BMMB 852: Applied Bioinformatics

Week 10, Lecture 20

István Albert

Bioinformatics Consulting Center
Penn State, 2015

Sequencing coverage a single uniform population

Coverage for an even mixture of two bacteria

Coverage for different abundances of two bacteria
In class exercise:
Coverage for unknown abundance of unknown number of bacteria

In class exercise:
Coverage for a transcriptome: transcripts of variable lengths expressing at different levels

Realistic coverage measures

Neither of the models assumptions are correct

Empirical observation → raise the required coverage at least 10 fold

What part of the genome is coverable to begin with?

What part of the genome is uniquely coverable with a give read size?

Nomenclature: “accessible”, “mappable”, “effective” genome sizes

Introduce variations then observe their effects
Pileups, variant calling

• Pileup → show all bases at a given index

---

Homework 20

• Introduce one large scale modification into the Ebola genome.

• Simulate data from this mutated genome.

• Align the data to the non-mutated genome and visualize the alignment.

• What aspect of the data or visualization shows the changes that you have introduced?