Velvet:

https://www.ebi.ac.uk/~zerbino/velvet/

Use [sudo apt-get velvet]

???the question is, how to make 'MAXKMERLENGTH=99', after sudo installed velvet, ==!

I used vi editor to edit the Makefile in velvet\_1.2.10 folder

[vi Makefile ]

then follow the instruction in this page: http://www.cs.rit.edu/~cslab/vi.html

I updated both MAXKMERLENGTH=99 AND CATEGORIES=99;

VelvetOptimiser:

http://www.vicbioinformatics.com/VelvetOptimiser-2.2.5.tar.gz

I unziped the VelvetOptimiser into /home/gd44/CBB520/VelvetOptimiser-2.2.5 directory;

1st, download fastq or fasta raw data files into Ubuntu directory: /home/gd44/CBB520/data

------------Here I just use [wget http://url.duke.edu] commander.

2nd, run velveth, velvetg, and VelvetOptimiser.pl

[velveth ] , [velvetg], [/home/gd44/CBB520/VelvetOptimiser-2.2.5/VelvetOptimiser.pl ]

to check if each module works well.

3rd, run velveth

[ velveth auto\_data\_17/ 17 -shortPaired -fastq yjm993million.fastq yjm993million2.fastq ]

4th, run velvetg

[velvetg auto\_data\_17/ -cov\_cutoff auto -exp\_cov auto ]

===========================================================================================

repeat step 3 and step 4 for all K-mers between 17 and 31;

but, change the output directories correspondingly;

auto\_data\_17, auto\_data\_19, auto\_data\_21 ------------auto\_data\_29, auto\_data\_31

===========================================================================================

5th, run VelvetOptimiser.pl

[/home/gd44/CBB520/VelvetOptimiser-2.2.5/VelvetOptimiser.pl -s 17 -e 31 -f '-shortPaired -fastq yjm993million.fastq

yjm993million2.fastq' ]

AFTER step 5, we could see the VelvetOptimiser output, which will give the optimisized parameters for velveth and velvetg.

here we suppose the best Kmer is 21, best -exp\_cov = 4, -cov\_cutoff = 3.94

6th, run velvetg again, with the parameters just got:

[ velvetg auto\_data\_21/ -cov\_cutoff 3.94, -exp\_cov 4 ]

or

[velvetg auto\_data\_21/ -amos\_file yes -exp\_cov 3.94 -cov\_cutoff 4

7th, check the contig.fa and the stat.txt documents.

8th, transfer the outputed AMOS file into genomic-browser readable format;   
OR, use other script to align the contig.fa with the reference sequence he gave us... (the yjm993\_ve6h.\*\*.fasta)

For K-mer optimizer:

Final optimised assembly details:

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

Assembly id: 3

Assembly score: 1625289

Velveth timestamp: Sep 29 2014 18:13:32

Velvetg timestamp: Sep 29 2014 18:36:39

Velveth version: 1.2.10

Velvetg version: 1.2.10

Readfile(s): -shortPaired -fastq yjm993million.fastq yjm993million2.fastq

Velveth parameter string: auto\_data\_21 21 -shortPaired -fastq yjm993million.fastq yjm993million2.fastq

Velvetg parameter string: auto\_data\_21 -clean yes -exp\_cov 3 -cov\_cutoff 2.0497824

Assembly directory: /home/guang/auto\_data\_21

Velvet hash value: 21

Roadmap file size: 23554239

Total number of contigs: 13348

n50: 678

length of longest contig: 5740

Total bases in contigs: 5694345

Number of contigs > 1k: 1152

Total bases in contigs > 1k: 1625289

Paired Library insert stats:

Paired-end library 1 has length: 5, sample standard deviation: 13

Paired-end library 1 has length: 5, sample standard deviation: 13

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

Assembly output files are in the following directory:

/home/guang/auto\_data\_21

root@guang-HP-Pavilion-dv2-Notebook-PC:/home/guang# Downloads/VelvetOptimiser-2.2.5/VelvetOptimiser.pl -s 17 -e 33 -f '-shortPaired -fastq yjm993million.fastq yjm993million2.fastq'

Final optimised assembly details:

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

Assembly id: 21

Assembly score: 94022

Velveth timestamp: Sep 29 2014 18:53:07

Velvetg timestamp: Sep 29 2014 19:12:39

Velveth version: 1.2.10

Velvetg version: 1.2.10

Readfile(s): -shortPaired -fastq yjm993million.fastq yjm993million2.fastq

Velveth parameter string: auto\_data\_75 75 -shortPaired -fastq yjm993million.fastq yjm993million2.fastq

Velvetg parameter string: auto\_data\_75 -clean yes -exp\_cov 13 -cov\_cutoff 1.29608265764191

Assembly directory: /home/guang/auto\_data\_75

Velvet hash value: 75

Roadmap file size: 14313579

Total number of contigs: 348

n50: 935

length of longest contig: 13324

Total bases in contigs: 190025

Number of contigs > 1k: 21

Total bases in contigs > 1k: 94022

Paired Library insert stats:

Paired-end library 1 has length: 207, sample standard deviation: 431

Paired-end library 1 has length: 207, sample standard deviation: 431

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

Assembly output files are in the following directory:

/home/guang/auto\_data\_75

root@guang-HP-Pavilion-dv2-Notebook-PC:/home/guang# Downloads/VelvetOptimiser-2.2.5/VelvetOptimiser.pl -s 35 -e 75 -f '-shortPaired -fastq yjm993million.fastq yjm993million2.fastq

Final optimised assembly details:

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

Assembly id: 4

Assembly score: 89156

Velveth timestamp: Sep 29 2014 19:14:50

Velvetg timestamp: Sep 29 2014 19:19:26

Velveth version: 1.2.10

Velvetg version: 1.2.10

Readfile(s): -shortPaired -fastq yjm993million.fastq yjm993million2.fastq

Velveth parameter string: auto\_data\_81 81 -shortPaired -fastq yjm993million.fastq yjm993million2.fastq

Velvetg parameter string: auto\_data\_81 -clean yes -exp\_cov 15 -cov\_cutoff 2.00656292772518

Assembly directory: /home/guang/auto\_data\_81

Velvet hash value: 81

Roadmap file size: 13286464

Total number of contigs: 89

n50: 2882

length of longest contig: 13315

Total bases in contigs: 112858

Number of contigs > 1k: 26

Total bases in contigs > 1k: 89156

Paired Library insert stats:

Paired-end library 1 has length: 42, sample standard deviation: 65

Paired-end library 1 has length: 42, sample standard deviation: 65

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

Assembly output files are in the following directory:

/home/guang/auto\_data\_81

root@guang-HP-Pavilion-dv2-Notebook-PC:/home/guang# Downloads/VelvetOptimiser-2.2.5/VelvetOptimiser.pl -s 75 -e 93 -f '-shortPaired -fastq yjm993million.fastq yjm993million2.fastq'

**So, I choose 81 as the kmer; -exp\_cov 15 -cov\_cutoff 2.00656  
According to the output above, I think kmer=21 or kmer=75 should be ok.**

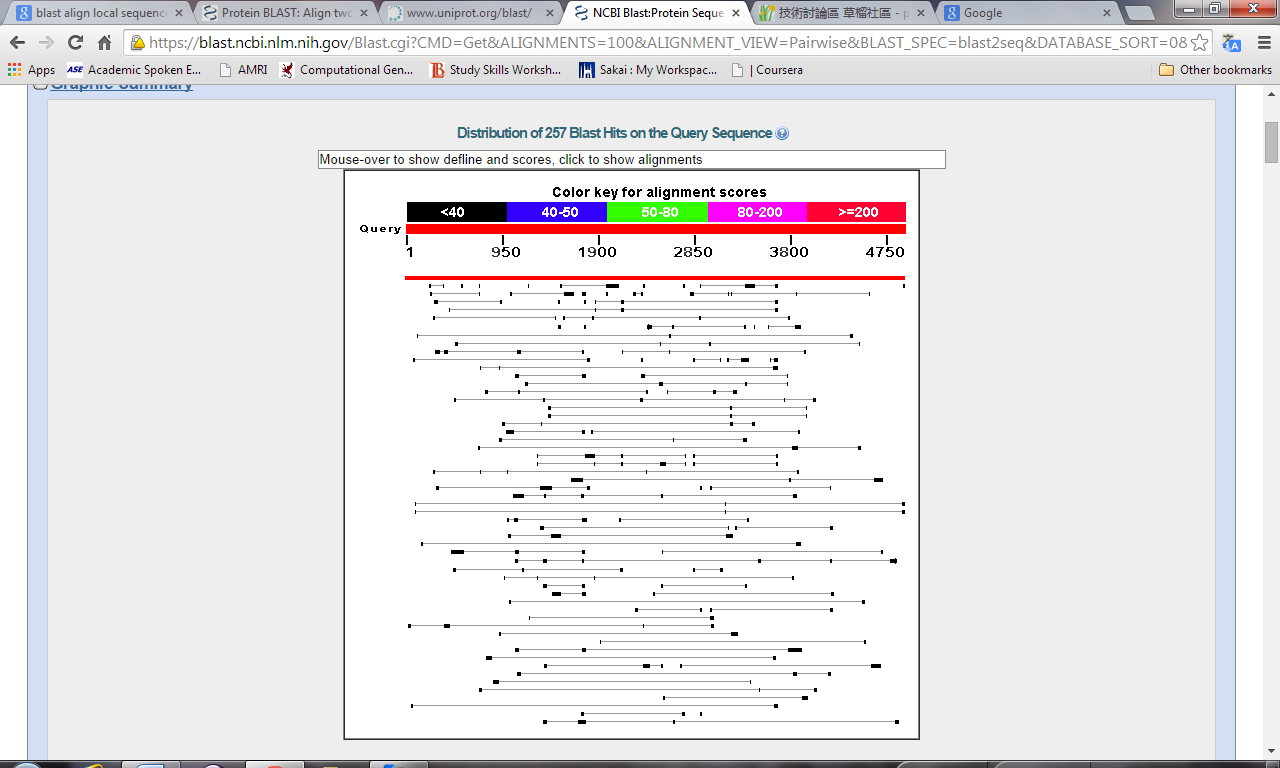
After the last run with Kmer=81, -exp\_cov 15 –cov\_cutoff 2.0065, I put the MDN1 sequence and the contig.fa data to BLASTN,

I got MDN1 protein sequence from url: <http://www.ncbi.nlm.nih.gov/protein/584374819?report=fasta>

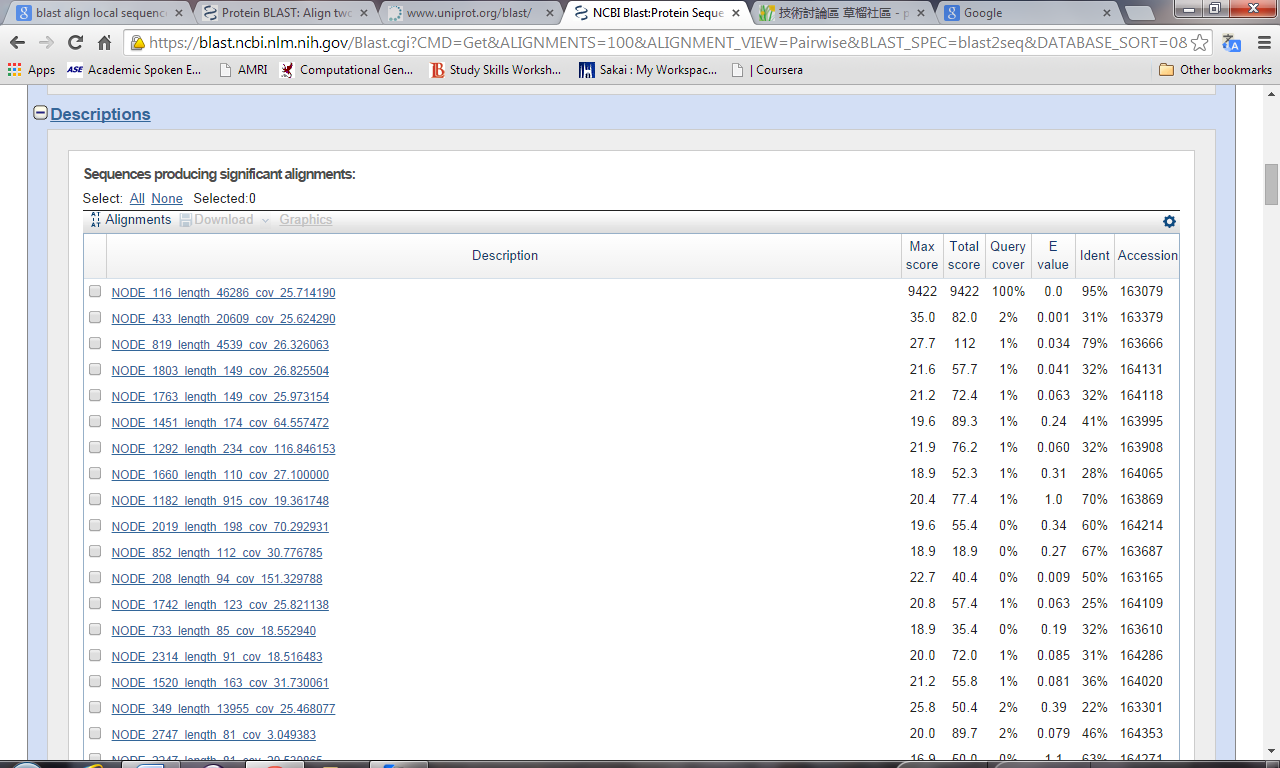
https://blast.ncbi.nlm.nih.gov

my url: https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&ALIGNMENTS=100&ALIGNMENT\_VIEW=Pairwise&BLAST\_SPEC=blast2seq&DATABASE\_SORT=0&DESCRIPTIONS=100&DYNAMIC\_FORMAT=on&FIRST\_QUERY\_NUM=0&FORMAT\_OBJECT=Alignment&FORMAT\_PAGE\_TARGET=&FORMAT\_TYPE=HTML&GET\_SEQUENCE=yes&I\_THRESH=&LINE\_LENGTH=60&MASK\_CHAR=2&MASK\_COLOR=1&NEW\_VIEW=yes&NUM\_OVERVIEW=100&OLD\_BLAST=false&PAGE=Translations&QUERY\_INDEX=0&QUERY\_NUMBER=0&RESULTS\_PAGE\_TARGET=&RID=2M5PKS7A11N&SHOW\_LINKOUT=yes&SHOW\_OVERVIEW=yes&STEP\_NUMBER=&WORD\_SIZE=3&DISPLAY\_SORT=3&HSP\_SORT=3

Then got just one node with 95% identical and 100% coverage with MDN1 gene, so, I think the assembly result is ok.

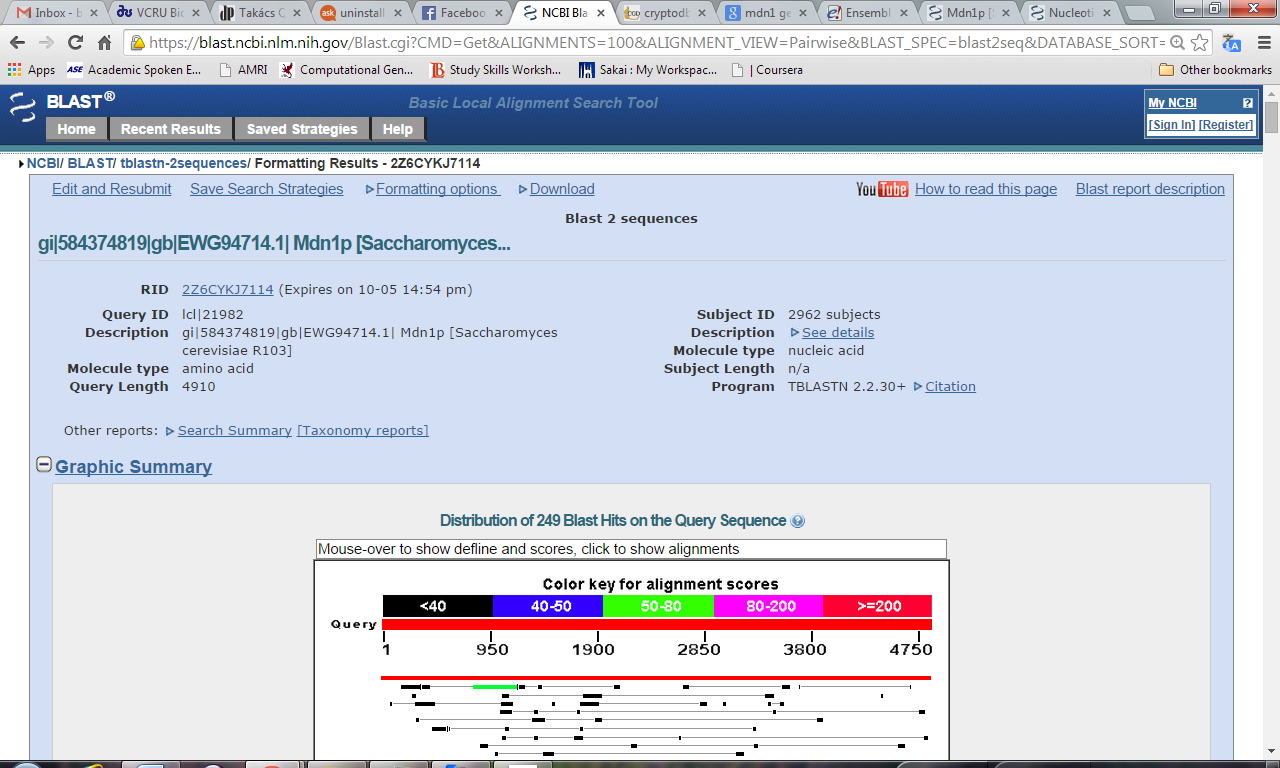


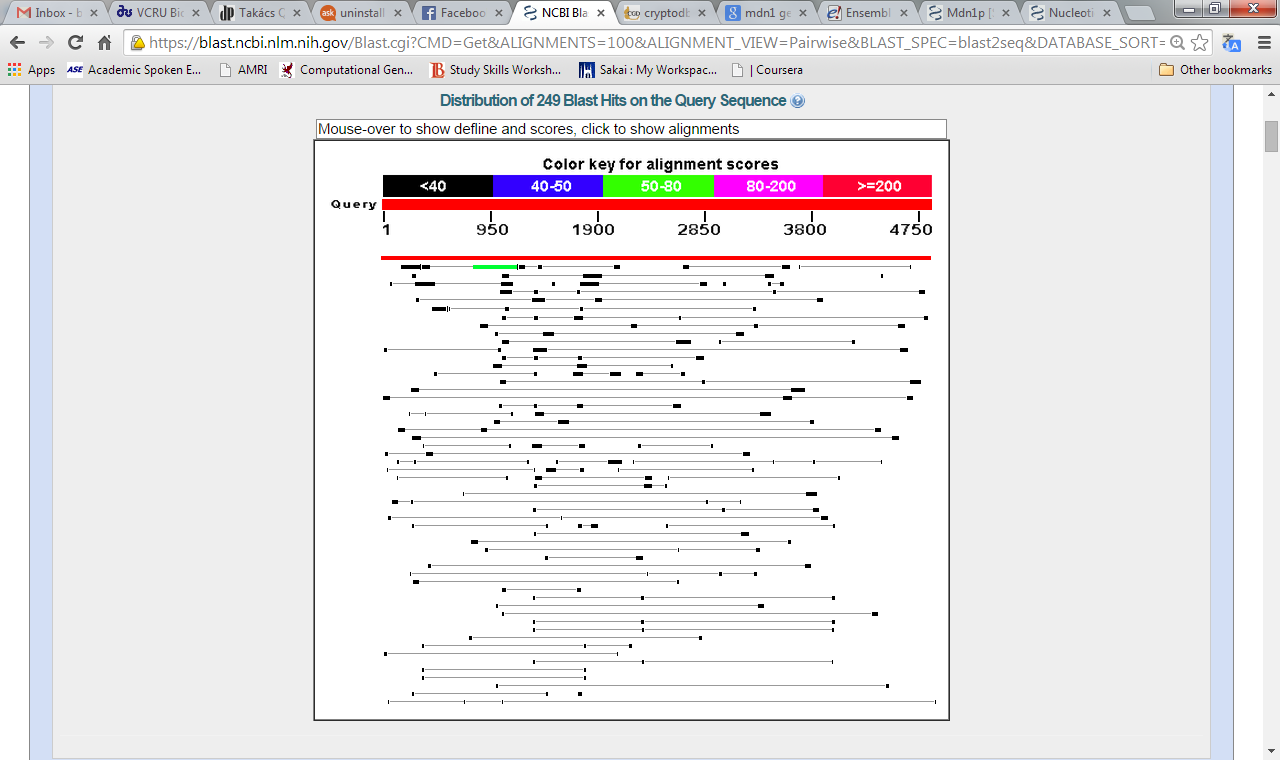
The red line represents the node (contig) which matches the MDN1 gene best.

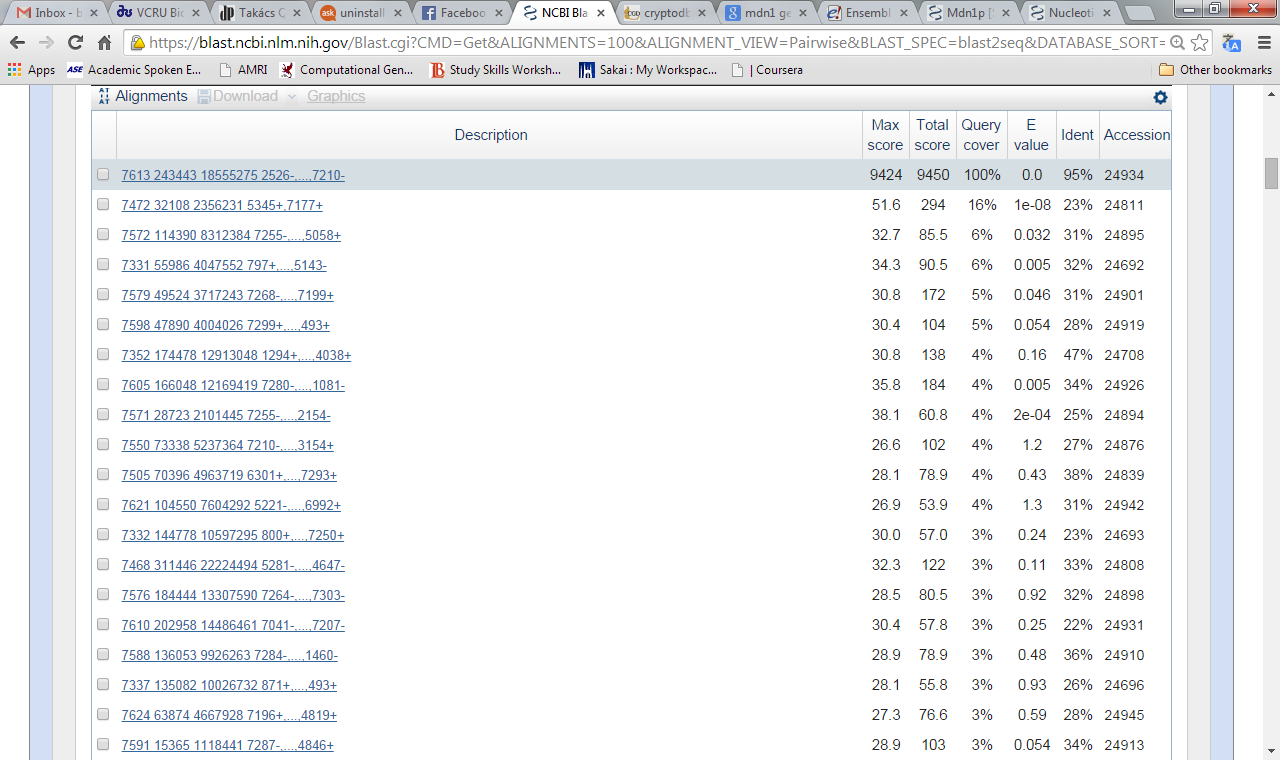


I will try Abyss the day after tomorrow.

For ABySS output: got 100% coverage and 95% identical too.







For ABySs, it’s kind of straight forwards, just run kmers from 17 to 64, then pick out the best one;

Sudo apt-get install abyss