Using SPAdes

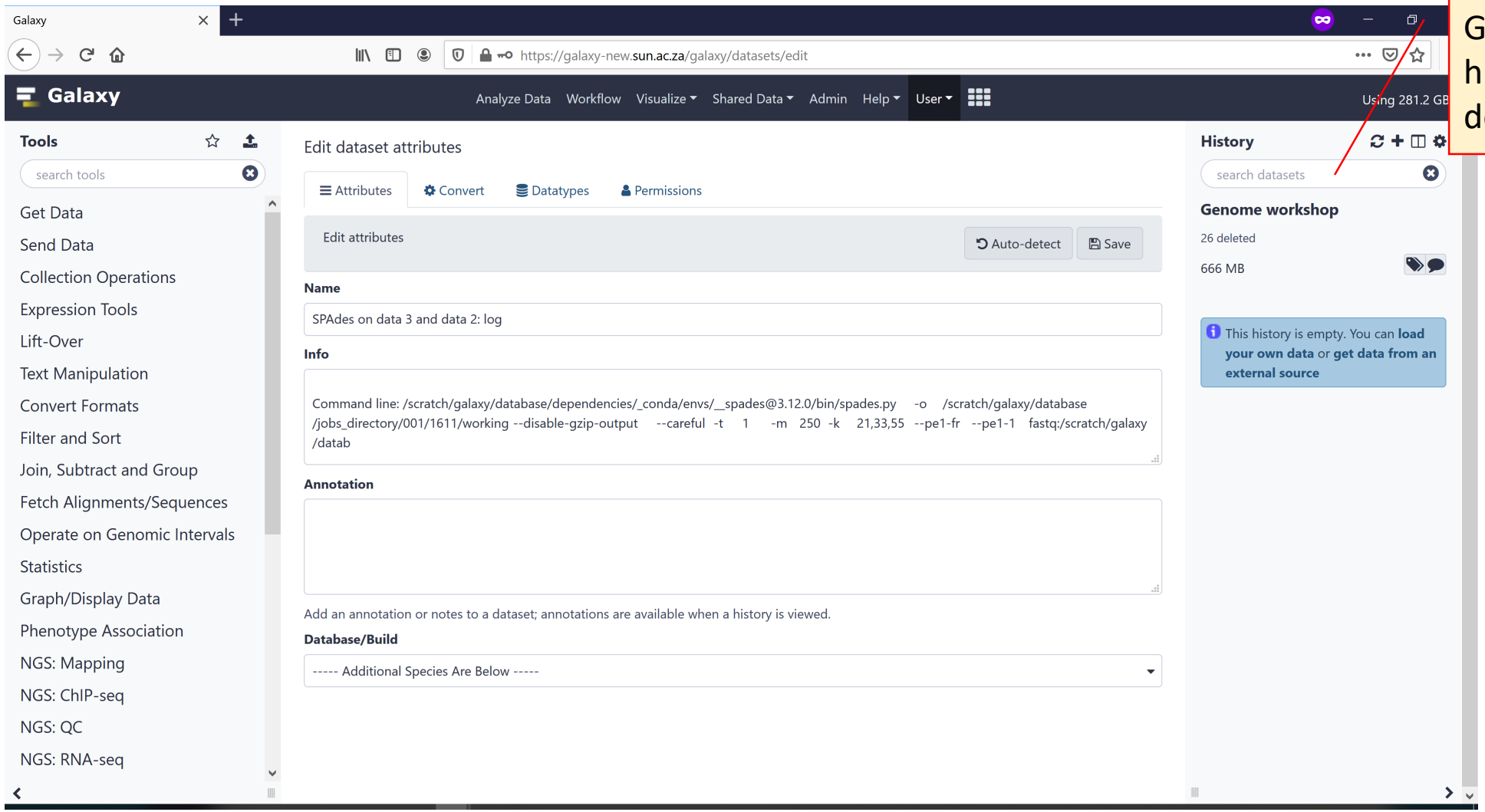
SPAdes pipeline

SPAdes comes in several separate modules

BayesHammer
read error correction tool for Illumina reads, which works well on both single-cell and standard data sets.
IonHammer
read error correction tool for IonTorrent data, which also works on both types of data.
SPAdes
iterative short-read genome assembly module; values of K are selected automatically based on the read length and data set type.
MismatchCorrector
a tool which improves mismatch and short indel rates in resulting contigs and scaffolds; this module uses the BWA tool

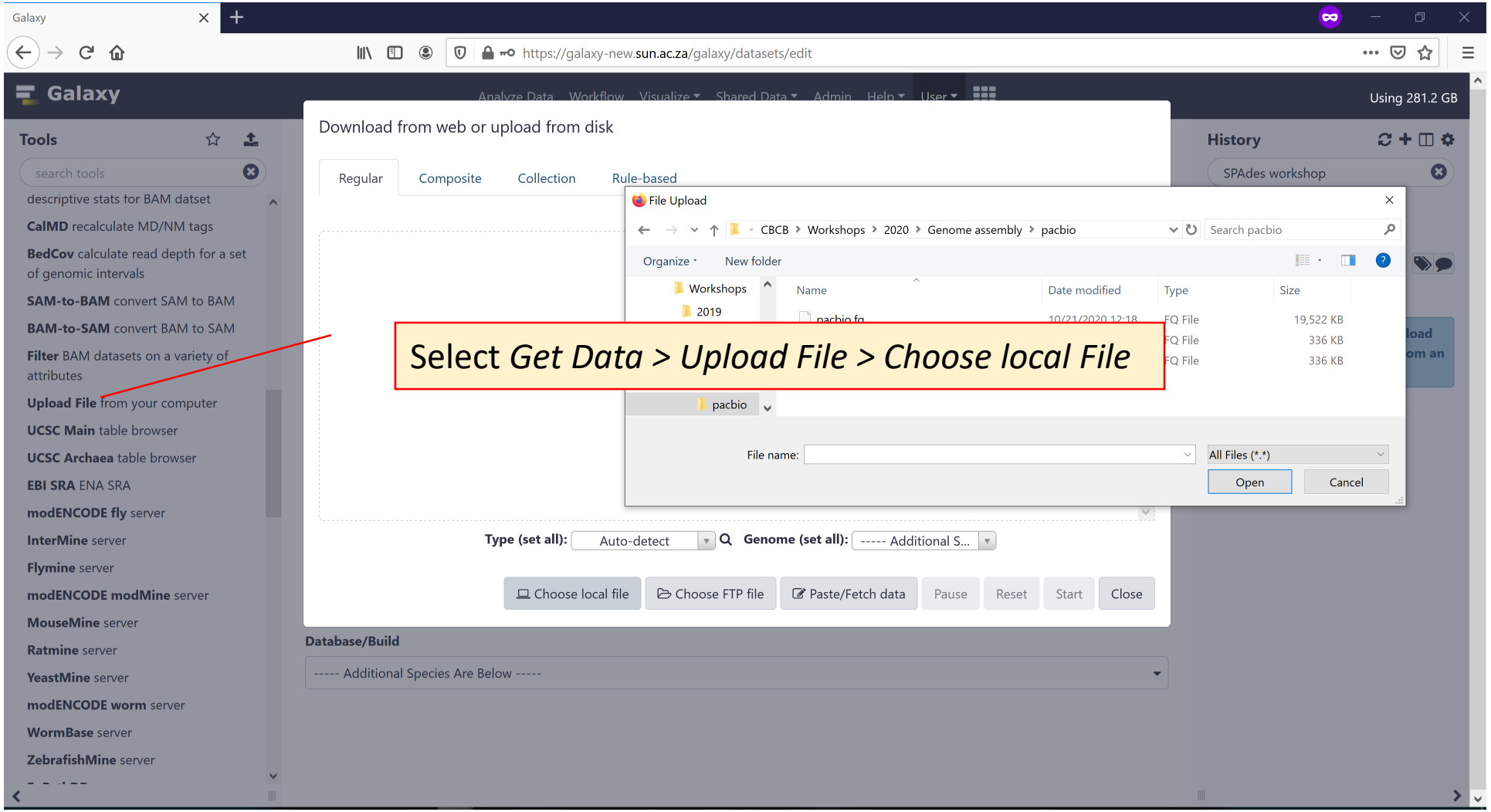
MismatchCorrector is turned off by default, but we recommend to turn it on
We recommend to run SPAdes with BayesHammer/IonHammer to obtain high-quality assemblies.
However, if you use your own read correction tool, it is possible to turn error correction module off. It is also possible to use only the read error correction stage, if you wish to use another assembler. S

Get a new, blank history

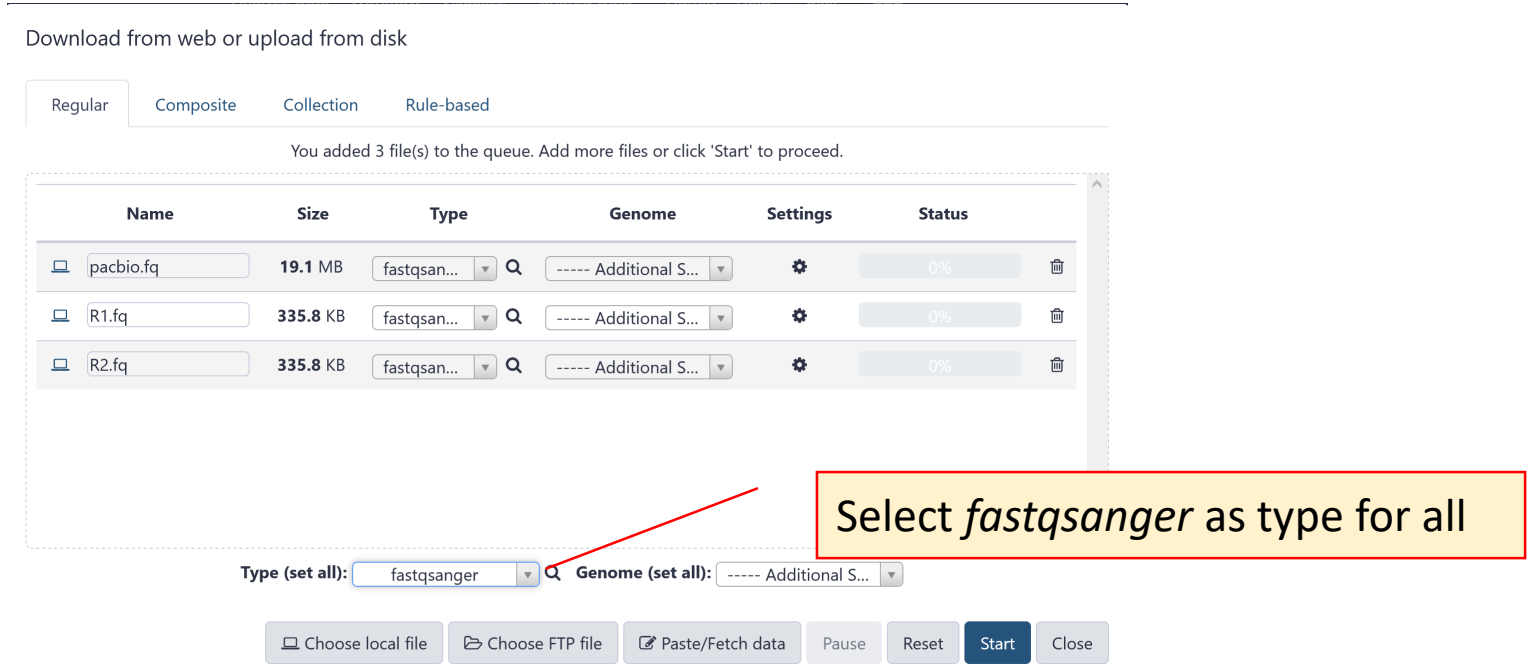


Give your new history a descriptive name

Gettings data to Galaxy

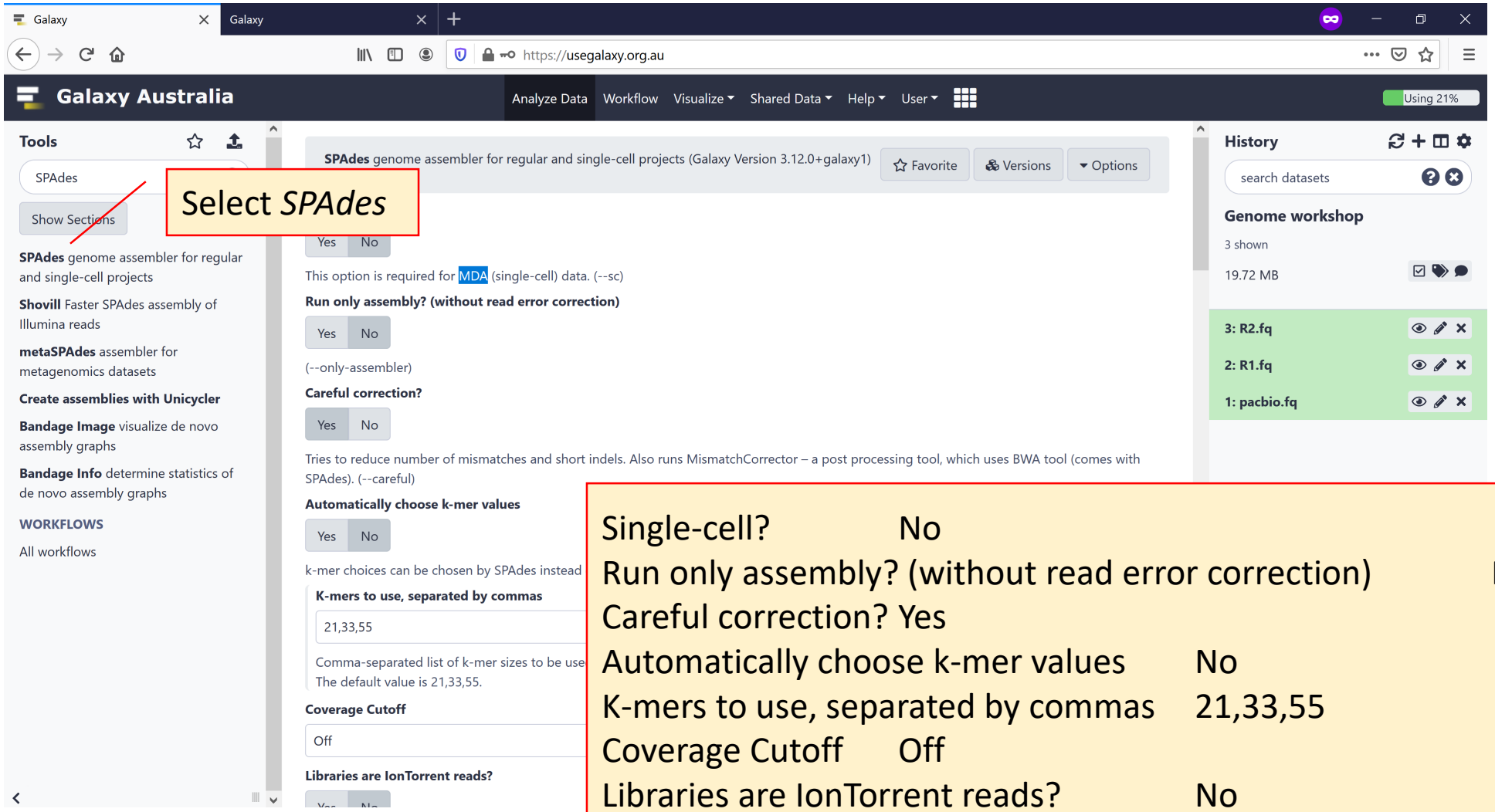


Select the file type



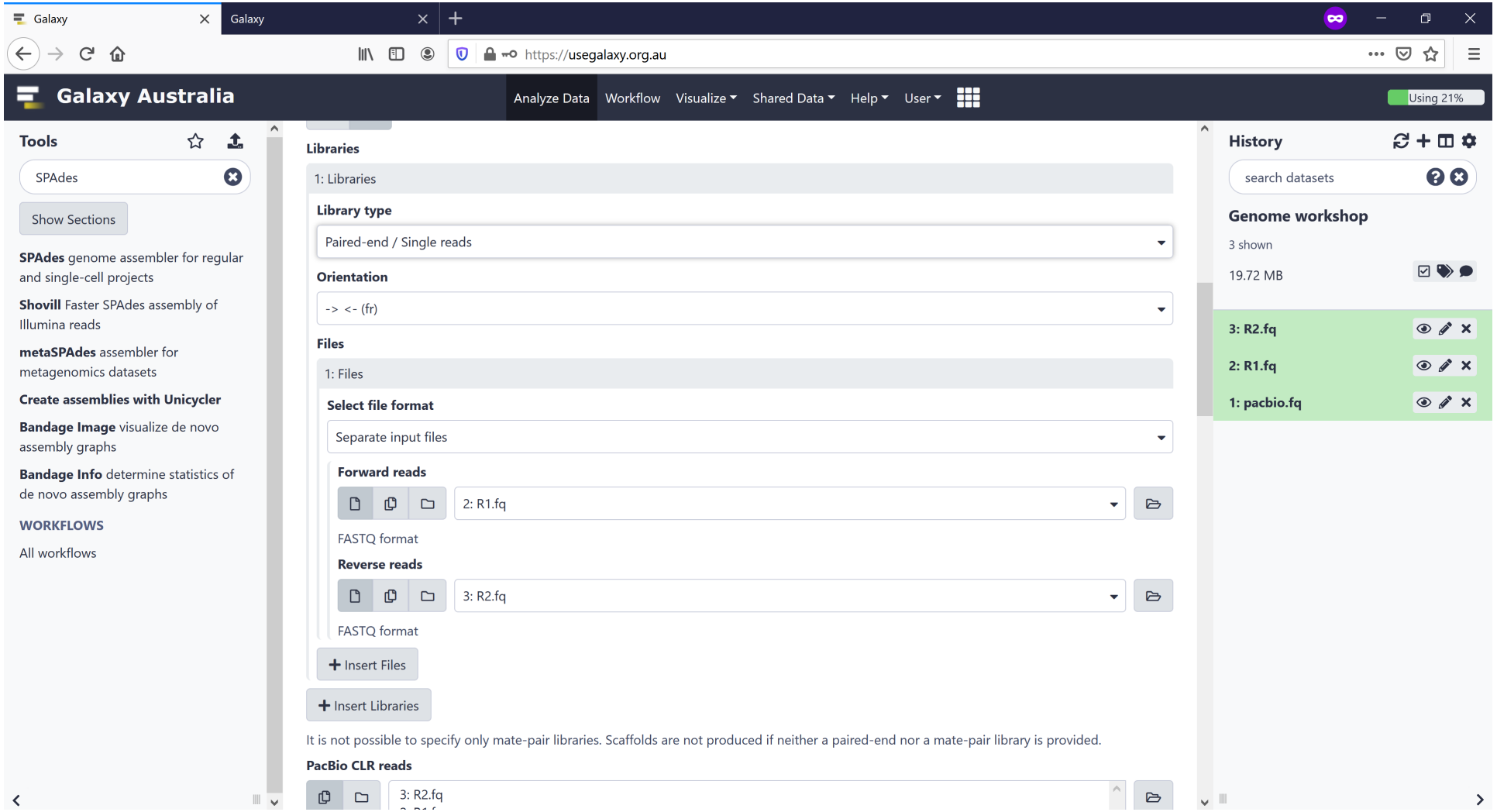
Select *fastqsanger* as type for all

SPAdes settings



Single-cell? No
Run only assembly? (without read error correction) No
Careful correction? Yes
Automatically choose k-mer values No
K-mers to use, separated by commas 21,33,55
Coverage Cutoff Off
Libraries are IonTorrent reads? No

Select the forward and reverse Illumina read files



Select

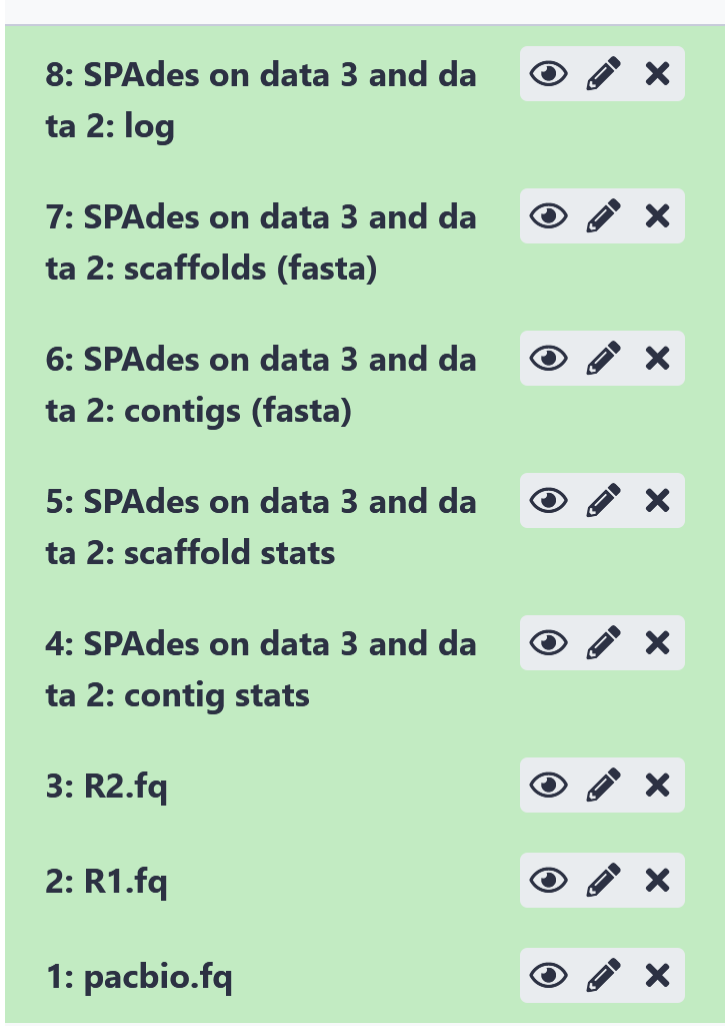
Data compression/decompression with gzip

Usage: gzip [OPTION]... [FILE]...
Compress or uncompress FILEs (by default, compress FILES in-place).

Mandatory arguments to long options are mandatory for short options too.

```
 -c, --stdout      write on standard output, keep original files unchanged
 -d, --decompress  decompress
 -f, --force       force overwrite of output file and compress links
 -h, --help        give this help
 -l, --list        list compressed file contents
 -L, --license     display software license
 -n, --no-name     do not save or restore the original name and time stamp
 -N, --name        save or restore the original name and time stamp
 -q, --quiet       suppress all warnings
 -r, --recursive   operate recursively on directories
 -S, --suffix=SUF  use suffix SUF on compressed files
 -t, --test        test compressed file integrity
 -v, --verbose     verbose mode
 -V, --version     display version number
 -1, --fast        compress faster
 -9, --best        compress better
  --rsyncable   Make rsync-friendly archive
```

SPAdes output



- There should now be 5 new files in your history
- You can select the eye icon to view the file
- Note the small contigs with very high coverage
- Scaffolds = contigs: np scaffold assembly

name	length	coverage
#name	length	coverage
NODE_1	132140	12.619781
NODE_2	35102	12.666048
NODE_3	5925	13.405451
NODE_4	4666	25.835610
NODE_5	790	65.717007
NODE_6	533	10.721757
NODE_7	297	21.004132
NODE_8	111	13.571429

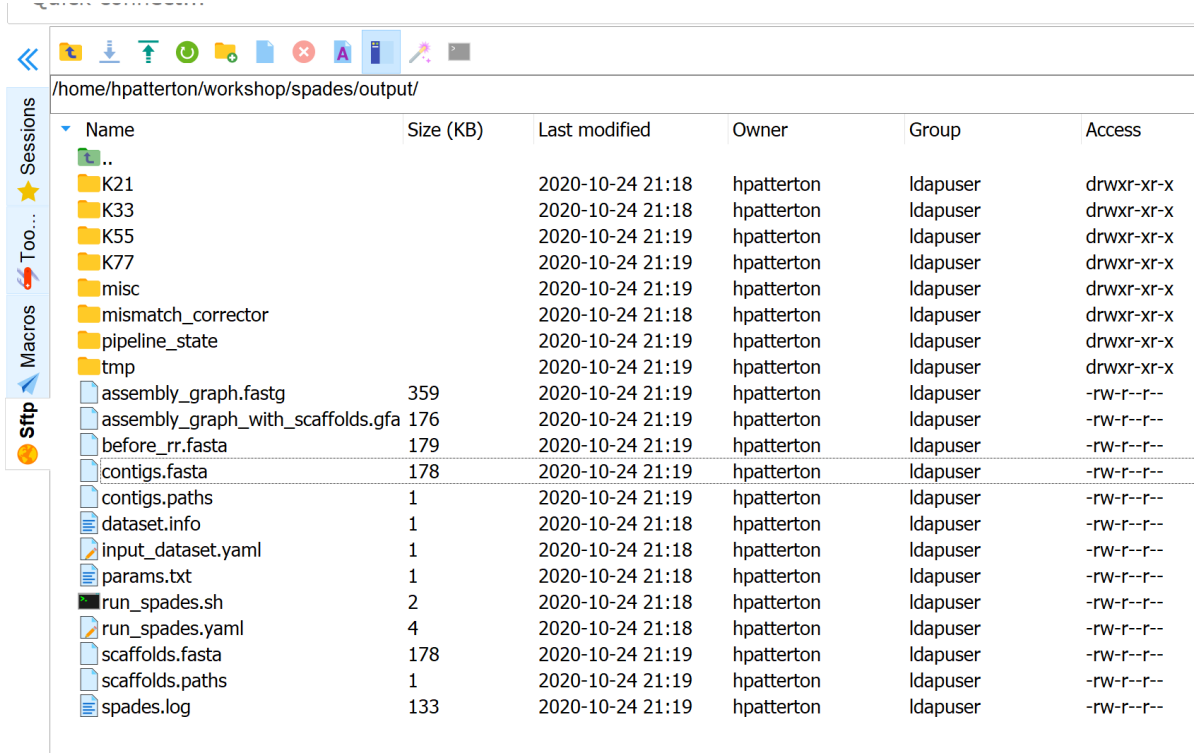
SPAdes on the HPC

```
#!/bin/bash

#PBD -N workshop_spades
#PBS -l walltime=1:00:00
#PBS -l ncpus=1
#PBS -l mem=1GB
#PBS -e myprog.err
#PBS -o myprog.out
#PBS -M hpatterton@sun.ac.za
#PBS -m abe

cd $PBS_O_WORKDIR
module load app/SPAdes/3.14.0
spades.py -o $PBS_O_WORKDIR --careful -1 $PBS_O_WORKDIR/reads/mutant_R1.fastq.gz
-2 $PBS_O_WORKDIR/reads/mutant_R2.fastq.gz
echo Done!
```

SPAdes output



Name	Size (KB)	Last modified	Owner	Group	Access
..					
K21		2020-10-24 21:18	hpatterton	ldapuser	drwxr-xr-x
K33		2020-10-24 21:18	hpatterton	ldapuser	drwxr-xr-x
K55		2020-10-24 21:19	hpatterton	ldapuser	drwxr-xr-x
K77		2020-10-24 21:19	hpatterton	ldapuser	drwxr-xr-x
misc		2020-10-24 21:19	hpatterton	ldapuser	drwxr-xr-x
mismatch_corrector		2020-10-24 21:18	hpatterton	ldapuser	drwxr-xr-x
pipeline_state		2020-10-24 21:19	hpatterton	ldapuser	drwxr-xr-x
tmp		2020-10-24 21:19	hpatterton	ldapuser	drwxr-xr-x
assembly_graph.fastg	359	2020-10-24 21:19	hpatterton	ldapuser	-rw-r--r--
assembly_graph_with_scaffolds.gfa	176	2020-10-24 21:19	hpatterton	ldapuser	-rw-r--r--
before_rr.fasta	179	2020-10-24 21:19	hpatterton	ldapuser	-rw-r--r--
contigs.fasta	178	2020-10-24 21:19	hpatterton	ldapuser	-rw-r--r--
contigs.paths	1	2020-10-24 21:19	hpatterton	ldapuser	-rw-r--r--
dataset.info	1	2020-10-24 21:18	hpatterton	ldapuser	-rw-r--r--
input_dataset.yaml	1	2020-10-24 21:18	hpatterton	ldapuser	-rw-r--r--
params.txt	1	2020-10-24 21:18	hpatterton	ldapuser	-rw-r--r--
run_spades.sh	2	2020-10-24 21:18	hpatterton	ldapuser	-rw-r--r--
run_spades.yaml	4	2020-10-24 21:18	hpatterton	ldapuser	-rw-r--r--
scaffolds.fasta	178	2020-10-24 21:19	hpatterton	ldapuser	-rw-r--r--
scaffolds.paths	1	2020-10-24 21:19	hpatterton	ldapuser	-rw-r--r--
spades.log	133	2020-10-24 21:19	hpatterton	ldapuser	-rw-r--r--

- Always make sure that myprog.err is empty, and that contents of myprof.out has no information that could indicate problems
- The contigs.fasta and scaffolds.fasta in the top directory are the final output files
- Assembly graph and assembly graph with scaffold files are also produces
- You can view these files in *bandage* (see later)

Output of contigs

```
>NODE_7_length_577_cov_6.114000
CATGCTAGCAAGTTAAGCGAACACTGACATGATAAATTAGTGGTTAGCTATATTTTTTTA
CTTTGCAACAGAACCGAAAATAATCTCTTCAATTTATTTTTATATGAATCCTGTGACTCA
ATGATTGTAATATCTAAAGATTTCAGTTCATCATAGACAATGTTCTTTTCAACATTTTTT
ATAGCAAATTGATTAAATAAATTCTCTAATTTCTCCCGTTTGATTTCACTACCATAGATT
ATATTATCATTGATATAGTCAATGAATAATGACAAATTATCACTCATAACAGTCCCAACC
CCTTTATTTTGATAGACTAATTATCTTCATCATTGTAAAACAAATTACACCCTTTAAATT
TAACTCAACTTAAATATCGACAAATTAAAAAACAATAAAATTACTTGAATATTATTCATA
ATATATTAACAACTTTATTATACTGCTCTTTATATATAAAATCATTAATAATTAAACAAG
CCTTAAAATATTTAACTTTTTTGTGATTATTACACATTATCTTATCTGCTCTTTATCACC
ATAAAAATAGAAAAAACAAGATTCCTAAAGAATATAG
>NODE_8_length_319_cov_11.933884
TCGTATTCTTCGACTGATAATTGCTCTCTAGATTCTAGCATATTTAAGTGTTTCTCTTTA
TCTAATGCTTTGTCATATCCTTTAACGATTGAACCACTAAAGATTTCTCCTACTGCTCCT
GAACCATAACTAAATAGACATACTTTCTCTTCTGGTTGGAATGTGTGGTTCTGTAATAAC
GAAATTAAACTTAAGTATAATGATCCTGTATAAATGTTACCAACATCTCTATTCCATAAT
ACGGTTCTGTTGCAAAGTTGAATTTATAGTATAATTTTAACAAAAAGGAGTCTTCTGTAT
GAACTATTTCAGATATAAA
>NODE_9_length_155_cov_7.538462
TTACTCTTTCAGCCTTTTTAAATTCAAGAATATGCAGAAGTTCAAAGTAATCAACATTAG
CGATTTTCTTTTCTCTCCATGGTCTCACTTTTCCACTTTTTGTCTTGTCCACTAAAACCC
TTGATTTTTCATCTGAATAAATGCTACTATTAGTA
```