

# MetaTrass: 基于共标签测序数据的人类肠道微生物高质量宏基因组组装工具

齐彦伟<sup>1,2,3#</sup>, 古盛强<sup>1,4#</sup>, 张跃<sup>1</sup>, 郭立东<sup>1,4</sup>, 徐梦阳<sup>1,2,3,5</sup>, 程小芳<sup>5,6</sup>,  
王欧<sup>5,6</sup>, 孙颖<sup>1</sup>, 陈建威<sup>1</sup>, 方晓东<sup>5,7</sup>, 刘心<sup>1,2,3</sup>, 邓礼<sup>1,2,3\*</sup>, 范广益<sup>1,2,3,5\*</sup>

<sup>1</sup> 青岛华大基因研究院

<sup>2</sup> 农业基因组学国家重点实验室

<sup>3</sup> 深圳国家基因库

<sup>4</sup> 中国科学院大学生命科学学院

<sup>5</sup> 深圳华大生命科学研究院

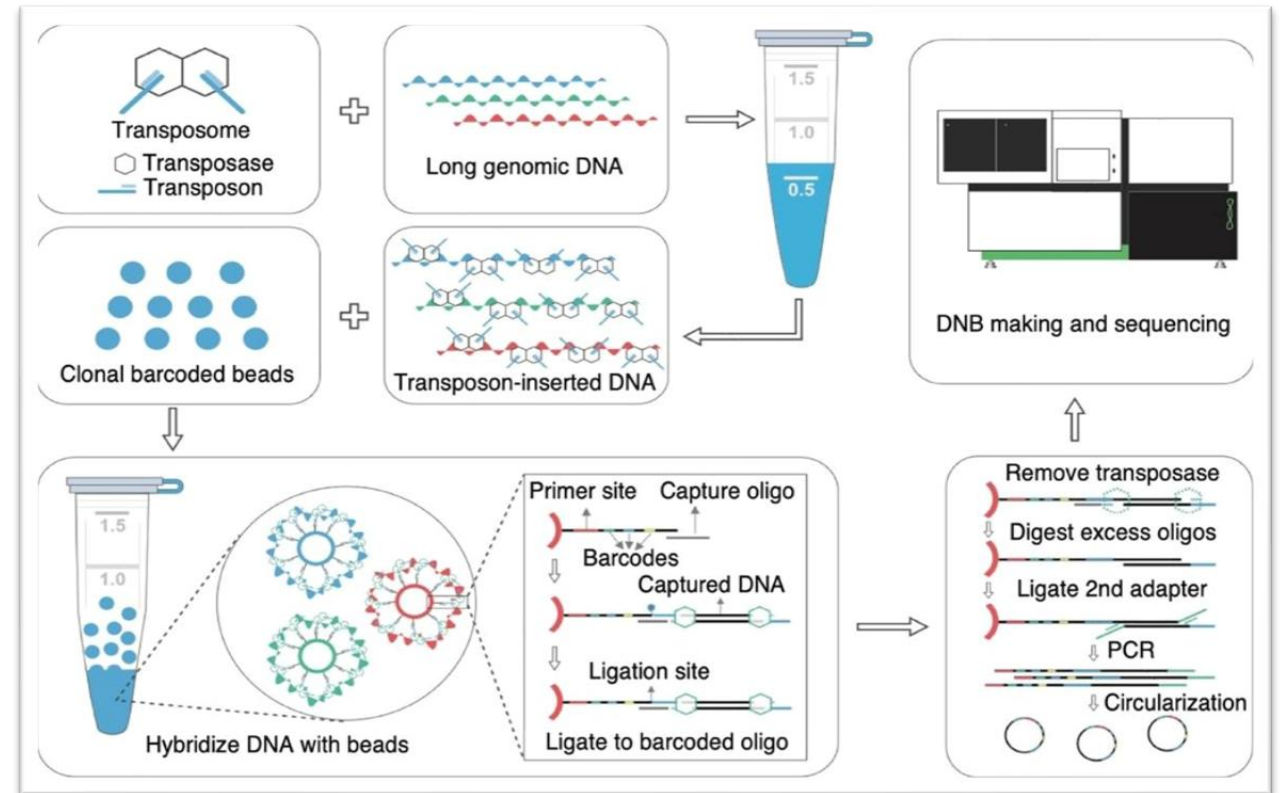
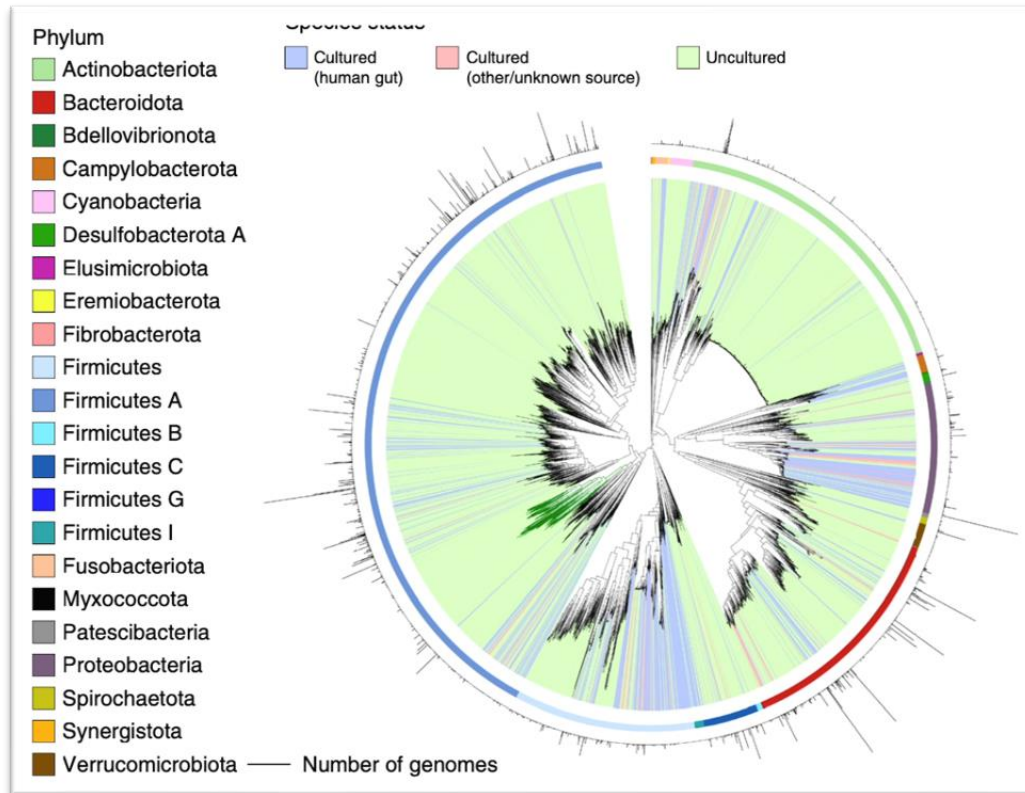
<sup>6</sup> 深圳华大智造科技有限公司

<sup>7</sup> 深圳华大基因股份有限公司



Yanwei Qi, Shengqiang Gu, Yue Zhang, Lidong Guo, Mengyang Xu, Xiaofang Cheng, Ou Wang, Ying Sun, Jianwei Chen, Xiaodong Fang, Xin Liu, Li Deng, Guangyi Fan, 2022. MetaTrass: a high-quality metagenome assembler of the human gut microbiome by cobarcode sequencing reads. *iMeta*. <https://doi.org/10.1002/imt2.46>

# 背景介绍



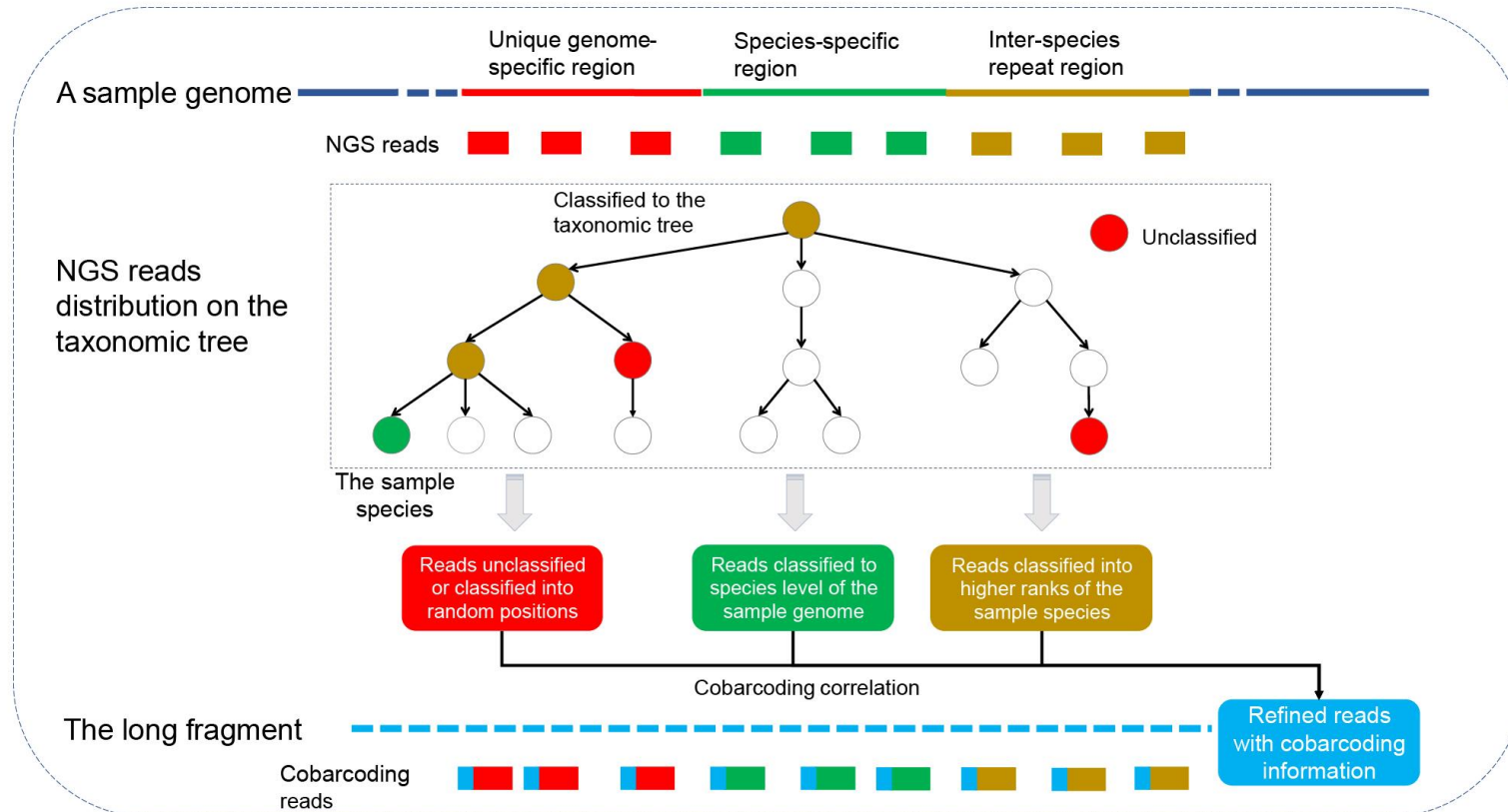
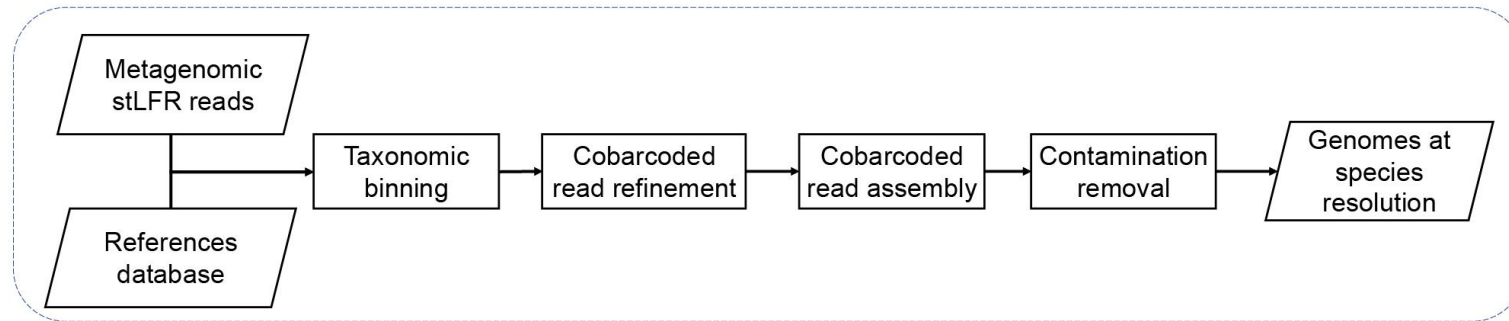
The Unified Human Gastrointestinal Genome (UHGG) collection, comprising 204,938 nonredundant genomes from 4,644 gut prokaryotes.

*Nature Biotechnology* 39 (2021)

single-tube long fragment read (stLFR), a technology that enables sequencing of data from long DNA molecules using economical second-generation sequencing technology.

*Genome Research* 29(5): 798–808. (2019)

# MetaTrass组装方法实现流程与原理解析



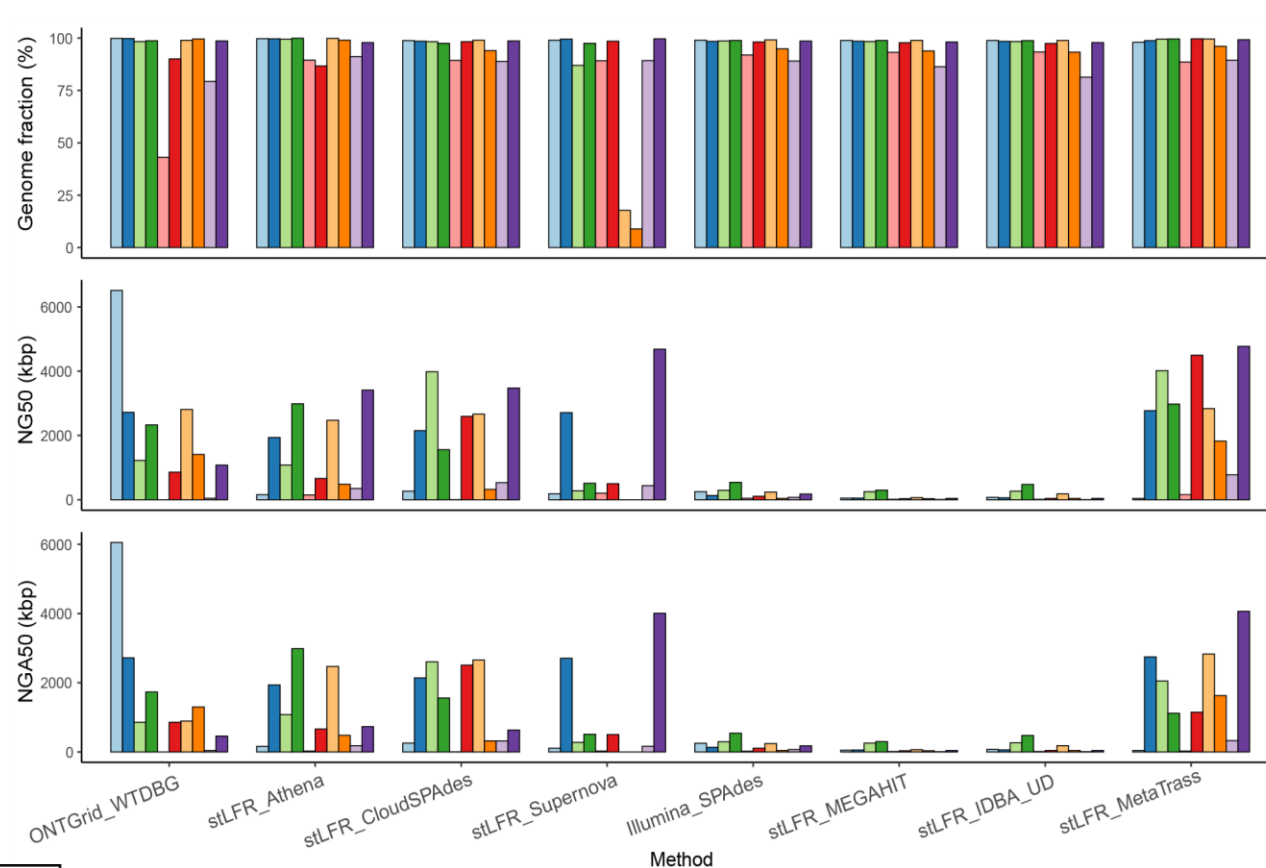
Metagenomic Taxonomy Read Assembly of Single Species (MetaTrass)

# 不同组装方法组装MOCK10标准品结果对比

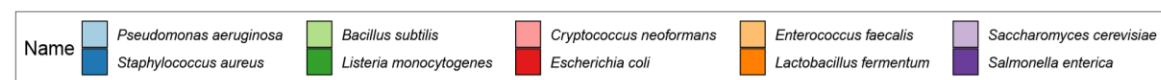
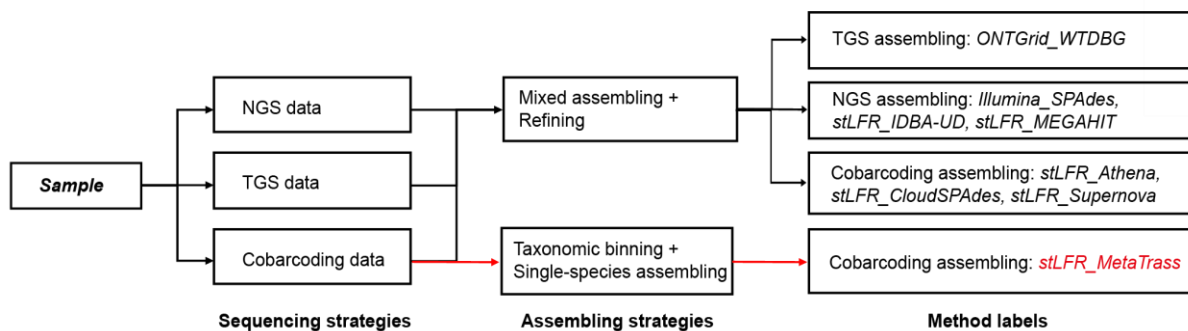
## • MOCK10标准品stLFR测序数据基本信息

Taxid	Species name	Theoretic genomic DNA	Reference length (bp)	Classified read count	Coverage depth
1280	<i>Staphylococcus aureus</i>	12	2730326	6725076	492.621
1351	<i>Enterococcus faecalis</i>	12	2845392	75166780	5283.4
1423	<i>Bacillus subtilis</i>	12	4045677	38992509	1927.61
1613	<i>Lactobacillus fermentum</i>	12	1905333	31651503	3322.41
1639	<i>Listeria monocytogenes</i>	12	2992342	22911609	1531.35
287	<i>Pseudomonas aeruginosa</i>	12	6792330	4103526	120.828
28901	<i>Salmonella enterica</i>	12	4809318	10133379	421.406
4932	<i>Saccharomyces cerevisiae</i>	2	12843354	54904223	854.983
5207	<i>Cryptococcus neoformans</i>	2	29176277	12497637	85.6699
562	<i>Escherichia coli</i>	12	4875441	18123742	743.471

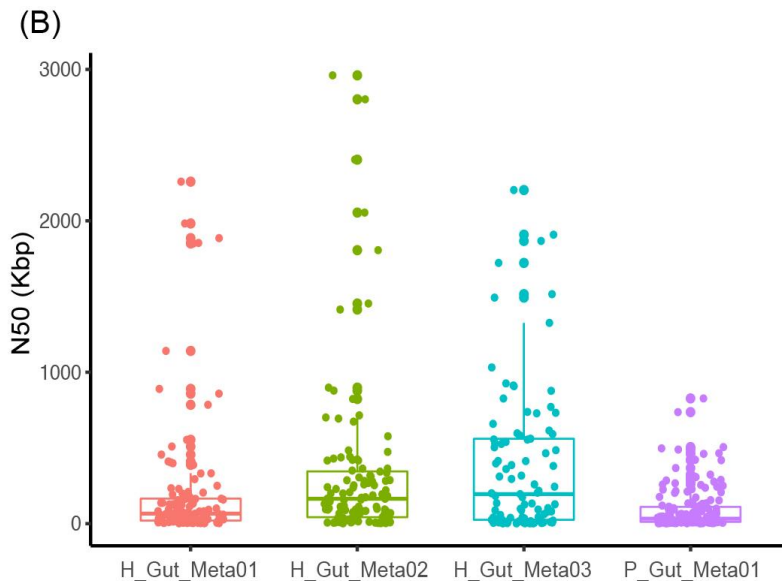
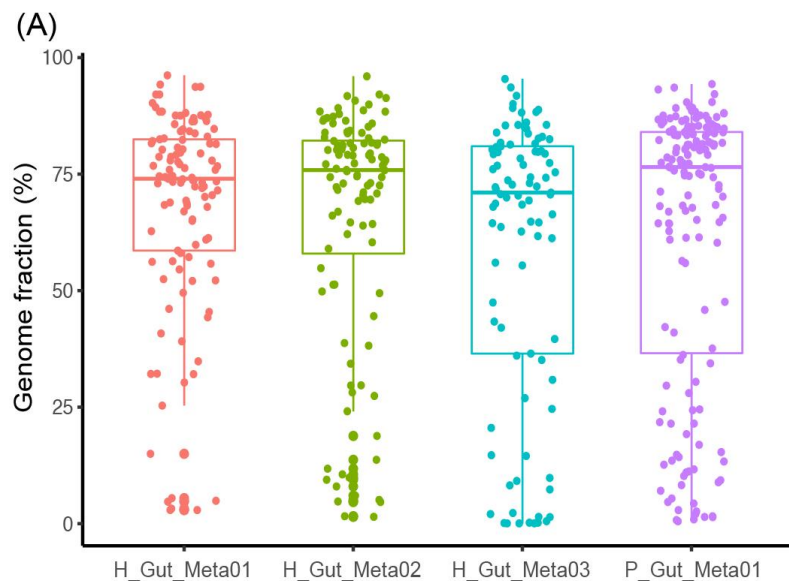
## • 对比不同组装方法组装MOCK10标准品的10个菌种基因组的结果评价



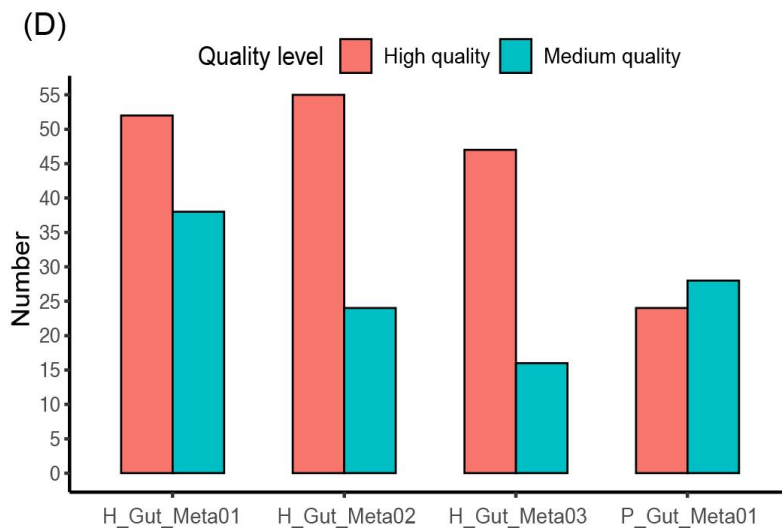
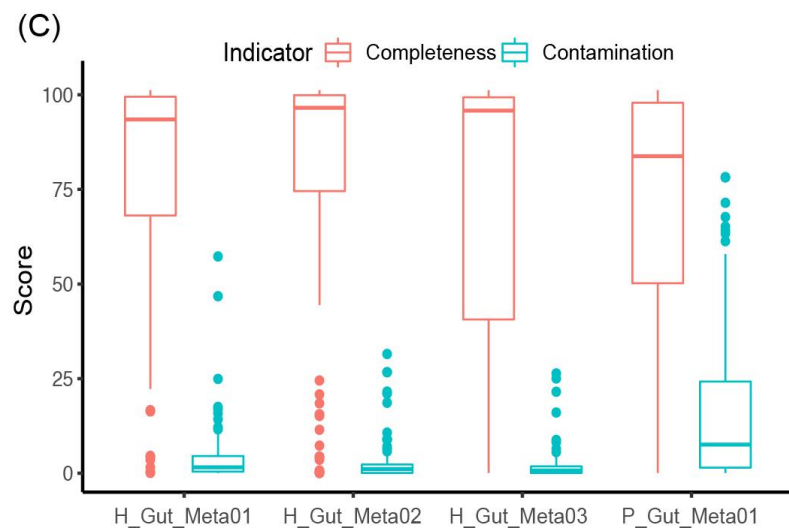
## • MOCK10标准品的测序数据类型和组装方法策略



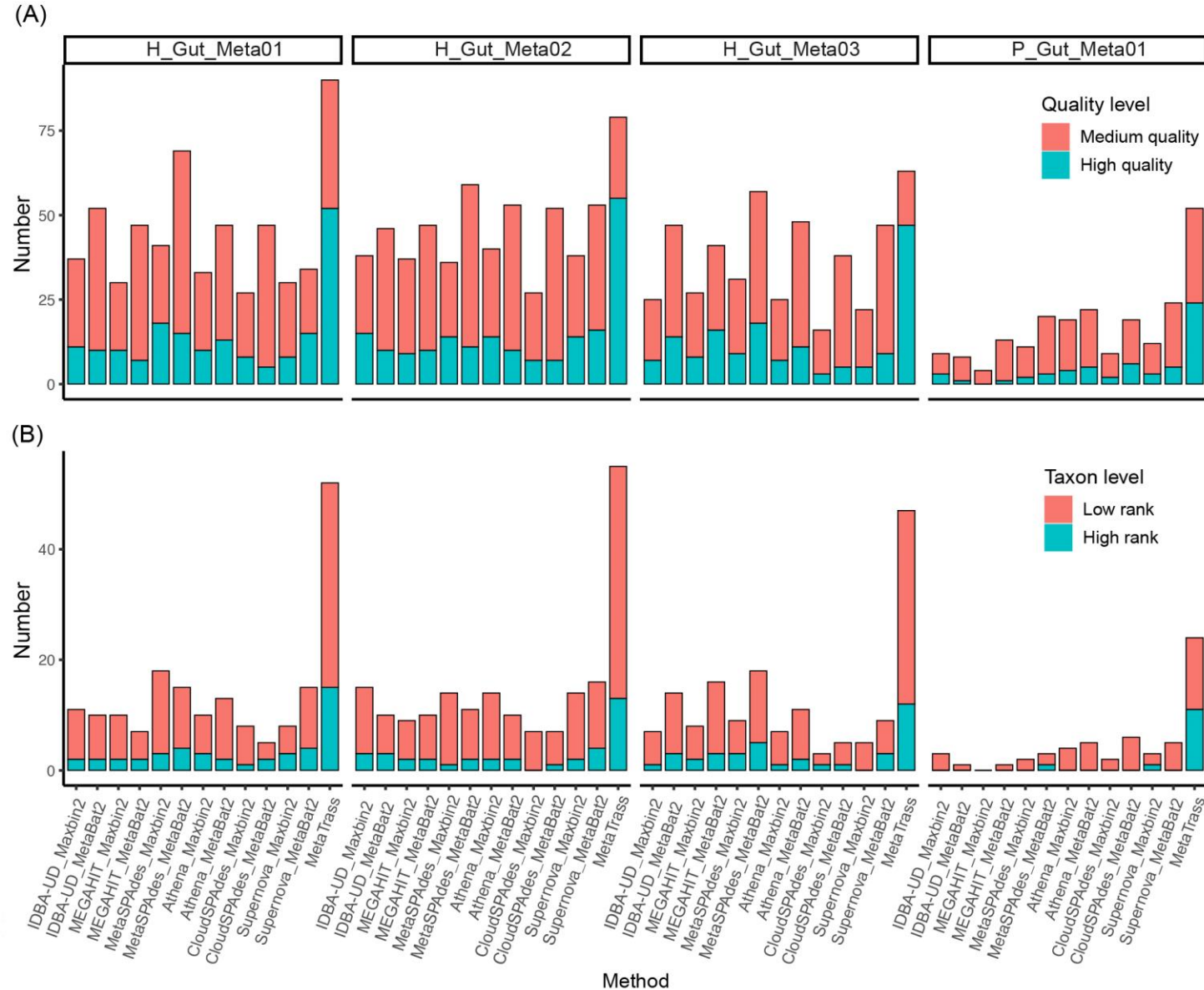
# MetaTrass组装人肠道宏基因组评估



- 获得较高且一致性的基因组百分数
- 得到Contig N50从Kb到Mb范围提升
- 总体上有较高的基因组完整性和低水平的污染度
- 获得较为可观数量的高质量和中质量单菌水平基因组



# 不同方法的宏基因组组装结果比较

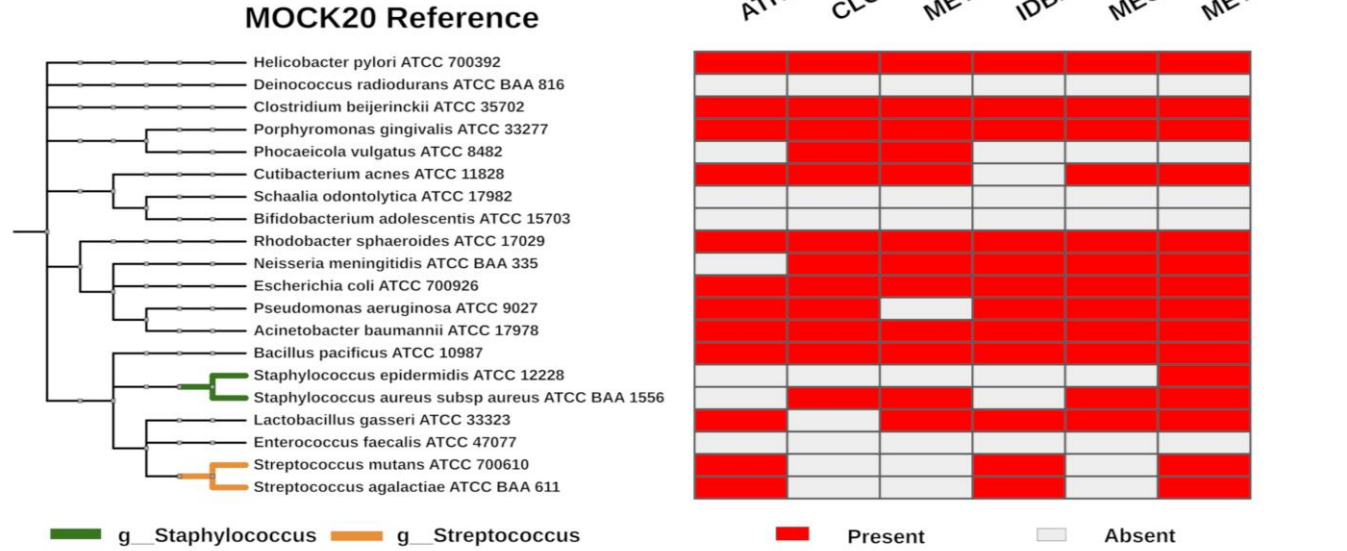


- MetaTrass获得高质量和中质量的菌种数目明显多于对比的12种传统方法
- MetaTrass获得高质量菌种中注释层级精度高于“目”层级几乎与其他方法总的高质量数目相当

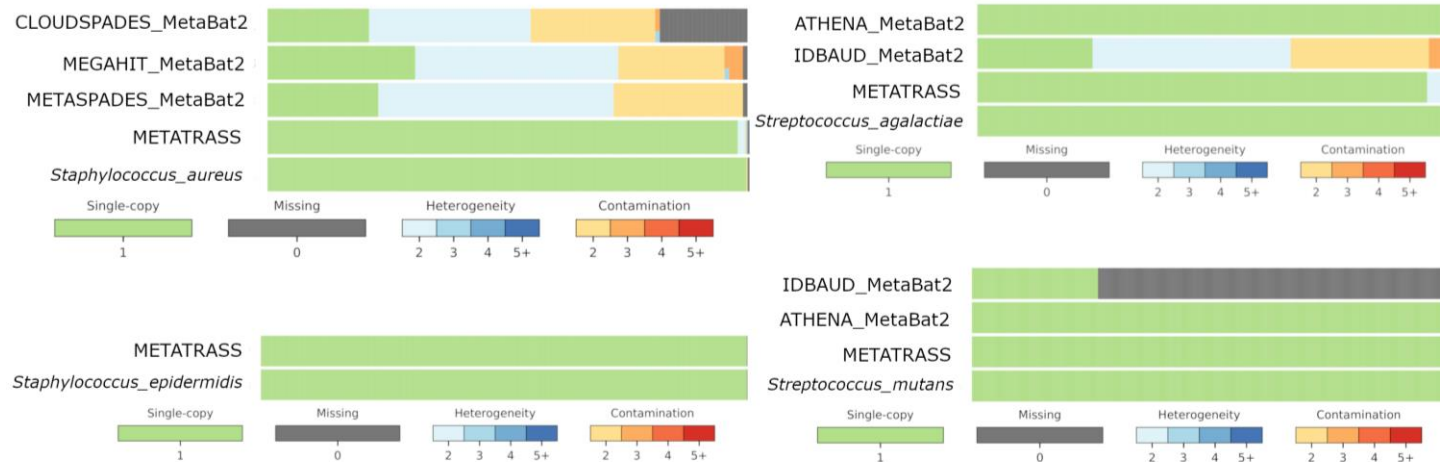


# MetaTrass应用于其他共标签数据 (MOCK20)

(A)



(B)



# Software performance and download

- Time consumption between different assembly tools

Assembler	Thread number		Runtime (min)			
	All samples		H_Gut_meta01	H_Gut_Meta02	H_Gut_Meta03	P_Gut_Meta01
IBDA-UD	6		863	884	911	2657
MEGAHIT	16		179	161	163	611
MetaSPAdes	16		1478	1289	1429	3459
CloudSPAdes	16		1024	1163	1039	2627
Supernova	8		1249	864	1098	6776
Athena	16		13813	8689	6361	--
MetaTrass	16		5145	2631	3147	8363

- MetaTrass' performance

Sample	Base number (Gb)	Peak RAM usage (Gb)	CPU max (thread)	Time (min)
H_Gut_Meta01	34.48	50.2	16	5,145
H_Gut_Meta02	35.33	55.1	16	2,631
H_Gut_Meta03	37.88	55.5	16	3,147
P_Gut_Meta01	97.2	71	16	8,363

- Freely download from Github repository

The screenshot shows the Github repository page for BGI-Qingdao / MetaTrass. The repository is public and has 154 commits. The file list includes folders like MetaTrass, Test, bin, config, images, and tools, as well as files like .gitattributes, LICENSE, README.md, and Trass.py. The README.md file is open, showing the title 'MetaTrass © BGI-Qingdao' and a description: 'MetaTrass is the abbreviation to Metagenomics Taxonomic Reads For Assembly Single Species. MetaTrass is based on high-quality references with a taxonomic tree and long-range information encoded within co-barcoding short-read sequences. The comprehensive use of co-barcoding information and references in our approach can reduce the false-negative effects of genome taxonomy to assemble high-quality metagenomes.'

<https://github.com/BGI-Qingdao/MetaTrass>

# 总结

- MetaTrass是第一个采用“先分箱后组装”策略的宏基因组组装工具，其源代码已在Github网址<https://github.com/BGI-Qingdao/MetaTrass>公开。
- MetaTrass通过综合利用共标签测序数据和参考基因组信息实现高效率地组装人类肠道微生物群落菌种水平的高质量基因组。
- MetaTrass利用二代测序短读长之间的共标签关联来减少传统分箱结果中的假阴性，并改善菌种基因组草图的组装连续性。

Yanwei Qi, Shengqiang Gu, Yue Zhang, Lidong Guo, Mengyang Xu, Xiaofang Cheng, Ou Wang, Ying Sun, Jianwei Chen, Xiaodong Fang, Xin Liu, Li Deng, Guangyi Fan, 2022. MetaTrass: a high-quality metagenome assembler of the human gut microbiome by cobar coding sequencing reads. *iMeta*. <https://doi.org/10.1002/imt2.46>



“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](mailto:office@imeta.science)



[iMeta](#)

