

MIMM4750G

What is bioinformatics?



Course syllabus

- Lectures and practicals
- Evaluation
 - In-class assignments (20%)
 - Lab assignments (40%)
 - Proposal (abstract) (10%)
 - Short paper (30%)

What is bioinformatics?

- New experimental equipment generating enormous amounts of complex data.
- You need computing skills to function in a modern biological lab.
- Bioinformatics is the computational management and analysis of biological data.

A spectrum of bioinformatics

Type		Usage	Understands methods/models
Casual end-user		Requires a push-button interface. Applying a defined analysis pipeline to their data.	None
Informed end-user		Can install programs with package manager, sometimes from source. Can run a given pipeline and change settings to fit their data/hypothesis.	Basic
Power user		Can resolve dependency issues when compiling programs from source. Can customize a pipeline for new uses.	Comprehensive
Developer		Builds new pipelines. Implements/adapts existing methods into new programs.	Expert
Computer scientist		Creates new methods and algorithms.	Specialist

Unlimited intersections in bioinformatics

Biology

Gene expression

Biological networks

Evolution

Protein structure

Systems biology

Microbiology

Immunology

Cancer biology

...

Computing

Database programming

Computational statistics

Dynamical systems modeling

Artificial intelligence

Text mining

Optimization

Simulation

Visualization

...

New technologies

- Whole genome sequencing (WGS), GWAS
- Next-generation sequencing (deep sequencing, barcoding, metagenomics, RNA-Seq, ChIP-Seq)
- single-cell transcriptomics
- live-cell imaging data analysis
- remote sensing (drones, satellites, swarm sensing)
- Non-relational, distributed databases ("big data")

The command line

- It is impossible to create a "Microsoft Excel" for bioinformatics
- Inevitably you will ask a question that there is no button for.
- Lab technologies move fast.
- No one has time to make nice interfaces.
- Bioinformatics requires "command line fluency", which generally means working with UNIX-like operating systems.
- The most current methods are generally available as open-source software.

Phylogenetics

Systems biology

System

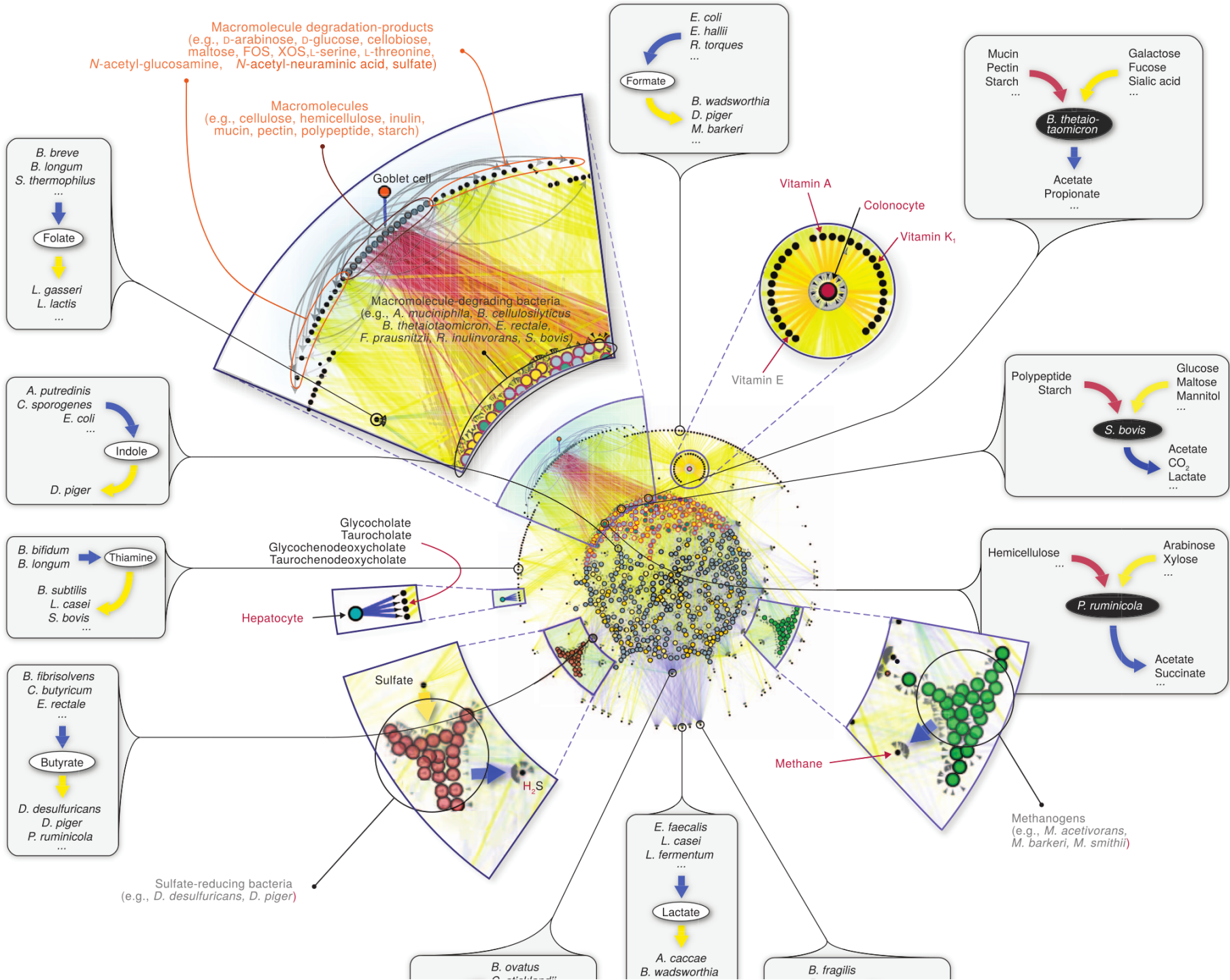


Image processing

Architectural transitions in *Vibrio cholerae* biofilms at single-cell resolution

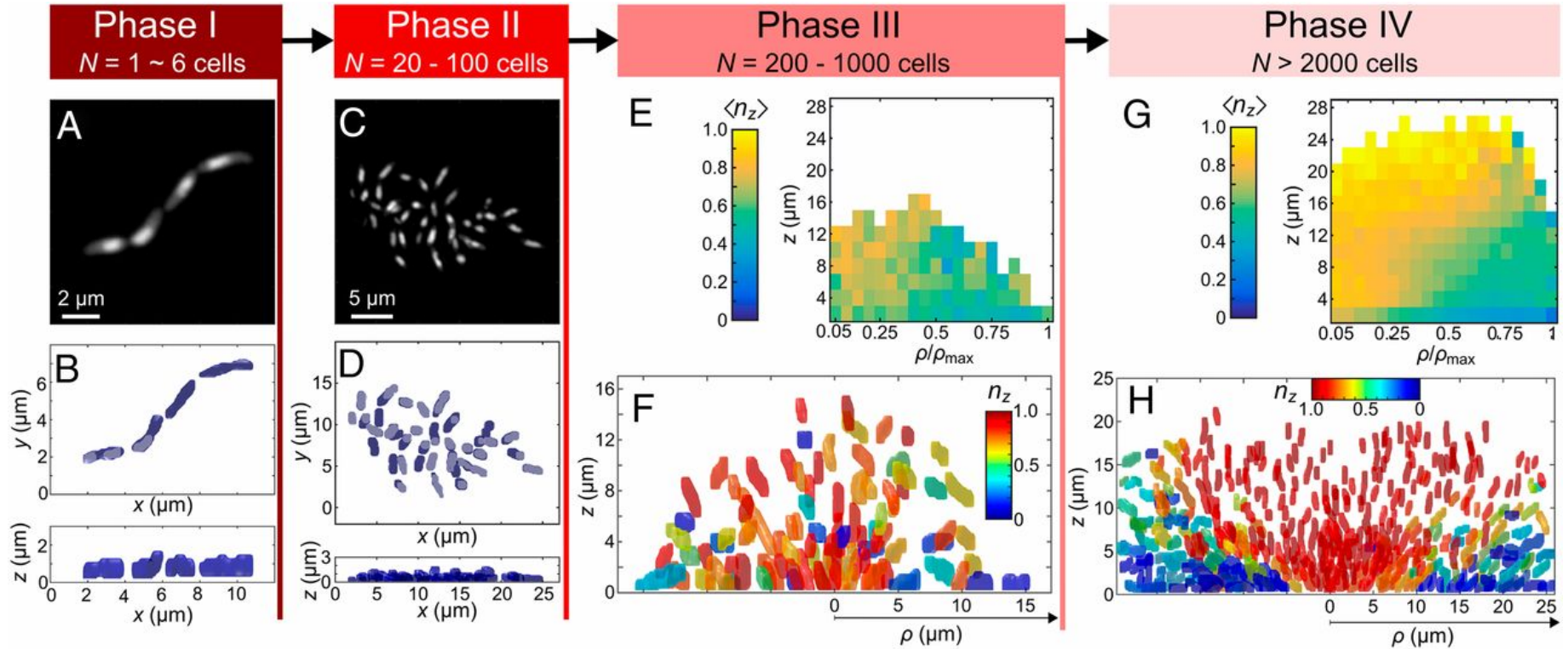


Figure 14, Drescher *et al.* 2016, Proc Natl Acad Sci 113.

Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host-pathogen transcriptomes

