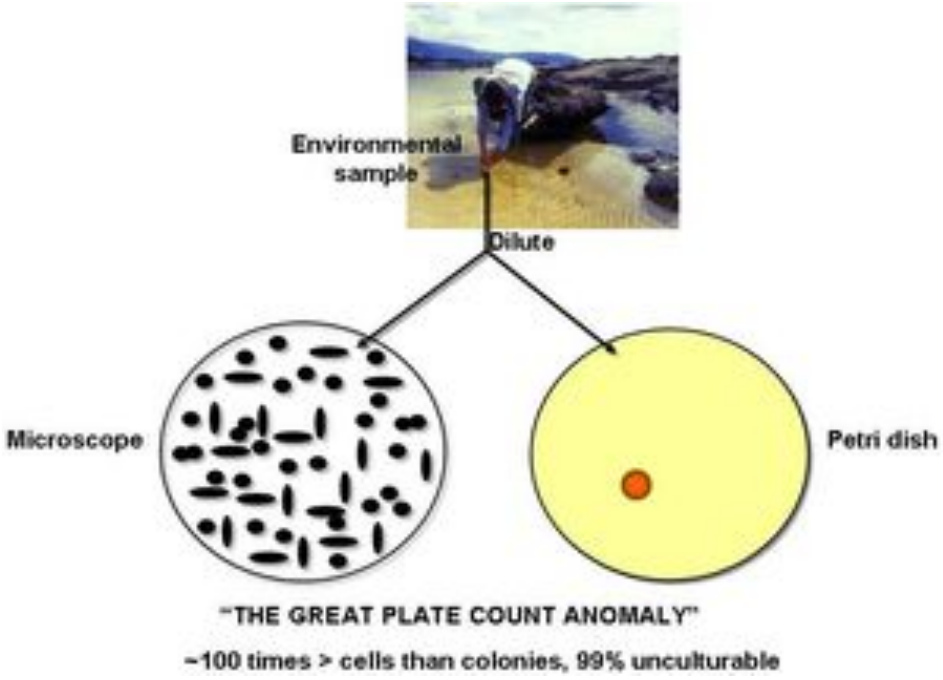
Introduction to Metagenomics

Dr. Jung Soh

657.000 SS 2017

Why metagenomics?

- "Great plate count anomaly" (Staley and Konopka, 1985)
  - The observation that most of the microbes seen in the microscope cannot currently be grown under laboratory conditions
- Almost impossible to culture all constituents of a given microbiome sample



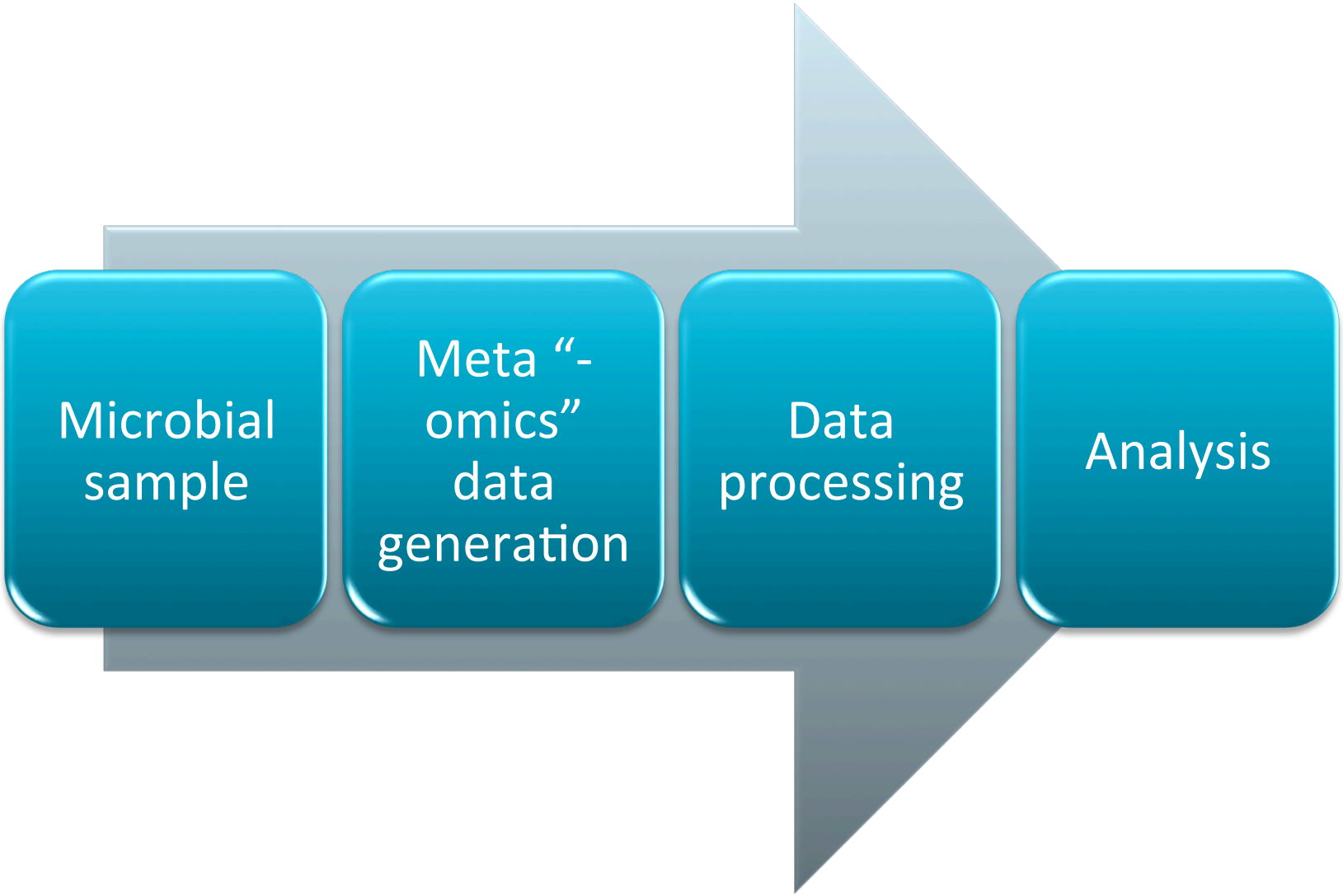
http://schaechter.asmblog.org/schaechter/2010/07/the-uncultured-bacteria.html

Staley, J.T., Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annual Review of Microbiology, 39:321–346.

Metagenomics

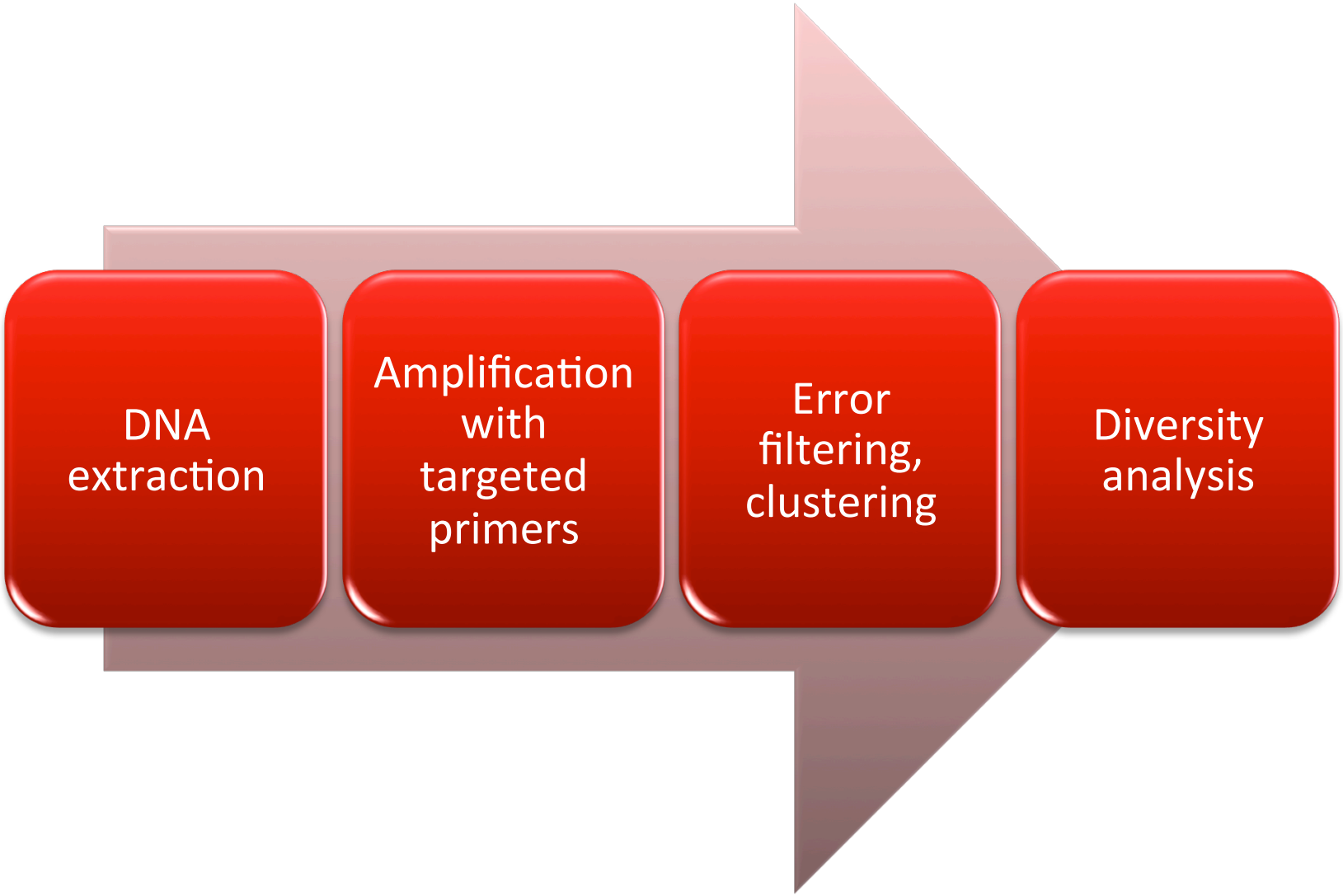
- Explore relationships between microbes and their habitats
- Combination of experimental and computational techniques
  - Marker genes
  - Metagenomes
  - Metatranscriptomes
  - Metaproteomes
  - Metametabolomes

Highest-level workflow



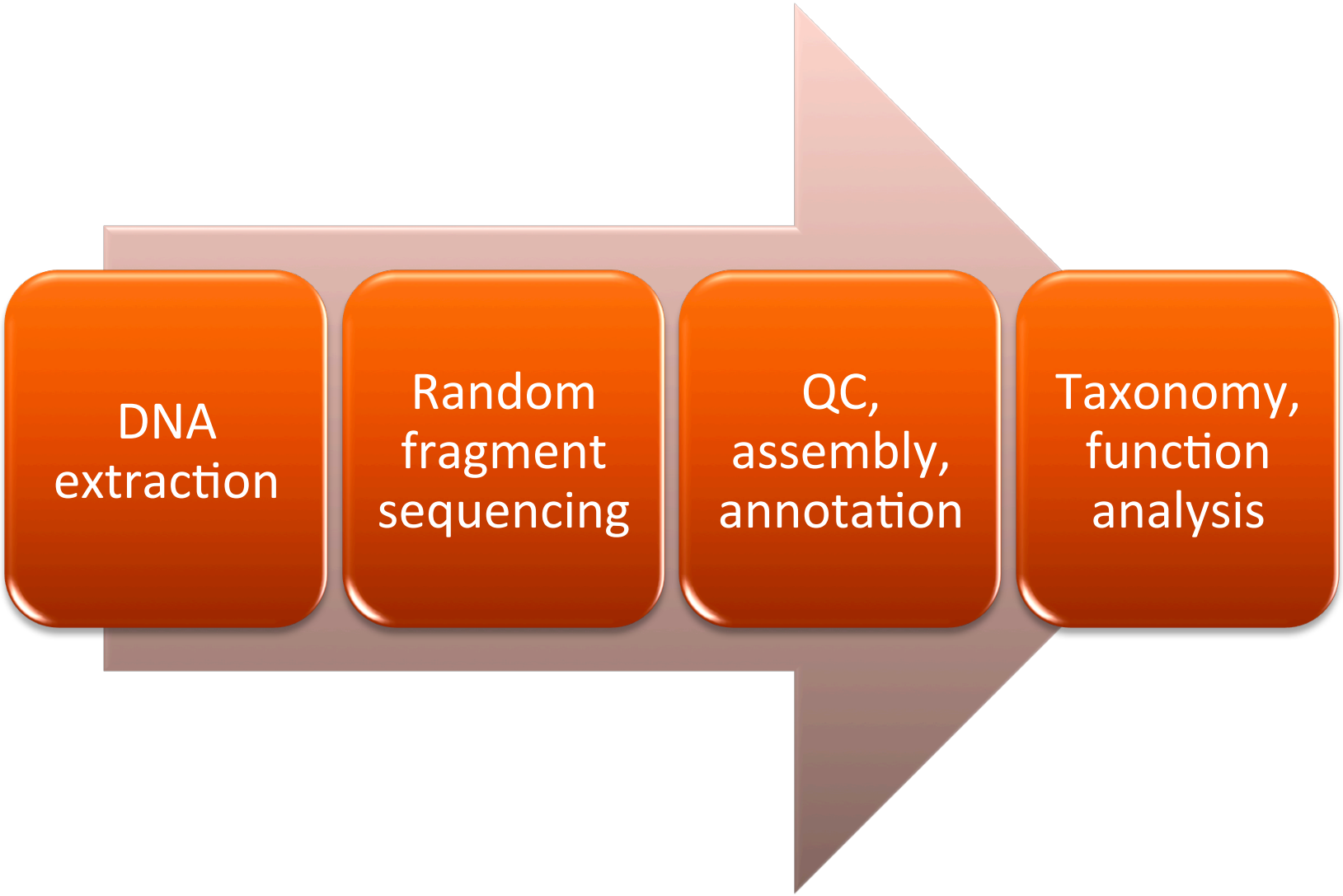
https://bioinformatics.ca/metagenomics2016module-1-ppt

Marker genes workflow



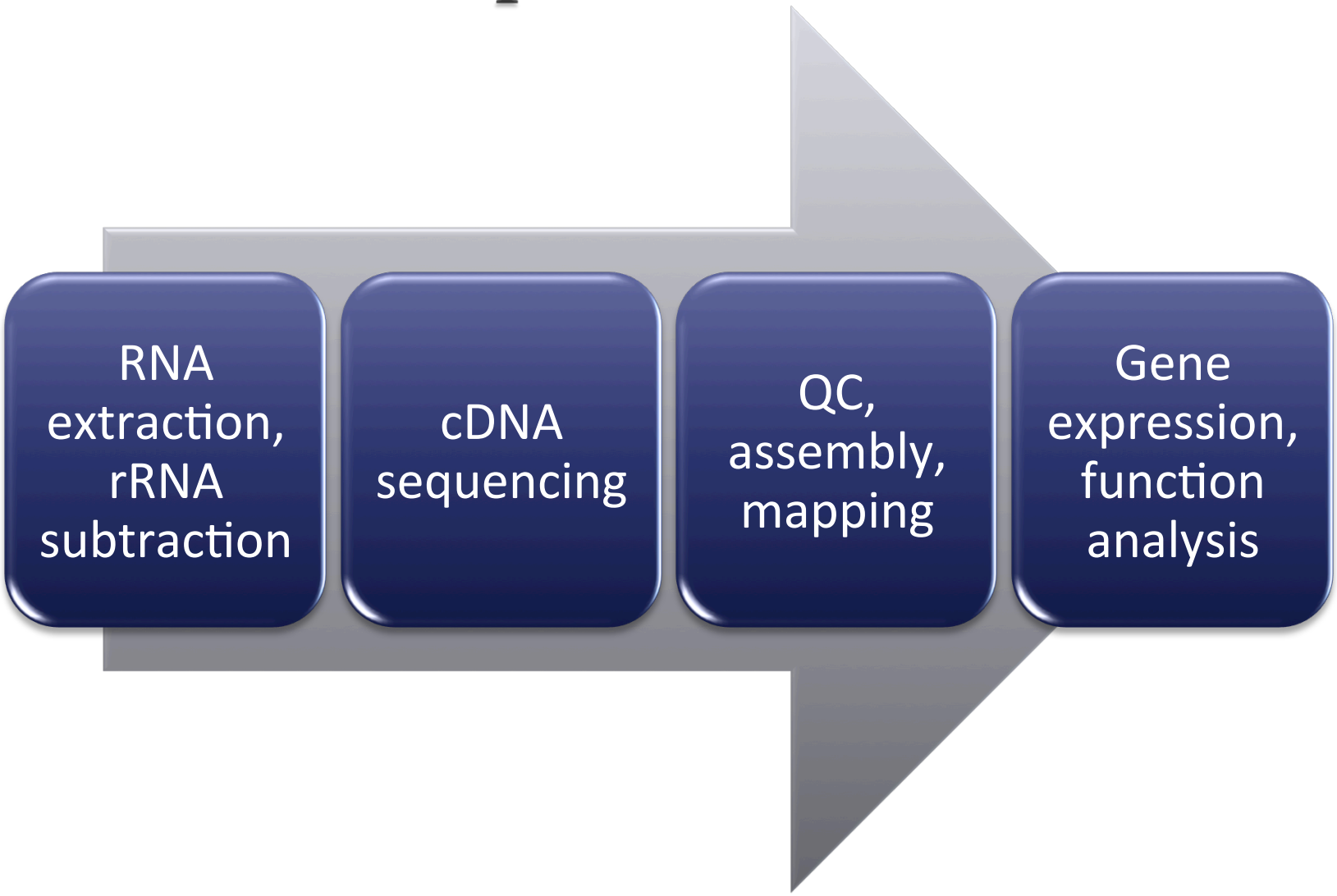
https://bioinformatics.ca/metagenomics2016module-1-ppt

Metagenomes workflow



https://bioinformatics.ca/metagenomics2016module-1-ppt

Metatranscriptomes



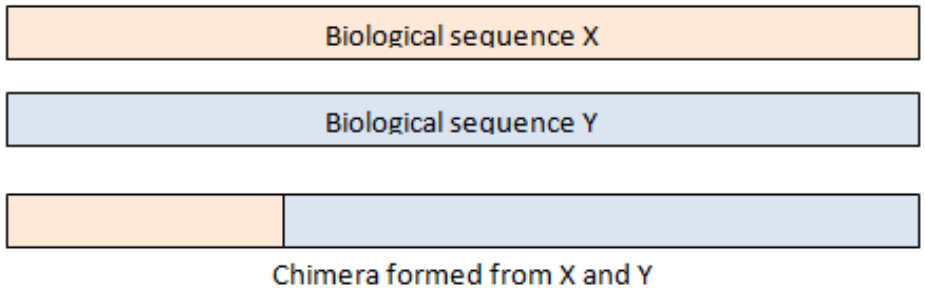
https://bioinformatics.ca/metagenomics2016module-1-ppt

Metagenomics questions

- Who's there?
  - Taxonomic classfication
- How many/much of them are there?
  - Community diversity analysis
- What do they do?
  - Functional analysis

Data quality issues

- Sequencing errors
- Chimeras
  - More common with amplicons than shotgun sequencing



drive5.com

- Metadata availability
  - Accuracy/consistency limits utility of sequencing data
  - Not available for many public datasets

Comparability issues

- Not reliably comparable across sequencing platforms
- Difference in target regions
  - Several V (variable) regions in 16S rRNA gene
- Complexity of tools and analyses
  - Ground truth is not available
  - Difficulty in evaluating tools

Taxonomy resolution issues

- Species/strain-level diversity is not usually attainable
  - Variation in sequences not distinguishable from sequencing errors
- To asssemble or not assemble?
  - Longer sequences provide more information
  - But at the cost of possible introduction of errors
    - Chimeric contigs created by assembly

Functional analysis issues

- Functions not assigned for majority of reads/contigs
  - Existing databases are from single organisms
- Functions assigned not specific enough
  - Only assigned to the top level
    - Amino acids and derivatives
    - Protein metabolism

Two types of metagenomics

- Marker gene metagenomics
  - Use one or more marker genes
  - Less resources
  - Mostly bacteria can be identified
  - No functional analysis

- Whole genome metagenomics
  - Sequence whole metagenomes
  - No target primer bias
  - More resources
  - All microbes can be identified
  - Functional analysis possible