

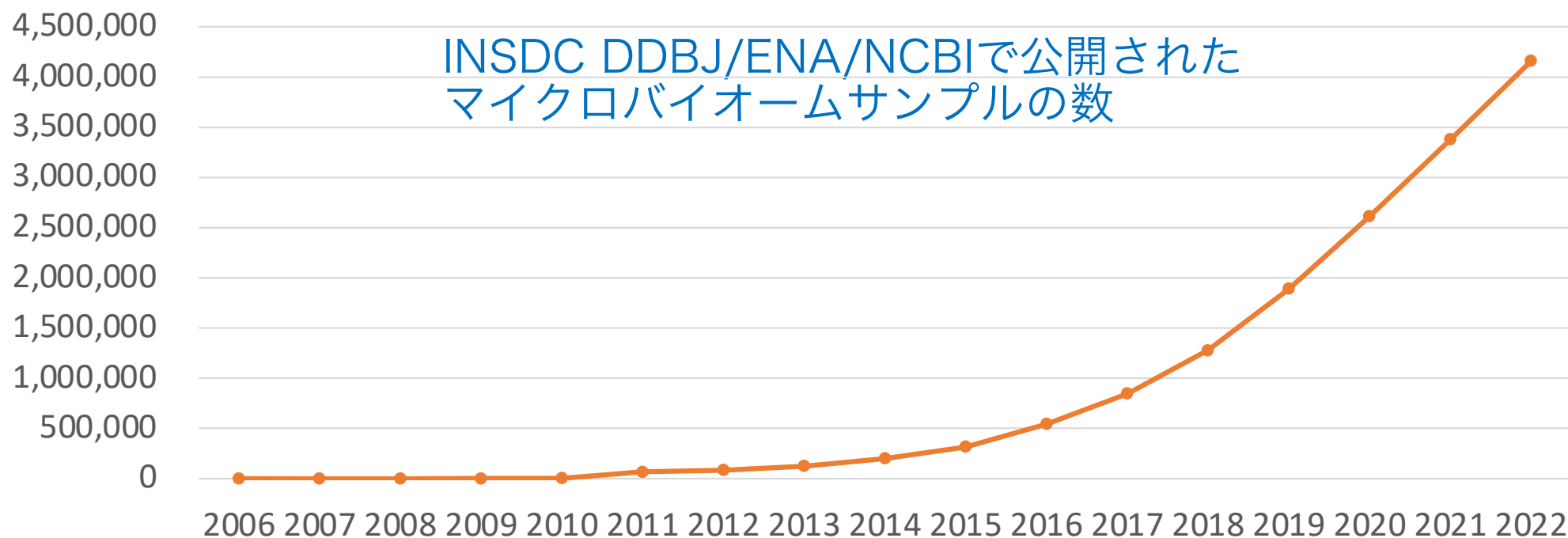
2022年10月5日
トーゴーの日シンポジウム2022

マイクロバイオーム研究を先導する ハブを目指した微生物統合データベース の特化型開発

国立遺伝学研究所
情報研究系
森 宙史

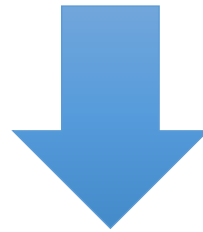
研究開発の背景

- 微生物は地球上の様々な場所に存在し環境と深く関係
- 微生物研究はバイオ分野に留まらず、他の多様な分野と連携可能
- 特に、微生物群集（マイクロバイオーーム）の研究が急速な勢いで進展している



今までのMicrobeDB.jp開発における目標

ゲノム情報を核として様々な微生物学上の知識を統合し、幅広い分野の研究者が微生物学の知見を利用できる微生物エンサイクロペディアになるような統合データベースの構築を目標とする。



微生物学分野のオミックス研究の発展に寄与
データ駆動型研究による新しい仮説の提唱



Microbe DB .JP

integrates lots of data related to microbes.

Especially, we integrate the microbial data that can be linked to **genomes**. **since 2011**



Microbe DB .JP

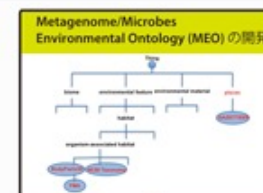
<https://microbedb.jp/>

Microbe DB.jp

MicrobeDB.jp プロジェクトでは様々な微生物学上の知識を、ゲノム情報を核として遺伝子、系統、環境の3つの軸に沿ってセマンティックウェブの技術を用いて整理統合し、幅広い分野での微生物学発展に資することの出来るデータベースの構築を目標としています。

中身は基本的にRDFデータ

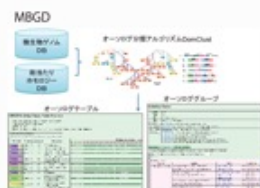
Ontology



Gene

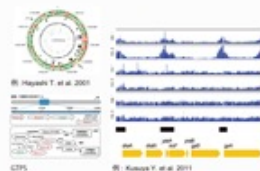
Taxon

Environment



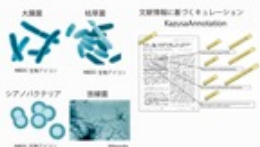
Ortholog: **MBGD**

オースログデータ



Genome: RefSeq

オミックスデータ



モデル微生物の高品質
アノテーションデータ



Taxonomy:
NCBI Taxonomy

系統分類データ



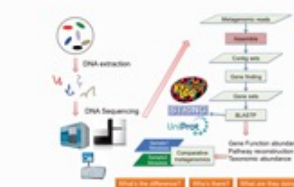
Culture Collection:
NBRC/JCM

菌株データ



Metadata:
INSDC DRA

環境のメタデータ



Metagenome:
INSDC DRA

メタゲノムデータ

Red color indicates our collaborators.

soil由来マイクロバイームサンプルの絞り込み検索



Home Document Analysis - e.g. hot spring, Enterococcus faecalis, psbA Search Sign Up Sign in

facet_metagenome_public8531

hasMetagenomeAnalysis

taxonomy8531

function119

Search id ...

attribute name

Search attribute name ...

attribute value

Search attribute value ...

hasMEO (Text)

soilX

hasMEO: Component

Component for environment8102

hasMEO: Env

Environment for microbes7117

taxonomy (Text)

Search taxonomy terms ...

taxonomy

root8531

hasHostTaxonomy (Text)

Search HostTaxonomy...

hasHostTaxonomy

root1410

pH

014

Temperature

-100150

HMADO (Text)

Search HMADO terms ...

HMADO

Human microbiome associated87 disease

HostEthnicity

Search HostEthnicity ...

Metagenomic samples8531 results found in 155ms

hasMetagenomeAnalysis: taxonomyXhasMEO (Text): soilXClear all filters

Previous1234...Next

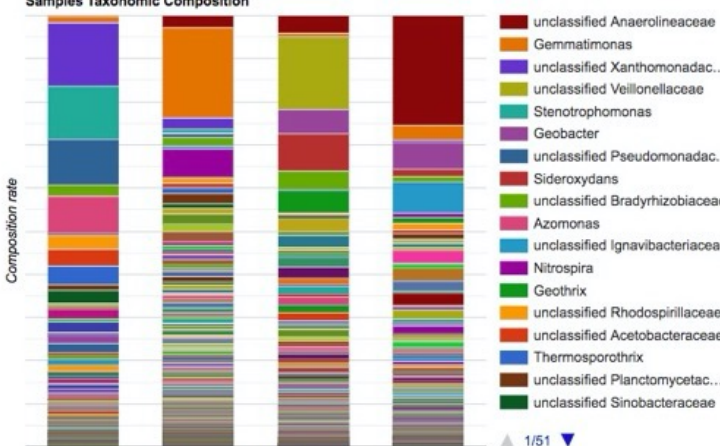
10Select AllDeselect All

Select	MDB SampleID	title	organism.name	organism.identifier	BioProjectID	SRAID	SRRID	BioSampleID	publishedDate
Remove	SAMD00003586	Urease gene-containing Archaea dominate autotrophic ammonia oxidation in two acid soils	soil metagenome	410658	PRJDB1924	DRS001577	DRR002212	SAMD00003586	2012-07-04T00:00:00.000
Remove	SAMD00009749	Active ammonia oxidizers in an acid soil are phylogenetically closely related to neutrophilic Nitrososphaera viennensis	soil metagenome	410658	PRJDB2274	DRS012638	DRR014314	SAMD00009749	2013-10-30T00:00:00.000
Add	SAMEA1559038	RMgM_Sample_16S	soil metagenome	410658	PRJEB3363	ERS184934	ERR186224	SAMEA1559038	2012-11-01T00:00:00.000
Remove	SAMEA1559037	RMR_Sample_16S	soil metagenome	410658	PRJEB3363	ERS184936	ERR186222	SAMEA1559037	2012-11-01T00:00:00.000
Add	SAMEA1559036	RMgR_Sample_16S	soil metagenome	410658	PRJEB3363	ERS184935	ERR186225	SAMEA1559036	2012-11-01T00:00:00.000
Add	SAMEA1559035	RRR_Sample_16S	soil metagenome	410658	PRJEB3363	ERS184933	ERR186223	SAMEA1559035	2012-11-01T00:00:00.000
Remove	SAMD00018981	MIMARKS Survey related sample from rhizosphere metagenome	rhizosphere metagenome	939928	PRJDB2986		DRR021946 DRR021947 DRR021948	SAMD00018981	2015-01-16T00:00:00.000
Add	SAMN02054434	MIMARKS Survey related sample from Soil metagenome		410658	PRJNA198445	SRS416341	SRR835396 SRR835397 SRR835398	SAMN02054434	2013-04-23T00:00:00.000
Add	SAMN02054433	MIMARKS Survey related sample from Soil metagenome		410658	PRJNA198445	SRS416342	SRR835399 SRR835400 SRR835401	SAMN02054433	2013-04-23T00:00:00.000
Add	SAMN02054432	MIMARKS Survey related sample from Soil metagenome		410658	PRJNA198445	SRS416340	SRR835402 SRR835403 SRR835404	SAMN02054432	2013-04-23T00:00:00.000

Metagenome sample comparison analysisCompare

Taxonomic composition (bar)Taxonomic composition (heatmap)Diversity indexHierarchical clusteringPCoAFunctional composition (bar)Functional composition (heatmap)

Samples Taxonomic Composition



unclassified Anaerolineaceae
Gemmatimonas
unclassified Xanthomonadac...
unclassified Veillonellaceae
Stenotrophomonas
Geobacter
unclassified Pseudomonadac...
Sideroxydans
unclassified Bradyrhizobiaceae
Azomonas
unclassified Ignavibacteriaceae
Nitrospira
Geothrix
unclassified Rhodospirillaceae
unclassified Acetobacteraceae
Thermosporothrix
unclassified Planctomycetac...
unclassified Sinobacteraceae

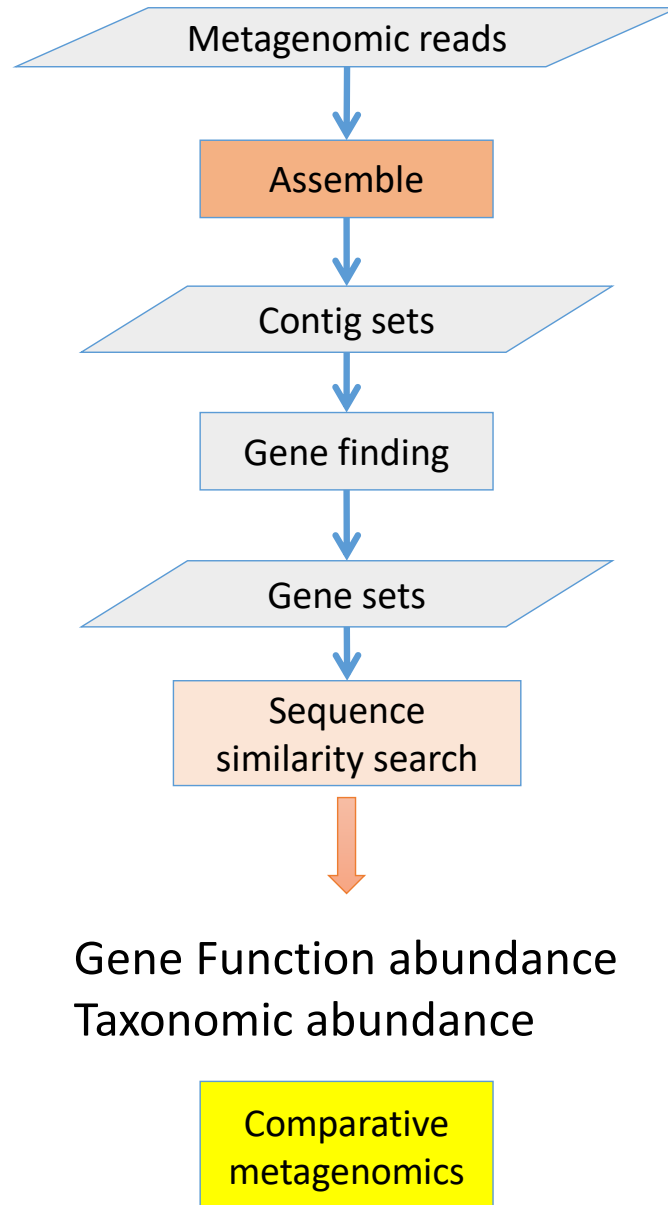
代表的なマイクロバイオームDB間の比較

	Admin	Database URL	Sequence data	Separate amplicon and shotgun?	Taxa	Function	Number of samples in Sept. 2022
NCBI Taxonomy + SRA	NCBI, USA	https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=408169	○	×	△	×	>4,150,000
GOLD	JGI, USA	https://gold.jgi.doe.gov/	×	×	×	×	155,000
IMG/M	JGI, USA	https://img.jgi.doe.gov/cgi-bin/m/main.cgi	○	×	○	○	29,961
MG-RAST	Chicago U. USA	https://www.mg-rast.org/	○	○	○	○	80,144
MGNify	EBI, EU	https://www.ebi.ac.uk/metagenomics/	×	○	○	○	422,983
MicrobeDB.jp v.3	NIG, Japan	https://microbedb.jp/	×	○	○	○	1,533,607 (66,604)

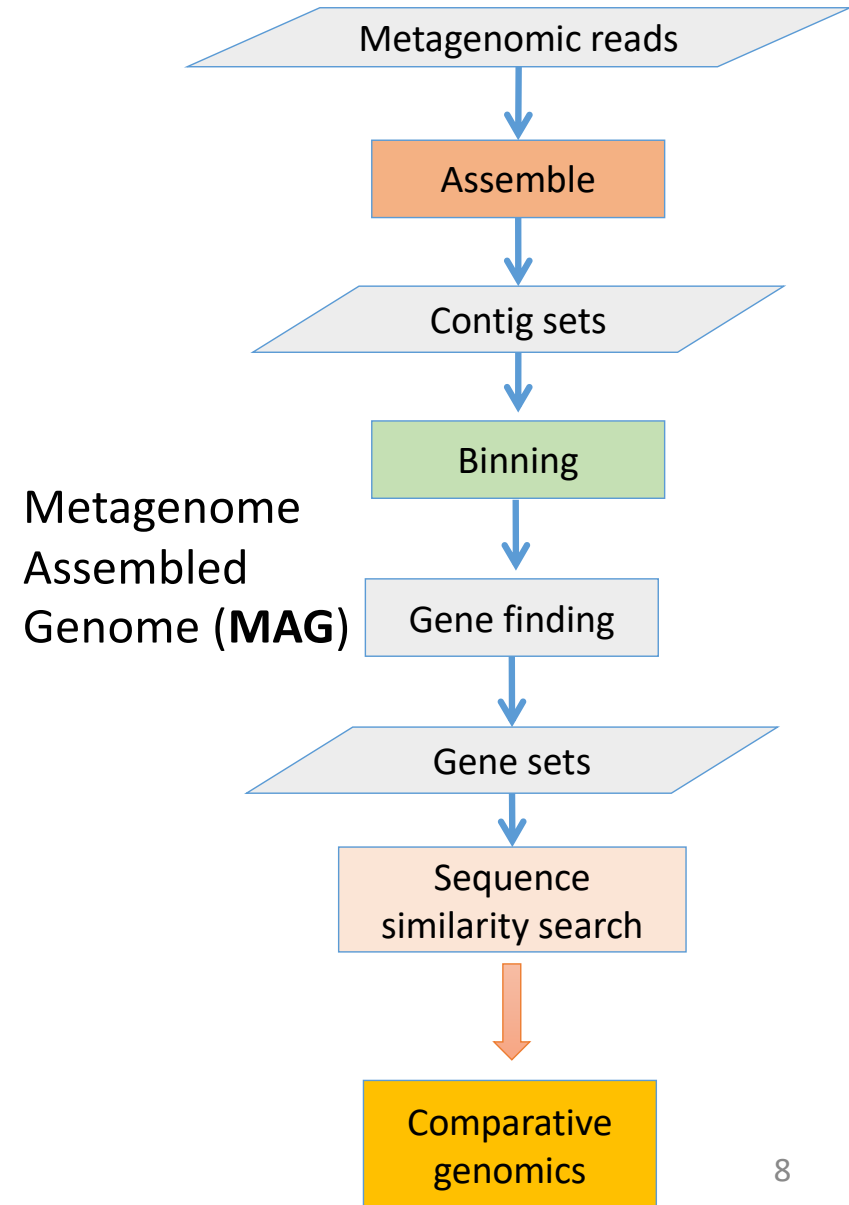
Metagenome Assembled Genome (MAG)

Metagenome analysis strategy

Assembly approach



Assembly + **Binning** approach



k-mer組成やcoverageを用いてメタゲノムContigを分けるツールの例

CONCOCT	Genome binner using differential coverage, tetranucleotide frequencies, paired-end linkage	Near complete (>95%) assignment of datasets at some cost for average genome purity and completeness.
MaxBin 2.0	Genome binner using multi-sample coverage, tetranucleotide frequencies	Largest average purity and completeness across entire abundance range. Recovery of 2 nd most genomes with high purity and completeness.
MetaBAT2	Genome binner using multi-sample coverage, tetranucleotide frequencies, paired-end linkage	Assignment of a large portion (>88%) of datasets at some costs for average genome purity and completeness.
MetaWatt-3.5	Genome binner using tetranucleotide frequencies	Recovery of the most genomes with high purity and completeness; near complete assignment of datasets at some cost for average genome purity and completeness.
MyCC	Genome binner using short k-mer frequencies, multi-sample coverage, and 40 universal phylogenetic marker genes	Near complete assignment of datasets at some cost for average genome purity and completeness.

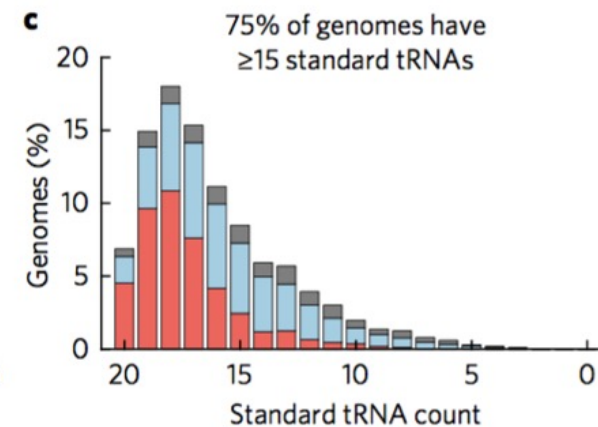
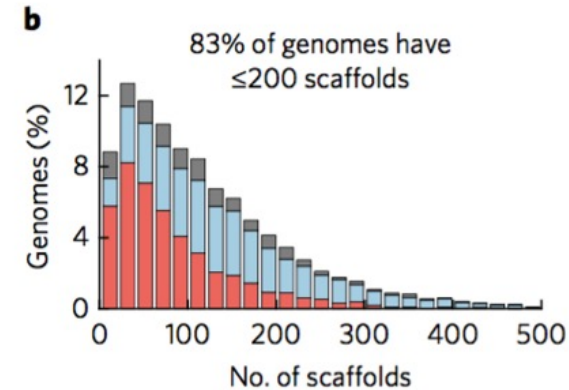
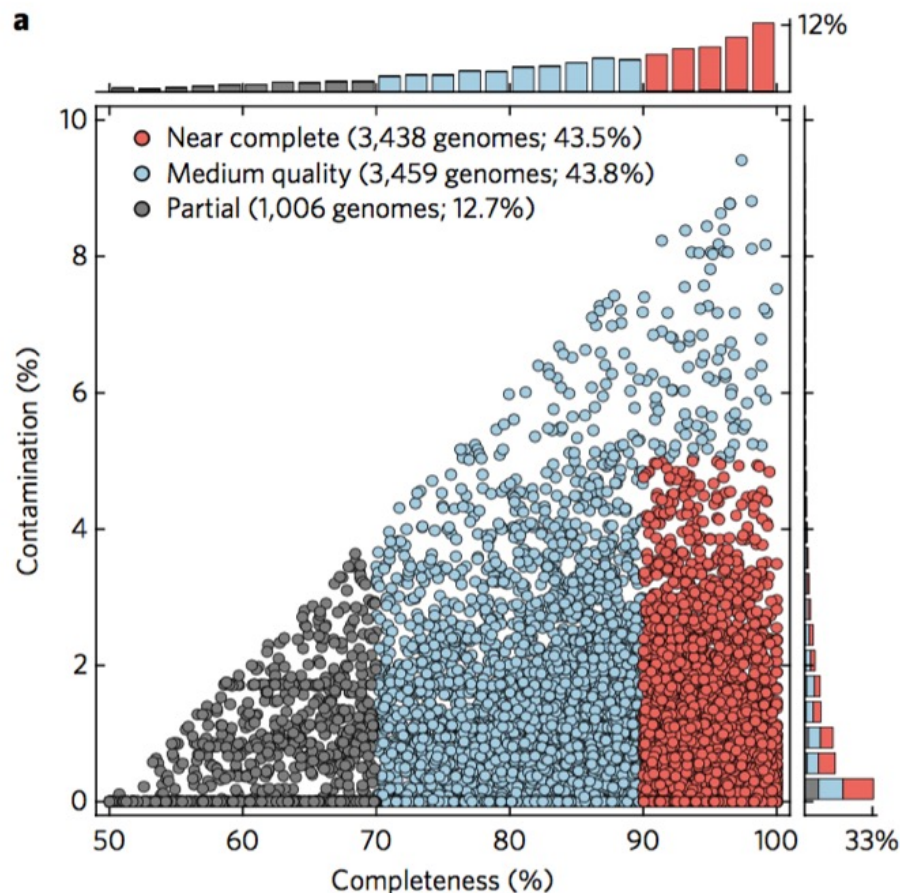
Sczyrba et al. 2017, Nature Methods

メタゲノムデータの全体像を議論するのではなく、優占系統のドラフトゲノム配列を抽出して各ゲノムが持つ機能について議論する

Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life

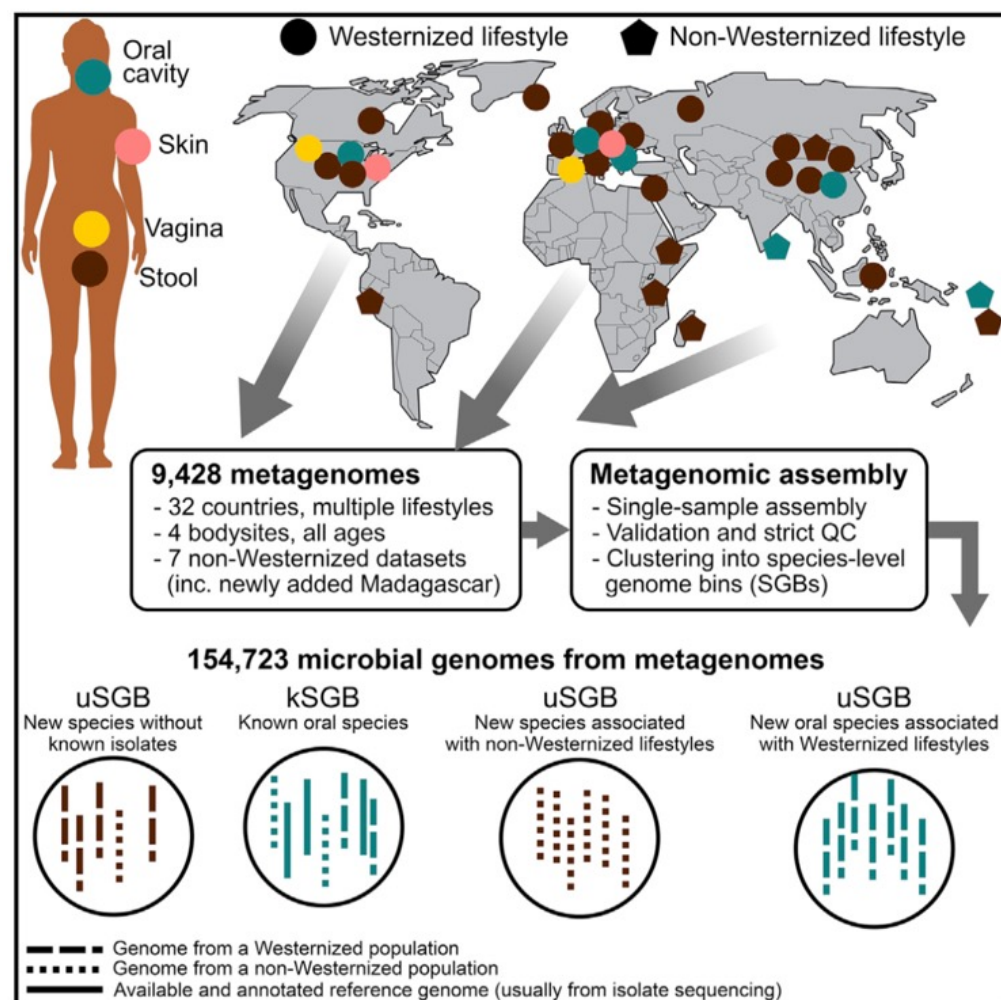
Donovan H. Parks , Christian Rinke , Maria Chuvpochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz * and Gene W. Tyson*

Metagenome Assembled Genomes (MAGs)



Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle

Graphical Abstract



Authors

Edoardo Pasolli, Francesco Asnicar, Serena Manara, ..., Christopher Quince, Curtis Huttenhower, Nicola Segata

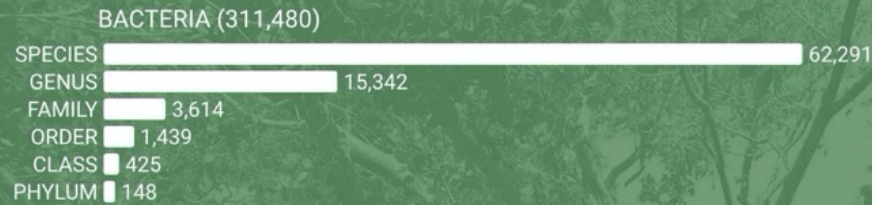
Correspondence

nicola.segata@unitn.it

In Brief

The human microbiome harbors many unidentified species. By large-scale metagenomic assembly of samples from diverse populations, we uncovered >150,000 microbial genomes that are recapitulated in 4,930 species. Many species (77%) were never described before, increase the mappability of metagenomes, and expand our understanding of global body-wide human microbiomes.

*** GTDB Release 207 is now available  download files ***



Welcome to GTDB

GENOME TAXONOMY DATABASE

317,542 genomes

Release 07-RS207 (8th April 2022)



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA



The Code of Nomenclature of Prokaryotes Described from Sequence Data

Edited by William B Whitman ^a, Maria Chuvochina ^b, Brian P Hedlund ^c, Philip Hugenholtz ^b, Kostas T Konstantinidis ^d, Alison E Murray ^e, Marike Palmer ^c, Donovan H Parks ^b, Alexander J Probst ^f, Anna-Louise Reysenbach ^g, Luis M Rodriguez-R ^h, Ramon Rossello-Mora ⁱ, Iain Sutcliffe ^j and Stephanus N Venter ^k

SeqCode Registry

cognitis nomina    

The SeqCode Registry is currently in **beta testing** phase, and changes to the layout are being actively introduced. If you have any questions or suggestions, [contact us](#).

SeqCode Registry

Part of the [SeqCode initiative](#), a path forward for naming the uncultivated



Register New Names or New Types



Search and Explore Registered
Names and Publications



Read and Learn More About the
SeqCode

Names 

Search



We have identified or registered **729 names** from **2707 publications** by **10793 authors**.

<https://disc-genomics.uibk.ac.at/seqcode/>

培養を経ずにメタゲノム配列データから 機能や進化的な類縁関係が推定された系統が多数存在

In Sept. 2022

Taxonomic rank	Cultured and validly described taxa	Genome-based putative taxa including uncultured taxa
Phylum	42	166
Class	147	477
Order	276	1,571
Family	656	4,070
Genus	3,865	16,686
Species	22,601	65,703

List of prokaryotic names with standing in nomenclature

(<https://lpsn.dsmz.de/text/numbers>)

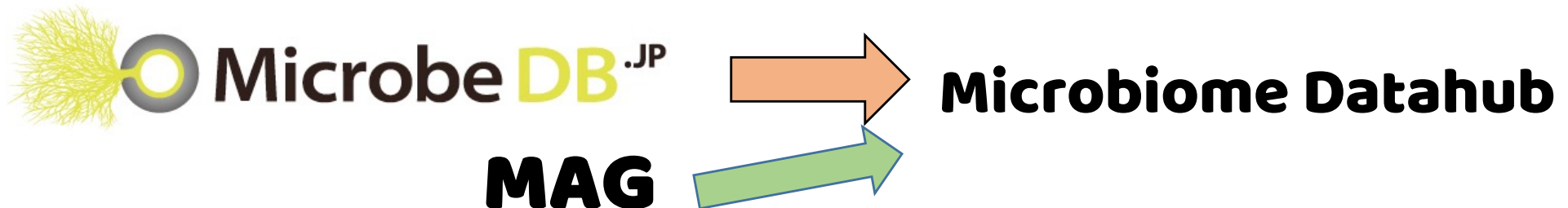
Genome Taxonomy Database (<https://gtdb.ecogenomic.org/stats>)

本研究開発の構想・目標

MicrobeDB.jpを
マイクロバイオームに特化した統合DBに作り変える

- ・爆発的なデータの増加に対応しいち早くデータを収録
- ・マイクロバイオームデータに特化した検索機能やUI
- ・様々な解析ワークフローに対応したデータ解析機能の実装

爆発的な勢いで増加するマイクロバイオームデータをいち早く収録し、検索・解析可能な統合DBとして、MicrobeDB.jpをマイクロバイオーム研究の国際的なデータハブへ発展させる。



Microbiome Datahub

MicrobeDB.jpの重要なデータや機能は引き継ぎつつ、
<https://mdatahub.org> で新DBを構築して今年度末公開予定

研究開発実施体制

国立遺伝学研究所

森宙史：プロジェクト全体の研究統括

中村保一：DDBJ & 他の統合DBとの連携

藤澤貴智：効率的な逐次DB更新・運用システムの”再”構築

東光一、谷澤靖洋：MAGデータの統合・キラーアプリケーションの開発

雇用予定研究員：マイクロバイオームデータに関する開発

基礎生物学研究所

内山郁夫：オーソログデータのマイクロバイオームデータとの統合

千葉啓和：オーソログ情報のRDF化と解析手法の効率化（DBCCLS）

西出浩世：オーソログデータの開発補助

東京工業大学

山田拓司：ヒトマイクロバイオームデータのキラーデータセットの開発

中川善一：ヒトマイクロバイオームデータのメタデータ整備

東京大学

松井求：微生物PhenotypeデータのRDF化とキラーアプリケーション開発

鈴木誉保、西村祐貴、雇用予定研究員：微生物Phenotypeデータの構築

Microbiome Datahubの 主な研究開発項目

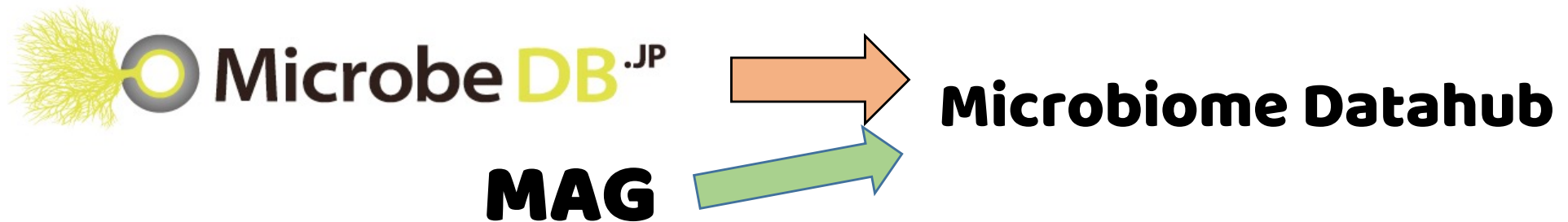
1. ゲノム・メタゲノム(MAG)の統合化
2. メタゲノムとリソース（菌株）との連結
3. データ爆発への対応
4. 効率的な逐次DB更新・運用システムの"再"構築
5. キラーデータセット・アプリケーションの開発
6. マルチオミックス連携

ゲノムとMAGの統合化、MAGのクオリティチェック、
メタゲノムと菌株の連結等、
MAGを中心にマイクロバイオーームデータを整理したDBとして
Microbiome Datahubの開発を進めております

まとめ

MicrobeDB.jpを
マイクロバイオーームに特化した統合DBに作り変える

爆発的な勢いで増加するマイクロバイオーームデータをいち早く収録し、検索・解析可能な統合DBとして、MicrobeDB.jpをマイクロバイオーーム研究の国際的なデータハブへ発展させる。



ポスター 第1部 14:20～15:20 10番