



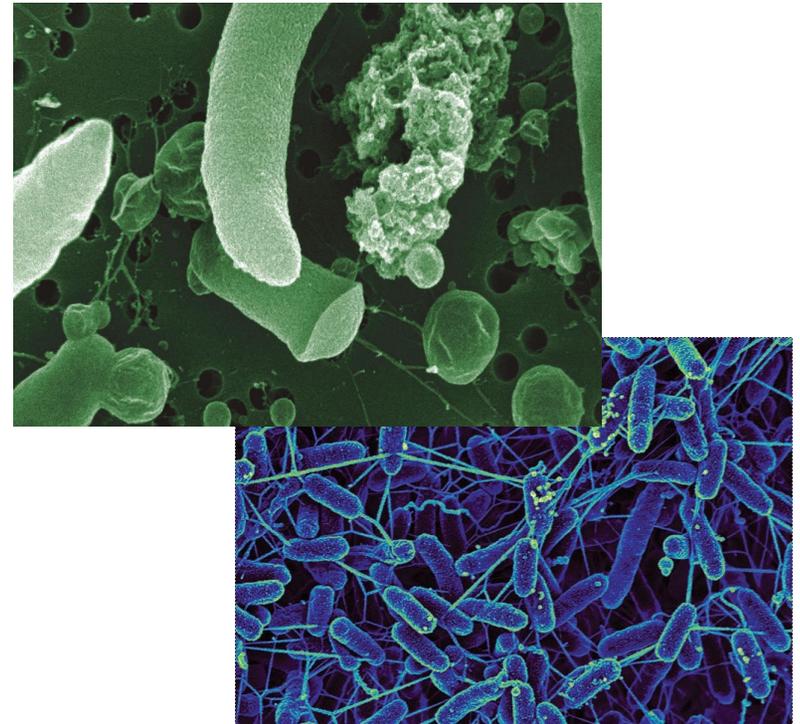
*DOE Systems Biology Knowledgebase*

**KBBase**

Integration and modeling  
for predictive biology

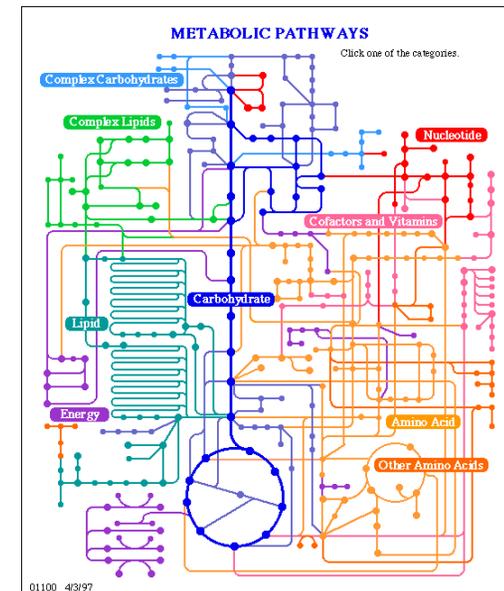
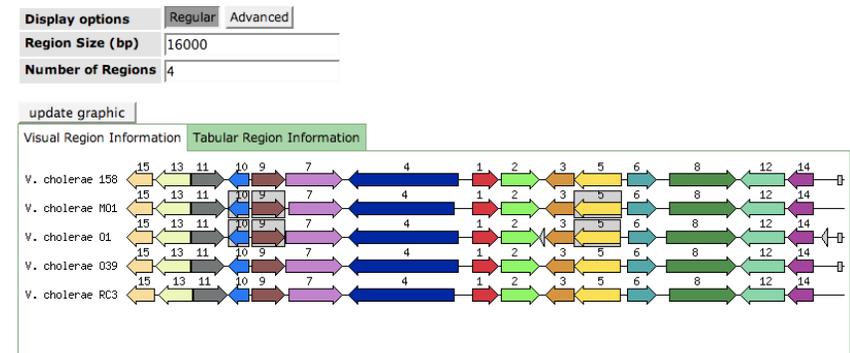
## Our overall goals are to:

- Reconstruct and predict metabolic and gene expression regulatory networks to manipulate microbial function
- Vastly increase the capability of the scientific community to communicate and utilize their existing data
- Enable the planning of effective experiments and maximizing understanding of microbial system function



## We propose to do this by:

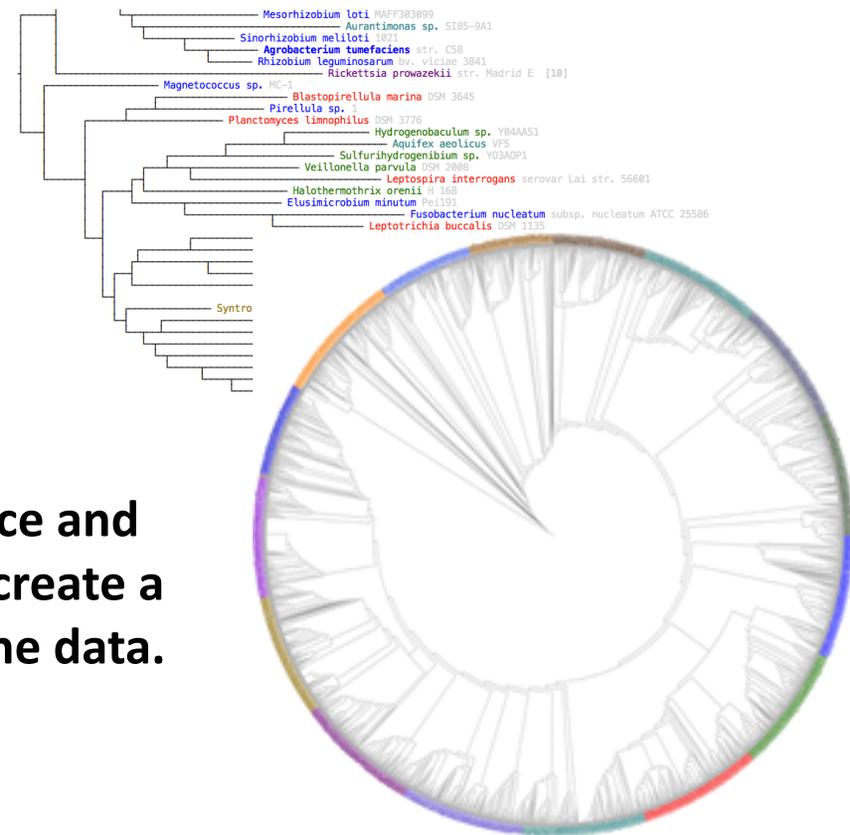
- Annotating genomes and assigning confidence
- Reconstructing metabolism and optimizing for function
- Reconstructing regulation and assessing agreement with expression data
- Integrating and standardizing -omics data from multiple data sources
- Constructing models of microbial organisms and interlinking models with data



**Within 13 months, we will be able to demonstrate use of the following:**

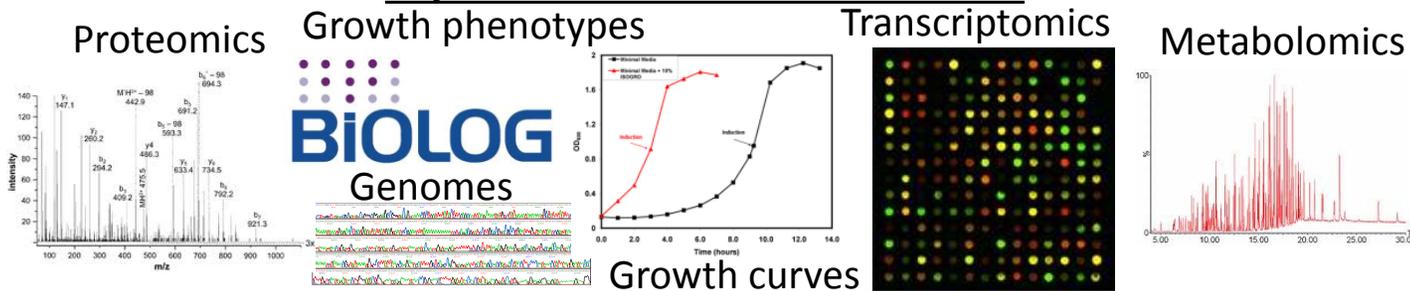
- Data integration and data model
- Next generation organism pages
- Phylogenetic tree services
- Next generation gene pages
- Metabolic modeling
- Regulatory/Transcriptional networks

**A microbiologist with a genome sequence and phenotypic growth data will be able to create a metabolic model fully reconciled with the data.**





## Experimental and Genomic Data



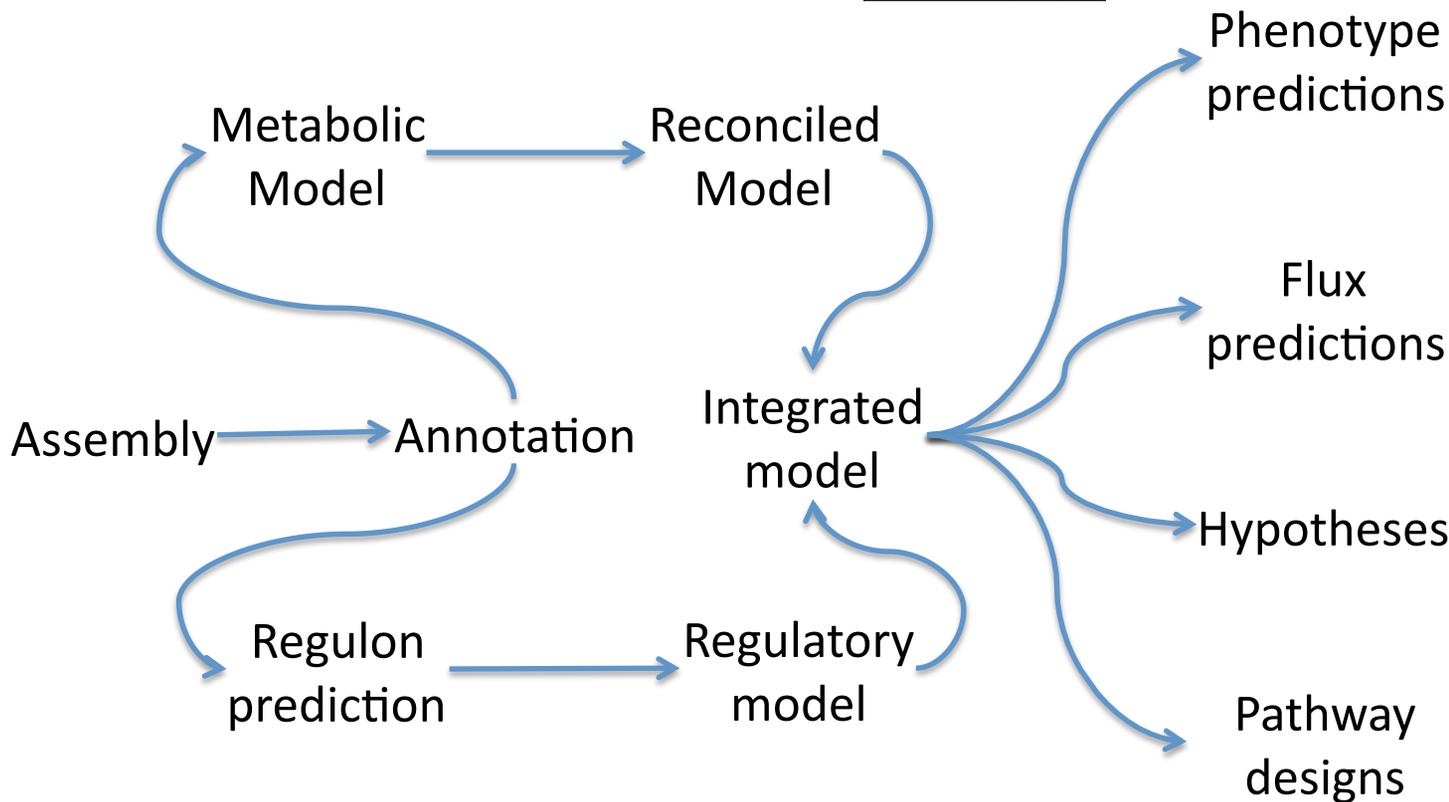
### Use experiences

### Web API

### IRIS



### GLAMM



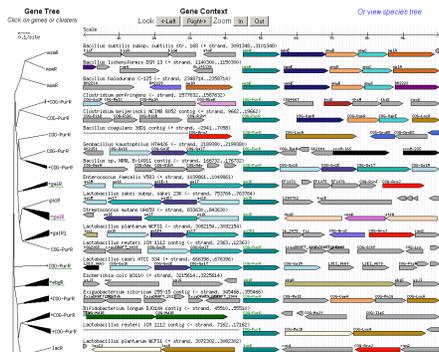
Compute

Web services

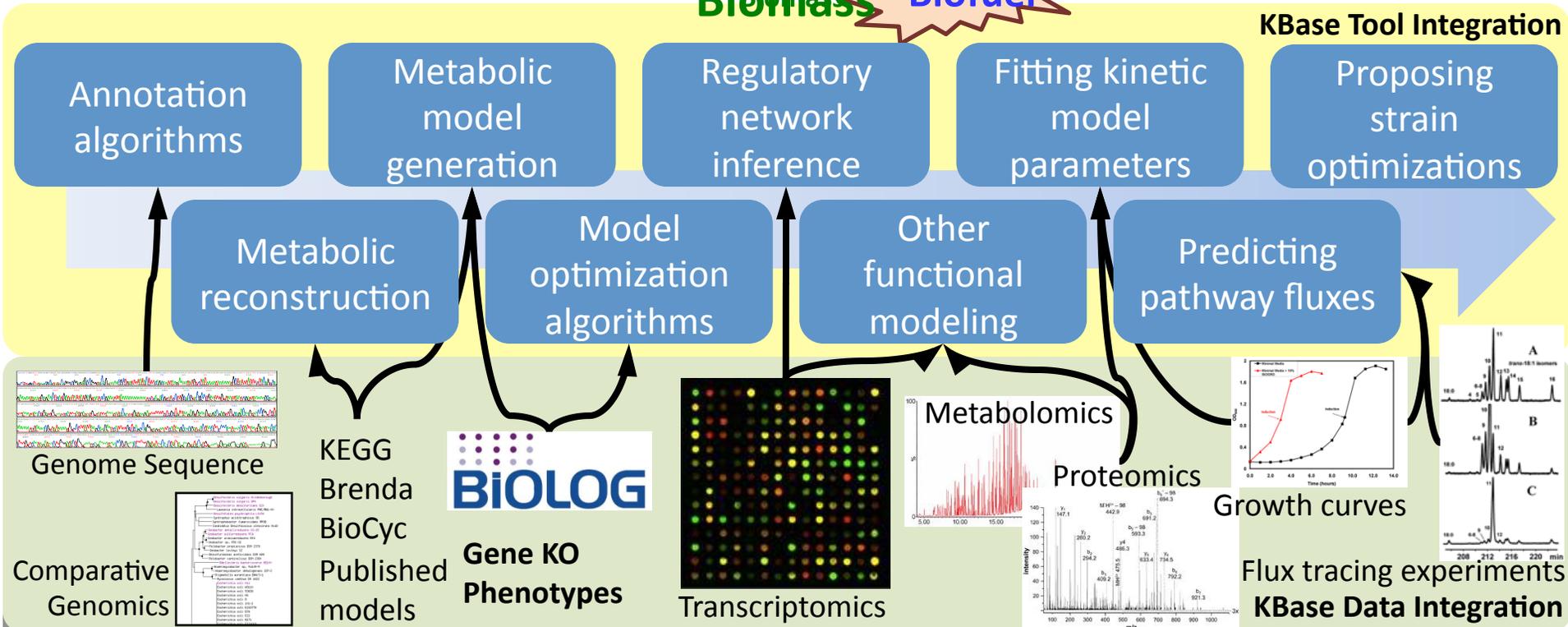
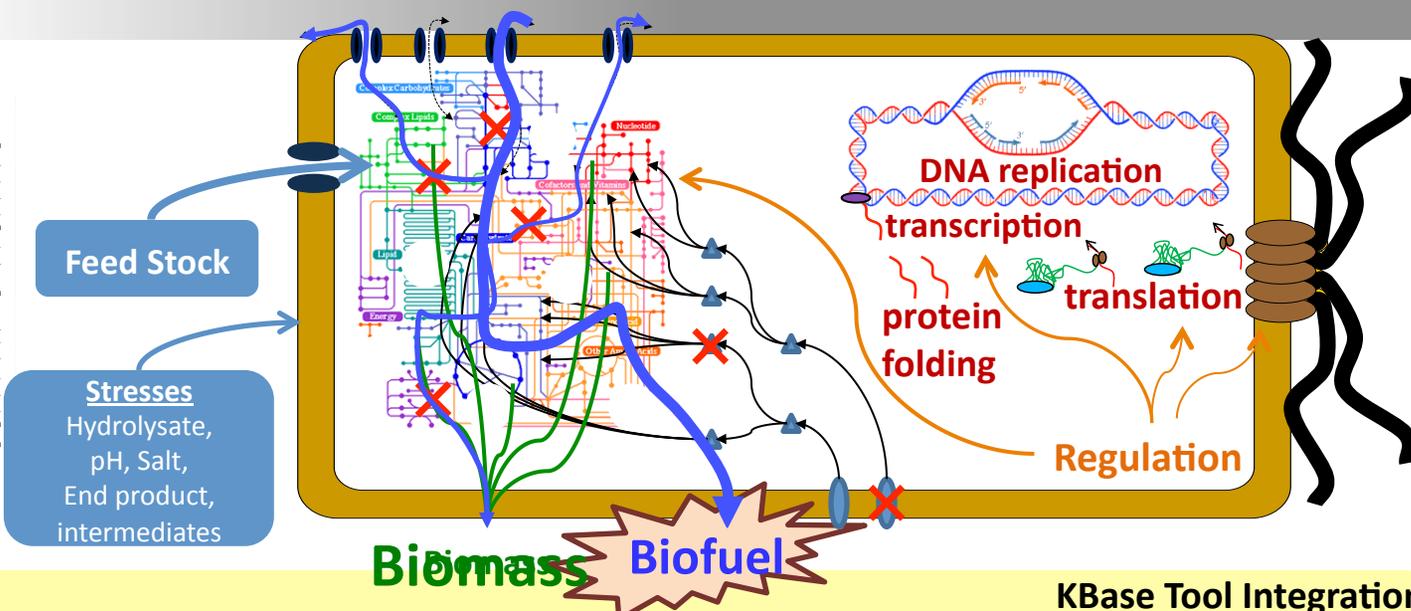
Central Store

User store

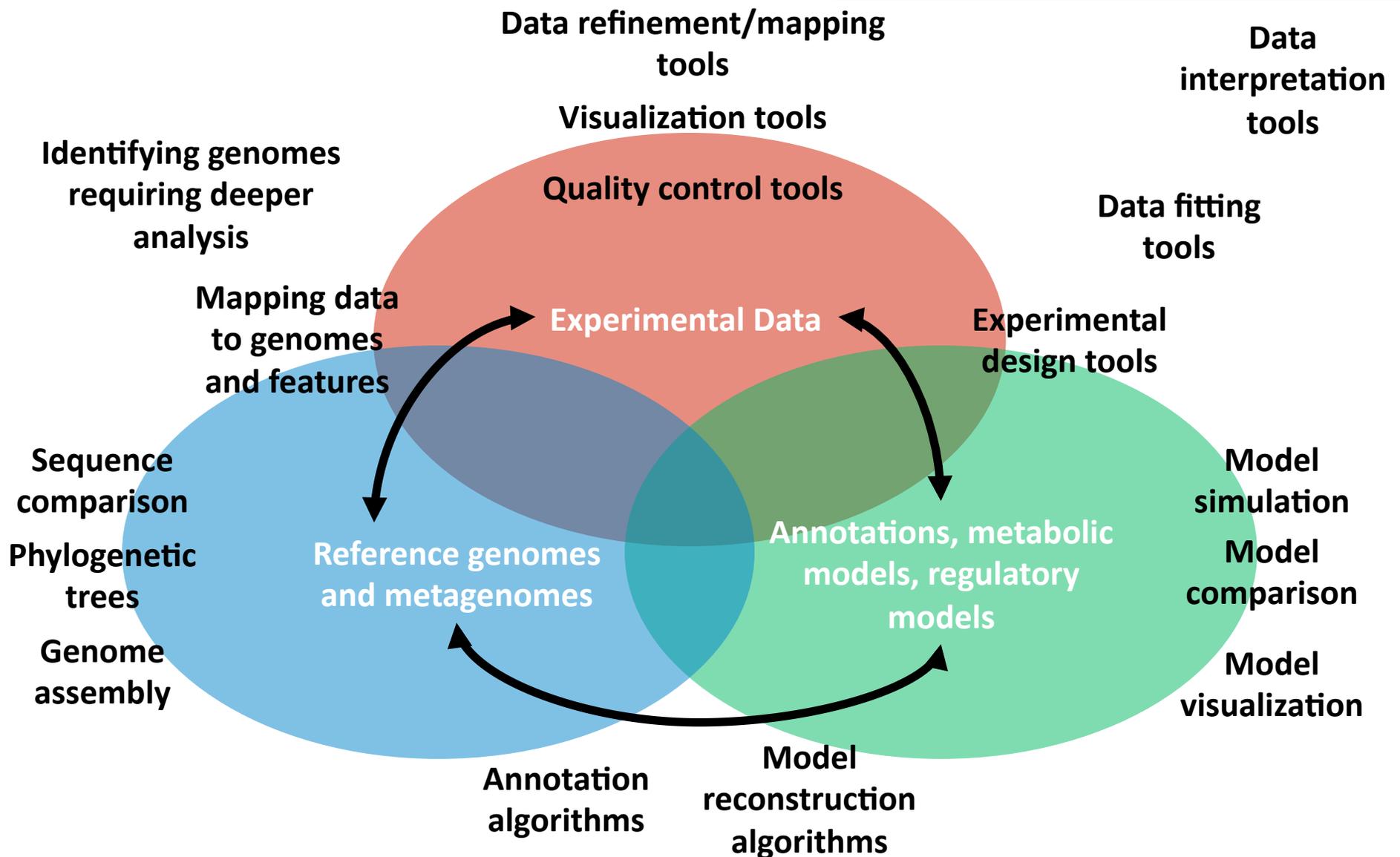
**Annotated Genome**



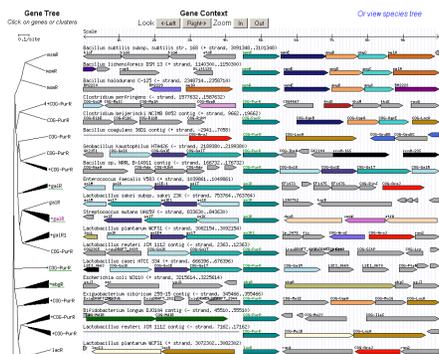
# DOE Knowledge Base: an example workflow



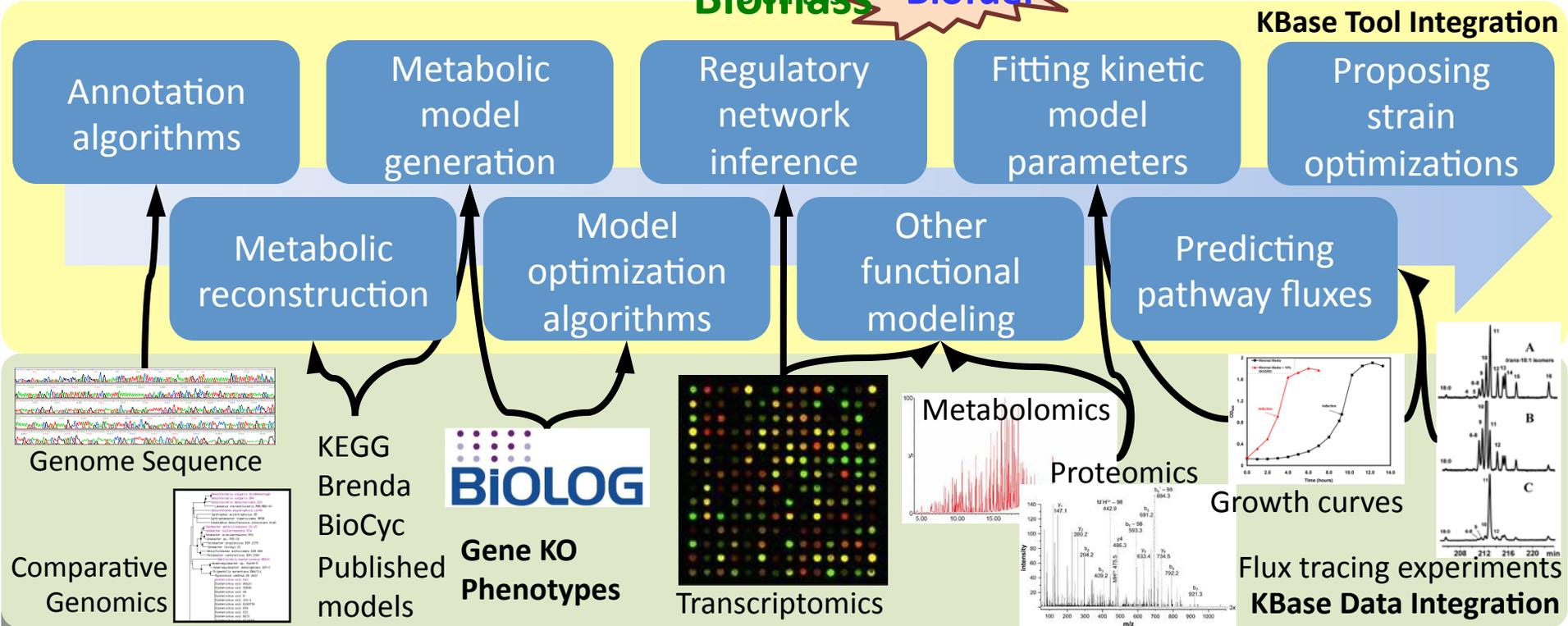
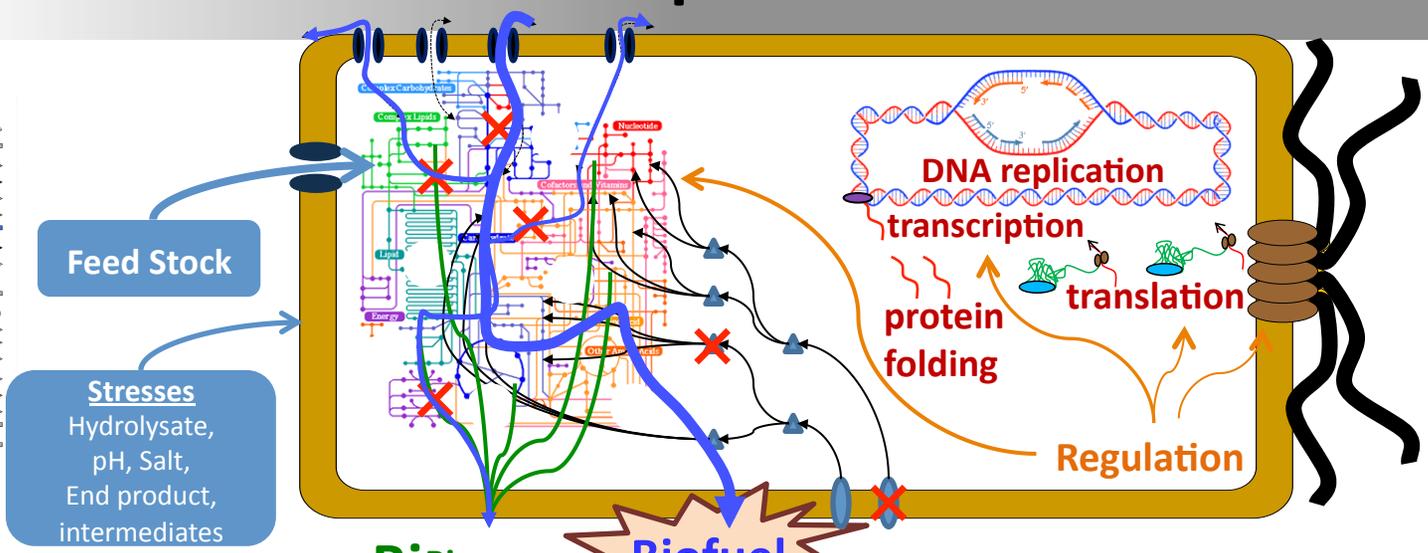
# Increasing Need for Integration of Data/Tools



### Annotated Genome



# Example Microbial Workflow



## Increasing Scale of Experiments

- JBEI Enigma project performing genome-wide transposon mutagenesis in hundreds of growth conditions
  - Large-scale identification of viable genes in growth conditions
  - 300,000+ phenotypes per genome
  - 10,000+ of growth curves in hundreds of conditions
- UMN and ANL culturing and sequencing 100s of soil microbes
  - Multiple growth curves for each microbe
  - Biolog phenotype arrays for each microbe
  - All genomes will be assembled, annotated, modeled, and reconciled
- Over 600 designed strains of *B. subtilis* knocking out all dispensable regions of the genome
  - 35% of the genome knocked out

## Increasing Scale of Available Reference Data and Models

- Over 4000 public prokaryotic genome sequences in GenBank
- 50,000 private genomes processed through RAST
- Over 15,000 metabolic models constructed in the Model SEED
- Enabling comparative analysis at a scale never before possible

Essential and metabolic genes in all models

