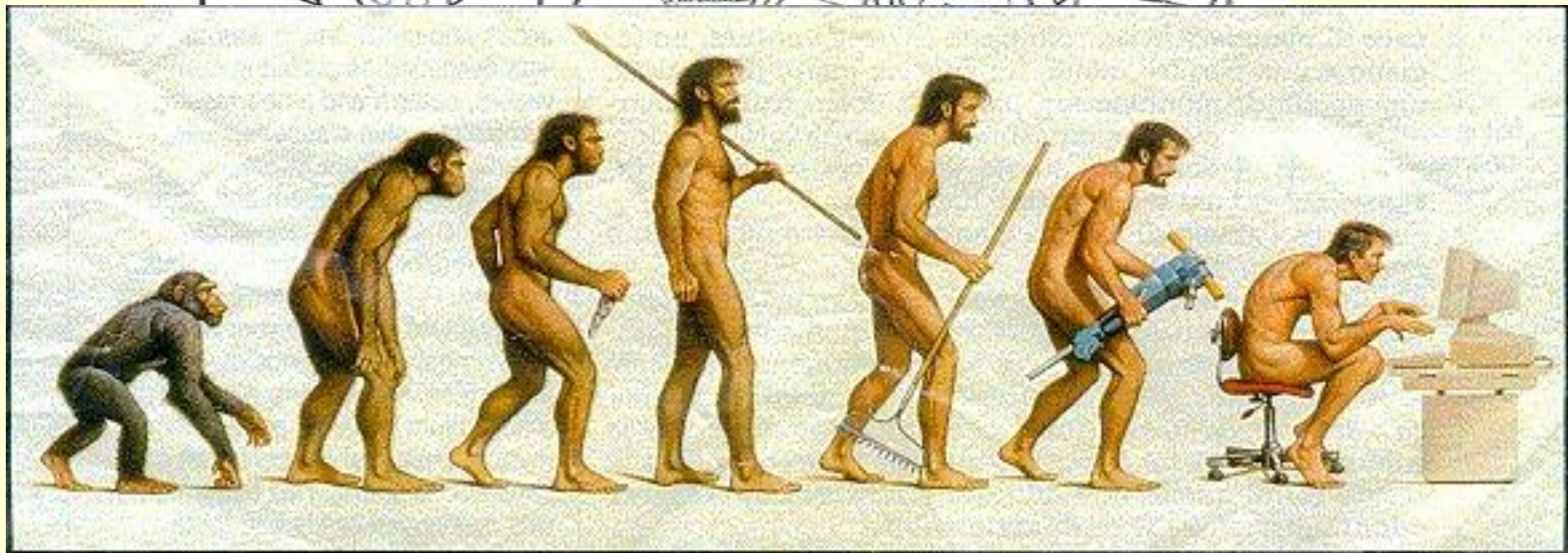


# 用系统的方法预测基因组进化

周 姚  
遗传学

# 引言：一路走来



**Somewhere, something went terribly wrong**



Ernst Haeckel 1866

引言： 下一步.....



# 内容纲要

- 模型：基因组水平的代谢模型
- 预测：
  - 突变效应分布及上位性互作
  - 基因组进化的一般方式
  - 进化轨迹
- 总结

# 模型：基因组水平代谢网络模型

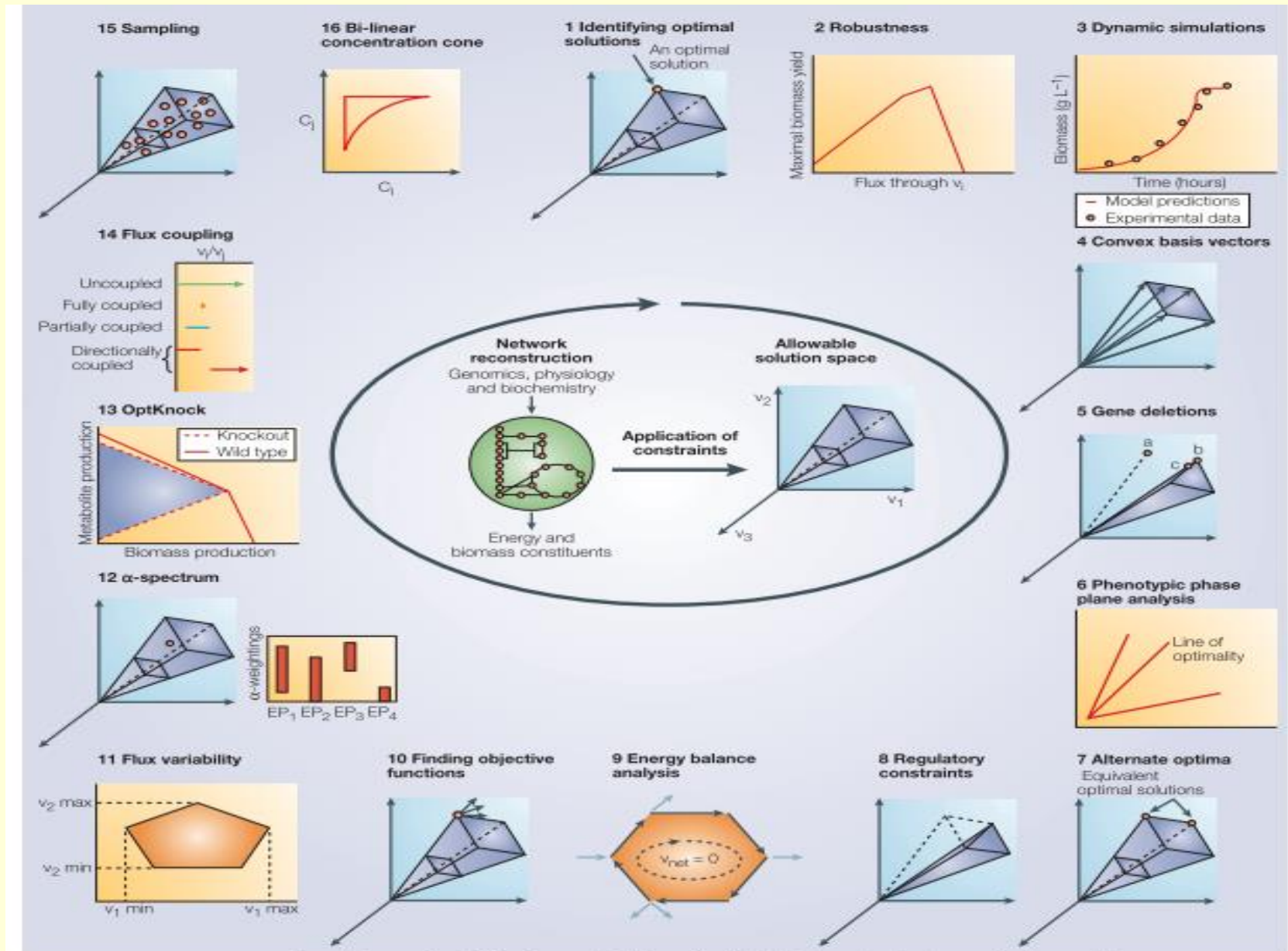
- 这是一类**基于限制**的模型（constraint-based model）。比如流平衡分析（FBA）。
- 限制分为四类：
- **物理-化学限制**（Physico-chemical constraints）：比如物质在细胞内的扩散，渗透压，反应速率等；
- **拓扑结构限制**（Topobiological constraints）：生物大分子的三维结构受其所在的细胞环境的影响，而结构和功能是密切相关的；
- **环境限制**（Environmental constraints）：营养物质、pH值、温度、电子受体数量等；
- **调控限制**（Regulatory constraints）：与以上三种硬性条件限制不同的是，这类限制是自我强加的。比如在某种环境中基因的表达与否，或者表达的酶活的强弱等。它保证了生物体能在艰难环境中的生存。

# 限制的数学表示

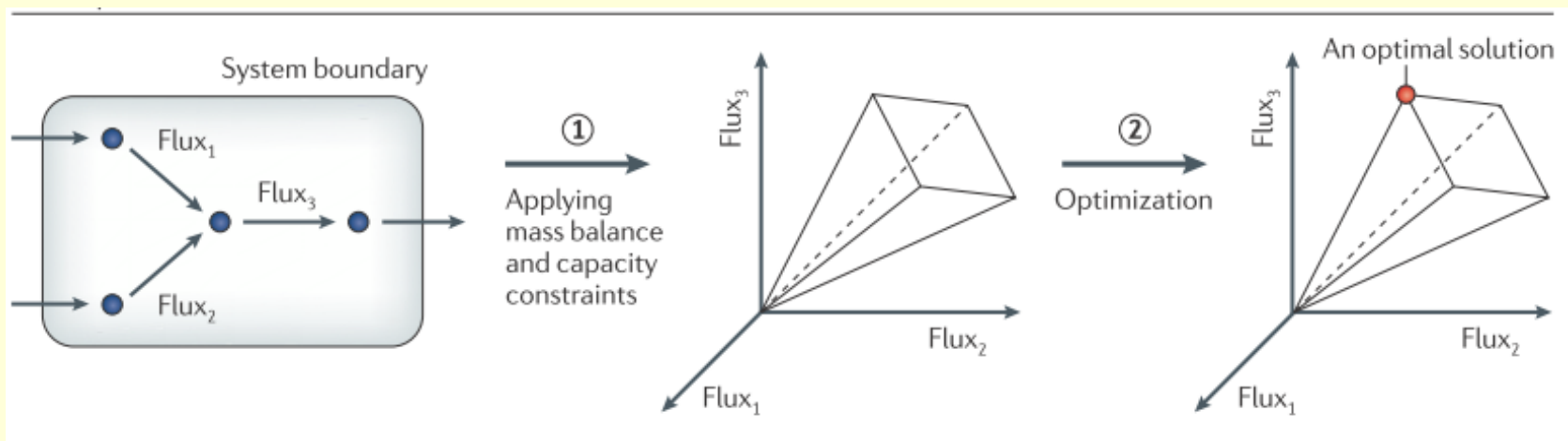
- 限制的两种基本类型：平衡（balance）和边界（bounds）
- 平衡： $S \cdot v = 0$
- 边界： $v_{\min} \leq v \leq v_{\max}$
- 这样，把一个平衡网络限制在一定的范围内，就给数学分析带来了可能。
- 解空间（solution space）：数学术语。所有可行解所组成的集合构成一个向量空间，就叫它解空间。



# 复杂的过程



# 简而言之





# 预测：三个层次

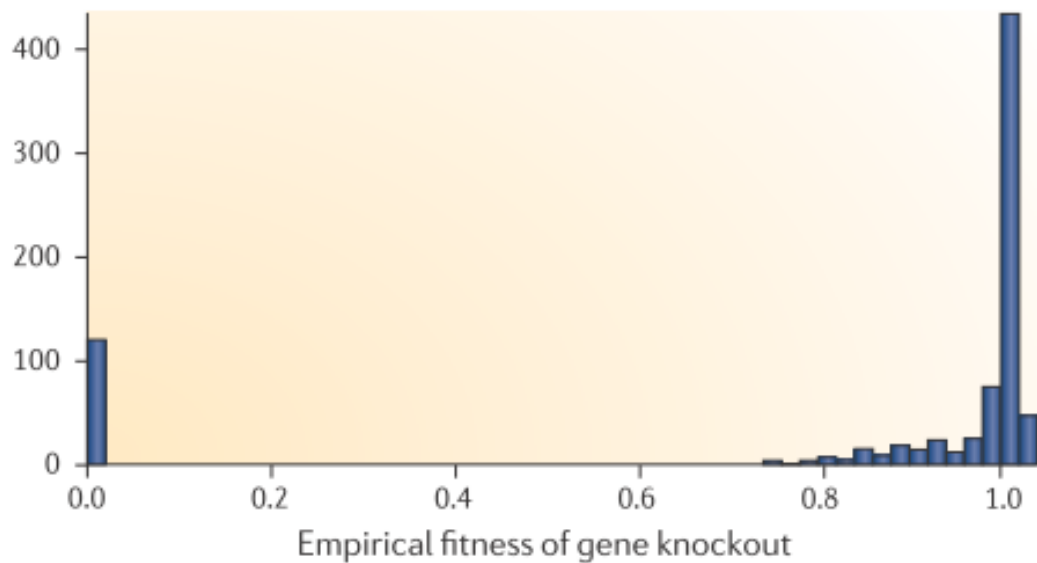
- 突变效应分布及上位性互作
- 基因组进化的一般方式
- 进化轨迹

# 突变效应分布

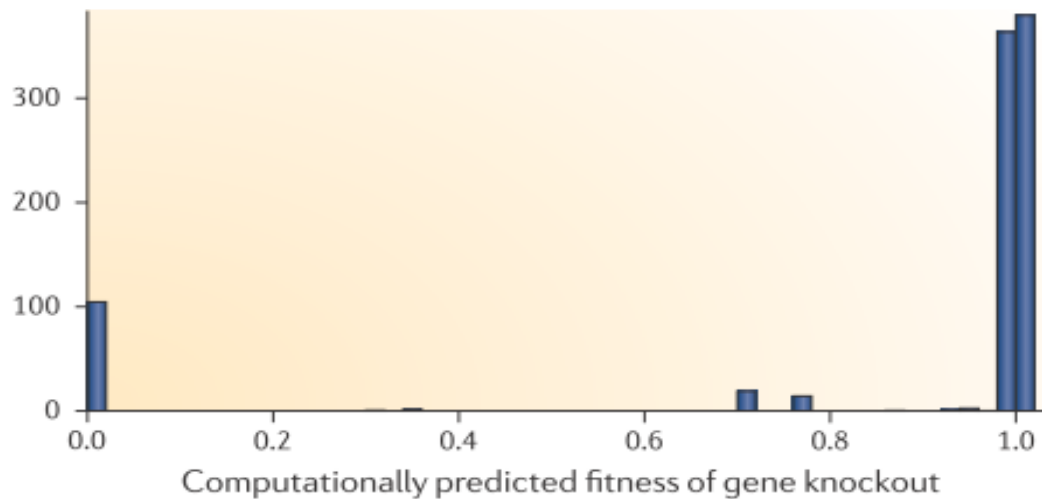
- 在进化过程中，突变是主要动力。
- 当把微生物基因组上的所有基因一个一个敲除的时候，会对微生物的适应性产生什么样的影响？

**Aa**

Number of genes

**Ab**

Number of genes



- **基因的非必需性**（gene dispensability）：  
在某种特定的环境条件下，当一个基因敲除时，并没有影响生物的适应性，我们称这个基因在这种环境中是非必需的。
- **基因敲除悖论**（gene knockout paradox）：  
生物的性状从根本上说是受基因决定的。但是，当你把一个特定的基因从生物体内敲除时，绝大多数时候，并不影响生物体的生存（适应性）。

WHY?

# WHY?

- **环境影响基因的表达**：一种基因在一种情况下是非必需的，但是在另一种情况下可能就是必需的。用模型进行模拟显示，占**37-68%**；实验结果显示约为**50%**；
- **存在多拷贝**：当敲除一个时，其它的拷贝会进行补充。预测比例占**15-28%**；实验表明约为**37.5%**
- **其他代谢途径补充**：一种表型可能不是仅仅由一条代谢进行的。存在其他的代谢途径对其进行补充。预测其比例占**4-17%**，实验结果为**12.5%**。

# 上位性互作

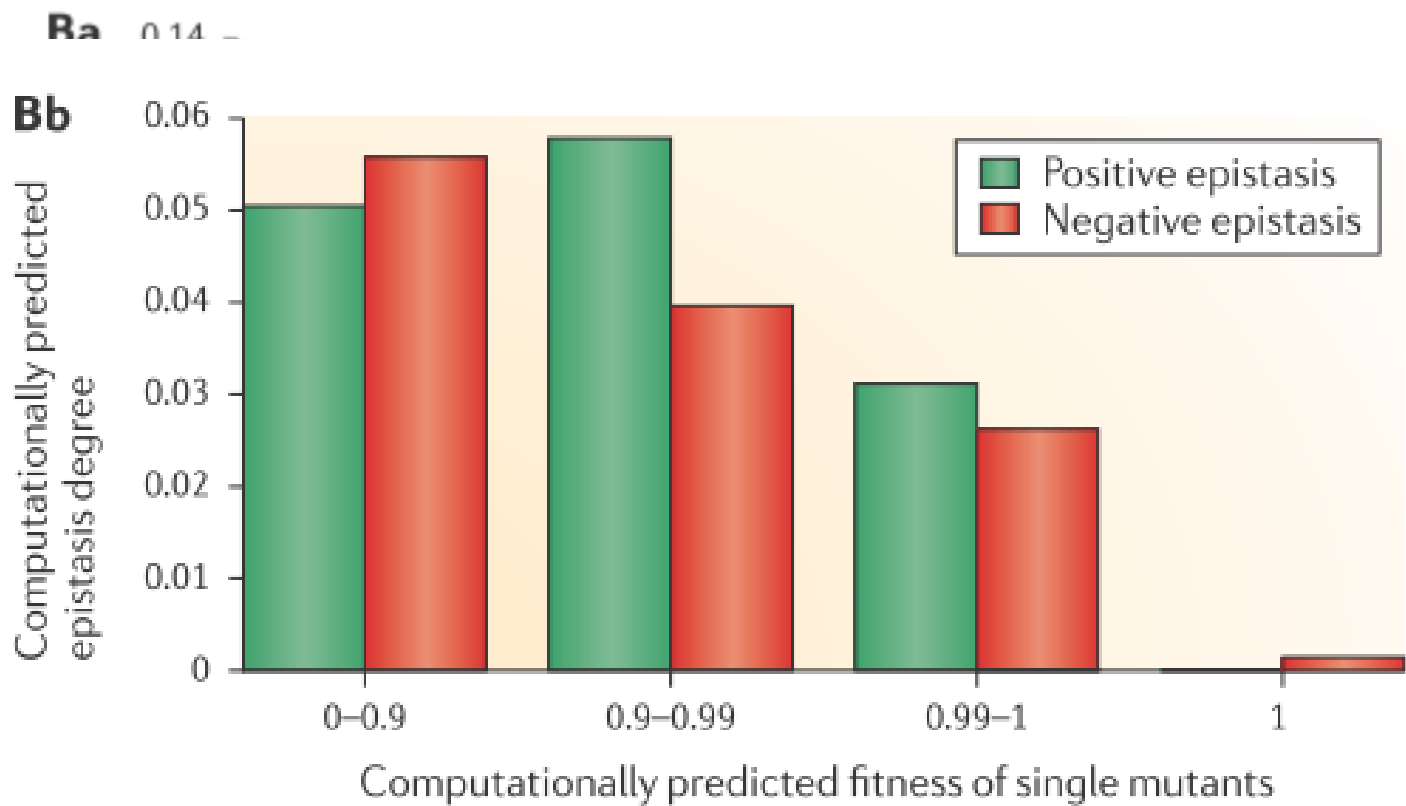
- 一对等位显性基因的表现受到另一对非等位基因的作用，这种非等位基因间的作用叫上位效应。
- 一个正在飞行的飞机如果它有两个发动机，一个罢工也许并没有什么，但是如果两个同时罢工，其所带来的影响恐怕不是一个发动机所带来影响的两倍。



# 数学分析

- 有两个位点：A和B
- $aB:1-s$
- $Ab:1-t$
- 那么  $ab$  呢？
- $(1-s) * (1-t) ?$
- $(1-s) * (1-t) + \epsilon$
- $\epsilon > 0$ , 正向互作；  $\epsilon < 0$ , 负向互作

- 一个分离群体，在没有选择压的存在下会因为上位性的作用而变得各不相同。
- 早期的研究把认为在一个群体中其上位性的均值为0；模型中也就没了 $\epsilon$ ；
- 然而，这种简化是与事实不同的。后来群体遗传学家的在研究中把 $\epsilon$ 当做一个定值建立模型。因为他们如果 $\epsilon$ 不确定，模型的计算将巨大的不可想象
- 但是，系统模型估计出了上位效应



- 预测用的模型是非常简单。从模型中可以得出了四个结论：
- 一对基因间可能存在正的或负的互作，也有可能不存在互作；
- 互作可能在一种情况下存在，而在另一种情况下消失；
- 具有多效性的基因是少量的；
- 上位互作是普遍存在的。

# 预测：三个层次

- 突变效应分布及上位性互作
- 基因组进化的一般方式
- 进化轨迹

# 基因组进化的一般方式

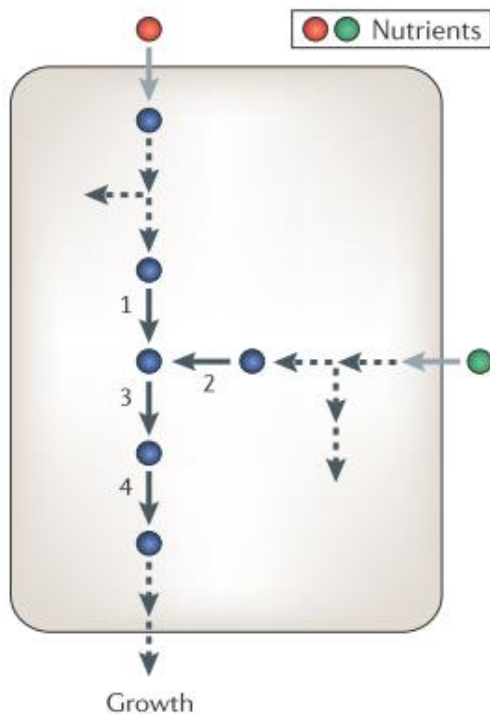
- 留下还是淘汰？（gene conservation）
- 是否需要复制？（gene duplicability）
- 基因的水平转移？（gene horizontal transfer）
- 序列该怎么变？（sequence and expressing evolution）



- 在代谢途径相关的基因中，系统模型通过整合基因敲除、酶活、代谢流分布及边界等信息，解释在基因保留、基因复制和序列进化等行为背后的分子机制。
- 如果基因组需要增加：
- 在微生物中，基因的水平转移在进化上占主要作用；
- 而在高等生物中，基因的复制占主要作用。

- 系统的模型能够显示不同的基因是同时加到一条代谢途径中的，还是在不同的时候；简单的原理如下：

Ca Asymmetric loss

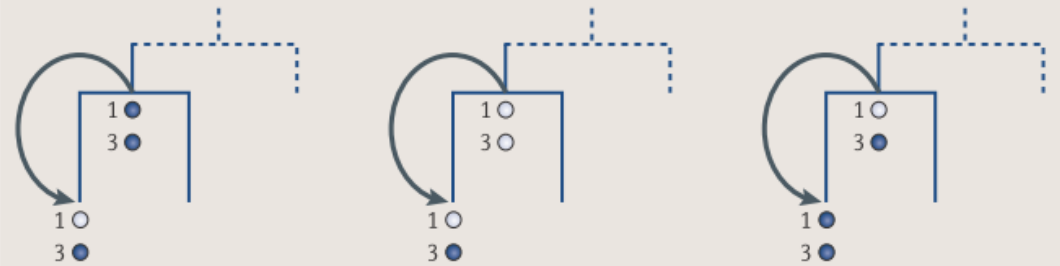


Cb Asymmetric loss

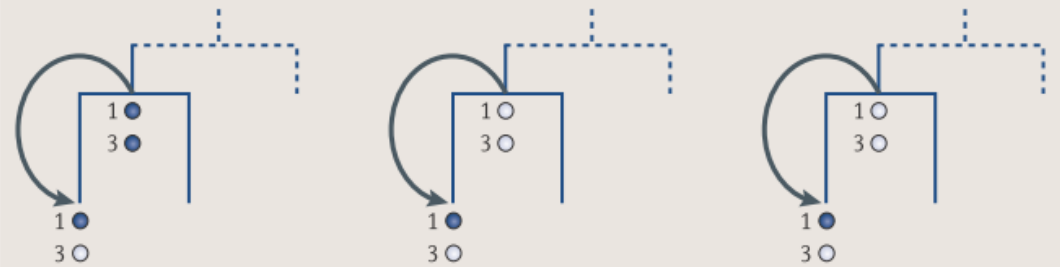
Asymmetric gain

Contingent gain

Most frequent scenario



Least frequent scenario



● Present gene ○ Absent gene

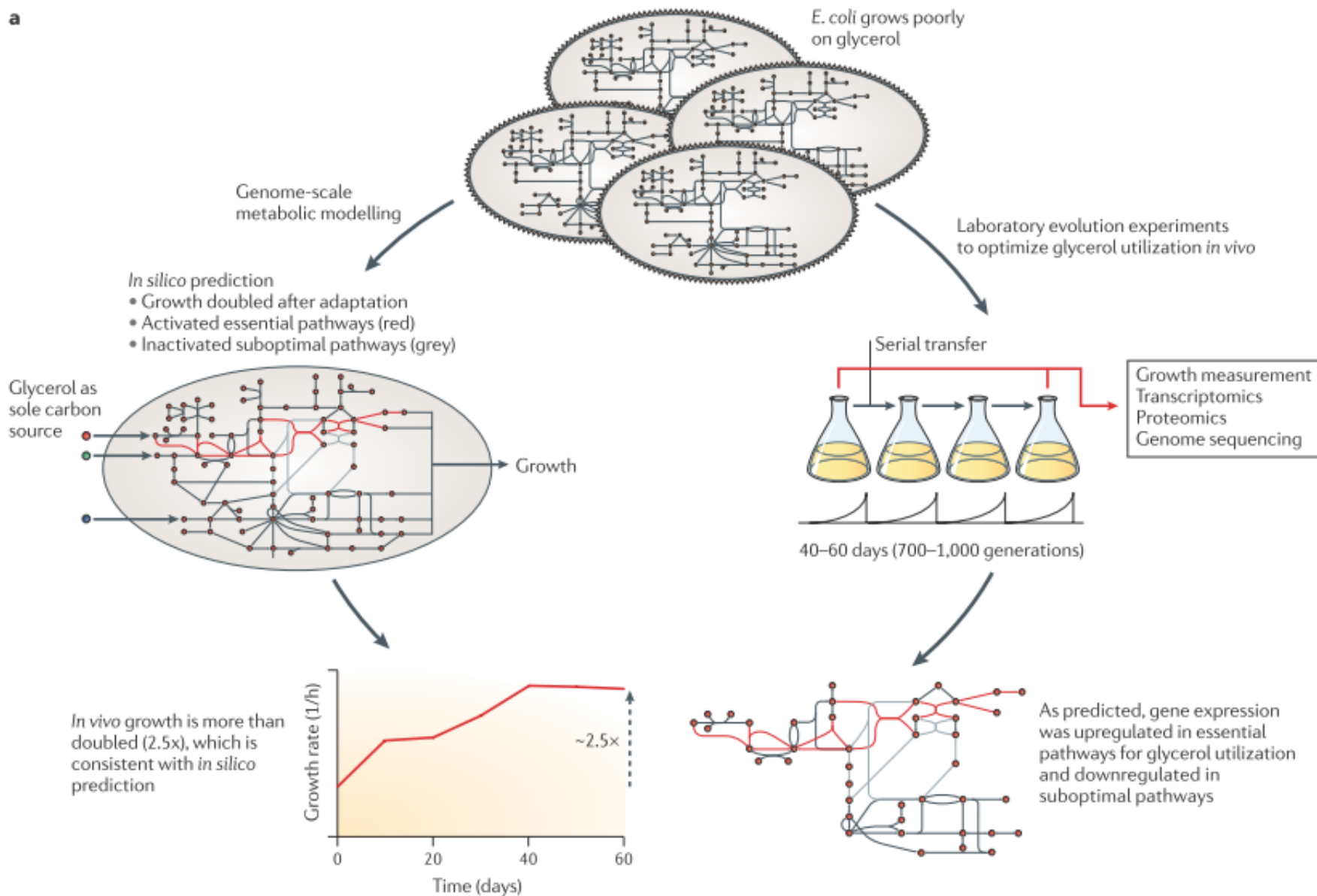
# 预测：三个层次

- 突变效应分布及上位性互作
- 基因组进化的一般方式
- 进化轨迹

# 进化轨迹

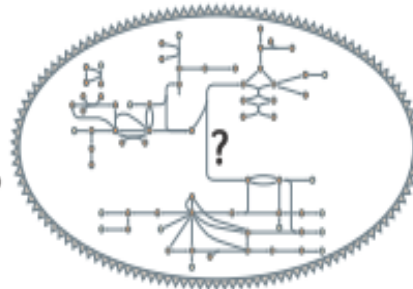
- 以上的研究都是从整体水平分析，那么到具体的代谢途径中，哪个基因会先丢失？什么样类型的基因最有可能加入进来？某个酶的活力将会是增强还是减弱（序列改变）？
- 实验室进化（Evolution in the laboratory）
- 基因组最小化演变（Evolution of genome minimization）

a



b

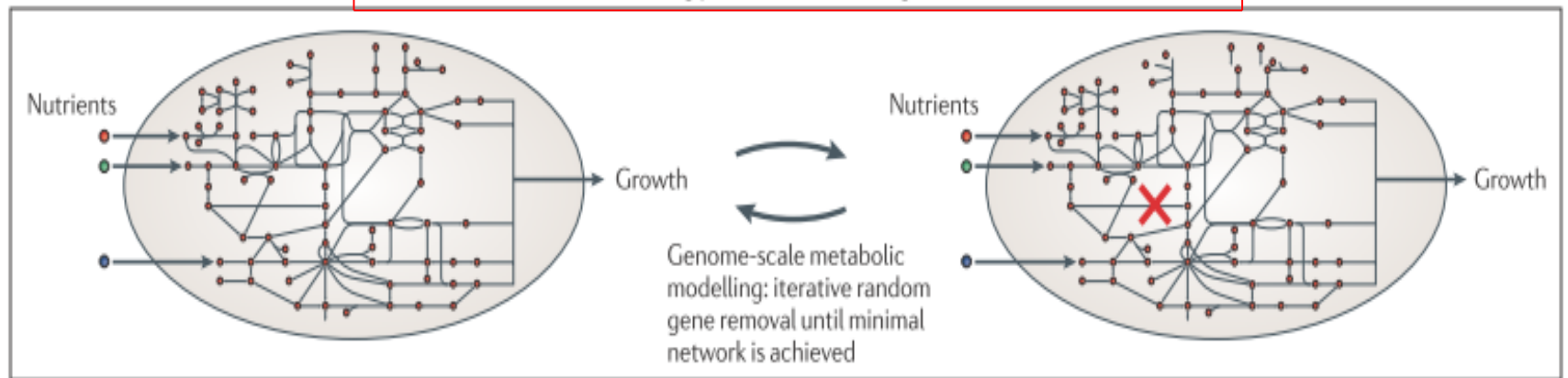
Is the metabolism  
of endosymbiotic  
bacteria predictable?



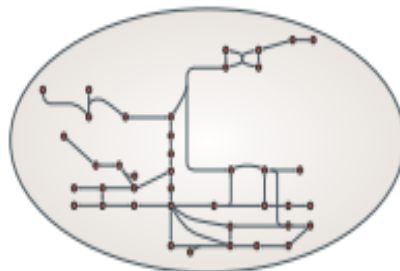
*Buchnera aphidicola*

In silico simulation of  
reductive evolution

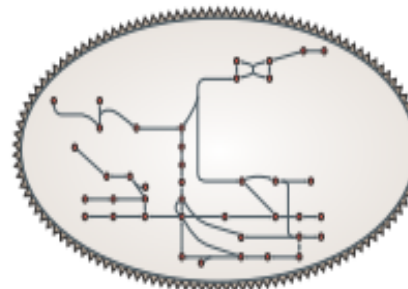
*E. coli* metabolic model as a starting point with mimicked growth environment of *Buchnera*



In silico minimal network



Comparison revealed  
~80% accuracy



Empirical network based  
on genome sequence



# 总结

- 用基于限制的系统模型，但我没告诉你具体的过程和原理；
- 做了突变效应分布及上位性互作、基因组进化的一般方式、进化轨迹三个层次上的进化分析，但我没有告诉你模型是怎么做到的
- 因为：我不懂模型，只懂进化中的几个名词而已

来源:

详细内容请参考文献:

*Systems-biology approaches for predicting  
genomic evolution*

Balázs Papp etc. NATURE REVIEWS

GENETICS; 2011,12: 591-602

谢谢!