

多菌株大肠杆菌代谢网络揭示环境适应

报告人 刘浩

翻译 罗志辉

Nov 26, 2014

问题背景

- 大肠杆菌作为一种模式生物在微生物实验中已做了大量的研究
- 随着测序技术的发展和應用，大肠杆菌的测序结果为大肠杆菌的研究带来了新的问题
- K-12 MG1655型大肠杆菌菌株是微生物试验中使用比较多的一种大肠杆菌菌株
- 最新的测序结果表明这种非致病性菌株只是大肠杆菌这个物种众多众多基因组中的一部分

问题背景

- 比如，O157:H7 EDL933型菌株的基因组大小比K-12 MG1655型的多出大约一百万碱基对（大约20%）
- 多基因组测序结果已经定义了一套大肠杆菌中常见的基因，但它们只是大肠杆菌基因池中的一小部分
- 越来越多的测序结果为我们带来了一个问题：什么是一个菌株？什么是一个物种？

方法的提出

- 最近通过对20个大肠杆菌菌株的研究发现这些菌株所共有的基因组元件中有很大大一部分与代谢功能有关
- 因此，通过描述编码核心代谢功能的基因可以帮助我们定义大肠杆菌这个物种核心代谢能力的通用特征
- 重建代谢网络是研究生物体间代谢基因组多样性的一种有效工具
- K-12 MG1655菌株是目前代谢重建最完整的菌株
- 因为越来越多的测序结果表明K-12 MG1655菌株只是大肠杆菌这个物种的一部分代表
- 因此需要为其它大肠杆菌菌株建立代谢基因组模型

研究目的

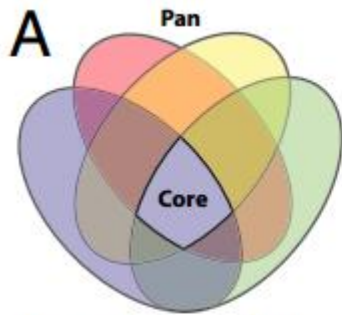
- 本研究的目的是为已被完整测序的大肠杆菌建立基因组模型，以此重建整个物种及特定菌株的代谢网络
- 16sRNA实验将志贺氏菌归类于大肠杆菌，因此已构建的模型覆盖了共生菌株、肠道内以及肠道外致病性菌株，这允许我们对大肠杆菌这个物种的代表性代谢能力进行综合的分析

研究结果

1. 核心代谢和全覆盖代谢内容的描述
2. 对不同营养物的分解能力区分不同菌株代谢模型
3. 不同基质区分致病性和共生（非致病性）菌株
4. 代谢模型结合Gap-Filling法促进对营养缺陷型菌株基因基础的研究
5. 单营养实验确认表明模型具有很高的准确性

核心代谢和全覆盖代谢内容的描述

- 55个大肠杆菌菌株基因组的重建可用于比较不同物种间基因、化学反应以及代谢内容的比较
- 根据代谢内容在所有菌株中出现的情况，将代谢内容分为核心代谢和全覆盖代谢
- 核心代谢代谢内容是所有菌株所有共有的，同样地，所有菌株总的代谢内容定义了全覆盖代谢，即所谓的交集与并集的关系

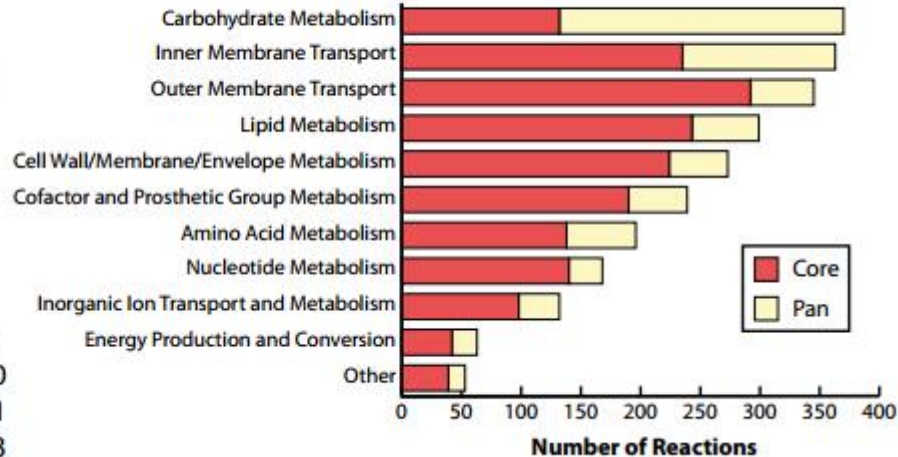


For 55 unique *E. coli* strains

	Core:	Pan:
Genes:	965	1,460
Reactions:	1,773	2,501
Metabolites:	1,665	2,043

B

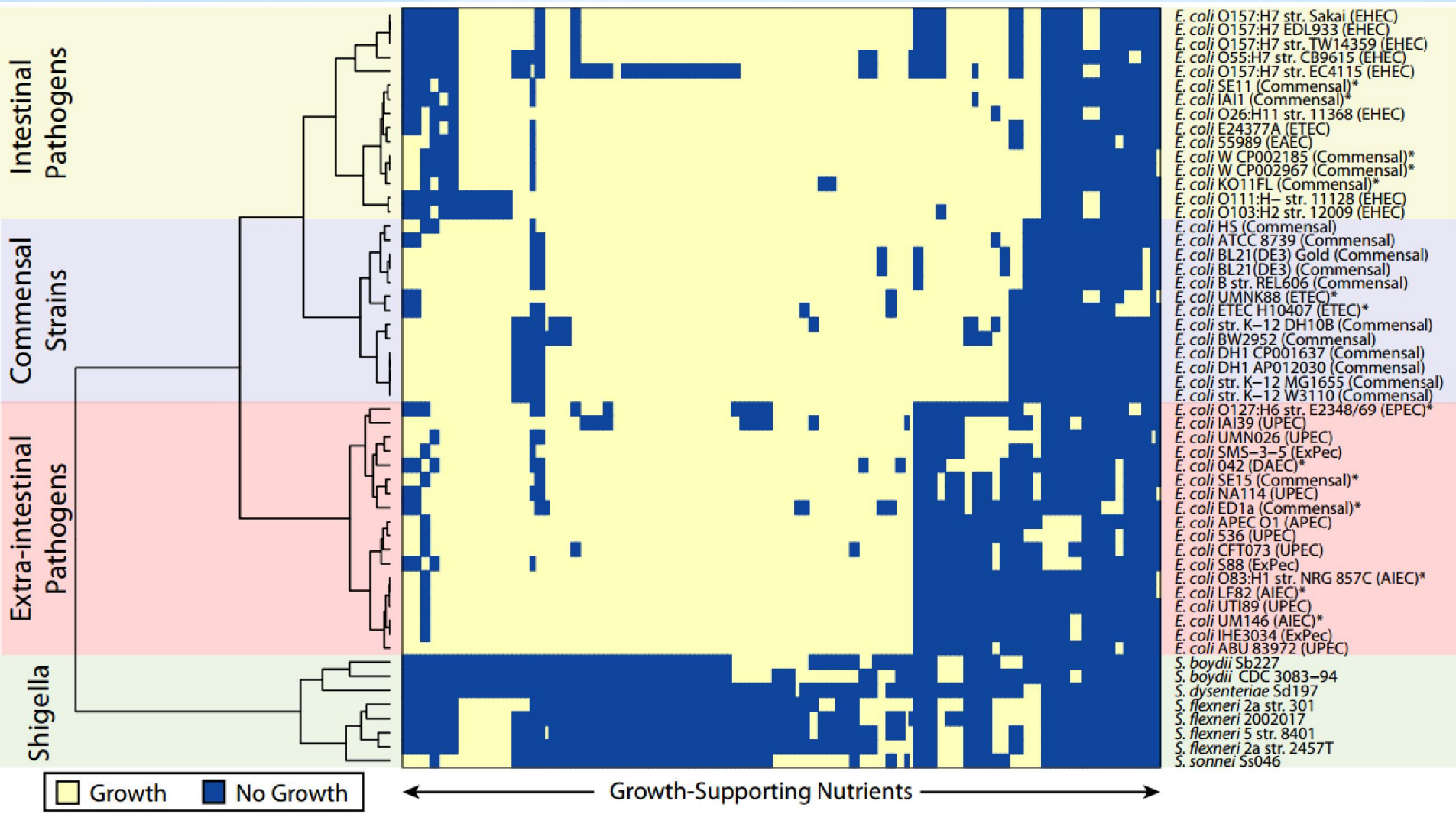
Reaction Distribution by Subsystem



- 脂类代谢、细胞壁、细胞膜、囊膜代谢，核酸代谢，辅因子以及辅基代谢是最保守的子系统，与这些代谢有关的反应占到了所有核心反应的80%，相反，碳水化合物代谢中的反应只有36%的是核心反应
- 而碳水化合物中的反应有64%属于全覆盖代谢，是所有全覆盖代谢的反应中最高的

对不同营养物的分解能力区分不同菌株代谢模型

- 将静态代谢网络重建转换为可用于计算的数学模型可用于以每个重建内容以为基础的表型计算
- 属于不同碳源代谢子系统的反应占了pan反应的大部分，因此，可以假设这些能力可以反映不同E. coli菌株对不同营养环境适应能力上的功能性差异
- 对这55个E. coli和Shigella模型在基本培养基上进行了654个不同生长条件的生长模拟，这些条件由唯一的生长支持物碳、氮、磷、硫和有氧、无氧环境组成



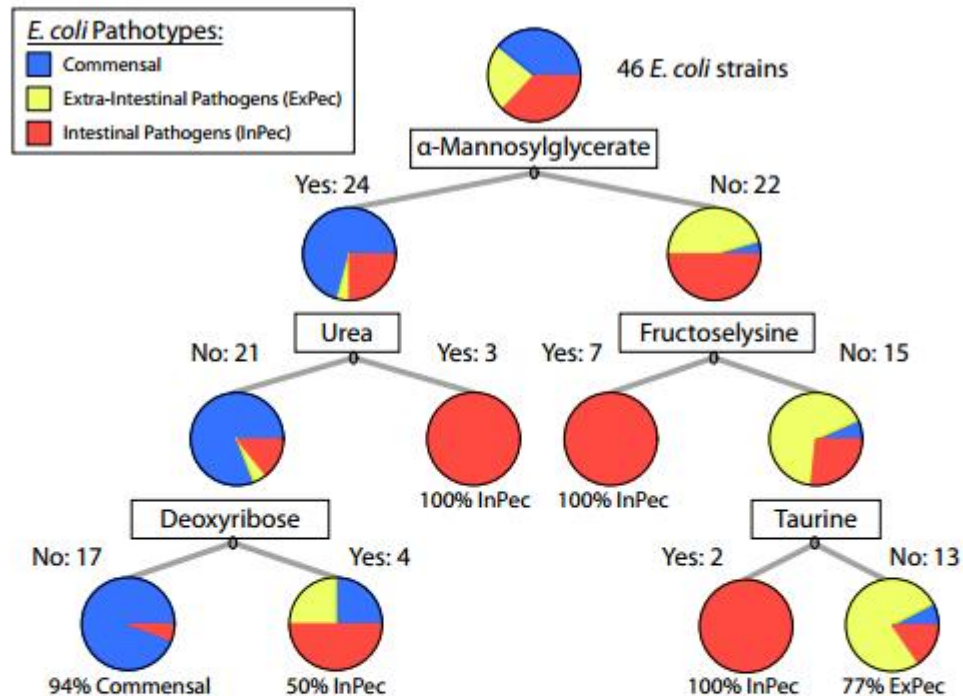


Fig. 3. Classification of *E. coli* pathotypes based on growth-supporting conditions. Growth-supporting nutrients were used to create a classification tree. This tree can be used to determine if an *E. coli* strain is commensal, an intestinal pathogen, or an extraintestinal pathogen. For example, following the tree to the right shows that 77% of *E. coli* strains that cannot grow on α -mannosylglycerate, fructoselysine, or taurine as sole carbon sources are expected to be extraintestinal pathogens. Thus, a small number of nutrient sources can be used to classify *E. coli* strains of different types.

不同基质区分致病性和共生（非致病性）菌株

- 根据模拟生长表型，研究人员从肠道内和肠道外的E. coli中观察到了共生菌株的普遍分离，这表明建立以代谢能力为基础的分类模式是可能的
- 为此，研究人员利用统计学方法分别对共生菌株和肠道外致病性菌株以及共生菌株和肠道内致病性菌株进行了比较分析

不同基质区分致病性和共生（非致病性）菌株

Table 1. Nutrients predicted to give commensal strains or extraintestinal pathogenic strains of *E. coli* a catabolic advantage

Source	Commensal (%)	ExPec (%)	<i>P</i> value
ExPec strain nutrients			
3-Phospho-D-glycerate	0	36	1.3E-2
L-Arginine	11	64	5.0E-3
Cellobiose	33	82	1.3E-2
<i>N</i> -acetyl-D-galactosamine	67	100	3.9E-2
Commensal strain nutrients			
Fructoselysine	89	0	2.2E-6
Psicoselysine	89	0	2.2E-6
Dopamine	89	0	2.2E-6
Phenethylamine	89	0	2.2E-6
Tyramine	89	0	2.2E-6
Phenylacetaldehyde	72	0	1.2E-4
α -Mannosylglycerate	94	9	5.7E-6
4-Hydroxyphenylacetate	56	9	1.3E-2
Cyanate	83	27	3.8E-3
Melibiose	78	27	9.7E-3
Phenylpropanoate	72	27	2.0E-2
3-(3-Hydroxy-phenyl)propionate	89	36	5.0E-3
3-Hydroxycinnamic acid	89	36	5.0E-3

不同基质区分致病性和共生（非致病性）菌株

Table 2. Nutrients predicted to give commensal strains or intestinal pathogenic strains of *E. coli* a catabolic advantage

Source	Commensal (%)	InPec (%)	<i>P</i> value
InPec strain nutrients			
Sucrose	33	65	5.0E-2
Raffinose	28	65	2.6E-2
L-Arginine	11	65	1.3E-3
D-Arabitol	0	24	4.6E-2
Ribitol	0	24	4.5E-2
Agmatine	0	47	1.0E-3
Urea	0	47	1.0E-3
Commensal strain nutrients			
Galactonate	100	65	7.6E-3
α-Mannosylglycerate	94	35	2.6E-4
Dopamine	89	41	3.5E-3
Phenethylamine	89	41	3.5E-3
Tyramine	89	41	3.5E-3
5-Dehydro-D-gluconate	78	24	1.6E-3
L-Idonate	78	24	1.6E-3
D-Allose	78	41	2.5E-2
Butyrate	78	47	5.0E-2
Acetoacetate	78	47	5.0E-2
4-Hydroxyphenylacetate	56	18	2.0E-2

代谢模型结合Gap-Filling法促进对营养缺陷型菌株基因基础的研究

- GEMs除了能用于对支持生长的营养物质的研究，还可用于对特定菌株营养缺陷型遗传学基础的研究
- 55个重构模型中的20个在M9葡萄糖培养基上不能产生关键的有机物
- 通过检测，8个Shigella菌落表现出了对维生素B3的营养缺陷型。模拟结果与文献结果一致，表明很多Shigella菌落不能在添加维生素B3的M9培养基中生长
- Gap分析表明这种营养缺陷型是由于天冬氨酸氧化酶的缺失，由nadB基因编码，对nadB基因的生物信息学分析表明，与E. coli相比，这是一个由于大量同义替换突变造成的假基因。

Strain	Folate	Niacin	L-Met	L-Trp	L-Leu	Thiamin
<i>Escherichia coli</i> CFT073				○		
<i>Escherichia coli</i> IAI39	○					○
<i>Escherichia coli</i> K-12 DH10B		○			○	
<i>Escherichia coli</i> UMN026		○				
<i>Shigella boydii</i> CDC 3083-94		○				○
<i>Shigella boydii</i> sb227	○	○				○
<i>Shigella dysenteriae</i> SD 197	○				○	
<i>Shigella flexneri</i> 5 str. 8401		○				
<i>Shigella flexneri</i> 2002017	○	○				
<i>Shigella flexneri</i> 2a str. 2457T	○	○				
<i>Shigella flexneri</i> 2a str. 301	○	○	○			
<i>Shigella sonnei</i> Ss046	○	○				

Fig. 4. Model predicted strain-specific auxotrophies. GEM predicted minimal media conditions for each auxotrophic strain. *Shigella* strains lack essential vitamin biosynthesis capabilities for niacin (vitamin B3), thiamin (vitamin B1), and folate (vitamin B9). Other strains have lost biosynthetic pathways for the essential amino acids methionine, tryptophan and leucine, thus becoming auxotrophic for these compounds.

单营养实验确认表明模型具有很高的准确性

- 为了验证模型的准确性，最后用实验进行验证
- 研究人员挑选了55种重建模型的12种进行6种营养物质的生长。生长情况通过接种48小时后用光密度进行评估，光密度大于0.08说明生长了。这12中菌落包括3肠道外致病性菌落、3个肠道内致病性菌落，5个共生菌落以及1个Shigella菌落，覆盖了上面所讨论到的表现型。
- 模型由真阳性和真阴性确认，即实验结果和模型的一致。相反，假阳性和假阴性表明模型的潜在错误。实验结果表明GEM预测结果与实验结果一致性的准确度达到了80%。

A

Tested 12 strains on 6
discriminating carbon substrates
(72 growth screens)
Overall accuracy = 80%

		GEM computation	
		+	-
Experimental observation	+	1) True Positive 51% (37)	2) False Negative 14% (10)
	-	3) False Positive 6% (4)	4) True Negative 29% (21)

B

Compound	Genes	Commensal Strains					Intestinal Pathogens			Extra-intestinal Pathogens		Shigella	Substrate accuracy	
		MG1655	8739	DH1	HS	KO11FL	O42	EDL933	Sakai	SMS3-5	CFT073	UMN026		2457T
Acetoacetate	<i>atoE, atoAD</i>	TP	TP	TP	TP	TN	TP	TN	TN	TP	TP	TP	TN	100%
Deoxyribose	<i>deoC, deoP, deoK</i>	TN	TN	TN	TN	TN	TP	TN	TN	TN	TP	TN	TN	100%
D-malate	<i>dctA, yeaU</i>	TP	TP	TP	FP	TP	TP	TP	TP	TP	FP	TP	TN	83%
Ferric dicitrate	<i>fecBCDE</i>	TP	TP	TP	TN	TN	TP	TN	TN	FN	TN	FP	TN	83%
Melibiose	<i>mela, melB</i>	TP	TP	FP	TP	TP	FN	TP	TP	TP	FN	TP	TP	75%
2x Phenyls	<i>tynA, feaB, paaK</i>	TP	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN	FN	42%
Strain accuracy:		100%	100%	83%	83%	100%	67%	83%	83%	67%	50%	67%	83%	

结论

- 大肠杆菌菌株的代谢基因模型是一种有力的工具，可用于：
 1. 比较分析大肠杆菌物种内不同菌株的核心和全覆盖代谢能力，阐述大肠杆菌对不同营养物质的适应性
 2. 区分共生、肠道内致病性、肠道外致病性大肠杆菌菌株
 3. 预测大肠杆菌营养缺陷型，并结合其它方法对其进行遗传基础研究

总结

- 该文献的创新点在于提出了利用计算模拟的方法对大肠杆菌分类极其生长条件的分析，其优点在于能同时处理大量菌株在大量不同营养条件下的表现，最后只需要用少量实验来验证计算模拟的准确性
- 对我的启发是，我们做生物信息平时处理了大量的数据，但很少有比较好的生物学问题，需要在这方面有所加强，好的方法和好的问题才能做出比较有意思的研究
- 对实验的改进是，对在实验部分提到的实验验证中出现的假阳性和假阴性没有进行进一步的挖掘，这个这个的挖掘研究可能发现比较好的生物学问题

谢谢