Day 6: Metabolic pathway analysis

MMB-114

Schedule

Day 1: Basics of UNIX and working with the command line
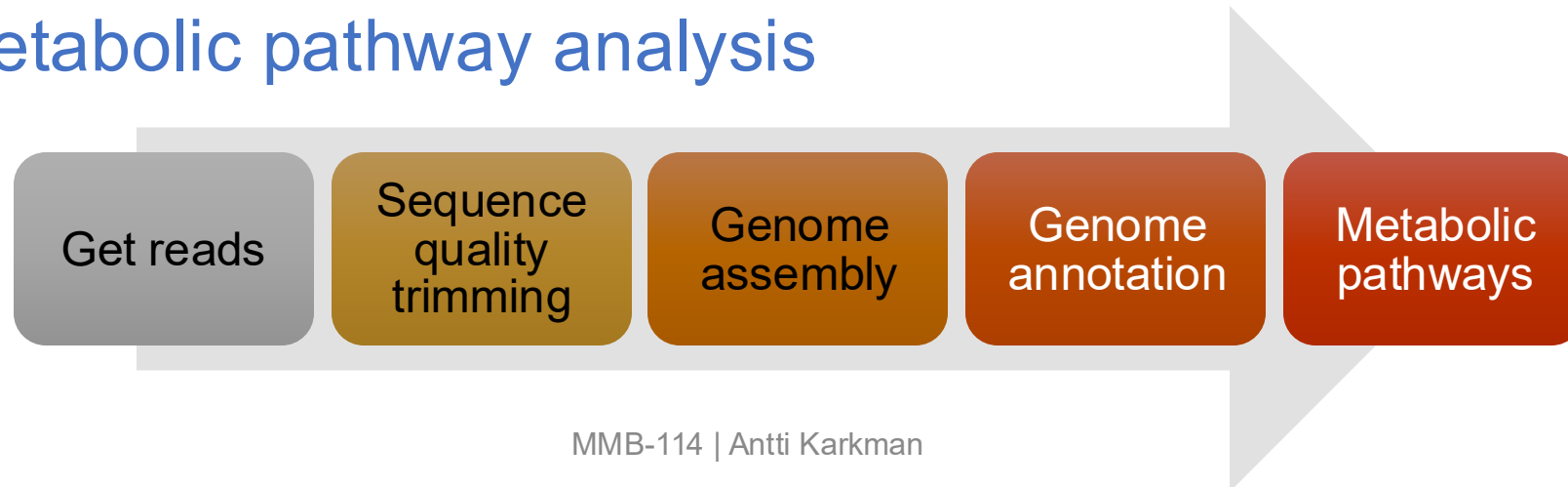
Day 2: Handling of Nanopore/Illumina data

Day 3: Check-up

Day 4: Genome assembly

Day 5: Genome annotation

Day 6: Metabolic pathway analysis

Get reads → Sequence quality trimming → Genome assembly → Genome annotation → Metabolic pathways

Metabolic pathways

- Biochemistry meets molecular biology
- Series of linked chemical reactions occuring within a cell
- Metabolism:
  - Catabolism: The processes by which a living organism obtains energy and raw materials from nutrients
  - Anabolism: The processes bu which energy and raw meterials are used to build macromolecules and cellular structures (biosynthesis)
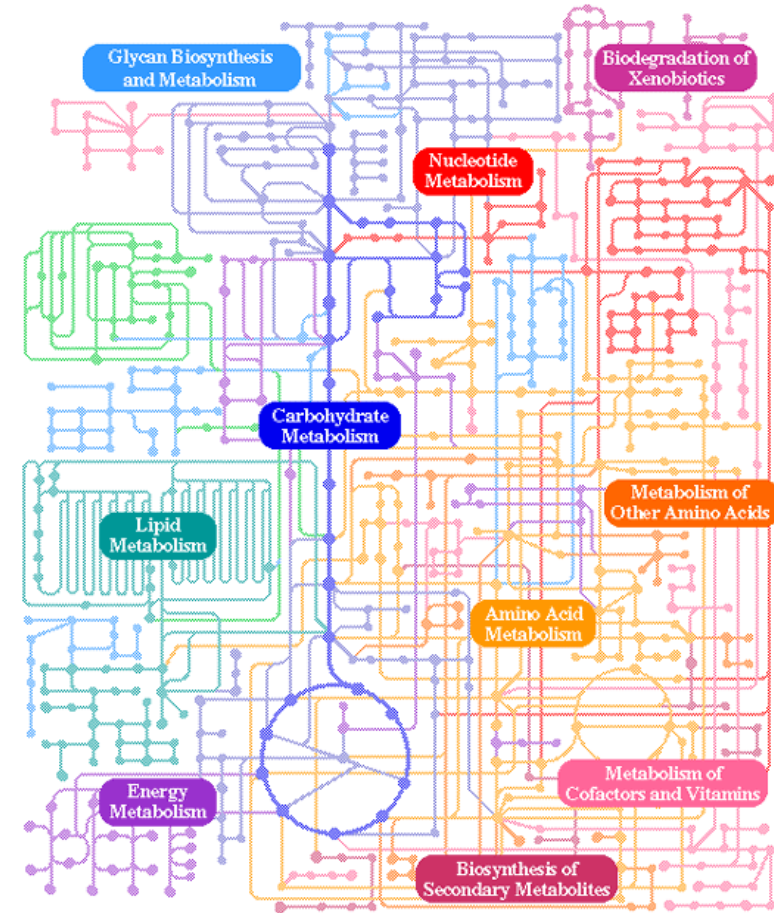
Gene databases

KEGG	Collection of databases dealing with genomes, biological pathways, diseases, drugs and chemical substances
UniProt	Aggregate of two databases: SwissProt with functional annotations obtained from the literature and subjected to human review and TrEMBL with functional annotations computationally assigned
Pfam	Curated database of protein families
Interpro	Curated database of protein families
Metacyc	Highly curated metabolic database that contains metabolic pathways, enzymes, metabolites, and reactions from all domains of life
GO	The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism. Three structured, controlled vocabularies (ontologies): biological processes, cellular components and molecular functions
SEED	A comparative genomics environment consisting of databases of protein families (FIGfam) and metabolic pathways (Subsystems)

KEGG: Kyoto Encyclopedia of Genes and Genomes

http://www.genome.jp/kegg

- Collection of databases dealing with genomes, biological pathways, diseases, drugs and chemical substances

- KEGG PATHWAYS: collection of manually drawn pathway maps representing our knowledge on the molecular interaction, reaction and relation networks

- KEGG MODULES: collection of manually defined functional units used for annotation and biological interpretation of sequenced genomes

antiSMASH

https://antismash.secondarymetabolites.org/

- Web server for **identification**, **annotation** and **analyis** of secondary **metabolite synthesis gene clusters in** bacterial and fungal genomes

- Glossary:
  - BGCs: Biosynthetic Gene Clusters
  - NPs: Natural Products (secondary metabolites)

- Full documentation at: https://docs.antismash.secondarymetabolites.org/

Alternative workflows for bacterial genomes

Kbase web platform

- Graphical user interface (GUI)

- Free and "easy" to use

- Workflows are called "Narratives", an interactive notebook of the analyses

- Easy to share data, workflows and results

- Limited number of "apps"

Bactopia

- "Flexible pipeline for complete analysis of bacterial genomes"

- Runs locally, no web interface

- No GUI, works on command line

SeqHub

- "Our mission is to make biological sequence data easy to find, understand, and share."

- Annotation server

- Free for academics

Genome announcement format

See examples at Microbiology Resource Announcements:

https://journals.asm.org/journal/mra

Or one of our recent announcements:

https://doi.org/10.1128/mra.00120-25

Let's see what your strain is capable of

Look for pathways of interest

How does the strain

- Gets energy
- Gets carbon and nitrogen
- Survives in stress
- Move around

https://github.com/karkman/MMB-114_Genomics

(**Day 6:** Metabolic pathway analysis)