

Setting up a portable metagenome assembly pipeline for CBIO - Support #38

How to download SRA files to hex

04/20/2018 03:35 PM - Katie Lennard

Status:	New	Start date:	10/07/2017
Priority:	Normal	Due date:	
Assignee:		% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:		Spent time:	0.00 hour
Description			
<p>How to download input files (testdata) from SRA to hex:</p> <p>Use sra-tools installed on hex (activate by doing 'load module software/sra-tools')</p> <p>NB: when using the commands: fastq-dump or prefetch the .sra files are automatically downloaded to your home directory (/home/kviljoen/ncbi/public...) which we don't want. We want to store data in fhgfs. So to change this we need to point our \$HOME elsewhere by e.g. 'export HOME=/researchdata/fhgfs/cbio/cbio/project08/team/katie/ncbi_PRJNA290380_metagenomics_testset' (you will have to set this for every session).</p> <ul style="list-style-type: none">Files downloaded with the following command (for a given sample/run), which splits the reads into two files – one for F reads, one for R reads (see https://edwards.sdsu.edu/research/fastq-dump/) fastq-dump --gzip --skip-technical --readids --read-filter pass --dumpbase --split-files --clip SRR4408075 Output: SRR4408075_pass_1.fastq.gz SRR4408075_pass_2.fastq.gzThe file integrity can be validated with the vdb-validate command: vdb-validate SRR4408075 NB: if you just do fastq-dump without prefetch it will still get the .sra file but I don't think it does the .cache file which you seem to need to do the vdb-validate checkLoop command to download multiple files: for j in \$(<sra_list); do echo \$j; echo fastq-dump --readids --read-filter pass --dumpbase --split-files --clip \$j; done The -clip flag is to remove Illumina-specific sequences sra_list is just a file with sampleIDs (one per line)Loop command to validate multiple files: for j in \$(> vdb_validate_results; done Note the >> vdb_validate_results didn't work to output results to file (also tried tee -a but didn't work) But all files were validated as output to screenInput files were downloaded to /researchdata/fhgfs/cbio/cbio/project08/team/katie/ncbi_PRJNA290380_metagenomics_testset/			