Genome assembly and annotation

Day 6: Metabolic pathway analysis

Antti Karkman

Department of Microbiology – UH

antti.karkman@helsinki.fi

Slides modified from Igor Pessi, with permission

Aims for this part of MMB-114

Day 1: Basics of UNIX and working with the command line

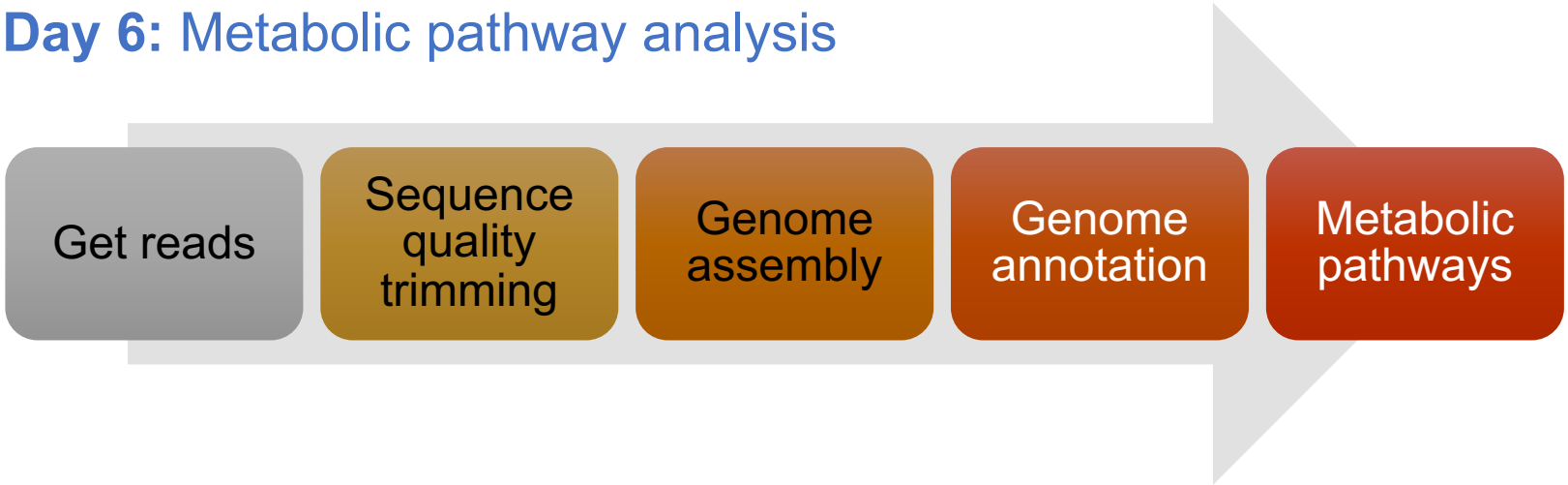
Day 2: Handling of Illumina data

Day 3: Genome assembly

Day 4: Check-up and report

Day 5: Genome annotation

Day 6: Metabolic pathway analysis



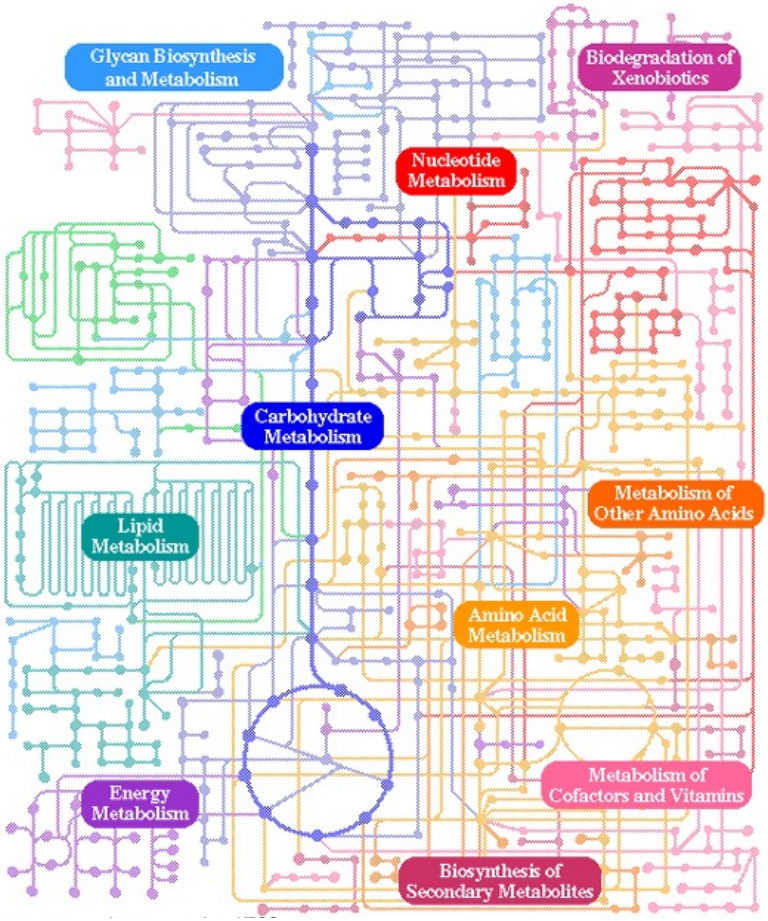
Metabolic pathways

Biochemistry meets molecular biology

Series of linked chemical reactions occurring within a cell

Metabolism

- **Catabolism:** The processes by which a living organism obtains its energy and raw materials from nutrients
- **Anabolism:** The processes by which energy and raw materials 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

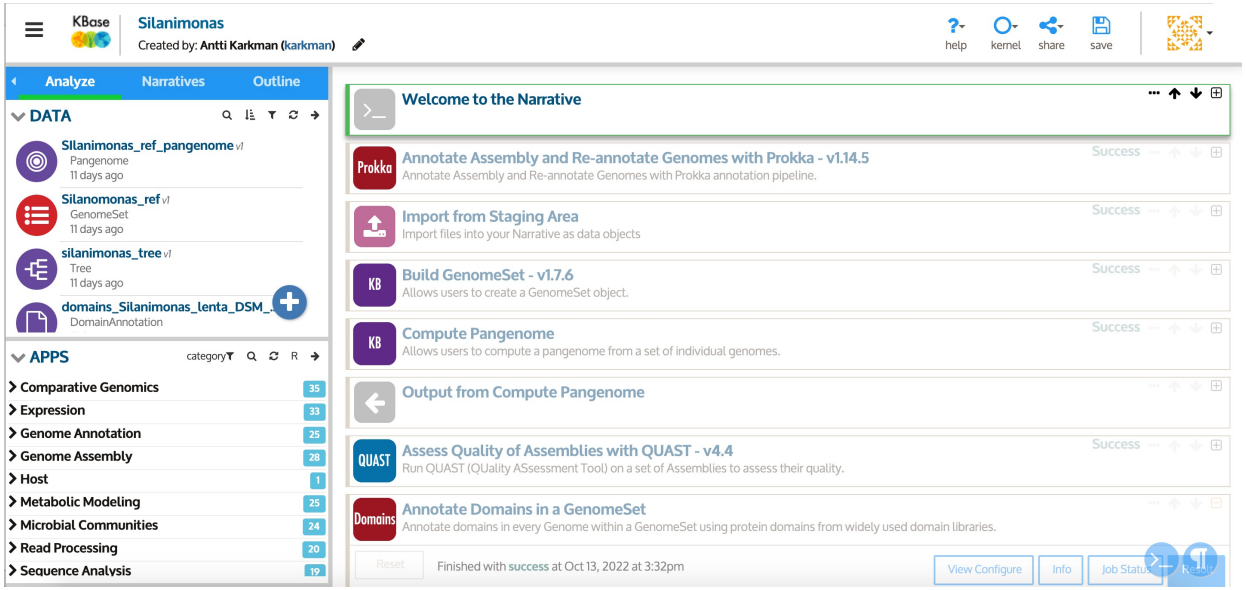
http://www.genome.jp/kegg

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

Kbase web platform



KBase
PREDICTIVE BIOLOGY

https://www.kbase.us/

- 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”

Let's see what our 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)