

初代与次代GSMM

<https://doi.org/10.1007/s00253-022-12066-y>

1. 摘要

2 首先明确GSMM是什么(文献里这都没说, 无语)

1 文献将GSMM分为了两类

1. 初代GSMM: 缺乏限制, 预测精度极其有限
2. 次代GSMM: 初代GSMM+整合组学数据+增加约束+整合不同生物学模型+构建全细胞模型

2. Introduction

2.1. GSMM概念与构建

1 什么是GSMM:

1. 简单理解为就是一种数学模型
2. 基于特定生物体的基因组信息构建, 包括了基因、蛋白质、代谢物以及它们之间的相互作用

2 GSMM的核心: 反应GPRs关系(基因-蛋白质-反应)

3 GSMM构建的方式: 可能有上百步骤, 但是大致为

1. 模型起草
2. 模型细化
3. 模型数学化
4. 模型验证

4 GSMM自动化构建工具: RAVEN Toolbox, Model SEED, Merlin, 但是构建出的模型仍需修剪

2.2. GSMM分析算法: 模型应用

所有item见[clik here](#)

1 代谢通量(Flux):

1. 单位时间内通过代谢网络中特定代谢途径(pathway)的物质的量
2. 文献中说它反应了分子通过代谢途径的转化率(我觉得是转化速率?)

2 FBA通量平衡分析(Flux Balance Analysis):

1. 概述: 其实是个很基础的方法, 在重建基因组规模代谢网络中模拟代谢的数学方法
2. 大致步骤: 假定稳态(代谢物质产销平衡)→设定目标函数(如最大化产量)→线性规划寻求最优解
3. 缺陷:
 - 只能模拟稳态下的通量分布

- 受到代谢底物摄取率的限制(constrained by the substrate uptake rate), 使得精度降低

4. 其他基于FBA的通量分析算法

FBA变种	描述	文献
rFBA	纳入转录调控的FBA	https://doi.org/10.1006/jtbi.2001.2405
MD-FBA	考虑代谢物稀释的FBA	https://doi.org/10.1186/gb-2010-11-4-r43
dFBA	考虑酶动力学的FBA	https://doi.org/10.1016/S0006-3495(02)73903-2
cFBA	考虑反应热力学的FBA	https://doi.org/10.1371/journal.pone.0064567

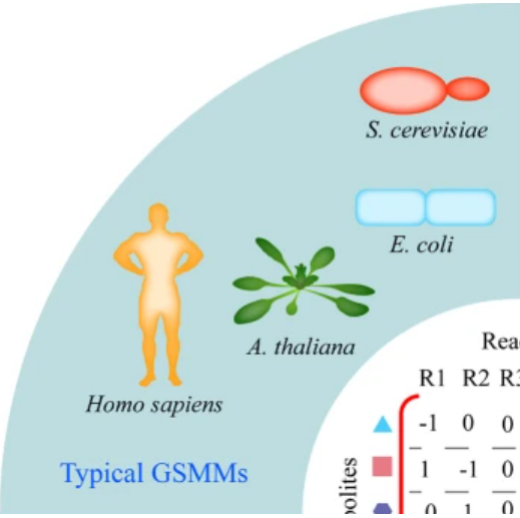
3 计算机辅助菌株设计算法(in silico strain design algorithm)

1. in silico来自拉丁语，字面意思是在硅中，引申为通过计算机模拟的实验
此外还有in vitro(在实验室)，in vivo(生物体内实验)
2. 该算法的目的：确定遗传修饰目标
3. 分类

算法类别	简要描述
基因删除	识别删除哪些基因能改善微生物性能
基因添加	添加新基因以改善微生物特性
基因上调/下调	调节特定基因表达水平，影响代谢途径活性和细胞代谢产物
添加异源途径	引入其他生物的代谢途径，生产新的代谢产物或提高已有产物生产

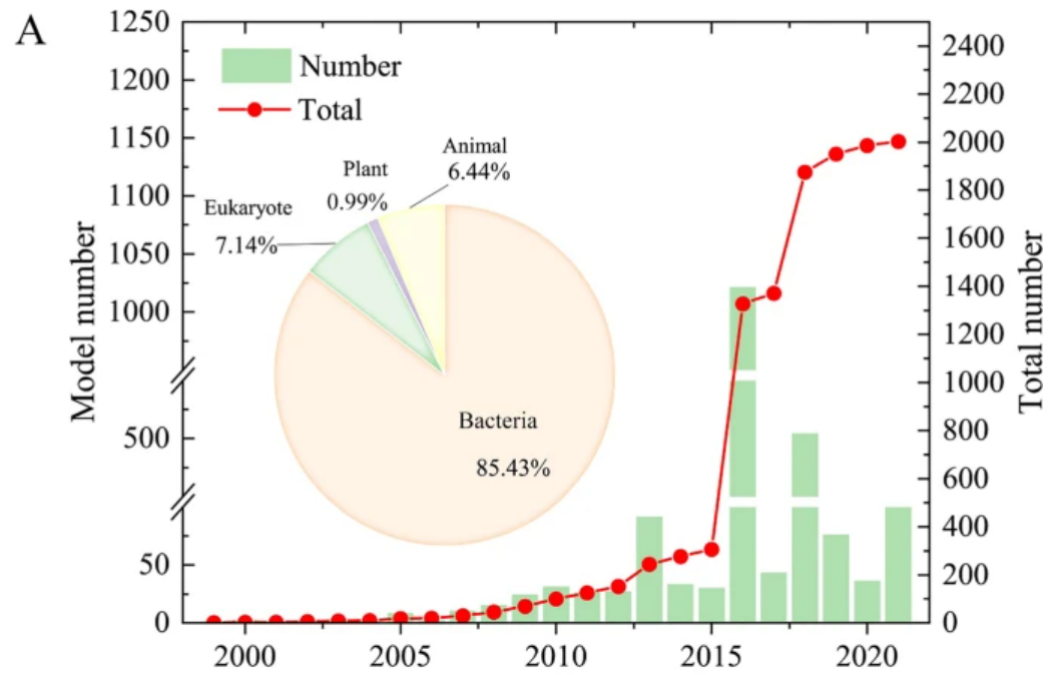
3. 初代GSMM

3.1. 典型GSMM



1 世界第一个GSMM：1999年流感嗜血杆菌的GSMM， <https://doi.org/10.1074/jbc.274.25.17410>

2 截至2021.8的GSMM数量和分布：eukaryotic(真核)



3 个别物种的GSMM

1. 玉米的GSMM具有最多的基因数量，最多的代谢物
2. 人类细胞的GSMM拥有最多的反应数

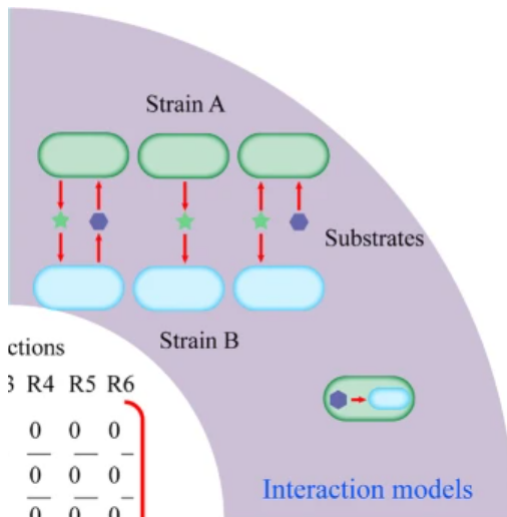
4 不同生物界的GSMM差异

1. 平均来讲，植物的GSMM拥有最多的基因
2. 平均来讲，动物的GSMM拥有最多的反应和代谢物

5 典型的模型生物有多个GSMM模型：这主要是源于GSMM的迭代升级，致使模型的规模增加，预测准确率也提升

3.2. 互作(Interaction)模型

两物种(微生物)间的关系



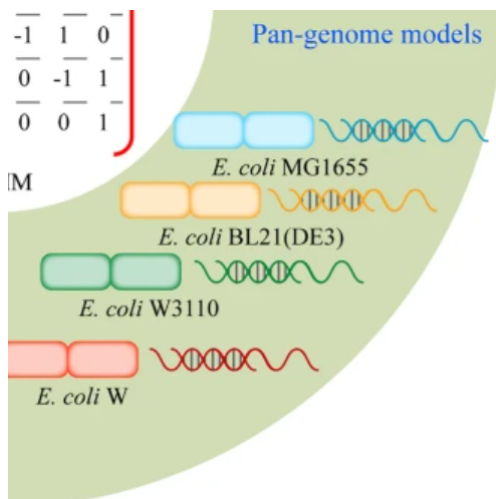
1 微生物的互相作用：互利共生(mutualism)、共生(commensalism)、寄生(parasitism)和竞争(competition)

2 模型的构成：三类不同的隔间(compartment)

1. GSMM隔间：每个GSMM有单独隔间，也就是每种微生物单独建模，自身的代谢封闭在自身的隔间
2. 共同体隔间：不同微生物的GSMM放在一起构成隔间，代表微生物之间互相发生代谢物交换
3. 环境隔间：模拟微生物与其周围环境之间的代谢物交换

3 用途：主要集中于模拟两种物种之间代谢物的交换，如联合发酵/共生

3.3. 泛基因组模型(Pan-genome models)



0 基因组：仅限于一个生物体内，所有遗传物质的总和

1 泛基因组：一个菌种内所有菌株的基因集合

1. 种类

- 核心基因组(core genome)：所有菌株共有的基因
- 可变基因组(dispensable genome)：一些菌株特有的基因

2. 作用：分析一个种内不同菌株的基因组特性，识别它们之间的表型差异
3. 构建：通过同属微生物

3 截至目前，只有三种泛基因组被发表，涉及的物种大多致病(由此泛基因组研究主要集中于，基于代谢能力的致病性分析)

3.4. 宏基因组模型(Meta-genome models)

1 宏基因组概述：

1. 概念：从特定环境样本中提取的所有遗传物质的总和，包括许多微生物的基因组
2. 指标：物种丰富度量化了微生物群落的物种数量，是宏基因组数据的重要指标
3. 构建：基于单微生物的基因组
4. 作用：评估菌株在不同环境下的代谢特征，预测不同物种(菌种)的交互

3 目前只有两种宏基因组模型被构建，每个模型需要构建几百个GSMM

4. 次代GSMM

4.0. 对初代GSMM的思考

1 初代GSMM其实也存在一个约束，就是底物的摄取速率，当然我觉得其实这是一种无奈的简化

2 初代GSMM还需要引入更多限制和调控，以提高精度

PS: GUR(葡萄糖摄取速率)和 μ (生长速率)的关系，GSMM预测的结果与实际严重不符

3 初代GSMM的改进集中于：

1. 整合组学数据
2. 添加约束条件
3. 整合不同生物网络
4. 构建全细胞模型

4.1. 组学数据的整合

4.1.1. 四种组学(omics)

1 基因组学(Genomics)

1. 含义：研究生物体的全部基因组，通过分析DNA序列来识别和定量基因/弄清功能和交互
2. 基因组注释(annotation)：标记基因组中的遗传元素及其功能，弄清哪些序列编码哪些蛋白/RNA
3. 在GSMM中：基因组学是GSMM的基础，基因组注释结果被用来预测每个基因的功能

2 转录组学(Transcriptomics)

1. 转录组：特定细胞、组织或生物体中，所有RNA分子(转录产物)的集合
2. 转录组学：研究基因在RNA层面的表达情况，即哪些基因转录成RNA，RNA在特定情况下翻译的水平

3 蛋白质组学(Proteomics)

- 1. 蛋白质组：特定范围(个体/细胞/组织)内，所有蛋白质的总集合
- 2. 蛋白质组学：量化表示所有蛋白的丰度(abundance)，研究蛋白的功能/结构/互相作用

4 通量组学(Fluxomics)：

- 1. 通量组：一个细胞内所有代谢通量的集合
- 2. 通量组学：量化分析通量组，测量分析流量

5 代谢组学(metabolomic)

- 1. 代谢物：指参与或由细胞代谢过程产生的小分子化合物，如糖类、脂类、氨基酸和核苷酸
- 2. 代谢组：一定范围内代谢物的集合
- 3. 代谢组学：代谢物的分析和定量

4.1.2. 组学应用于GSMM：提高预测精度

模型名称	组学数据整合
ME-model(代谢和基因表达模型)	转录组学
GIMME(基因通过代谢和表达调节而失活)	基因表达数据
iMAT(综合代谢分析工具)	转录组学和蛋白质组学
IOMA(整合组学-代谢分析算法)	蛋白质组学和代谢组学
GSMM结合13C通量组学数据	通量组学

4.2. 添加约束条件

1 所谓限制是什么限制

- 1. 整合动力学(kinetics)：如反应速率常数和底物浓度对酶活性的影响
- 2. 整合热力学(thermodynamics)：考虑化学反应的热力学特性，如吉布斯自由能变化，以确保模型中的代谢反应在热力学上是可行
- 3. 考虑酶学特性：如酶的活性，特异性，异构效应
- 4. 整合蛋白质(酶)3D模型：了解其功能和与其他分子的相互作用

2 实例

名称	限制内容	基本描述
k-ecoli457	动力学 (km和kcat值)	基于大肠杆菌核心GSMM
GSMMs (多种细菌)	热力学 ($\Delta rG'o$)	基于不动杆菌ADP1的GSMMs
GECKO	酶学限制	MATLAB/Python工具包
ecYeast7	酶学限制	\

名称	限制内容	基本描述
ec_iML1515	酶学限制	\
ijO1366 (E. coli)	3D结构信息	将参与代谢的蛋白质3D结构信息整合到模型中

4.3. 整合不同生物网络

1 其实除了GSMM，还有其他生物网络(biological networks)，以下为几种

1. 基因调控网络(GRNs, Gene Regulatory Networks)
 - 描述基因之间的调控关系
 - 例如一个转录因子可以激活或抑制特定基因的转录，从而影响该基因的表达水平
2. 蛋白质相互作用网络(PPIs, Protein-Protein Interaction Networks)
 - 描述蛋白质之间的相互作用
 - 例如信号传导、代谢途径和细胞结构维持
3. 信号转导网络(STNs, Signal Transduction Networks)
 - 涉及细胞内的信号传递过程
 - 这部分详见[here](#).

2 实例

名称	结合内容	基本描述
MTBPROM2.0	GSMM与GRN	更准确模拟基因删除效应
iFBA模型	GSMM, GRN, STN	动态模拟代谢、调控、信号

我其实觉得有必要多了解[iFBA模型](#)，目测有用机器学习爆改的潜力

4.4. 构建全细胞模型(whole-cell model)

简单理解就是GSMM-pro-plus-max

1 全细胞模型是什么

1. 最为复杂的生物学网络
2. 作用之一：描述细胞大分子(cellular macromolecular)的形成
3. 作用之二：研究生物大分子和代谢系统间的互相作用机制
4. 构建：将所有细胞活动(代谢模块化)→搞清模块之间的相互作用→数字化建模细胞活动

2 根据文中提到的文献，全细胞模型的功能为

1. 预测细胞行为
2. 资源分配计算
3. 基因与表型关系
4. 描述细胞生命周期

3 构建全细胞模型的挑战：要很多数据，数据一致性，数据整合，计算资源大，模型验证，需合作

5. GSMM在生物工业的应用

(industrial biotechnology)

5.1. 预测细胞表型(phenotypes)

- 1 表型：细胞在特定条件下的形态、大小、结构、代谢活动
- 2 通过计算不同条件下的代谢通量分布，比如基于不同底物摄取率，来模拟细胞的最大生长率
- 3 识别关键基因，如通过单基因删除算法来确定必需基因
- 4 GSMM可以模拟在不同条件下的细胞生长，为发酵过程的优化提供参考

5.2. 指导代谢工程

- 1 方法：模拟全局层面上的不同策略，去除竞争途径/增强合成途径/消除反馈抑制/引入异源途径
- 2 效果：引导合理的菌株设计，提高代谢工程的效率

6. GSMM在人类健康的应用

- 1 研究代谢失调
- 2 发现生物标志物和药物靶标
- 3 癌症研究：预测针对癌细胞增殖的潜在途径

7. GSMM的未来

关注这一点：

- 1 传统GSMMs中大多数参数是固定的，这限制了模型预测的准确性
- 2 机器学习方法的发展，提供了一种不需要固定约束参数、可以自动修改参数的新途径，从而提高模型的准确性
- 3 尝试将机器学习方法与GSMMs结合，以精确预测代谢通量的调控和复杂条件下的相互作用