|  |  |
| --- | --- |
| **Term** | **Purpose** |
| Absolute path /home/paa9/aegypti\_genome/ae\_aegypti.gff3  Relative path from /home/paa9 : aegypti\_genome/ae\_aegypti.gff3 | Paths |
| ../NAME | Relative name, move up a directory level and then go to another document within the same level |
| pwd | Tell what directory we are in |
| ls | Directory |
| ls -l | Directory with more info |
| mkdir | Make a directory |
| rmdir | Remove empty directory |
| rm -r NAME | Remove directory that is not empty |
| use tab | Autocomplete |
| cd | Change directory |
| cd ../ | Move up a directory |
| cp NAME ../NEW LOCATION | Copy a file into another location |
| rm | Remove a file |
| mv | Move file |
| \* | Wildcards, autocomplete anything that ends with those characters  Example: \*.txt refers to everything that ends with txt  if we want to move four files that end with .xt, say mv \*.txt ../Homework 2 |
| /what you’re looking for | Search for text |
| head NAME | Gives first ten lines of the file |
| tail NAME | Gives last ten lines of the file |
| less | See less lines |
| more | Shows more lines |
| head -50 NAME | See the first 50 lines |
| tail -50 NAME | See last 50 lines |
| less NAME | start us at the top, and then use return or space bar to scroll further down |
| more | same thing, subtle differences that we don/t need to know |
| > sequence ID header followed up info | shows start of the sequence and the sequence name, reference ID and other info about where on the chromosome the sequence is located, etc. some files have multiple sequences you can see when scrolling |
| q | escape |
| exit | log out |
| arrow on keyboard | go back to previous commands |
| https://console.cloud.google.com/compute/instances?authuser=0&project=gcp-gu-hpc-medusa&instancessize=50 | log onto bananas controller |
| gcloud compute ssh bananas-controller | log onto bananas controller from terminal app |
| grep “X” FILE NAME | gives you every line of a file where the term in parentheses comes up.  example: grep “>” NAME gives you the headers of all of the sequences in the file |
| wc | line, words, character counts in that order |
| wc -l FILE NAME | just gives you the number of lines |
| | | combines multiple commands  example: grep “>” NAME | wc -l  will give you number of sequences in a file because first you get headers and then tells you number of lines |
| cat NAME1 NAME2 | combines content of files to spit out onto screen only |
| > | file redirect. puts in new file |
| cat NAME1 NAME2 > NEW NAME | combines the files! file redirect and names new file a new file |
| nano NEW FILE NAME | makes a new file! editor opens up, start typing text |
| return | yes |
| nano OLD FILE NAME | edit files that have already been created! |
| control x | exit nano editor |
| ^X^C | kill process |
| wget url\_to\_file | download a file into your current working directory based on a URL that points to the file |
| https://www.ncbi.nlm.nih.gov/ | find genome, go to see genome assembly and annotation, most recent assembly is usually best, click on assembly, FTP for RSEQ (on right side), |
| gunzip | decompress downloaded |
| control x | exit a nano article |
| gsutil ls gs://gu-biology-dept-class | access class storage bucket |
| gsutil cp gs://gu-biology-dept-class/file\_nam  e . | copy a file from the bucket into your working directory |
| echo | ? |
| sbatch NAME OF FILE | runs the sbatch file |
| command n | open a new window |
| sort file name | sorts rows by column one alphabetically |
| grep "protein\_coding\_gene" VectorBase-48\_AgambiaePEST.gff |cut -f9|more -d "-" -f1|sort|uniq|wc -l | find and count total number of protein coding genes |
| sort -r file name | sorts rows by column one anti alphabetically |
| sort -k2 file name | sort by second column |
| sort -k2n file name | sort based on numerical values for column 2, smallest to biggest |
| sort -k2nr file name | reverse, does biggest numbers to smallest |
| cut -f1 file name | cuts out everything but order one |
| cut -f1,2 file name | cuts out everything but columns one and 2 |
| cut -d “WHATEVER THE SPACER THING IS” -f1 file name | if file has spacer that is something else, specify what we are using to separate columns and this gives us just column 1 |
| uniq WORD file name | collapses all identical adjacent entries |
| sort FILE|uniq | collapses all identical ones, condenses to non redundant gene entries to a gff |
| module avail | tells you which modules are available on hpc |
| module load NAME | loads the module, when you’re done do the module unload |
| bedtools --help | pulls up the help file! |
| srun --pty bash | go to compute node |
| bowtie2 | check options for bowtie usage |
| bowtie2-build -h | check options to build index |
| bowtie2-build C\_albicans\_SC5314\_A22\_current\_chromosomes.fasta c\_alb | on compute node!! build the index files |
| module load samtools | load samtools module |
| samtools view -S -b sample.sam > sample.bam | convert sam to bam |
| samtools view sample.bam|more  samtools view -H sample.bam|more | view to visualize alignments, need  H to give you the headers also |
| samtools view sample.bam|wc -l | count number of alignments |
| samtools sort sample.bam sample.sorted.bam | sort |
| samtools index sample.sorted.bam | index! gives an output file is .bam.bai |

|  |  |
| --- | --- |
| cufflinks | for transcript assembly! |
| makeblastdb -dbtype prot -in FILE NAME | make index files for protein reference file for blast |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |