



Université de
Sherbrooke

ESTIMATING COMMUNITY COMPOSITION OF UNDERSTUDIED MICROBIOMES THROUGH NOVEL SPECIES GENOME ASSEMBLIES

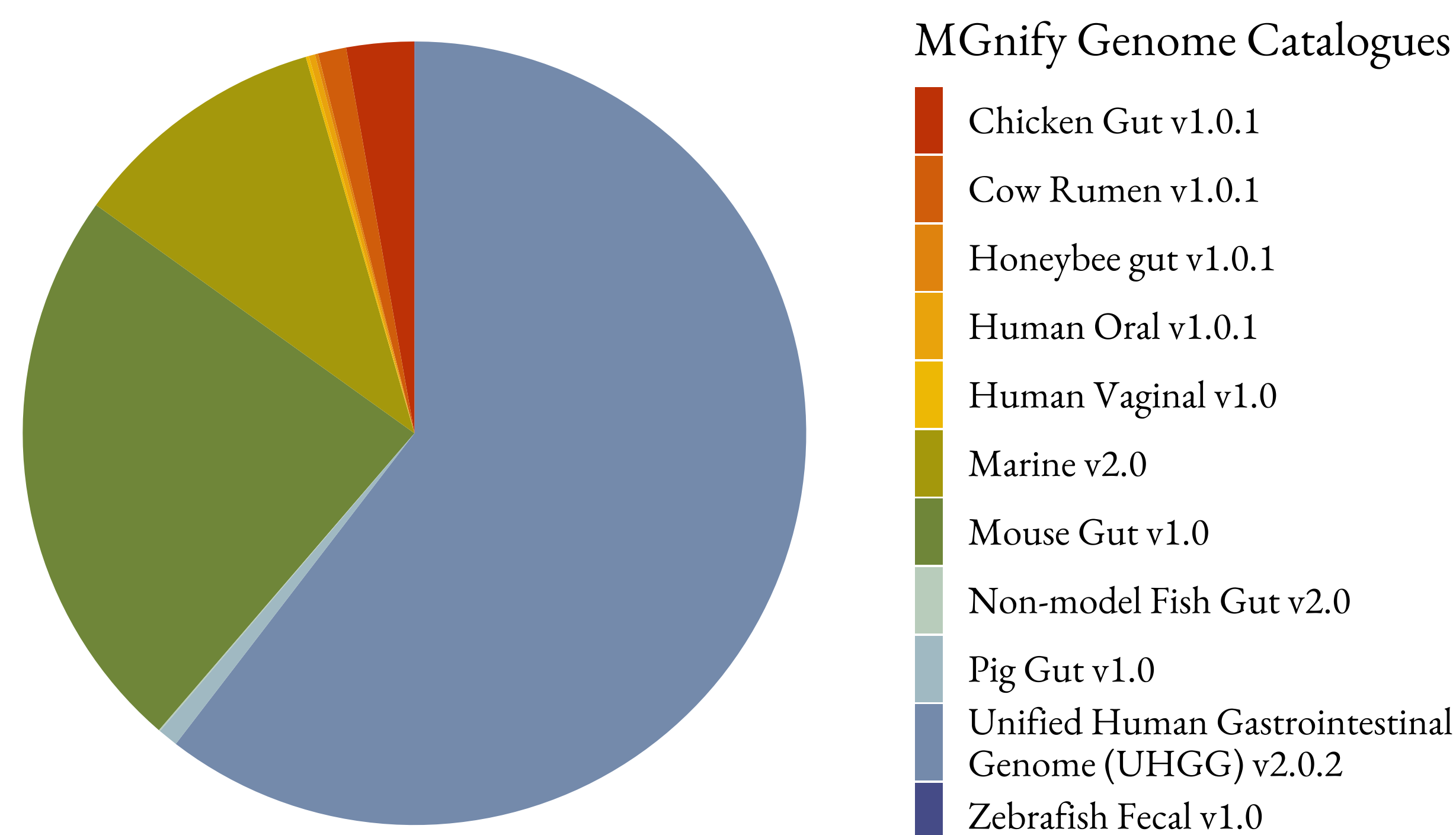
Jonathan Rondeau-Leclaire*, Pierre-Étienne Jacques, Isabelle Laforest-Lapointe.

Département de Biologie, Faculté des Sciences, Université de Sherbrooke.

*Corresponding author : jonathan.rondeau-leclaire@usherbrooke.ca

UNDERSTUDIED MICROBIOMES ABOUND

Most microbiome studies focus on humans, especially the gut. Microbes living elsewhere are underrepresented in genomic databases. **Community composition estimation** is confined to known taxa, undermining the breadth of microbiome research. To better study microbiomes, **unknown species** must be accounted for.



SPECIES LACKING A REFERENCE GENOME

Identified through *de novo* (co)assembly of samples into **metagenome-assembled genomes** (MAGs);

Added to reference databases to better estimate sample composition;

A study-specific effort is necessary for understudied microbiomes.

OBJECTIVES

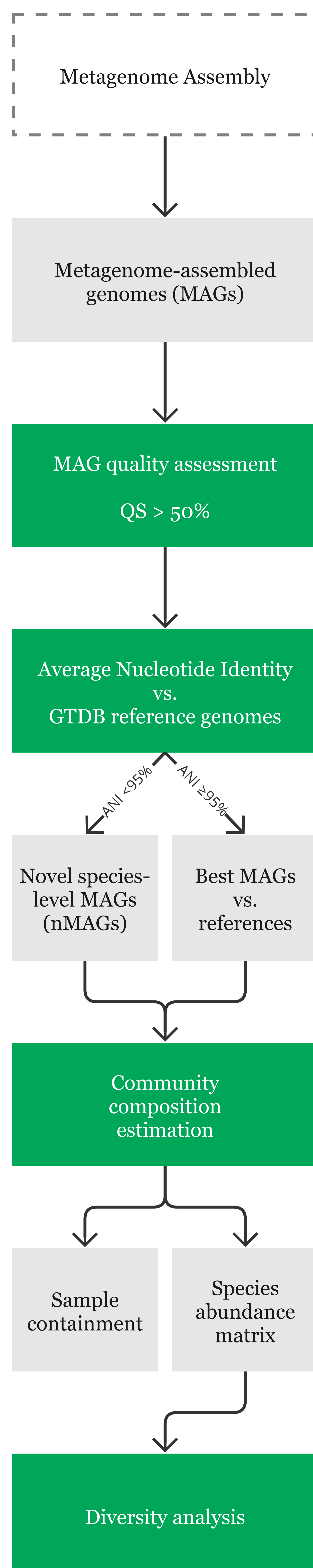
BETTER ESTIMATE COMMUNITY COMPOSITION BY IDENTIFYING NOVEL SPECIES-LEVEL MAGs

1. Identify **novel species** from a set of MAGs;
2. Evaluate the effect of nMAGs addition to genome database on increasing the **containment** therein of the samples from which the nMAGs were assembled;
3. Explore the effects of better **community composition estimation** on a simple comparative diversity analysis.

METHODS

1. MAG quality score (QS; Completeness – 5 × Contamination) > 0.50
2. Novel species-level MAGs (**nMAGs**) : Average Nucleotide Identity (ANI) < 95 % with every known representative genome (GTDB r214)
3. MAGs with >= 95 % ANI are substituted if better QS than reference
4. Estimate sample containment* and community composition with default and improved reference database (Sourmash gather)
5. Explore effects of improvement on sample diversity (Shannon, rarefaction)

* Jaccard containment : abundance-weighted intersection of sample and reference database k-mers.



NOVEL AND BETTER MAGs IDENTIFIED FROM TWO BIOMES

	Human saliva 26 samples (13 persons)	Boreal mosses 131 samples (4 species)
Species representative MAGs	130	153
Good quality MAGs (QS > 0.50)	110	107
nMAGs (ANI < 95%)	12	102
non-novel species (ANI ≥ 95%)	98	5
of which QS > reference QS	20	1

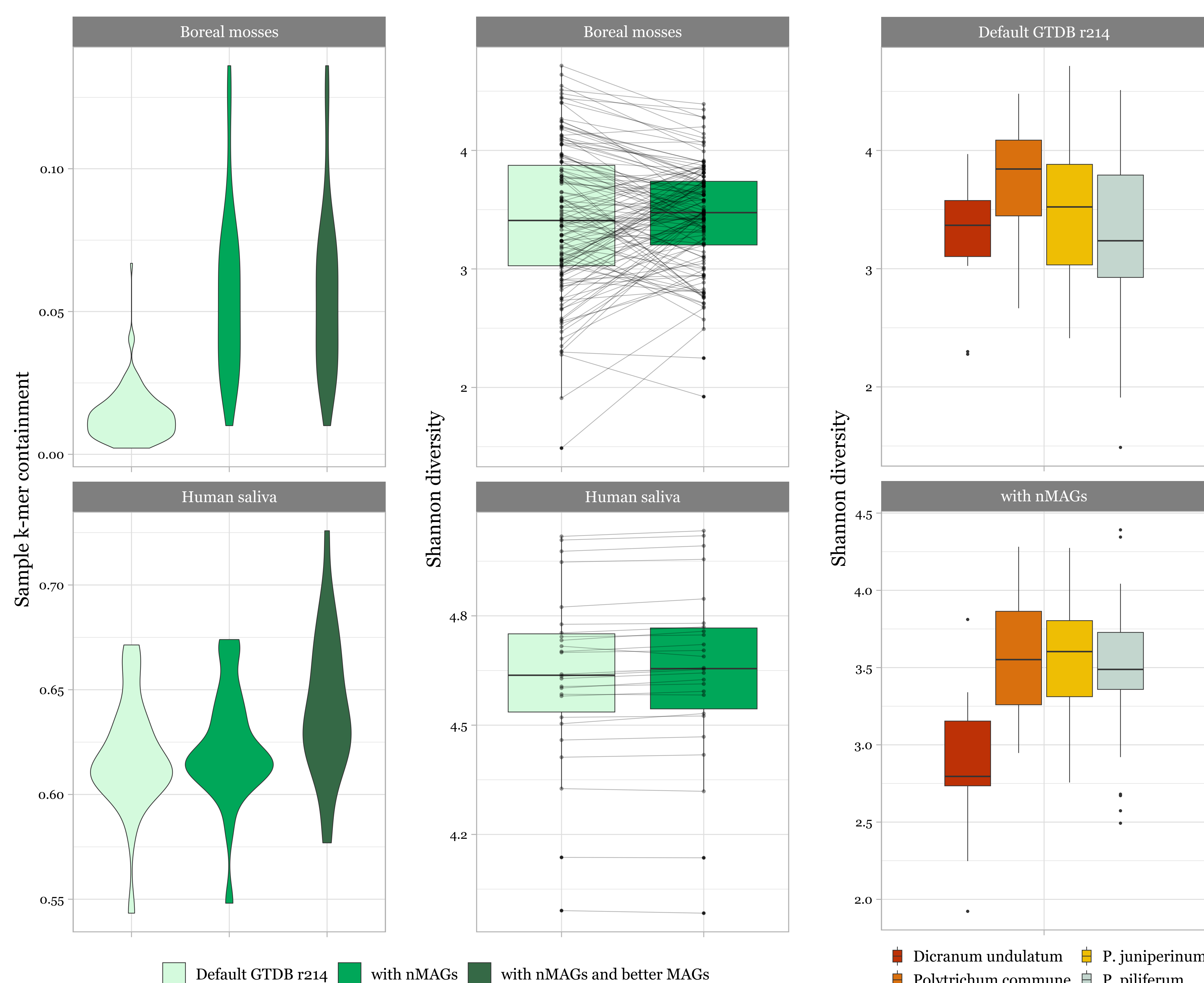
NOVEL AND BETTER MAGs IMPROVE SAMPLE CONTAINMENT

nMAGs increase sample containment only in Moss dataset; better MAGs increase only in saliva.

Human saliva: 1.04-fold mean increase (Cohen's paired $d = 1.03$)

Boreal mosses: 4.84-fold mean increase (2.1 to 9.1-fold for 90% samples*; Cohen's paired $d = 1.72$)

* 95% confidence, 1000 bootstraps



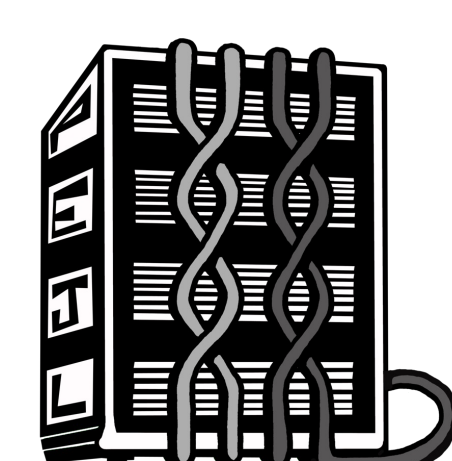
NOVEL MAGs DIFFERENTIALLY AFFECT DIVERSITY ESTIMATIONS

Mean diversity doesn't change, but **variance halves** (from 0.37 to 0.20 when adding nMAGs)

Sample-wise diversities shift **without pattern** (mean change 2.6 % ± 17.7 %)

Variance explained by linear regression on host species **triples** (R^2 : 0.04, $p = 0.04$ to R^2 : 0.18, $p < 10^{-5}$)

Including nMAGs **reduces noise** and could provides a better estimate of diversity.



We acknowledge
the support of :



REFERENCES

1. SEEK **INFORMATION** on the protocols in place in your faculty for presenting references.
2. **RESPECT** the protocol the event has requested.
3. **BE** consistent in your presentation of the various references.

