

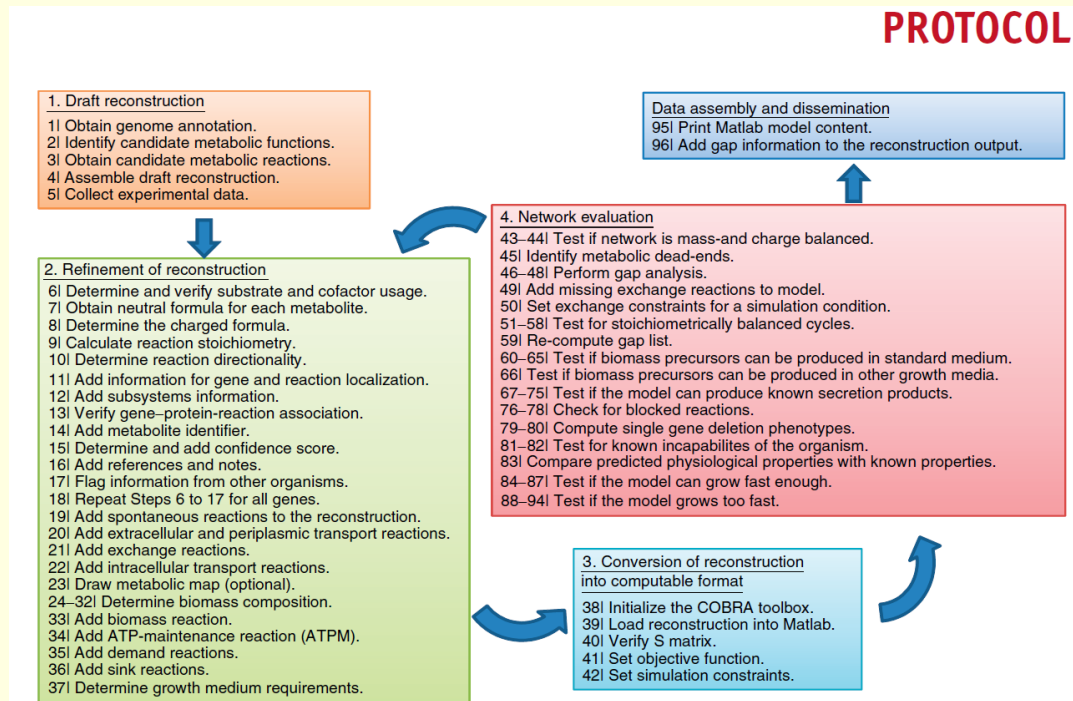
# High-throughput generation, optimization and analysis of genome-scale metabolic models

Christopher S. Henry



# 96 steps to model generation

## palsson (nprot 2009)



**Figure 1** | Overview of the procedure to iteratively reconstruct metabolic networks. In particular, Stages 2–4 are continuously iterated until model predictions are similar to the phenotypic characteristics of the target organism and/or all experimental data for comparison are exhausted.

The seed pipeline replicates 73 of the first 82 steps in the protocol.

# the Model SEED

<http://www.theseed.org/models/>

---

- A web-based resource designed to speed the creation of new metabolic models.
- Currently works on non compartmentalized species mostly bacteria
- Creates “Draft Models”
- ‘Analysis-ready’ models (as they can simulate the production of biomass from transportable nutrients)
- Supplies tools for manual alteration of the models (GUI + SBML alterations)



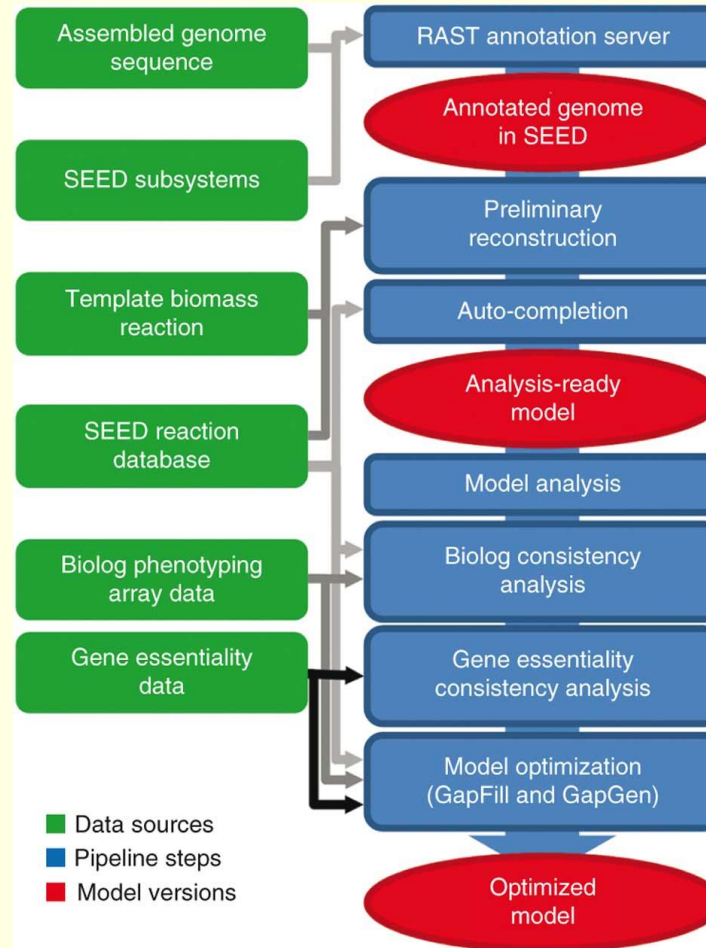
# Core logic

---

- Integration & augmentations of technologies for genome annotation
- Construction of gene-protein-reaction (GPR) associations
- Generation of biomass reactions
- Reaction network assembly
- Thermodynamic analysis of reaction reversibility
- Model optimization



# The protocol



# AVAILABLE METABOLIC MODELS – LAST YEAR

## Industrially relevant organisms

*E. coli*

- 2077 Reactions
- 1260 Genes



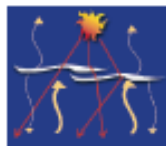
*S. cerevisiae*

- 1402 Reactions
- 910 Genes



*M. barkeri*

- 619 Reactions
- 692 Genes



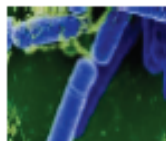
*G. sulfurreducens*

- 522 Reactions
- 588 Genes



*B. subtilis*

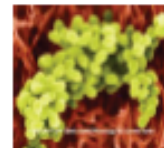
- 1020 Reactions
- 844 Genes



## Pathogens

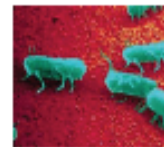
*S. aureus*

- 640 Reactions
- 619 Genes



*S. typhimurium*

- 898 Reactions
- 826 Genes



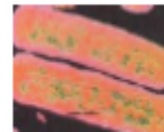
*H. pylori*

- 558 Reactions
- 341 Genes



*H. influenzae*

- 472 Reactions
- 376 Genes



*M. tuberculosis*

- 939 Reactions
- 661 Genes



## Mammalian cells

*H. sapiens*

- 3280 Reactions
- 1484 Genes



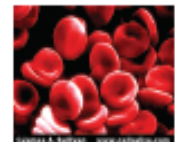
Human mitochondria

- 218 Reactions

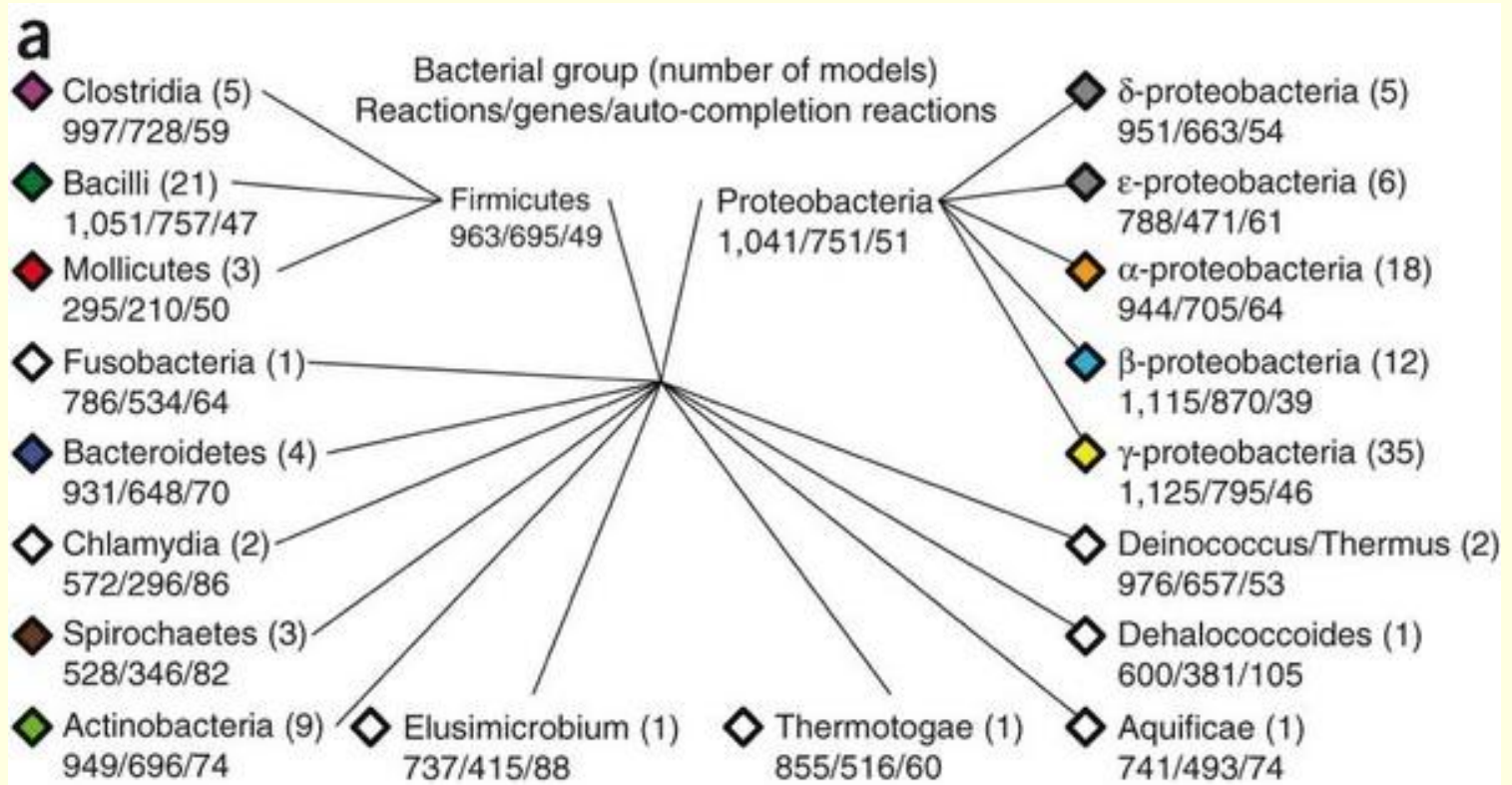


Red blood cell

- 39 Reactions



# 130 metabolic models were published by the seed project



Models are displayed along with the number of models contained within each group and the average number of reactions, genes and auto-completion reactions in the Group members.

# Inactive reactions

---

Reactions that cannot carry flux during simulated growth (FVA) are indicative of gaps in the metabolic network where additional manual curation is required.

- In the 130 SEED models, the average fraction of inactive reactions is 31.7%
- In manually refined published models the average fraction of inactive reactions is 16%.

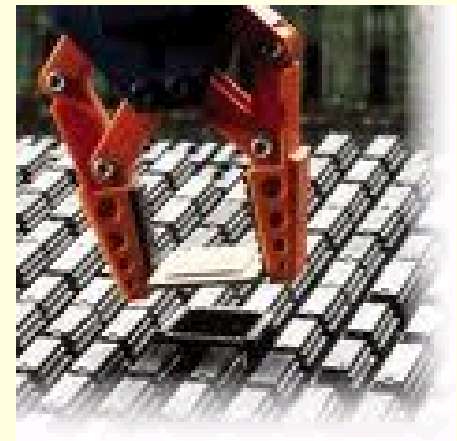




# Auto completion

---

- Average of 56 reactions per model (on the 130 models)
- Number of reactions added increased as the total number of reactions in the model decreased

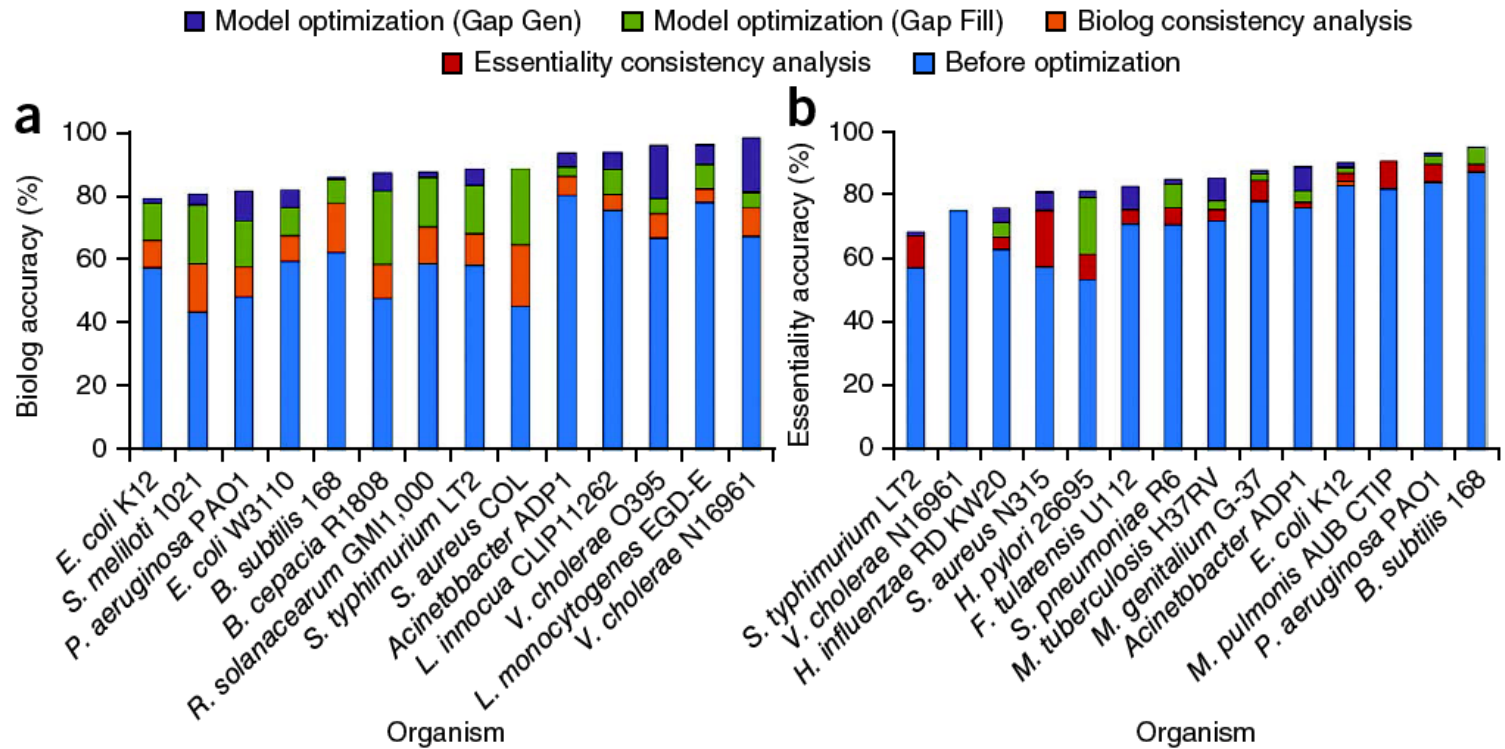


# Auto completion method

$$\text{Minimize } \sum_{i=0}^R \left( 1 + P_{T,i} + P_{K,i} + P_{SS,i} + P_{F,i} - f_{SS,i} - f_{p,i} \right) z_i$$

- $z_i$  is a binary variable created for any reaction (including direction) not currently included in the model.
- $P_{T,i}$  is a penalty on the addition of transport reactions during the auto-completion process.
  - 4 for transport reactions involving compounds in the biomass reaction,
  - 2 for all other transport reactions
- $P_{K,i}$  is a penalty favoring addition of KEGG reactions.
- $P_{SS,i}$  is a penalty favoring the addition of reactions mapped to SEED functional roles and subsystems.
- $P_{f,i}$  is a penalty on the addition of reactions proceeding in a thermodynamically unfavorable direction.
- $f_{SS,i}$  is a bonus applied to reactions involved in subsystems already well represented in the preliminary model.
- $f_{p,i}$  is a bonus applied to reactions involved in short linear pathways

# Models Accuracy



# Models Accuracy

**Table 1 Prediction accuracy of SEED models**

Organism	Published model exists	Biolog accuracy (%)		Essentiality accuracy (%)	
		Original	Optimized	Original	Optimized
<i>B. cepacia</i> R1808	No	47.5	87.3	–	–
<i>E. coli</i> W3110	No	59.3	81.8	–	–
<i>F. tularensis</i> U112	No	–	–	70.9	82.5
<i>L. innocua</i> CLIP11262	No	75.5	93.8	–	–
<i>L. monocytogenes</i> EGD	No	77.8	96.0	–	–
<i>M. pulmonis</i> AUB CTIP	No	–	–	81.8	90.5
<i>R. solanacearum</i> GMI	No	58.6	87.7	–	–
<i>S. meliloti</i> 1021	No	43.2	80.6	–	–
<i>S. pneumoniae</i> R6	No	–	–	70.6	84.8
<i>V. cholerae</i> N16961	No	67.1	98.2	75.0	75.0
<i>V. cholerae</i> O395	No	66.5	95.7	–	–
New model average	No	61.9	90.1	74.6	83.2
<i>Acinetobacter</i> ADP1	Yes	80.0	93.3	75.7	88.8
<i>B. subtilis</i> 168	Yes	62.0	86.0	87.2	95.0
<i>E. coli</i> K12	Yes	57.1	79.3	82.7	89.9
<i>H. influenzae</i> RD KW20	Yes	–	–	62.9	75.7
<i>H. pylori</i> 26695	Yes	–	–	53.2	80.9
<i>M. genitalium</i> G-37	Yes	–	–	77.7	87.5
<i>M. tuberculosis</i> H37RV	Yes	–	–	71.9	85.1
<i>P. aeruginosa</i> PAO1	Yes	48.1	81.5	83.9	92.9
<i>S. aureus</i> COL	Yes	45.2	88.7	–	–
<i>S. aureus</i> N315	Yes	–	–	57.3	80.6
<i>S. typhimurium</i> LT2	Yes	58.0	88.6	57.0	68.2
Models with published counterpart average	Yes	58.4	86.2	71.0	84.5

Empty elements in the table indicate a lack of Biolog or essentiality data for the corresponding organism.

# Interesting results

---

- The number of essential metabolic genes remained relatively constant around an average value of 237 (others used encoding secondary metabolic functions, transcriptional control and signaling mechanisms to improve versatility).
- Only 47 reactions were associated with essential genes in nearly every model analyzed whereas 740 reactions were associated with essential genes in fewer than ten models analyzed
- the number of essential nutrients decreases as the number of reactions in the models increases





**Thanks!**