NGS Bioinformatics
Practical assignment

Module topic: genome assembly
Contact session title: Module10_Day1
Trainer: Eugene Gardner
Participant: <*write your name here*>
Date: <*write today's date here*>

Module 10 Genome assembly Day 2

Please note

- **Hand-in information** please upload your completed assignment to the Vula 'Assignments' tab. Take note of the final hand-in date for each assignment, which will be indicated on Vula.

Please ONLY provide answers to the exercise questions in the practical assignment document. The numbering for the questions in each section are provided below:

3. Assembly algorithms

1. What is the contig sequence?
2. What was difficult here?

4. Illumina Genome Assembly

Write down the results for each assembly made using different k-mer sizes. Which one looks the best?
Question: What is the best choice for k?

k-mer	nodes	n50	average contig	largest contig
41				
49				
55				

Question: How does the contig N50 compare to the scaffold N50 for each of your assemblies?

k-mer	nodes	contig n50	scaffold n50
41			
49			

55			

5. Assembly estimation

Q1. What is the predicted heterozygosity?

Q2. What is the predicted genome size?

Q3. Does this seem reasonable?

6. Pac Bio Genome assembly cont.d

6.1 How does it compare to the Illumina assembly?

6.2 How does the wtdbg2 assembly compare to the canu assembly?

What do you notice in terms of the number of SNP and indel calls?

6.3 When running this analysis on these polished genomes, do we still get variants? More or less than with the raw canu and wtdbg2 assemblies? Why?