

メタゲノム解析における スーパーコンピュータの利用例

森 宙史

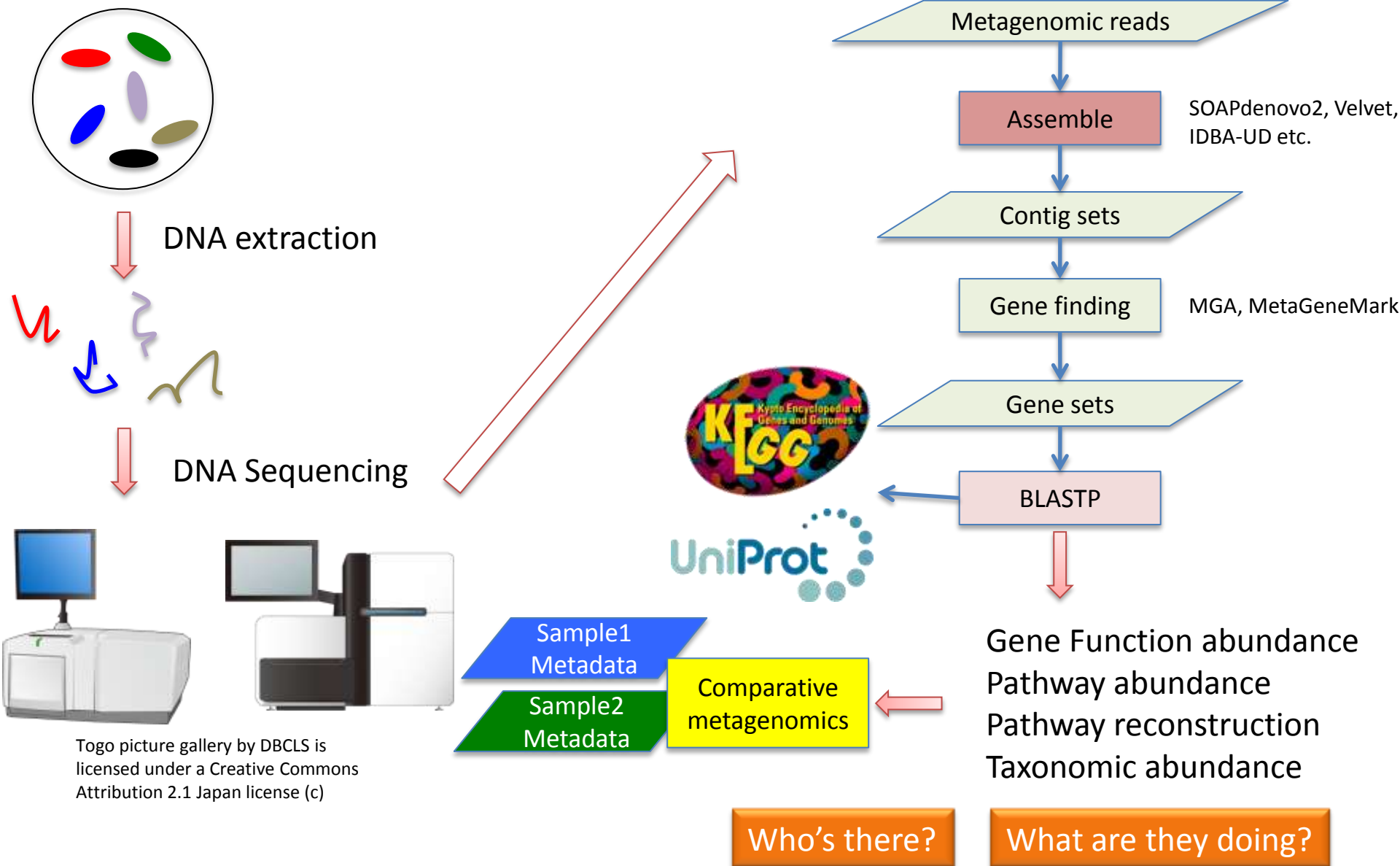
Hiroshi Mori, Ph.D.

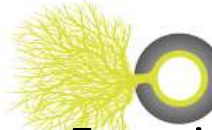
東京工業大学

大学院生命理工学研究科

生命情報専攻 黒川・中島・山田研

Metagenomic sequencing analysis (メタゲノム解析)



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



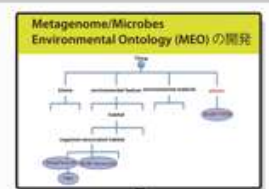
Microbe DB .JP

<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利用して整理統合し、幅広い分野での微生物学の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。

Ontology

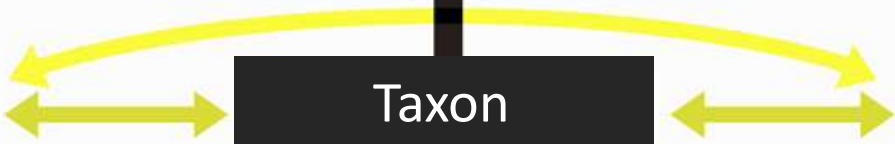
オントロジー: 検索タームの柔軟化&明確化



Gene

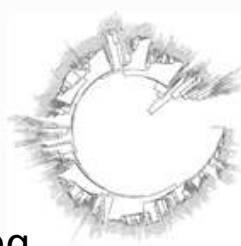
Taxon

Environment



MBGD
オースログデータ

Ortholog: **MBGD**



Taxonomy:
NCBI Taxonomy

環境のメタデータ

Metadata:
INSDC SRA

Genome: **GTPS/RefSeq**
オミックスデータ
Annotation:
TogoAnnotation
モデル微生物の高品質アノテーションデータ

Genome: **GTPS/RefSeq**

Culture Collection:
NBRC/JCM
菌株データ
菌株保存情報(培養条件含む)

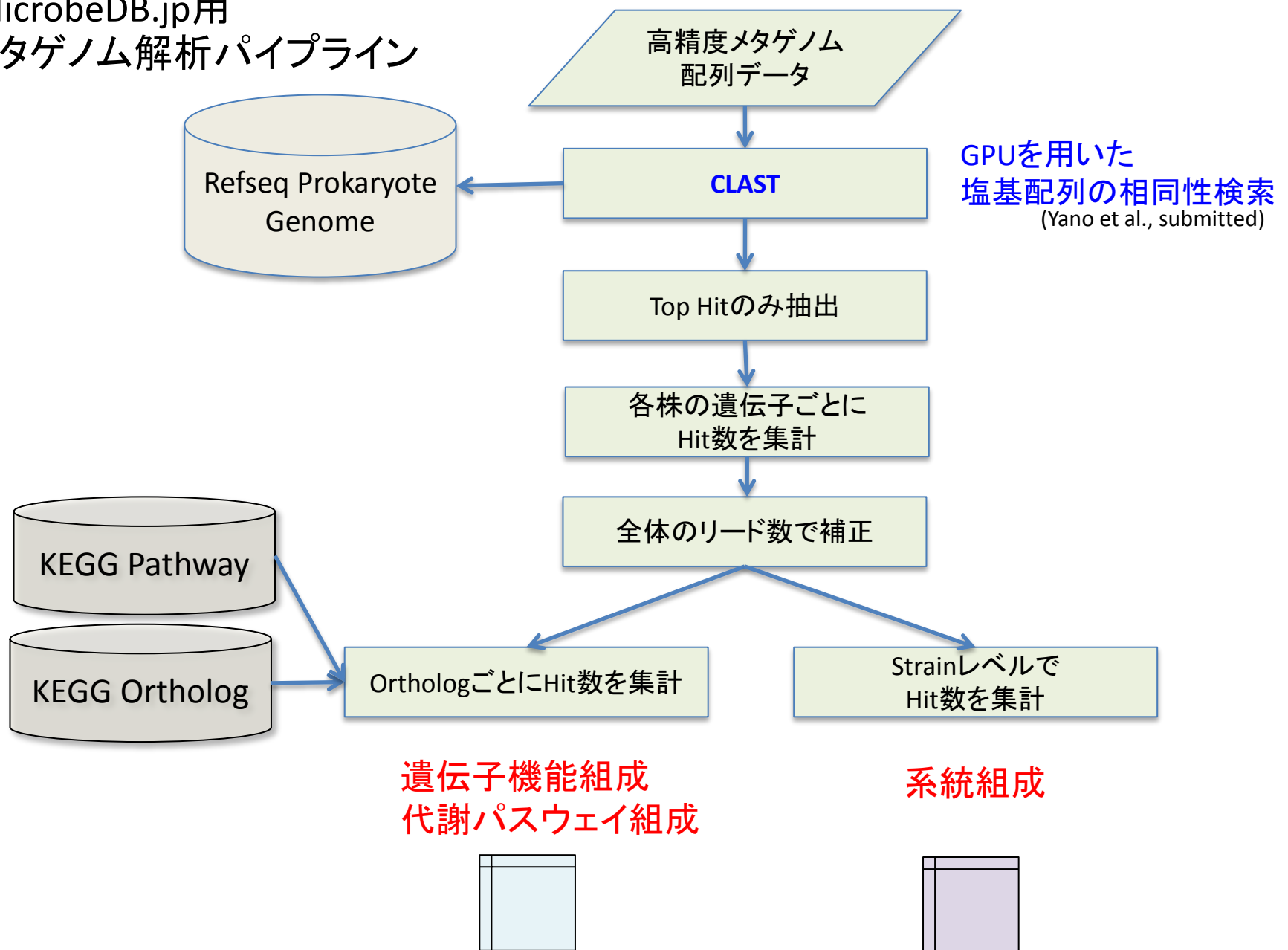
Culture Collection:
NBRC/JCM

Metagenome:
INSDC SRA
メタゲノムデータ

Metagenome:
INSDC SRA

Red color indicates our collaborators.

MicrobeDB.jp用 メタゲノム解析パイプライン



- ・メタゲノム解析では近縁種のReference genome配列が未知の場合が多いので、感度は高いが時間のかかる配列相同性検索を行う必要がある
Thinノードでクエリのアミノ酸配列データを分割して100-300並列で計算
- ・数千サンプルの比較解析を目的として、GPUノードで数十並列で計算

DRAの.lite.sra or .fastqファイルをスパコン上で読み込みのみできると、とても助かります。

GPUを使った計算で遺伝研スパコンを使う理由は、無料で使えることが大きい。

各Jobを課金制にすると、探索目的での計算が実行しにくくなる。

BLASTで多数の並列Jobの投げ方の指針を出した方が良くも。