(base) leonid@leonid-Aspire-A315-42:~$ pip install bio

Collecting bio

Downloading bio-1.3.9-py3-none-any.whl (270 kB)

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 270.9/270.9 kB 1.9 MB/s eta 0:00:00

Collecting biopython>=1.79

Using cached biopython-1.79-cp38-cp38-manylinux\_2\_5\_x86\_64.manylinux1\_x86\_64.whl (2.3 MB)

Requirement already satisfied: tqdm in ./anaconda3/lib/python3.8/site-packages (from bio) (4.50.2)

Collecting mygene

Downloading mygene-3.2.2-py2.py3-none-any.whl (5.4 kB)

Requirement already satisfied: requests in ./anaconda3/lib/python3.8/site-packages (from bio) (2.24.0)

Requirement already satisfied: numpy in ./anaconda3/lib/python3.8/site-packages (from biopython>=1.79->bio) (1.22.3)

Collecting biothings-client>=0.2.6

Downloading biothings\_client-0.2.6-py2.py3-none-any.whl (37 kB)

Requirement already satisfied: idna<3,>=2.5 in ./anaconda3/lib/python3.8/site-packages (from requests->bio) (2.10)

Requirement already satisfied: certifi>=2017.4.17 in ./anaconda3/lib/python3.8/site-packages (from requests->bio) (2020.6.20)

Requirement already satisfied: chardet<4,>=3.0.2 in ./anaconda3/lib/python3.8/site-packages (from requests->bio) (3.0.4)

Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in ./anaconda3/lib/python3.8/site-packages (from requests->bio) (1.25.11)

Installing collected packages: biopython, biothings-client, mygene, bio

Successfully installed bio-1.3.9 biopython-1.79 biothings-client-0.2.6 mygene-3.2.2

(base) leonid@leonid-Aspire-A315-42:~$ wget https://raw.githubusercontent.com/vanya-antonov/hse22-project/main/zhunt3-alan.c

--2022-06-08 17:04:11-- https://raw.githubusercontent.com/vanya-antonov/hse22-project/main/zhunt3-alan.c

Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.109.133, 185.199.108.133, 185.199.111.133, ...

Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.109.133|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 23973 (23K) [text/plain]

Saving to: ‘zhunt3-alan.c’

zhunt3-alan.c 100%[===================================================================================>] 23,41K --.-KB/s in 0,001s

2022-06-08 17:04:12 (44,5 MB/s) - ‘zhunt3-alan.c’ saved [23973/23973]

(base) leonid@leonid-Aspire-A315-42:~$ gcc zhunt3-alan.c -lm -o zhunt3

zhunt3-alan.c: In function ‘user\_regret’:

zhunt3-alan.c:336:9: warning: implicit declaration of function ‘gets’; did you mean ‘fgets’? [-Wimplicit-function-declaration]

336 | gets(tempstr);

| ^~~~

| fgets

/usr/bin/ld: /tmp/ccmJcbMv.o: in function `user\_regret':

zhunt3-alan.c:(.text+0xba7): warning: the `gets' function is dangerous and should not be used.

(base) leonid@leonid-Aspire-A315-42:~$ conda install -c bioconda bedtools

Collecting package metadata (current\_repodata.json): done

Solving environment: done

## Package Plan ##

environment location: /home/leonid/anaconda3

added / updated specs:

- bedtools

The following packages will be downloaded:

package | build

---------------------------|-----------------

bedtools-2.30.0 | hc088bd4\_0 14.0 MB bioconda

conda-4.13.0 | py38h06a4308\_0 896 KB

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

Total: 14.9 MB

The following NEW packages will be INSTALLED:

bedtools bioconda/linux-64::bedtools-2.30.0-hc088bd4\_0

The following packages will be UPDATED:

conda conda-forge::conda-4.11.0-py38h578d9b~ --> pkgs/main::conda-4.13.0-py38h06a4308\_0

Proceed ([y]/n)? y

Downloading and Extracting Packages

conda-4.13.0 | 896 KB | ########################################################################################################################### | 100%

bedtools-2.30.0 | 14.0 MB | ########################################################################################################################### | 100%

Preparing transaction: done

Verifying transaction: done

Executing transaction: done

(base) leonid@leonid-Aspire-A315-42:~$ vim merge\_bed.sh

(base) leonid@leonid-Aspire-A315-42:~$ bash merge\_bed.sh

Z\_10\_falciparum.bed

Z\_10\_gaboni.bed

Z\_10\_knowlesi.bed

Z\_10\_vivax.bed

Z\_10\_yoelii.bed

Z\_11\_falciparum.bed

Z\_11\_gaboni.bed

Z\_11\_knowlesi.bed

Z\_11\_vivax.bed

Z\_11\_yoelii.bed

Z\_12\_falciparum.bed

Z\_12\_gaboni.bed

Z\_12\_knowlesi.bed

Z\_12\_vivax.bed

Z\_12\_yoelii.bed

Z\_13\_falciparum.bed

Z\_13\_gaboni.bed

Z\_13\_knowlesi.bed

Z\_13\_vivax.bed

Z\_13\_yoelii.bed

Z\_14\_falciparum.bed

Z\_14\_gaboni.bed

Z\_14\_knowlesi.bed

Z\_14\_vivax.bed

Z\_14\_yoelii.bed

Z\_1\_falciparum.bed

Z\_1\_gaboni.bed

Z\_1\_knowlesi.bed

Z\_1\_vivax.bed

Z\_1\_yoelii.bed

Z\_2\_falciparum.bed

Z\_2\_gaboni.bed

Z\_2\_knowlesi.bed

Z\_2\_vivax.bed

Z\_2\_yoelii.bed

Z\_3\_falciparum.bed

Z\_3\_gaboni.bed

Z\_3\_knowlesi.bed

Z\_3\_vivax.bed

Z\_3\_yoelii.bed

Z\_4\_falciparum.bed

Z\_4\_gaboni.bed

Z\_4\_knowlesi.bed

Z\_4\_vivax.bed

Z\_4\_yoelii.bed

Z\_5\_falciparum.bed

Z\_5\_gaboni.bed

Z\_5\_knowlesi.bed

Z\_5\_vivax.bed

Z\_5\_yoelii.bed

Z\_6\_falciparum.bed

Z\_6\_gaboni.bed

Z\_6\_knowlesi.bed

Z\_6\_vivax.bed

Z\_6\_yoelii.bed

Z\_7\_falciparum.bed

Z\_7\_gaboni.bed

Z\_7\_knowlesi.bed

Z\_7\_vivax.bed

Z\_7\_yoelii.bed

Z\_8\_falciparum.bed

Z\_8\_gaboni.bed

Z\_8\_knowlesi.bed

Z\_8\_vivax.bed

Z\_8\_yoelii.bed

Z\_9\_falciparum.bed

Z\_9\_gaboni.bed

Z\_9\_knowlesi.bed

Z\_9\_vivax.bed

Z\_9\_yoelii.bed

(base) leonid@leonid-Aspire-A315-42:~/PycharmProjects/python/minor\_2022/project$

(base) leonid@leonid-Aspire-A315-42:~$ pip install dna\_features\_viewer

Collecting dna\_features\_viewer

Downloading dna\_features\_viewer-3.1.1-py3-none-any.whl (30 kB)

Requirement already satisfied: Biopython in ./anaconda3/lib/python3.8/site-packages (from dna\_features\_viewer) (1.79)

Requirement already satisfied: matplotlib>=3 in ./anaconda3/lib/python3.8/site-packages (from dna\_features\_viewer) (3.3.2)

Requirement already satisfied: packaging in ./anaconda3/lib/python3.8/site-packages (from dna\_features\_viewer) (20.4)

Requirement already satisfied: certifi>=2020.06.20 in ./anaconda3/lib/python3.8/site-packages (from matplotlib>=3->dna\_features\_viewer) (2020.6.20)

Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in ./anaconda3/lib/python3.8/site-packages (from matplotlib>=3->dna\_features\_viewer) (2.4.7)

Requirement already satisfied: kiwisolver>=1.0.1 in ./anaconda3/lib/python3.8/site-packages (from matplotlib>=3->dna\_features\_viewer) (1.3.0)

Requirement already satisfied: numpy>=1.15 in ./anaconda3/lib/python3.8/site-packages (from matplotlib>=3->dna\_features\_viewer) (1.22.3)

Requirement already satisfied: cycler>=0.10 in ./anaconda3/lib/python3.8/site-packages (from matplotlib>=3->dna\_features\_viewer) (0.10.0)

Requirement already satisfied: python-dateutil>=2.1 in ./anaconda3/lib/python3.8/site-packages (from matplotlib>=3->dna\_features\_viewer) (2.8.1)

Requirement already satisfied: pillow>=6.2.0 in ./anaconda3/lib/python3.8/site-packages (from matplotlib>=3->dna\_features\_viewer) (8.0.1)

Requirement already satisfied: six in ./anaconda3/lib/python3.8/site-packages (from packaging->dna\_features\_viewer) (1.15.0)

Installing collected packages: dna\_features\_viewer

Successfully installed dna\_features\_viewer-3.1.1

(base) leonid@leonid-Aspire-A315-42:~$ apt-get install proteinortho

E: Could not open lock file /var/lib/dpkg/lock-frontend - open (13: Permission denied)

E: Unable to acquire the dpkg frontend lock (/var/lib/dpkg/lock-frontend), are you root?

(base) leonid@leonid-Aspire-A315-42:~$ sudo apt-get install proteinortho

[sudo] password for leonid:

Reading package lists... Done

Building dependency tree

Reading state information... Done

The following package was automatically installed and is no longer required:

libjpeg62

Use 'sudo apt autoremove' to remove it.

The following additional packages will be installed:

blends-common diamond-aligner med-config

Suggested packages:

blends-doc

The following NEW packages will be installed:

blends-common diamond-aligner med-config proteinortho

0 upgraded, 4 newly installed, 0 to remove and 6 not upgraded.

Need to get 879 kB of archives.

After this operation, 2 867 kB of additional disk space will be used.

Do you want to continue? [Y/n] Y

Get:1 http://archive.ubuntu.com/ubuntu focal/universe amd64 blends-common all 0.7.2ubuntu3 [15,3 kB]

Get:2 http://archive.ubuntu.com/ubuntu focal/universe amd64 diamond-aligner amd64 0.9.30-3 [644 kB]

Get:3 http://archive.ubuntu.com/ubuntu focal/universe amd64 proteinortho amd64 6.0.14+dfsg-1 [207 kB]

Get:4 http://archive.ubuntu.com/ubuntu focal/universe amd64 med-config all 3.5.1 [11,9 kB]

Fetched 879 kB in 2s (443 kB/s)

Preconfiguring packages ...

/tmp/med-config.config.H8Wj1V: 13: .: Can't open /usr/share/blends/blend-utils

med-config failed to preconfigure, with exit status 127

Selecting previously unselected package blends-common.

(Reading database ... 235233 files and directories currently installed.)

Preparing to unpack .../blends-common\_0.7.2ubuntu3\_all.deb ...

Unpacking blends-common (0.7.2ubuntu3) ...

Selecting previously unselected package diamond-aligner.

Preparing to unpack .../diamond-aligner\_0.9.30-3\_amd64.deb ...

Unpacking diamond-aligner (0.9.30-3) ...

Selecting previously unselected package proteinortho.

Preparing to unpack .../proteinortho\_6.0.14+dfsg-1\_amd64.deb ...

Unpacking proteinortho (6.0.14+dfsg-1) ...

Selecting previously unselected package med-config.

Preparing to unpack .../med-config\_3.5.1\_all.deb ...

Unpacking med-config (3.5.1) ...

Setting up diamond-aligner (0.9.30-3) ...

Setting up proteinortho (6.0.14+dfsg-1) ...

Setting up blends-common (0.7.2ubuntu3) ...

Setting up med-config (3.5.1) ...

Processing triggers for menu (2.1.47ubuntu4) ...

Processing triggers for man-db (2.9.1-1) ...

(base) leonid@leonid-Aspire-A315-42:~$ wget -O falciparum\_protein.faa.gz https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/002/765/GCA\_000002765.3\_GCA\_000002765/GCA\_000002765.3\_GCA\_000002765\_protein.faa.gz

--2022-06-13 19:57:22-- https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/002/765/GCA\_000002765.3\_GCA\_000002765/GCA\_000002765.3\_GCA\_000002765\_protein.faa.gz

Resolving ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)... 130.14.250.10, 130.14.250.11, 2607:f220:41f:250::230, ...

Connecting to ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)|130.14.250.10|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 2413927 (2,3M) [application/x-gzip]

Saving to: ‘falciparum\_protein.faa.gz’

falciparum\_protein.faa.gz 100%[===================================================================================>] 2,30M 1,23MB/s in 1,9s

2022-06-13 19:57:25 (1,23 MB/s) - ‘falciparum\_protein.faa.gz’ saved [2413927/2413927]

(base) leonid@leonid-Aspire-A315-42:~$ wget -O vivax\_protein.faa.gz https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/002/415/GCA\_000002415.2\_ASM241v2/GCA\_000002415.2\_ASM241v2\_protein.faa.gz

--2022-06-13 19:58:17-- https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/002/415/GCA\_000002415.2\_ASM241v2/GCA\_000002415.2\_ASM241v2\_protein.faa.gz

Resolving ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)... 130.14.250.10, 130.14.250.11, 2607:f220:41f:250::230, ...

Connecting to ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)|130.14.250.10|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 2293848 (2,2M) [application/x-gzip]

Saving to: ‘vivax\_protein.faa.gz’

vivax\_protein.faa.gz 100%[===================================================================================>] 2,19M 1023KB/s in 2,2s

2022-06-13 19:58:20 (1023 KB/s) - ‘vivax\_protein.faa.gz’ saved [2293848/2293848]

(base) leonid@leonid-Aspire-A315-42:~$ wget -O gaboni\_protein.faa.gz https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/602/025/GCA\_001602025.1\_ASM160202v1/GCA\_001602025.1\_ASM160202v1\_protein.faa.gz

--2022-06-13 19:59:08-- https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/602/025/GCA\_001602025.1\_ASM160202v1/GCA\_001602025.1\_ASM160202v1\_protein.faa.gz

Resolving ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)... 130.14.250.10, 130.14.250.11, 2607:f220:41f:250::230, ...

Connecting to ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)|130.14.250.10|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 2227871 (2,1M) [application/x-gzip]

Saving to: ‘gaboni\_protein.faa.gz’

gaboni\_protein.faa.gz 100%[===================================================================================>] 2,12M 1,09MB/s in 2,0s

2022-06-13 19:59:11 (1,09 MB/s) - ‘gaboni\_protein.faa.gz’ saved [2227871/2227871]

(base) leonid@leonid-Aspire-A315-42:~$ wget -O knowlesi\_protein.faa.gz https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/006/355/GCA\_000006355.3\_GCA\_000006355.2/GCA\_000006355.3\_GCA\_000006355.2\_protein.faa.gz

--2022-06-13 19:59:47-- https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/006/355/GCA\_000006355.3\_GCA\_000006355.2/GCA\_000006355.3\_GCA\_000006355.2\_protein.faa.gz

Resolving ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)... 130.14.250.10, 130.14.250.11, 2607:f220:41f:250::230, ...

Connecting to ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)|130.14.250.10|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 2413718 (2,3M) [application/x-gzip]

Saving to: ‘knowlesi\_protein.faa.gz’

knowlesi\_protein.faa.gz 100%[===================================================================================>] 2,30M 1,95MB/s in 1,2s

2022-06-13 19:59:49 (1,95 MB/s) - ‘knowlesi\_protein.faa.gz’ saved [2413718/2413718]

(base) leonid@leonid-Aspire-A315-42:~$ wget -O yoelii\_protein.faa.gz https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/002/385/GCA\_900002385.2\_GCA\_900002385/GCA\_900002385.2\_GCA\_900002385\_protein.faa.gz

--2022-06-13 20:00:28-- https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/002/385/GCA\_900002385.2\_GCA\_900002385/GCA\_900002385.2\_GCA\_900002385\_protein.faa.gz

Resolving ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)... 130.14.250.13, 165.112.9.229, 2607:f220:41f:250::229, ...

Connecting to ftp.ncbi.nlm.nih.gov (ftp.ncbi.nlm.nih.gov)|130.14.250.13|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 2246156 (2,1M) [application/x-gzip]

Saving to: ‘yoelii\_protein.faa.gz’

yoelii\_protein.faa.gz 100%[===================================================================================>] 2,14M 1,80MB/s in 1,2s

2022-06-13 20:00:31 (1,80 MB/s) - ‘yoelii\_protein.faa.gz’ saved [2246156/2246156]

(base) leonid@leonid-Aspire-A315-42:~$ gzip -d falciparum\_protein.faa.gz

(base) leonid@leonid-Aspire-A315-42:~$ gzip -d vivax\_protein.faa.gz

(base) leonid@leonid-Aspire-A315-42:~$ gzip -d gaboni\_protein.faa.gz

(base) leonid@leonid-Aspire-A315-42:~$ gzip -d knowlesi\_protein.faa.gz

(base) leonid@leonid-Aspire-A315-42:~$ gzip -d yoelii\_protein.faa.gz

(base) leonid@leonid-Aspire-A315-42:~$

(base) leonid@leonid-Aspire-A315-42:~$ proteinortho falciparum\_protein.faa vivax\_protein.faa gaboni\_protein.faa knowlesi\_protein.faa yoelii\_protein.faa

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Proteinortho with PoFF version 6.0.14 - An orthology detection tool

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Detected 8 available CPU threads (adjust this with -cpus), Detected 'diamond' version 0.9.30

Checking input files.

Checking falciparum\_protein.faa... ok

Checking vivax\_protein.faa... ok

Checking gaboni\_protein.faa... ok

Checking knowlesi\_protein.faa... ok

Checking yoelii\_protein.faa... ok

\*\*Step 1\*\*

Generating indices.

Building database for 'yoelii\_protein.faa' (6039 sequences)

Building database for 'vivax\_protein.faa' (5389 sequences)

Building database for 'falciparum\_protein.faa' (5387 sequences)

Building database for 'gaboni\_protein.faa' (5356 sequences)

Building database for 'knowlesi\_protein.faa' (5326 sequences)

\*\*Step 2\*\* using diamond

Running blast analysis: 100% (10/10)

[OUTPUT] -> written to myproject.blast-graph

\*\*Step 3\*\*

Clustering by similarity (Proteinortho mode) using up to 4402.5 MB of memory (75% of total memory) and 8 cpu core(s). Adjust this behaviour with the -mem option.

Reading myproject.blast-graph

5 species

23896 paired proteins

43509 bidirectional edges

Done

[OUTPUT] -> Orthologous groups are written to myproject.proteinortho.tsv

You can extract the fasta files of each orthology group with 'proteinortho\_grab\_proteins.pl -tofiles myproject.proteinortho.tsv yoelii\_protein.faa vivax\_protein.faa falciparum\_protein.faa gaboni\_protein.faa knowlesi\_protein.faa'

(Careful: This will generate a file foreach line in the file myproject.proteinortho.tsv).

[OUTPUT] -> Orthologous pairs are written to myproject.proteinortho-graph

[OUTPUT] -> Summary is written to myproject.proteinortho-graph.summary

[OUTPUT] -> Orthologous groups are written to myproject.proteinortho.html

All finished.

(base) leonid@leonid-Aspire-A315-42:~$

(base) leonid@leonid-Aspire-A315-42:~$ intersectBed -a falciparum\_tss.bed -b final\_Z\_falciparum.bed -wb > falciparum\_intersection.bed

(base) leonid@leonid-Aspire-A315-42:~$ intersectBed -a vivax\_tss.bed -b final\_Z\_vivax.bed -wb > vivax\_intersection.bed

(base) leonid@leonid-Aspire-A315-42:~$ intersectBed -a gaboni\_tss.bed -b final\_Z\_gaboni.bed -wb > gaboni\_intersection.bed

(base) leonid@leonid-Aspire-A315-42:~$ intersectBed -a knowlesi\_tss.bed -b final\_Z\_knowlesi.bed -wb > knowlesi\_intersection.bed

(base) leonid@leonid-Aspire-A315-42:~$ intersectBed -a yoelii\_tss.bed -b final\_Z\_yoelii.bed -wb > yoelii\_intersection.bed

(base) leonid@leonid-Aspire-A315-42:~$ touch cluster1.faa

(base) leonid@leonid-Aspire-A315-42:~$ touch cluster2.faa

(base) leonid@leonid-Aspire-A315-42:~$ touch cluster3.faa

(base) leonid@leonid-Aspire-A315-42:~$ touch cluster4.faa

(base) leonid@leonid-Aspire-A315-42:~$ touch cluster5.faa

(base) leonid@leonid-Aspire-A315-42:~$