U.S. FOOD & DRUG ADMINISTRATION

Human Foods Program



Metagenomically Assembled Genome Generation with Integrated Classification

Kranti Konganti, Padmini Ramachandran | HFP, FDA

Metagenomics at HFP

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

- Progress at HFP - Precision serotyping with isolate independent metagenomic samples.
- Adopting several metagenomic pipelines for WGS and vice versa
- LFFM and ORTS lab have started evaluating some of the metagenomic based serotyping pipelines
- Probiotics/Dietary supplements - Used metagenomics data for regulatory action.

Culture-Based Surveillance Misses the Unseen

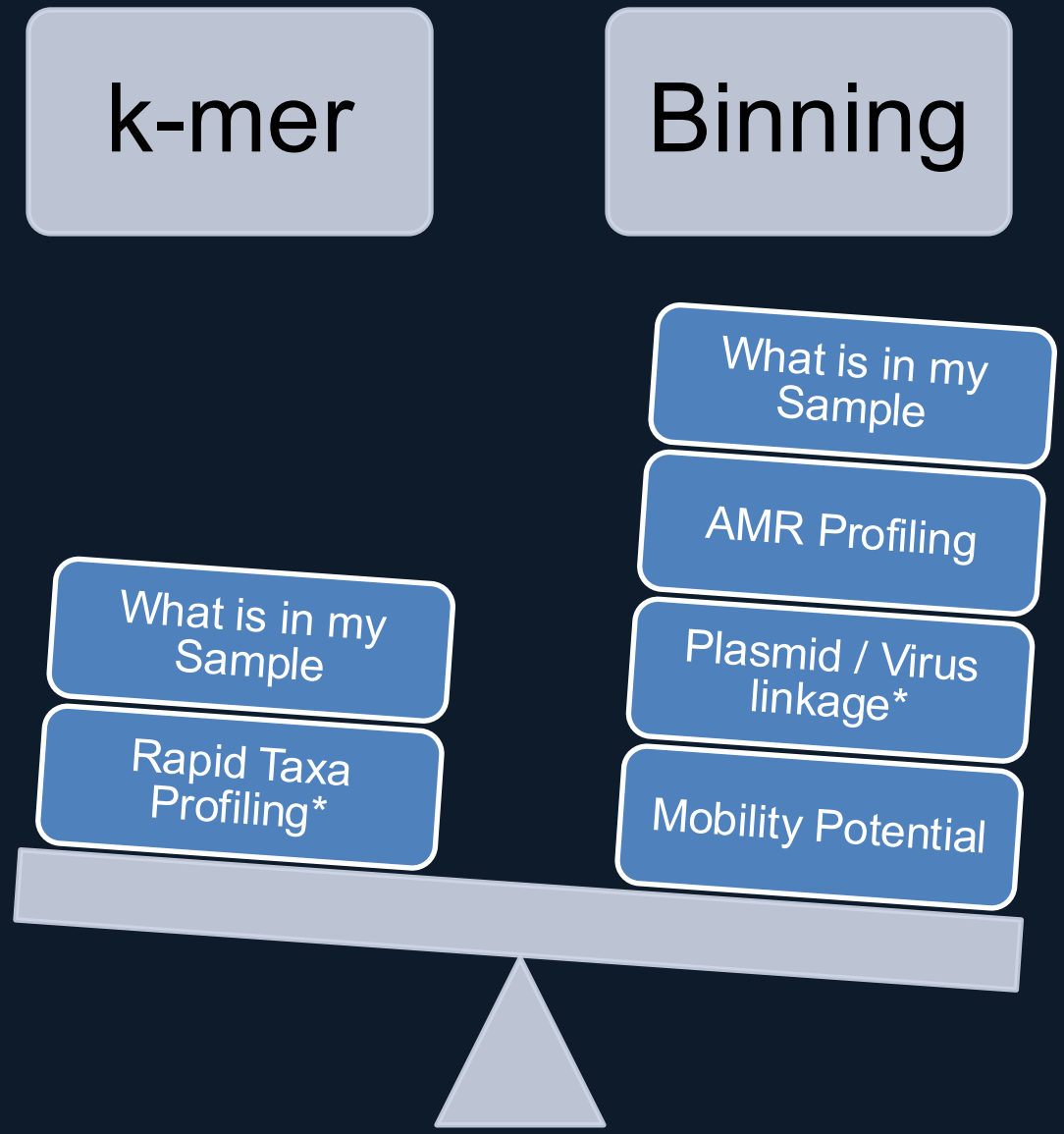
U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

- Metagenomics captures full microbial communities without culture bias
- AMR genes, mobile elements, and unculturable organisms become visible
- But reconstructing genomes from metagenomes is the binning problem

Han et al: Benchmarking metagenomic binning tools ….: Nature Communications | *Yazhini et al: Evaluation of metagenomic binning… : Briefings in Bioinformatics*

To Bin or Not ?

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



Pantiukh et al: Metagenome-assembled genomes ...; mSystems 2026 | Oskolkov: Refining filtering criteria of Kraken family of tools... Frontiers in Microbiology 2026

No Single Binning Tool Recovers All Genomes

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

41%

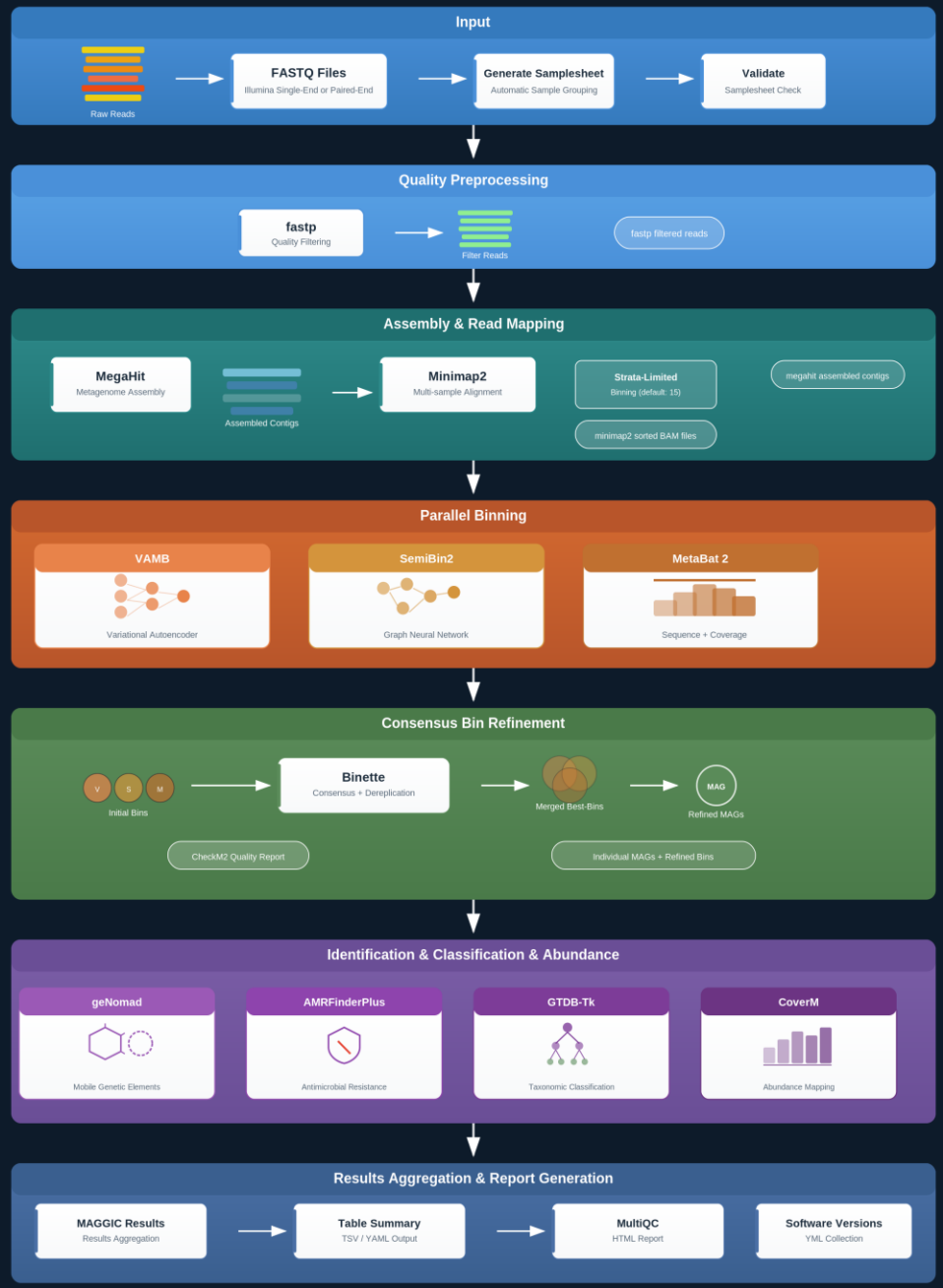
more species recovered with multi-sample binning

- Composition-based vs coverage-based tools complement each other
- Consensus refinement reduces chimerism and contamination
- Multi-sample binning at scale requires alternative strategies

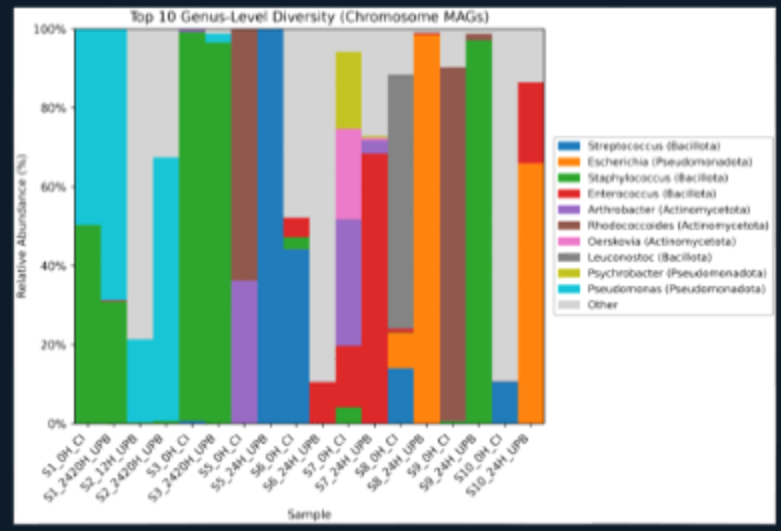
Han et al: Benchmarking metagenomic binning tools …; Nature Communications

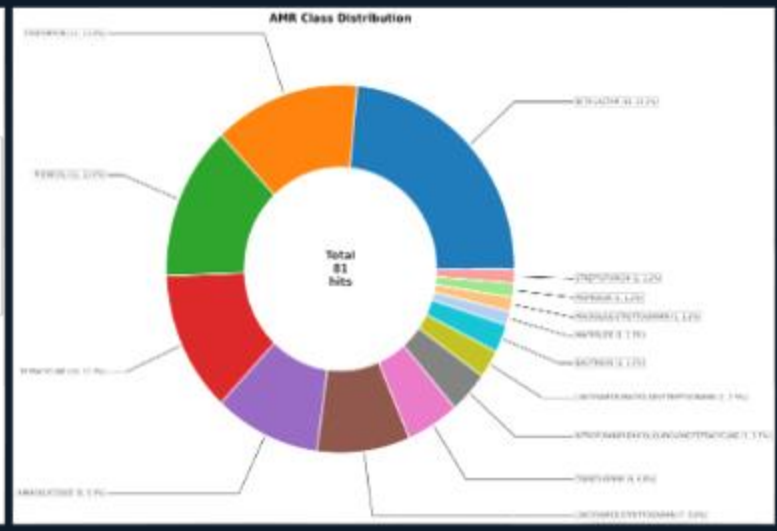
MAGGIC: MAG Generation with Integrated Classification

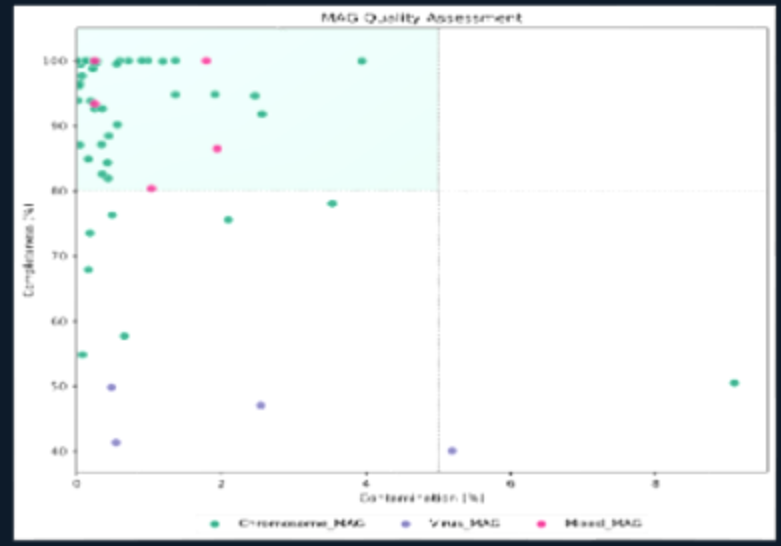
U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

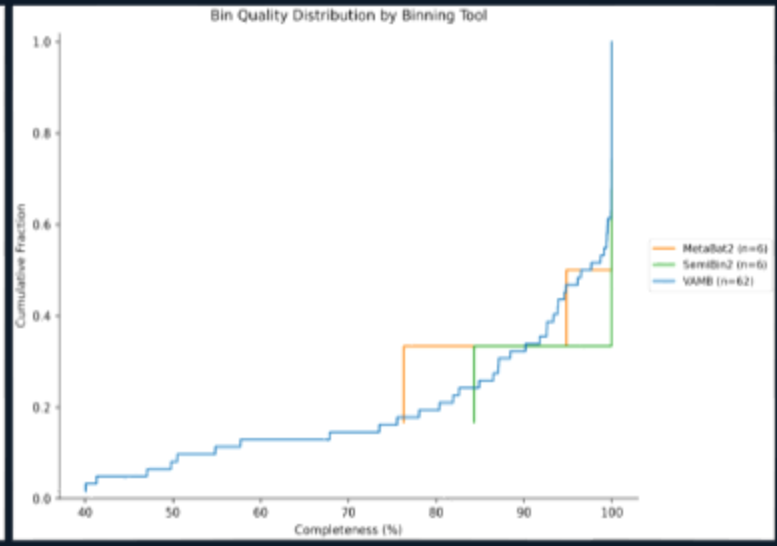


https://github.com/CFSAN-Biostatistics/MAGGIC









Preprocessing and Filtering

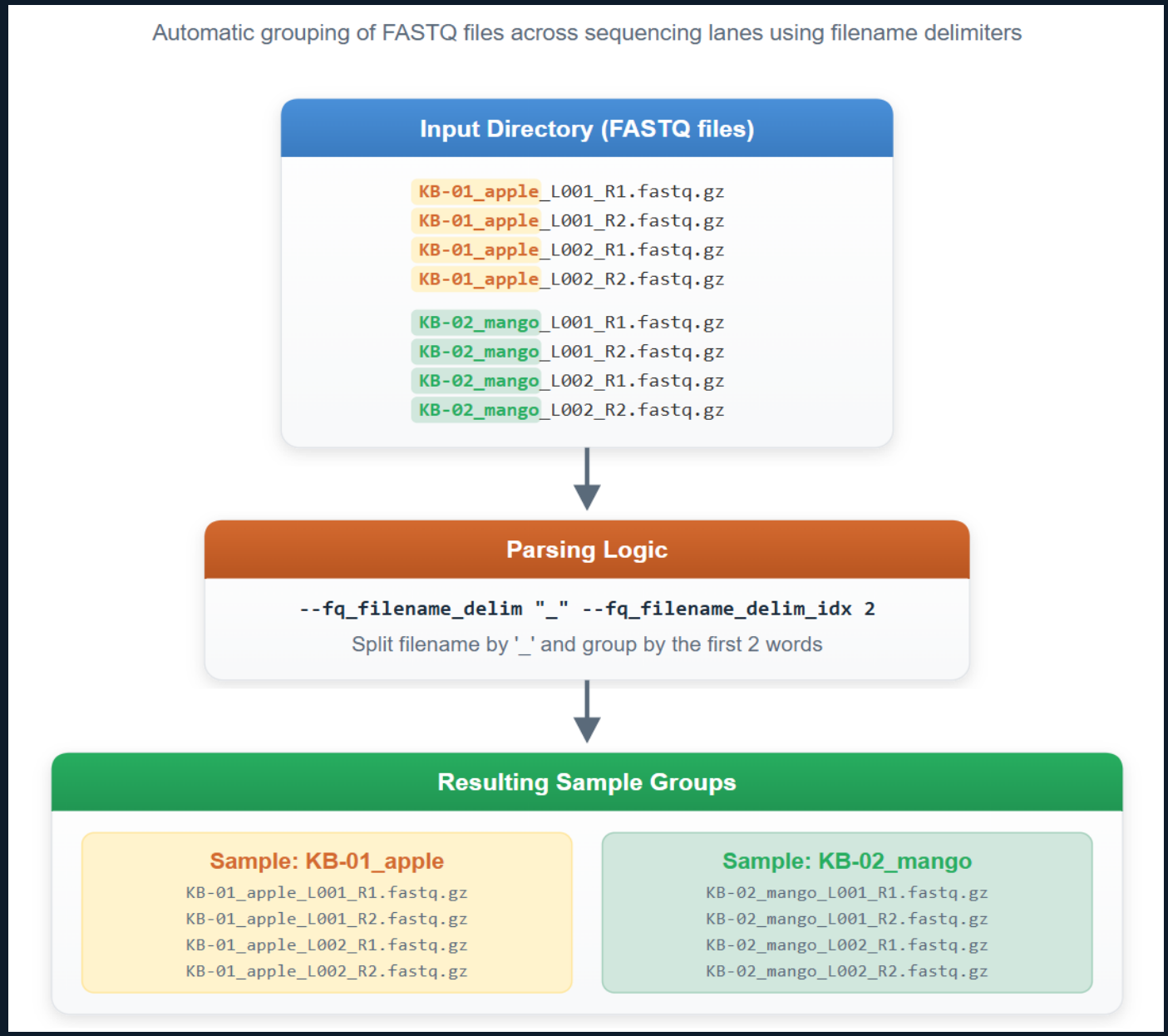
U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



10 stages | Containerized | Nextflow-driven

Preprocessing and Filtering

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



Preprocessing and Filtering

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



10 stages | Containerized | Nextflow-driven

Scalable Multi-Sample Binning

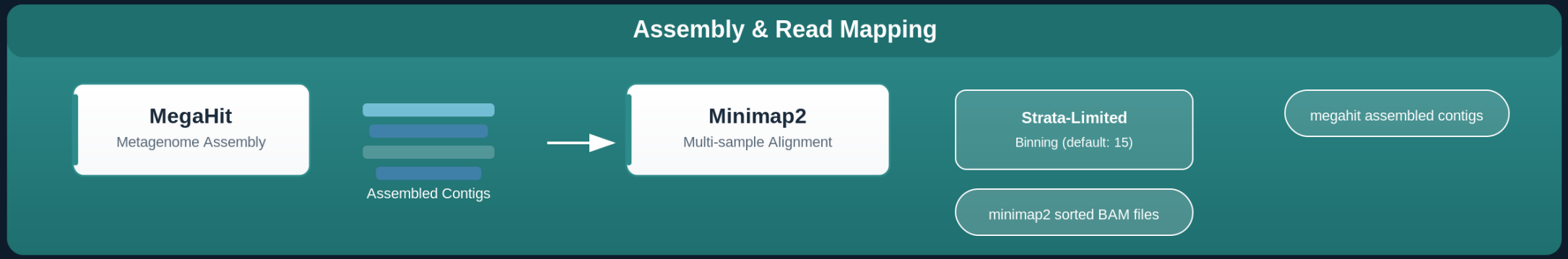
U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

250,000

BAM files for 500 samples (all-vs-all)

7,500 (33x)

BAM files for 500 samples (Strata Control)



10 stages | Containerized | Nextflow-driven

Three complementary binning strategies

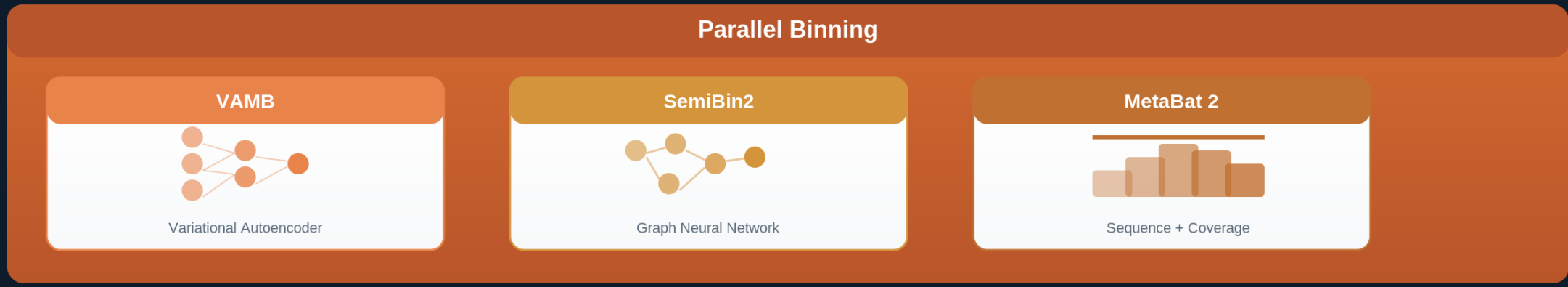
U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

3 binners

Parallel binners + Binette consensus

nf-core/mag

- No integrated AMR detection
- No plasmid or virus classification
- COMEBin consensus takes 7+ days at scale

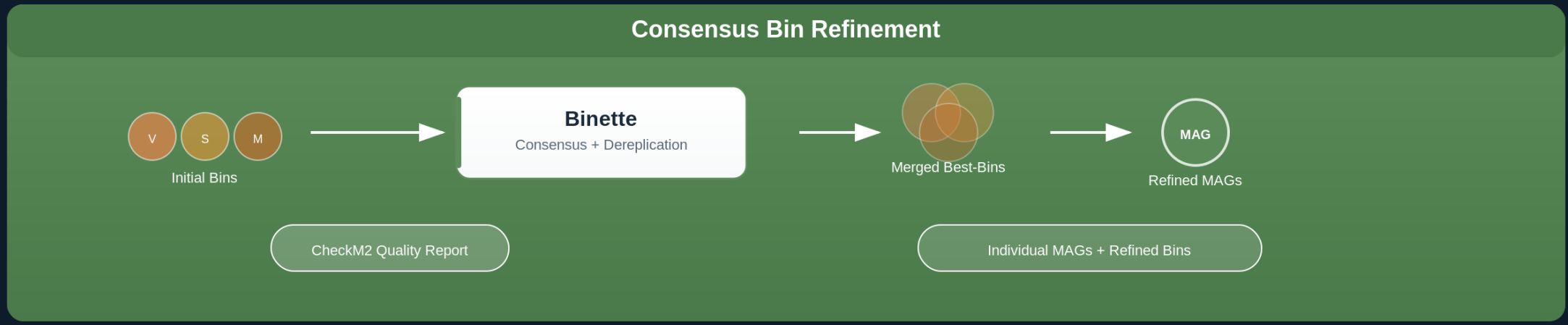


MAGGIC: MAG focused pipeline automated end-to-end for many FDA runs.

Consensus Bin Refinement

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

- Binette merges overlapping bins from all three binners
- CheckM2 estimates completeness and contamination per refined MAG
- Single-copy gene markers distinguish chromosome from plasmid bins*

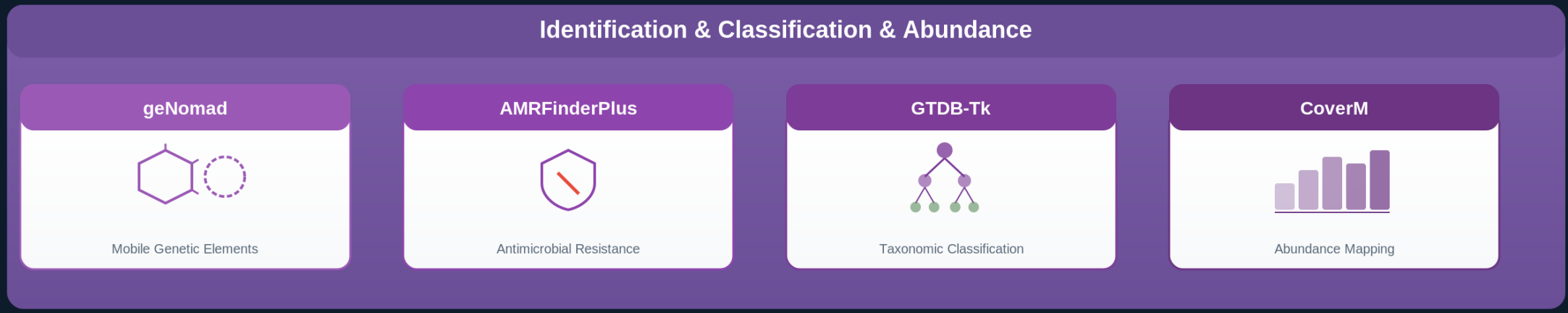


10 stages | Containerized | Nextflow-driven

Taxonomy, MGE detection, and resistance profiling

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

- GTDB-Tk places MAGs on the genome tree of life with ANI to references
- geNomad ensemble scoring for plasmid_score and virus_score per contig
- AMRFinderPlus profiles resistance genes
- CoverM TMM



10 stages | Containerized | Nextflow-driven

MAGGIC Bin Classification

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

Bin Type	Completeness	geNomad Signal	Criteria
Chromosome_MAG	>= 50%	None / provirus only	CheckM2 chromosomal markers present
Mixed_MAG	>= 50%	Plasmid fraction >= 0.2	plasmid_score >= 0.75 in 20%+ of contigs
Plasmid_MAG	< 50%	Plasmid positive	plasmid_summary.tsv has entries
Virus_MAG	< 50%	Virus positive	virus_summary.tsv has entries (checked first)

Level	Completeness	Contamination	ANI	AF
High	>= 90%	< 5%	>= 95%	>= 0.65
Medium	>= 50%	< 10%	>= 80%	>= 0.50
Low	Below Medium thresholds			

MIMAG+

MAGGIC MGE Hard Filters

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

70.8%
plasmid precision

geNomad
per-contig scoring engine, like PlasMAAG

97.3%
virus F1 score

https://github.com/CFSAN-Biostatistics/MAGGIC/wiki

Metric	Formula	High Threshold	Medium Threshold
Length_Weighted_Score	`sum(s_i x l_i) / sum(l_i)`	>= 0.9	>= 0.7
Signal_Uniformity (plasmid)	LW score + min_score	LW>=0.9, min>=0.8	LW>=0.7, min>=0.5
Signal_Uniformity (virus)	Same + provirus penalty	High + no proviruses	High scores with proviruses
Mobility_Potential	Evidence summary	plasmid:High|mob:Relaxed	Pipe-separated components

Camargo et al: geNomad…; Nature Biotechnology | *Lindez et al: PlasMAAG …; Nature Biotechnology* | *Mainguy et al: Binette …; bioRxiv*

maggic-wand: Publication Quality Plots

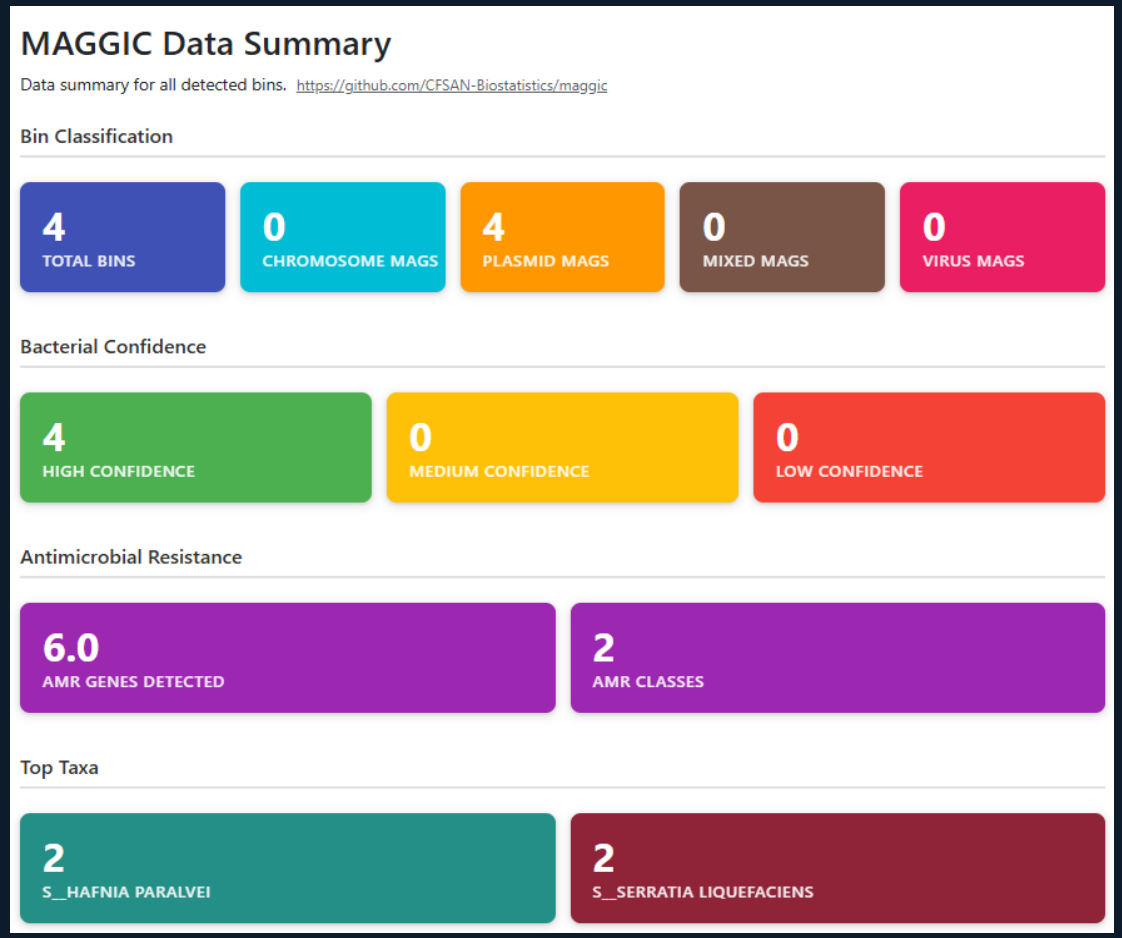
U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

- Input is MAGGIC results folder
- 10 plot types: stacked bars, heatmaps, ECDF, donut, wordcloud, radar, scatter
- Genus and Species level diversity with hierarchical clustering and phylum-colored side bars
- Taxonomy fallback system handles unassigned ranks with named parent lineage labels
- Integrated into the MAGGIC pipeline

https://pypi.org/project/maggic-wand

Beyond chromosomes: plasmids, viruses, and mixed bins

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



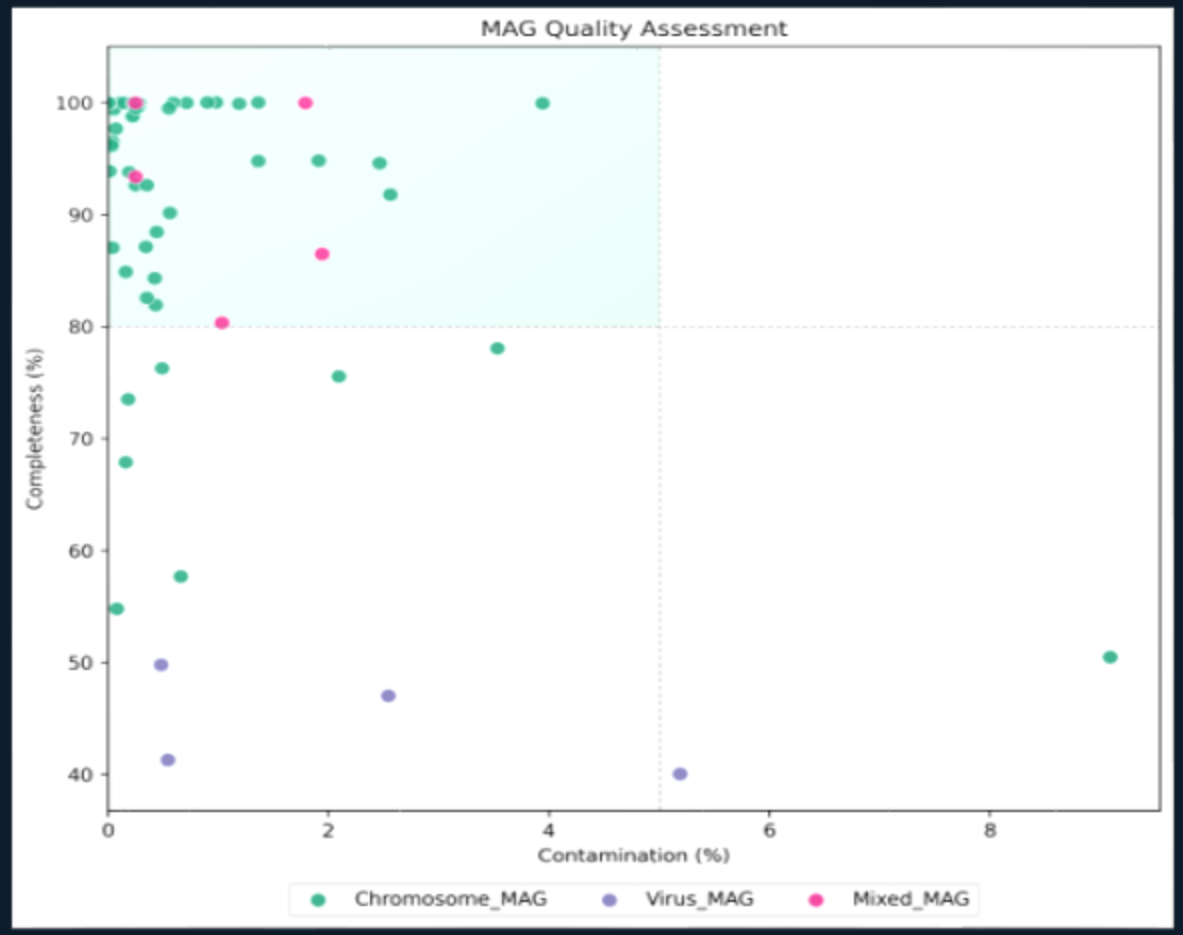
MAGGIC Data Summary

- Chromosome MAG: CheckM2 completeness >= 50%, chromosomal markers
- Plasmid MAG: geNomad plasmid_score + PlasMAAG weighting
- Virus MAG: geNomad virus_score with signal uniformity

Chklovski et al: CheckM2; Nat Methods | *Camargo et al: geNomad…; Nature Biotechnology* | *Lindez et al: PlasMAAG …; Nature Biotechnology*

Beyond chromosomes: plasmids, viruses, and mixed bins

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



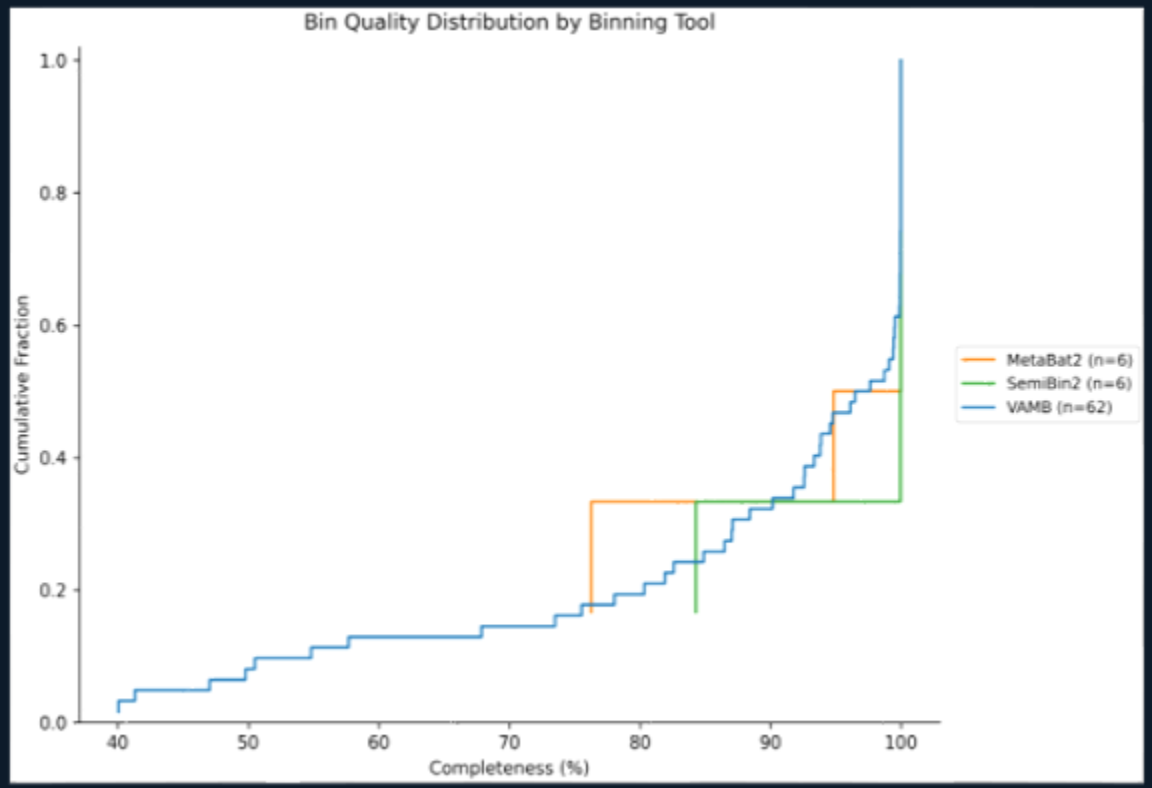
CheckM2 completeness vs Contamination by confidence level

- Chromosome MAG: CheckM2 completeness >= 50%, chromosomal markers
- Plasmid MAG: geNomad plasmid_score + PlasMAAG weighting
- Virus MAG: geNomad virus_score with signal uniformity

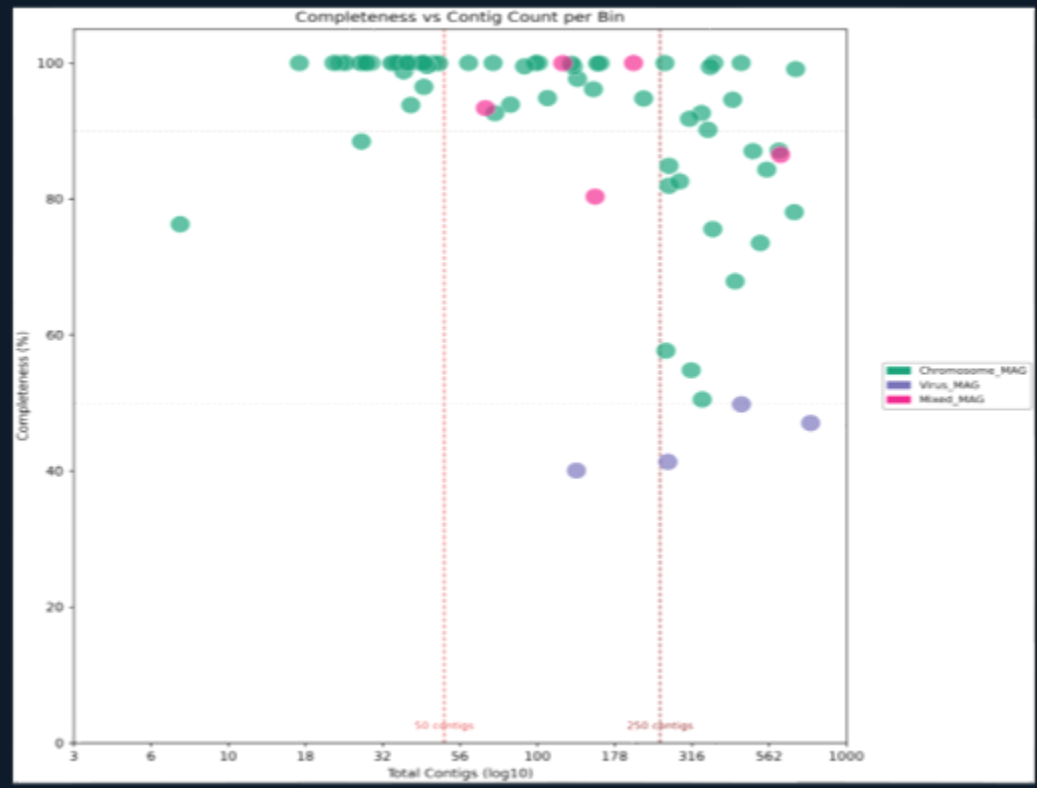
Chklovski et al: CheckM2; Nat Methods | *Camargo et al: geNomad…; Nature Biotechnology* | *Lindez et al: PlasMAAG …; Nature Biotechnology*

Quality metrics across all recovered MAGs

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



Cumulative distribution of quality metrics

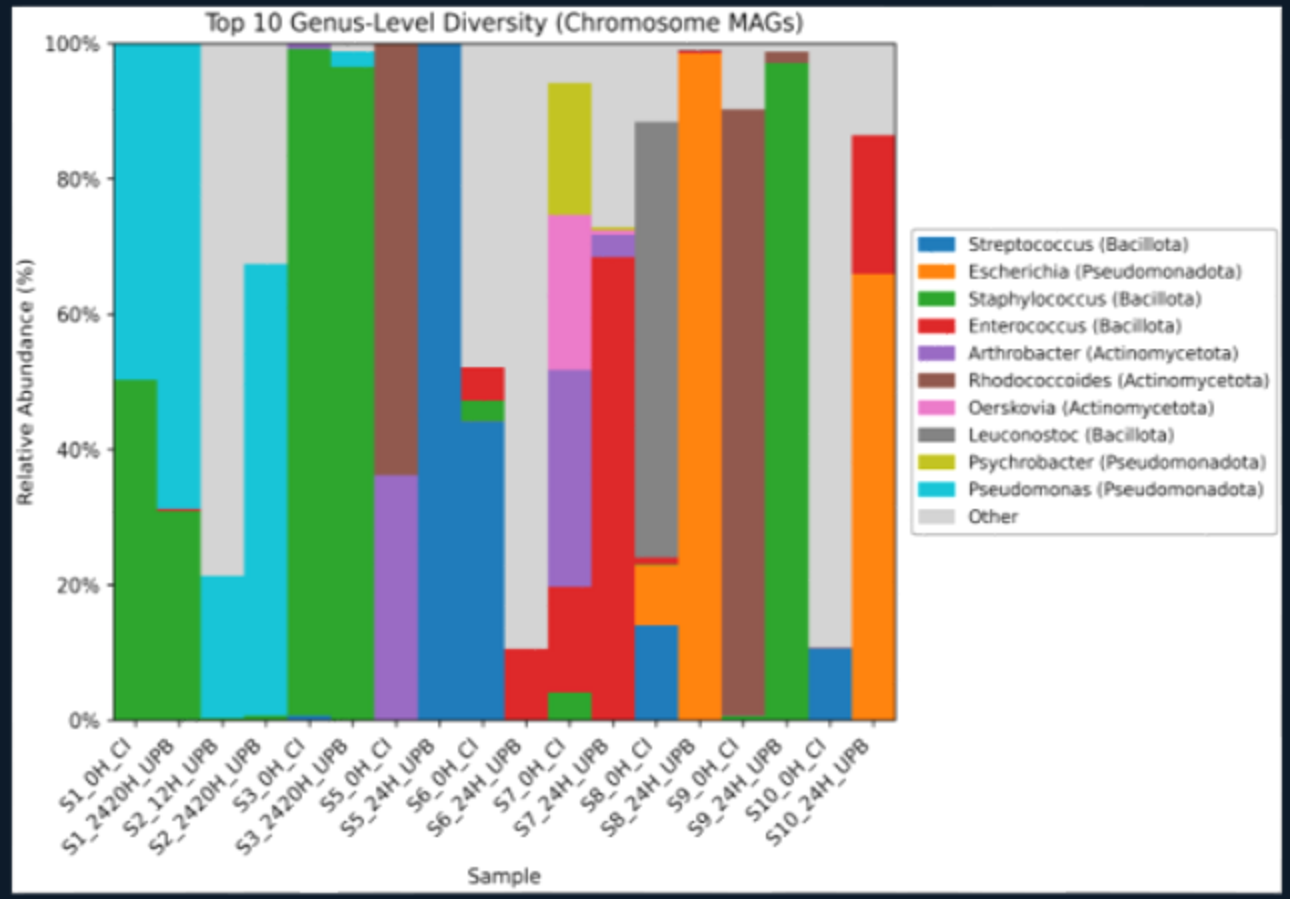


Completeness vs Contig length distribution

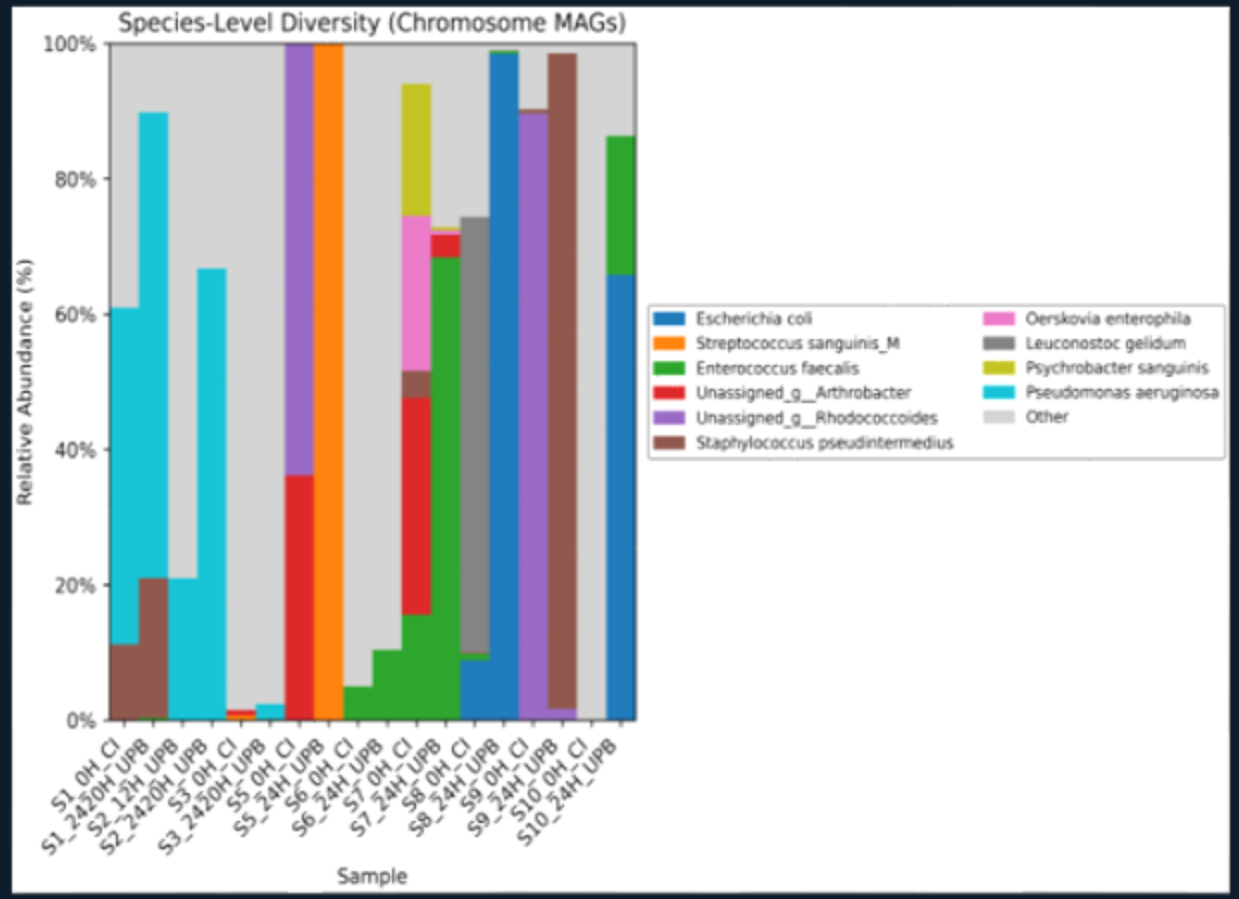
30-column unified TSV | Quality + Taxonomy + MGE + AMR

Genus and Species diversity across samples

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



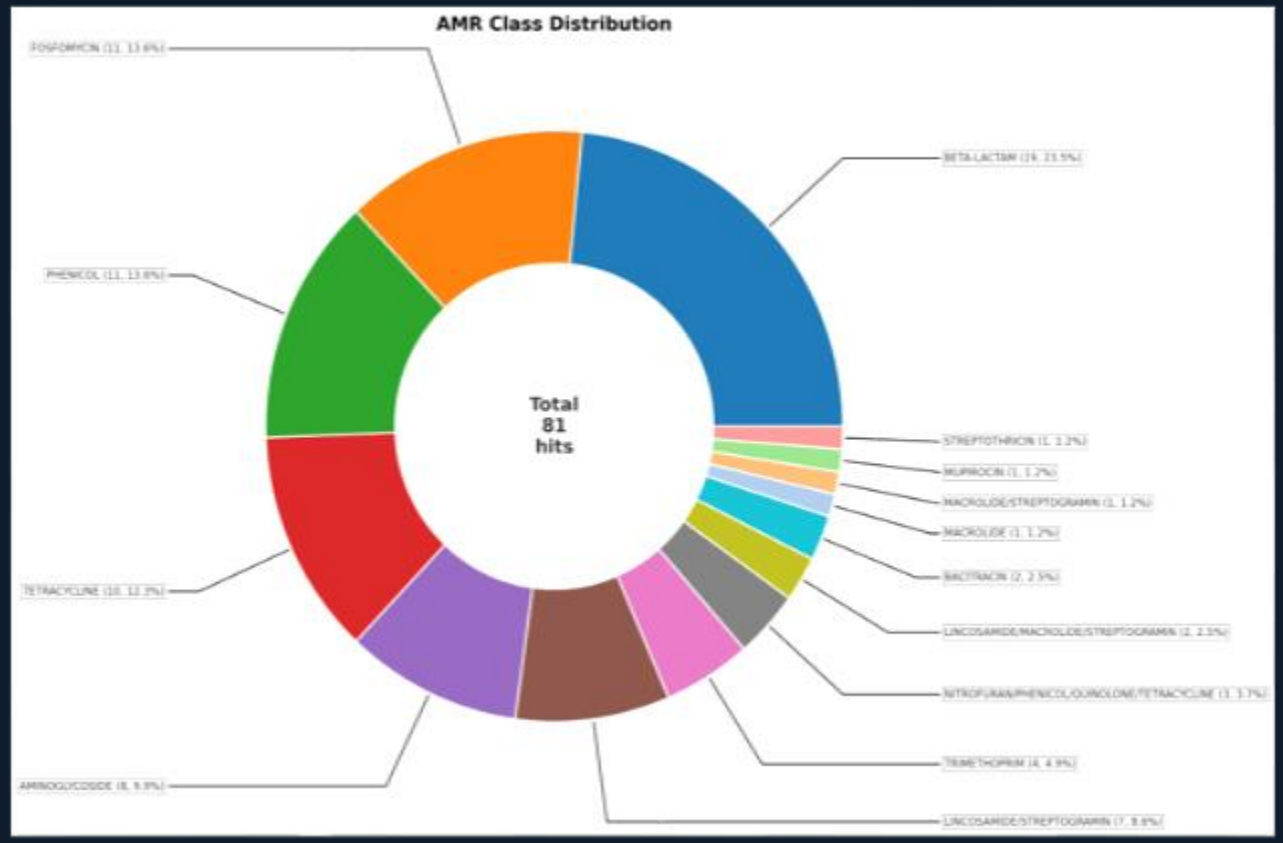
Genus-level TMM-normalized abundance



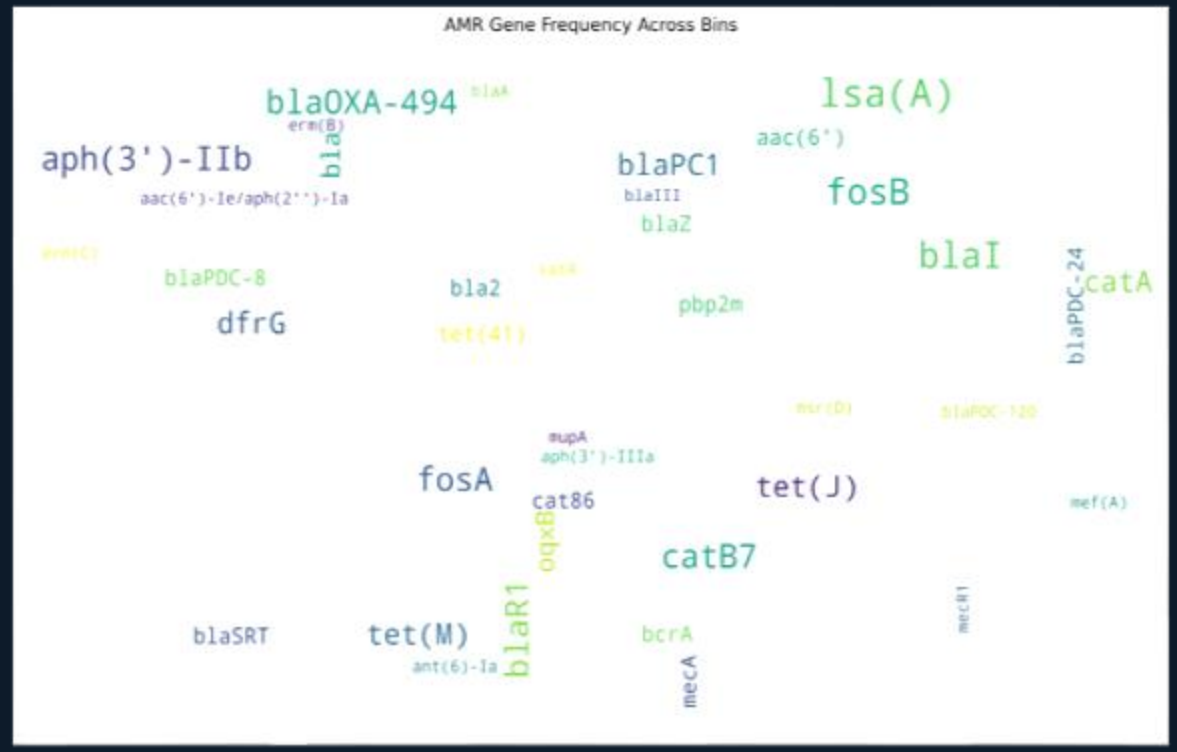
Species-level diversity across samples

AMR Gene Profiling

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program



AMR gene distribution by resistance class



Individual AMR gene frequency

Feldgarden et al: AMRFinderPlus; Nat Sci Reports

Evivo With MCT Oil - Adverse Event Reports

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

EVIVO ACTIVATED B. INFANTIS EVC001 INFANT PROBIOTIC IN MCT OIL

Brand	Number of cases	Microorganisms	Outcomes as self reported into CAERS*
EVIVO WITH MCT OIL (n=5)	2023 – 5 cases	*BIFIDOBACTERIUM INFANTIS* EVC001	• Death – 1 • Sepsis/Bacteremia – 5 • Gastric/Intestinal perforation – 2 • Necrotizing Enterocolitis (NEC) - 0 • Pyocele - 1

Bifidobacterium longum subspecies *infantis*, is one of the first successful colonizers of the infant gut and produces enzymes which break down human milk oligosaccharides, promoting a healthy intestinal microbiome.

B. longum infantis EVC001: isolated, characterized, described, sequenced and marketed by UC Davis/Evolve Biosystems; product contains 8 billion CFU in 0.5 ml vial

Evivo with MCT oil CAERS – Medical record review

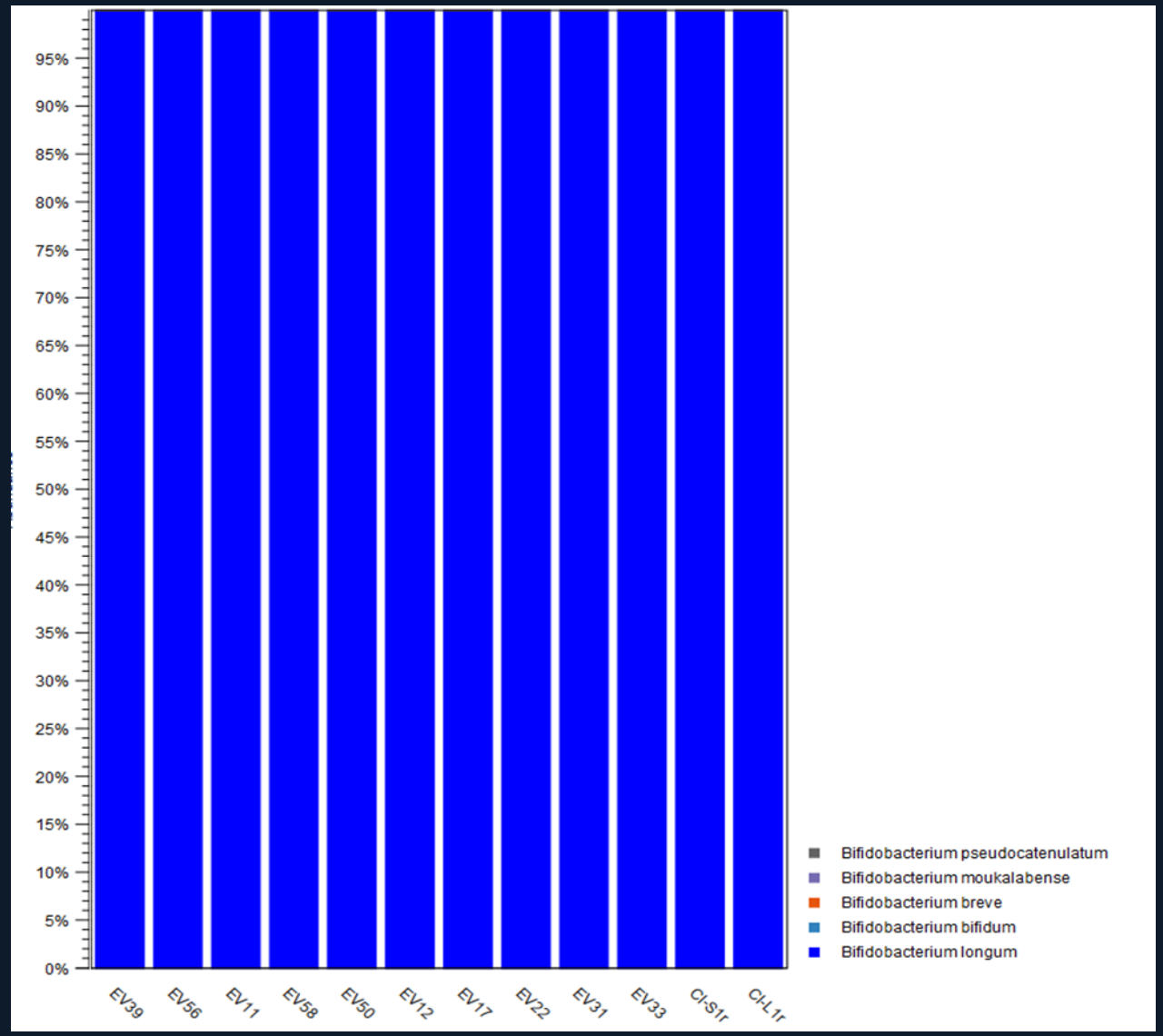
U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

CAERS #	Sex	Gestational age at birth (weeks)	# Evivo doses (prior to Adverse Event)	Notable Adverse Event(s)	Outcome (by end of medical records)	Date of Whole Genome Sequence (WGS) Match	Other
2023-CFS-007355 (case 1)	**Female**	**24 1/7**	**7**	**Sepsis**	**Death**	**9/13/2023**	
2023-CFS-010294 (case 2)	Male	31	3	Bacteremia, gastric perforation, s/p ex-lap, repair and partial omentectomy	Transferred; discharged	10/26/2023	Monochorionic /diamniotic twin B
2023-CFS-010743 (case 3)	Male	28 4/7	2	Sepsis, spontaneous intestinal perforation, s/p ex-lap, small bowel resection	Transferred; recovering inpatient	10/26/2023	dichorionic/ diamniotic twin B (cases 3 & 4 are related)
2023-CFS-010752 (case 4)	Male	28 4/7	3	Sepsis; concern for NEC	Transferred; recovering inpatient	10/26/2023	dichorionic/ diamniotic twin A (cases 3 & 4 are related)
2023-CFS-011201 (case 5)	Male	29 4/7	22 (not confirmed)	Pyocele, *Bifidobacterium* bacteremia	Transferred; discharged	NA	

Taxonomic classification of assembled metagenome: 2023-CFS-007355 (case 1)

U.S. FOOD & DRUG ADMINISTRATION
Human Foods Program

	strain	contig No.	reads	N50	genome size (bp)	coverage X	GC content
clinical	CI-L1r	58	4,904,637	109,535	2,790,074	379	59.8
	CI-S1r	67	6,077,305	90,709	2,791,792	487	59.8
product	EV11	65	91,286,664	93,736	2,791,037	4,504	59.8
	EV12	69	60,404,583	93,541	2,790,431	2,894	59.8
	EV17	70	72,897,275	88,337	2,789,076	3,548	59.8
	EV22	64	80,726,202	88,366	2,790,098	4,007	59.8
	EV31	64	146,304,603	109,403	2,793,338	7,171	59.8
	EV33	68	95,644,861	87,368	2,791,496	4,587	59.8
	EV39	67	97,207,528	87,279	2,790,676	4,725	59.8
	EV50	70	89,969,419	87,373	2,792,290	4,221	59.8
	EV56	68	60,113,341	87,383	2,790,362	2,914	59.8
	EV58	67	73,788,419	88,391	2,791,994	3,744	59.8



product clinical



MAGGIC-EVAL: An AI/LLM assisted evaluation of metagenomics results

MAGGIC-EVAL: LLM ASSISTED RESULTS REVIEW

Co-abundance & correlation

Synthetic example of CLR correlations with shrinkage estimation and false-discovery control.

ILLUSTRATIVE SPECIES × SPECIES MATRIX

	Eco	Kpn	Sau	Efa	Pae	Bsu
E. coli	1.00	+0.70	-0.40	-0.40	-0.45	-0.49
K. pneumoniae	+0.70	1.00	-0.41	-0.34	-0.47	-0.46
S. aureus	-0.40	-0.41	1.00	+0.69	-0.39	-0.36
E. faecalis	-0.40	-0.34	+0.69	1.00	-0.42	-0.37
P. aeruginosa	-0.45	-0.47	-0.39	-0.42	1.00	+0.69
B. subtilis	-0.49	-0.46	-0.36	-0.37	+0.69	1.00

Circled cells: selected pairs passing illustrative BH q < 0.05

STATISTICAL STEPS

01 **CLR transform**
Compositional abundance → log ratios

02 **Ledoit-Wolf shrinkage**
Stabilized covariance / correlation

03 **Benjamini-Hochberg**
FDR correction across pairs

CORRELATION r

−1 0 +1

SYNTHETIC EXAMPLE ONLY • CO-ABUNDANCE DOES NOT PROVE INTERACTION

CLR → LEDOIT-WOLF → BH FDR

Quinn TP et al. propr: An R-package for identifying proportionally abundant features using compositional data analysis. Sci Rep. 2017;7(1):16252. doi:10.1038/s41598-017-16520-0.

Comas-Cufí, M. (2026). coda.base: A basic set of functions for compositional data analysis (R package version 1.0.7). Comprehensive R Archive Network (CRAN). r-project.org

MAGGIC-EVAL: LLM ASSISTED RESULTS REVIEW

Four attention tiers

Species-level findings are assigned a review priority for expert interpretation.

01 **Needs Immediate Attention**
Urgent clinical or safety follow-up

02 **Needs Attention**
Significant finding to investigate

03 **Further Confirmation Needed**
Ambiguous or low-confidence finding

04 **Congruent Analyses Needed**
Cross-check with other diagnostic assays

REVIEW WORKFLOW



Scientists review the findings

MAGGIC-EVAL: LLM ASSISTED RESULTS REVIEW

TASK IS STRUCTURED
species taxonomy + quality metrics + PubMed context => clinical significance

NO OPEN-ENDED REASONING (NO HALLUCINATION)
3-sentence structure (identity, clinical significance, drive-home) with paragraph limit

PUBMED GROUNDING MAKES LARGE PARAM COUNT UNECESSARY
facts come from literature, *not* learned model weights

<30B PARAMETERS SUFFICIENT FOR TEXT SUMMARIZATION
synthesis, scientific reasoning and classification tasks

MAGGIC-EVAL: LLM ASSISTED RESULTS REVIEW

	Runs/Year	Input (M tok)	Output (M tok)	Claude ($5/$25/M)	OpenAI ($3/$15/M)
1 Scientist (5spc/run)	600	4.2	0.6	$3.6	$18
1 Scientist (15spc + text)	600	120	3	$90	$450
10 Scientists (15spc + text)	6,000	1,200	30	$9,300	$4,500
50 Scientists (15spc + text)	30,000	6,000	150	$46,500	$22,500
50 Scientists + monthly re-runs	60,000	12,000	300	$93,000	$45,000

ACKNOWLEDGMENTS

KRANTI KONGANTI

CHRISTOPHER GRIM

ANDREA OTTESEN

BRANDON KOCUREK

SWAYAM ADMALA

JACLYN SMITH

NARJOL GONZALEZ

MARIA HOFFMANN

HFP REEDLING HPC Scientific Engineering Team

https://github.com/CFSAN-Biostatistics/MAGGIC

https://www.abstractsonline.com/pp8/#!/21597/presentation/258

https://aphl.org/events-communities/event-calendar/event-detail/2026-genometrakr-meeting