BST281: Genomic Data Manipulation, Spring 2017

Monday 04: Sequencing

Sign up for journal club!

 Email instructional team with group, paper, date, and justification by end of Wednesday.

 20 minutes + 5 minutes questions on March 1, 6, or 8, all group members must present.

 Background, overview/results, methods, discussion/interpretation, questions.

Sequencing technologies have evolved from southern blots:

 Radiolabeling + sequencing by synthesis with random termination.

 Sanger sequencing: sequencing by synthesis with four fluorescent dyes (clone libraries).

 1000s of NTs, essentially no error.

 454 emulsion PCR to avoid cloning. 100s of NTs, homopolymer runs.

 Illumina to increase throughput + reduce cost. 100s of NTs, very low error.

 PacBio and Nanopore using competing single-cell technologies to increase read lengths.

 10000s of NTs, variable error modalities.

Because all reads are shorter than whole genomes, hierarchical shotgun sequencing is ubiquitous.

 Sequencing proceeds in read mate pairs read from opposite ends of short fragments.

 Read quality degrades (in most technologies) as it proceeds.

 Dephasing, degradation, cross-talk, homopolymers (or equivalent).

 Insert sizes = length of physical nucleotides between the read portions.

 Different sizes for different strategies, local vs. global assembly.

# Python

Loops: while and for

 while condition: body, repeats while condition is true

 for variable in collection: body (using variable), repeats once with variable set to each element in collection

break stops the current loop immediately, continues below

continue stops the current loop immediately, continues next iteration

# Textbooks

Sequencing: Pevsner, Chapter 2 p19-58, stop before "Access to Biomedical Literature";

Chapter 3, p69-78, stop before "Pairwise Alignment, Homology, and Evolution";

p104-112, start at "Pairwise Alignment with Dot Plots";

Chapter 5, p167-168, stop before "Specialized BLAST Sites"

Python loops: Haddock and Dunn, Chapter 9 p141-172

# Literature

[The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. Sela et al, PNAS 2008](https://www.ncbi.nlm.nih.gov/pubmed/19033196)

[Scaffolding of a bacterial genome using MinION nanopore sequencing. Karlsson et al, Nature SciRep 2015](https://www.ncbi.nlm.nih.gov/pubmed/26149338)