

微生物の MAG データの統合データベース Microbiome Datahub

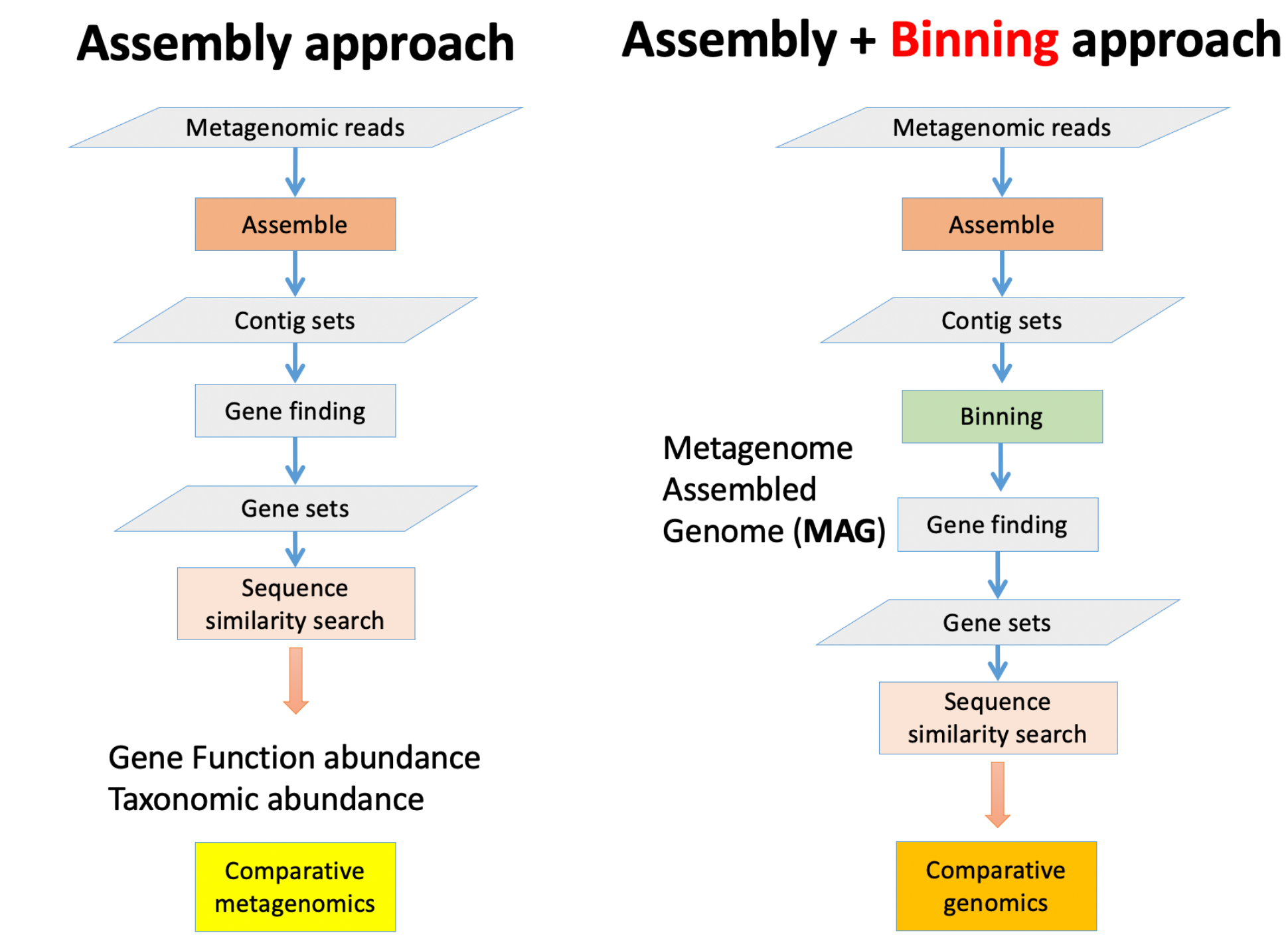


○森宙史（遺伝研）、藤澤貴智（遺伝研）、東光一（ROIS・ALIS）、谷澤靖洋（遺伝研）、中川善一（東京科学大）、西出浩世（基生研）、中村保一（遺伝研）、山田拓司（東京科学大）、松井求（京都大）、内山郁夫（基生研）

Microbiome Datahub (<https://mdatahub.org>)

Mori H. et al. Microbiome. 2026

Metagenome analysis strategy



In April 2026		
Taxonomic rank	Cultured and validly described taxa	Genome-based putative taxa including uncultured taxa
Phylum	46	186
Class	124	641
Order	255	2,343
Family	665	6,725
Genus	3,961	37,503
Species	22,175	199,923

List of prokaryotic names with standing in nomenclature (<https://psn.dsmz.de/text/numbers>)
Genome Taxonomy Database R232 (<https://gtdb.ecogenomic.org/stats>)

Microbiome Datahub project member

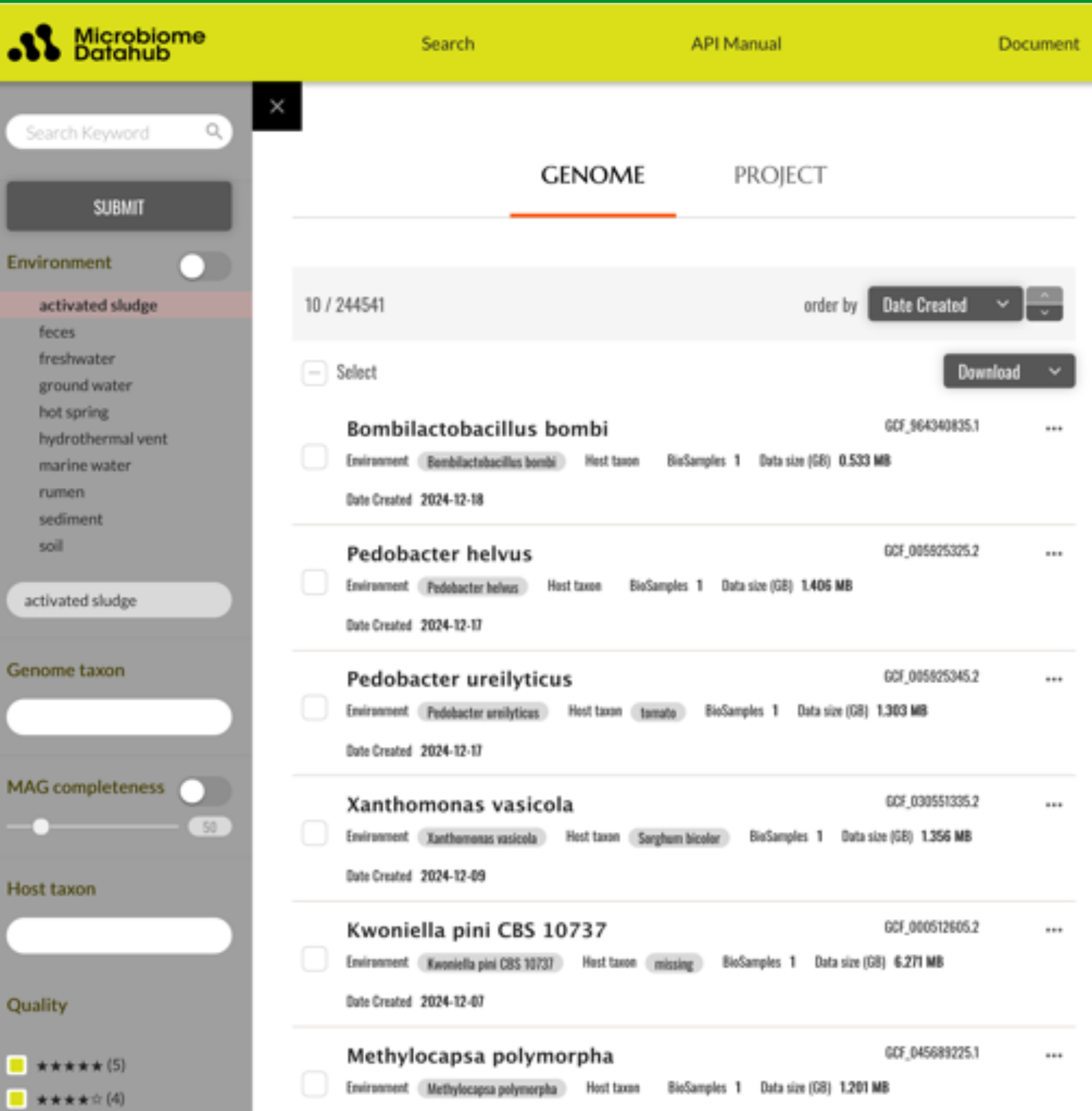
NIG (国立遺伝学研究所)
Hiroshi Mori, Takatomo Fujisawa, Koichi Higashi, Yasuhiro Tanizawa, Yasukazu Nakamura

NIBB (基礎生物学研究所)
Ikuo Uchiyama, Hirokazu Chiba, Hiroyo Nishide

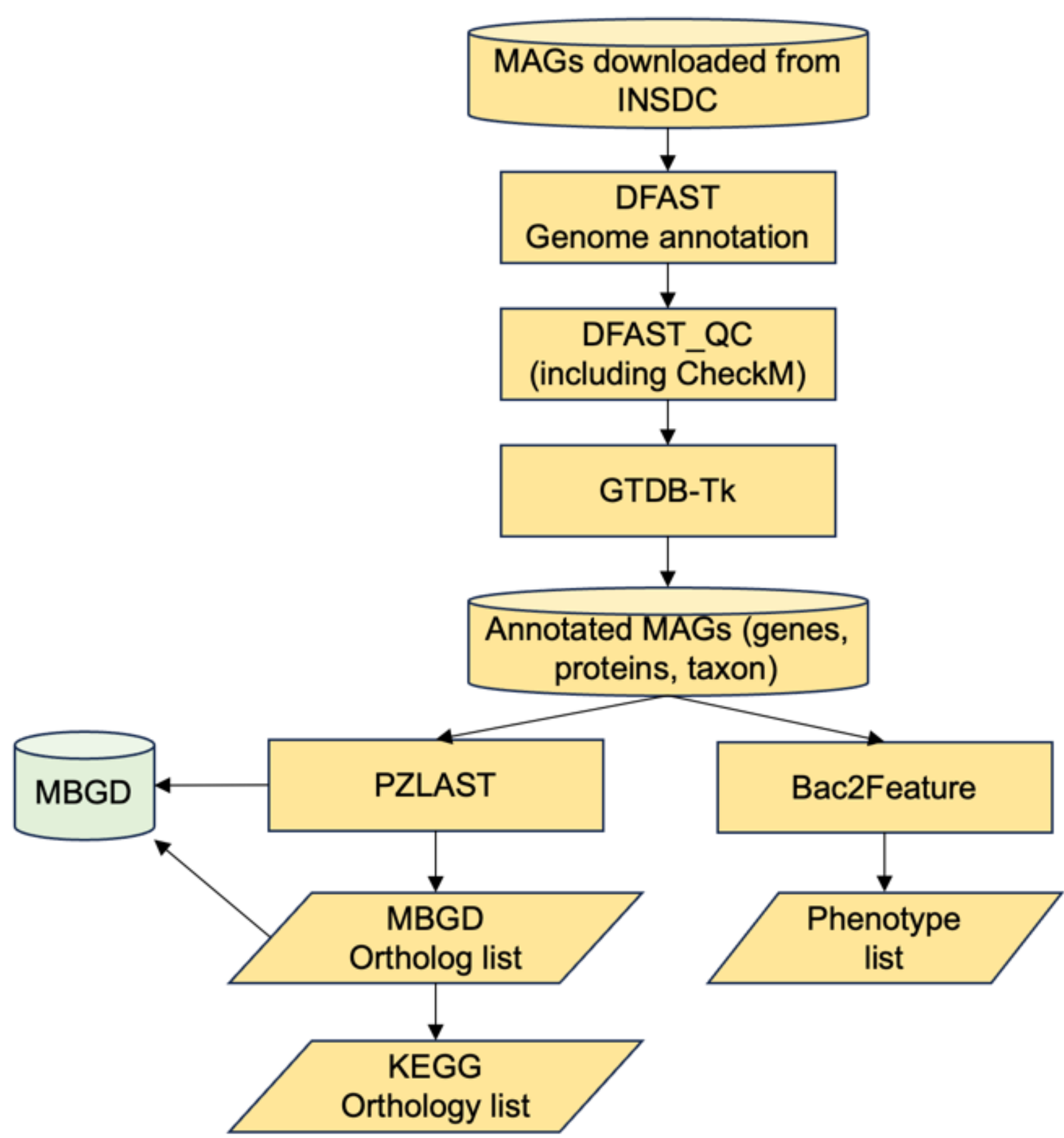
Institute of Science Tokyo (東京科学大学)
Takuji Yamada, Zenichi Nakagawa

Kyoto Univ. (京都大学)
Motomu Matsui, Takao Suzuki, Yuki Nishimura

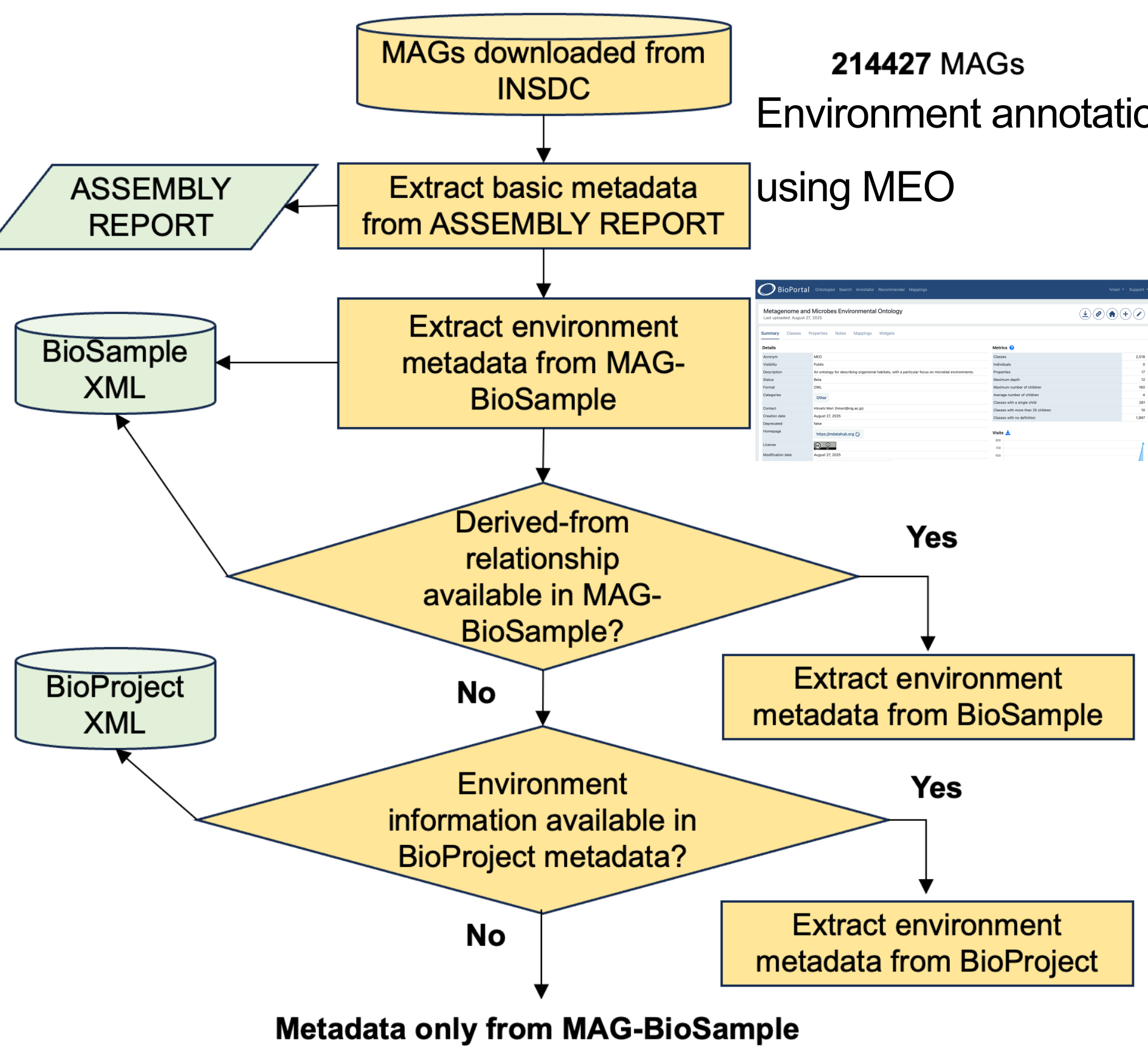
国立研究開発法人 科学技術振興機構 NBDC 統合化推進プログラム



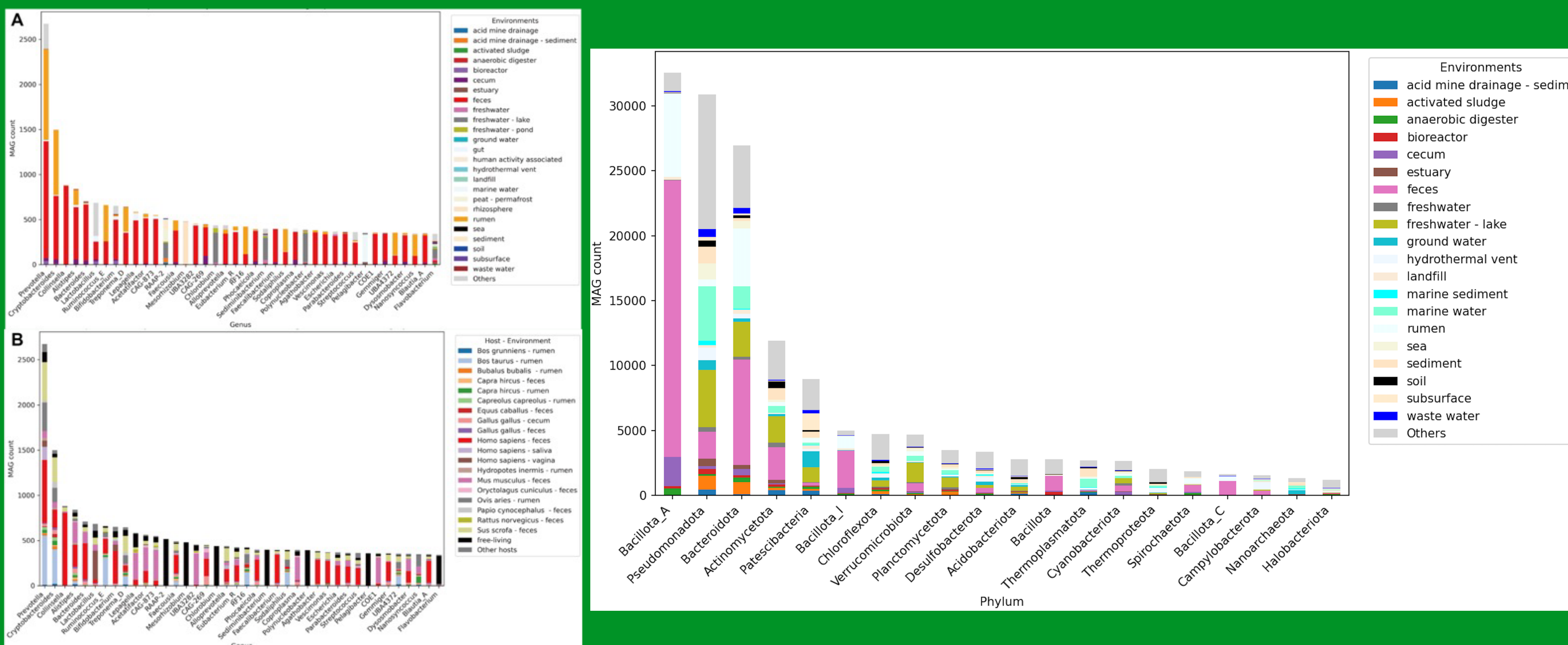
MAG data annotation workflow



MAG metadata annotation

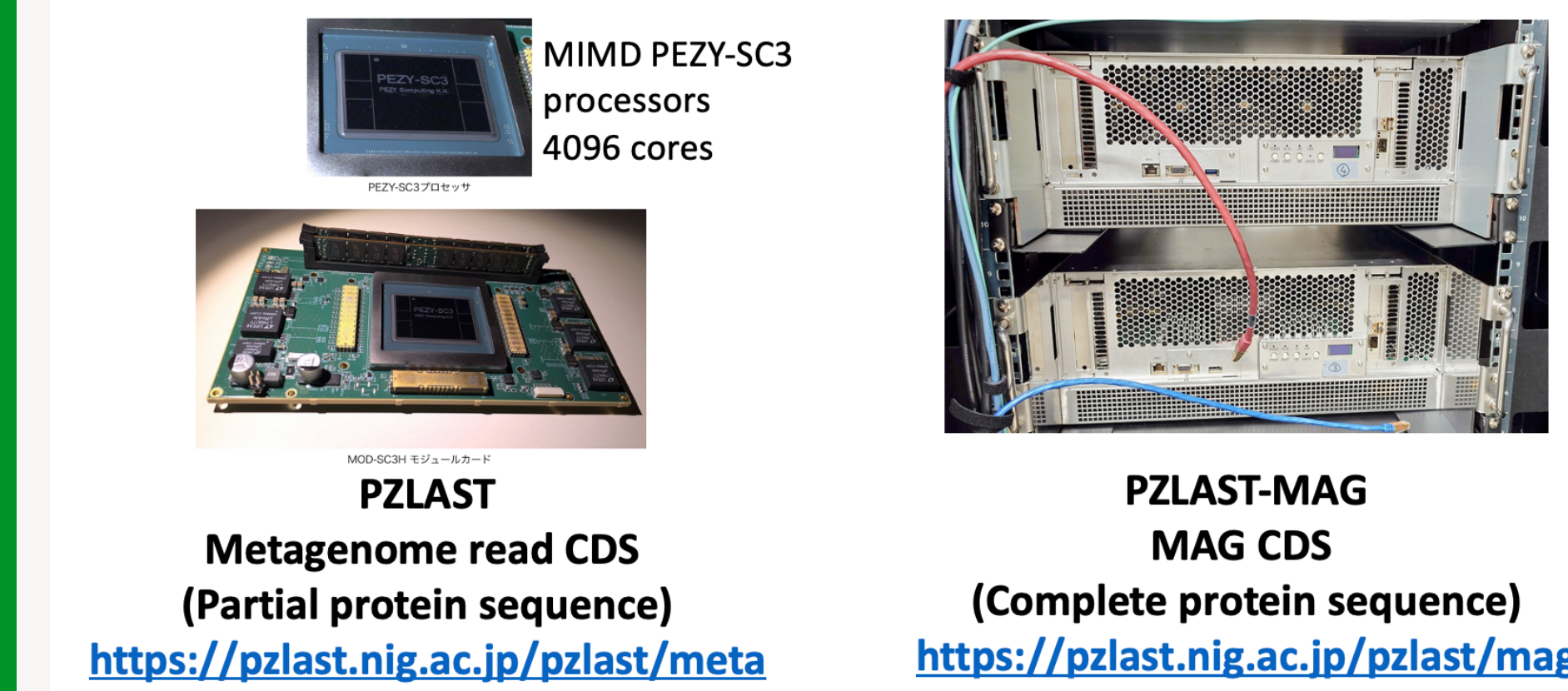


	Microbiome Datahub	MGnify	IMG/M	SPIRE
Developer	NIG	EMBL	JGI	EMBL
Number of MAGs	214,427	518,533	268,973	1,158,553
Number of environments	123	18	178	111
Number of phyla	213	182	197	182
Number of proteins	454,799,231	2,455,939,992	791,399,378	2,452,036,556
Total nucleotides (Gb)	533 Gb	1,291 Gb	784 Gb	2,627 Gb
Number of projects	1,759	5,189	8,929	739
Average completeness (%)	80.59	87.07	76.89	81
Average contamination (%)	1.83	1.03	2.10	1.77
Average MAGs per project	121.9	99.9	30.1	1567.7
Minimum completeness (%)	N.A.	50	50	N.A.
Minimum contamination (%)	N.A.	5	10	N.A.
Last data update date	May, 2023	Nov, 2025	Dec, 2025	Sep, 2023

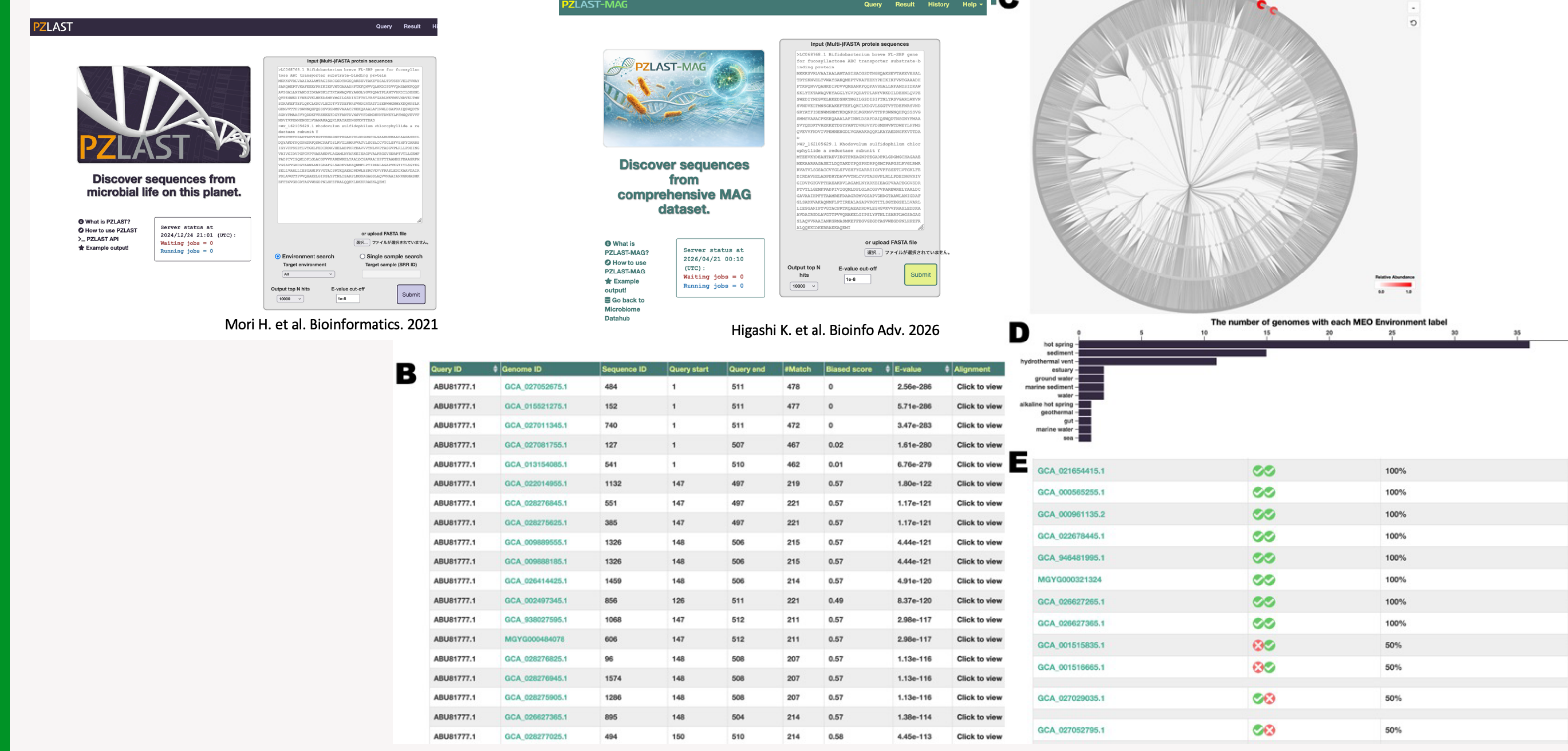


MAG protein sequence similarity search server (PZLAST-MAG)

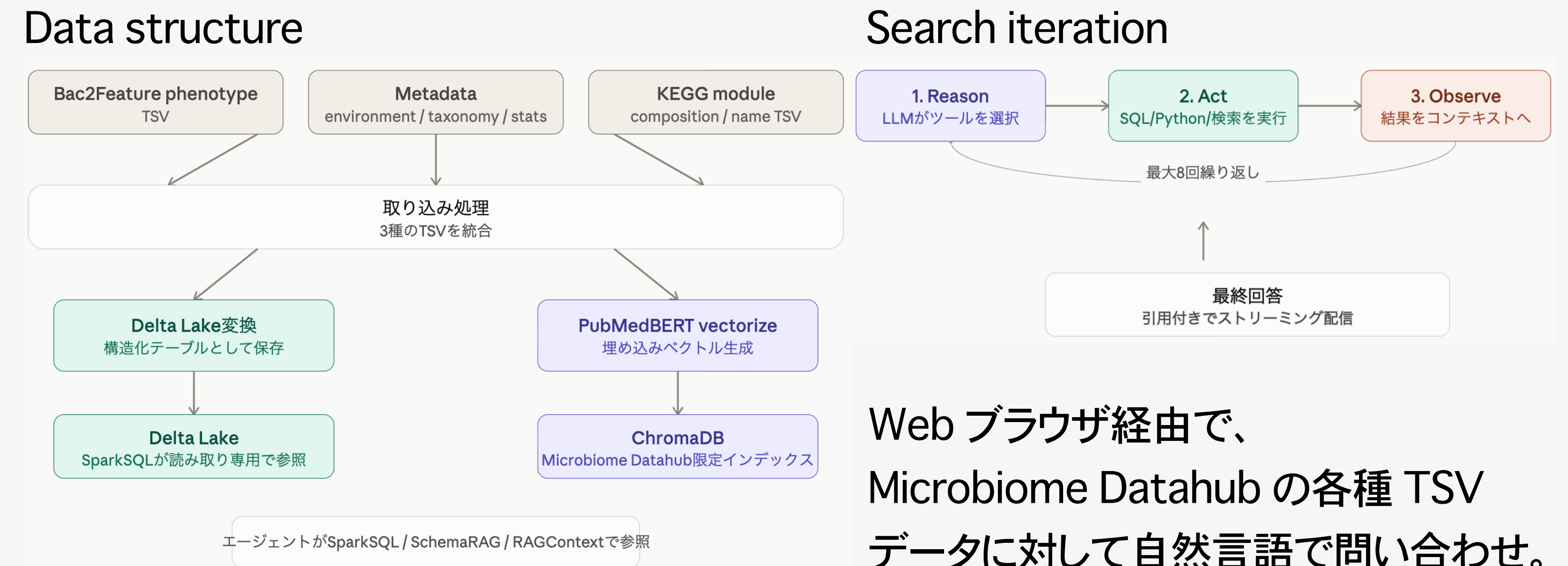
Ultrafast amino acid sequence similarity search web server against metagenome sequences (PZLAST)



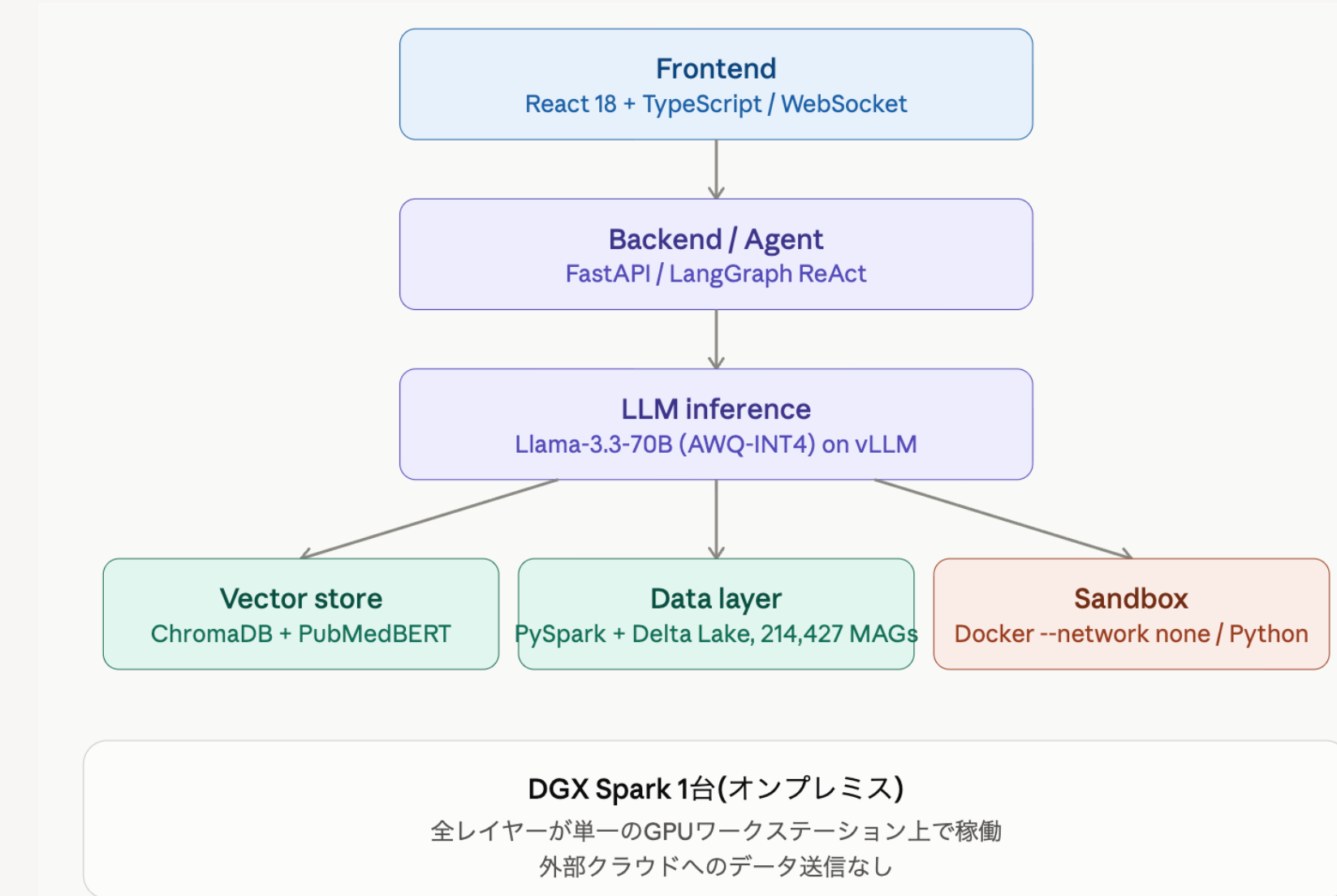
PZLAST-MAG protein vs protein sequence similarity search against 0.45 billion MAG proteins with several minutes



Microbiome Datahub search & quick analysis AI agent



AI agent structure



Web ブラウザ経由で、Microbiome Datahub の各種 TSV データに対して自然言語で問い合わせ。

自然言語での回答、または Python でグラフ描いて結果の視覚化を行える AI agent を開発。現状、DGX Spark で試験的に運用。セキュリティや使い勝手が問題無さそうなら一般利用可能にする予定。

Collaboration of gLM development by ALIS

Microbiome Datahub の 1 種 1 ゲノムの高精度 MAG データ 69586 ゲノムを用いて、一塩基解像度での自己回帰型 DNA 言語モデルを開発。Evo や Evo2 などと比較した性能評価を行い中。

MAG/SAG データの年ごとの増加

Microbiome Datahub data update
Total high quality MAGs/SAGs: 691,210
Bacteria: 657,152
Archaea: 34,014

Unique BioProjects: 4,851
Complete genomes: 5,005

全ての MAG/SAG の塩基配列データとメタデータをダウンロードし、現在遺伝子機能アノテーションの計算を実行中。完了次第 Web サイトをアップデート予定

