Genomics

Freie Universität Berlin, Institut für Informatik
Peter Robinson, Rosario Piro
Wintersemester 2015/2016

2. Übungsblatt
Diskussion am 29.10.15

Exercise 1.

Recall our definitions of our idealized genome from lecture #2.

- G: genome size (e.g., 3.2×10^9 nucleotides for humans)
- L: read length (e.g., 100 nucleotides for a typical Illumina run)
- N: read number
- n_b: total number of total bases (analogous for k-mers)
- d_b: mean base depth (analogous for k-mers)

In the lecture, we calculated the sequencing depth needed such that at least 99% of the genome would be covered by at least one read. Under the same slightly unrealistic assumptions as in the lecture, calculate the number of reads required to ensure that 99.9% of the entire genome is covered. You may assume that the human genome is exactly 3 Gigabases in length.

Exercise 2.

a) Explain the difference between a `contig` and a `scaffold`?

b) Explain why demanding a larger overlap between reads in order to combine reads into contigs is associated with a lower expected contig size?

Exercise 3.

a) Recall from the lecture, our construction of a de Bruijn graph from the Dr Seuss sentence, "The more that you read, the more things you will know. The more that you learn, the more places you'll go." Recall that we created nodes from all "2-mers" from this sentence (where two mer refers to a sequence of two words). Two mer nodes were connected if the last word of one word was the same as the first word of the other node. Let us now follow the example of Michael Schatz and create a de Bruijn graph from the first part of the first sentence of Dicken's *A Tale of Two Cities*:

> It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness

To do so, create 4-mer nodes and conenct them if the last three words of one node are the same as the first three words of another node. Draw the resulting de Bruijn graph. Now try to reconsruct the original sentence from you graph. How many possible "sequences" can you find?

b) Now simplify your graph

Exercise 4.

a) Does the following graph have a Eulerian tour? Cycle?

b) Does the following graph have a Eulerian tour? Cycle?

b) Does the following graph have a Eulerian tour? Cycle?