

Lets start by downloading a genome to mine! You can find genome assemblies on [www.ncbi.nlm.nih.gov/genome/browse#!/overview/](https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/)

do [Running] Tue 11:13

Genome List - Genome - NCBI - Mozilla Firefox

Genome List - Genome - X +

https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Log in

Genome > **Genome Information by Organism**

Organism name (common or scientific) or Accession (Assembly, BioProject) Search

Download Reports from FTP site

Overview (37743): Eukaryotes (6135); Prokaryotes (146802); Viruses (17632); Plasmids (12989); Organelles (11731)

Filters Download

Choose Columns Page 1 of 755 50 View 1 - 50 of 37,743

#	Organism Name	Organism Groups	Size(Mb)	Chromosome	Organelles	Plasmids	Assemblies
1	'Brassica napus' phytoplasma	Bacteria;Terrabacteria group;Tenericutes	0.743598	-	-	-	1
2	'Candidatus Kapabacterium' thiocyanatum	Bacteria;FCB group;Bacteroidetes/Chlorobi group	3.27299	-	-	-	1
3	'Chrysanthemum coronarium' phytoplasma	Bacteria;Terrabacteria group;Tenericutes	0.739592	-	-	-	1
4	'Echinacea purpurea' witches'-broom phytoplasma	Bacteria;Terrabacteria group;Tenericutes	0.545427	-	-	-	1
5	'Oseidax' symbiont bacterium Rs2_46_30_T18	Bacteria;unclassified Bacteria;unclassified Bacteria (miscellaneous)	4.02183	-	-	-	1
6	ANME-1 cluster archaeon	Archaea;Euryarchaeota;Methanomicrobia	1.80585	-	-	-	2
7	ANME-1 cluster archaeon ex4572_4	Archaea;Euryarchaeota;Methanomicrobia	1.01808	-	-	-	1
8	ANME-2 cluster archaeon	Archaea;Euryarchaeota;Methanomicrobia	3.60934	-	-	-	1
9	ANME-2 cluster archaeon HR1	Archaea;Euryarchaeota;Methanomicrobia	2.19546	-	-	-	1
10	ANMV-1 virus	Viruses;unclassified archaeal viruses;unclassified	0.038465	1	-	-	1
11	Abaca bunchy top virus	Viruses;ssDNA viruses;Nanoviridae	0.006422	6	-	-	1
12	Abalone herpesvirus Victoria/AUS/2009	Viruses;dsDNA viruses, no RNA stage;Malacoherpesviridae	0.211518	1	-	-	1
13	Abalone shriveling syndrome-associated virus	Viruses;dsDNA viruses, no RNA stage;unclassified	0.034952	1	-	-	1
14	Abditibacterium utsteinense	Bacteria;Terrabacteria group;Alditibacteriota	3.60633	-	-	-	1
15	Abelson murine leukemia virus	Viruses;Other;Retroviridae	0.005894	1	-	-	1
16	Abeotoma whistleri	Eukaryota;Protists;Other Protists	101.554	-	-	-	1
17	Abiotrophia	Bacteria;Terrabacteria group;Firmicutes	2.02078	-	-	-	1
18	Abiotrophia defectiva	Bacteria;Terrabacteria group;Firmicutes	2.04344	-	-	-	1
19	Abisko virus	Viruses;Other;Other	0.010187	1	-	-	1
20	Absidia glauca	Eukaryota;Fungi;Other Fungi	48.7464	-	-	-	1
21	Absidia repens	Eukaryota;Fungi;Other Fungi	47.4229	-	-	-	1
22	Absidia dolichum	Bacteria;Terrabacteria group;Firmicutes	2.19105	-	-	-	2
23	Abutilon Brazil virus	Viruses;ssDNA viruses;Geminiviridae	0.005271	2	-	-	1
24	Abutilon golden mosaic virus	Viruses;ssDNA viruses;Geminiviridae	0.002629	1	-	-	1
25	Abutilon mosaic Bolivia virus	Viruses;ssDNA viruses;Geminiviridae	0.005399	2	-	-	1
26	Abutilon mosaic Brazil virus	Viruses;ssDNA viruses;Geminiviridae	0.005282	2	-	-	1
27	Abutilon mosaic virus	Viruses;ssDNA viruses;Geminiviridae	0.005217	2	-	-	1

Lets download the Sloth genome:

Activities Firefox Web Browser Tue 11:14

Genome List - Genome - NCBI - Mozilla Firefox

https://www.ncbi.nlm.nih.gov/genome/browse#/overview/choloepus

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Log in

Genome > **Genome Information by Organism**

choloepus Search Download Reports from FTP site

Overview (1); Eukaryotes (1); Prokaryotes (0); Viruses (0); Plasmids (0); Organelles (2) Filters Download

#	Organism Name	Organism Groups	Size(Mb)	Chromosome	Organelles	Plasmids	Assemblies
1	Choloepus hoffmanni	Eukaryota/Animals/Mammals	3,286.01	-	-	-	1

Choose Columns Page 1 of 1 50 View 1 - 1 of 1

Popular Connect Resources Actions

- PubMed Facebook Literature Submit
- PMC Twitter Health Download
- PubMed Health LinkedIn Genomes Learn
- Bookshelf YouTube Genes Develop

Activities Firefox Web Browser Tue 11:15

Choloepus hoffmanni (ID 907) - Genome - NCBI - Mozilla Firefox

https://www.ncbi.nlm.nih.gov/genome/907

NCBI Resources How To Sign in to NCBI

Genome Genome Limits Advanced Help

Choloepus hoffmanni (Hoffmann's two-fingered sloth)  
Representative genome: Choloepus hoffmanni (assembly C\_hoffmanni-2.0.1)  
Download sequences in FASTA format for genome  
Download genome annotation in GFF, GenBank format  
BLAST against Choloepus hoffmanni genome

Display Settings: Overview

Organism Overview; Organelle Annotation Report [1]

**Choloepus hoffmanni (Hoffmann's two-fingered sloth)**  
Hoffmann's two-toed sloth is an arboreal mammal of Central and South America. These animals are unusual for conducting nearly all their activities while hanging up by their limbs. In their native habitat, they are capable of vigorous swimming and self-defense. [More...](#)

Lineage: Eukaryota[3067]; Metazoa[1021]; Chordata[478]; Craniata[470]; Vertebrata[163]; Xenarthra[2]; Pilosa[1]; Folivora[1]; Megalonychidae[1]; Choloepus[1]; Choloepus hoffmanni[1]

**Summary**

Submitter: Broad Institute  
Assembly level: Scaffold  
Assembly: GCA\_000164785.2\_C\_hoffmanni-2.0.1 scaffolds: 269,084 contigs: 370,041 N50: 64,490 L50: 13,390  
BioProjects: PRJNA30809  
Whole Genome Shotgun (WGS): INSDC: ABVD000000000.2 total length (Mb): 3286.01  
Statistics: GC%: 40

Opening GCA\_000164785.2\_C\_hoffmanni-2.0.1\_genomic.fna.gz

You have chosen to open:

GCA\_000164785.2\_C\_hoffmanni-2.0.1\_genomic.fna.gz  
which is: Gzip archive (990 MB)  
from: ftp://ftp.ncbi.nlm.nih.gov

What should Firefox do with this file?

☐ Open with Archive Manager (default)

☒ Save File

☐ Do this automatically for files like this from now on.

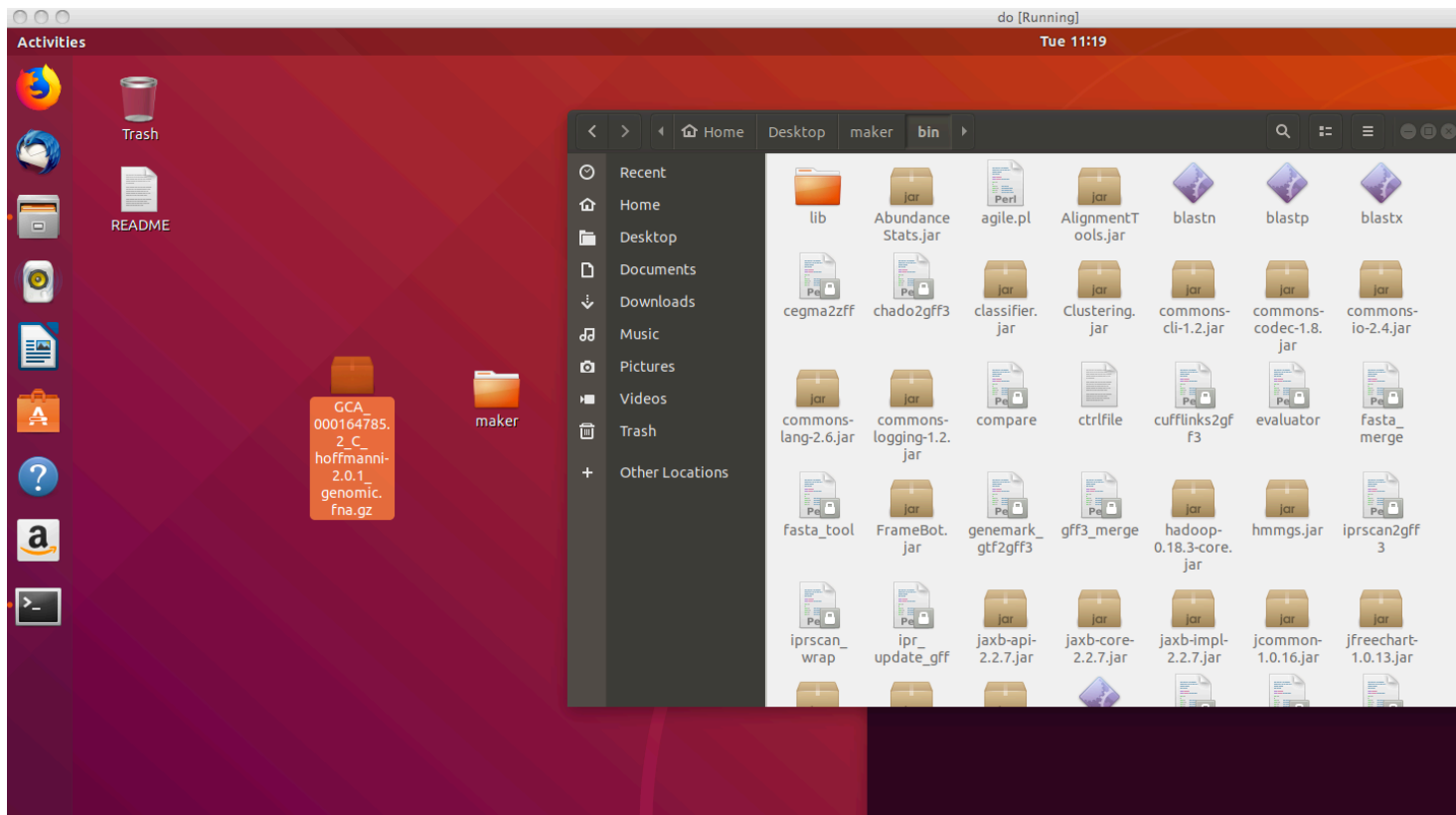
Cancel OK

Choloepus hoffmanni

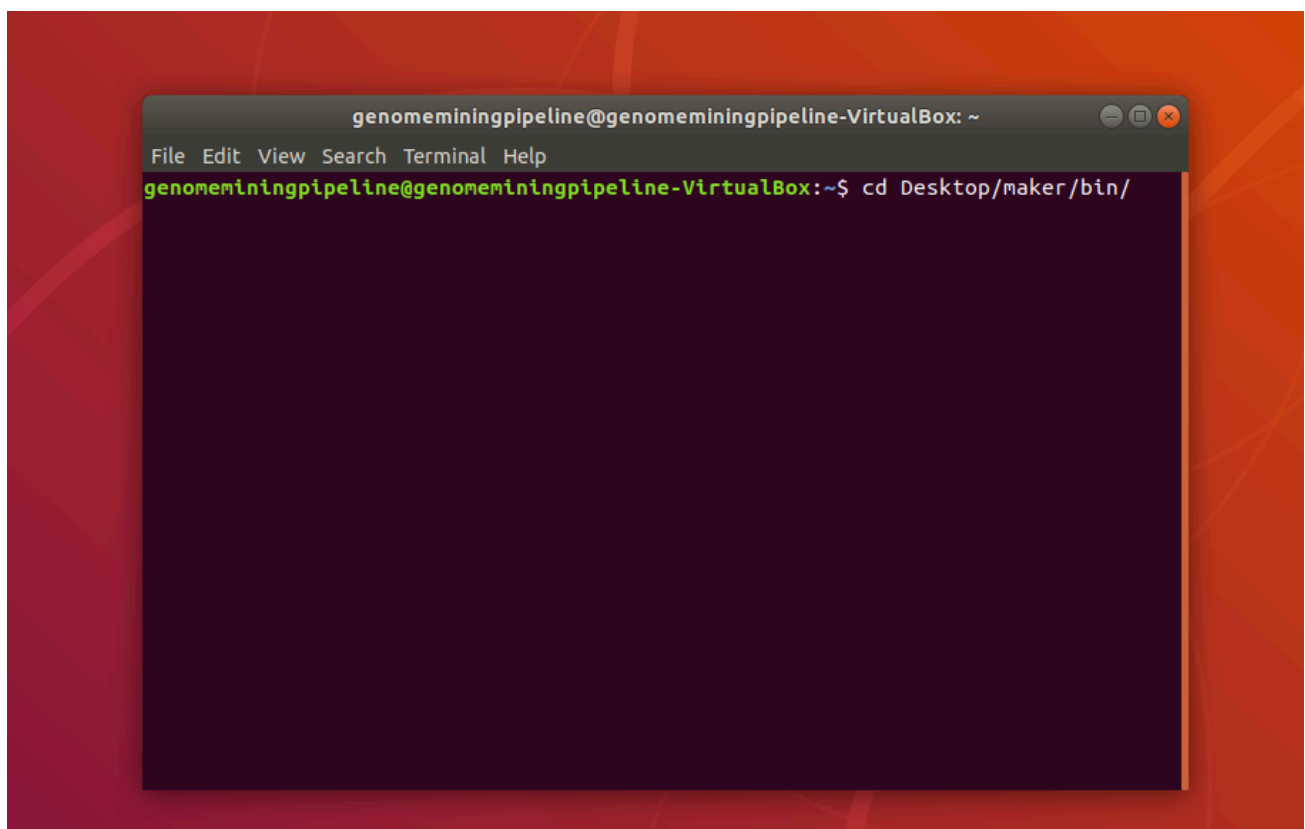
Genome

See more...

Now lets place the downloaded genome file into the /maker/bin/ directory:



Using terminal, lets move into the /maker/bin/ directory with “cd Desktop/maker/bin/”

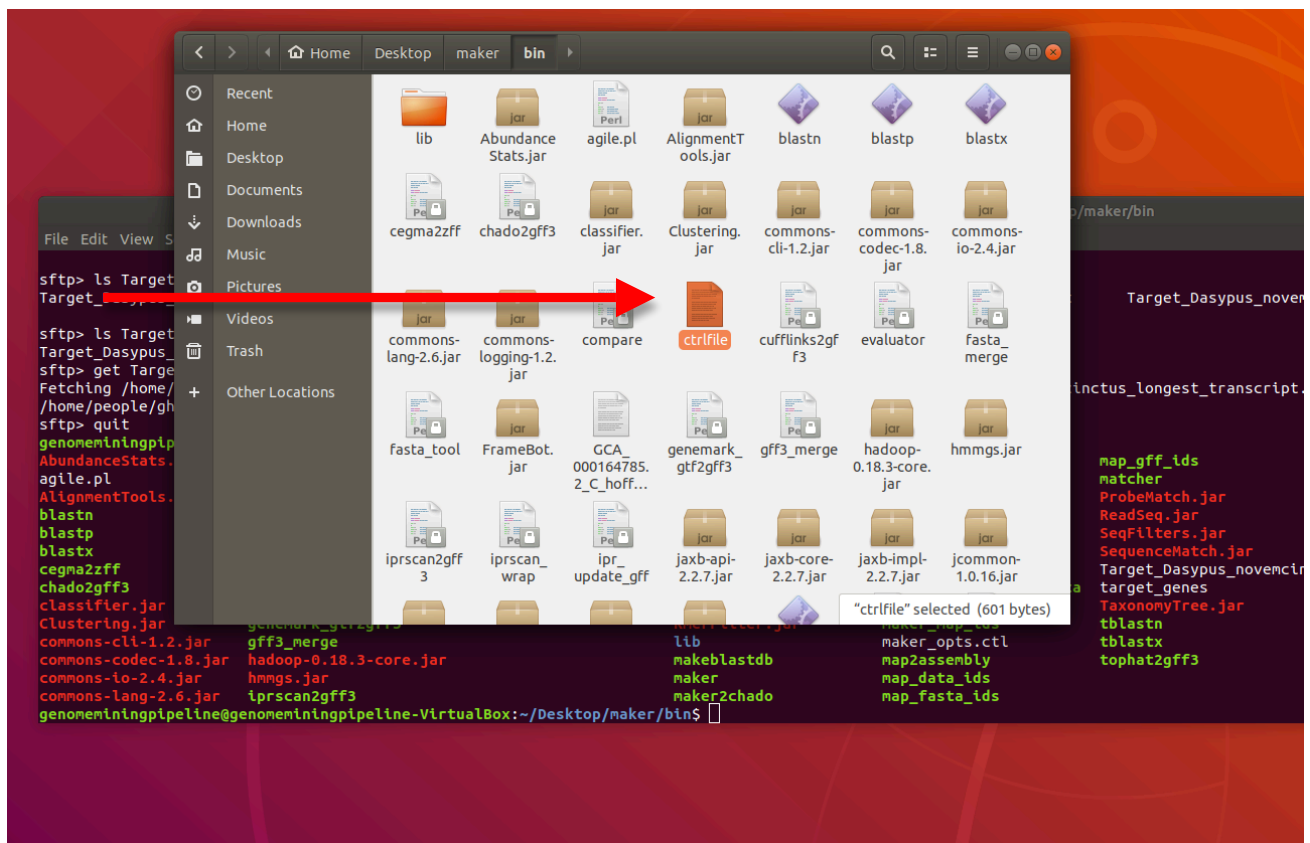


We unzip the genome to mine with “gunzip”:

```
genomeminingpipeline@genomeminingpipeline-VirtualBox: ~/Desktop/maker/bin
File Edit View Search Terminal Help
genemark_gtf2gff3
gff3_merge
hadoop-0.18.3-core.jar
hmms.jar
iprscan2gff3
iprscan_wrap
ipr_update_gff
jaxb-api-2.2.7.jar
jaxb-core-2.2.7.jar
jaxb-impl-2.2.7.jar
jcommon-1.0.16.jar
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$ tar -zxvf lib/^
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$ ls
AbundanceStats.jar  commons-io-2.4.jar  gff3_merge  hadoop-0.18.3-core.jar  junit-4.8.2.jar  maker_functional  SeqFilters.jar
agile.pl  commons-lang-2.6.jar  hmms.jar  iprscan2gff3  KmerFilter.jar  maker_functional_fasta  SequenceMatch.jar
AlignmentTools.jar  commons-logging-1.2.jar  iprscan_wrap  ipr_update_gff  maker  maker_functional_gff  target_genes
blastn  compare  jaxb-api-2.2.7.jar  jaxb-core-2.2.7.jar  maker2chado  maker_map_ids  TaxonomyTree.jar
blastp  ctrlfile  jaxb-impl-2.2.7.jar  jcommon-1.0.16.jar  maker2eval_gtf  maker_opts.ctl  tblastn
blastx  cufflinks2gff3  jfreechart-1.0.13.jar  jsr173_api-1.0.jar  maker2jbrowse  map_data_ids  tblastx
cegma2zff  evaluator  jcommon-1.0.16.jar  jfreechart-1.0.13.jar  maker2wap  map_gff_ids  tophat2gff3
chado2gff3  fasta_merge  jfreechart-1.0.13.jar  jsr173_api-1.0.jar  maker2zff  matcher  ReadSeq.jar
classifier.jar  FrameBot.jar  jfreechart-1.0.13.jar  jsr173_api-1.0.jar  maker_bopts.ctl  ProbeMatch.jar
Clustering.jar  GCA_000164785.2_C_hoffmanni-2.0.1_genomic.fna.gz  jfreechart-1.0.13.jar  jsr173_api-1.0.jar  maker_exe.ctl  ReadSeq.jar
commons-cli-1.2.jar  genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$ gunzip GCA_000164785.2_C_hoffmanni-2.0.1_genomic.fna.gz
commons-codec-1.8.jar  genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$
```

So, we have downloaded our target genome to mine, placed it in the correct directory and unzipped it. Now we need to tell AGILE where to look for all the necessary files.

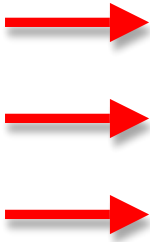
Since all the binaries, including all associated FrameBot files (all .jar files in red), are in the maker/bin/ directory, as EMBOSS matcher, lets point AGILE to the maker/bin/ directory for these. Open the “ctrlfile”:



We can get the exact path to our directory using the “pwd” command:

```
commons-io-2.4.jar      hmmgs.jar              maker
commons-lang-2.6.jar   iprscan2gff3          maker2chado
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$ pwd
/home/genomeminingpipeline/Desktop/maker/bin
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$
```

Once opened, we can paste the path as the location of each program binary. We also add the name of the genome file. Note that “Target\_genes”, “Genome File” and “SisterTaxa” also have names:



```
Overhang:"0";
ShearLength:"4000";
Maker_Path:"/home/genomeminingpipeline/Desktop/maker/bin";
Blastn_Path:"/home/genomeminingpipeline/Desktop/maker/bin";
Makeblastdb_Path:"/home/genomeminingpipeline/Desktop/maker/bin";
Target_genes:"target_genes";
Matcher_Path:"/home/genomeminingpipeline/Desktop/maker/bin";
Framebot_Path:"/home/genomeminingpipeline/Desktop/maker/bin";
GenomeFile:"GCA_000164785.2_C_hoffmanni-2.0.1_genomic.fna";
SisterTaxa:"Target_Dasypus_novemcinctus_longest_transcript.fa";
TestName:"SlothTest";
OverlapDiff:"20";
Evaluate:"1e-5";
Score:"40";
Ref_Coverage:"70";
Augustus_model:"human";
```

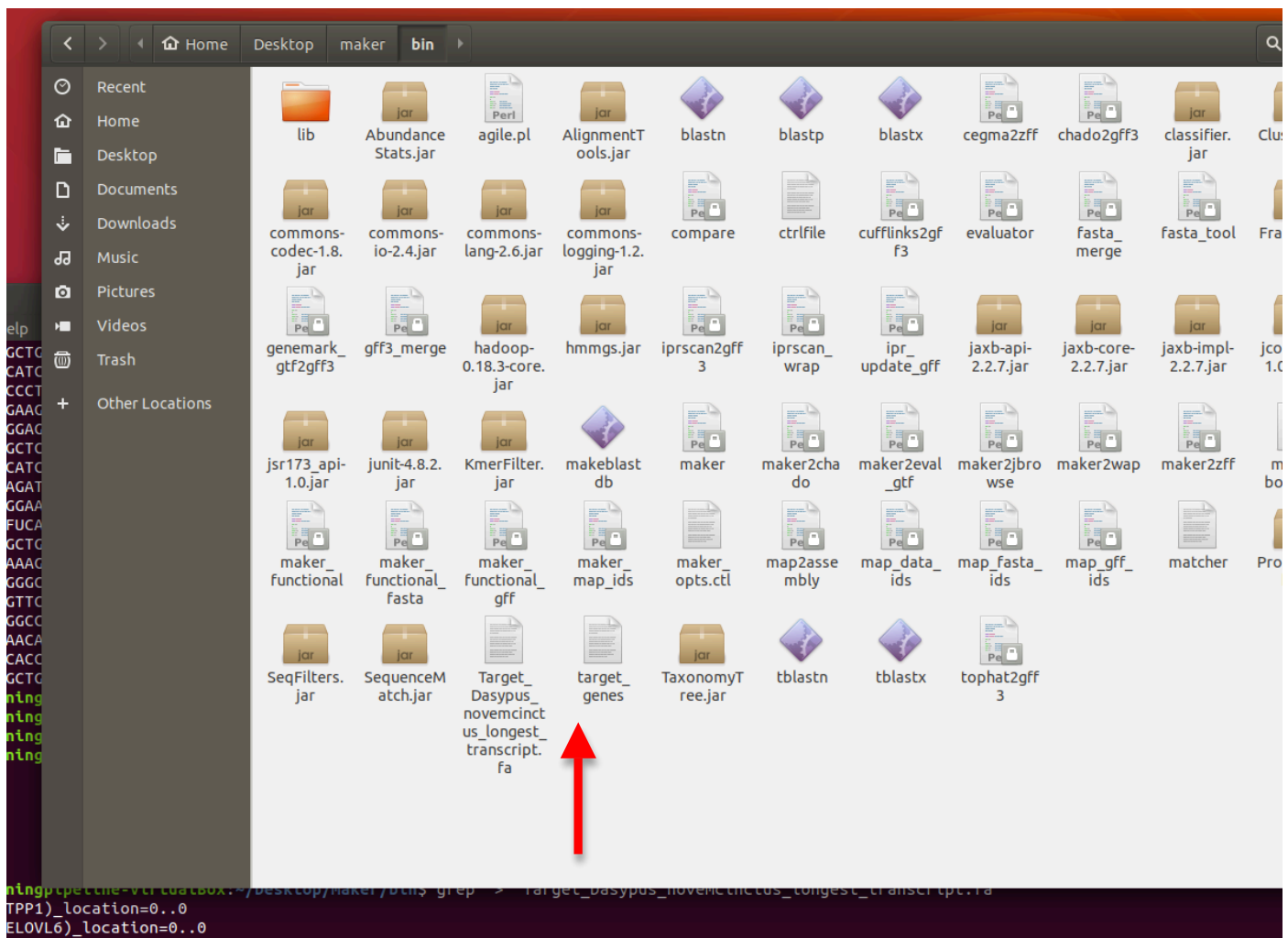


Lets create and open a file named “Target\_Dasypus\_novemcinctus\_longest\_transcripts.fna”. We can generate this file using the ‘RefSeq\_get.pl’ script found here: [https://github.com/batlabucd/GenomeMining/tree/master/RefSeq\\_Mining](https://github.com/batlabucd/GenomeMining/tree/master/RefSeq_Mining)

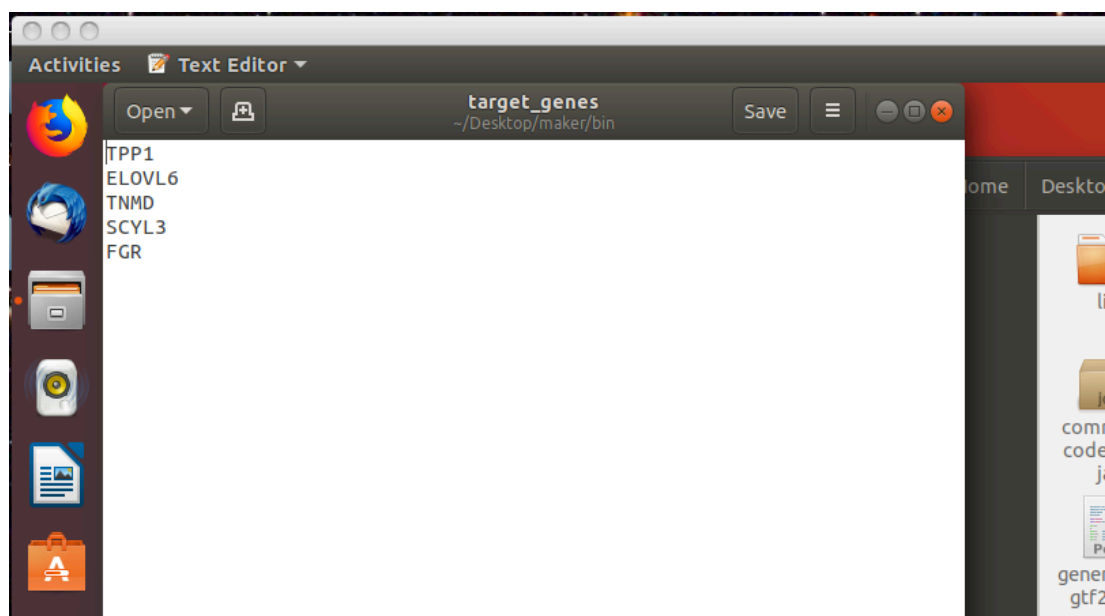
These are the genes we will use as a reference. Note that all fasta headers contain the gene name in the format “(GENE)”, eg (TPP1). This is essential as AGILE will look for this format.

```
Open Target_Dasypus_novemcinctus_longest_transcript.fna Save
~/Desktop/maker/bin
>Dasypus_novemcinctus_Genes_(TPP1)_location=0..0
ATGTGTCACGTGACAGCAGAGCTGCAGAAGACGGAATGGGCTCCGAGCCTGCCCTCTCGGGCTCCTTGCCTCATCTACTGGAAGTCAGTTACAGCCCGGAGCCTGACCAGCAATGGACAGTGTCCCCAGGCTGGGTGTCCCTGGGCCGCGCGGACCTGAGGAAGATCTGAGT
CTCACCCTTTGCCCTGAGACAGCAGAACCTGGAGAGACTCTCTGAGCTGGTGCAAGCTGTGTGCGACCTTGCCTCTCCTGCTACGGAAATACCTGAGCCTAGAGGATGTAGCTGAGCTGGTCCAGGCATCACCCTACCCGCCACACAGTCCAAAAATGGCTCATGGCAGCTGGAGCC
TGGAACTGTCACTTTCAGTGACCACACAGGACTTTCTGACTTGTCTGCTGAGCATCCGACAGGCGAGAGCTGCTGCTTTCTGGGGCTGAGTTTCATCGCTATGTGGGGGGACCTACAGACACCTAGCTTTTGGGTCTCCACATCCCTACAGCTCCACAGGCCATTGGCACCCCATGTGGAC
TTTGTGGGGGGGCTGCACGCTTTCCCTCTATATCATCTTGAAGCAGCGCCCTGAGCCACAGGTGGCAGGAACCTGTGGGCTGCATCTGGGGTGACCCCATCCGTGGTCCGTGACGATACAACTTGACGGGCCAAGATGTAGGCTCTGGTACCACCAACACAGCCAGGCCCTGTGCC
CAGTCTCTGGAGCAGTACTTCATGAAGCAGACCTGACTGAGTTTCATGCCCTCTTTGGTGAAGAAGCTTGCACACCAGACATCAGTAGCCGTGGTGGTGGAAACAGGGCCGAGGCAGGGCTGGTATTGAGGCCAGCTAGATGTGGAGTACCTGATGAGTCTGGTGCCACATCTCC
ACCTGGGTCTATAGTAGCCCTGGCCGGCATGAGGCACAAGAGCCCTTCTACAGTGGCTCCTGCTGCTCAGTAATGAGTCAGCCCTGCCACAGGTGCACACCGTGAGCTATGGGACGACGAGGACTCCCTCAGCAGCGCTACATCCAGCGGTCAACAGCGAGTTTCATGAAGGCTGCT
TCTCGGGGTCTCACCCTGCTCTTTGGCTCAGGTGACAGTGGTGTGGGTGTGGTCTGTGCTGGAAGACACGGTTCGGTCCGAGCTTCCCTGCTCCAGCCCTATGTACACAGTAGGAGGCACATCTTCCAGAAACCTCTTTGTGCACAAATGAGATTGTTGACTATATCAGT
GGTGGTGGCTTCAGCAATGTCTTCCACAGCCTTCATATCAGGAGGAAGCAGTAGGCCAGTTCTGAGCTCCAGCCCACTGCCCACATCCAGCTACTTCAATGCCAGTGGCCGTGCTTACCACAGATGGCTGCACCTCTCTGATGGGTACTGGGTGGTCAGCAATAGAGTGCCCAAT
CCATGGATATCTGGCACCTCGGCTCTACTCCAGTGTGGAGGGTCTGTCCCTGGTAAATGAGCAGAGAATACTCAATGGCCGCCCTCTCGGCTTCTCAACCCAAAGGCTTACCAACAGCGTGGGCGAGGACTCTTTGATGTGACCCAGGCTGTATGCATCTCTGTCTGAAT
GAAGAAGTGACGGGAGCATTGTGTCCCTCCCTCCCTCCGCACTCTCTGGGCAATTTGGCGTGGAGGGCGCAGGGTGCTTCCCACTCCACTACGATCAATTTGTTGTTGTTGCCACTACGATCTTTAAAGCTGACTGA
>Dasypus_novemcinctus_Genes_(ELOVL6)_location=0..0
ATGAATTCGAAAAGCAGTTCAACGAGAGTGAAGCCATTTCGATGATGCAAGAAAATGATGTAGTGCTCTGCCAGCACCTCGGCTCTTCCAGCTGTATCATTCCTACTCTTGACTCCGACAGAGCAAAAGCGGAATGCTTGCACGTGGCAGCTGCTTGTGAAAATGTTTGTCCAGAAGAAG
TCGTTCTCTCTTCTGCTCTGTATGCTGCCCTTATATTTGGTGTAGGCACCTAAATGAACAAACGGGCCAAGTTTGAACTAGGAAGCCGTAGTGCTCTGGTCTCTCACCCTTGACGCTTTCAGTATATTCGGTGCTCTTGAACCTGGTGCTTATATGGGTACATTTTGTAGCCAAA
GGCCTGAAGGAGTCAGTATGTGACCCAGAGTTTTTACATTGGACCTGTCAACCAATTCGGCTTATGCTTTGTACTAAGCAAAAGCAGCTGAACTAGGGGACACGATTCATTATTCTGAGGAAGCAGAAGTTGATCTTCTGCACTGGTACCACACATCACTGTGCTCTGTACTCT
TGGTACTCTCACAAGACATGGTTGCTGGGGGTGGTGGTTCATGACTATGAACATATACCGTGACGCGGTGATGTACTCTACTATGCTTGGCGGACAGCGGTTCCGAGTCTCCCGGAAGTTTGCCATGTTTCATCACCCTTGTCCACAGATCACCACAGATGGTGATGGGCTGTGTATC
AATACCTGCTCTTCTCTGGTTCAGAGCAAGACAGTGCATCCCACTTTCAGAACATCTTCTGGCTCTCACTCATGTACCTCAGCTACTTTTGCTCTTCTGCCATTTCTCTTTGAGGCCCTACATCAGCAAAATGAGGAAAATGAGAAAGGCTGAATAG
>Dasypus_novemcinctus_Genes_(TNMD)_location=0..0
ATGGCAAAAATCCTCTCAGATAACTATGAGGACTGCCACATCTTAAATCGAGGAGCACTTAAATCCAAGAAGATATGTAATCACTTAAGATTTTGGGAATGATTTTGGCATCTCGGCCCTAACTTTAATAGTCTGTTTTTGGGGAGTAAGCACTTCTGGCCAAAGGCACCAAAAAAC
ACCTATGACATGGAGCACACTTTCTATAGCAATGGAGAGAAGAAGATTACATGAAATTTGATCCTGTGACCAAGACTGAAATATTGAGAAGTGGAAATGGCACTGATGAAACACTGGAAGTATATGACTTTAAAAATGGATACACTGGCATCTACTTTGTAGGCTCGCAAAAATGT
TTTCATCAAAATCAAAATTAAGTGATTCTGAAATTTTCTGATCCAGAGGCAGAAATGATGAGAATGAAGAAATACCACAACTTCTTTGAACAATCTGTGATTGGTGCCAGCAGAAAAGCCTATTGAAACCCAGACCTTCTTAAAAATCCAAAATTCGGAGATTTGTGATAAT
GTGACCATGTATGGATCAGTCCCACTAATGACAGTTTCCGAGTTACAAGACTTTGAGGAGGATGGTGAAGATCTTCCCTTTCTACCAATGAAAAAAGGAATGAAACAAATGAGCAGTGGGTGGTGCCTCAATCAAGGTGGAGAAAACCCGTATGCCAGACAGCAACTGAG
GAAGAACCTTCCAATAAATGACTATCTGAAAAATGGAATCGAATTTGACCCCATGCTGGATGAGAGAGGATATTTGTTGATTACTGTGCTGAGGCAACCCGCTACTGTGCCGTGTCTGTGAACCTTTACTAGGTTATTATTCATATCCATATCTACTGCTACCAAGGAGGACGGGTCACTGTG
CGTGTATCATGCCTTGAACCTGGTGGTGGCCCGCATCTGGCAGGGTCTAA
>Dasypus_novemcinctus_Genes_(SCYL3)_location=0..0
ATGGGATCAGAGAACAGTGTCTTAAAGAGCTATACACTGAAAGAACCCACATTTACCTTACCCTCTGGACTTGTGTTTTTCCAGCTGTTCTGCAAGATGGCAAAATTTGCTTCTGTTTTTGTATATAAGAGAGAAAATGAAGACAAGTTAATAAGCAGCCAGCACCTGAAGACACTACGA
CACCTTGTCTTGTCAAGATTTTTGTCTTGTACTGTGGAAGCAGATGGCATTCATCTTGTCACTGAGCAGTGCAGCCTCTGGAAGTGGCTTTGGAAGCTGTCTTCTGCAGAGGTCTGTGCTGGGATCTATGACATATTGCTGGCTCTTACCTTCTCTCATGACAGAGGAACTTAACACAC
AACAACGCTCGCTGTCTGTCTGTGTTGCTGAGTGAAGCAGGCGACTGGAAGCTAGGAGGGATGGAGCGCTGTGTAAGTTCCGGAGGCCACGCCAGAGTTTCTGAGGAGTATTCACTCAGTAAGAGACCAGCATCTGTCCCTCTGGAAGAGATGTCTCCGAATTCACCTACTCTGCCAGAG
TCACATGGACACGCCGAGATGCCTATTCATTTGGAACATTTGGTGAAGTTTGTCTCACAGCTTAAAGTGAACAGGTTTCAGCGGATGTTCTTCCAGCTTTCAGCAGACCTTGCACTCATCTTGTGTAATACCATTCCAAAATGTGGGCCACGCTCTGCACCTTACTGTCCCATGACTTC
TTCCAGAAATGATTTCTAGAAGTTGTGAATTTCTTGAAGTTTAAACATTGAAGACTGAAGAGGAAAAAATCAATTTTCAAGTTCTGTGGACAGAGTCAAGCTGCTTATCGGAGGAATTAATAGCTTCAAGGTTGGTGCTCTTCTGCTTAAATCAGTTGGTGTTCGCAACCTGTAGCT
GTTAAGAGTTTTCTTCTCATCTGCTTGGCCGAAAAAAGATAATGACACAGGGAGAAACCCATTGCTTGTCTCTCCAGCCCTGTTCCAGTCTCGGGTTATCCCTGTGCTTCTGAGTTGTTTGAAGTTTCAGGAGGACATGTCCGGATGGTGCTGTCTCAGATTGAGGCCATATGGAAG
CACTTCACTCAGGAGCAGCTGAAGAAAGTCATCTTGCCACAGGTATTACTGGGCTTGGCTGACACAAGTGATTCAATTGTGGGATTACTCTTATAGCTCTCGAGTGTGGTCTCTGCTTGGACAGAGGTGGTTGTGGGAGGAGAAAGAACCAAGATCTTCAACAGTACTGCTCCAAGT
TTTACCAAAATCTAGCTTCTTCCAGAAGATTCTCCATGCAGCTGTCTGACGCCAGCAGCTCAGCTCTGCCAATCTTGGAGAACCCTCTCTAGCATATTCCTAAATGTTTATTTTCTGGCAACATGCCCATCAACAGCAAGAACACATACAGCGAGATTACTACAATACTCTT
TTACAGACAGGTGATGATGTTTTCTCAGCTATTAAATTTCCCATGAATGGAATCTCAGATGTGAAAAATACTACAGGAGACAGCGAAAGCTCAATTTCTAAGAGTCTGAAGAGTGGCTGACTGGAGTGAGCCTGAGGAACCTGAAAAACAAAGCTGTTAAACATACAGATTTGGCCTAGAGAG
CCATGTGATCTGCCAATTCCTGTTTACTAACTTGAATACAGAGGAGGAGCTTGGGATGACTTTGAGCTCAGCAGCTTGGATAAAAAATAAAGCAGGAAGTGAATCACTGCTACAATACCTGTTTACCTCAGGGAGGACAAAGCCATTCTACTTTGTTTCCACTTGTCTCAGGAGGCT
AAGCCTTTGAAATCAAGCTGTCCAATCTTGTCCAAGCAGGATGACCCAAAGCAGATCAAGTCATCAAAACTGTGCTCACAAGAAAAACCTTTAAAGTTCCATCAGAACTTGGTTTAGGAGAGGAGTTTACCATTCAGTAAAAAAGAAAGCAGCTGCAAGATCTGAAATGGACTGGTTT
GCTGATATGATCCAGAAATTAACCGTCAGCTGCTTTCTTATTTTACCTGAAGTGAAGCAGACATGATGGTTCACCAAGACAGATGCTCATCAGTGAATGCAGTTTCTCAAAATTTGCTGCAGCAGAAATTAAGTGAAGGAGGCTGAAGGCTGGGAGAGAAGGGGAGCTGAAC
TGGGAAGATAACTGGTGA
>Dasypus_novemcinctus_Genes_(FGR)_location=0..0
ATGGGCTGTGTGTTCTGCAAGAGGTGGAGCTGGGGCCCAAGGAGGATGTTGATCTGGAAGGGGACAGCAGGGGCTTGGGGCTGCAGATGCTACGGCCCGGAACCCACCAGGECGGGCTGTGCTCTCTTTCGCCCAACCCCTCACTACGGCAACTTTACCCCTCAGCCACACAGCCCT
GCCCTCTCGATGGGGCACCATCAGGGGCTCTCAGGCACAGGGTGACCTGTTTCATCGCTTGTATGACTATGAGGCCGGACAGAGGATGACCTCACCCTACCAAGGCGAGAAGTTCCACATCTGAACAAATCGGAAGGTGACTGGTGGGAGGCTCGGTCCCTCAGCTCTGGACAC
ACTGGTTTACATTTCCAGCAATTAACGTGGCCCTGTGGACTCCATCCAGGCTGAAGAGTGGTACTTTGGAAGATCAGCCGGAAGGATGCGGAGAGGCACTGTGCTCCCGGGCAACCCCGGGGGGCTTCTTCTTCTGGGAGAGTGAGACACCAAAAGGAGCTACTCCTCTGCTCACTCCGG
GATTGGGACAGGCTAAAGGCGATCAGCTGAAGCATTACAAGATCCGCAAACTGACACGGGCGGCTACTACATCACCACGCGGGGCCAGTTTGACTCTGTGCAGGAGCTGGTGCAGCACTATGCTGAAGTGAACGACGGCTTGTGCACCTGCTCAGCGTGGCTCGGCCACTGAAAGGCGG
CAGACATGGGCTGGCAGGACGCTGGGAGATCAGCCGACGCTCATCGCTGACGCGCGCTGGCCACCGGCTGTTTGGGGAGCTGTGGCTGGGACGTTGGAACGGCTGACCAAGGTGGCGGTGAAGACGCTGAAGCGGGGACCATGTCCCCGAAGACTTCTGAGGAGGGCG
CAGATCATGAAGCTGCTGCGGCACGACAAGCTCGTGACGCTGACGAGTGGTGTGGAGGAGCCATCTACATCGTGACGGAGTTCATGTGCCACGGCAGCTGCTGGATTCTCTCAAGTACCCGAGGGGACAGGATTGAAGCTGCCCAACTGGTGGACATGGCAGCCAGGTAGCCCGG
GGCATGGCTACATGGAAGCTATGAATACATCCACCGCGACCTGCGCGCAGCCAACTTCTGGTGGCCAGCGGCTGGTGTGCAAGATCGCGGACTTTGGGCTGGCCGCTCATCAAGGACATGAGTACAACCCCGGCAAGGGGCAAGTTCCCATCAAGTGGACAGCCCGAGGGCT
GCTCTCTATGGCAGATTCAACATCAAGTCAGATGTGGTCTTTTGGGATCTGCTCACTGAATTCATCAGCAAGGGCCGAATCCCTACCCAGGCATGACCAAGCGGGAAGTGTGGAACAGGTCAGGACGCGGTACCACATGCCGTGCCCTCCGAGCTGCCCTGCATCCCTGTACGAGGCA
ATGAGACAGACTGGCGTCTGGACCCAGAGGAAAGGCCGACCTTCAATACCTCGAGTCTTCTAGAGGACTACTTACCTCCACGGAACCCCACTACCAGCTGGGATGAGACATAG
```

Now lets create/modify the “target\_genes” file:



We now add a list of all the genes found in the headers of the genes in our reference file:



Lets recap:

- we have downloaded, and unzipped our target genome
- we have placed it into the maker/bin/ directory
- we have modified our ctrlfile to let AGILE know where to look for each program (for convenience, all files are in the same directory)
- we have generated a file with the fasta nucleotide sequence of the genes we want but from a closely related sister taxon to our target genome to mine.
- we have listed all the genes we want in our 'target\_gene' file

Now we can run AGILE with “perl agile.pl”

```
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$  
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$  
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$  
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$  
genomeminingpipeline@genomeminingpipeline-VirtualBox:~/Desktop/maker/bin$ perl agile.pl
```