# Team Exercise

There are three exercises, which can be done in any order.

The goal of each exercise is to run QC tests to figure out why the datasets cannot be assembled into a single genome. If the cause cannot be determined with certainty, then suggest hypotheses consistent with the evidence. The final step is to identify how one can estimate the genome size of the sample organism.

There are five points for each exercise.
1 point for identifying the issue
1 point for identifying some of the evidence
1 point for identifying all of the evidence
1 point for describing how to identify the genome size
1 point for the correct estimation.

## Exercise 1

Genome: Enterococcus
This is a public dataset http://www.ebi.ac.uk/ena/data/view/PRJNA60717.

### Data:

SRR492065_{1,2}.fastq.gz

### Questions:

What is the issue with this dataset?
What is the underlying evidence?
What steps should be taken to identify the genome size?
What is the estimated size of the genome?

## Exercise 2

Genome: Spironucleus salmonicida
This is a public dataset http://www.ebi.ac.uk/ena/data/view/PRJNA60811.

### Public Dataset ():

SRR948594_{1,2}.fastq.gz
SRR948595_{1,2}.fastq.gz

### Questions:

What is the issue with this dataset?
What is the underlying evidence?
What steps should be taken to identify the genome size?
What is the estimated size of the genome?

## Exercise 3

This is a bacteria of unknown origin.

### Dataset:

Bacteria_001_{1,2}.fastq.gz
Bacteria_002_{1,2}.fastq.gz

### Questions:

What is the issue with this dataset?
What is the underlying evidence?
What steps should be taken to identify the genome size?
What is the estimated size of the genome?