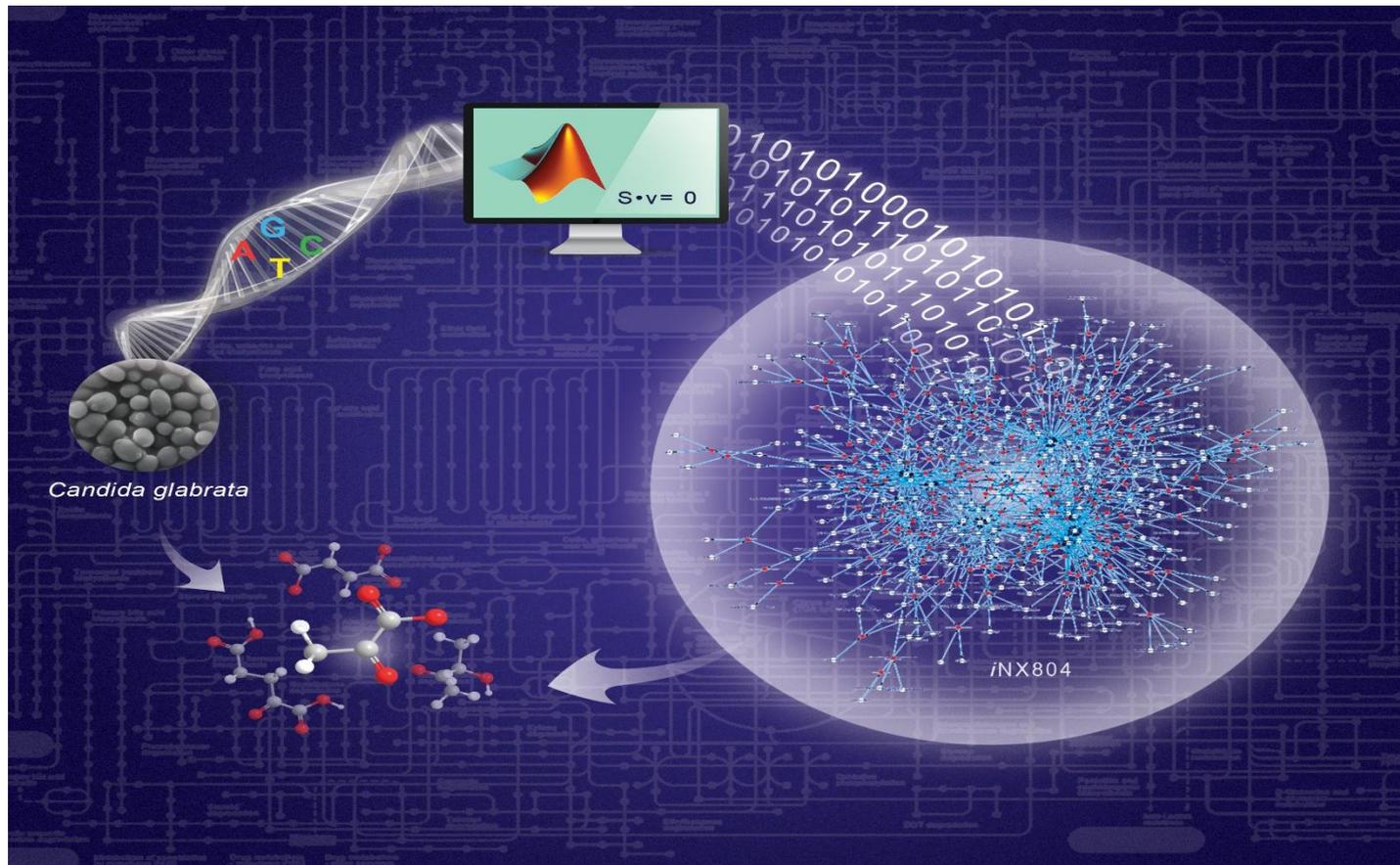


光滑球拟酵母基因组规模生物模型的构建与应用



□ 微生物基因组规模生物网络模型

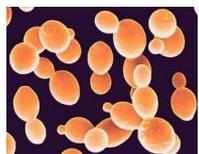
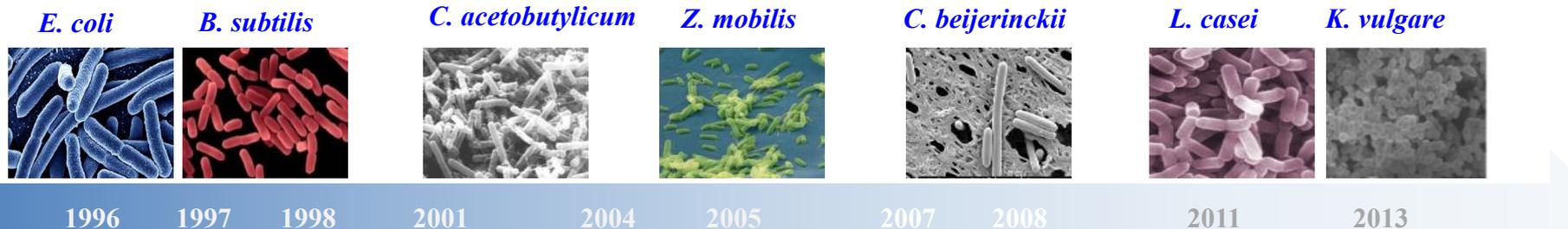
Genome-scale biological models, GSBMs 通过模拟物种的细胞行为以及物种和环境间相互作用，从而真实地反映生物系统的复杂性和全面性。



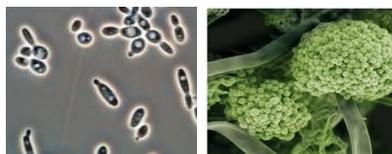


后基因组时代的组学技术

典型工业微生物全基因组测序

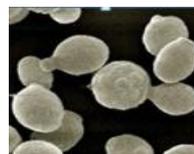


S. cerevisiae



Y. lipolytica

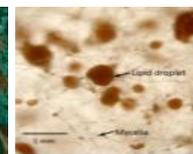
A. niger



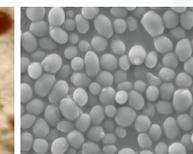
S. stipitis



P. rubens



M. alpina



C. glabrata

2330个微生物已完成，7556个正在进行.....

后基因组时代下的组学大数据



基因组数据



基因结构
和功能



转录组数据



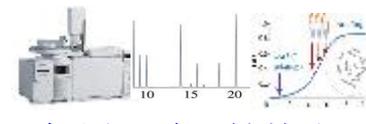
RNA量化
调控机制



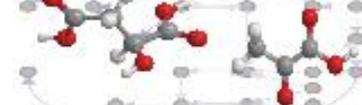
蛋白组数据



蛋白质状态
和相互作用



流量&代谢物组

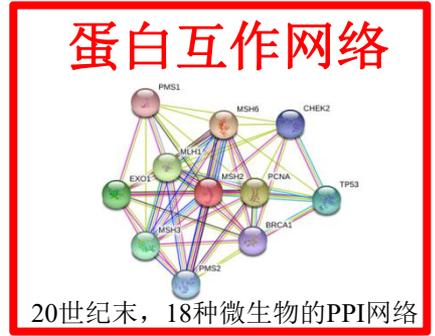
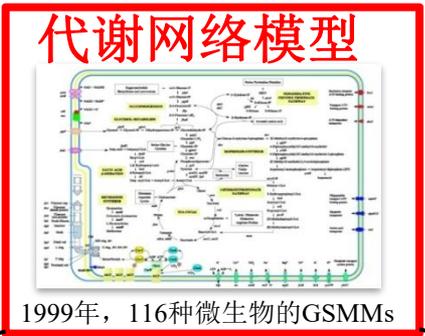
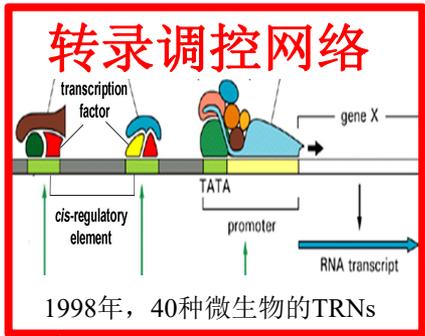


结合功能基因组
模拟代谢网络



基因组规模生物网络的发展

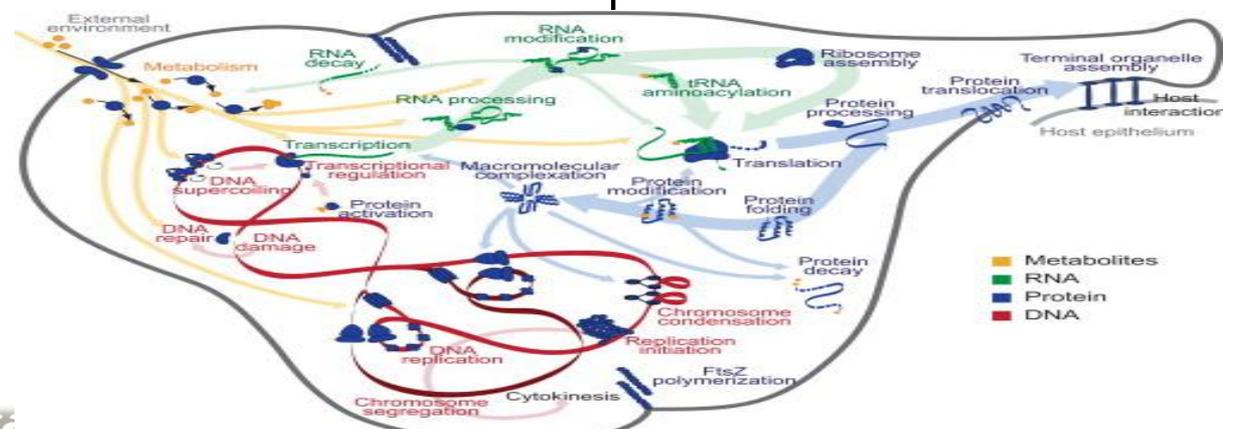
层级1: 物种特异性模型



层级2: 以 GSMs 为中心



层级3: 全细胞网络模型





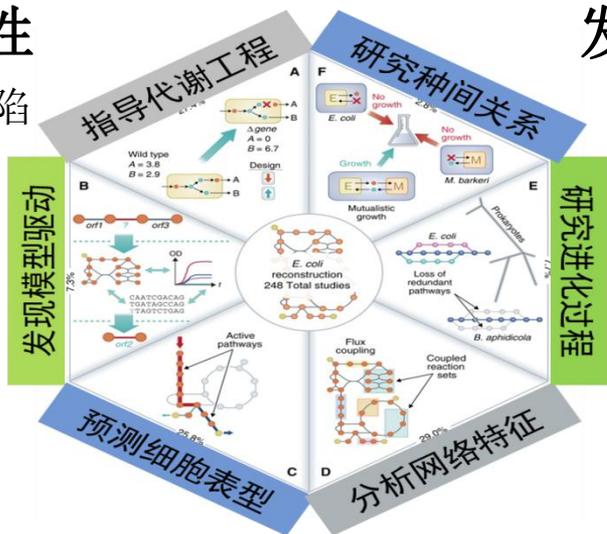
基因组规模生物网络模型在工业微生物中的应用

解析工业微生物的生理特性

- 鉴定目标微生物底物谱和营养缺陷
- 预测微生物细胞生长能力
- 鉴定微生物必需基因与必需反应
- 解析微生物非代谢生化过程

发展生产性能优化策略

- 鉴定产物合成模块
- 促进产物高效合成
- 确保产物合成鲁棒性
- 调节产物合成模块的兼容性



在发酵工业中的应用

鉴定与**高产性状**相关的分子标记和基因组排列特征

确定与**副产物形成**相关的基因组冗余片段

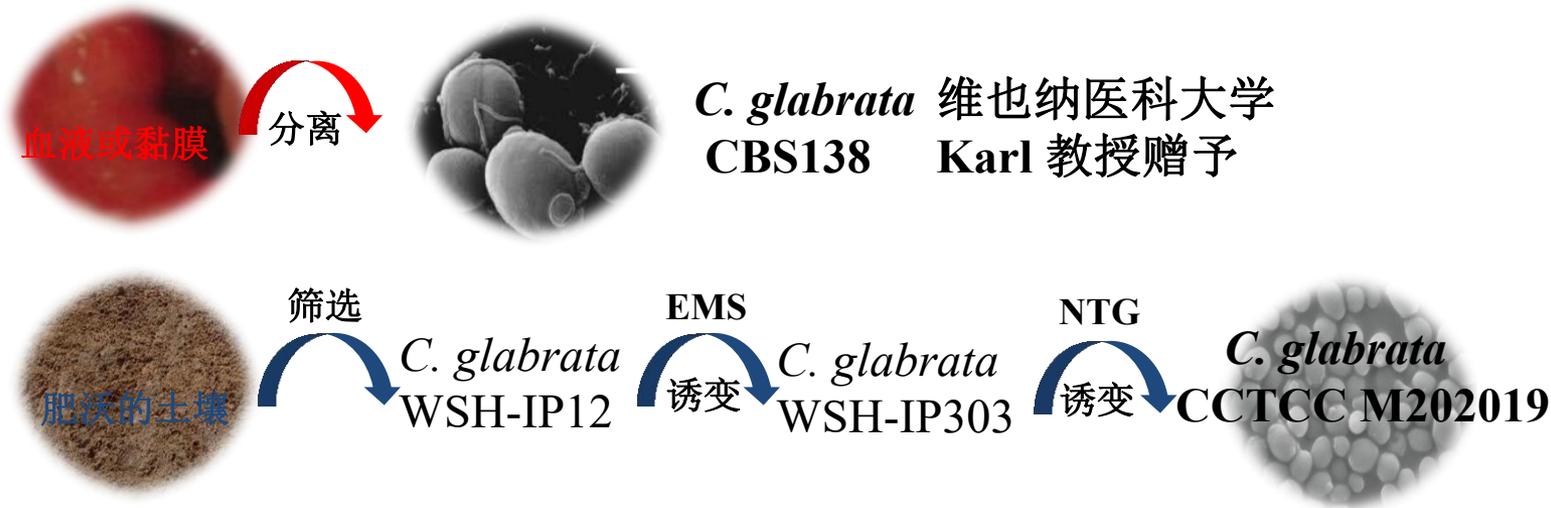
阐释与**生产效率**相关的最佳代谢及关键调控元件

解析与**目标产物最终浓度**密切相关的全局调控组合



光滑球拟酵母(*Candida glabrata*)生理特性的研究进展

光滑球拟酵母的来源



光滑球拟酵母生理性能

菌种选育

- 强化丙酮酸合成
- 削弱丙酮酸分解
- 解除代谢产物反馈
- 增强菌株环境耐受性

营养条件优化

- 营养物质的优化
- 中间代谢物的补给

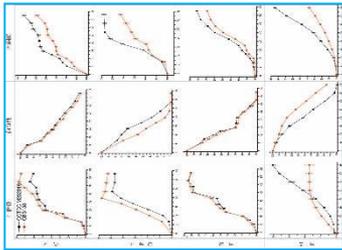
能量代谢改造

- 理性降低胞内ATP水平
- 改造胞内还原力水平





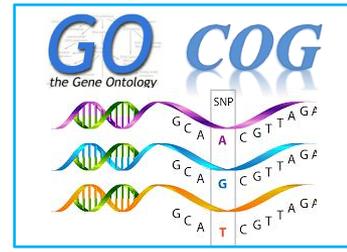
第二章光滑球拟酵母全基因组测序与特征分析



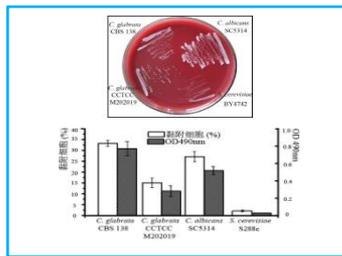
两菌生产表型差异



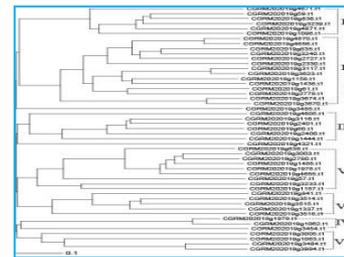
全基因组测序



两菌基因组比较



实验验证毒性表型差异

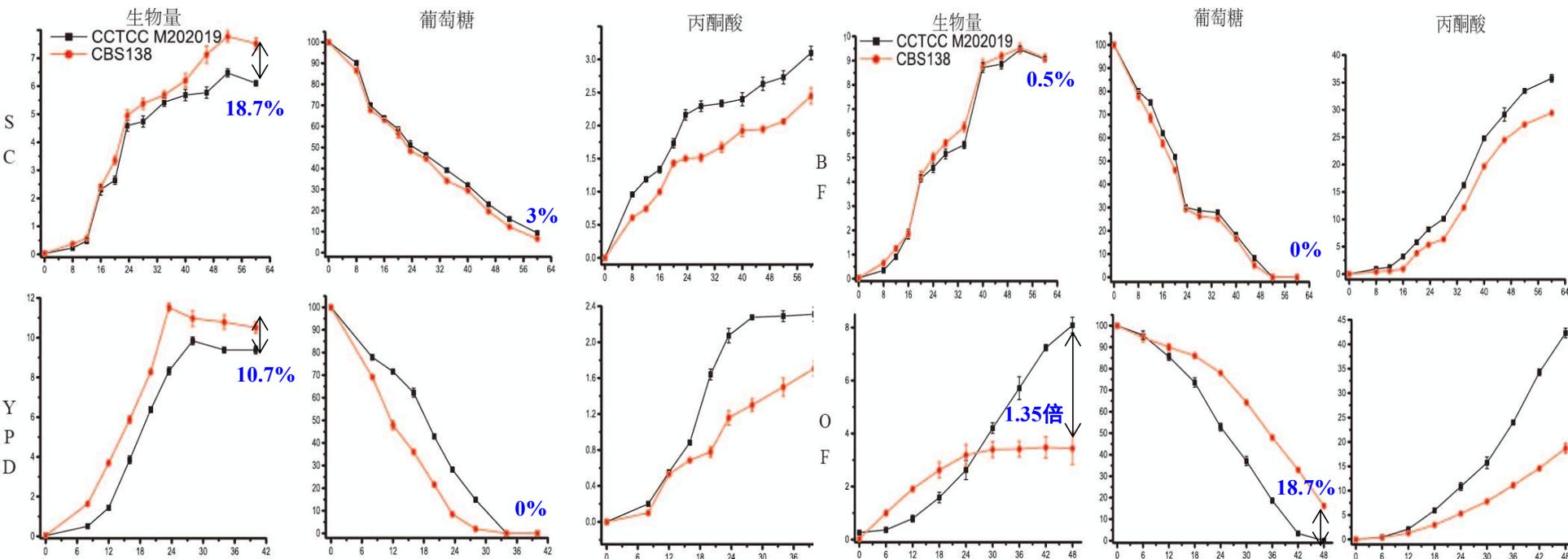


遗传信息深入解析

利用基因组测序和比较基因组分析，鉴定 *C. glabrata* CCTCC M20219 高产丙酮酸和低毒性相关的基因特征。



丙酮酸发酵性能评估



参数	全合成培养基(SC)		酵母浸出粉葡萄糖培养基(YPD)		基础发酵培养基(BF)		优化发酵培养基(OF)	
	M202019	CBS138	M202019	CBS138	M202019	CBS138	M202019	CBS138

培养时间 (h)	60	60	40	40	60	60	48	48
----------	----	----	----	----	----	----	----	----

***C. glabrata* CCTCC M202019 的丙酮酸发酵性能高于CBS138**

葡萄糖消耗速率 ($\text{g}\cdot\text{L}^{-1}\cdot\text{h}^{-1}$)	1.51	1.56	2.50	2.50	1.67	1.67	2.08	1.75
丙酮酸产量 ($\text{g}\cdot\text{L}^{-1}$)	3.10	2.45	2.31	1.71	35.76	29.45	42.32	18.70
丙酮酸生产强度 ($\text{g}\cdot\text{L}^{-1}\cdot\text{h}^{-1}$)	0.052	0.041	0.058	0.043	0.60	0.49	0.88	0.40
丙酮酸得率 ($\text{g}\cdot\text{g}^{-1}$)	0.034	0.026	0.23	0.17	3.58	2.95	0.42	0.22



全基因组测序及基本特征

总体特征

基因组大小(Mb)	12.1
GC含量(%)	38.47
Contigs数目	111
Contig N50	659495
Scaffolds数目	74
Scaffold N50	775409
重复序列长度(%)	1.15

基因注释特征

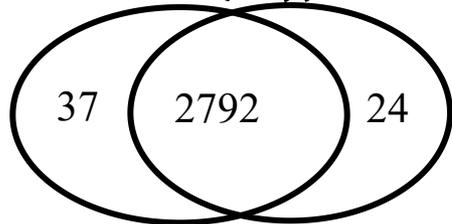
编码蛋白	5345
tRNA 基因	191
rRNA 基因	6
有EC号基因	961



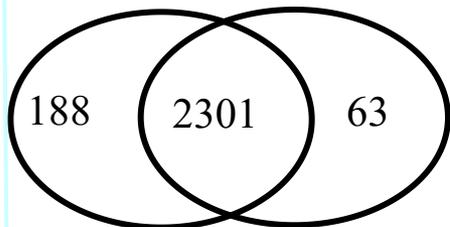
第二章光滑球拟酵母全基因组测序与特征分析

功能基因组比较

GO注释



KOG分类



M202019 CBS138

基于BLASTP基因同源性

相似度	M202019	CBS138
100%	94.6%	95.6%
[90%,1)	275	159
(0,90%)	16	43

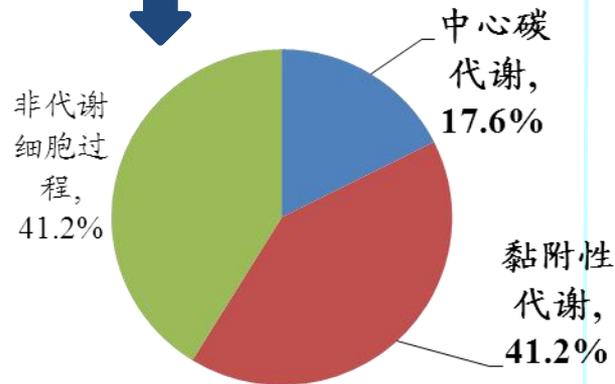
74.3%与黏附性代谢相关

单核苷酸多肽性位点

205个SNPs

144个
51个基因
编码区

61个
25个基因的
调控区



两菌基因组高度相似，与功能代谢相关的差异集中在中心碳代谢和黏附性代谢



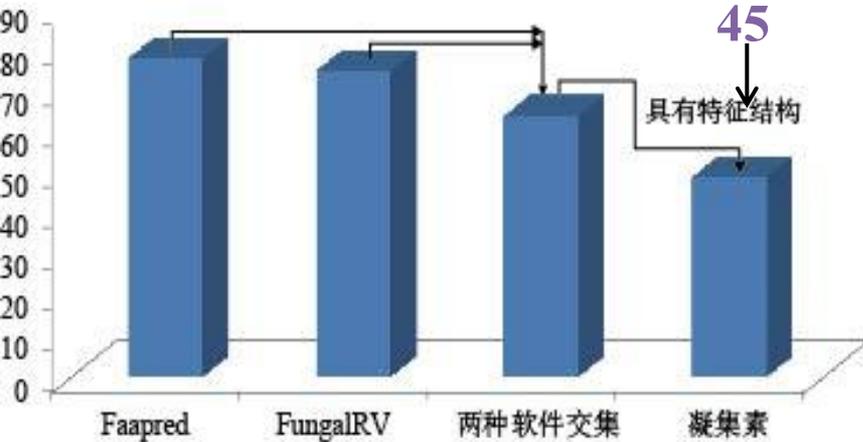
□ M202019 丙酮酸积累相关基因特征

代谢亚系统	功能	基因
转运	葡萄糖	<i>hxt3</i> 和 <i>hxt 4/6/7</i>
	烟酸	<i>trn1</i> 和 <i>trn2</i>
	乙酸	<i>CAGL0M03465g</i>
	二元羧酸	<i>CAGL0J04114g</i>
氧化磷酸化	细胞色素C 氧化酶	<i>cox1</i> 、 <i>cox2</i> 、 <i>cox7a</i> 、 <i>cox7c</i> 和 <i>cox17</i>
	细胞色素C 还原酶	<i>cytb</i> 和 <i>ocr10</i>
	F-类型 ATP 酶	<i>atp8</i> 、 <i>j</i> 和 <i>k</i>
	ETC 复合体 IV	<i>pet309</i>
下游代谢	乙酰CoA 水解酶	<i>ach1</i> 和 <i>gln1</i>
	谷氨酸合成酶	<i>glt1</i>

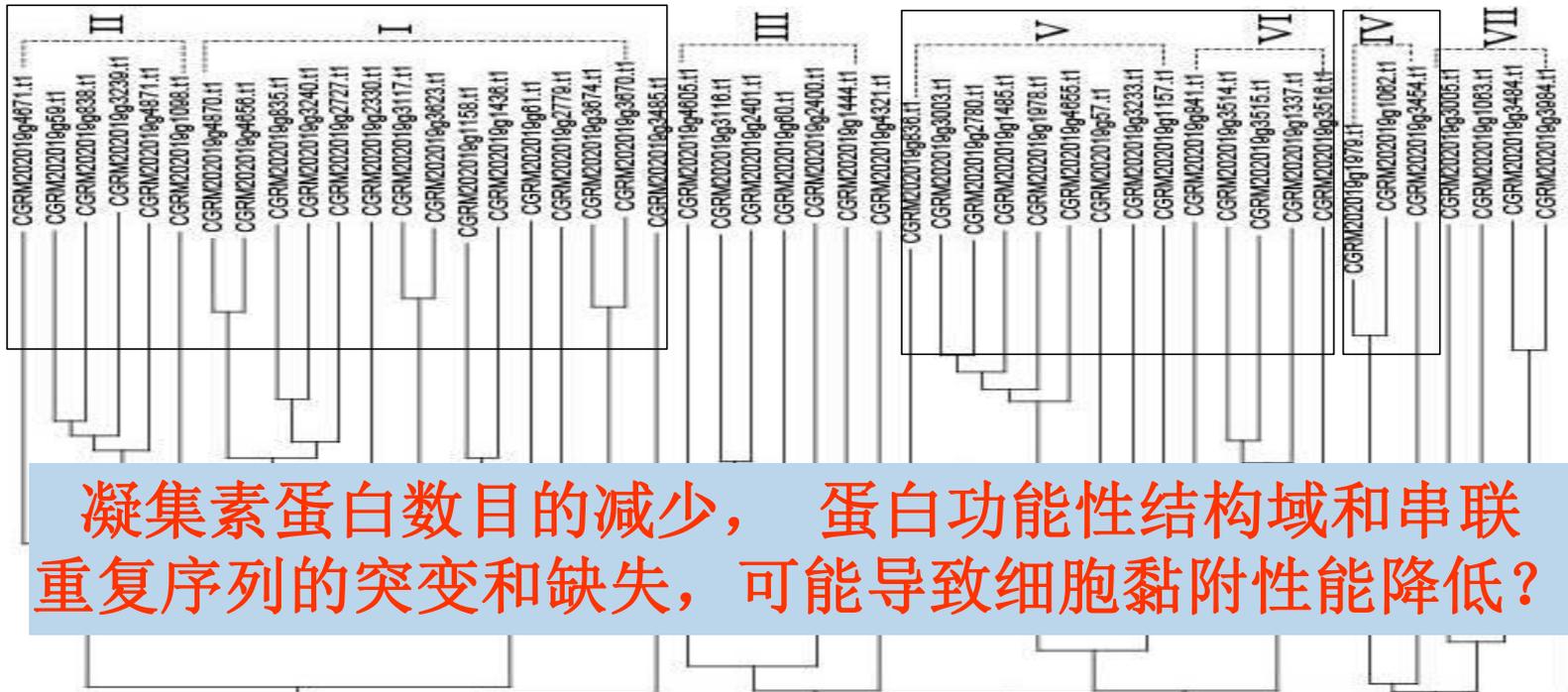
中心碳代谢中基因差异集中在丙酮酸的合成、分解相关反应



□ M202019黏附性相关的基因特征

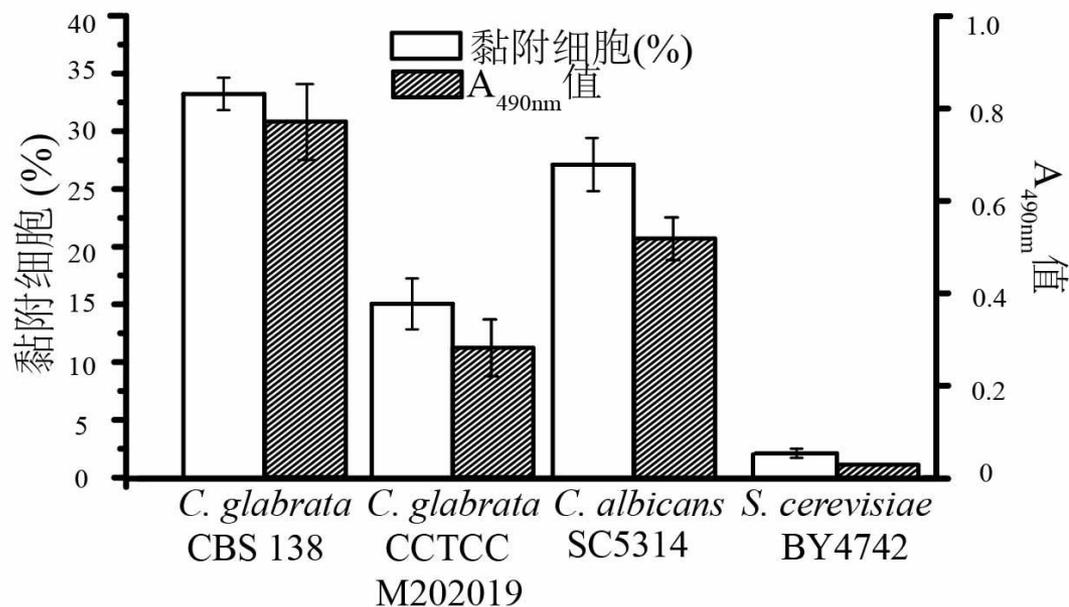
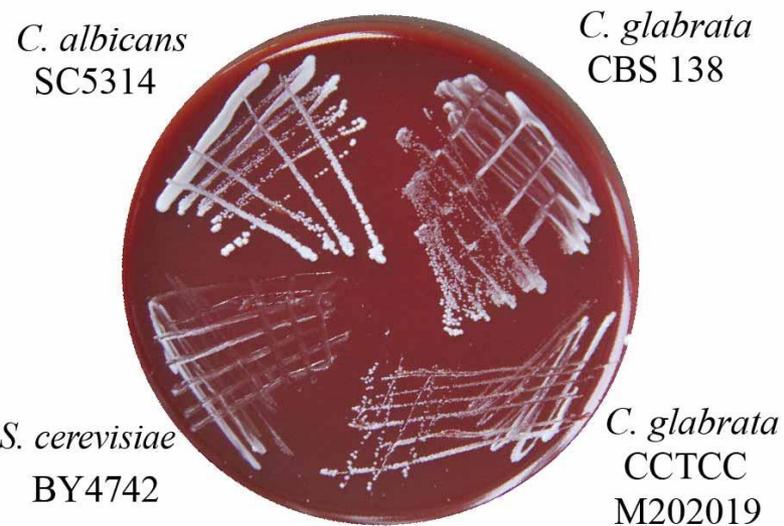


- PA14及其类似的结构域
- 磷脂酰肌醇锚蛋白(GPI anchor)
- 菌丝调节胞壁蛋白(Hyphal_reg_CWP)
- Flocculin重复序列
- 富含S/T的低复杂度重复序列 (III、VII)
- 保守的(V)SHITT/TTVVT氨基酸模体 (I、III、IV、V)





M202019 生产安全性评价



***C. glabrata* CCTCC M202019 的细胞毒性和黏附性均弱于CBS138, M202019更具生物安全性**

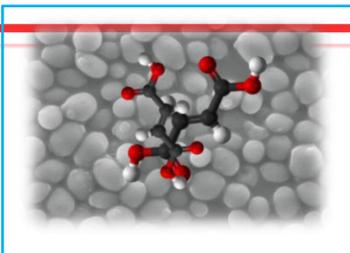


本章小结(第二章)

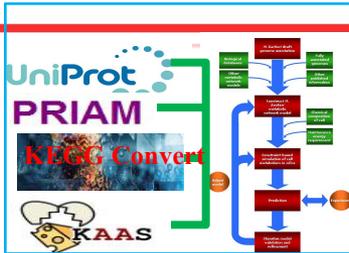
1. 运用二代高通量测序技术，解析了丙酮酸生产菌株 *C. glabrata* CCTCC M202019 全基因组序列，其特征包括12.1Mbp的基因组、38.47%的GC含量、5345个基因、191个tRNA、6个rRNA和1.15%的重复序列等；
2. 基于与 *C. glabrata* CBS138 的基因组比较，发现两菌基因组具有高度相似性，而差异主要分布在①中心碳代谢中营养物质和二羧酸的转运、氧化磷酸化和丙酮酸下游碳代谢；②黏附性代谢的凝集素蛋白的低复杂度重复区和功能结构域的变化；
3. 通过发酵实验和细胞毒性实验，证明了 *C. glabrata* CCTCC M202019 比 CBS138 菌株具有更强的丙酮酸生产能力和更弱的黏附性和细胞毒性。



第三章光滑球拟酵母的GSMM iNX804的构建与应用



已测序的基因组

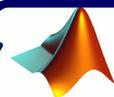


粗模型的构建

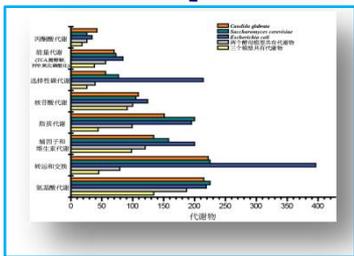


模型的精细化

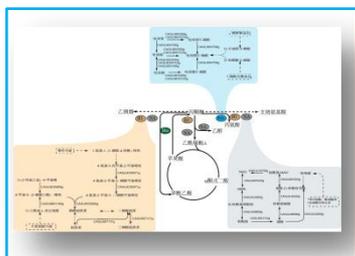
补gaps



功能性模型



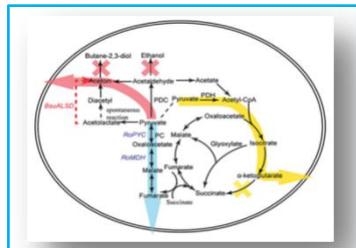
网络分析



途径挖掘



生理表型



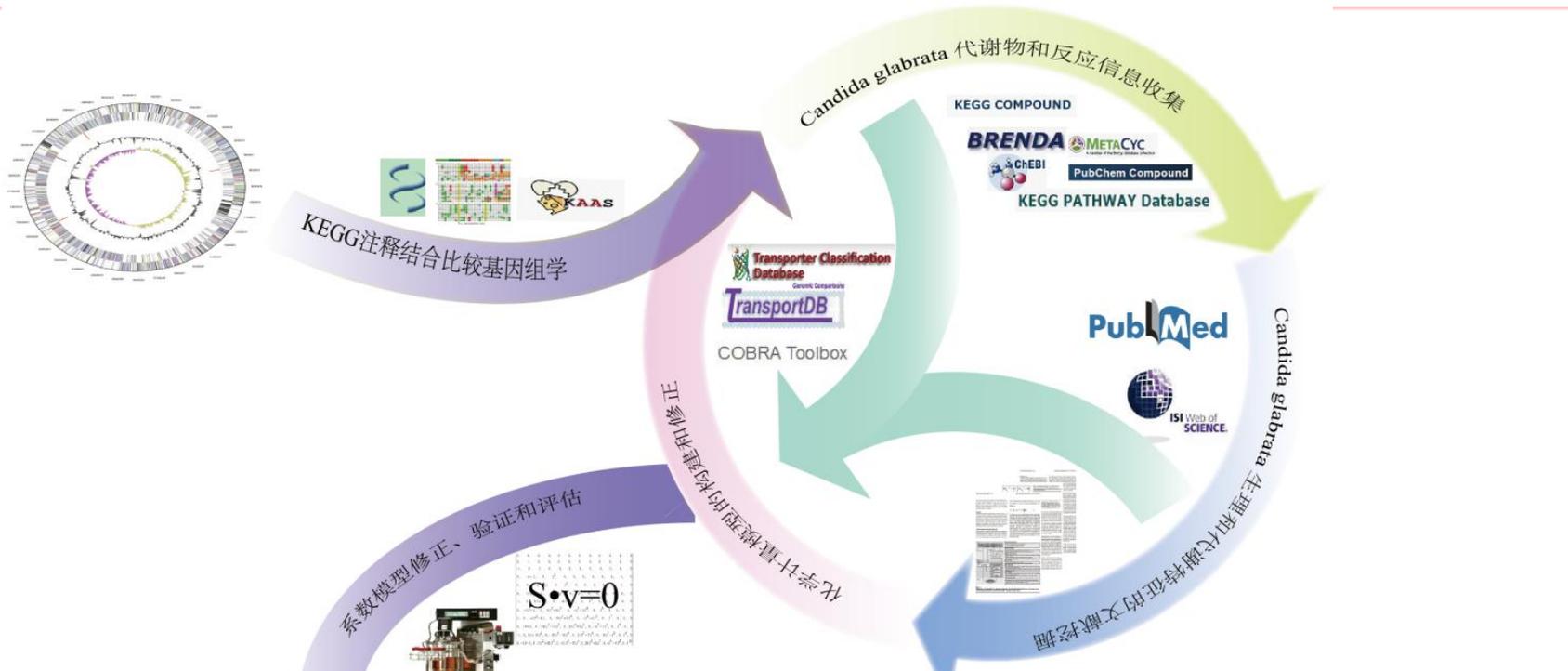
代谢改造

从*C. glabrata*蛋白序列出发，构建基因组规模代谢网络模型，解析其生长、生理性能，预测*C. glabrata*作为细胞工厂的潜能。

子应建注 激心到上



模型iNX804的构建流程和基本特征



模型iNX804
804个基因
1025个代谢物
1287个反应



细胞器	反应	代谢物	基因
细胞质	645	579	523
线粒体	165	238	154
胞外区间	216	108	82
过氧化物酶体	39	62	32
高尔基体	2	12	2
液泡	3	25	3
交换空间	217		121



模型*iNX804*准确性评价

① 60种底物利用的定性模拟

底物种类	一致性	<i>C. glabrata</i> 底物利用特点
40种碳源	95%	碳源利用谱相对较窄
20种氮源	100%	仅赖氨酸和半胱氨酸不支持生长

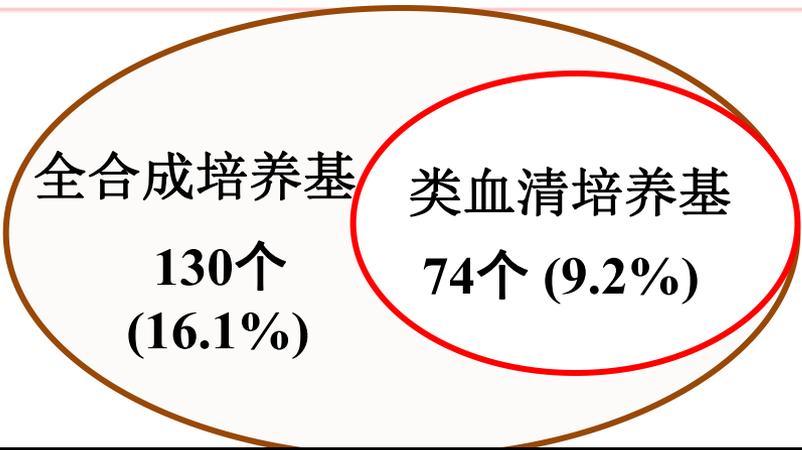
② *C. glabrata* 细胞生长的定量模拟

限制 ($\text{mmol}\cdot\text{gDW}^{-1}\cdot\text{h}^{-1}$)						生长速率 (h^{-1})	
GUR	OUR	CER	PPR	EPR	GPR	实验值	模拟值
4.15	13.61	10.02	0.74	0.36	-	0.31	0.32
6.43	21.77	9.37	0.54	5.15	-	0.44	0.44
8.71	14.31	10.04	2.84	6.81	-	0.40	0.42
11.02	6.31	18.93	0.007	18.33	1.62	0.12	0.13

模型*iNX804*能够准确模拟*C. glabrata*的生长代谢



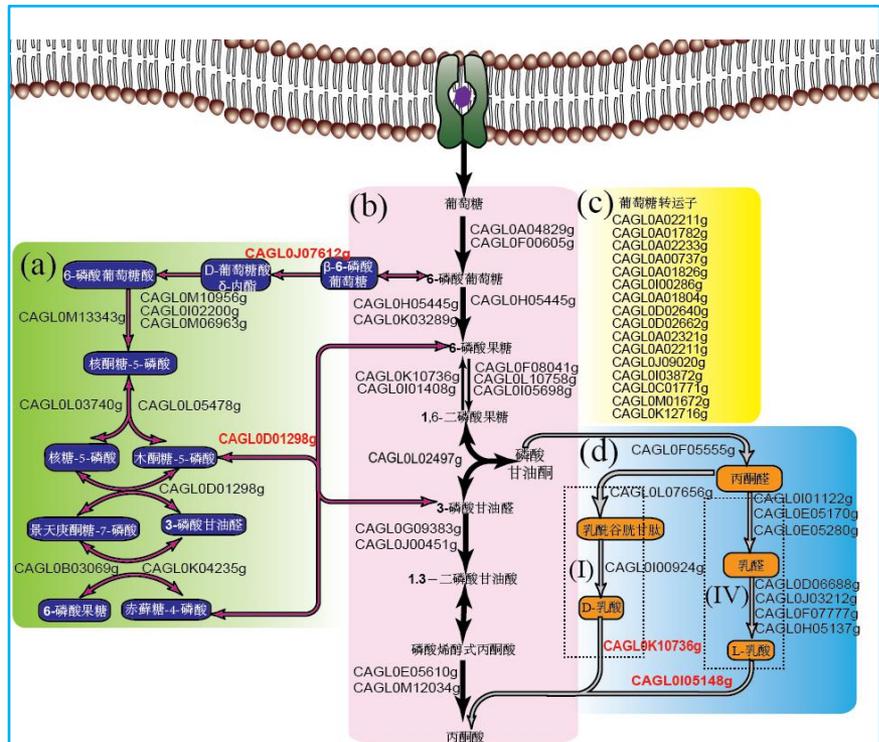
□ 生长必需基因的预测和验证



代谢途径	药物靶点	基因	M1	M2	代谢途径	药物靶点	基因	M1	M2
	几丁质合成酶	CAGL0B04389g	NE	NE	类固醇生物合成	麦角甾醇	CAGL0F01793g	NE	NE
		CAGL0I04840g					CAGL0E04334g	E	NE
		CAGL0A02904g					CAGL0M07656g	NE	NE
细胞壁合成	1,3-葡聚糖合成酶FKS1组分	CAGL0G01034g	NE	NE			CAGL0M07095g	E	NE
							CAGL0D05940g	NE	NE
							CAGL0H04653g	NE	NE
							CAGL0L00319g	E	NE
	1,6-葡聚糖合成酶	KRE9, KNH1			细胞色素P450 51		CAGL0E04334g	E	NE
							CAGL0D04114g		
						角鲨烯单加氧酶	CAGL0D05940g	NE	NE
叶酸代谢	二氢叶酸还原酶	CAGL0J00385g;	NE	NE	其他	胞质和线粒体辅酶A合成酶	CAGL0L00649g	NE	NE
		CAGL0J03894g;					CAGL0B02717g		
		CAGL0L11044g				过氧化氢酶	CAGL0K10868g	NE	NE
艾利希途径	二氢叶酸合成酶	CAGL0J07920g	E	E	核苷酸代谢	胸苷酸合成酶	CAGL0K05467g	E	E
	芳香氨基转移酶	CAGL0G01254g	NE	NE					



基于模型 iNX804 对丙酮酸代谢途径解析



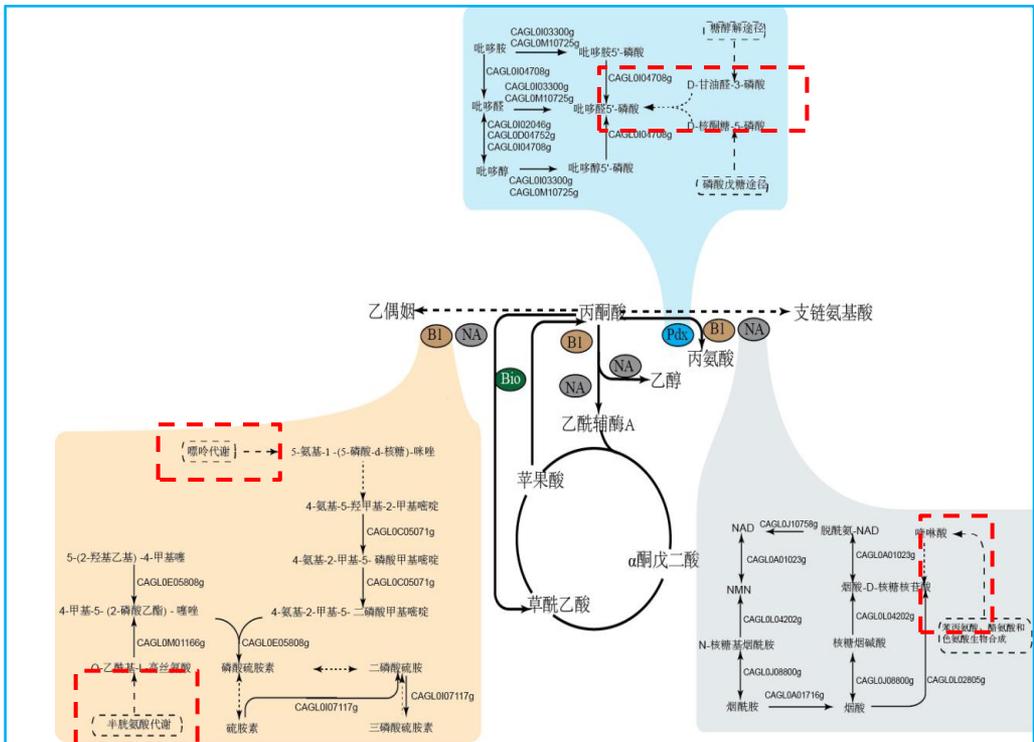
丙酮酸合成途径

高效的葡萄糖转运能力

从葡萄糖出发三条丙酮酸合成途径

糖酵解、磷酸戊糖途径、丙酮醛途径

被“弱化”的丙酮酸降解



丙酮酸分解代谢

VB₆、VB₁、NA、Bio合成途径不完整

甘油醛-3-磷酸和-核酮糖-5-磷酸 \rightarrow 5-磷酸-吡哆醛

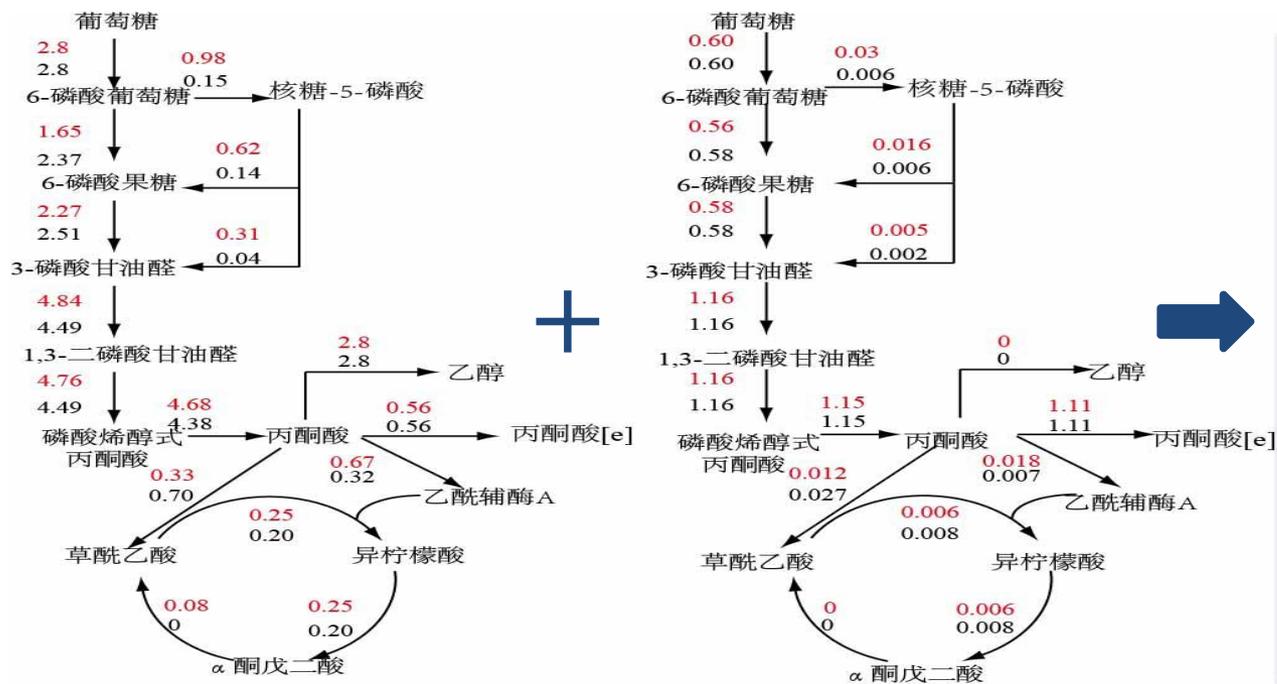
嘧啶代谢和半胱氨酸代谢 \rightarrow 二磷酸硫胺素

喹啉酸 \rightarrow 核糖核苷酸





丙酮酸合成的模拟和预测



- ✓ 生长阶段，33.3%的碳流量通过EMP途径流向乙醇，10%流向丙酮酸；
 - ✓ 产酸阶段，92.5%的碳流量从葡萄糖经过EMP途径合成丙酮酸。
- 糖酵解是丙酮酸合成的主要途径**

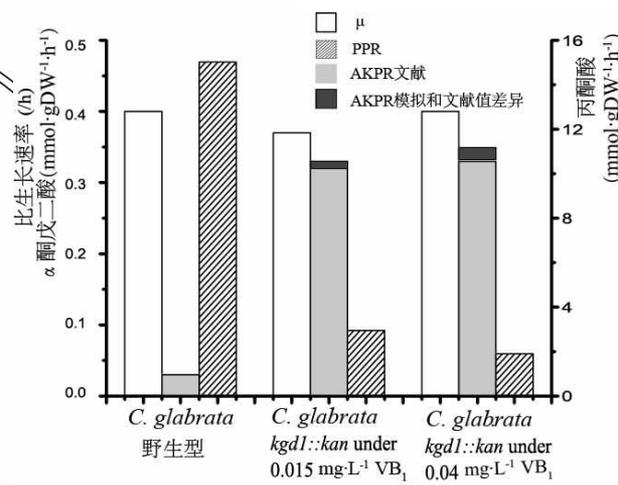
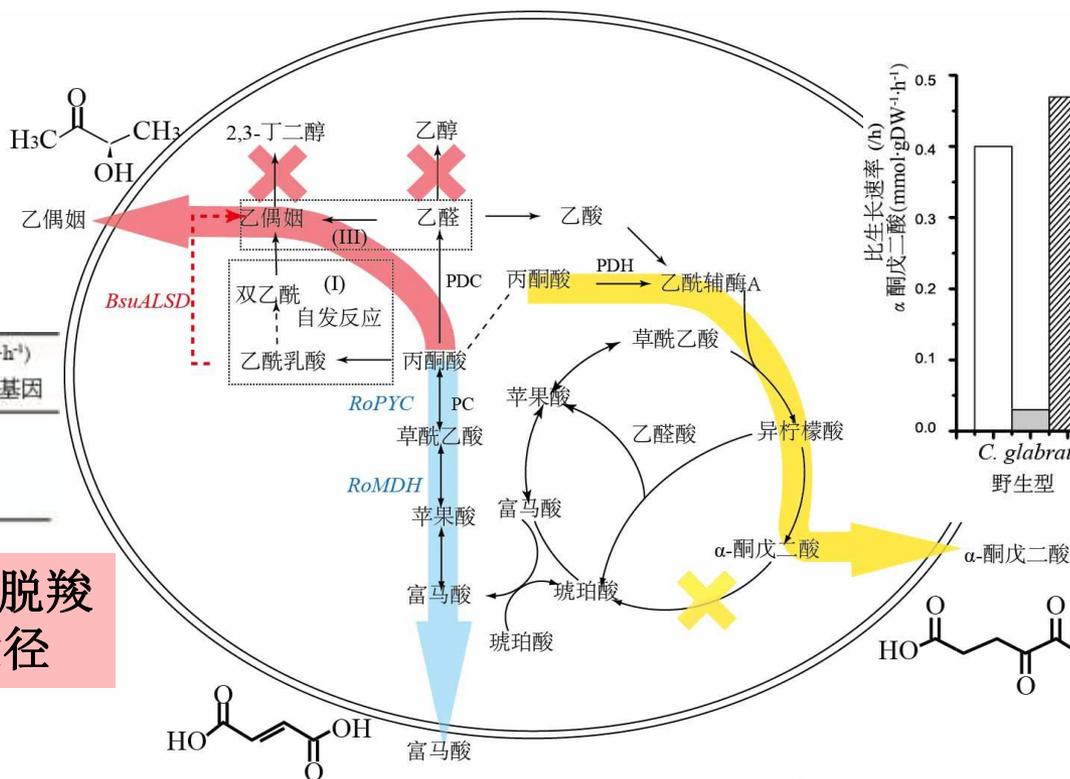
菌株	生物量(h ⁻¹)	丙酮酸(mmol·gDW ⁻¹ ·h ⁻¹)	乙醇(mmol·gDW ⁻¹ ·h ⁻¹)
对照	0.19	10.66	5.19
Δzwf	0.19	0	16.98
$\Delta tk1$	0.19	0	15.59
$\Delta zwf\Delta tk1$	0.19	13.62	2.94

敲除丙酮酸和其他两类合成途径连接处非必需基因，可增强丙酮酸合成



精细化学品的代谢途径设计与构建

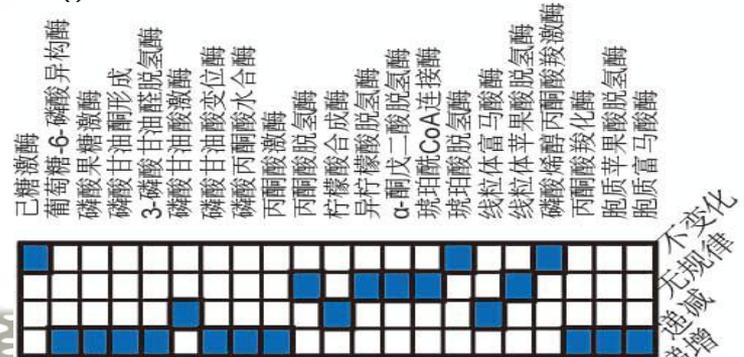
乙偶姻生产 (mmol·gDW ⁻¹ ·h ⁻¹)		
目标函数	消除副产物	表达alsD基因
生物量形成	0	1.01
乙偶姻交换	6.51	7.83



抑制 α -酮戊二酸脱氢酶、增加硫胺素含量和胞内ATP水平

异源表达 α -乙酰乳酸脱羧酶构建丙酮酸裂解途径

富马酸产量会随着糖酵解途径和胞质还原路径而递增



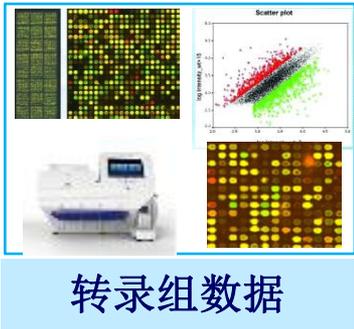
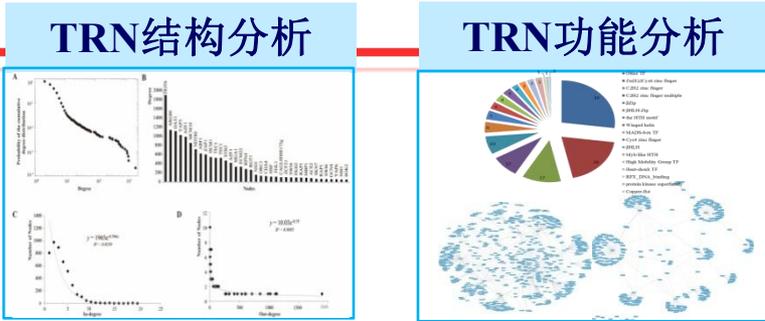


本章小结(第三章)

1. 采用半自动的模型构建方法，构建了*C. glabrata*的GSMM模型*iNX804*，包括804个基因、1023个代谢物和1284个生化反应；
2. 基于*C. glabrata*对60种底物利用和细胞生长速率的定量模拟，评价模型准确性；
3. 运用必需基因分析，预测*C. glabrata*在全合成培养基和类血清培养基上的130个和74个必需基因，部分已被验证为*C. glabrata*药物靶点；
4. 利用模型*iNX804*解析和优化*C. glabrata*生产丙酮酸、 α -酮戊二酸、富马酸、乙偶姻生产策略，并得到湿实验验证。



第四章光滑球拟酵母转录调控网络的构建与分析



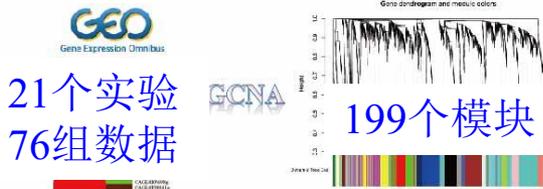
基于C. glabrata大量转录组数据，构建C. glabrata转录调控网络，结合代谢模型探索细胞生长代谢和毒性代谢的转录调控关系。



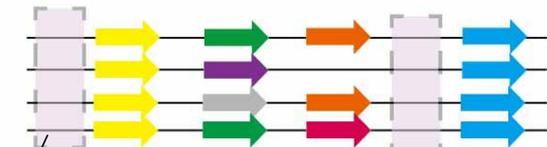
转录调控网络模型的构建

基于基因表达数据从头反向推测

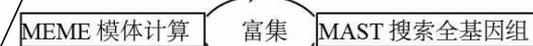
鉴定基因共表达模块



提取共调节基因模块上游1000bp区域序列



查找转录因子结合位点(TFBS)



0	0.04	0.92	0.04
0.04	0.92	0	0.04
0	0	1	0

3164个转录模体

0.08	0.04	0	0.88
0.08	0.15	0.27	0.5

同公开TFBS数据库比较注释模体

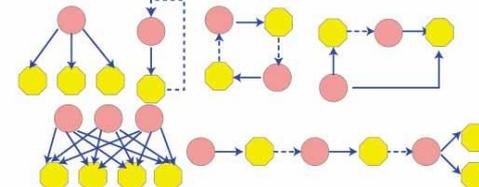


有参照TRN的比较基因组构建

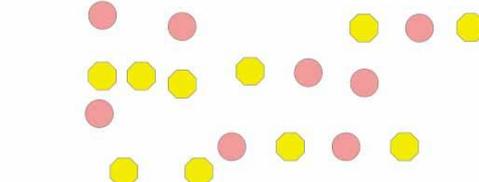
与目标微生物亲缘关系近
参照模型应该由实验方法构建

Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*

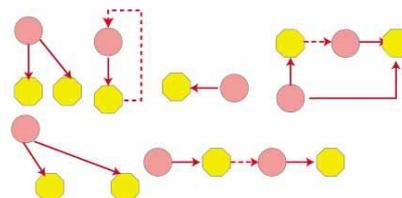
Url: papers.lee.inh.ac.uk/200709-094/



同源比对模板基因组和目标基因组



依据同源比对结果确定转录调控关系



97个TFs和
1425个TGs组
成2294组转
录调控关系

评价和修正转录调控网络

根据GO功能注释的网络内聚性评价

$$P\text{-value} = \sum_{i=1}^T \frac{\binom{B}{i} \binom{G-B}{T-i}}{\binom{G}{T}}$$

依据文献报道验证和修正TRN

启动子结构 基因调控关系 Chip-seq等湿实验 基因表达数据 相同功能代谢途径

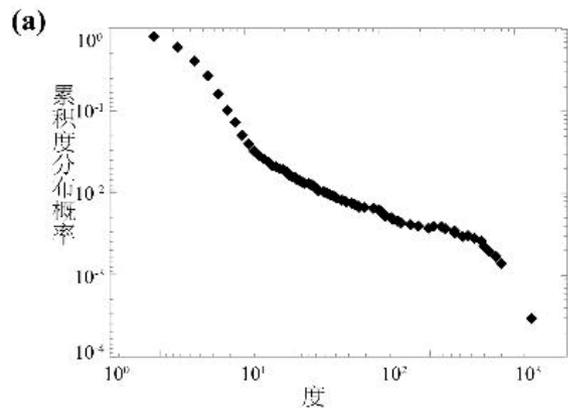


C. glabrata 145个转录因子和3239个靶基因间6655种相互作用

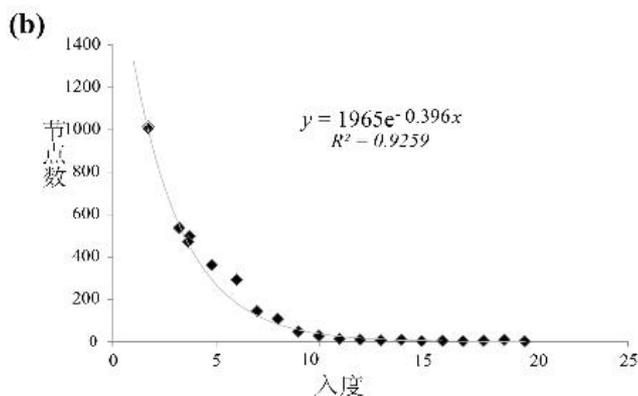


转录调控网络模型的基本特征

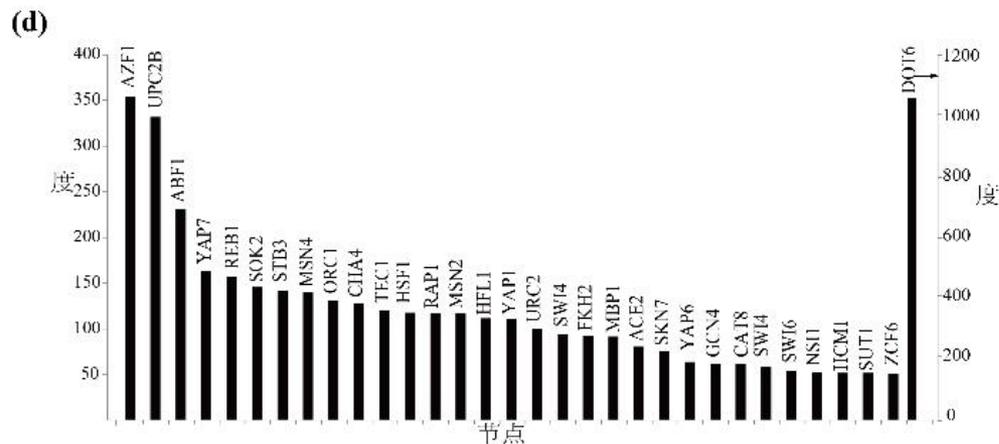
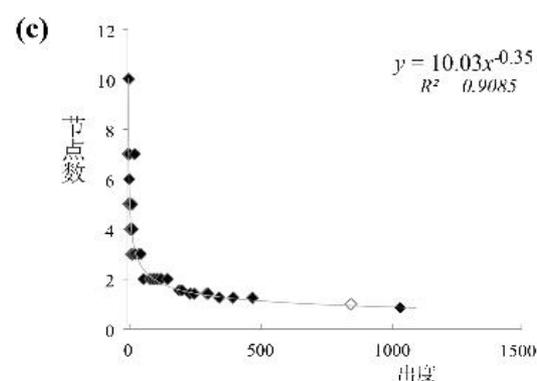
累积度分布服从幂律分布



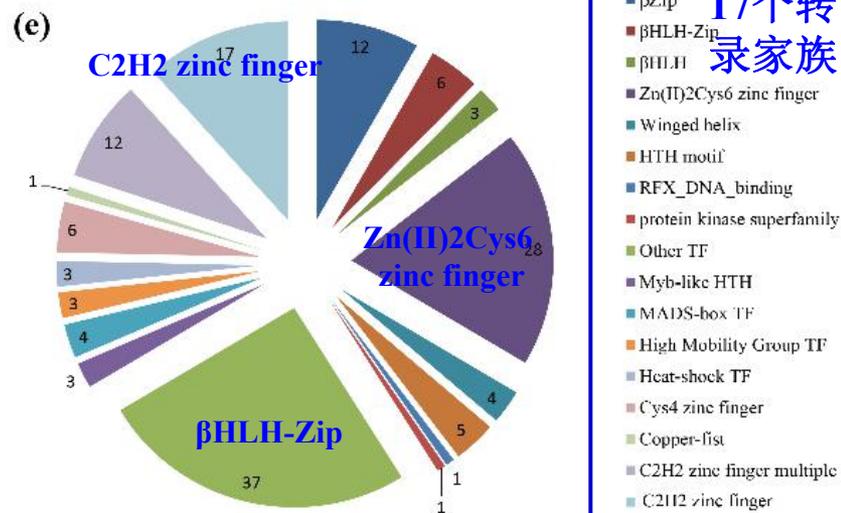
入度连接性呈指数分布



出度连接性呈自由度分布



连接度大于50的节点全部为转录因子



73.3%的TFs调控细胞代谢



转录调控功能性模块的分析

转录模块	GO分类	GO分类基因数	模块内基因数	交集	P值
碳代谢	糖酵解过程	16	939	14	1.93e-09
氮代谢	蛋氨酸生物合成	9	396	6	9.76e-06
胁迫响应	真菌型细胞壁组织	33	1648	19	8.81e-04
固醇代谢	麦角甾醇生物合成	18	348	4	5.22e-03
毒性代谢	对药物的反应	51	319	8	8.73e-03
菌丝形成	核糖体大亚基合成	21	1459	12	3.07e-03

<0.01

***C. glabrata* 转录调控功能网络具有显著的网络内聚性**

- ### 碳代谢
- 葡萄糖效应(8)
 - 中心碳代谢 (3)
 - 肌醇代谢 (3)
 - 糖转运 (1)
 - 多糖降解 (1)

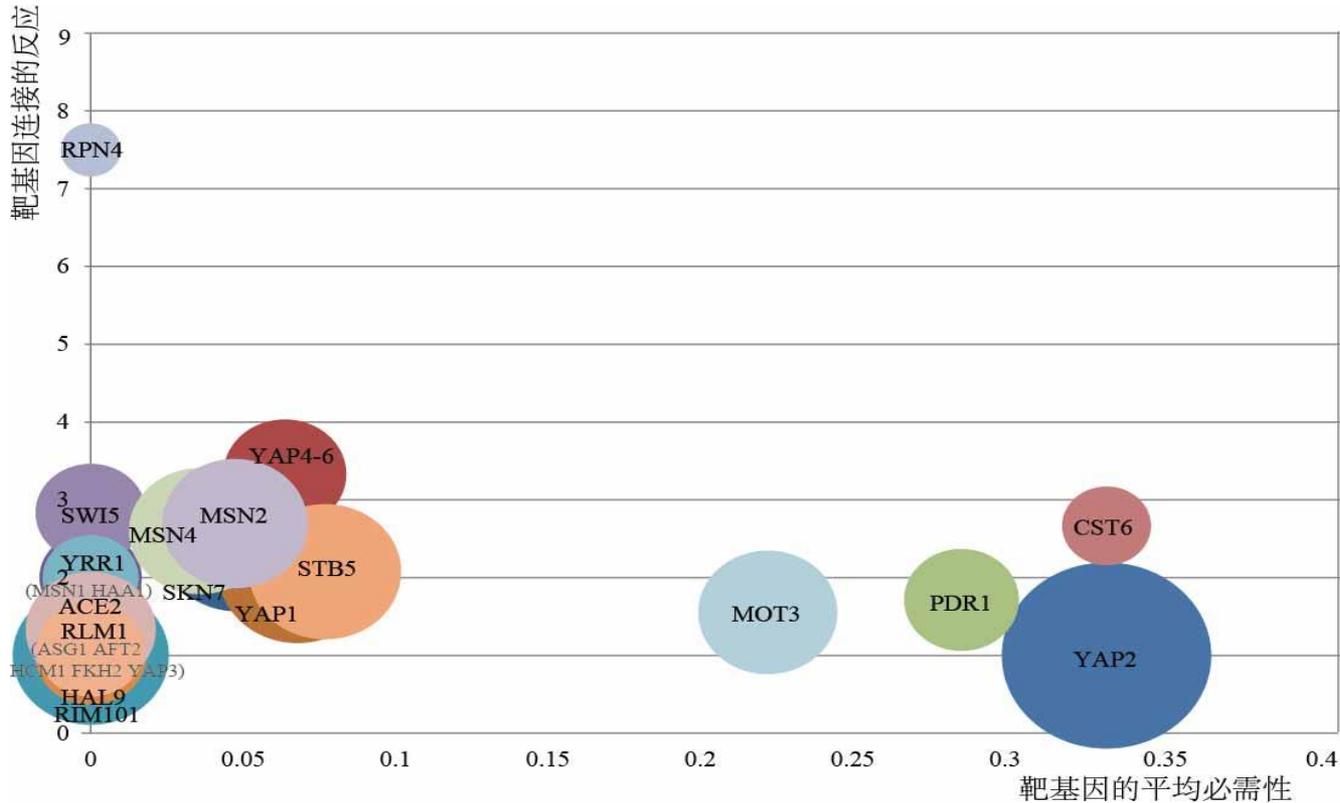
- ### 氮代谢
- 氮源阻遏效应(3)
 - 通用氨基酸合成(6)
 - 氨基酸转运(5)
 - 特定氨基酸代谢(7)

- ### 胁迫响应
- 酸性(5)
 - 高渗(6)
 - 氧化 (5)
 - 磷酸盐(2)
 - 铁离子(3)
 - 烟酸(1)

- ### 其他代谢
- 假菌丝形成(8)
 - 嘧啶代谢(2)
 - 脂质代谢(4)
 - Zn、Cu离子(4)
 - 毒性代谢 (14)



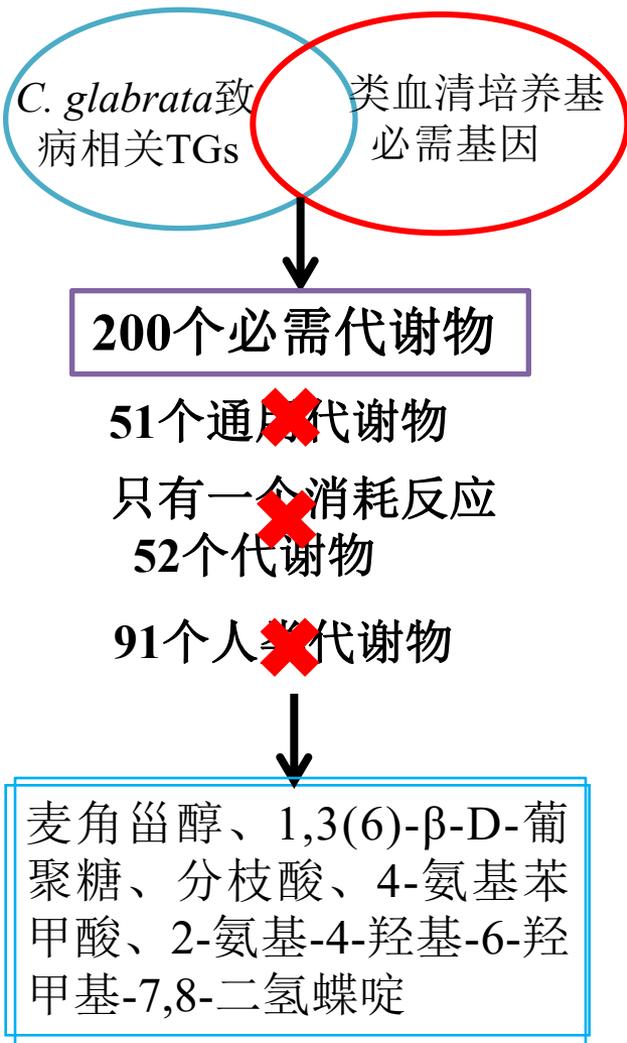
□ *C. glabrata* 致病性转录调控因子评价



- ① TGs在*C. glabrata* GSMM *i*NX804中比例(一半TFs的TGs>20%，*YAP2*的TGs 60%);
- ② TGs连接的反应数(22个TFs<4个，例外*RPN4*);
- ③ TGs的平均必需性(82.6% TFs < 0.15，*YAP2*和*CST6*第一，其次是*PDR1*和*MOT3*。



光滑球拟酵母药物靶点的预测



代谢物	酶	代谢途径	转录因子& 功能
ERGOST	24(241)-固醇还原酶	类固醇合成	PDR1-毒性
13GLUCAN	1,3- β -葡聚糖合成酶	细胞壁的生物合成	DAT1-胁迫, AZF1-碳源阻遏效应, SKN7-过氧化胁迫,
16GLUCAN	1,6- β -葡聚糖合成酶	细胞壁的生物合成	DAT1-胁迫, AZF1-碳源阻遏效应
CHOR	氨基脱氧分支酸合成酶	叶酸合成	DOT6-假菌丝
	邻氨基苯甲酸合成酶	苯丙氨酸、酪氨酸和色氨酸的代谢	DOT6, TEC1-假菌丝, UPC2B-固醇合成, CRZ1-钙信号, GCN4-胁迫
PABA	二氢叶酸合成酶	叶酸合成	DOT6, TEC1-假菌丝
AHHMP	二氢叶酸合成酶	叶酸合成	TEC1-假菌丝
	二氢新蝶呤醛缩酶	叶酸合成	TEC1-假菌丝



本章小结(第四章)

1. 采用从头反向工程和同源比较方法，首次构建了*C. glabrata*基因组规模转录调控模型，包括145个转录因子和3239个靶基因间6655种相互作用；
2. *C. glabrata*的TRN中145个转录因子分布于17个转录家族，73.3%的转录因子调控细胞代谢过程；
3. *C. glabrata*转录调控子网络的功能性具有显著的网络内聚性(P值在 $8.73e-03$ 和 $1.93e-09$ 之间)；
4. 结合*C. glabrata*的GSMM *iNX804*，针对致病性转录调控模块，得到麦角甾醇、1,3- β -D-葡聚糖、1,6- β -D-葡聚糖、分枝酸、4-氨基苯甲酸和2-氨基-4-羟基-6-羟甲基-7,8-二氢蝶啶这6种必需代谢物相关的8种酶可作为有效的药物靶点。