

StrainPanDA: 使用泛基因组分解技术对宏基因组数据进行菌株组成和关联的基因成分谱的重建

胡函，谭宇翔，李陈浩，陈俊宇，寇岩，徐振江
刘洋彧，谭彦，戴磊

中科院深圳先进技术研究院合成生物学研究所定量合成中心
深圳未知君生物科技有限公司
哈佛医学院麻省总医院计算与整合生物中心
南昌大学食品科学与技术国家重点实验室
哈佛医学院和布莱根妇女医院



Hu, Han, Yuxiang Tan, Chenhao Li, Junyu Chen, Yan Kou, Zhenjiang Zech Xu, Yang-Yu Liu, Yan Tan, and Lei Dai. 2022. “StrainPanDA: Linked reconstruction of strain composition and gene content profiles via pangenome-based decomposition of metagenomic data.” *iMeta*. e41. <https://doi.org/10.1002/imt2.41>

简介

动机

- 微生物群落中，同一菌种内常常同时存在多个变异体菌株，这些变异体在基因成分上存在很大差别。
- 菌种内变异体可能有着显著的表型差异，并且在微生物跨环境适应性以及宿主微生物相互作用中扮演重要作用。

当前方法

- 大多数菌株分析工具关注使用单核苷酸变异信息对菌株组成进行分析。
- 当前的泛基因组工具比如PanPhlAn仅仅推导宏基因组样本中的主导菌株的基因成分。

解决方案

- 希望有一套方法，从宏基因组数据中同时对共存菌株的组成和对应的基因成分进行重建。

简介

StrainPanDA (Strain-level Pangenome Decomposition Analysis , 菌株水平泛基因组分解分析)

Gene family abundance data matrix (D)

	Sample1	Sample2	Sample3	... SampleS
GF 1	5	1	3	10
GF 2	1	2	1	7
GF 3	1	1	0	2
...	...	...	...	...
GF N	5	1	1	1

=

Gene content profile matrix (P)

	Strain1	Strain2	Strain3	... StrainK
GF 1	0	1	1	1
GF 2	1	1	1	1
GF 3	0	0	1	1
...	...	...	...	...
GF N	1	1	1	0

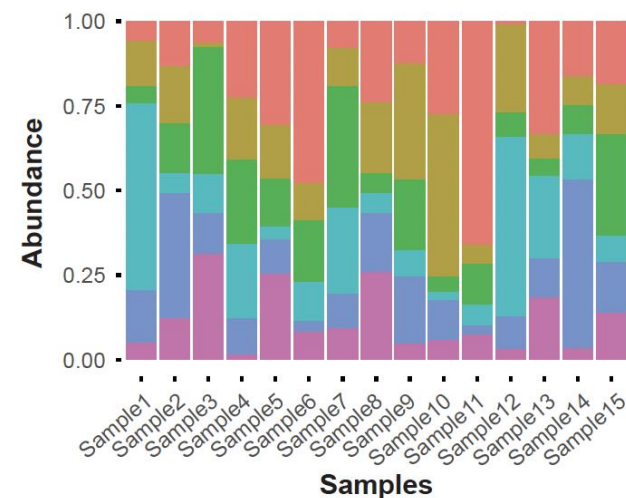
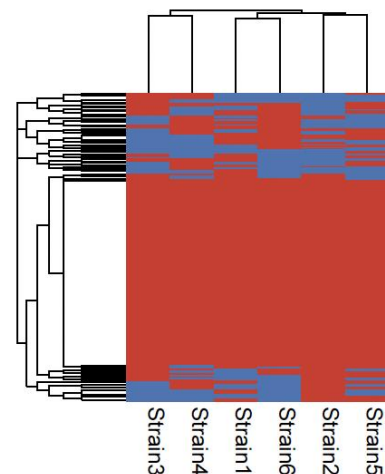
$\times$

Strain composition matrix (S)

	Sample1	Sample2	Sample3	... SampleS
Strain 1	0.1	0.2	0	0.1
Strain 2	0.1	0.1	0.3	0.4
Strain 3	0.3	0.3	0.3	0
...	...	...	...	...
Strain K	0.2	0	0.1	0

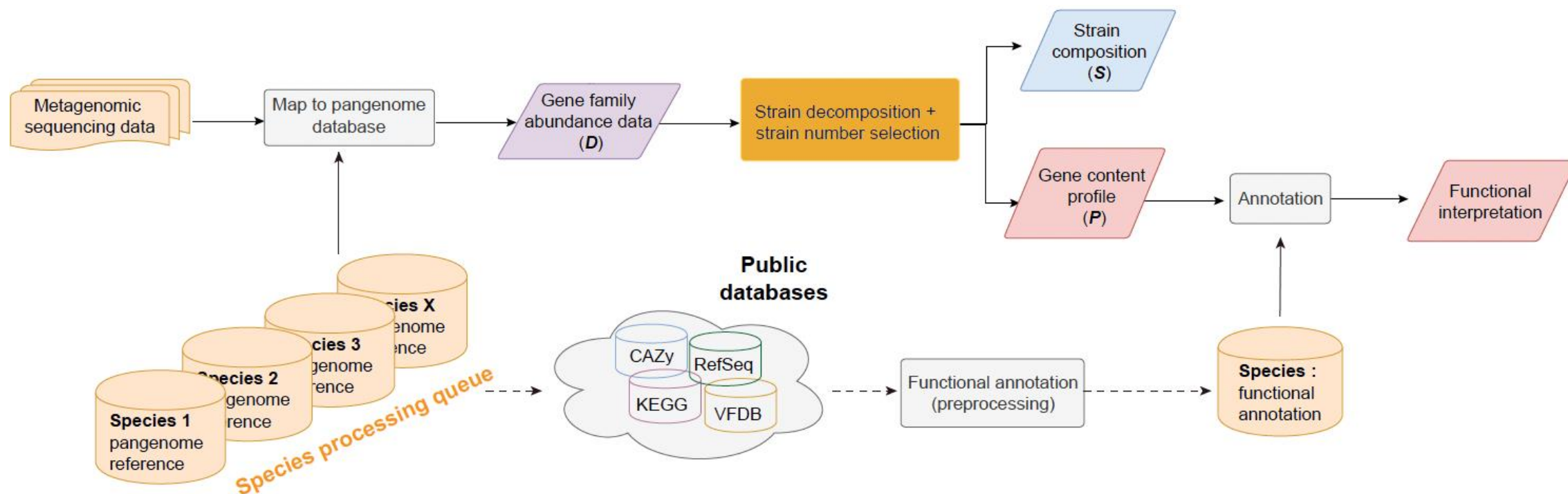
通过对基因家族丰度矩阵进行矩阵分解重建菌株组成和关联的基因成分谱。

■ Present
■ Absent



简介

StrainPanDA workflow



研究设计

验证：多菌株混合的模拟体系

- 大肠杆菌：考虑测序错误，测序深度以及宏基因组背景噪音对结果的影响。
- 其他菌种：长双歧杆菌，艰难梭菌，粪肠球菌，普拉梭菌，普雷沃氏菌。

评测：与其他菌株分析工具比较

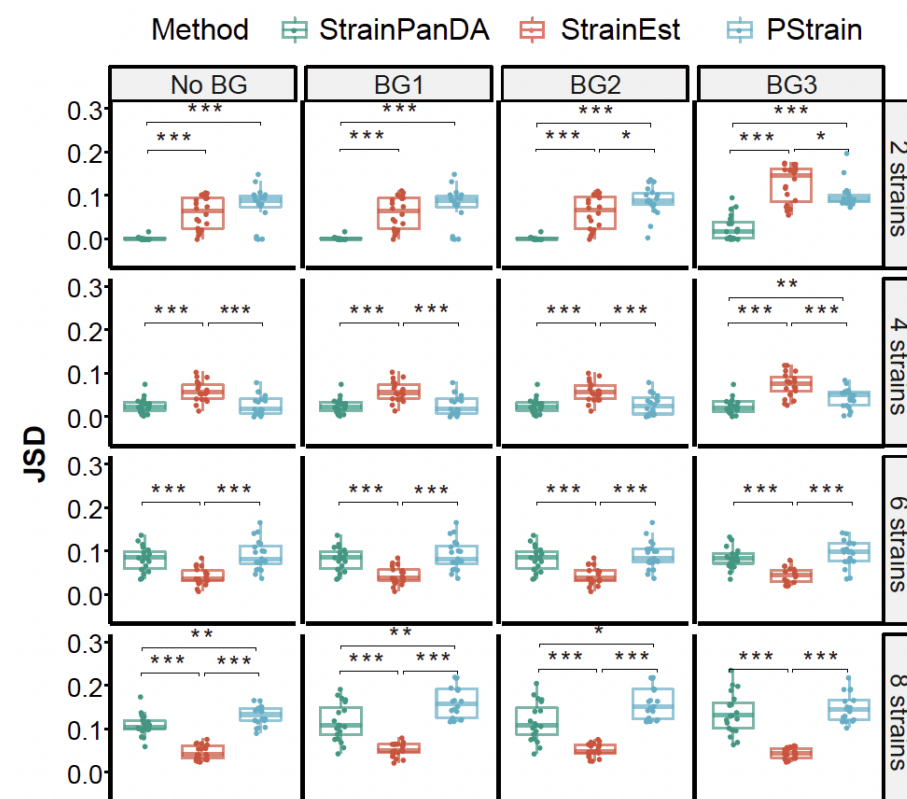
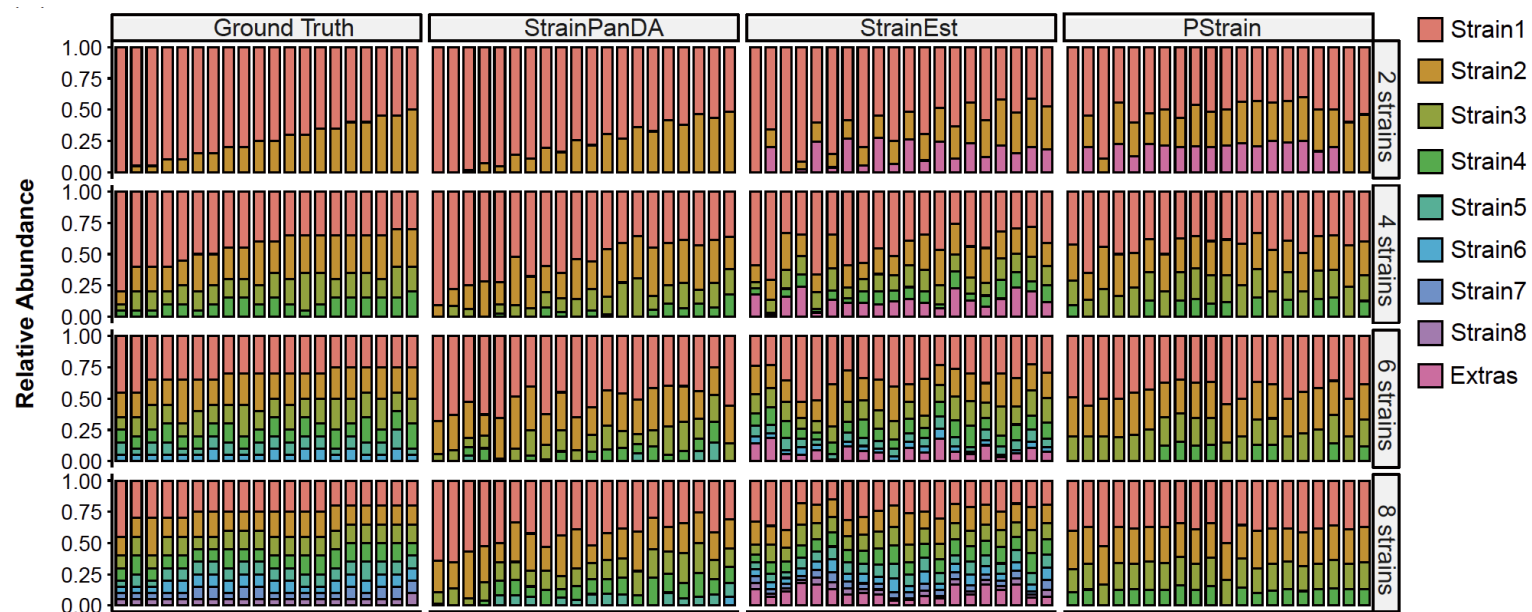
- StrainEst, PStrain, PanPhlAn

应用：纵向宏基因组数据集

- 婴儿肠道微生物 (Bäckhed *et al.* 2015)
- FMT干预后肠道微生物 (Kong *et al.* 2020)

结果

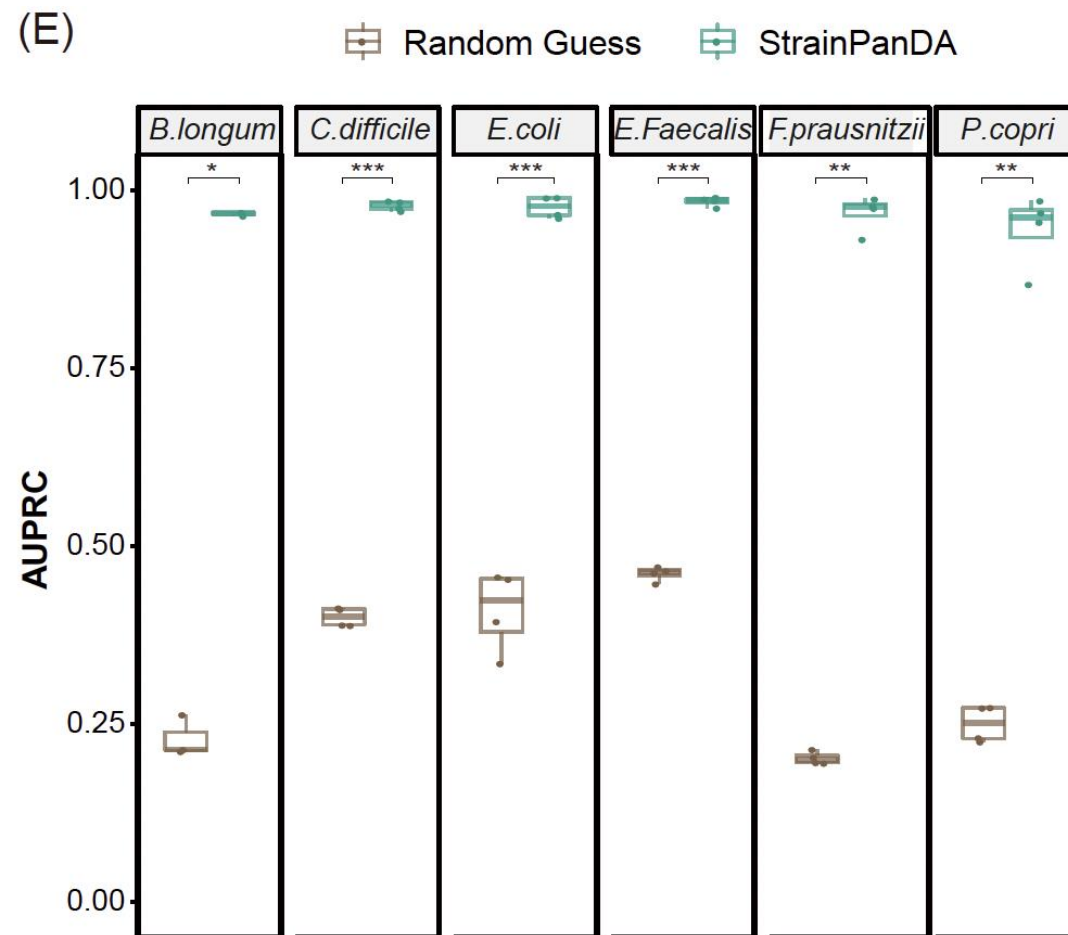
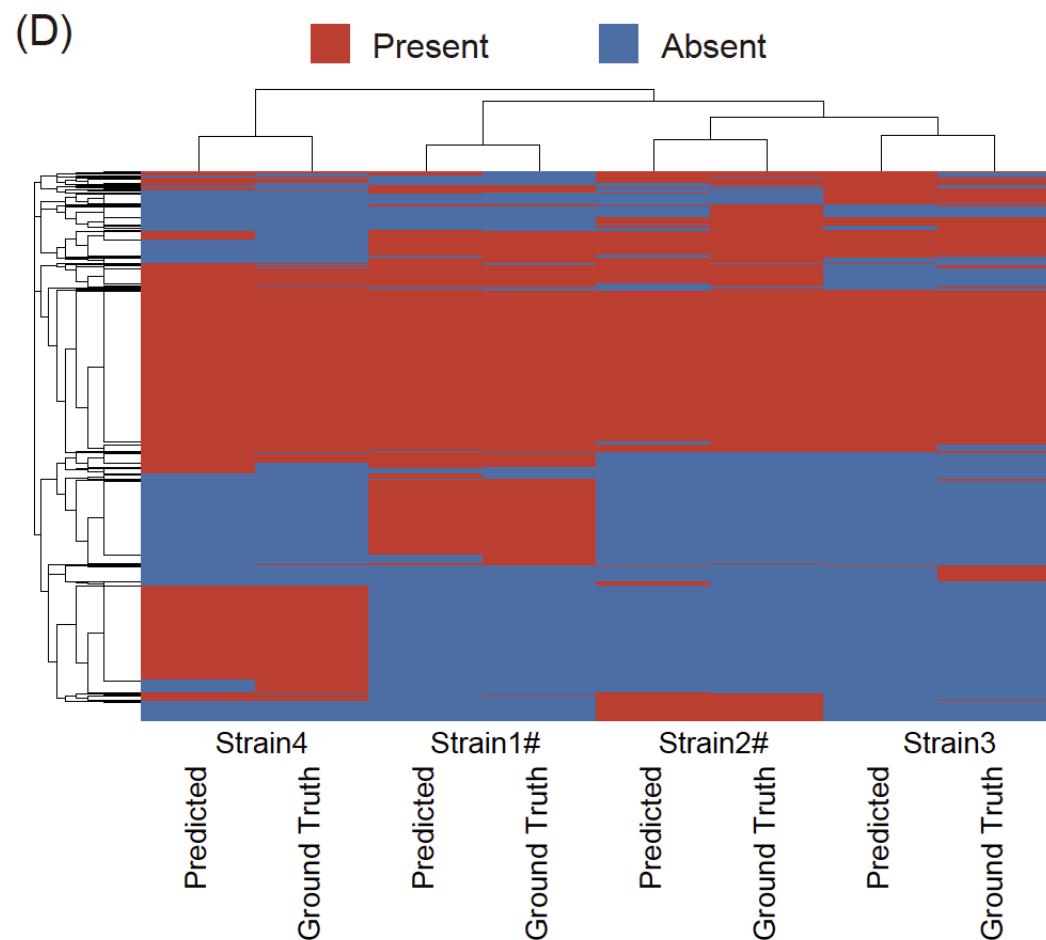
StrainPanDA 准确推导菌株组成



JSD: Jensen-Shannon divergence, JS散度

结果

StrainPanDA 可以准确推导共存菌株基因成分

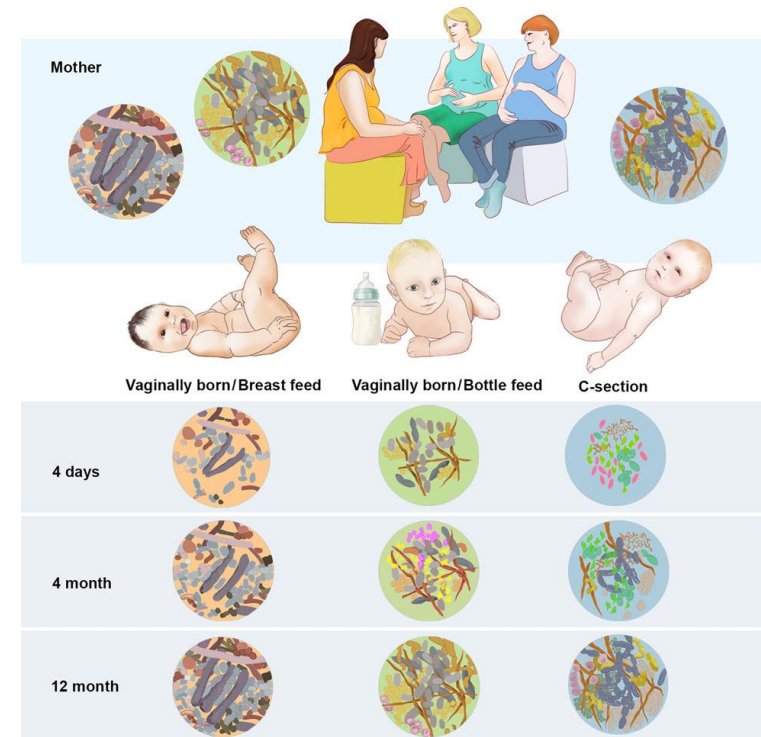
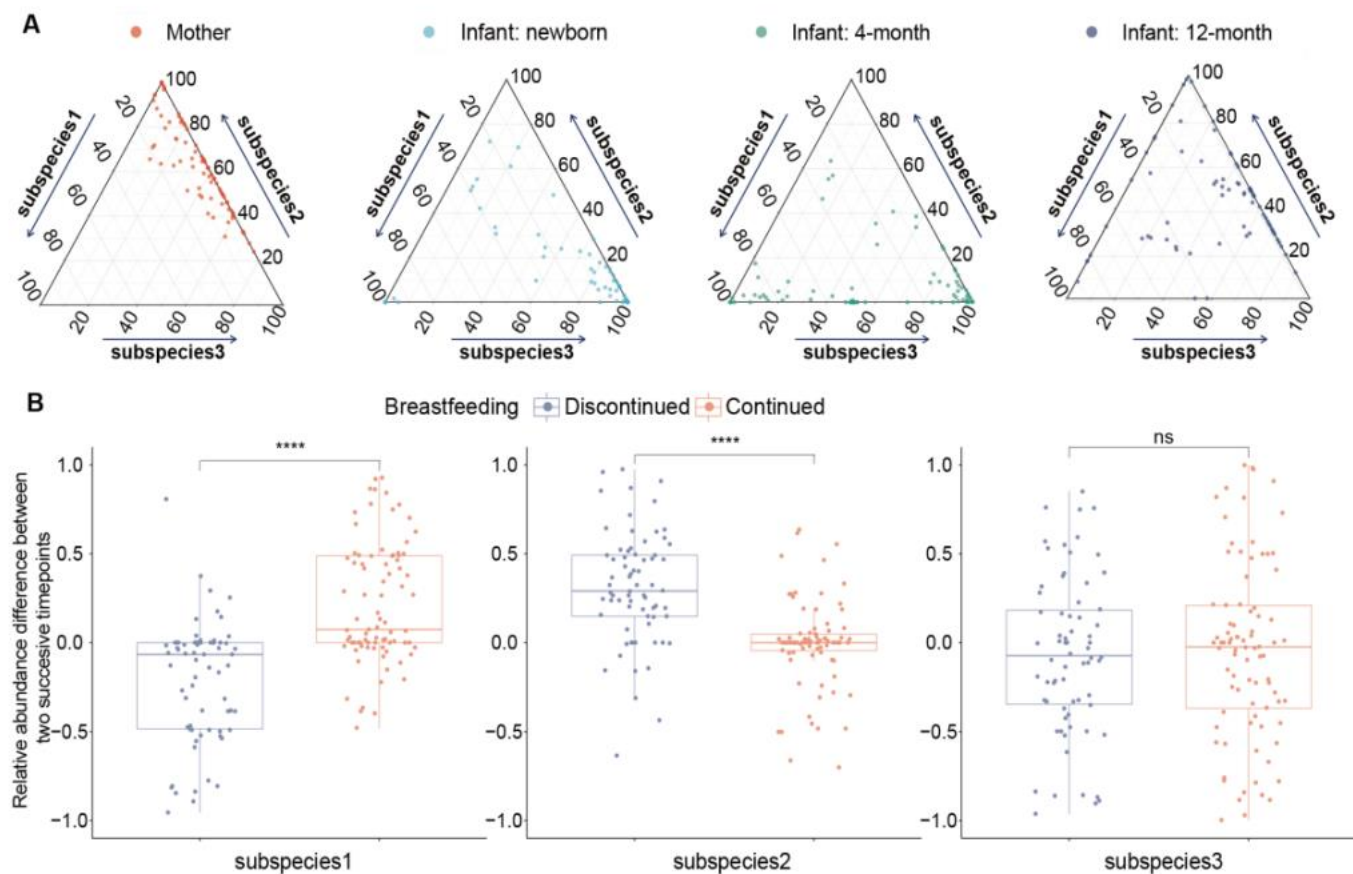


AUPRC: area under the Precision-Recall Curve, 精度/召回曲线下面积

结果

案例#1:长双歧菌亚种在婴儿肠道菌群中的演替

长双歧菌(*B. longum*)亚种

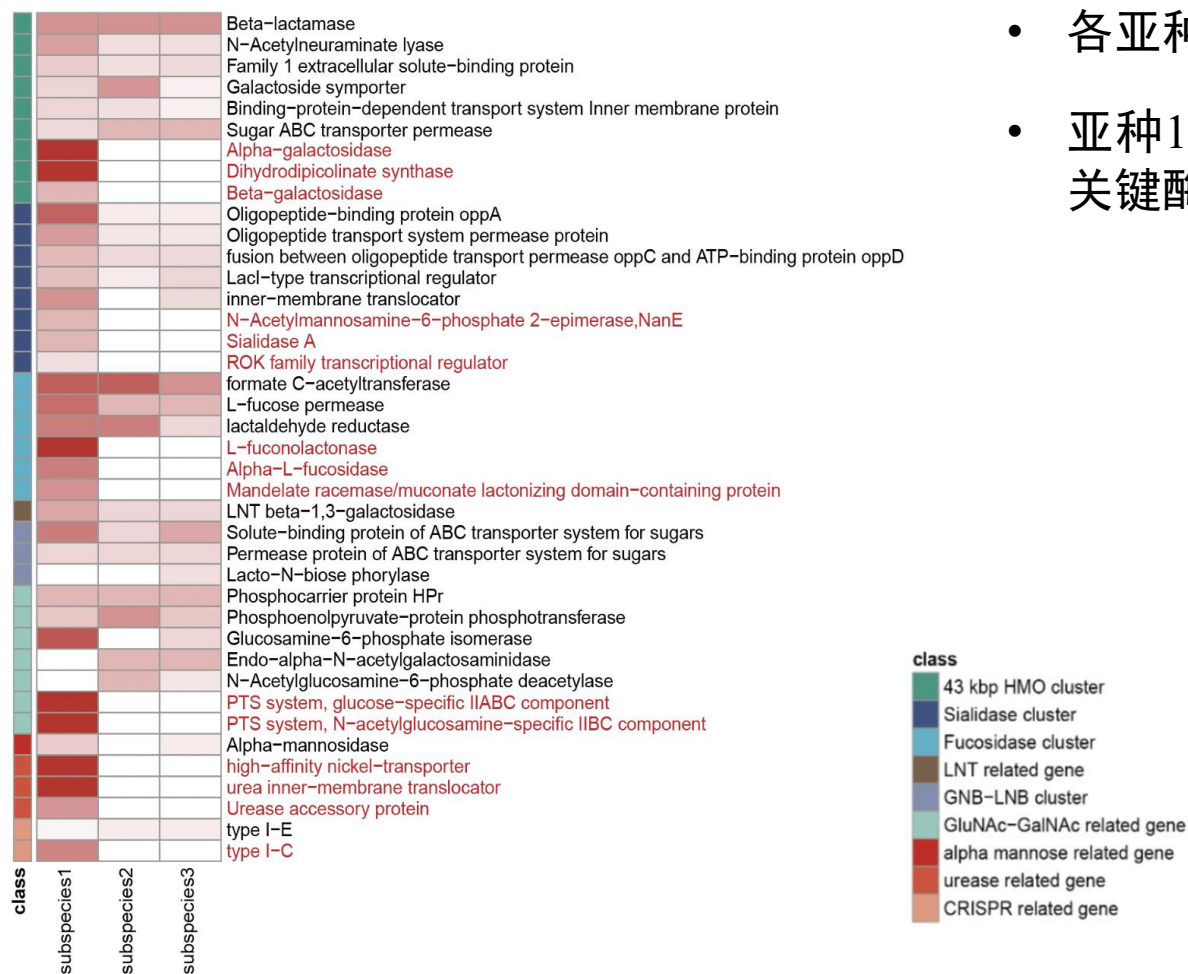


- 母乳喂养状态的改变(中断或者继续)与长双歧菌优势亚种的转变相关联.

结果

案例#1:长双歧杆菌亚种在婴儿肠道菌群中的演替

StrainPanDA重建的基因成分谱

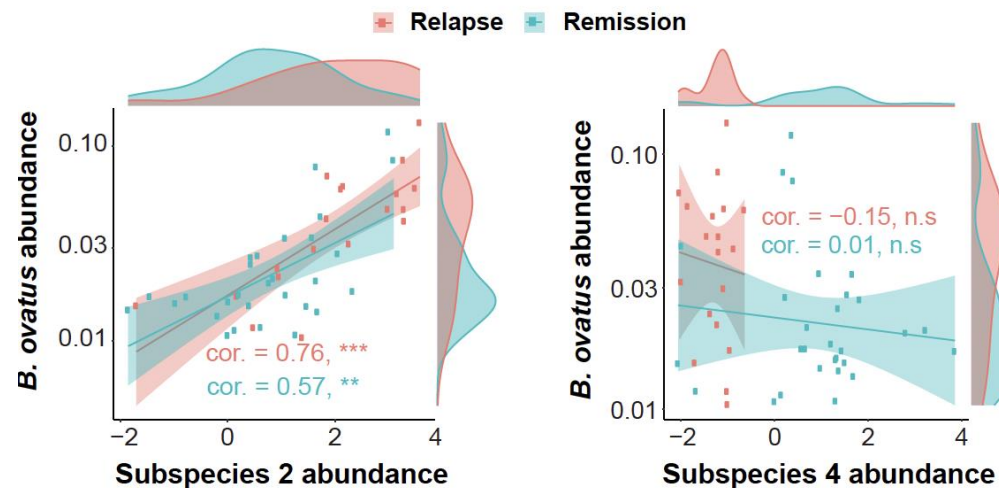
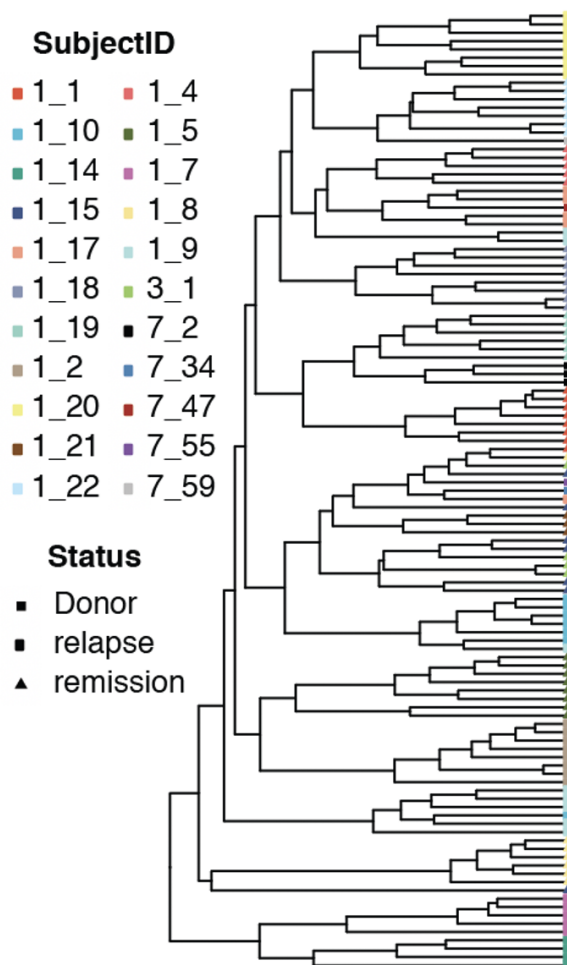


- 各亚种具有不同的营养利用相关的基因
- 亚种1特有的一些基因家族属于编码人乳寡糖降解需要的关键酶的基因

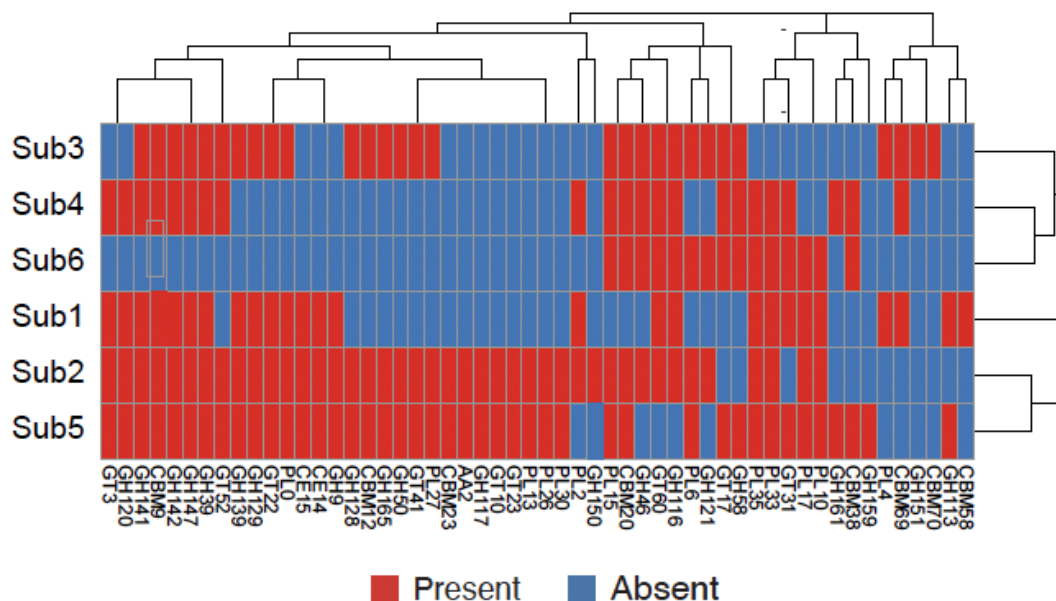
结果

案例#2：接受FMT干预的克罗恩病患者

卵形拟杆菌 (*B. ovatus*) 亚种



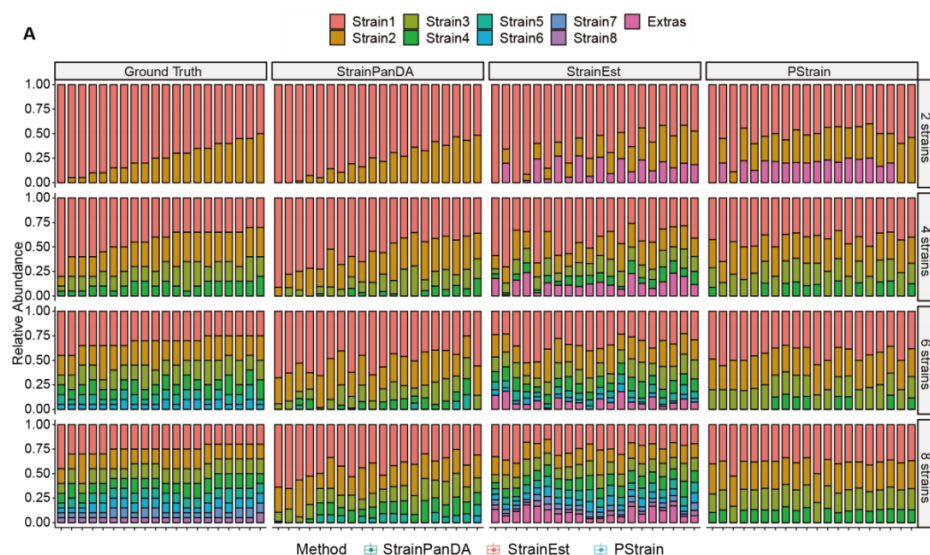
- 亚种组成具有显著的个体特征；
- 卵形拟杆菌的两个亚种丰度与其菌种的丰度相关性差别很大；
- 亚种2具有更多的碳水化合物活性酶(CAZy)基因以及特异的毒力因子基因，可能与其竞争优势和FMT后病情反复存在关联。



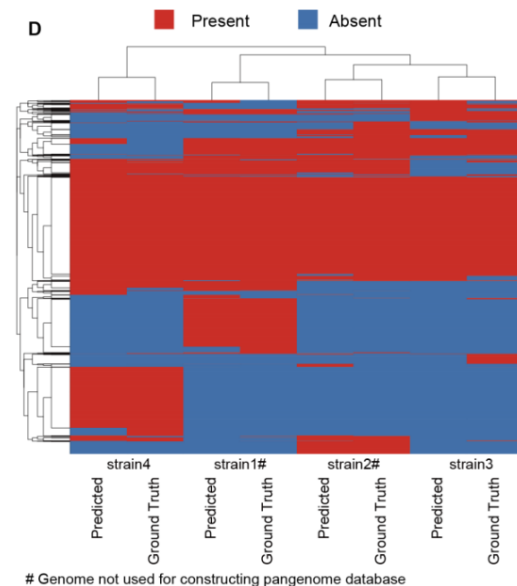
总结

- 与主流菌株分析工具相比，StrainPanDA 提供了一种高准确高鲁棒性重建菌株组成与菌株基因成分谱的新方法。
- 菌株组成与关联基因成分谱的重建对于理解微生物适应性变化以及菌株特有功能极其重要。
- StrainPanDA可以通过GitHub获取：<https://github.com/xbiome/StrainPanDA>

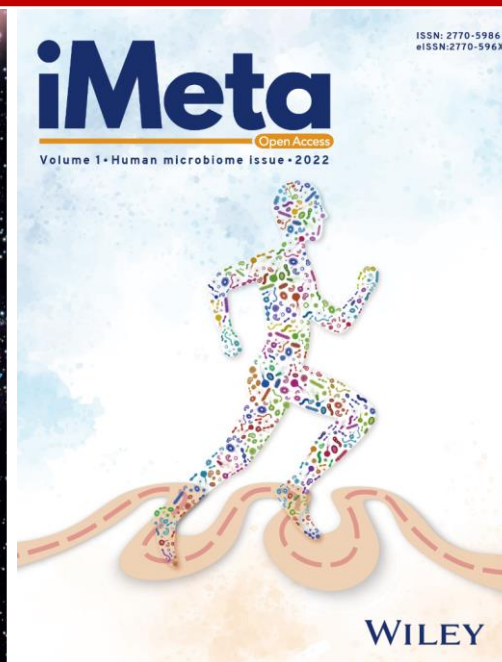
菌株组成



菌株基因成分谱



Hu, Han, Yuxiang Tan, Chenhao Li, Junyu Chen, Yan Kou, Zhenjiang Zech Xu, Yang-Yu Liu, Yan Tan, and Lei Dai. 2022. “StrainPanDA: Linked reconstruction of strain composition and gene content profiles via pangenome-based decomposition of metagenomic data.” *iMeta*. e41. <https://doi.org/10.1002/imt2.41>



“iMeta” 是由威立、肠菌分会和本领域数百位华人科学家合作出版的开放获取期刊，主编由中科院微生物所刘双江研究员和荷兰格罗宁根大学傅静远教授担任。目的是发表原创研究、方法和综述以促进宏基因组学、微生物组和生物信息学发展。目标是发表前10%(IF > 15)的高影响力论文。期刊特色包括视频投稿、可重复分析、图片打磨、青年编委、前3年免出版费、50万用户的社交媒体宣传等。2022年2月正式创刊发行！



主页: <http://www.imeta.science>

出版社: <https://onlinelibrary.wiley.com/journal/2770596x>



投稿: <https://mc.manuscriptcentral.com/imeta>



office@imeta.science



[iMeta](#)

