My completed code and data is contained in the folder `assembler/`.

1

The source code is contained in `src`. The program can be compiled by running `ant` and run with `bin/assembler <args>`. I set up a script `run_trials` which I used to easily playing with the parameters on my runs.

2

Results from the small first-week genome are contained in `reduced_genome/`.

3

Results from the second week reads are contained in the `full_genome` folders and labeled by coverage.

4

The approximate value of parameter a (coverage) required to get c contigs with the parameters $l = 50$ and $G = 10000$ is calculated as follows:

Given

$$c = Ne^{\frac{Nl}{G}}$$
$$a = \frac{Nl}{G}$$

we find

$$c = \frac{G}{l}ae^{-a}$$
$$ae^{-a} = \frac{cl}{G}$$

which cannot be easily simplified.

5