger.ac.uk>

Usage:   bwa <command> [options]

Command: index         index sequences in the FASTA format
         mem           BWA-MEM algorithm
         fastmap       identify super-maximal exact matches
         pemerge       merge overlapping paired ends (EXPERIMENTAL)
         aln           gapped/ungapped alignment
         samse         generate alignment (single ended)
         sampe         generate alignment (paired ended)
         bwasw         BWA-SW for long queries

         fa2pac        convert FASTA to PAC format
         pac2bwt       generate BWT from PAC
         pac2bwtgen    alternative algorithm for generating BWT
         bwtupdate     update .bwt to the new format
         bwt2sa        generate SA from BWT and Occ

Note: To use BWA, you need to first index the genome with `bwa index'. There are
      three alignment algorithms in BWA: `mem', `bwasw' and `aln/samse/sampe'. If
      you are not sure which to use, try `bwa mem' first. Please `man ./bwa.1' for
      for the manual.
```

3.14 Blasr

"*BLASR (Basic Local Alignment with Successive Refinement) rapidly maps reads to genomes by finding the highest scoring local alignment or set of local alignments between the read and the genome.*"

Install dependencies

```
# apt-get install libhdf5-serial-dev
```

Download

```
# wget https://codeload.github.com/PacificBiosciences/blasr/zip/master
# mv master blasr-master.zip
# unzip blasr-master.zip
```

```
# mv blasr-master /usr/local/
# cd /usr/local/blasr-master/
```

Edit alignment/Makefile to solve compiling error due to -lpthread flag position. [5.2.1]

Compile and install

```
# make
# ln -s /usr/local/blasr-master/alignment/bin/blasr /usr/local/bin/
# chown -R root:root /usr/local/blasr-master
```

Check installation

```
$ blasr -h
```

```
Options for blasr
   Basic usage: 'blasr reads.{fasta,bas.h5} genome.fasta [-options]
 option Description (default_value).
...
To cite BLASR, please use: Chaisson M.J., and Tesler G., Mapping
single molecule sequencing reads using Basic Local Alignment with
Successive Refinement (BLASR): Theory and Application, BMC
Bioinformatics 2012, 13:238 .
```

3.15 HMMER

"*HMMER is used for searching sequence databases for homologs of protein sequences, and for making protein sequence alignments. It implements methods using probabilistic models called* ***profile hidden Markov models*** *(profile HMMs).*"

Download, compile and install

```
# cd /root/
# wget ftp://selab.janelia.org/pub/software/hmmer3/3.1b1/hmmer-3.1b1-linux-intel-
  x86_64.tar.gz
# tar -xvzf hmmer-3.1b1-linux-intel-x86_64.tar.gz
# cd hmmer-3.1b1-linux-intel-x86_64/
# ./configure
# make
# make check
# make install

# cd /db
# wget ftp://ftp.sanger.ac.uk/pub/databases/Pfam/current_release/Pfam-A.hmm.gz
# gunzip Pfam-A.hmm.gz
```

Check installation

```
$ hmmalign -h
```

```
# hmmalign :: align sequences to a profile HMM
# HMMER 3.1b1 (May 2013); http://hmmer.org/
# Copyright (C) 2013 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3).
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Usage: hmmalign [-options] <hmmfile> <seqfile>

Basic options:
  -h     : show brief help on version and usage
  -o <f> : output alignment to file <f>, not stdout

Less common options:
  --mapali <f>    : include alignment in file <f> (same ali that HMM came from)
  --trim          : trim terminal tails of nonaligned residues from alignment
  --amino         : assert <seqfile>, <hmmfile> both protein: no autodetection
  --dna           : assert <seqfile>, <hmmfile> both DNA: no autodetection
  --rna           : assert <seqfile>, <hmmfile> both RNA: no autodetection
  --informat <s>  : assert <seqfile> is in format <s>: no autodetection
  --outformat <s> : output alignment in format <s>  [Stockholm]
```

```
Sequence input formats include:   FASTA, EMBL, GenBank, UniProt
Alignment output formats include: Stockholm, Pfam, A2M, PSIBLAST
```

3.16 Trinotate

"*Trinotate is a comprehensive annotation suite designed for automatic functional annotation of transcriptomes, particularly de novo assembled transcriptomes, from model or non-model organisms.*"

3.16.1 SignalP

"*SignalP 4.1 server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes. The method incorporates a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks.*"

Download from http://www.cbs.dtu.dk/cgi-bin/nph-sw_request?signalp

```
# cat signalp-4.1c.Linux.tar.Z | uncompress | tar xvf -
# mv signalp-4.1 /usr/local/bin/
# nano -w /usr/local/bin/signalp-4.1/signalp
```

```
#############################################################################
#               GENERAL SETTINGS: CUSTOMIZE TO YOUR SITE
#############################################################################

# full path to the signalp-4.1 directory on your system (mandatory)
BEGIN {
    $ENV{SIGNALP} = '/usr/local/bin/signalp-4.1';
}

# determine where to store temporary files (must be writable to all users)
my $outputDir = "/home/temp";

# max number of sequences per run (any number can be handled)
my $MAX_ALLOWED_ENTRIES=2000000;
```

```
ln -s /usr/local/bin/signalp-4.1/signalp /usr/local/bin/
chown -R root:root /usr/local/bin/signalp-4.1
chmod -R 755 /usr/local/bin/signalp-4.1
cp /usr/local/bin/signalp-4.1/signalp.1 /usr/local/share/man/man1/
```

Check installation

```
$ signalp -t euk -f short /usr/local/bin/signalp-4.1/test/euk10.fsa >
  euk10.fsa.short_out
$ diff euk10.fsa.short_out /usr/local/bin/signalp-4.1/test/euk10.fsa.short_out

$ signalp -t euk -f long /usr/local/bin/signalp-4.1/test/euk10.fsa >
  euk10.fsa.long_out
$ diff /usr/local/bin/signalp-4.1/test/euk10.fsa.long_out euk10.fsa.long_out

$ signalp -t euk -f all /usr/local/bin/signalp-4.1/test/euk10.fsa >
  euk10.fsa.all_out
$ diff /usr/local/bin/signalp-4.1/test/euk10.fsa.all_out euk10.fsa.all_out

$ signalp -t euk -f summary /usr/local/bin/signalp-4.1/test/euk10.fsa >
  euk10.fsa.summary_out
$ diff /usr/local/bin/signalp-4.1/test/euk10.fsa.summary_out euk10.fsa.summary_out

$ man signalp
```

```
$ signalp -h
```

```
Description: Predict signal peptide and cleavage site.

Usage: /usr/local/bin/signalp -f <format>  -p <graphics-type> -k -s <networks> -t <organism-type> -m <fasta-
 file> -n <gff-file> -v -l <logfile> -u <value>  -U <value> -w -h -c <value> -T <temp dir> -V <fasta-file(s)>
Options:
-f   Setting the output format ('short', 'long', 'summary' or 'all'). Default: 'short'
-g   Graphics 'png' or 'png+eps'. Default: 'Off'
-k   Keep temporary directory. Default: 'Off'
-s   Signal peptide networks to use ('best' or 'notm'). Default: 'best'
-t   Organism type> (euk, gram+, gram-). Default: 'euk'
-m   Make fasta file with mature sequence. Default: 'Off'
-n   Make gff file of processed sequences. Default: 'Off'
-T   Specify temporary file directory. Default: /home/temp
-w   web predictions. Default: 'Off'
-u   user defined D-cutoff for noTM networks
-U   user defined D-cutoff for TM networks
-M   Minimal predicted signal peptide length. Default: [10]
-c   truncate to sequence length - 0 means no truncation. Default '70'
-l   Logfile if -v is defined. Default: 'STDERR'
-v   Verbose. Default: 'Off'
-V   Print SignalP version and exit
-h   Print this help information
```

3.16.2 TMHMM

"*Prediction of transmembrane helices in proteins*"

Download from http://www.cbs.dtu.dk/cgi-bin/nph-sw_request?tmhmm

```
# cat tmhmm-2.0c.Linux.tar.gz | uncompress | tar xvf -
# mv tmhmm-2.0c /usr/local/bin/
# chown -R root:root /usr/local/bin/tmhmm-2.0c
# chmod -R 755 /usr/local/bin/tmhmm-2.0c
# nano -w /usr/local/bin/tmhmm-2.0c/bin/tmhmm
```

```
#!/usr/bin/env perl

# This is version 2.0c of tmhmm
...
# full path to the main directory of the software
#$opt_basedir = "/usr/cbs/packages/tmhmm/2.0c/tmhmm-2.0c/";
$opt_basedir = "/usr/local/bin/tmhmm-2.0c/";
```

```
# nano -w /usr/local/bin/tmhmm-2.0c/bin/tmhmmformat.pl
```

```
#!/usr/bin/env perl

# This is version 2.0c of tmhmmformat.pl
```

```
# ln -s /usr/local/bin/tmhmm-2.0c/bin/tmhmm /usr/local/bin/
# ln -s /usr/local/bin/tmhmm-2.0c/bin/tmhmmformat.pl /usr/local/bin/
# ln -s /usr/local/bin/tmhmm-2.0c/bin/decodeanhmm.Linux_x86_64 /usr/local/bin/tmhmm-
  2.0c/bin/decodeanhmm
# ln -s /usr/local/bin/tmhmm-2.0c/bin/decodeanhmm /usr/local/bin/
```

Check installation

```
$ nano -w seq.fasta
```

```
>5H2A_CRIGR you can have comments after the ID
MEILCEDNTSLSSIPNSLMQVDGDSGLYRNDFNSRDANSSDASNWTIDGENRTNLSFEGYLPPTCLSILHL
QEKNWSALLTAVVIILTIAGNILVIMAVSLEKKLQNATNYFLMSLAIADMLLGFLVMPVSMLTILYGYRWP
LPSKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGVSMPIPVF
GLQDDSKVFKQGSCLLADDNFVLIGSFVAFFIPLTIMVITYFLTIKSLQKEATLCVSDLSTRAKLASFSFL
PQSSLSSEKLFQRSIHREPGSYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNE
HVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCQYKENRKPLQLILVNTIPALAYKSSQLQA
GQNKDSKEDAEPTDNDCSMVTLGKQQSEETCTDNINTVNEKVSCV
```

```
$ tmhmm seq.fasta
```

```
# 5H2A_CRIGR Length: 471
# 5H2A_CRIGR Number of predicted TMHs:  7
# 5H2A_CRIGR Exp number of AAs in TMHs: 159.47336
# 5H2A_CRIGR Exp number, first 60 AAs:  0.01677
# 5H2A_CRIGR Total prob of N-in:        0.00629
5H2A_CRIGR      TMHMM2.0        outside      1    76
5H2A_CRIGR      TMHMM2.0        TMhelix     77    99
5H2A_CRIGR      TMHMM2.0        inside     100   111
5H2A_CRIGR      TMHMM2.0        TMhelix    112   134
5H2A_CRIGR      TMHMM2.0        outside    135   148
5H2A_CRIGR      TMHMM2.0        TMhelix    149   171
5H2A_CRIGR      TMHMM2.0        inside     172   191
5H2A_CRIGR      TMHMM2.0        TMhelix    192   214
5H2A_CRIGR      TMHMM2.0        outside    215   233
5H2A_CRIGR      TMHMM2.0        TMhelix    234   256
5H2A_CRIGR      TMHMM2.0        inside     257   324
5H2A_CRIGR      TMHMM2.0        TMhelix    325   347
5H2A_CRIGR      TMHMM2.0        outside    348   356
5H2A_CRIGR      TMHMM2.0        TMhelix    357   379
5H2A_CRIGR      TMHMM2.0        inside     380   471
```

```
$ tmhmm --short < seq.fasta >tmhmm.out
```

```
5H2A_CRIGR      len=471 ExpAA=159.47    First60=0.02    PredHel=7       Topology=o77-99i112-134o149-171i192-
  214o234-256i325-347o357-379i
```

```
$ cat seq.fasta  | decodeanhmm -f /usr/local/bin/tmhmm-2.0c/lib/TMHMM2.0.options
  -modelfile /usr/local/bin/tmhmm-2.0c/lib//TMHMM2.0.model
```

```
/usr/local/bin/tmhmm-2.0c/lib//TMHMM2.0.model
decodeanhmm 1.1g
Copyright (C) 1998 by Anders Krogh
Mon Nov 18 10:56:39 2013

Model in file "/usr/local/bin/tmhmm-2.0c/lib//TMHMM2.0.model" parsed successfully.
>5H2A_CRIGR you can have comments after the ID
%len 471
%lett A:29 C:14 D:20 E:18 F:25 G:18 H:6 I:37 K:23 L:55 M:14 N:28 P:17 Q:19 R:15 S:50 T:27 V:34 W:7 Y:15
%score BG 1983.221633 (4.210662 per character)
%score FW 1939.519064 (4.117875 per character)
%score NB(0) 1945.771111 (4.131149 per character)
%score LO(0) 37.450523 (0.079513 per character)
%pred NB(0): o 1 76, M 77 99, i 100 111, M 112 134, o 135 148, M 149 171, i 172 191, M 192 214, o 215 233, M 234
  256, i 257 324, M 325 347, o 348 356, M 357 379, i 380 471
   MEILCEDNTSLSSIPNSLMQVDGDSGLYRNDFNSRDANSSDASNWTIDGENRTNLSFEGYLPPTCLSILHLQ
?0 oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo

   EKNWSALLTAVVIILTIAGNILVIMAVSLEKKLQNATNYFLMSLAIADMLLGFLVMPVSMLTILYGYRWPLP
?0 ooooMMMMMMMMMMMMMMMMMMMMiiiiiiiiiiiiMMMMMMMMMMMMMMMMMMMMooooooooooo

   SKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGVSMPIPVFGLQ
?0 ooooMMMMMMMMMMMMMMMMMMMMiiiiiiiiiiiiiiiiiiiiMMMMMMMMMMMMMMMMMMMMoo

   DDSKVFKQGSCLLADDNFVLIGSFVAFFIPLTIMVITYFLTIKSLQKEATLCVSDLSTRAKLASFSFLPQSS
?0 oooooooooooooooooMMMMMMMMMMMMMMMMMMMMiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii

   LSSEKLFQRSIHREPGSYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEHVIGA
?0 iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiMMMMMMMMMMMMMMMMMMMMooooooooooMMMM

   LLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCQYKENRKPLQLILVNTIPALAYKSSQLQAGQNKDS
?0 MMMMMMMMMMMMMMMMMMMiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii

   KEDAEPTDNDCSMVTLGKQQSEETCTDNINTVNEKVSCV
?0 iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii
```

3.16.3 RNAMMER

"*The RNAmmer 1.2 server predicts 5s/8s, 16s/18s, and 23s/28s ribosomal RNA in full genome sequences.*"

Install Hmmer v2

```
# wget ftp://selab.janelia.org/pub/software/hmmer/2.3.2/hmmer-2.3.2.tar.gz
# tar -xvzf hmmer-2.3.2.tar.gz
# cd hmmer-2.3.2/
# ./configure
```

```
# make
# make check
# cd ..
# mv hmmer-2.3.2 /usr/local/bin/
# chown -R root:root /usr/local/bin/hmmer-2.3.2
# chmod -R 755 /usr/local/bin/hmmer-2.3.2
# ln -s /usr/local/bin/hmmer-2.3.2/src/hmmsearch /usr/local/bin/hmmsearch2

$ hmmsearch2 -h
```

Download from http://www.cbs.dtu.dk/cgi-bin/sw_request?rnammer

```
# mkdir rnammer-1.2
# cd rnammer-1.2/
# cat ../rnammer-1.2.src.tar.Z | uncompress | tar xvf -
# cd ..
# mv rnammer-1.2 /usr/local/bin/
# chown -R root:root /usr/local/bin/rnammer-1.2
# chmod -R 755 /usr/local/bin/rnammer-1.2
# ln -s /usr/local/bin/rnammer-1.2/rnammer /usr/local/bin/
# nano -w /usr/local/bin/rnammer-1.2/rnammer
```

```
## PROGRAM CONFIGURATION BEGIN

# the path of the program
my $INSTALL_PATH = "/usr/local/bin/rnammer-1.2";

# The library in which HMMs can be found
my $HMM_LIBRARY = "$INSTALL_PATH/lib";
my $XML2GFF = "$INSTALL_PATH/xml2gff";
my $XML2FSA = "$INSTALL_PATH/xml2fsa";

# The location of the RNAmmer core module
my $RNAMMER_CORE     = "$INSTALL_PATH/core-rnammer";

# path to hmmsearch of HMMER package
chomp ( my $uname = `uname` );
my $HMMSEARCH_BINARY;
my $PERL;
if ( $uname eq "Linux" ) {
        $HMMSEARCH_BINARY = "/usr/local/bin/hmmsearch2";
        $PERL = "/usr/bin/perl";
} elsif ( $uname eq "IRIX64" ) {
        $HMMSEARCH_BINARY = "/usr/local/bin/hmmsearch2";
        $PERL = "/usr/sbin/perl";
} else {
        die "unknown platform\n";
}
```

```
# nano -w core-rnammer [remove --cpu ]
# cp /usr/local/bin/rnammer-1.2/man/rnammer.1 /usr/local/share/man/man1/
```

Check installation

```
$ mkdir example
$ cd example/
$ cp /usr/local/bin/rnammer-1.2/example/ecoli.fsa ./
$ rnammer -S bac -m lsu,ssu,tsu -xml ecoli.xml -gff ecoli.gff -h ecoli.hmmreport <
  ecoli.fsa
```

check result against /usr/local/bin/rnammer-1.2/example/example

NOTE: In command

```
$TRINOTATE_HOME/util/rnammer_support/RnammerTranscriptome.pl --transcriptome
  Trinity.fasta --path_to_rnammer /usr/bin/software/rnammer_v1.2/rnammer
```

use `/usr/local/bin/rnammer` as rnammer path

3.16.4 Trinotate

Dependencies

```
# apt-get install sqlite3 libdbd-sqlite3-perl
```

Download and install

```
# wget http://garr.dl.sourceforge.net/project/trinotate/Trinotate_r20131110.tar.gz
# tar -xvzf Trinotate_r20131110.tar.gz
# mv Trinotate_r20131110 /usr/local/bin/
# chown -R root:root /usr/local/bin/Trinotate_r20131110
# chmod -R 755 /usr/local/bin/Trinotate_r20131110
# ln -s /usr/local/bin/Trinotate_r20131110 /usr/local/bin/trinotate
```

4 Webography

[1] AMOS http://sourceforge.net/projects/amos/ visited December 2013

[2] Blast+ and Blast legacy, http://blast.ncbi.nlm.nih.gov/Blast.cgi visited December 2013

[3] Bioconductor http:/www.bioconductor.org/ visited December 2013

[4] Bioperl site, http://www.bioperl.org/wiki/Main_Page visited December 2013

[5] Blasr GitHub https://github.com/PacificBiosciences/blasr visited December 2013

[6] BWA, http://bio-bwa.sourceforge.net/ visited December 2013

[7] HMMER http://hmmer.janelia.org/ visited December 2013

[8] How much compute power do you need for next-gen sequencing? , http://ivory.idyll.org/blog/how-much-compute-ngs.html, in Living in an Ivory Basement Stochastic thoughts on science, testing, and programming. Wed 20 July 2011, By C. Titus Brown , visited December 2013

[9] Mira site http://sourceforge.net/p/mira-assembler/wiki/Home/ visited December 2013

[10] Mummer site, http://mummer.sourceforge.net/ visited December 2013

[11] Pacific Biosciences, http://www.pacificbiosciences.com/products/software/algorithms/ visited December 2013

[12] Pandaseq site https://github.com/neufeld/pandaseq visited December 2013

[13] Quiime, http://qiime.org visited December 2013

[14] R, http://www.r-project.org/ visited December 2013

[15] Rapporto tecnico Ceris-CNR n.37 Storage in HA: cluster attivo/passivo open source. Giancarlo Birello, Ivano Fucile, Valter Giovanetti, Anna Perin, June 2011, http://www.digibess.it/fedora/repository/openbess:TO094-00158, visited December 2013.

[16] Rapporto tecnico Ceris-CNR n.41 Storage in HA: manutenzione ordinaria e straordinaria. Giancarlo Birello, Ivano Fucile, Valter Giovanetti, Anna Perin, February 2012, http://www.digibess.it/fedora/repository/openbess:TO094-00162, visited December 2013.

[17] Rapporto tecnico Ceris-CNR n.44 KVM: an open-source framework for virtualization. Giancarlo Birello, Ivano Fucile, Valter Giovanetti, Anna Perin, September 2013, http://www.digibess.it/fedora/repository/openbess:TO094-00284, visited December 2013.

[18] Riding the wave of biological data, http://www.cell.com/current-biology/abstract/S0960-9822%2811%2900281-8,in Current Biology, Volume 21, Issue 6, R204-R206, by Michael Gross , visited December 2013

[19] RNAmmer, http://www.cbs.dtu.dk/services/RNAmmer/ visited December 2013

[20] Rstudio, http://www.rstudio.com/ visited December 2013

[21] SignalP, http://www.cbs.dtu.dk/services/SignalP/ visited December 2013

[22] SMALT http://www.sanger.ac.uk/resources/software/smalt/ visited December 2013

[23] SSAHA2 http://www.sanger.ac.uk/resources/software/ssaha2/ visited December 2013

[24] The International Nucleotide Sequence Database Collaboration, http://www.ncbi.nlm.nih.gov/pubmed/23180798,in Nucleic Acids Res. ,2013 Jan, 41(Database issue):D21-4, by Nakamura Y, Cochrane G, Karsch-Mizrachi I; visited December 2013.

[25] TMHMM http://www.cbs.dtu.dk/services/TMHMM/ visited December 2013

[26] Tophat, http://tophat.cbcb.umd.edu/ visited December 2013

[27] Trinity, http://trinityrnaseq.sourceforge.net/ visited December 2013

[28] Trinotate, http://trinotate.sourceforge.net/ visited December 2013

[29] WGS-assembler, http://sourceforge.net/apps/mediawiki/wgs-assembler visited December 2013

5 Appendix

5.1 Qiime

5.1.1 /usr/share/qiime-1.7.0/activate.sh

```
export QIIME_CONFIG_FP=/usr/share/qiime-1.7.0/qiime-1.7.0-release/../qiime_config
export LAPACK=None
export PYTHONPATH=/usr/share/qiime-1.7.0/emperor-0.0.0-repository-48cac57c/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/emperor-0.0.0-repository-48cac57c/lib/:/usr/share/qiime-
  1.7.0/qiime-galaxy-0.0.1-repository-5b980770/lib/:/usr/share/qiime-1.7.0/qiime-1.7.0-
  release/lib/python2.7/site-packages:/usr/share/qiime-1.7.0/qiime-1.7.0-
  release/lib/:/usr/share/qiime-1.7.0/pprospector-1.0.1-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/tax2tree-1.0-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/pynast-1.2-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/matplotlib-1.1.0-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/ipython-latest-repository-23255a83/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/pycogent-1.5.3-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/biom-format-1.1.2-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/MySQL-python-1.2.3-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/sphinx-1.0.4-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/setuptools-0.6c11-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/mpi4py-1.2.2-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/tornado-2.2.1-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/gdata-2.0.17-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/SQLAlchemy-0.7.1-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/pyzmq-2.1.11-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/numpy-1.5.1-release/lib/python2.7/site-
  packages:/usr/share/qiime-1.7.0/cython-0.17-release/lib/python2.7/site-packages:$PYTHONPATH
export BLAS=None
export TEST_DB=1
export BLASTMAT=/usr/share/qiime-1.7.0/blast-2.2.22-release/data
export ATLAS=None
export PYRO_LOOKUP_FILE=/usr/share/qiime-1.7.0/ampliconnoise-1.27-release/Data/LookUp_E123.dat
export RDP_JAR_PATH=/usr/share/qiime-1.7.0/rdpclassifier-2.2-release/rdp_classifier-2.2.jar
export PYCOGENT=/usr/share/qiime-1.7.0/pycogent-1.5.3-release/./
export SEQ_LOOKUP_FILE=/usr/share/qiime-1.7.0/ampliconnoise-1.27-release/Data/Tran.dat
export PATH=/usr/share/qiime-1.7.0/emperor-0.0.0-repository-48cac57c/bin:/usr/share/qiime-
  1.7.0/qiime-galaxy-0.0.1-repository-5b980770/scripts:/usr/share/qiime-1.7.0/qiime-1.7.0-
  release/bin:/usr/share/qiime-1.7.0/r-2.12.0-release/bin:/usr/share/qiime-1.7.0/pprospector-
  1.0.1-release/bin:/usr/share/qiime-1.7.0/tax2tree-1.0-release/bin:/usr/share/qiime-
  1.7.0/cytoscape-2.7.0-release/.:/usr/share/qiime-1.7.0/ipython-latest-repository-
  23255a83/bin:/usr/share/qiime-1.7.0/pynast-1.2-release/bin:/usr/share/qiime-1.7.0/biom-format-
  1.1.2-release/bin:/usr/share/qiime-1.7.0/tornado-2.2.1-release/bin:/usr/share/qiime-
  1.7.0/cython-0.17-release/bin:/usr/share/qiime-1.7.0/pyzmq-2.1.11-
  release/bin:/usr/share/qiime-1.7.0/blast-2.2.22-release/bin:/usr/share/qiime-1.7.0/uclust-
  1.2.22-release/.:/usr/share/qiime-1.7.0/cdhit-3.1-release/.:/usr/share/qiime-
  1.7.0/rdpclassifier-2.2-release/.:/usr/share/qiime-1.7.0/parsinsert-1.0.4-
  release/.:/usr/share/qiime-1.7.0/muscle-3.8.31-release/.:/usr/share/qiime-1.7.0/mothur-1.25.0-
  release/.:/usr/share/qiime-1.7.0/bwa-0.6.2-release/.:/usr/share/qiime-1.7.0/infernal-1.0.2-
  release/bin:/usr/share/qiime-1.7.0/fasttree-2.1.3-release/.:/usr/share/qiime-1.7.0/cdbtools-
  10.11.2010-release/.:/usr/share/qiime-1.7.0/python-2.7.3-release/bin:/usr/share/qiime-
  1.7.0/vienna-1.8.4-release/.:/usr/share/qiime-1.7.0/chimeraslayer-4.29.2010-
  release/ChimeraSlayer:/usr/share/qiime-1.7.0/chimeraslayer-4.29.2010-release/NAST-
  iEr:/usr/share/qiime-1.7.0/drisee-1.2-release/.:/usr/share/qiime-1.7.0/sourcetracker-0.9.5-
  release/.:/usr/share/qiime-1.7.0/ampliconnoise-1.27-release/Scripts:/usr/share/qiime-
  1.7.0/ampliconnoise-1.27-release/bin:/usr/share/qiime-1.7.0/raxml-7.3.0-
  release/.:/usr/share/qiime-1.7.0/blat-34-release/.:/usr/share/qiime-1.7.0/rtax-0.983-
  release/.:/usr/share/qiime-1.7.0/clearcut-1.0.9-release/.:/usr/share/qiime-1.7.0/usearch-
  5.2.236-release/.:/usr/share/qiime-1.7.0/pplacer-v1.1.alpha14-Linux-3.2.0/.:/usr/share/qiime-
  1.7.0/pplacer-v1.1.alpha14-Linux-3.2.0/scripts/.:$PATH
export SOURCETRACKER_PATH=/usr/share/qiime-1.7.0/sourcetracker-0.9.5-release/.
export LD_LIBRARY_PATH=/usr/share/qiime-1.7.0/python-2.7.3-release/lib
export QIIME=/usr/share/qiime-1.7.0/qiime-1.7.0-release/./
```

5.2 Blasr

5.2.1 alignment/Makefile

```
#
# common.mk contains the configuration for this build setup.
#
#
ifeq ($(origin PREFIX), undefined)
  ifeq ($(origin SEYMOUR_HOME), undefined)
    PREFIX = /opt/seymour
  else
    PREFIX = $(SEYMOUR_HOME)
  endif
endif
ANALYSIS_HOME=$(PREFIX)/analysis

PBCPP_DIR = ..

include ../common.mk

#
# Define the targets before including the rules since the rules contains a target itself.
#

EXECS = wordCounter printReadWordCount blasr sdpMatcher swMatcher kbandMatcher sawriter saquery samodify
   printTupleCountTable cmpPrintTupleCountTable malign removeAdapters tabulateAlignment samatcher saprinter
   buildQualityValueProfile guidedalign extendAlign sals pbmask

# DISABLE for now
#cmpMatcher


all: bin make.dep $(EXECS)

BUILTEXECS = $(addprefix bin/, $(EXECS))
DISTRIB_SET = blasr swMatcher kbandMatcher sawriter samodify printTupleCountTable cmpPrintTupleCountTable
   removeAdapters sdpMatcher pbmask
DISTRIB_EXECS = $(addprefix bin/, $(DISTRIB_SET))
INSTALL_EXECS = $(addprefix install-, $(DISTRIB_SET))


include ../make.rules

ifneq ($(wildcard make.dep),)
    include make.dep
endif

wordCounter:        bin/wordCounter
printReadWordCount: bin/printReadWordCount
blasr:        bin/blasr
cmpMatcher:         bin/cmpMatcher
sdpMatcher:         bin/sdpMatcher
samatcher:          bin/samatcher
saprinter:          bin/saprinter
sals:               bin/sals
swMatcher:          bin/swMatcher
kbandMatcher:       bin/kbandMatcher
sawriter:           bin/sawriter
saquery:            bin/saquery
samodify:           bin/samodify
printTupleCountTable:bin/printTupleCountTable
cmpPrintTupleCountTable:bin/cmpPrintTupleCountTable
malign:             bin/malign
removeAdapters:     bin/removeAdapters
tabulateAlignment:  bin/tabulateAlignment
buildQualityValueProfile: bin/buildQualityValueProfile
guidedalign: bin/guidedalign
extendAlign: bin/extendAlign
pbmask: bin/pbmask

bin/sawriter: bin/SAWriter.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/pbmask: bin/Mask.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@ -L$(HDF5LIBDIR) -l$(HDF5LIBCPP) -l$(HDF5LIB) -lz -lpthread

bin/guidedalign: bin/GuidedAlign.o
	$(CPP) $(CPPOPTS) $^ $(STATIC) -o $@ -L$(HDF5LIBDIR) -l$(HDF5LIBCPP) -l$(HDF5LIB) -lz -lpthread

bin/extendAlign: bin/ExtendAlign.o
	$(CPP) $(CPPOPTS) $^ $(STATIC) -o $@ -L$(HDF5LIBDIR) -l$(HDF5LIBCPP) -l$(HDF5LIB) -lz -lpthread

bin/saquery: bin/SAQuery.o
	$(CPP) $(CPPOPTS) $^ $(STATIC) -o $@

bin/samodify: bin/SAModify.o
	$(CPP) $(CPPOPTS) $^ $(STATIC) -o $@
```

```
bin/wordCounter: bin/WordCounter.o
	$(CPP) $(CPPOPTS) $^ $(STATIC) -o $@

bin/printReadWordCount: bin/PrintReadWordCount.o
	$(CPP) $(CPPOPTS) $^ $(STATIC) -o $@

#
# Add nonstandard compilation rules to build compressed dna aligners
#

bin/CmpReadMatcher.o: ReadMatcher.cpp
	$(CPP) -c $(CPPOPTS) -DCOMPRESSED $< $(STATIC) -o $@

bin/cmpMatcher: bin/CmpReadMatcher.o bin/PositionTable.o
	$(CPP) $(CPPOPTS) $< bin/PositionTable.o $(STATIC) -o $@ -L$(HDF5LIBDIR) -l$(HDF5LIBCPP) -l$(HDF5LIB)
  -lz -lpthread

bin/CmpPrintTupleCountTable.o: PrintTupleCountTable.cpp
	$(CPP) -c $(CPPOPTS) -DCOMPRESSED $^ $(STATIC) -o $@

bin/cmpPrintTupleCountTable: bin/CmpPrintTupleCountTable.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

ifneq ($(shell uname -s),Darwin)
    LRT = -lrt
endif
bin/blasr: bin/Blasr.o
	$(CPP) $(CPPOPTS) $< -L$(HDF5LIBDIR) -l$(HDF5LIBCPP) -l$(HDF5LIB) $(LINK_PROFILER) -lpthread -lz $(LRT)
  -ldl $(STATIC) -o bin/blasr

bin/samatcher: bin/SAMatcher.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/saprinter: bin/SAPrinter.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/sals: bin/SALS.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/sdpMatcher: bin/SDPMatcher.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/malign: bin/MAlign.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/swMatcher: bin/SWMatcher.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/removeAdapters: bin/RemoveAdapters.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/kbandMatcher: bin/KBandMatcher.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/printTupleCountTable: bin/PrintTupleCountTable.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/tabulateAlignment: bin/TabulateAlignment.o
	$(CPP) $(CPPOPTS) $< $(STATIC) -o $@

bin/buildQualityValueProfile: bin/BuildQualityValueProfile.o
	$(CPP) -g $(CPPOPTS) $< $(STATIC) -o $@ -L$(HDF5LIBDIR) -l$(HDF5LIBCPP) -l$(HDF5LIB) -lz -lpthread

#
# Set up a default value for the install dir if one does
# not exist.
#
INSTALL_DIR ?= $(ANALYSIS_HOME)/bin
BUILD_DIR ?= bin
install:
	/usr/bin/install -d $(INSTALL_DIR)
	/usr/bin/install -m 555 $(DISTRIB_EXECS) $(INSTALL_DIR)

install-%:
	/usr/bin/install -d $(INSTALL_DIR)
	/usr/bin/install -m 555 $(DISTRIB_EXECS) $(INSTALL_DIR)
```