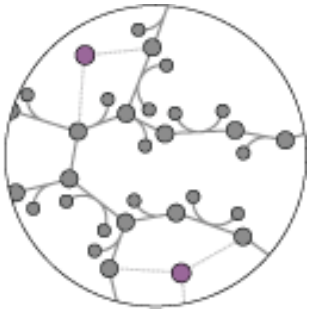


UCSD Systems Biology Research Group

Tutorial on Flux Balance Analysis and Genome-scale Modelling

Jonathan Monk
Yara Seif
CJ Norsigian

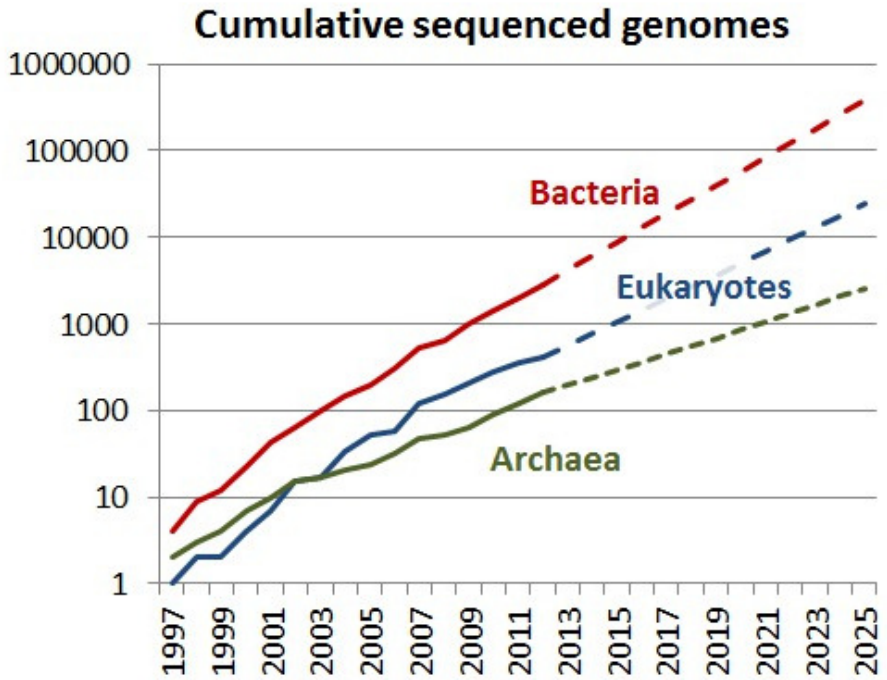
U01 AMR Modelling Working Group
September 26, 2017

Outline

- Intro to flux balance analysis (FBA) and genome-scale modelling
- Overview of FBA tools (COBRA and BiGG)
- Jupyter notebook live tutorials
  - 1. Reconstruction example
  - 2. Simple model simulations
  - 3. Application: gene essentiality and growth capabilities on different substrates
- Discussion
- Brief update of *C. diff* reconstruction progress and data collection efforts

Problem: Big Data

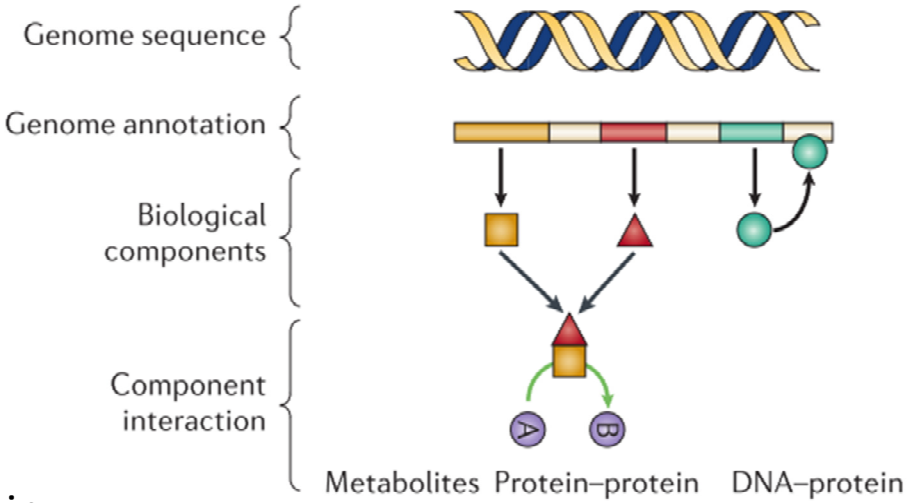
- Rapid advances in DNA sequencing technology
  - Exponential increase in data generation
- Sequencing “data deluge”
- Human Genome Project
  - 2003: Cost $2.7 billion
  - Today: $1000
- The exponential trend is expected to continue
- Huge opportunity
  - Sequences span the phylogenetic Tree of Life



- Genomes Online Database (GOLD)

One tool: network reconstructions

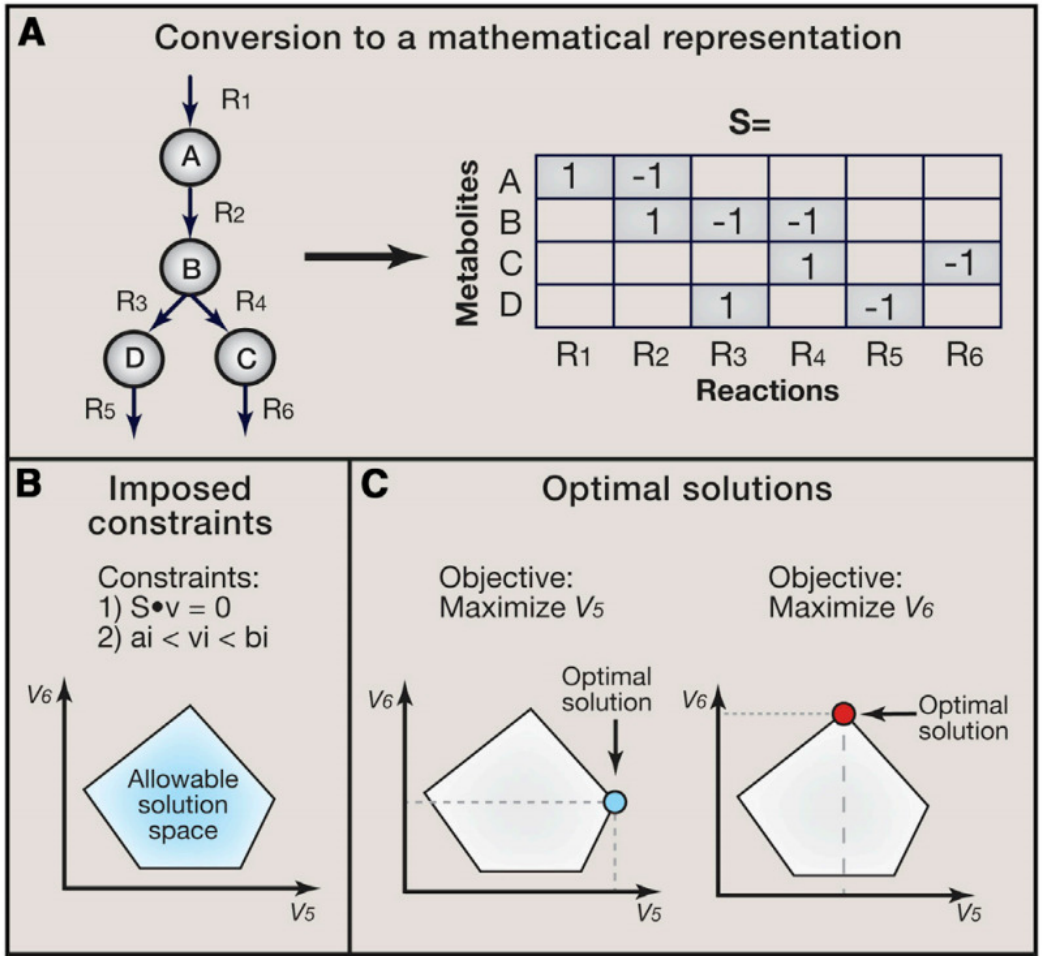
- Data is collected from many sources:
  - Genome annotation
  - Literature
  - Experiments
- Defines biological components
  - Assembled into a structured representation
  - Links components to each other via known interactions
- Example: Metabolic reconstruction:
  - Genes linked to their encoded protein
  - Proteins linked to their catalyzed reaction
  - Reactions linked by metabolites



Reed JL, Famili I, Thiele I, & Palsson BO (2006) **Towards multidimensional genome annotation.** *Nat Rev Genet* 7(2):130-141.

Constraints Based Modelling

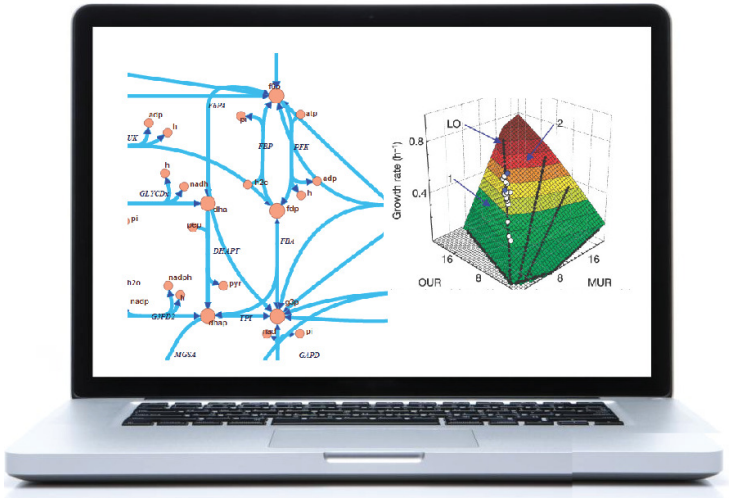
1. Reconstruct network
2. Convert to mathematical representation
3. Apply constraints
   1. Mass balance
   2. Energy balance
- 4. Define an objective:
  - For example: growth
- 5. Use linear optimization to solve

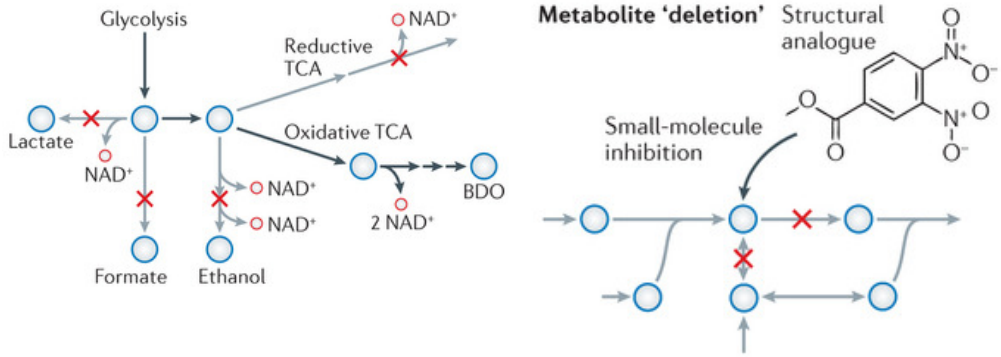


O'Brien EJ, Monk JM, Palsson BO: **Using Genome-scale Models to Predict Biological Capabilities**. *Cell* 2015, **161**(5):971-987

Utility of this Application

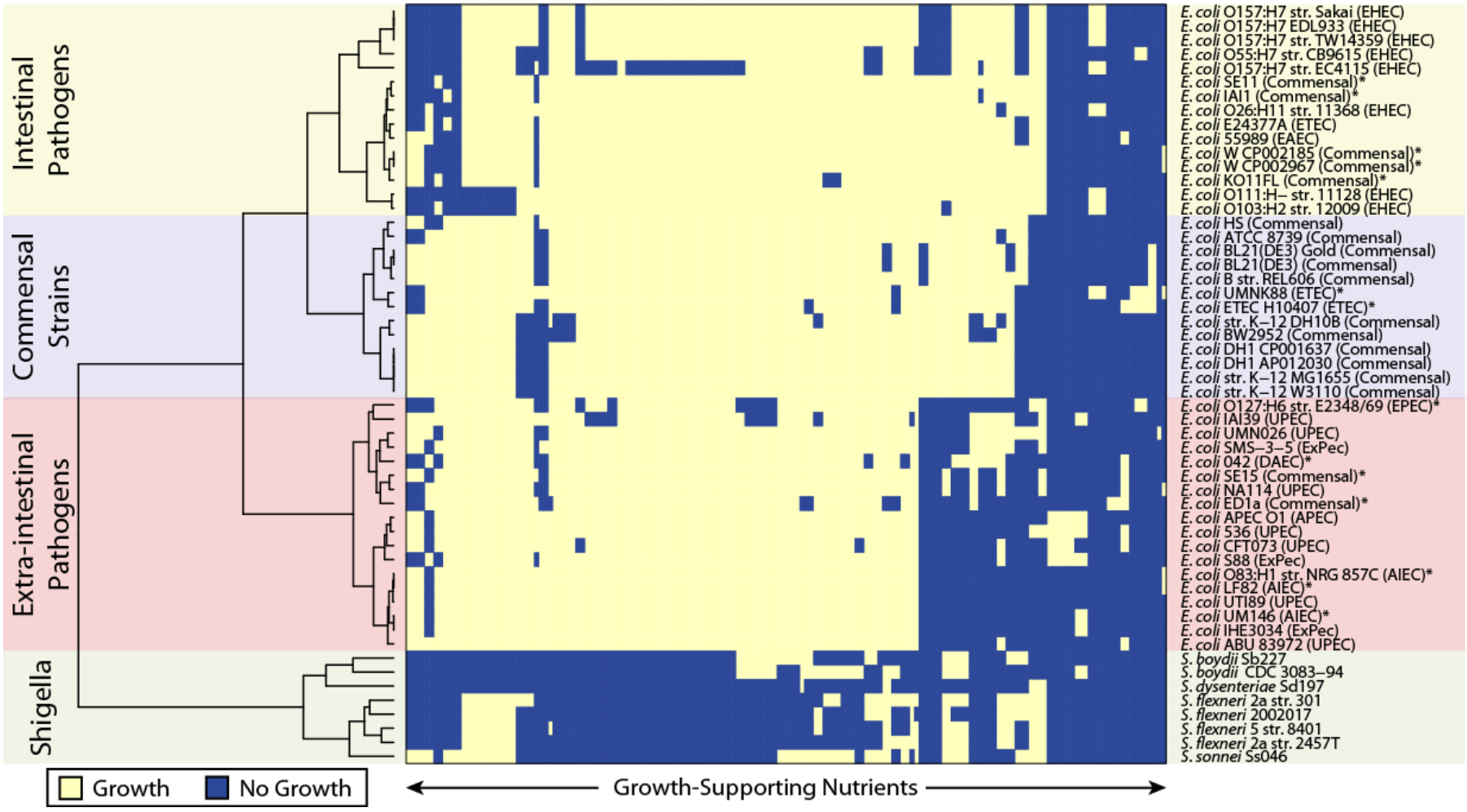
- Allows for simulation of the behavior of a cell in a computer
  - In different growth conditions
  - genetic perturbations
- Applications in:
  - Industrial biology
    - Metabolic Engineering
  - Medicine
    - Drug target discovery





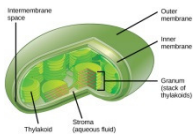
Bordbar A, Monk JM, King ZA, Palsson BO: **Constraint-based models predict metabolic and associated cellular functions**. *Nat Rev Genet* 2014, **15**(2):107-120.

Comparing growth capabilities for strains in a species (e.g. *E. coli*)

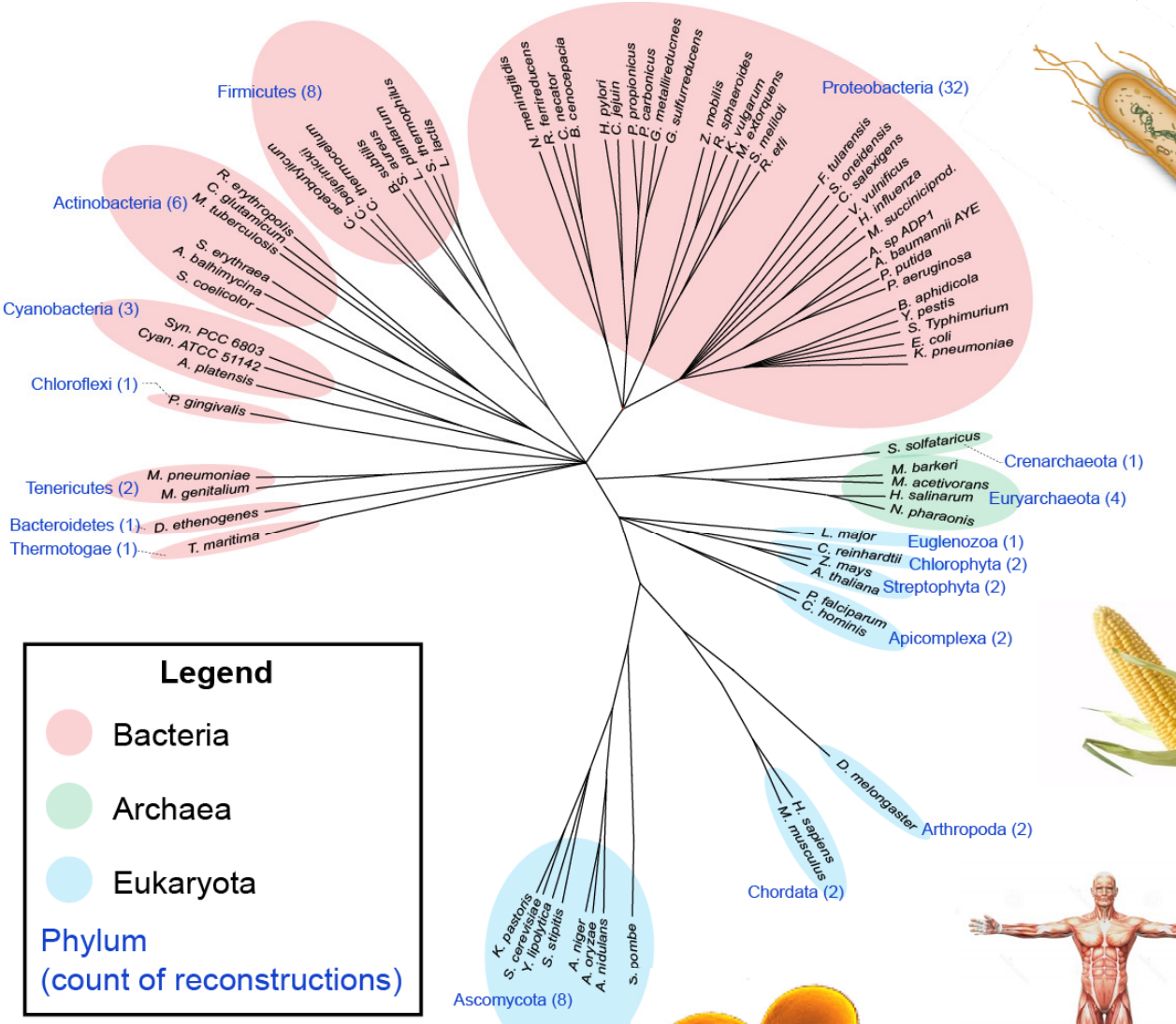


Monk, J. M. *et al.* **Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments**. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 20338–43 (2013).

Reconstructions exist for species across the tree of life

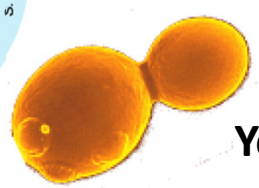


Algae



E. coli

Corn



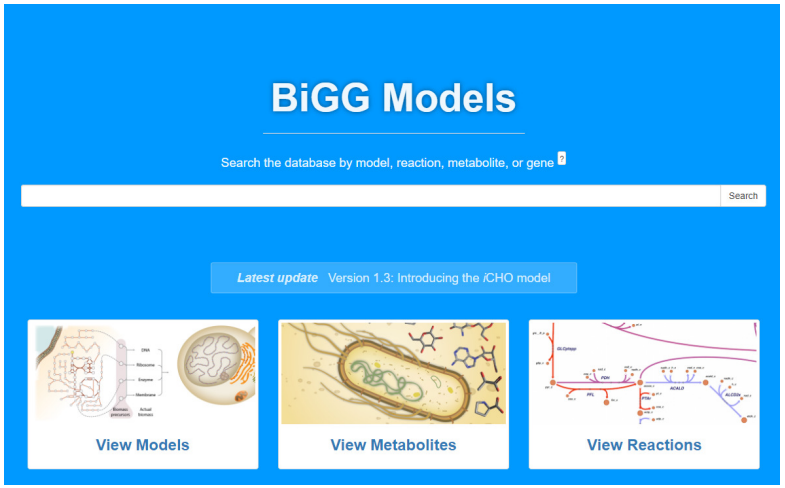
Yeast

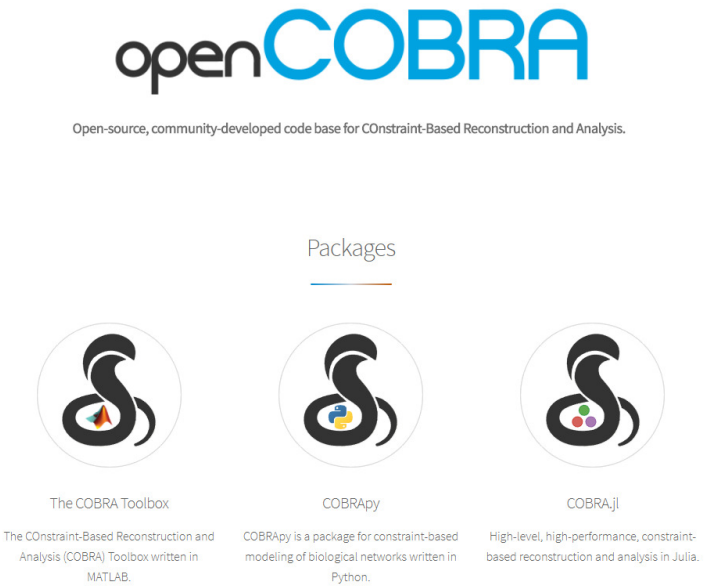
Human

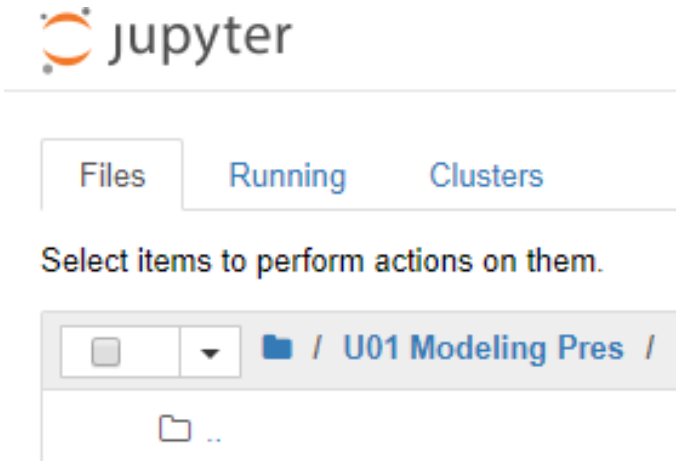
Monk, J., Nogales, J. & Palsson, B. O. **Optimizing genome-scale network reconstructions**. *Nat. Biotechnol.* **32,** 1–7 (2014).

Resources and tools

- BiGG: Repository for genome-scale models
  - Total of 80 available
  - http://bigg.ucsd.edu
- COBRA toolbox:
  - https://opencobra.github.io/
  - Today focusing on python version







Jupyter notebook live tutorial

1. Reconstruction example
2. Simple model simulations
3. Application: gene essentiality and growth capabilities on different substrates

Discussion

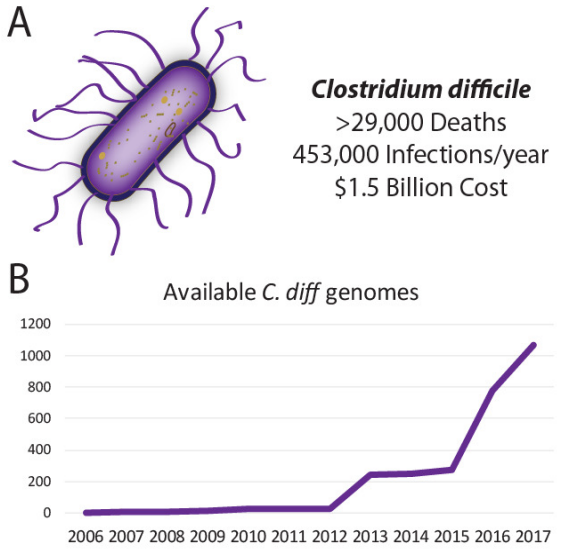
- Are jupyter notebooks a good format for sharing scripts?
- Do other groups want to contribute modelling tutorials?
- Should these be hosted on a dedicated server?
- Future presentations focused on using data generated from the U01 to model with?

Update on *C. diff* reconstruction

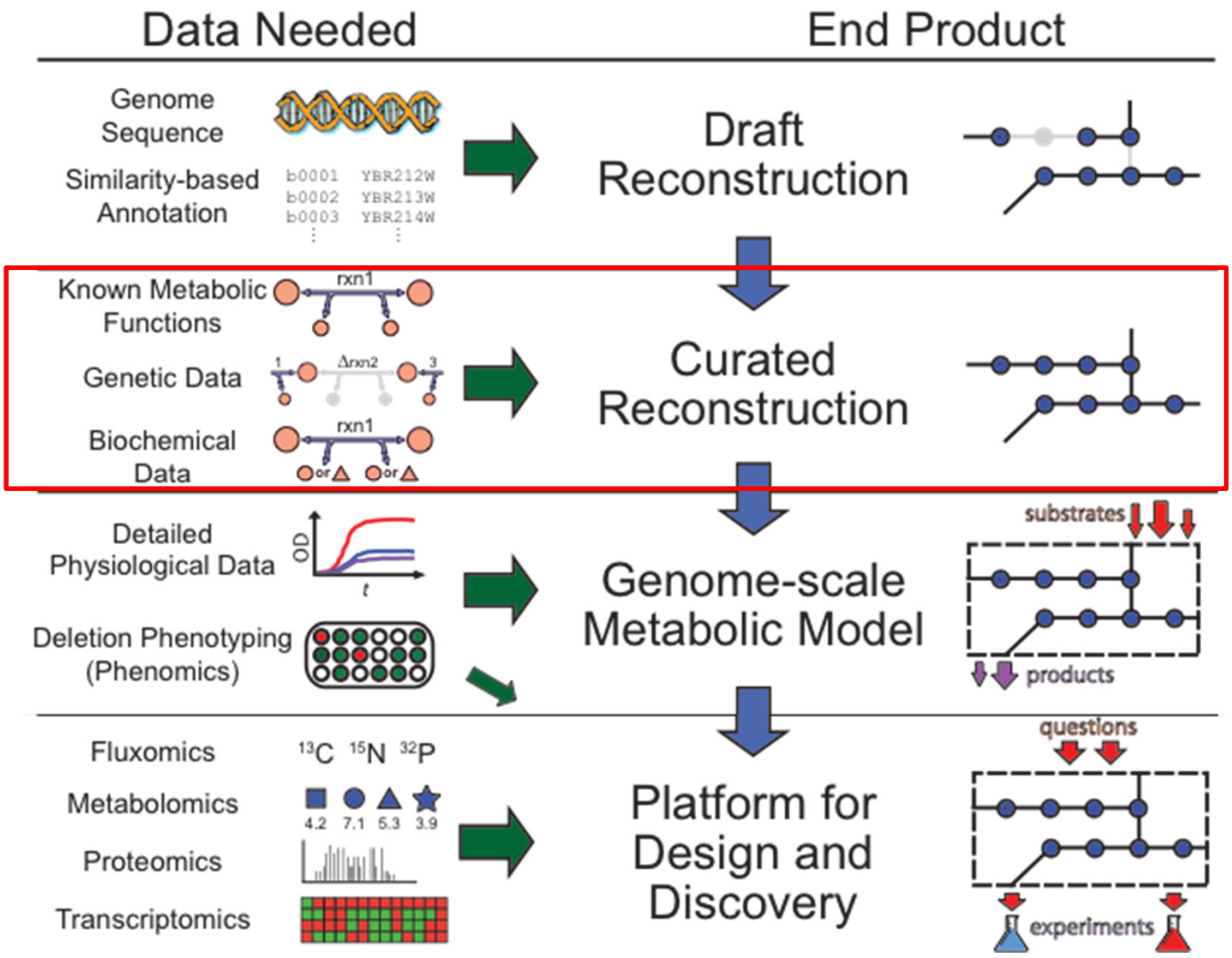
- Building draft reconstruction for *C. diff* 630 based on previous reconstruction
  - Kashari et al. 2017
- Mapping and standardizing reaction and metabolite IDs to SBRG standard
- Collecting data and genome sequences for model expansion and improvement
  - Baylor proteomics data sets (Tor Savidge)
  - Additional sequences?
  - Biolog data?
  - Others available?

Making life difficult for Clostridium difficile: augmenting the pathogen's metabolic model with transcriptomic and codon usage data for better therapeutic target characterization

Sara Saheb Kashaf[1]*, Claudio Angione[2] and Pietro Lió[1]



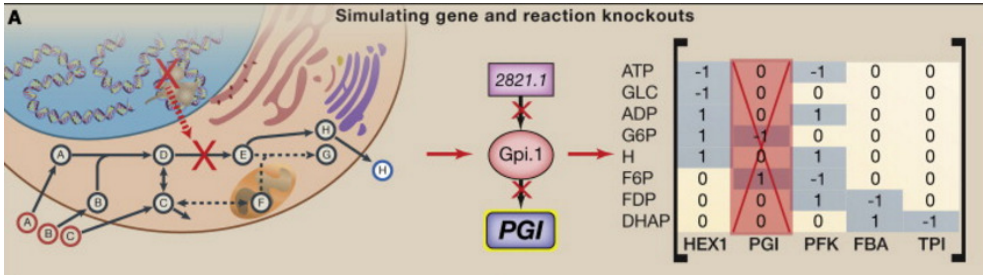
Model building pipeline

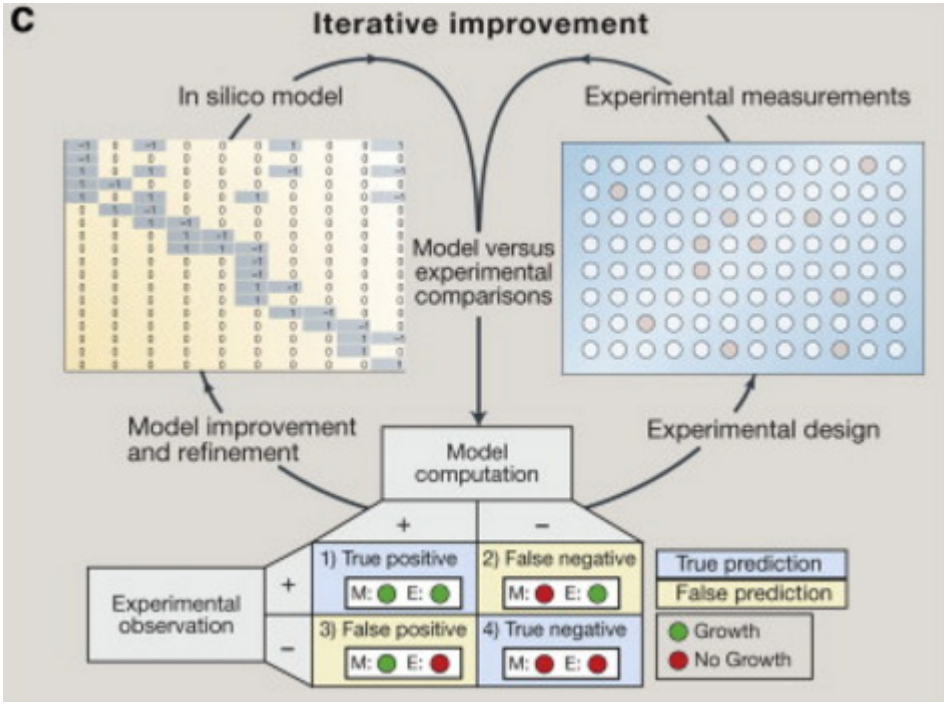


Nature Reviews Microbiology, Feb 2009

Model improvement

- Models can always be improved and expanded
- Models link genotype to phenotype
- Experimental validation identifies model errors
  - Can be used to iteratively update the model





O'Brien EJ, Monk JM, Palsson BO: **Using Genome-scale Models to Predict Biological Capabilities**. *Cell* 2015, **161**(5):971-987

Omics data types requested for *C. difficile* Genome-scale network reconstruction

Date: July 25, 2017

1. Phenomics:
    a. Genome-scale models can predict the effect of different inputs (i.e. media types) on a strain's phenotype. Thus, growth experiments in various media types are helpful build and validate a model. One of the best datasets for this purpose is Biolog phenotype microarrays.
    b. The effect of gene knockouts can also be readily evaluated by a genome-scale model. For this reason gene knockout experiments (or TNseq with estimated fitness measurements) are also very valuable for model building. Keep in mind that the KO experiments should be performed in media that is as fully defined as possible
2. Transcriptomics and proteomics:
    a. Transcriptomics (e.g. RNAseq) and proteomics data can be used to validate and test the accuracy of genome-scale models. Various treatments can be applied to the cells (e.g. different antibiotics, etc) however it is important to perform the experiments in as close to fully defined media types as possible.
3. Metabolomics:
    a. Quantitative metabolomics (e.g. measurements with internal standards as a references) is preferred for integration with genome-scale models. Both endo- and exo-metabolomic data can be integrated with genome-scale models.
4. Genomics:
    a. Finally, to construct strain-specific models of *C. difficile* metabolism we will require genomic sequences of several different *C. diff* strains. Strains with data associated with them (such as that listed above) will be most valuable.

We hope that U01 centers studying *C. difficile* antibiotic resistance will help us in collecting these data types to build consensus metabolic network reconstructions of *C. difficile* metabolism.