

METHODOLOGICAL CHOICES INFLUENCE COMMUNITY ANALYSES IN TAXONOMY-ORIENTED SHOTGUN METAGENOMICS



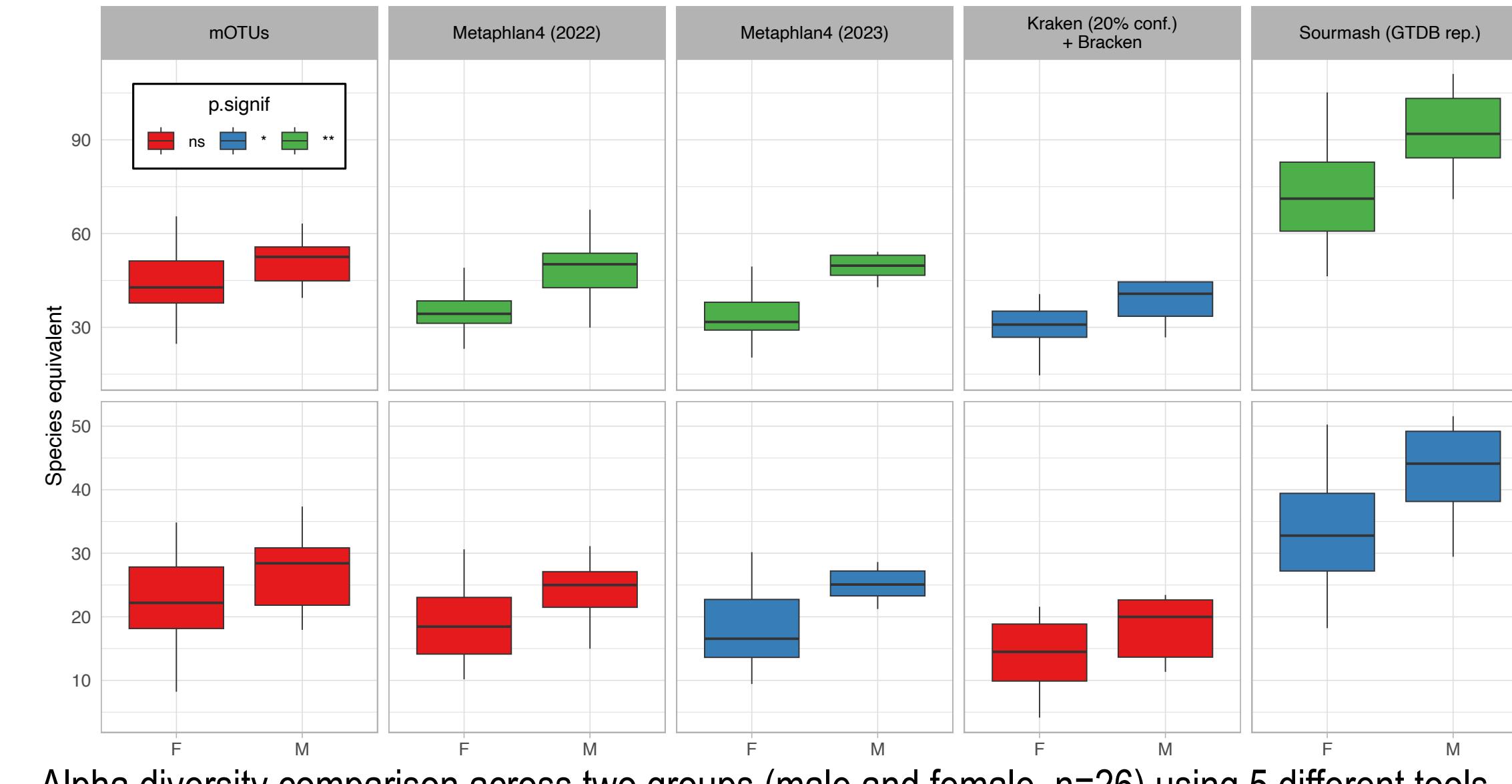
Jonathan Rondeau-Leclaire*, Pierre-Étienne Jacques and Isabelle Laforest-Lapointe

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

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

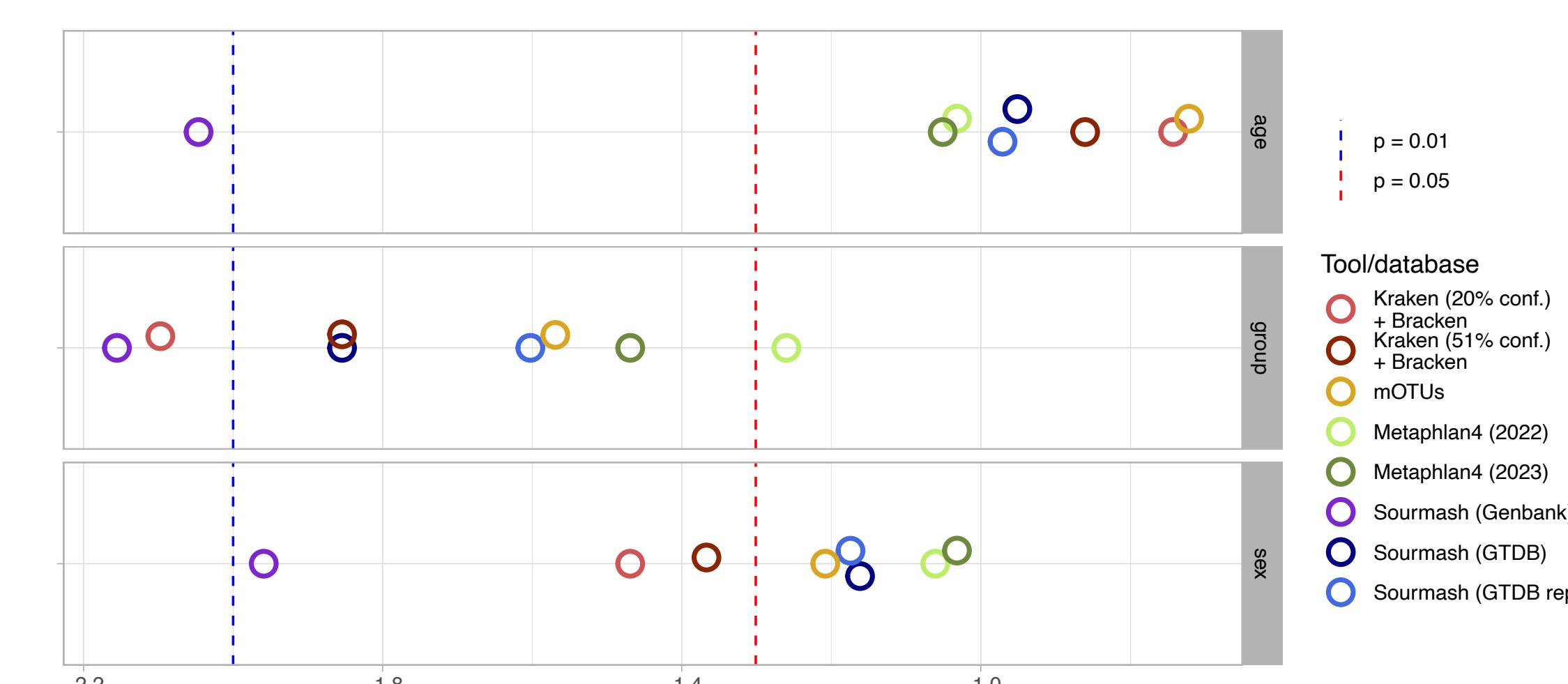
COMPOSITION ESTIMATION AFFECTS DIVERSITIES

Diversity and dissimilarity analysis results vary depending on choice of community composition estimation tools.



Alpha diversity comparison across two groups (male and female, n=26) using 5 different tools.

H_0 : no difference between groups of samples.



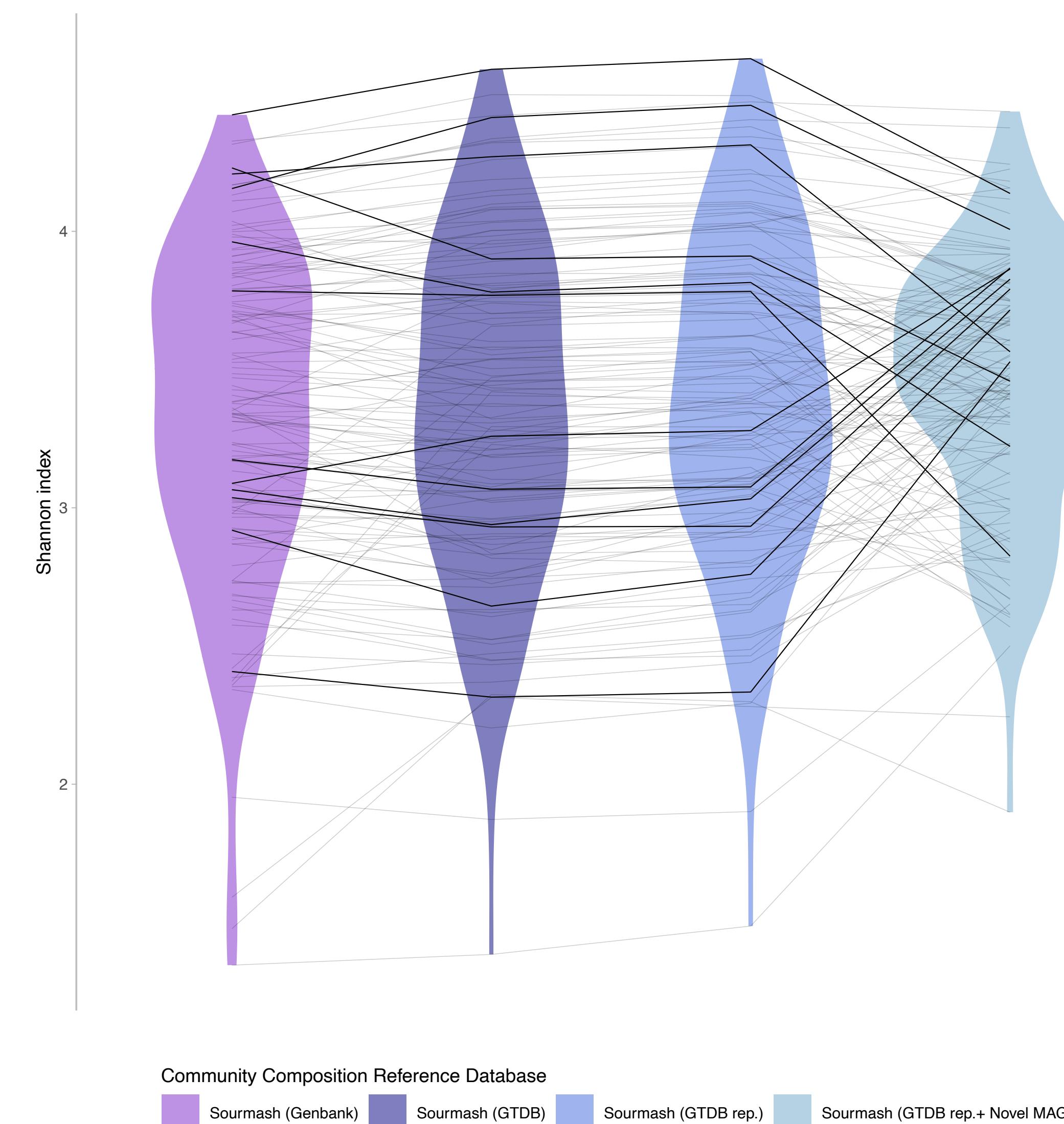
permANOVA assessing the significance of covariates in explaining community dissimilarities yields different interpretations depending on community composition estimation tool.

H_0 : no effect of a covariate on multivariate response (the dissimilarity matrix).

ADDING NOVEL MAGs TO REFERENCE DATABASE AFFECTS SAMPLE DIVERSITY IN UNDERSTUDIED MICROBIOME (BOREAL MOSES)

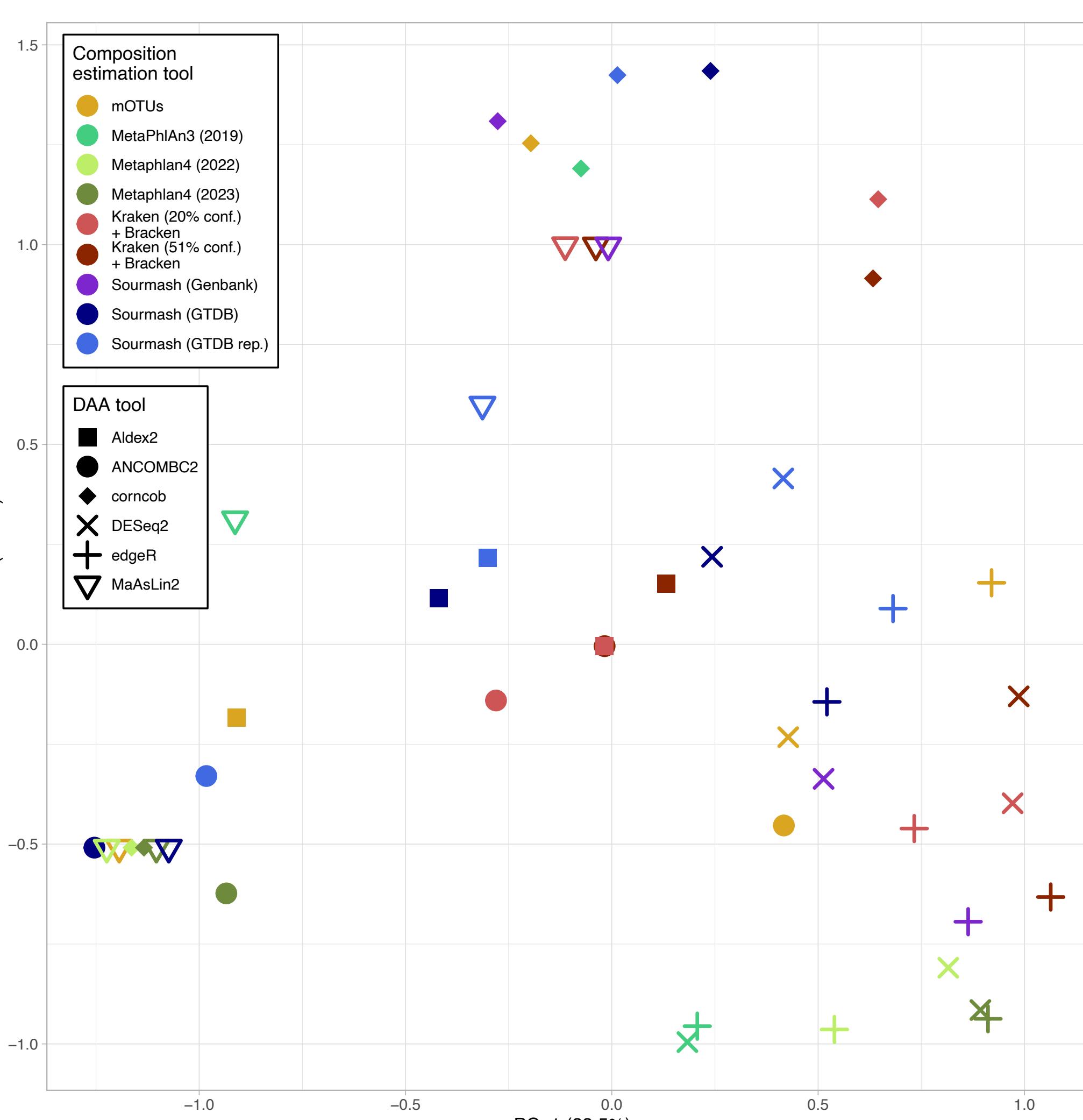
4.8-fold mean sample containment increase in reference database.

Diversity variance halves (mean increases 2.7 % \pm 17 %)

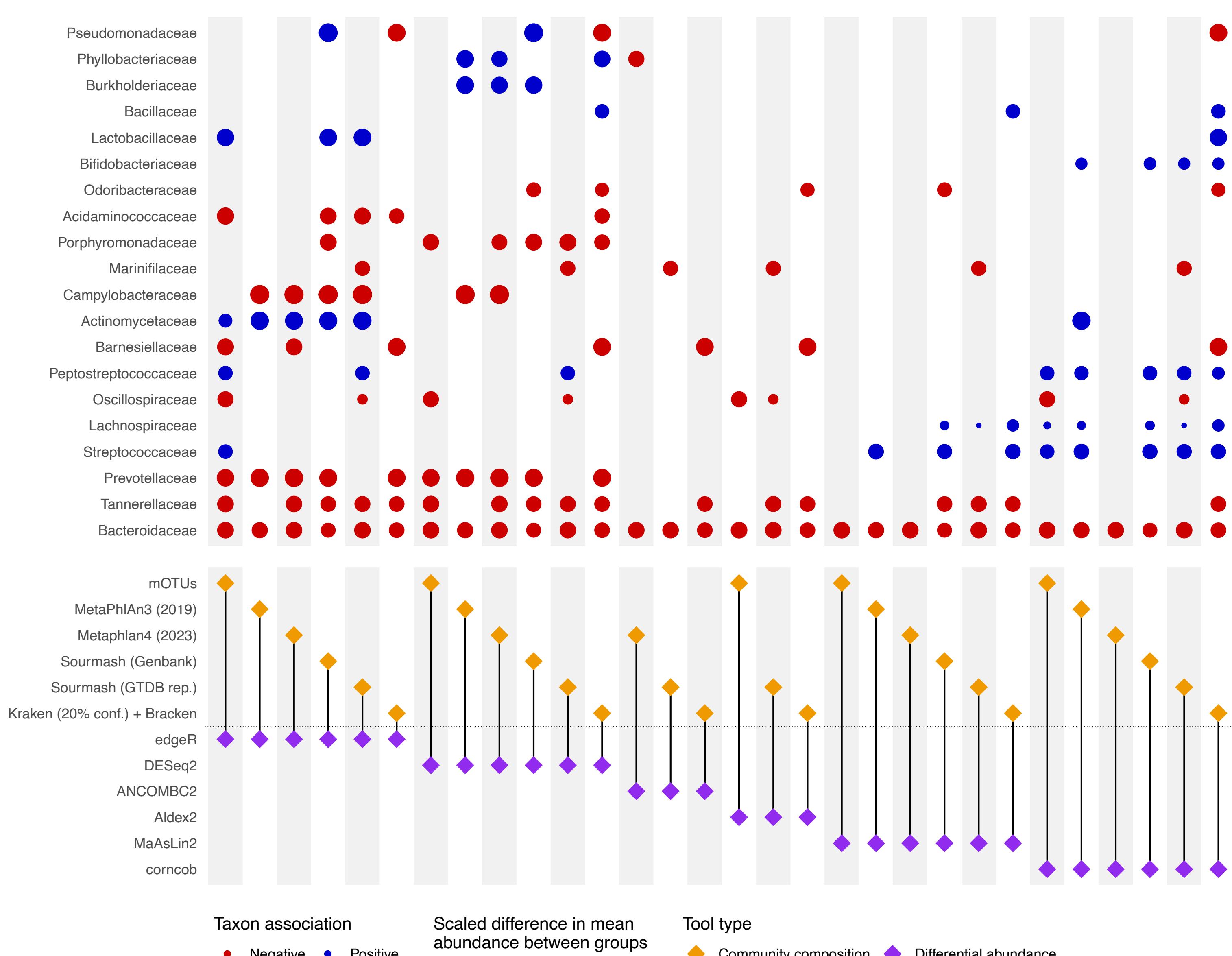


DIFFERENTIAL ABUNDANCE ANALYSIS (DAA) IS HIGHLY SENSITIVE TO TOOL COMBINATION

Choice of community composition estimation approach and DAA methods both influence taxa set identified as significantly differentially abundant. DNA-to-DNA and DNA-to-Marker profiling tools tend to cluster together, but some DAA tools cluster regardless of the profiling approach.



Principal coordinates analysis on Dice-Sørensen dissimilarities of families identified as differentially abundant (1) or not (0) for 54 combinations of community composition estimation tool (or database) and differential abundance analysis approach. permANOVA on distance matrix controlling for database taxonomy (NCBI vs. GTDB) and whether or not DAA tool models taxon-specific biases shows community composition estimation approach (DNA-to-DNA vs. DNA-to-Markers) to explain 12% of overall variance ($p<0.001$).

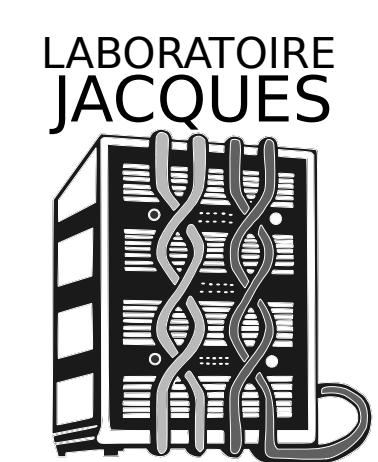
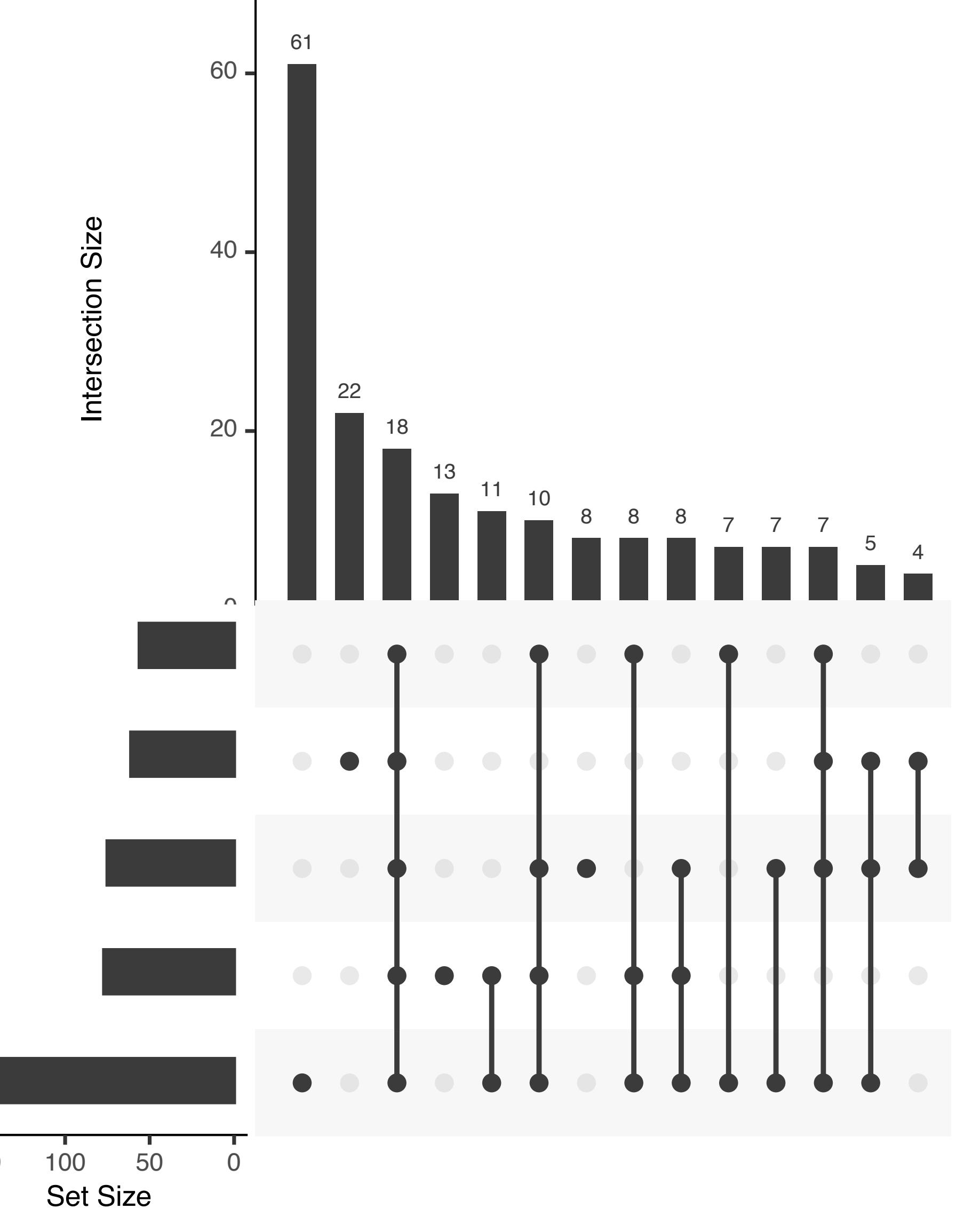
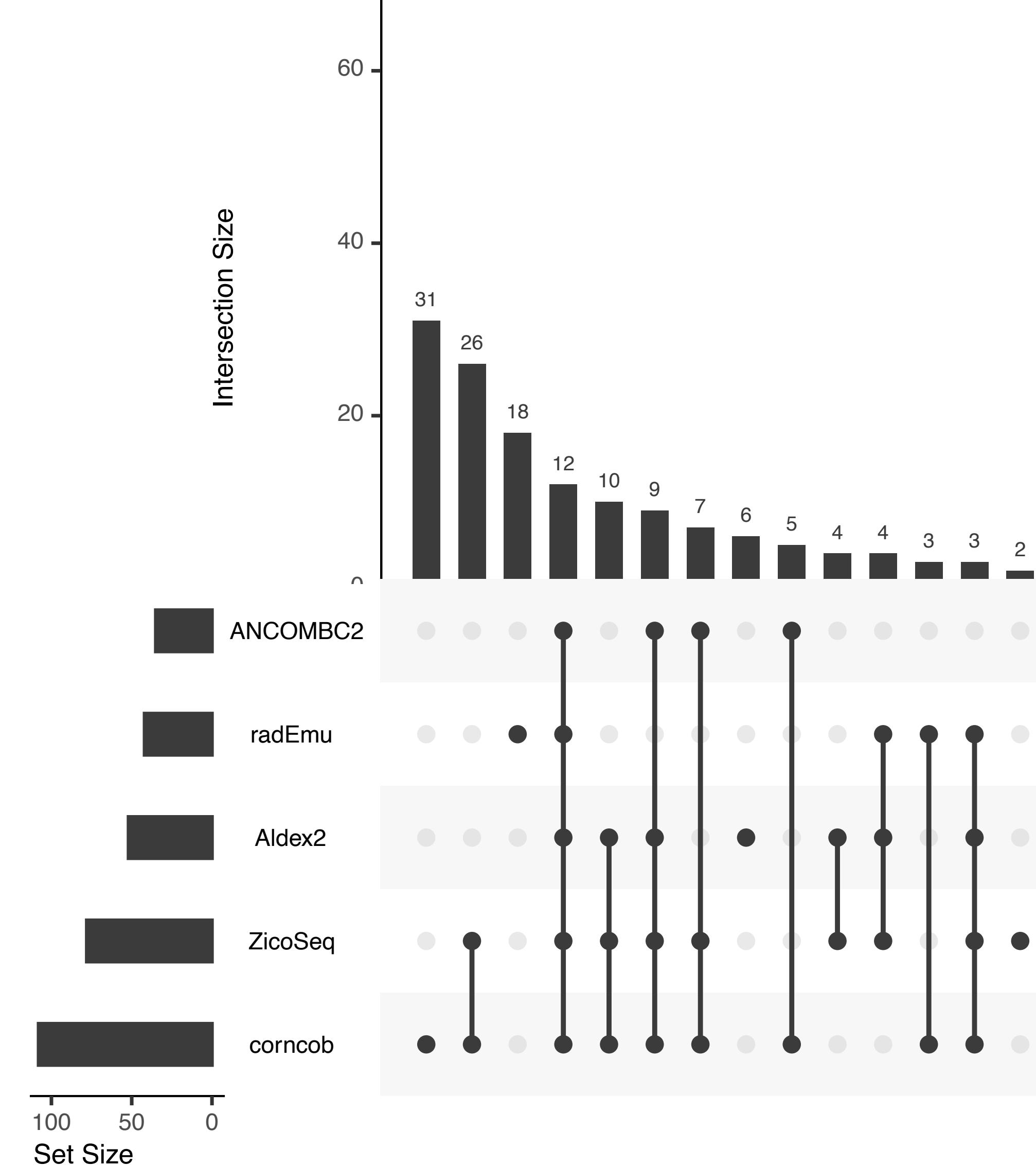
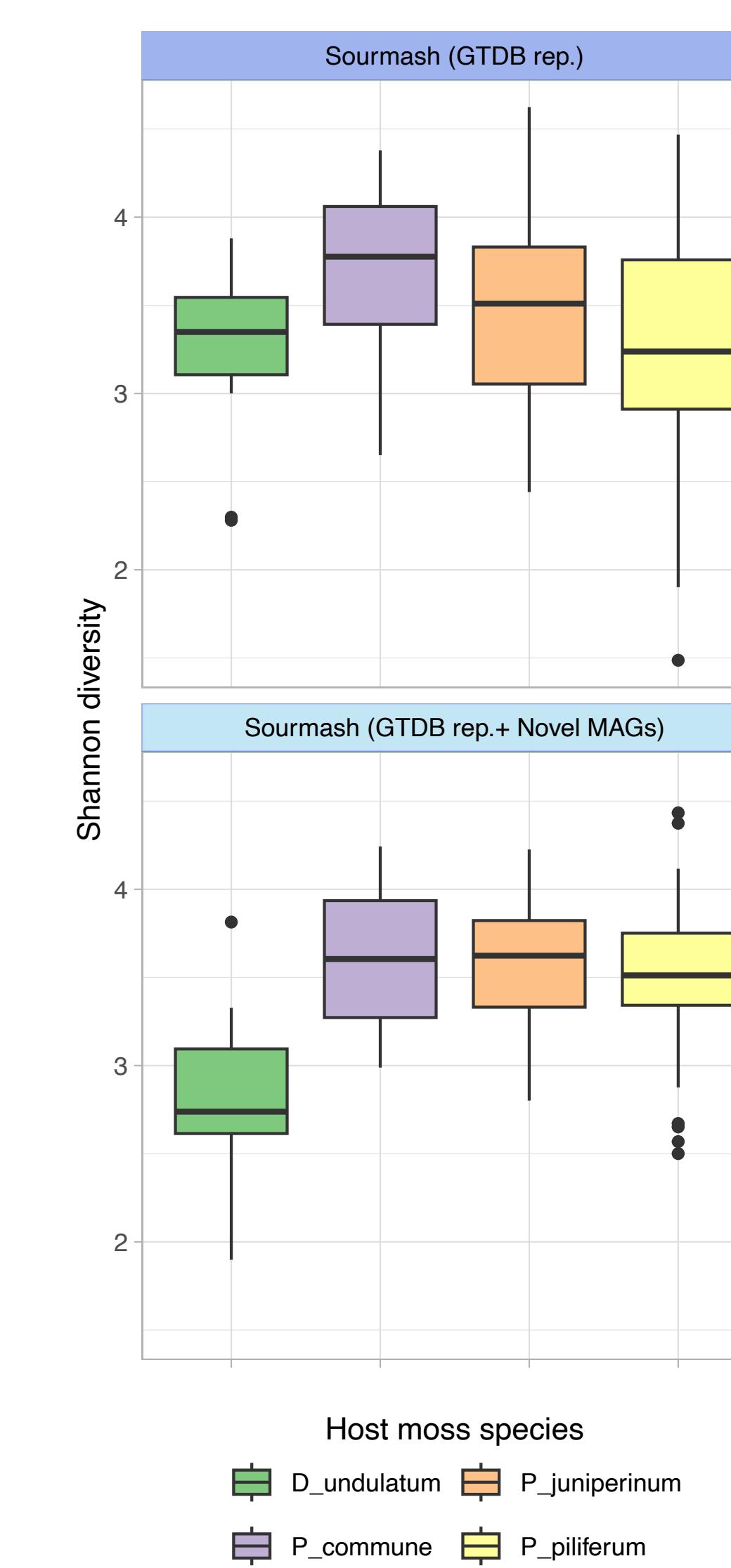


Top 20 overall detected families identified as differentially abundant across all combinations of 6 community composition estimation approaches and 6 differential abundance testing methods. Enriched in gut microbiome of persons with NAFLD (blue) or of healthy controls (red). 17 samples in each group, adj. $p < 0.05$.

DAA TOOL CONSENSUS INCREASES WITH NOVEL MAGs ADDITION

DAA TOOL CONSENSUS INCREASES WITH NOVEL MAGs ADDITION

Full consensus increases from 12 to 18 significantly differentially abundant species across two sample types in 131 boreal moss metagenomes. Four-tool consensus increases from 25 to 37.



We acknowledge the support of :

