
DTU

DTU Health Technology
Bioinformatics

Week 6 Recap

Gisle Vestergaard
Associate Professor
Section of Bioinformatics
Technical University of Denmark
gisves@dtu.dk

Last Week

- Quiz I or Deliverable V
- Sequence alignment
- 16s rRNA amplicon analysis
- Quantitative metagenomics including counting animals of the savanna
- Exercises…

Today

- Lessons from Quiz I
- Work on exercises
- Metagenome diversity exercise
- Metagenomic *de novo* assembly
- More computer exercises

Last weeks quiz

- How many lines is in a fastq file? What does each of the four lines contain?

Four lines...I literally gave you the answer in the next line

Line 1: @sequence identifier

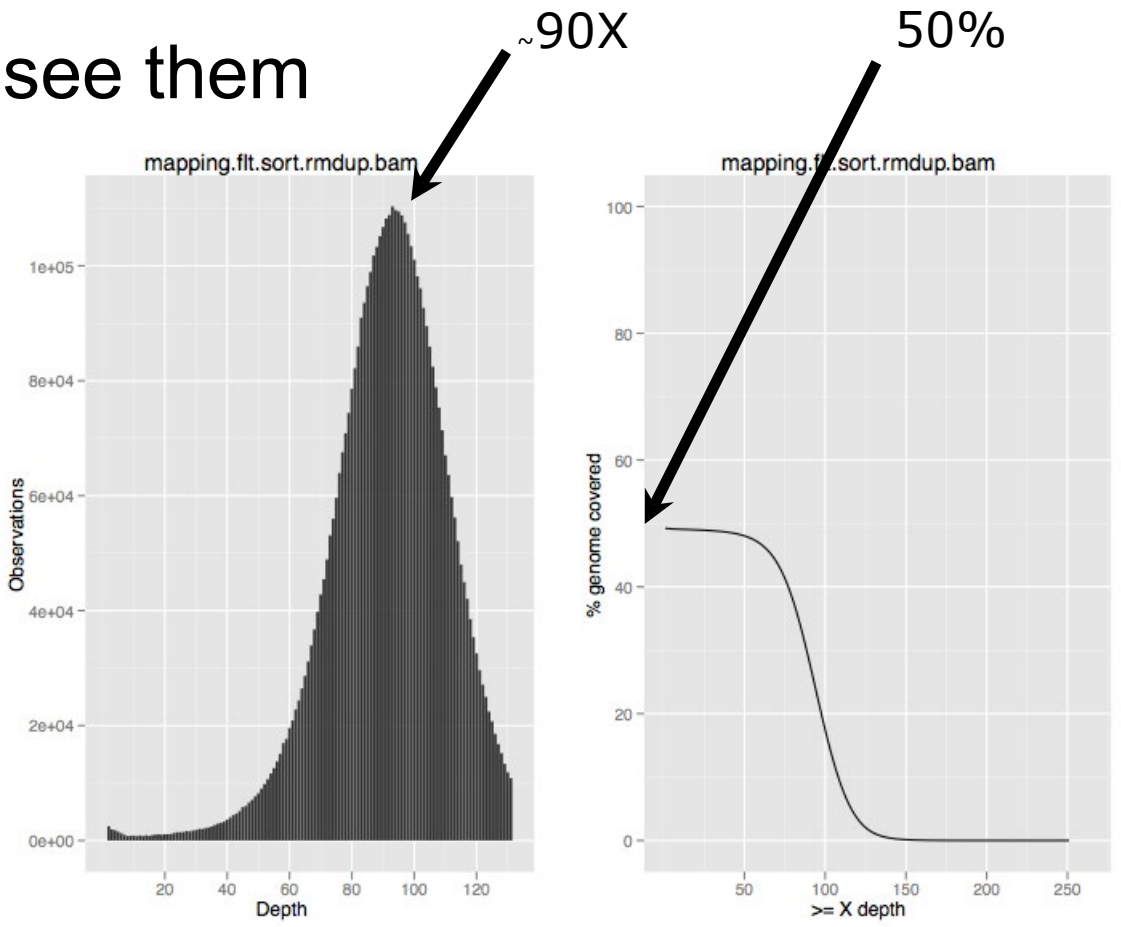
Line2: raw sequence

Line3: + (seldomly also the sequence identifier

Line4: Sequence quality score. Must (obviously) contain the same number of scores as letters in the raw sequence

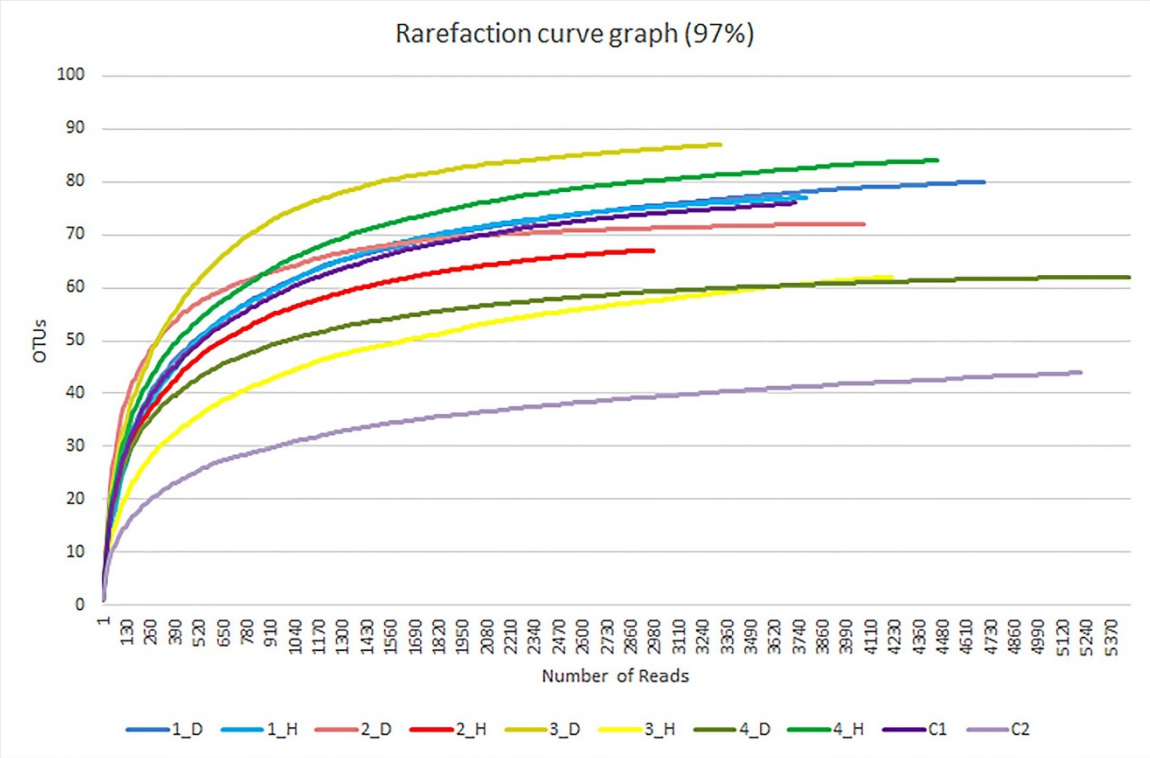
Genome coverage using kmers

- How many 15-mers do we observe
- How often (depth) do we see them
- Avg. depth ~ 90X
- Range from 0-250X
- Only 50% of the genome covered



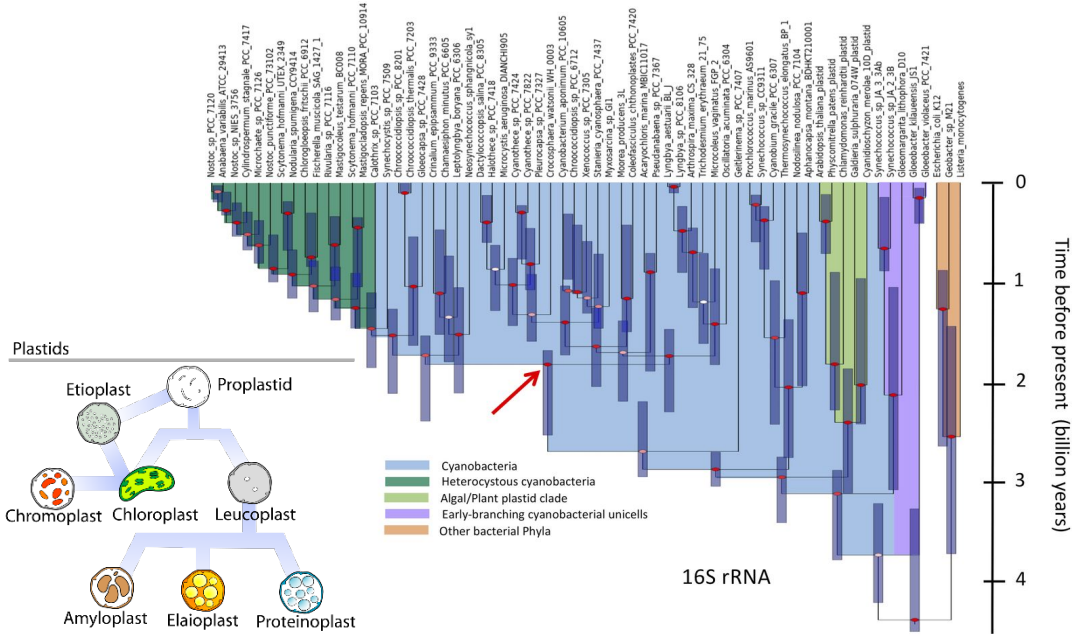
5. How does one check a 16s rRNA amplicon sample for adequate sequencing depth?

- Rarefaction plot



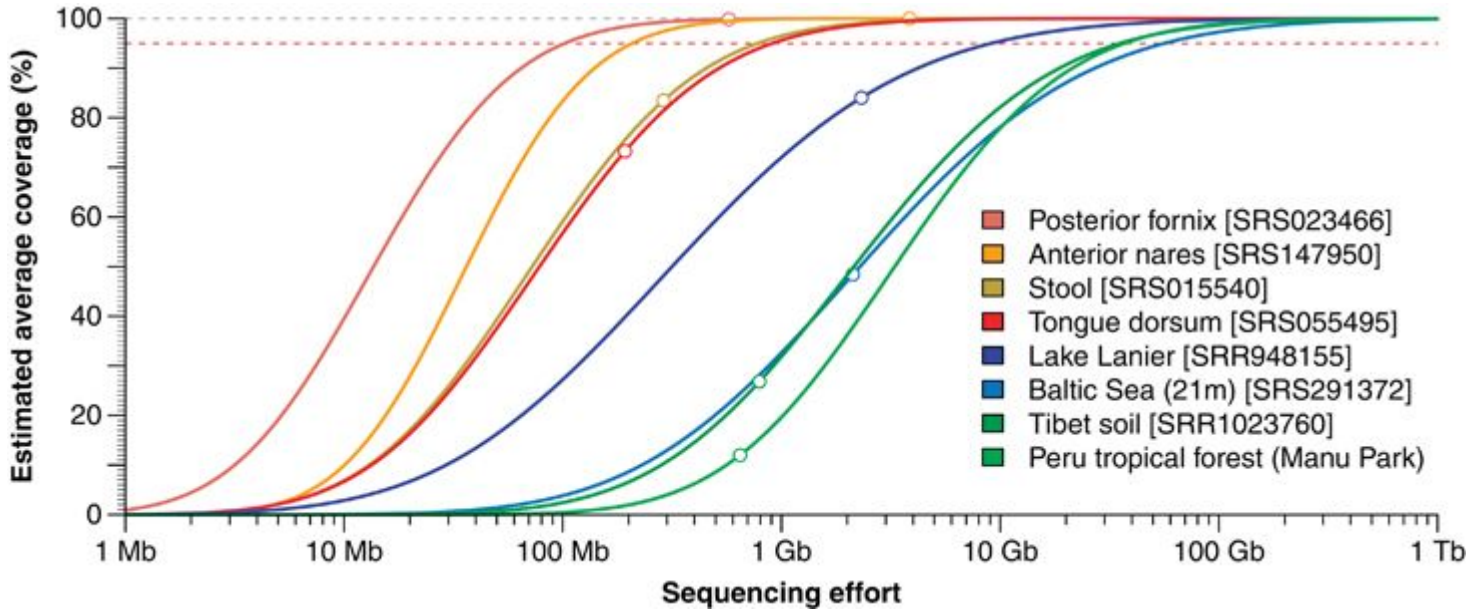
Remember your basic biology

- Some 16s rRNA studies of plants find a lot of cyanobacteria…can you guess why?
- It's ancient bacteria that are now part of the plant



6. How does one check a shotgun metagenome sample for sequencing depth?

- Nonpareil curves. Nonpareil uses the redundancy of the reads in a metagenomic dataset to estimate the average coverage and predict the amount of sequences that will be required to achieve "nearly complete coverage".



Why should we check for sequencing depth in a metagenomic study?

- We can give an honest estimate for how descriptive our dataset really is.

- We can see how much sequence is needed to describe an entire microbiome, thus avoiding over-sequencing. Great for pilot studies!