

 Microbe DB<sup>.JP</sup> データを用いた  
メタゲノム解析Webアプリケーションの開発

森 宙史  
Hiroshi Mori

東京工業大学  
大学院生命理工学研究科  
生命情報専攻 黒川・山田研 助教

# OUR MICROBIAL PLANET

**MICROBES**—life forms too tiny to see—play a surprisingly large role in life on Earth. Microbes are everywhere, and they do a lot of good for human health and our planet. In fact, disease-causing microbes make up only a very tiny fraction of the millions of types of microbes. Microbes...

**Think microbes are bad guys? Think again.**



**Keep us healthy.** Amazingly, only about 1 out of 10 cells in the human body is actually a human cell; most of the cells in our bodies are microbes! Some of the microbes living in our bodies actually help us fight disease-causing microbes by competing against them for space. This mutually beneficial relationship helps to protect us from getting diseases while giving the "good" microbes a place to live.

**Make air breathable.** Without microbes, we wouldn't have oxygen to breathe. This is because many microbes are photosynthetic—like plants, they harvest their energy from the sun, releasing oxygen into the air. Billions of years ago, photosynthetic microbes gradually added oxygen to Earth's atmosphere, making it possible for larger forms of life—including humans—to live.

**Provide sources of new medicines.** Hundreds of medicines available today were derived from chemicals first found in microbes. Microbes naturally produce an amazing variety of chemicals, which scientists can use to create new medicines.

**Help us digest food.** Many of the foods we eat would be indigestible without the 10-100 billion microbes living within our guts. Microbes also play a major role in creating many of the foods we love, such as cheese, yogurt and bread.



**Keep our environment clean.**

Because of their special adaptations, some microbes can help clean up gasoline leaks, oil spills, sewage, nuclear waste, and many other types of pollution.



**Support and protect crops.**

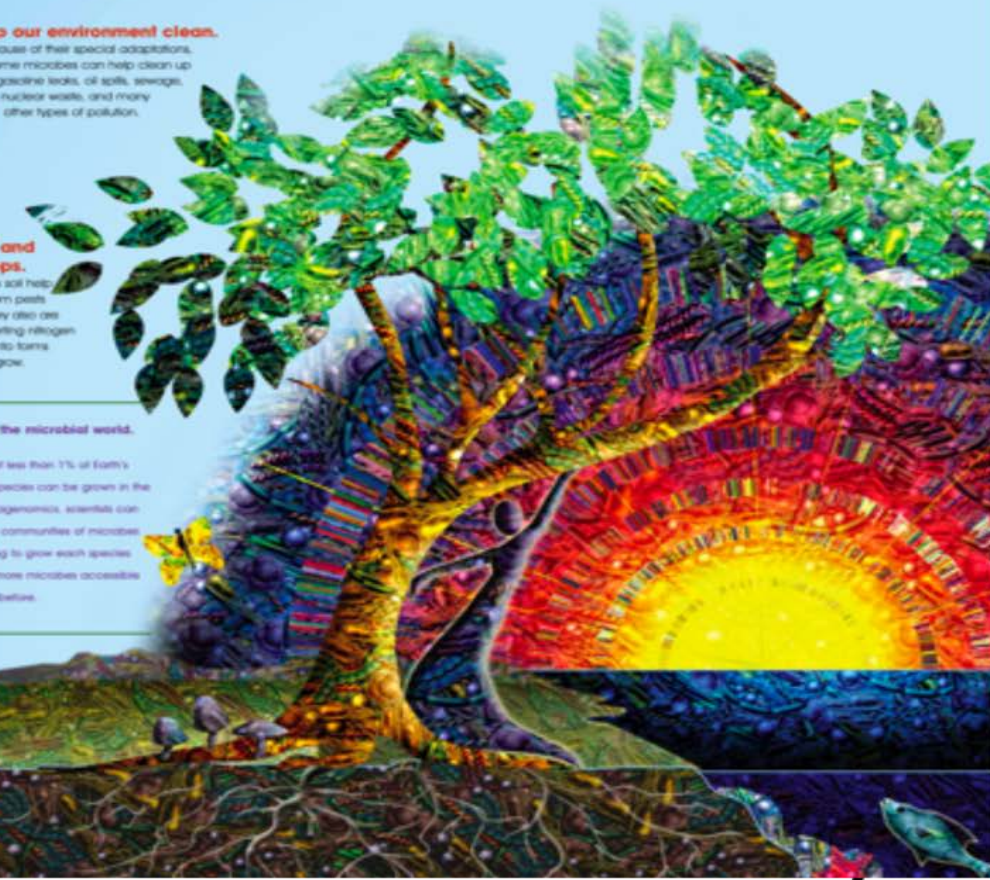
Microbes living in soil help protect plants from pests and diseases. They also are essential for converting nitrogen and other nutrients into forms that plants can use to grow.

**The science of metagenomics is shedding new light on the microbial world.**



Scientists estimate that less than 1% of Earth's millions of microbial species can be grown in the laboratory. Using metagenomics, scientists can now study how whole communities of microbes function without having to grow each species separately—making more microbes accessible to science than ever before.

**Living in a microbial world...**



Metagenome



This Figure is obtained from National Research Council (USA)

Obtained from Togo Picture

# 16S rRNA gene amplicon sequencing analysis (メタ16S解析)



DNA extraction

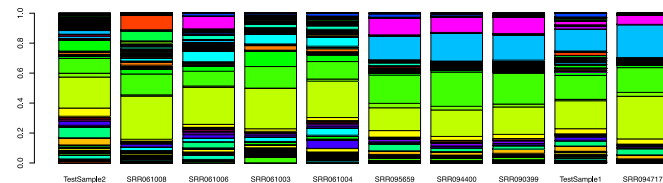
PCR amplification

DNA Sequencing

Pre-analysis (Remove Primer, Chimera etc.)

Sequence clustering with species level  
by CD-HIT-EST or UCLUST, etc.

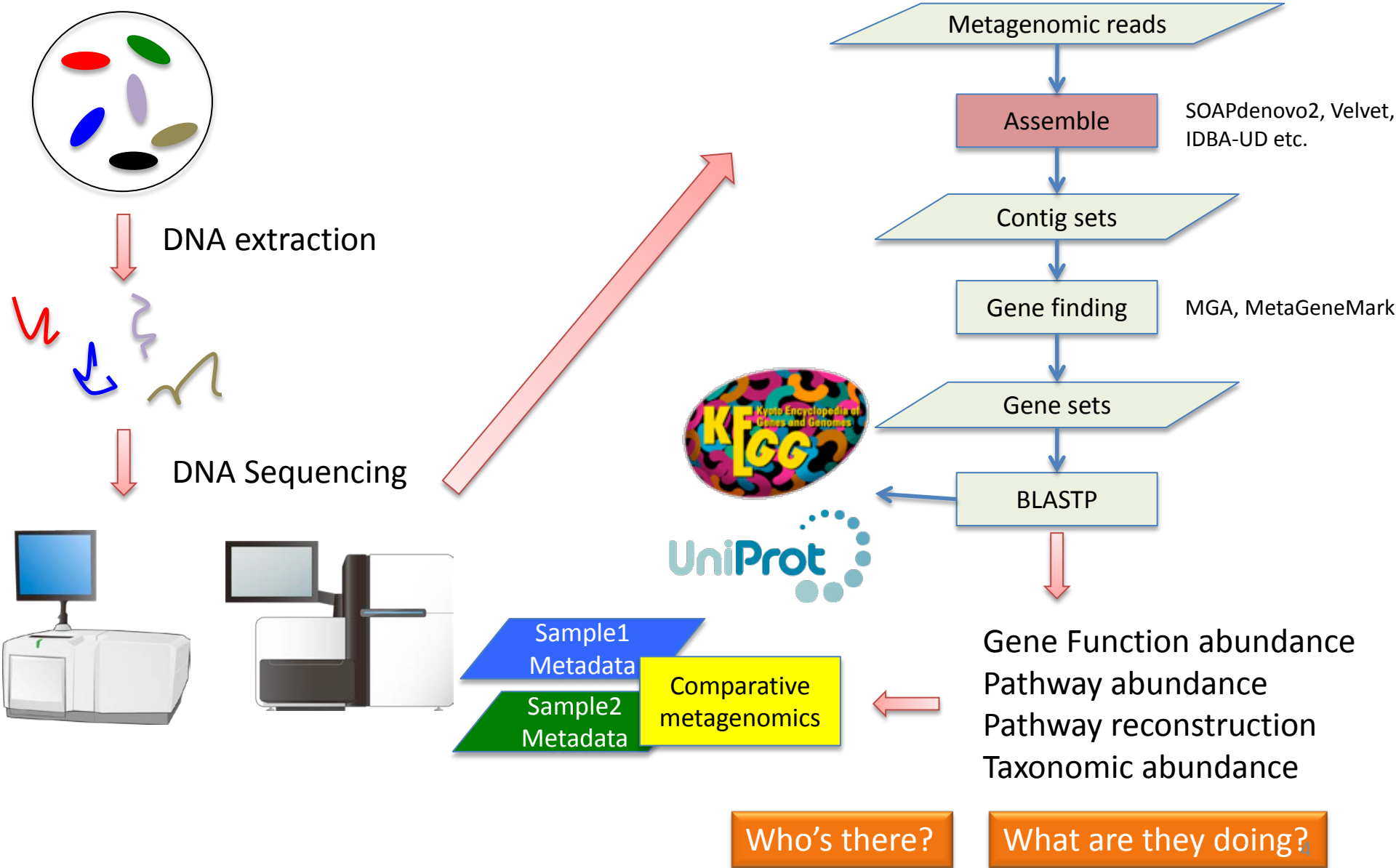
Taxonomic assignment and  
Comparison between samples



Who's there?



# Metagenomic sequencing analysis (メタゲノム解析)

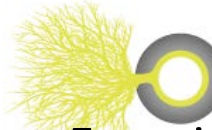


- メタ16S・メタゲノム解析データ(塩基配列データ)から、容易に系統組成および遺伝子機能組成の情報を抽出したい
- 組成情報を既存のメタ16S・メタゲノム解析サンプルと比較したい



その環境の細菌群集は他のサンプルと比較するとどのような共通性・違いがあるか？

その環境のメタデータと自分のサンプルのメタデータを比べて、何が異なるのか？

 **Microbe DB .JP** integrates lots of data related to microbes.  
Especially, we integrate the microbial data that can be linked to **genomes**.



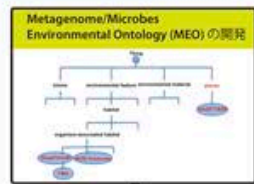
# Microbe DB .JP

<http://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使用して整理統合し、幅広い分野での微生物学の見解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。

## Ontology

オントロジー: 検索タームの柔軟化&明確化



Gene

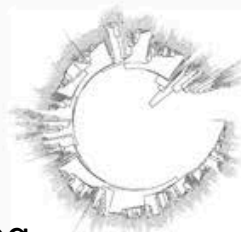
Taxon

Environment



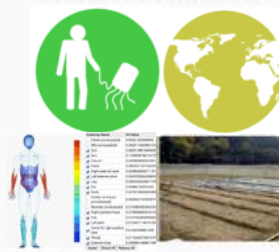
Ortholog: **MBGD**

オースログデータ



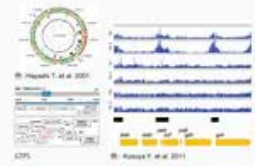
Taxonomy:  
NCBI Taxonomy

系統分類データ



Metadata:  
INSDC SRA

環境のメタデータ



Genome: **GTPS/RefSeq**

オミックスデータ



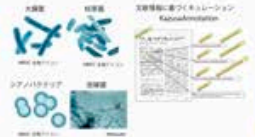
Culture Collection:  
**NBRC/JCM**

菌株データ  
菌株採得情報(培養条件含む)



Metagenome:  
INSDC SRA

メタゲノムデータ



Annotation:  
**TogoAnnotation**

モデル微生物の高品質  
アノテーションデータ



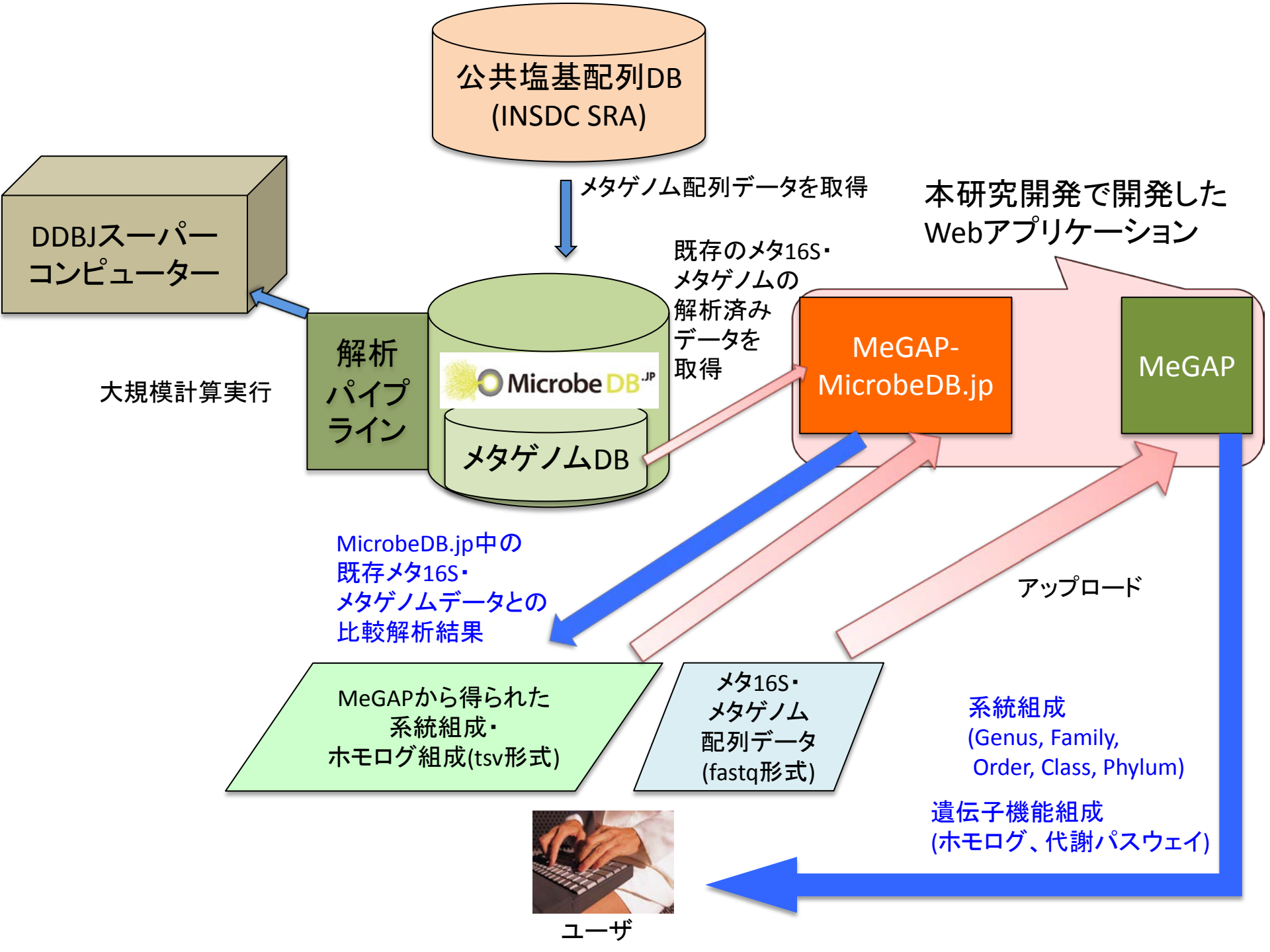
Red color indicates our collaborators.

# 既存のメタゲノム・メタ16Sデータの整理・統合

微生物の情報とその生息環境の情報を結びつける

1. 公共DB(SRA)からメタ16S・メタゲノム配列データとメタデータを取得
2. 配列データから系統組成・遺伝子機能組成情報を計算
3. メタ16S・メタゲノム論文から微生物の生息環境についてのメタデータを抽出
4. 生息環境を記述するために同義語、語彙間の関連性を記述したオントロジーMEOを構築
5. MEOとメタ16S・メタゲノムのメタデータの対応付け

サンプルの種類	サンプル数
メタ16S	23952
メタゲノム	1078





# MeGAP (MetaGenome Annotation Pipeline)

<http://fs2.bio.titech.ac.jp/megap/>

# MeGAP

English

MeGAP is a MetaGenome Annotation Pipeline.

## Try MeGAP

16S rRNA gene Amplicon Sequencing FASTQ file:  ファイルが選択されていません。

File format1:  flat file  .gz

ID:  (use [A-Za-z0-9-\_])

Email:

Metagenome Sequencing FASTQ file:  ファイルが選択されていません。

File format1:  flat file  .gz

ID:  (use [A-Za-z0-9-\_])

Email:

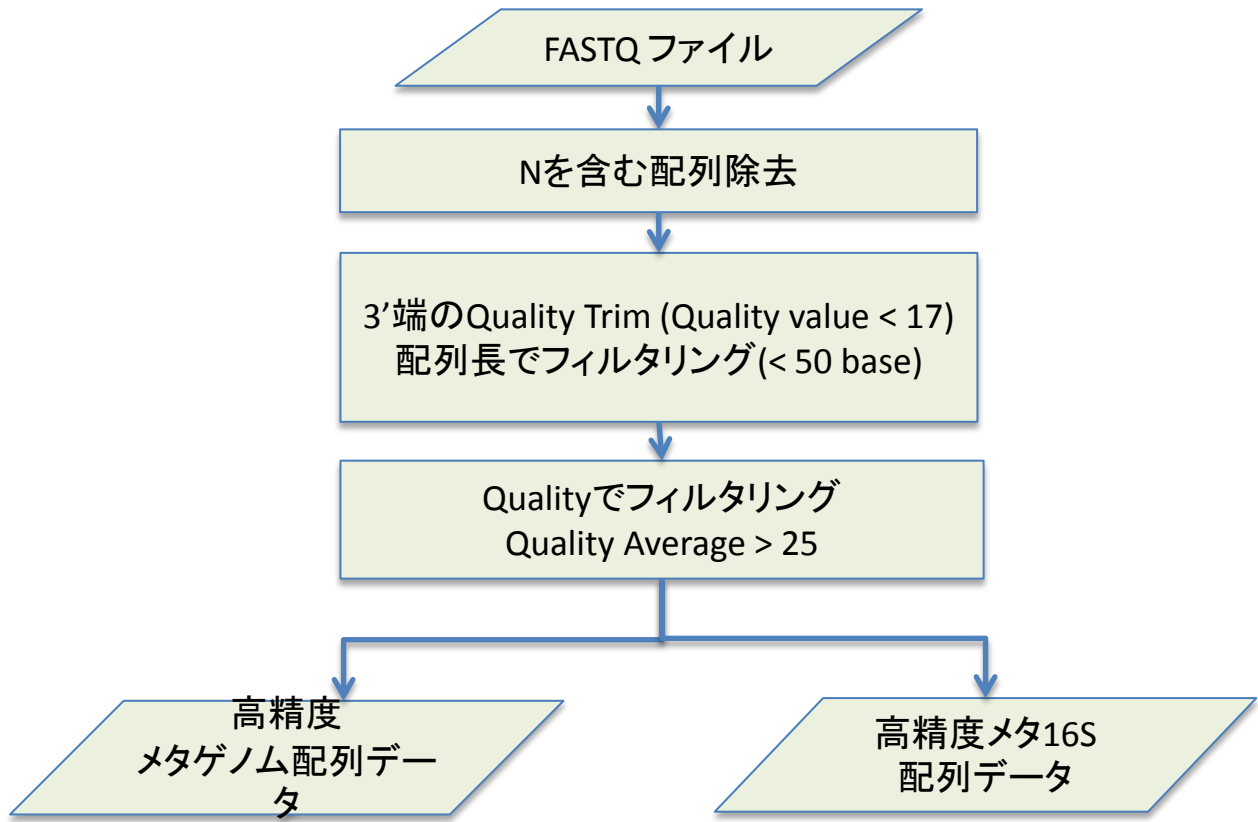
## How to use

### 1. Input data

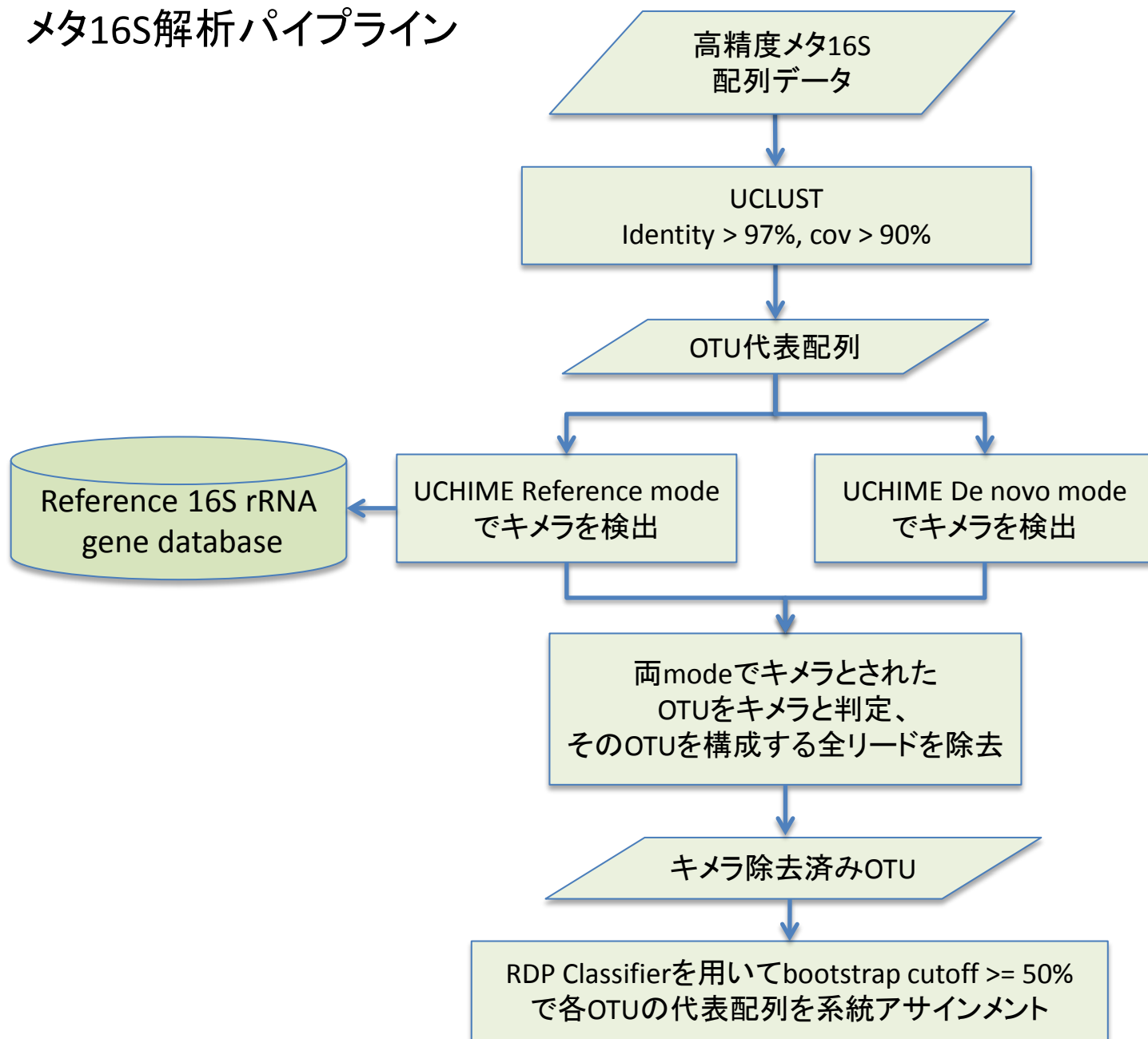
Both of a FASTQ file and gzipped FASTQ file are acceptable for the input data in the MeGAP. Sample 16S rRNA gene Amplicon sequencing fastq data and Sample metagenome sequencing fastq.gz data.

# 配列の前処理 (目的: メタゲノムとメタ16Sデータの区別 & 高精度配列データの抽出)

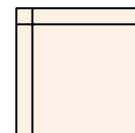
ユーザがメタ16S or メタゲノムの情報を入力



# メタ16S解析パイプライン



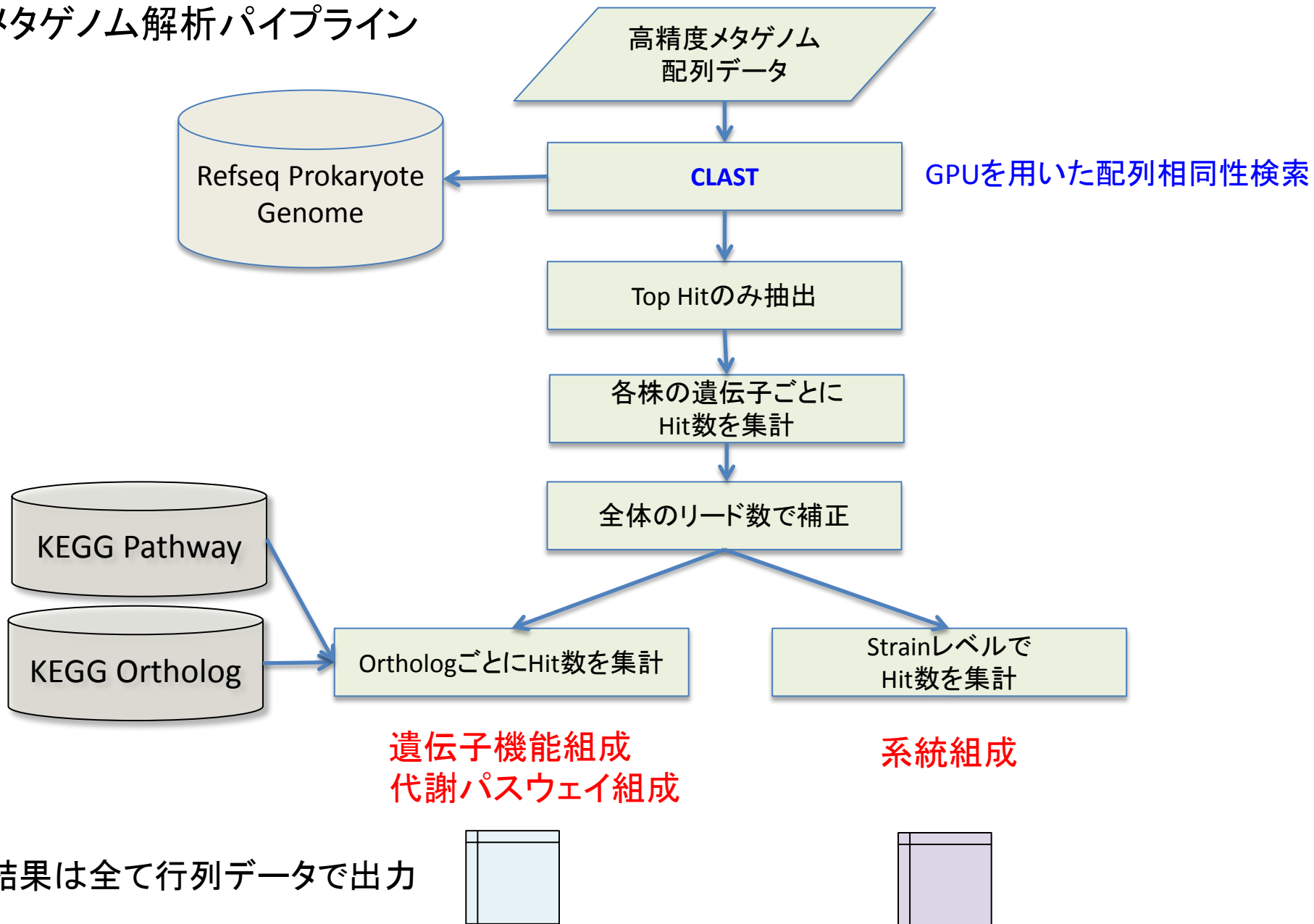
Genus-Phylumレベルの系統組成



# 系統組成

Classified	18000	
UnClassified	9200	
Abiotrophia	1	
Achromobacter	1	
Actinobacillus	5	
Actinomyces	8	
Anaeroglobus	113	
Asticcacaulis	15	
Atopobium	33	
Bifidobacterium	86	
Brevibacillus	2	
Bulleidia	3	
Campylobacter	93	
Capnocytophaga	7	
Catonella	6	
Centipeda	19	
Clostridium	XIX	5
Dialister	37	
Eikenella	34	
Escherichia/Shigella		234
Eubacterium	35	
Fusobacterium	2	
Gemella	180	
Gp19	1	
Granulicatella	10	
Haemophilus	76	
Hallella	4	
Howardella	2	
Johnsonella	1	
Kocuria	2	
Leptotrichia	10	
Megasphaera	21	
Methylobacterium	14	
Mogibacterium	10	
Mycoplasma	13	
Neisseria	56	
Nicoletella	1	
Novosphingobium	5	
Olsenella	12	
Oribacterium	13	
Paludibacter	21	
Paraprevotella	38	

# メタゲノム解析パイプライン



Chlamydia trachomatis D/UW-3/CX, complete genome - 1..1042519 0  
 Thermotoga maritima MSB8 chromosome, complete genome - 1..1860725 0  
 Aeropyrum pernix K1, complete genome - 1..1669696 0  
 Pyrococcus abyssi GE5 chromosome, complete genome - 1..1765118 0  
 Haemophilus influenzae Rd KW20 chromosome, complete genome - 1..1830138 0.022  
 Mycoplasma genitalium G37, complete genome - 1..580076 0  
 Methanocaldococcus jannaschii DSM 2661 chromosome, complete genome - 1..1664970 0  
 Synechocystis sp. PCC 6803 chromosome, complete genome - 1..3573470 0  
 Mycoplasma pneumoniae M129 chromosome, complete genome - 1..816394 0  
 Escherichia coli str. K-12 substr. MG1655 chromosome, complete genome - 1..4639675 0.0166  
 Sinorhizobium fredii NGR234 plasmid pNGR234a, complete sequence - 1..536165 0  
 Helicobacter pylori 26695 chromosome, complete genome - 1..1667867 0  
 Methanothermobacter thermoautotrophicus str. Delta H chromosome, complete genome - 1..1751377 0  
 Archaeoglobus fulgidus DSM 4304, complete genome - 1..2178400 0  
 Aquifex aeolicus VF5, complete genome - 1..1551335 0  
 Treponema pallidum subsp. pallidum str. Nichols chromosome, complete genome - 1..1138011 0  
 Helicobacter pylori J99 chromosome, complete genome - 1..1643831 0  
 Chlamydia pneumoniae CWL029 chromosome, complete genome - 1..1230230 0  
 Borrelia burgdorferi B31 plasmid cp32-1, complete sequence - 1..30750 0  
 Borrelia burgdorferi B31 plasmid cp32-3, complete sequence - 1..30223 0  
 Borrelia burgdorferi B31 plasmid cp32-4, complete sequence - 1..30299 0

## 遺伝子機能組成

140223213131test1

## 代謝パスウェイ組成

140223213131test1

K00001	0.01	ko00010	0.61
K00002	0.03	ko00020	0.32
K00003	0.11	ko00030	0.44
K00004	0	ko00040	0.11
K00005	0.03	ko00051	0.42
K00006	0.05	ko00052	0.24
K00007	0.01	ko00053	0.02
K00008	0	ko00061	0.12
K00009	0.02	ko00062	0.42
K00010	0.07		

## MeGAP-MicrobeDB.jp

MeGAP-MicrobeDB.jp is a comparative metagenomic analysis tool.

### Try MeGAP-MicrobeDB.jp

16S rRNA gene Amplicon Sequencing MeGAP Genus composition file:  ファイルが選択されていません。

File format1:  flat file  .gz

ID:  (use [A-Za-z0-9-\_])

Email:

Metagenome Sequencing MeGAP KO composition file:  ファイルが選択されていません。

File format1:  flat file  .gz

ID:  (use [A-Za-z0-9-\_])

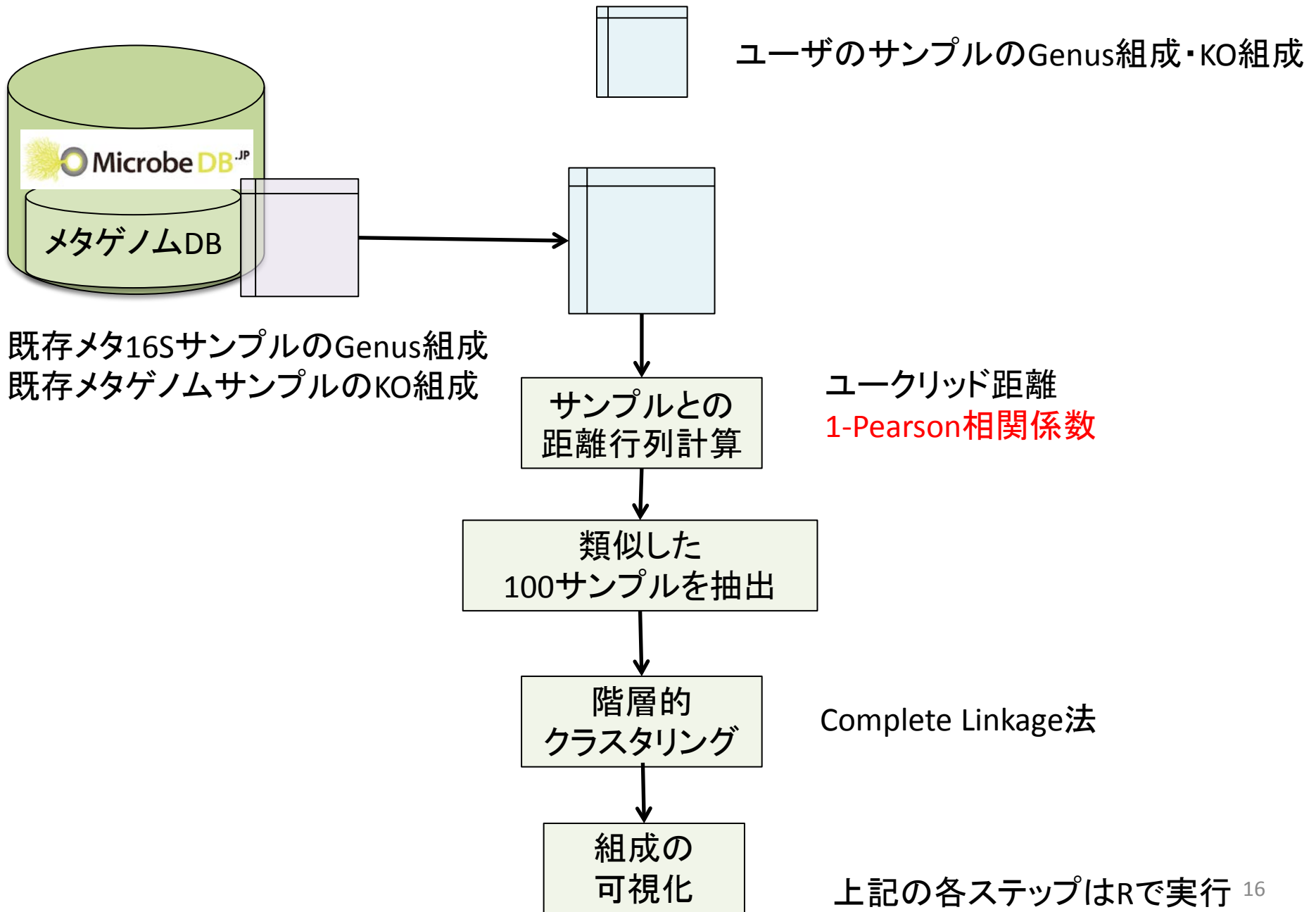
Email:

### How to use

#### 1. Input data

Input data is a MeGAP output tsv file of Genus or KO composition (Sample KO matrix) of your sample.

# MicrobeDB.jp中の既存サンプルとの比較解析





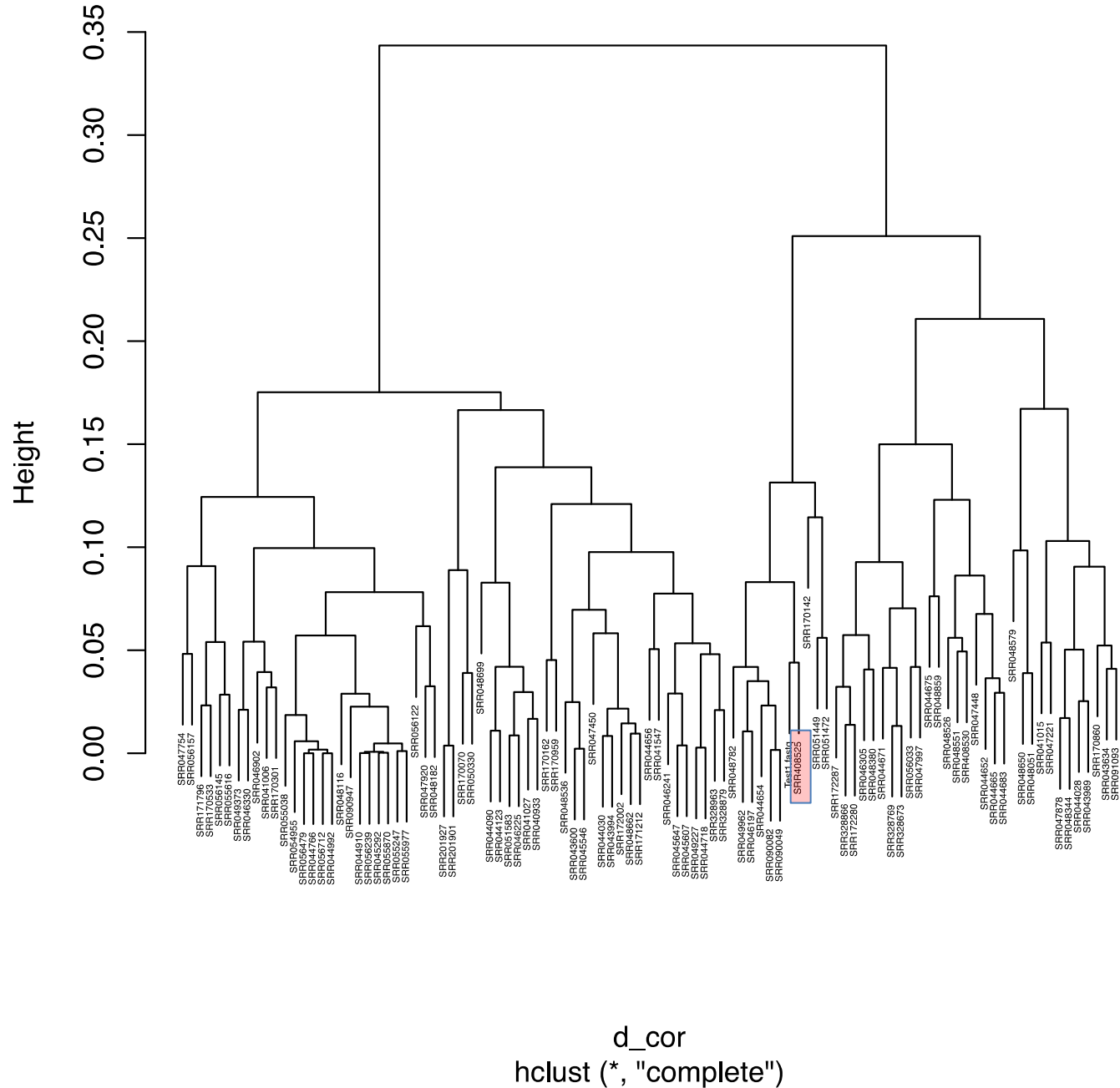
## 類似した100サンプルとのKO組成 (メタゲノム)

```
X140223213131test1 SRS257937 SRS085663 SRS000238 SRS140432 SRS150249
K00001 0 0 0 0 0 0 0 0 0 0.00298127845146256 0 0 0.000196596821000062 (
K00002 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2.36348510058993e-06 0 0.000318674
K00003 0 0 0 0 0 0 0.000847637211273575 0.000359324469996406 0.00490:
K00004 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
K00005 0 0 0 0 0 0 0 0 0 0 0.000124223602484472 2.93428091044868e-06
K00006 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
K00007 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
K00008 0 0 0 0 0 0 0.000179662234998203 0 0 0 2.05399663731408e-05
K00009 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
K00010 0 0 0 0 0 0 0 0.000410677618069815 0 0 1.46714045522434e-06
K00011 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
K00012 0 0 0.000119602918311207 0.00101419878296146 0 0 0.000538984
K00013 0 0 0.000358808754933621 0 0 8.46353064644447e-06 0.000211909:
K00014 0 0 0 0 0.000562904587672389 0 0.000635727908455182 0.00017964
K00015 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

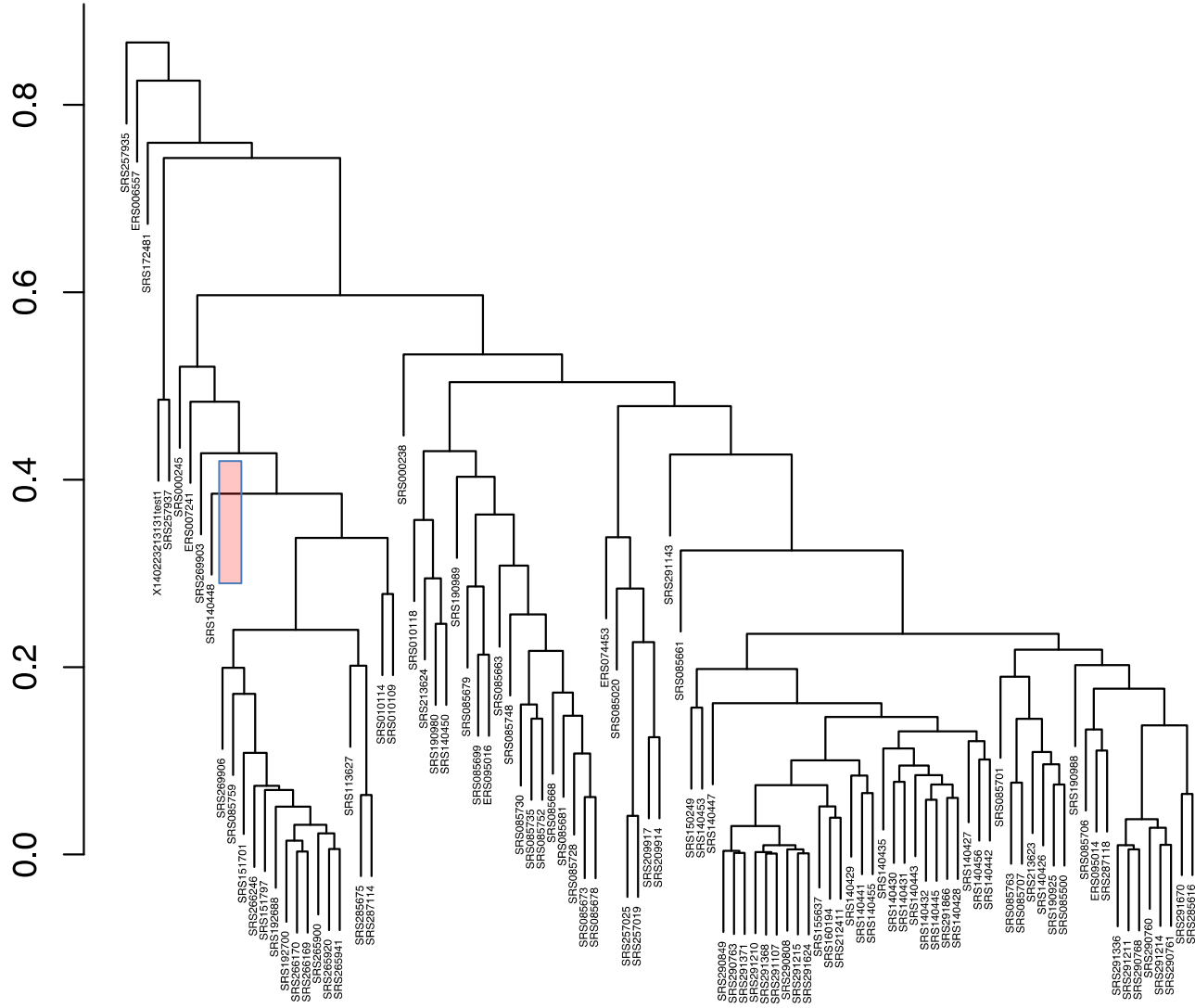
## 類似した100サンプルとのGenus組成 (メタ16S)

```
X.tmp.uploadcomp..140222080444test2 SRR040709 SRR046225 SRR050268
Abiotrophia 1 1 0 0 0 0 0 0 0 0 0 5 0 0 1 0 0 1 0 3 0 0 2 0 0 0 0 6
Acanthopleuribacter 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Acaricomies 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Acetanaerobacterium 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Acetitomaculum 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Acetivibrio 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Acetoanaerobium 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Acetobacter 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Acetobacterium 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

# Genus組成によるサンプル間階層的クラスタリングの結果 (メタ16S)



# KO組成によるサンプル間階層的クラスタリングの結果 (メタゲノム)



d\_cor  
hclust (\* "complete")

## まとめ

- メタ16S・メタゲノム解析データから、容易に  
系統組成および遺伝子機能組成の情報を抽出したい

### **MeGAP**

<http://fs2.bio.titech.ac.jp/megap/>

- 組成情報を既存のメタ16S・メタゲノム解析サンプルと比較したい

### **MeGAP-MicrobeDB.jp**

<http://fs2.bio.titech.ac.jp/megaptomicrobedb.jp/>

その環境の細菌群集は他のサンプルと比較するとどのような共通性・違いがあるか？

その環境のメタデータと自分のサンプルのメタデータを比べて、何が異なるのか？