

Background

We provide gene annotation for the human genome.

The latest assembly available is GRCh38 but some users still use the previous assembly, GRCh37, and want to convert coordinates between the two assemblies.

You can find more information about genome assemblies here:

<https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/>

For this exercise, we use the latest Ensembl release, 104.

SM note 3: This question is about conversion of gene coordinates between two assemblies.

<https://www.ncbi.nlm.nih.gov/assembly/88331>

✓ GRCh38 - hg38 - Genome - Assembly - NCBI

<https://www.ncbi.nlm.nih.gov/assembly/88331>

1. GRCh38 Genome Reference Consortium Human Build 38 Organism: Homo sapiens (human) Submitter: Genome Reference Consortium Date: 2013/12/17 Assembly type: haploid-with-alt-loci Assembly level: Chromosome Genome representation: full Synonyms: hg38 GenBank assembly accession: GCA_000001405.15 (replaced) RefSeq assembly accession: GCF_000001405.26 (replaced) IDs: 88331[UID] 883148 [GenBank ...

<https://www.ncbi.nlm.nih.gov/assembly/2758/>

✓ GRCh37 - hg19 - Genome - Assembly - NCBI

<https://www.ncbi.nlm.nih.gov/assembly/2758>

1. GRCh37 Genome Reference Consortium Human Build 37 (GRCh37) Organism: Homo sapiens (human) Submitter: Genome Reference Consortium Date: 2009/02/27 Assembly type: haploid-with-alt-loci Assembly level: Chromosome Genome representation: full Synonyms: hg19 GenBank assembly accession: GCA_000001405.1 (replaced) RefSeq assembly accession: GCF_000001405.13 (replaced) IDs: 2758[UID] 2468 [GenBank ...

end of SM note 3

Ensembl Perl API

The Ensembl Perl API is used to generate annotation as well as access it programmatically.

You can find more information about it here:

http://www.ensembl.org/info/docs/api/core/core_tutorial.html

SM note 4: Setup and installation for Exercise. Ensembl Perl API tutorial (with video)

http://uswest.ensembl.org/info/docs/api/api_installation.html was very helpful.

Presently, I did not make the additional installations for 'Variation genotype and frequency data'.

I used Ubuntu Release 20.04. Typing command 'lsb_release -a' on Terminal displays this information. *Please, note that a) text after '#' is for human not computer, b) text I greyed out is the command that needs to be typed on or copied to Terminal or Atom as the case maybe and c) after typing or copying the command hit Return/Enter key for it to execute.*

lsb_release -a #Displays version of Ubuntu d) Clicking on Terminal window and pressing Ctrl+C or Cmd+C ends any code running on Terminal.

Step 1: Install pre-requisite Perl language <http://www.ubuntu-server.com/ubuntu/how-to-install-perl-on-ubuntu/> and Perl modules using apt-get or apt <https://phoenixnap.com/kb/apt-vs-apt-get>

#prepare 'apt' for installation, 'apt-get' can also be used

sudo apt update

sudo apt upgrade

perl -v #if perl is installed this will provide version of perl, but if don't have perl then it will say 'not found'.

sudo apt install perl #install with this line if perl not available or skip to next line

sudo apt-get install libdbi-perl #installs DBI module

sudo apt-get install libmysqlclient-dev #installs mysql module

sudo apt-get install libdbd-mysql-perl #installs DBD mysql module

```
~/bashrc is being used by Terminal when Terminal is opened
shrad@shrad-GF75-Thin-9SC:~$ perl -v

This is perl 5, version 30, subversion 0 (v5.30.0) built for x86_64-linux-gnu-thread-multi
(with 50 registered patches, see perl -V for more detail)

Copyright 1987-2019, Larry Wall

Perl may be copied only under the terms of either the Artistic License or the
GNU General Public License, which may be found in the Perl 5 source kit.

Complete documentation for Perl, including FAQ lists, should be found on
this system using "man perl" or "perldoc perl". If you have access to the
Internet, point your browser at http://www.perl.org/, the Perl Home Page.

shrad@shrad-GF75-Thin-9SC:~$ sudo apt-get install libdbi-perl
[sudo] password for shrad:
Reading package lists... Done
Building dependency tree
Reading state information... Done
shrad@shrad-GF75-Thin-9SC:~$ sudo apt-get install libmysqlclient-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
shrad@shrad-GF75-Thin-9SC:~/embl/ensembl/misc-scripts$ sudo apt-get install libdbd-mysql-perl
Reading package lists... Done
Building dependency tree
Reading state information... Done
```

Step 2: Click this link http://uswest.ensembl.org/info/docs/api/api_installation.html for information about installation of API and BioPerl. I used gzipped tar format download and installation from Terminal over cloning from github.

Ref relevant for github <https://phoenixnap.com/kb/how-to-install-git-on-ubuntu>, <https://www.howtogeek.com/451360/how-to-clone-a-github-repository/> and http://uswest.ensembl.org/info/docs/api/api_git.html.

Step 3: Type the following commands on Terminal to download gzipped tar formats and install API and BioPerl.

cd #go to home directory

pwd #check directory location

mkdir embl #create new folder for this project

cd embl #go to embl directory to do the installation

wget http://ftp.ensembl.org/pub/ensembl-api.tar.gz #downloads ensembl-api into embl folder

wget https://cpan.metacpan.org/authors/id/C/CJ/CJFIELDS/BioPerl-1.6.924.tar.gz #downloads Bio-Perl into embl folder

```
shrad@shrad-GF75-Thin-9SC:~/Desktop$ cd
shrad@shrad-GF75-Thin-9SC:~$ pwd
/home/shrad
shrad@shrad-GF75-Thin-9SC:~$ cd
shrad@shrad-GF75-Thin-9SC:~$ mkdir embl
shrad@shrad-GF75-Thin-9SC:~$ cd embl
shrad@shrad-GF75-Thin-9SC:~/embl$ wget http://ftp.ensembl.org/pub/ensembl-api.tar.gz
shrad@shrad-GF75-Thin-9SC:~/embl$ wget https://cpan.metacpan.org/authors/id/C/CJ/CJFIELDS/BioPerl-1.6.924.tar.gz
```

Step 4 Now type the following commands on Terminal to unzip the downloaded API and BioPerl.

```
tar xzvf ensembl-api.tar.gz #extract API
tar xzvf BioPerl-1.6.924.tar.gz #extract BioPerl
```

Step 5 Add the above downloaded libraries in path of “.bashrc” file. Its a configuration file Terminal uses every time we open it. Adding libraries to path ensures that every time we run perl script or write codes in Terminal we don’t need to tell perl where the installations are it will know from the bashrc file itself. <https://www.linuxfordevices.com/tutorials/linux/bashrc-and-bash-profile> Feel free to use ‘.bash_profile’ or ‘.profile’ instead.

ls -a #view hidden files name starting with .

atom ~/.bashrc #this opens file in atom editor so can add lines of code or use any editor you like such as ‘nano’. Instead of ‘bashrc’ can also use ‘profile’ file.

#in atom add the following lines at the end of the file

```
PERL5LIB=${PERL5LIB}:${HOME}/embl/BioPerl-1.6.924
PERL5LIB=${PERL5LIB}:${HOME}/embl/ensembl/modules
PERL5LIB=${PERL5LIB}:${HOME}/embl/ensembl-compara/modules
PERL5LIB=${PERL5LIB}:${HOME}/embl/ensembl-variation/modules
PERL5LIB=${PERL5LIB}:${HOME}/embl/ensembl-funcgen/modules
export PERL5LIB
```

#save and close file in atom when done and go back to Terminal window.

source ~/.bashrc #put changes to ‘bashrc’ or ‘profile’ file into effect

or just restart terminal to make changes take effect. To exit terminal type ‘exit’

```
~/.bashrc is being used by Terminal when Terminal is opened
shrad@shrad-GF75-Thin-9SC:~$ ls -a
.          .config          .local          .Rhistory
..         .continuum       .mongorc.js     Save
.anaconda  .dbshell         .mozilla        snap
anaconda3  Desktop         Music           .ssh
.anaconda_backup Documents        .node_repl_history .sudo_as_admin_successful
.atom      Downloads       Pictures        Templates
.bash_history embl            .pki           .thunderbird
.bash_logout .gnupg         .profile       Videos
.bashrc     .gphoto        Public         .vscode
.bashrc.save .ipython       PythonSpyderStudy .wget-hsts
.cache      .jupyter
.conda      .kite
.condarc    'LivePatch Ubuntu Advantage for Infrastructure.pdf'
shrad@shrad-GF75-Thin-9SC:~$ atom ~/.bashrc
shrad@shrad-GF75-Thin-9SC:~$ source ~/.bashrc
```

Step 6 Now lets check if installation has worked by typing following commands in Terminal.

echo \$PERL5LIB #This displays all the lines we added to ‘bashrc’ or ‘profile’ file

cd embl/misc-scripts/ #This takes us into the ‘misc-scripts’ directory that has the test script

./ping_ensembl.pl #Will say ‘installation good’ so all is ok, but if any problems it will say so.

#This takes a little while to finish executing.

```

~/.bashrc is being used by Terminal when Terminal is opened
shrad@shrad-GF75-Thin-9SC:~$ echo $PERL5LIB
:/home/shrad/embl/bioperl-1.6.924:/home/shrad/embl/ensembl/modules:/home/shrad/embl/ensembl-comp
ara/modules:/home/shrad/embl/ensembl-variation/modules:/home/shrad/embl/ensembl-funcgen/modules
shrad@shrad-GF75-Thin-9SC:~$ cd ./embl/ensembl/misc-scripts
shrad@shrad-GF75-Thin-9SC:~/embl/ensembl/misc-scripts$ ./ping_ensembl.pl
shrad@shrad-GF75-Thin-9SC:~$ cd ./embl/ensembl/misc-scripts
shrad@shrad-GF75-Thin-9SC:~/embl/ensembl/misc-scripts$ ./ping_ensembl.pl
Installation is good. Connection to Ensembl works and you can query the human core database
shrad@shrad-GF75-Thin-9SC:~/embl/ensembl/misc-scripts$

```

Type 'top' in a new Terminal window or tab to see progress. Also, if in the Terminal window the cursor is blinking means its working.

```

shrad@shrad-GF75-Thin-9SC: ~
~/.bashrc is being used by Terminal when Terminal is opened
shrad@shrad-GF75-Thin-9SC:~$ top

```

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
1951	shrad	20	0	4918660	264464	105492	S	7.0	0.4	2:42.44	gnome-shell
1792	shrad	20	0	1180532	91444	46884	S	3.0	0.1	2:38.69	Xorg
7169	shrad	20	0	4790320	265568	102888	S	1.0	0.4	0:16.20	atom
97	root	20	0	0	0	0	I	0.3	0.0	0:12.99	kworker/1:1-events
184	root	0	-20	0	0	0	I	0.3	0.0	0:06.93	kworker/u25:0-i915_flip
1268	mongodb	20	0	1514840	101704	41860	S	0.3	0.2	0:20.19	mongod
1272	kernoops	20	0	11264	444	0	S	0.3	0.0	0:00.11	kerneloops
1691	shrad	20	0	1629380	188468	86288	S	0.3	0.3	0:12.33	kited
4115	root	20	0	0	0	0	I	0.3	0.0	0:01.02	kworker/u24:0-events_unbound
6913	root	0	-20	0	0	0	I	0.3	0.0	0:02.37	kworker/u25:3-i915_flip

end of SM note 4