

MetaTrass: a high-quality metagenome assembler of the human gut microbiome by cobarcoding sequencing reads

Yanwei Qi^{1,2,3#}, Shengqiang Gu^{1,4#}, Yue Zhang¹, Lidong Guo^{1,4}, Mengyang Xu^{1,2,3,5},
Xiaofang Cheng^{5,6}, Ou Wang^{5,6}, Ying Sun¹, Jianwei Chen¹, Xiaodong Fang^{5,7},
Xin Liu^{1,2,3}, Li Deng^{1,2,3*}, Guangyi Fan^{1,2,3,5*}

¹BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China

²State Key Laboratory of Agricultural Genomics, BGI-Shenzhen

³China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

⁴College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

⁵BGI-Shenzhen, Shenzhen 518083, China

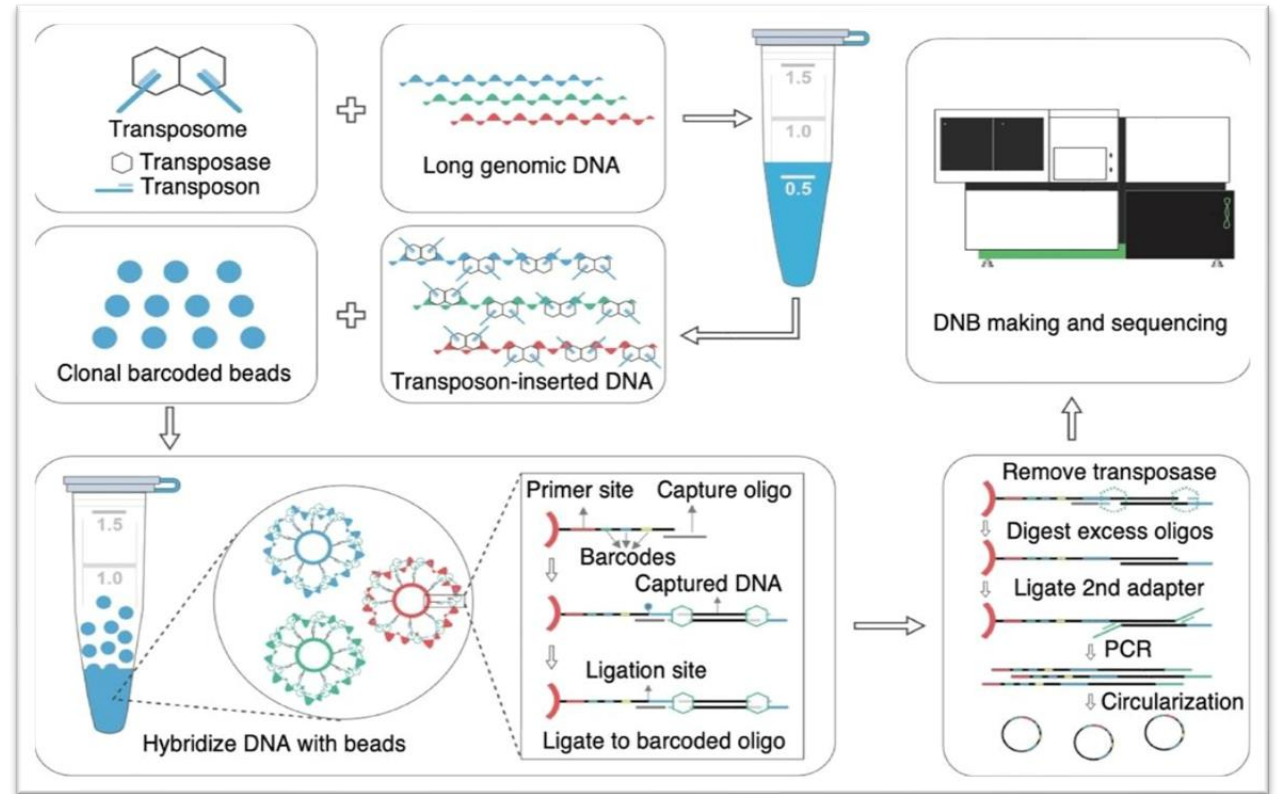
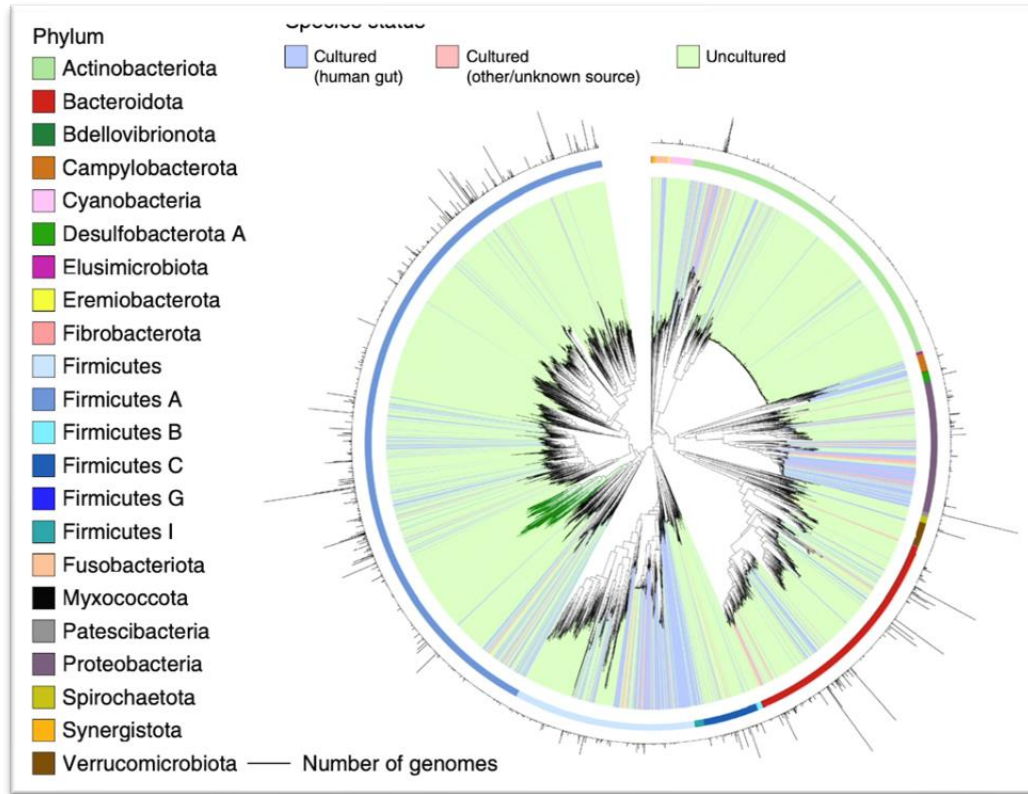
⁶MGI, BGI-Shenzhen, Shenzhen 518083, China

⁷BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China



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 cobarcoding sequencing reads. *iMeta*. <https://doi.org/10.1002/imt2.46>

Introduction



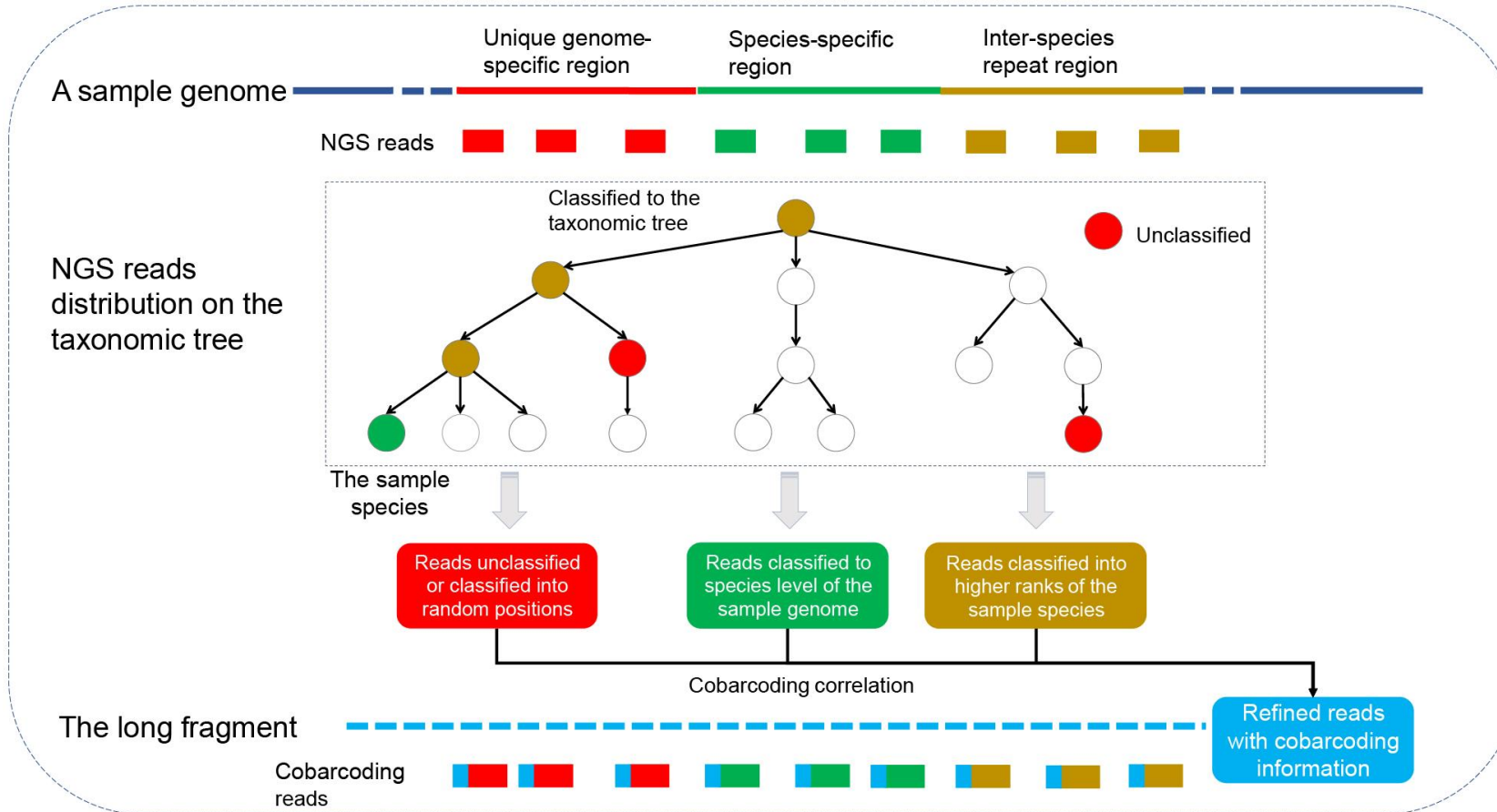
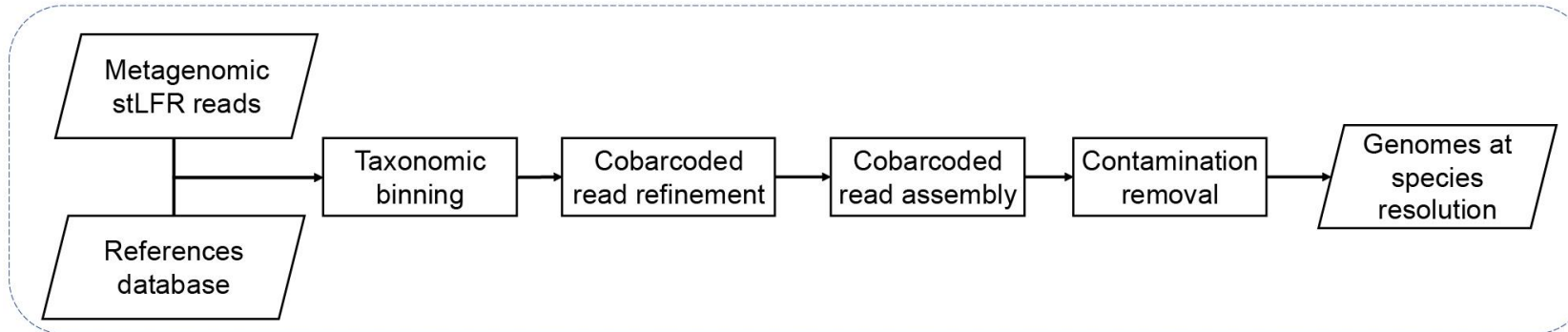
The Unified Human Gastrointestinal Genome (UHGG) collection, comprising 204,938 nonredundant genomes from 4,644 gut prokaryotes.

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 assembler pipeline and its principle

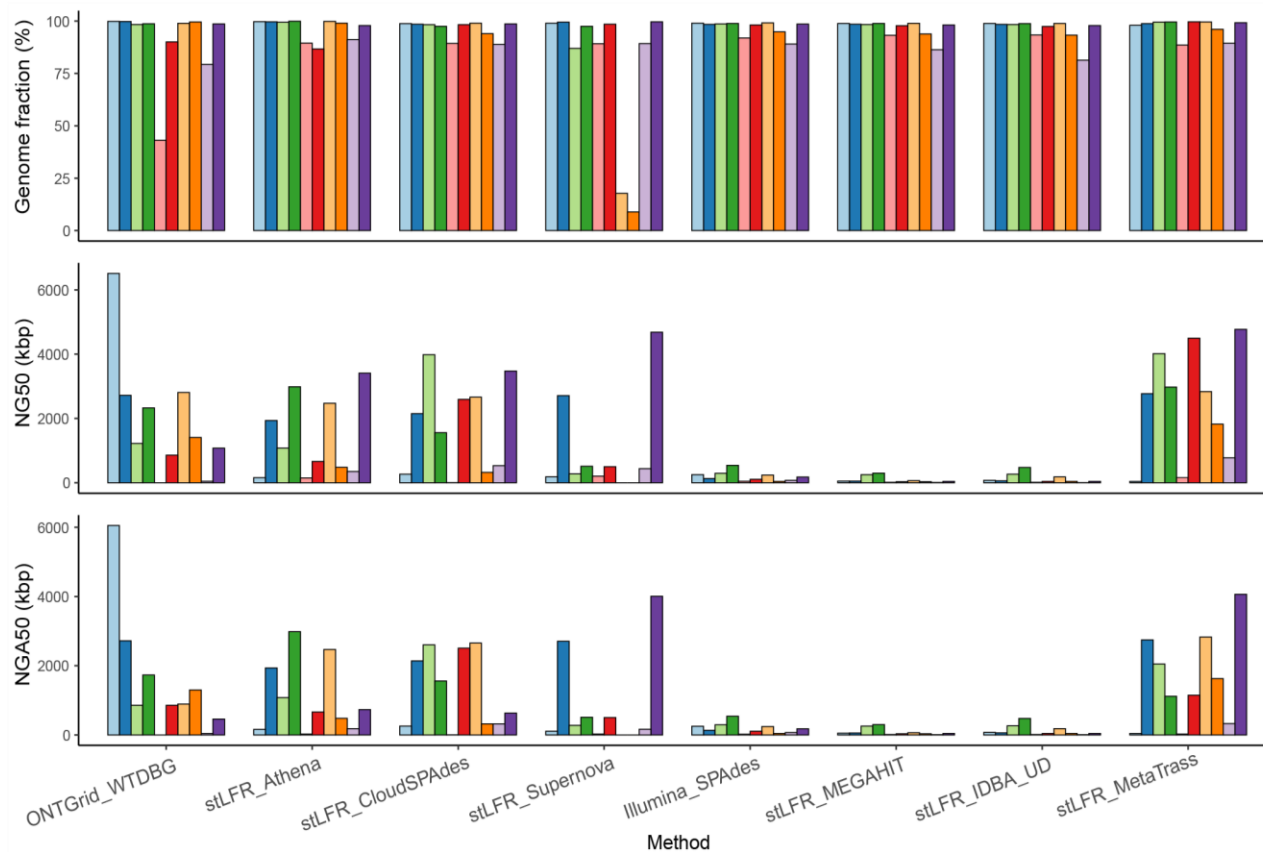


Benchmark tests by using MOCK10 standard sample

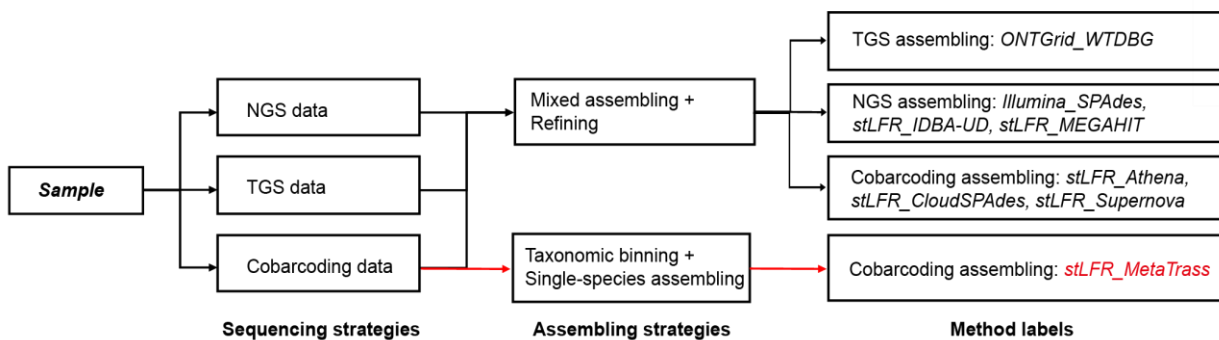
Table. stLFR sequencing basic information for MOCK10

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

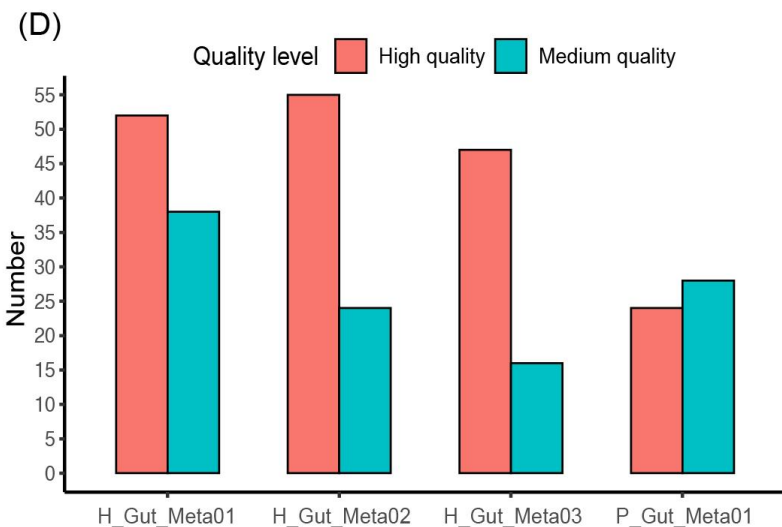
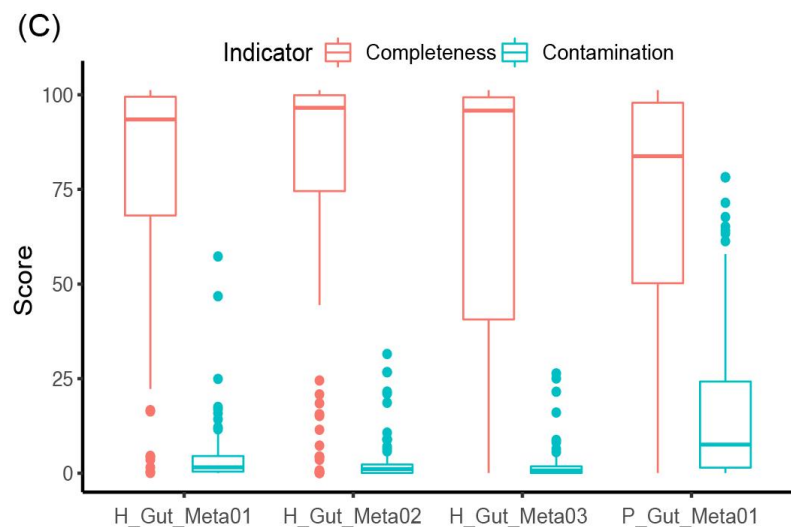
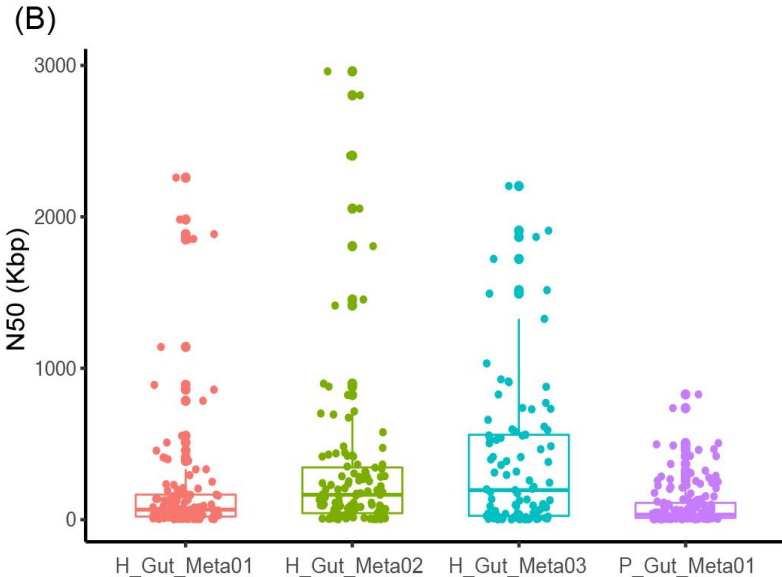
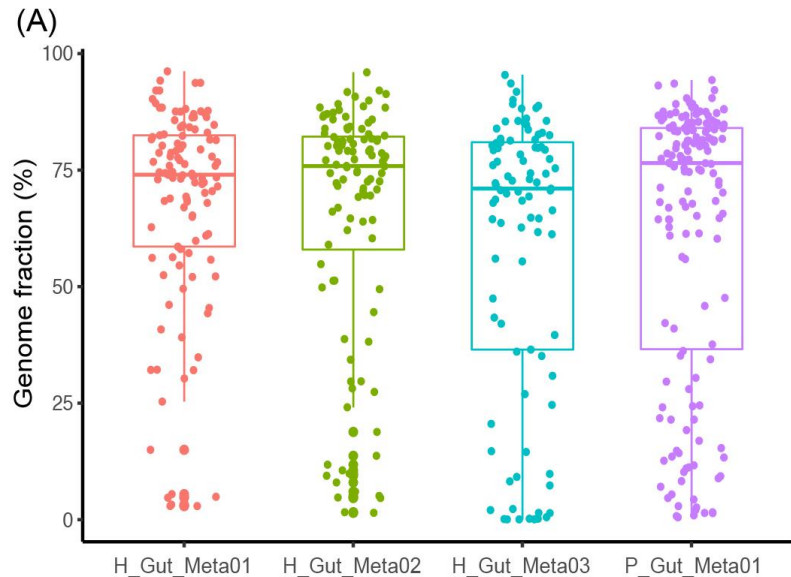
(B). Different method assembly results assessment for MOCK10



(A). MOCK10 data and assembly methods

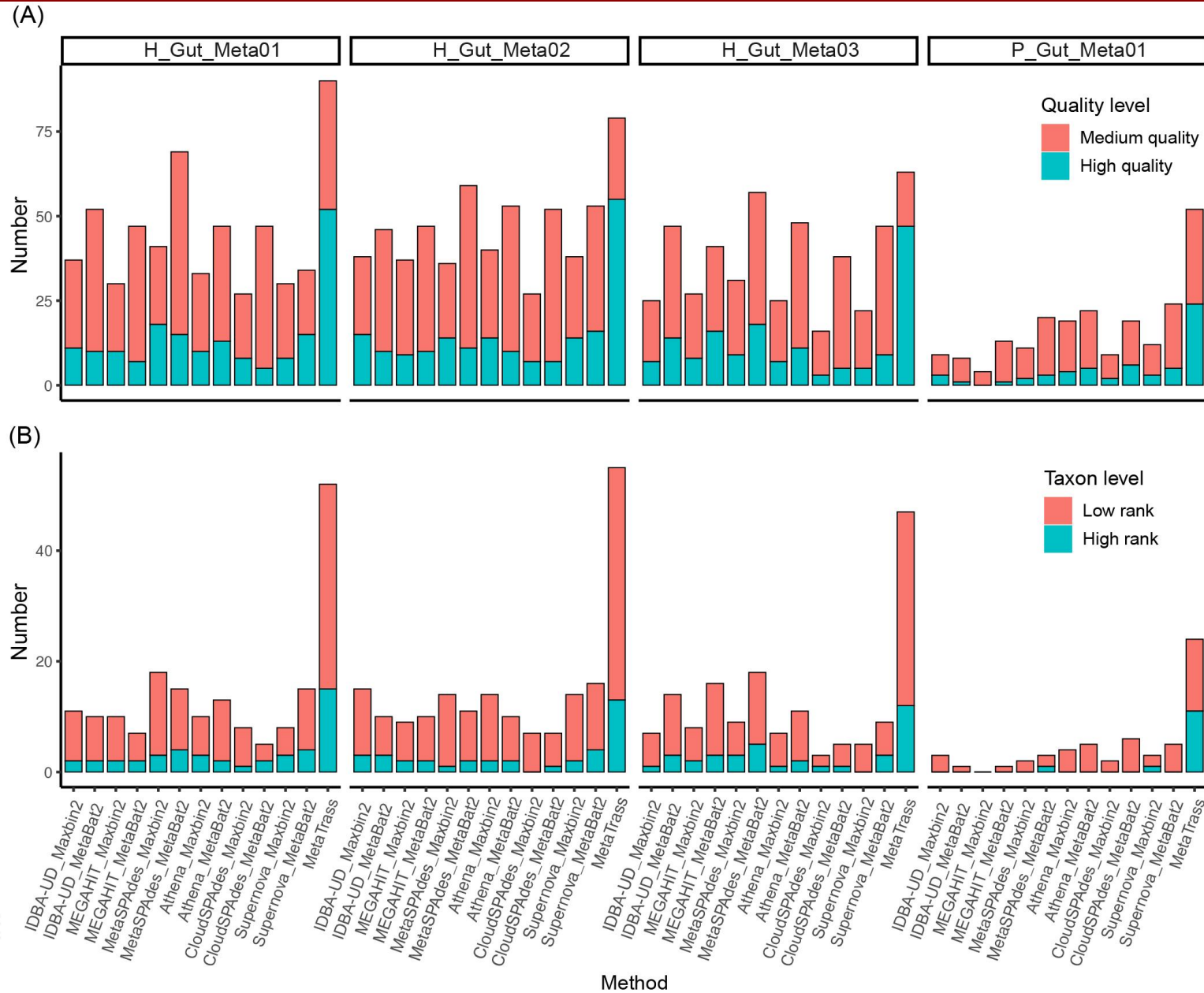


MetaTrass deployed on four human gut sample



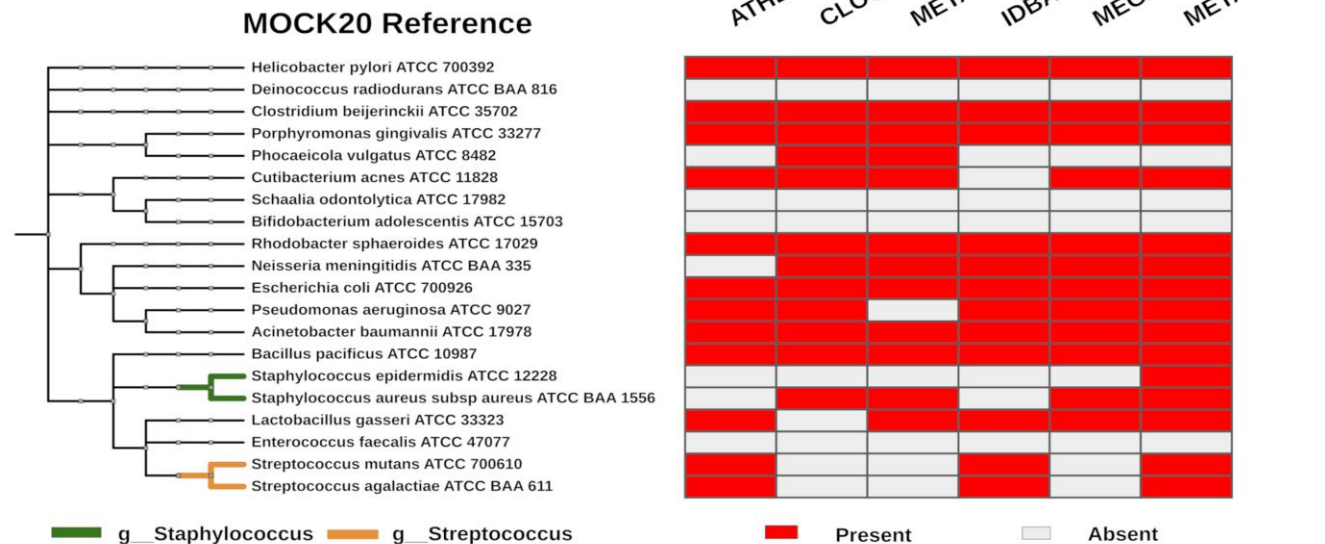
- Obtain a high and consistent percentage of the genome
- Get Contig N50 boost from Kb to Mb range
- Overall high genome completeness and low levels of contamination
- Obtain a considerable amount of high- and medium-quality genomes at the single bacterial level

Method comparison with genome quality and taxon level

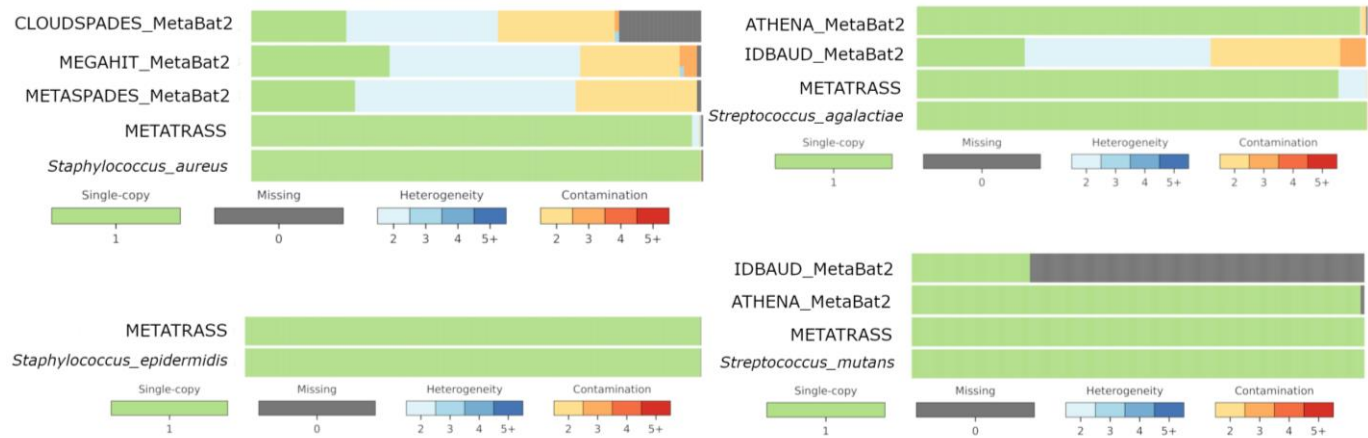


MetaTrass applied on 10X genomics data (MOCK20)

(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

BGI-Qingdao / MetaTrass Public

< Code Issues 1 Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags

Go to file Add file Code

QYanwei Merge branch 'main' of https://github.com/BGI-Qingdao/MetaTrass 466a2c3 7 days ago 154 commits

- MetaTrass fixed depth argument 9 months ago
- Test add demo test data and result last month
- bin add demo script 9 months ago
- config add lane.lst file to config 10 months ago
- images add output structure pic2 11 months ago
- tools add notes and TBarefiner_NOS 29 days ago
- .gitattributes Initial commit 15 months ago
- LICENSE Initial commit 15 months ago
- README.md Update README.md 18 days ago
- Trass.py fixed depth argument 9 months ago

README.md

MetaTrass © BGI-Qingdao

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>

Summary

- MetaTrass is the first metagenome assembly tool implemented by binning-first-assembly-later strategy and it is freely available at <https://github.com/BGI-Qingdao/MetaTrass>.
- MetaTrass integrates the cobarcoding and the reference genome information to achieve high capability of generating high-quality genomes at species level for the human gut microbial communities.
- MetaTrass uses the cobarcoding correlation between short length reads to reduce false negatives in conventional taxonomic binning and improve the continuity of draft assemblies.

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 cobarcoding sequencing reads. *iMeta*. <https://doi.org/10.1002/imt2.46>



“iMeta” is an open-access Wiley partner journal and launched by scientists of the Chinese Academy of Sciences. iMeta aims to promote metagenomics, microbiome and bioinformatics development by publishing original researches, methods or protocols, and reviews. The goal is to publish highly quality papers (Top 10%, IF > 15) targeting broad audience. Unique features including video submission, reproducible analysis, figure polishing, APC waiver, and promotion by social media with 500,000 followers. The first issue released in March 2022.

 Society: <http://www.imeta.science>
Publisher: <https://onlinelibrary.wiley.com/journal/2770596x>

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

 office@imeta.science

 [iMeta](#)

 [iMetaScience](#)

 [iMetaScience](#)
[iMetaJournal](#)