Contents

1. Connecting to Biowulf
   1. With a Mac computer
   2. With a Windows PC
2. The command line
3. Handy unix commands
4. Unix tips and tricks
5. File transfer connections to Biowulf
   1. Globus.org
   2. Mount a drive
   3. WinSCP (windows)
   4. scp at the command line (Mac)
   5. FileZilla (Mac and PC but be very careful downloading this file from the web)
6. File formats
   1. FASTA, FASTQ
   2. SAM, BAM
7. Modules (blast, fastxtoolkit, fastqc, bowtie, samtools)
8. IGV (Integrated Genome Viewer)
9. Additional resources
10. Connecting to Biowulf
    1. On Mac, open the “Terminal” program and type

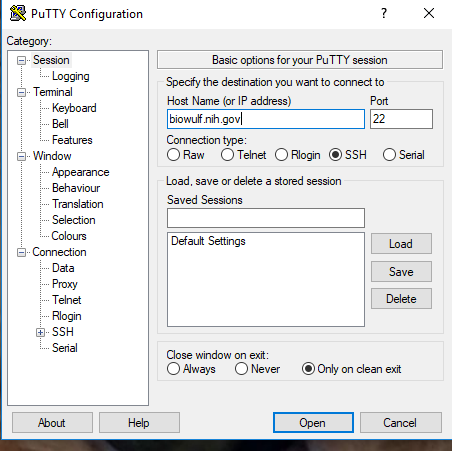
“ssh [username@biowulf.nih.gov](mailto:username@biowulf.nih.gov)” (without the quotes)

* 1. On PC, you will need to download and install PuTTY from:

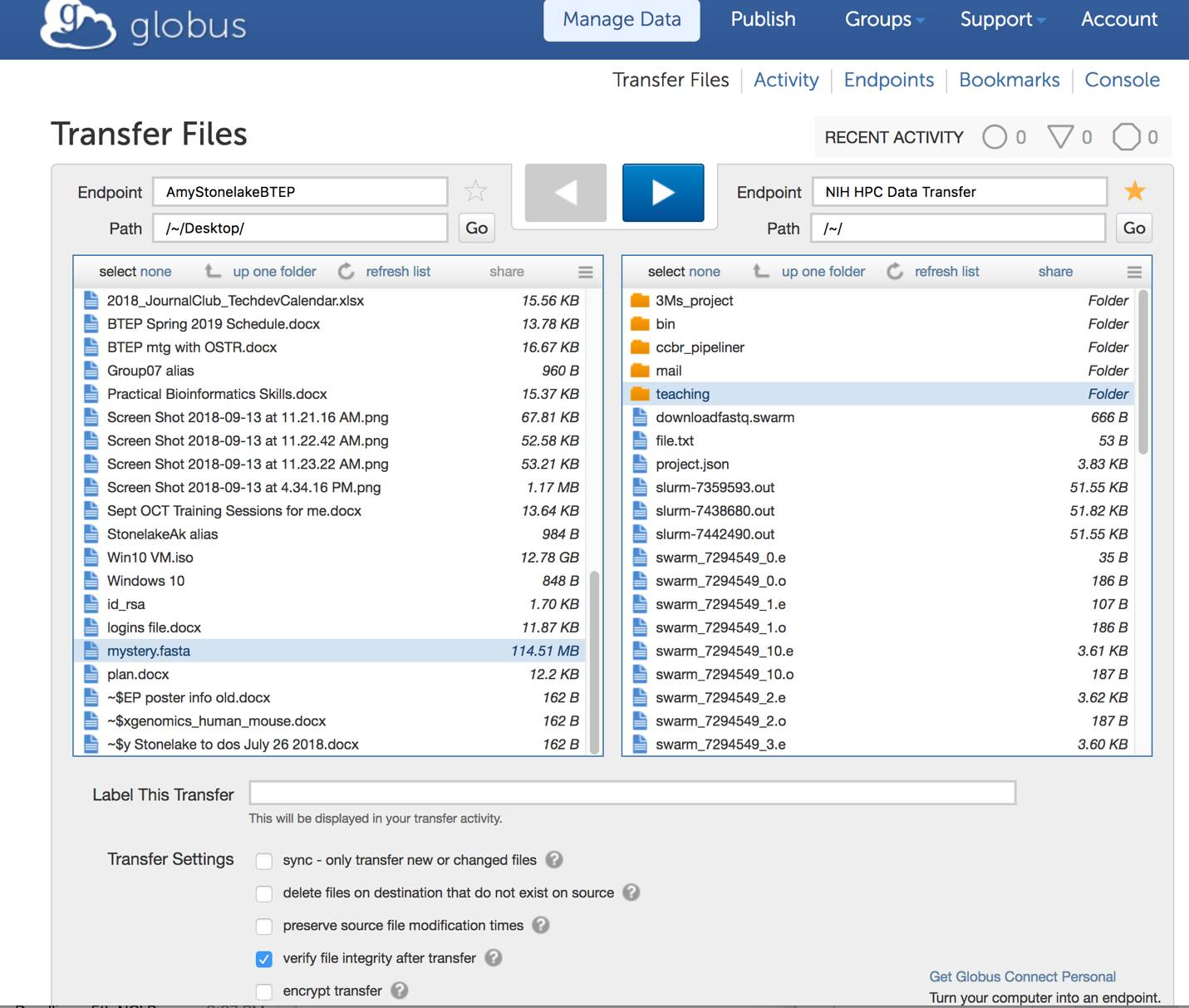
<https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html>

Host Name: biowulf.nih.gov

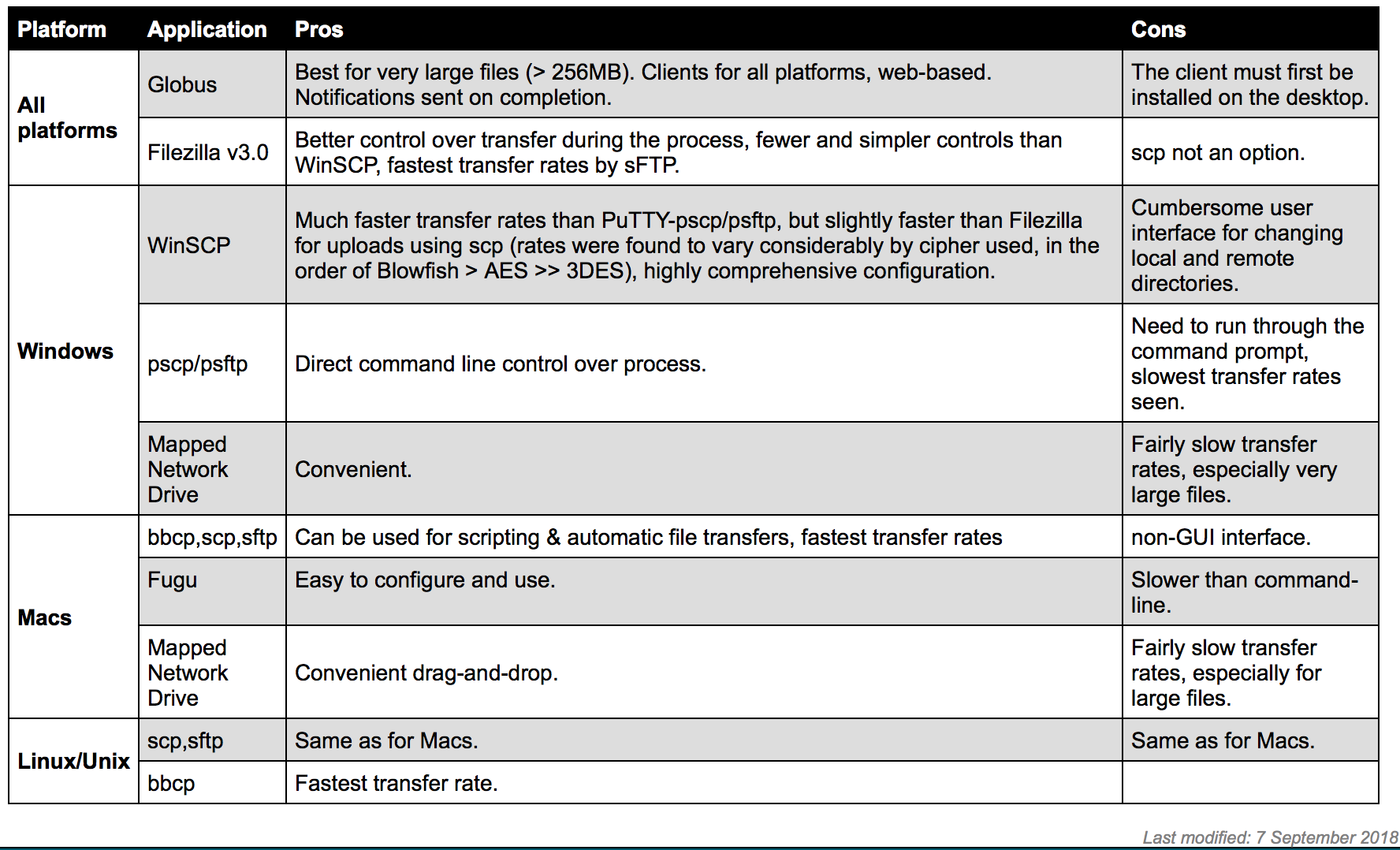
Connection Type: SSH



1. The Command line [username@biowulf ~] $
   1. “username” is your username
   2. @biowulf means you are logged into biowulf
   3. “~” indicates your home directory
   4. When you see the dollar sign “$”, you know you are at the command line
   5. If you don’t see the dollar sign, something is going on (running a program)
2. Handy unix commands
   1. pwd (print working directory)
   2. ls (list contents)
   3. cd (change directory)
   4. cd .. (go to home directory)
   5. cd /home/pathtofile
   6. less – peek inside a file, press “q” to quit
   7. ls -l (list details)
   8. ls -a (list hidden . files)
   9. rm (remove file)
   10. rmdir (remove directory)
   11. mkdir (make directory)
   12. nano file.txt (nano editor for creating files)
   13. move – move or rename files
3. Unix tips and tricks
   1. Use the “up arrow” to go back to the previous command
   2. Use “tab complete” to finish typing a uniquely named file, directory or program
4. File transfer connections to Biowulf
   1. Globus.org
      1. Setup your Globus endpoint (only need to do this one time)
      2. Open Globus Connect Personal (need to do this every time)
      3. Go to globus.org
      4. Choose your personal endpoint
      5. Choose a folder on biowulf
      6. Click the blue arrow
      7. You get an e-mail when it’s done!
      8. Need to have helix/biowulf account to get globus



* 1. Mount a drive
     1. Mac – “Go” -> ”Connect to server”
     2. PC - ”Computer”, “Tools” then “Map Network Drive” tab
     3. Be sure to set host as “smb://biowulf.nih.gov””
     4. See instructions on hpc.nih.gov (Biowulf) – “How To – Transfer Files”, “Transferring data to/from the NIH HPC systems



* 1. Secure ftp (sftp) or secure copy protocol (scp)
     1. FileZilla – be sure to get a clean copy!
     2. Mac OSX:

<http://packages.partek.com/bin/filezilla/fz-osx.app.tar.bz2>

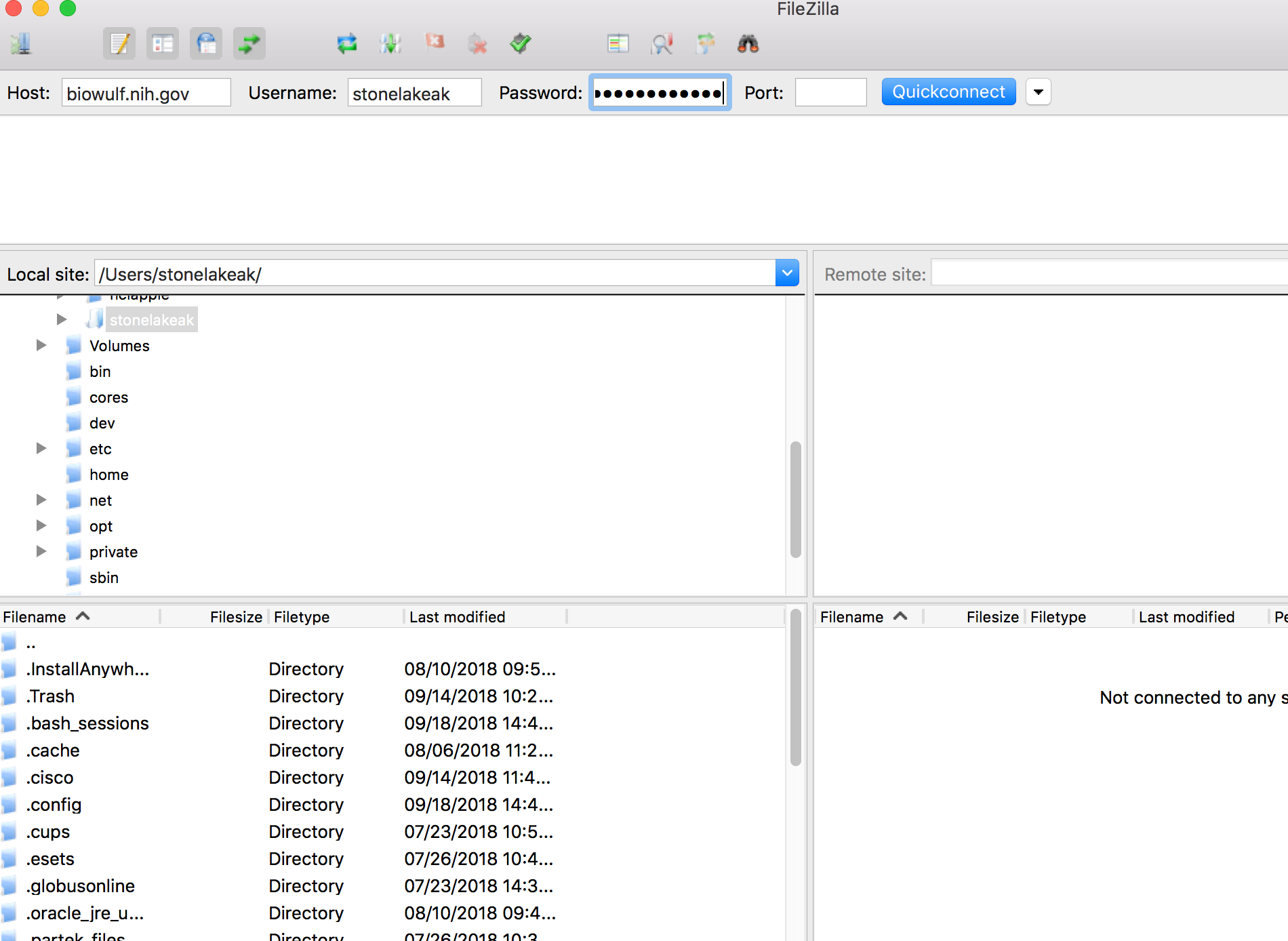
* + 1. Windows 32-bit:

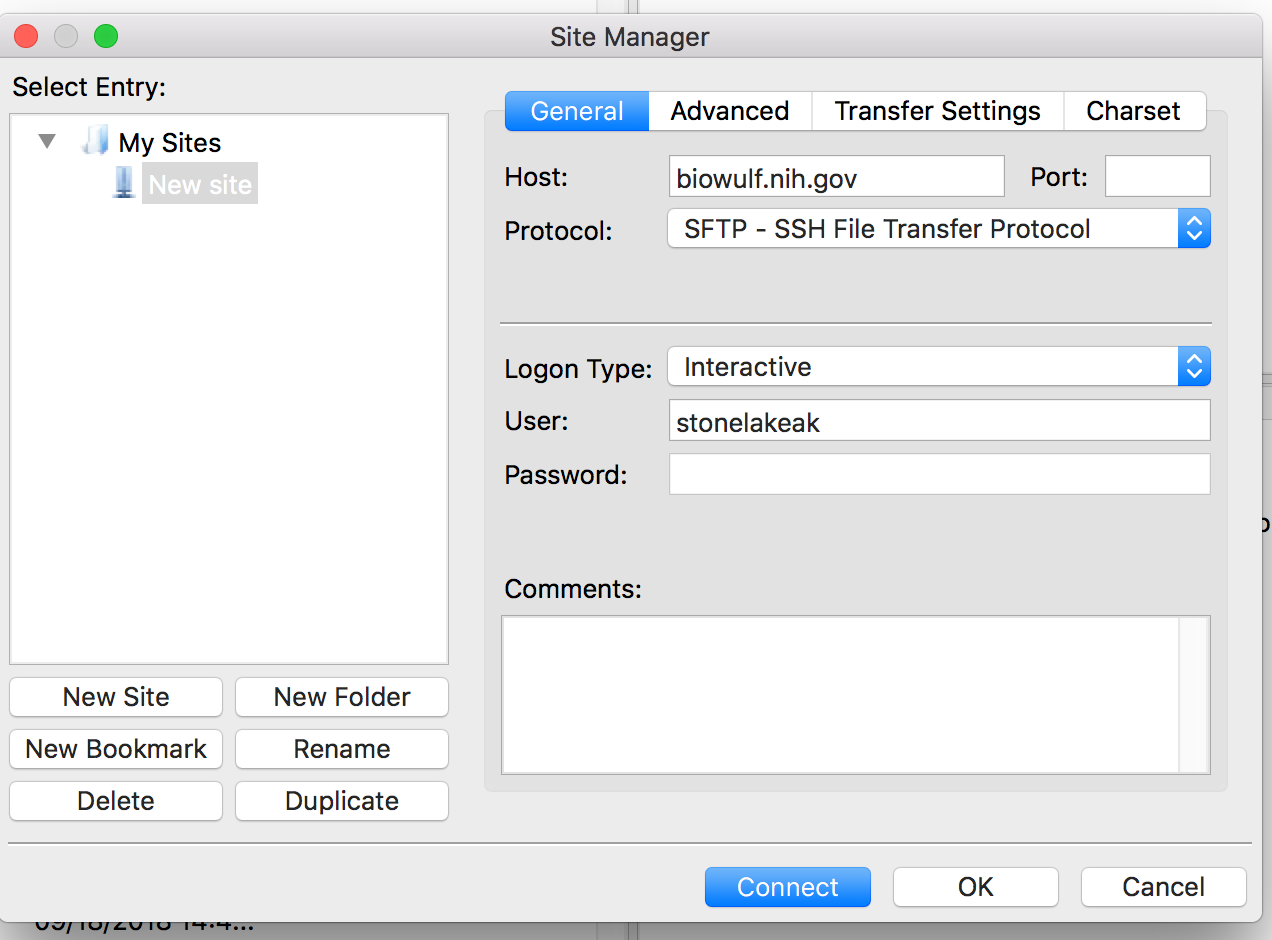
<http://packages.partek.com/bin/filezilla/fz-win32.exe>

* + 1. Windows 64-bit:

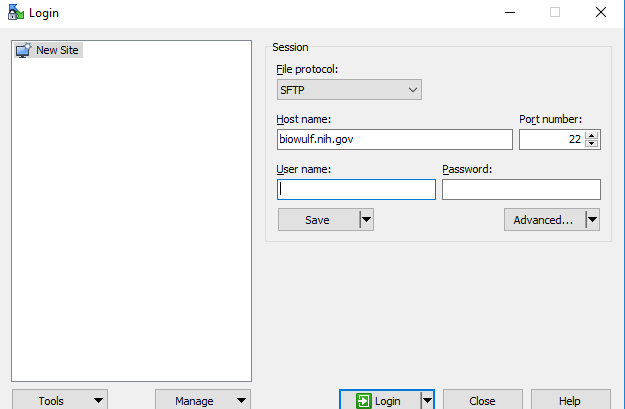
<http://packages.partek.com/bin/filezilla/fz-win64.exe>

FileZilla



FileZilla

WinSCP



1. File Formats
   1. FASTA format format has a header line followed by sequence data

>this\_is\_the\_name\_of\_the\_sequence

ACAGATAGACCAGATGACAGACAGATT

* 1. FASTQ – contains both sequence data and quality scores
  2. SAM – stores biological sequence aligned to a reference
  3. BAM – binary format of SAM, can be indexed

1. Modules (fastqc, fastxtoolkit, blast, samtools, bowtie)
   1. Use “module load” command
   2. “module avail” to see list of all modules
   3. “module spider” to do text matching on module name
   4. module load fastqc
      1. fastqc read1.fastq
      2. fastqc read2.fastq
      3. Generates html report you can download



* 1. module load fastxtoolkit
     1. fastq\_to\_fasta -i input\_file – o output\_file
  2. module load bowtie/2

bowtie2 -x /fdb/bowtie2.DELETE/hg19 -1 read1.fastq -2 read2.fastq -S aligned.sam

* 1. module load samtools

samtools view -b aligned.sam >aligned.bam

samtools sort aligned.bam >aligned\_sorted.bam

samtools index aligned\_sorted.bam

* 1. module load blast
     1. easyblast for interactive mode
     2. create .swarm file for large batches of sequence data

8. IGV (Integrated Genome Viewer) - [software.broadinstitute.org](http://software.broadinstitute.org/)

9. Additional resources

* Book ->“Unix and Perl to the Rescue, A Field Guide for the Life Sciences (and Other Data-rich Pursuits)”, Keith R. Bradnam & Ian Korf, 2012
* Web Site -> korflab.ucdavis.edu, Unix and Perl Primer for Biologists, Korf Lab, UC Davis
* hpc.nih.gov (Biowulf)
* Unix cheat sheet (Fosswire.com)

9. Additional resources

* Book ->“Unix and Perl to the Rescue, A Field Guide for the Life Sciences (and Other Data-rich Pursuits)”, Keith R. Bradnam & Ian Korf, 2012
* Web Site -> korflab.ucdavis.edu, Unix and Perl Primer for Biologists, Korf Lab, UC Davis
* hpc.nih.gov (Biowulf)
* Unix cheat sheet (Fosswire.com)