



**HAL**  
open science

# Mapler: Assessing assembly quality in taxonomically-rich metagenomes sequenced with HiFi reads

Nicolas Maurice, Clémence Frioux, Claire Lemaitre, Riccardo Vicedomini

► **To cite this version:**

Nicolas Maurice, Clémence Frioux, Claire Lemaitre, Riccardo Vicedomini. Mapler: Assessing assembly quality in taxonomically-rich metagenomes sequenced with HiFi reads. 2025 - Journées scientifiques Agroécologie et Numérique, Jan 2025, Dijon, France. , pp.1-1, 2025. hal-04941137

**HAL Id: hal-04941137**

**<https://hal.science/hal-04941137v1>**

Submitted on 11 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Mapler: Assessing assembly quality in taxonomically-rich metagenomes sequenced with HiFi reads



Nicolas Maurice<sup>1,2</sup>, Clémence Frioux<sup>2</sup>, Claire Lemaitre<sup>1</sup>, Riccardo Vicedomini<sup>1</sup>,  
<sup>1</sup>Inria/IRISA Genscale, Campus de Beaulieu, 35042 Rennes, France; <sup>2</sup>Inria/INRAE Pleiade, Campus de Talence, 33405 Bordeaux, France  
 nicolas.maurice@inria.fr



## Introduction

### Plant microbiomes

Plants recruit their microbiomes from millions ( $10^3$ - $10^8$ /g) of soil micro-organisms, to regulate key functions such as [1]:

- Nutrition and growth,
- Abiotic stress,
- Biotic stress.

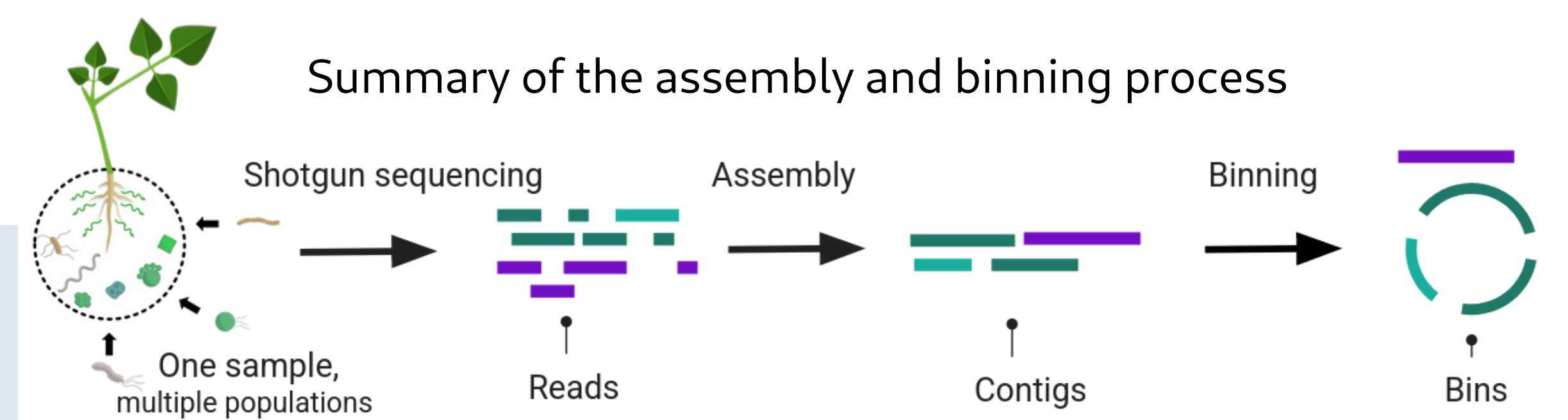
Yet, those taxonomically-rich ecosystems are poorly characterised. It is unclear how much of the improvements in sequencing and assembly scale to those complex ecosystems.

### Metagenome evaluation

In taxonomically-rich ecosystems, metagenome evaluation is a challenging task [2], due to:

- Low quality assemblies
- A lack of references databases

We present Mapler, a metagenome assembly and evaluation pipeline for HiFi reads in taxonomically-rich ecosystems.



## How to evaluate the low-quality metagenome assemblies in soil and plant microbiomes ?

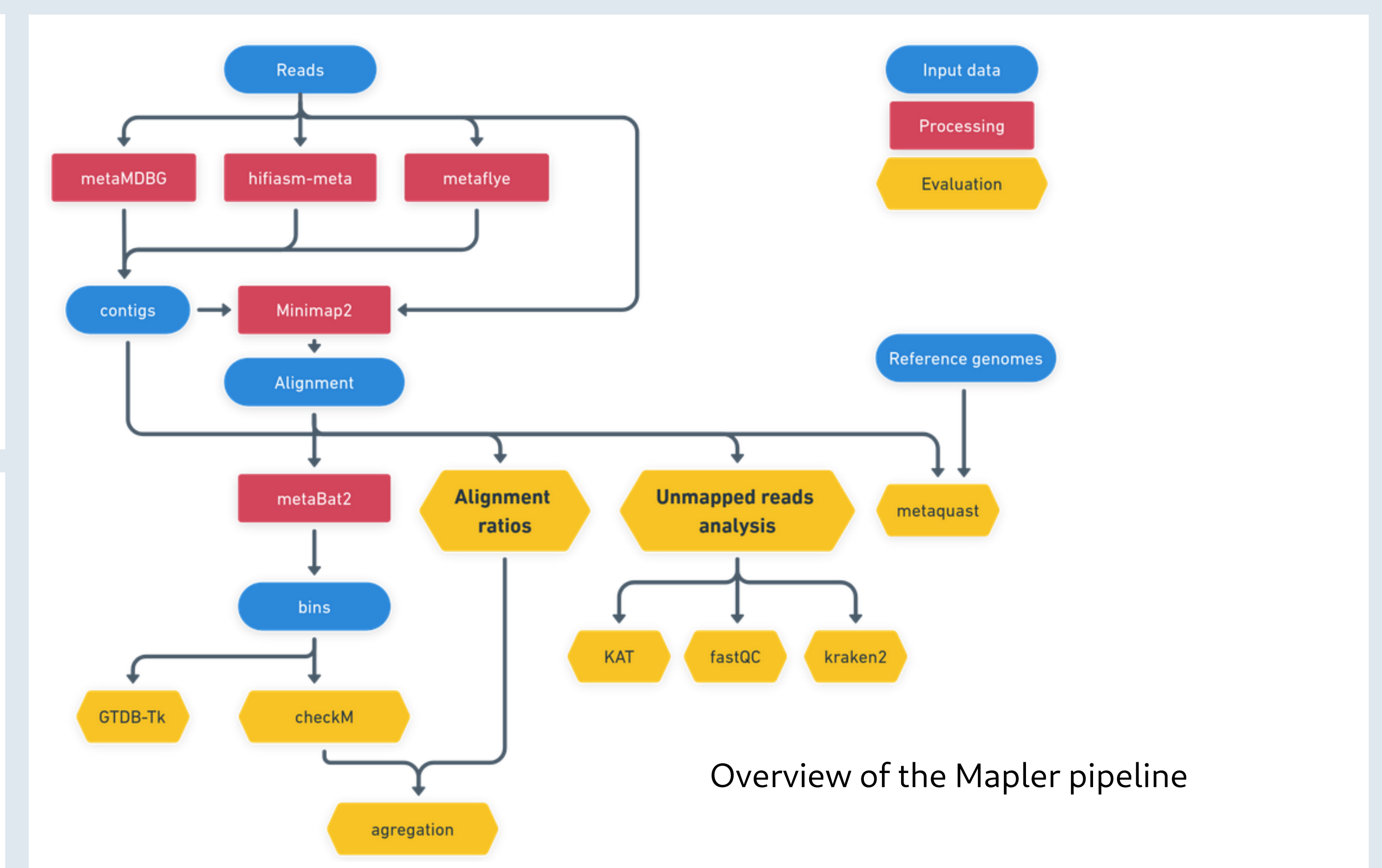
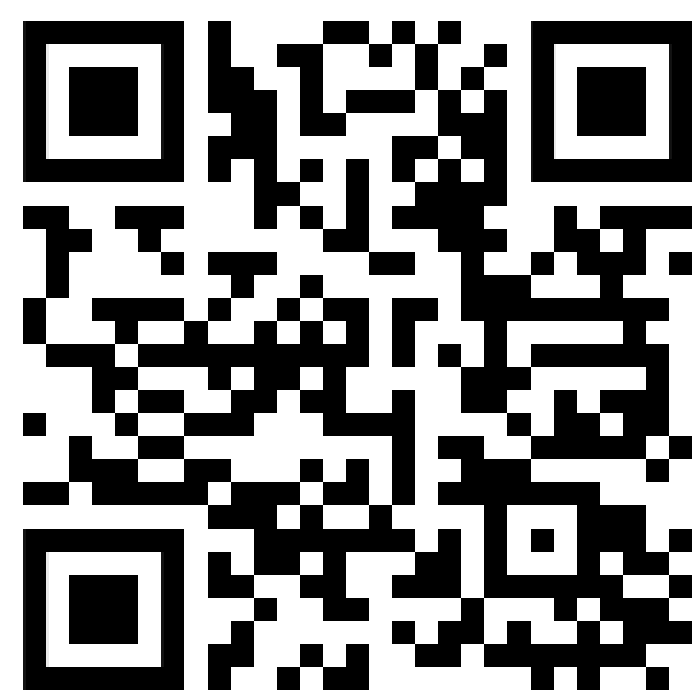
## Software description

### Overview

Mapler assembles, bins and evaluate multiple samples in parallel locally or on a cluster. After downloading it and snakemake, it installs dependencies itself.

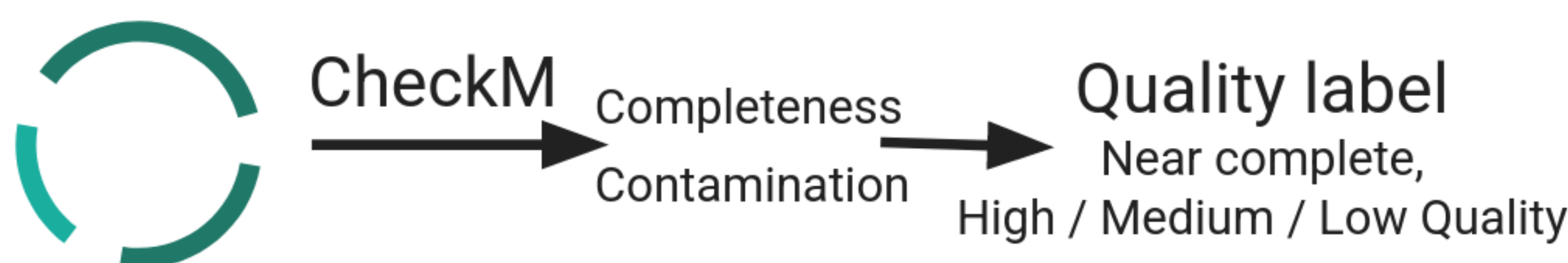
It can be found in an upcoming paper and at:

[gitlab.inria.fr/mistic/mapler](https://gitlab.inria.fr/mistic/mapler)



### Bin quality

The higher the completeness and the lower the contamination, the higher the quality.



### Alignment ratios



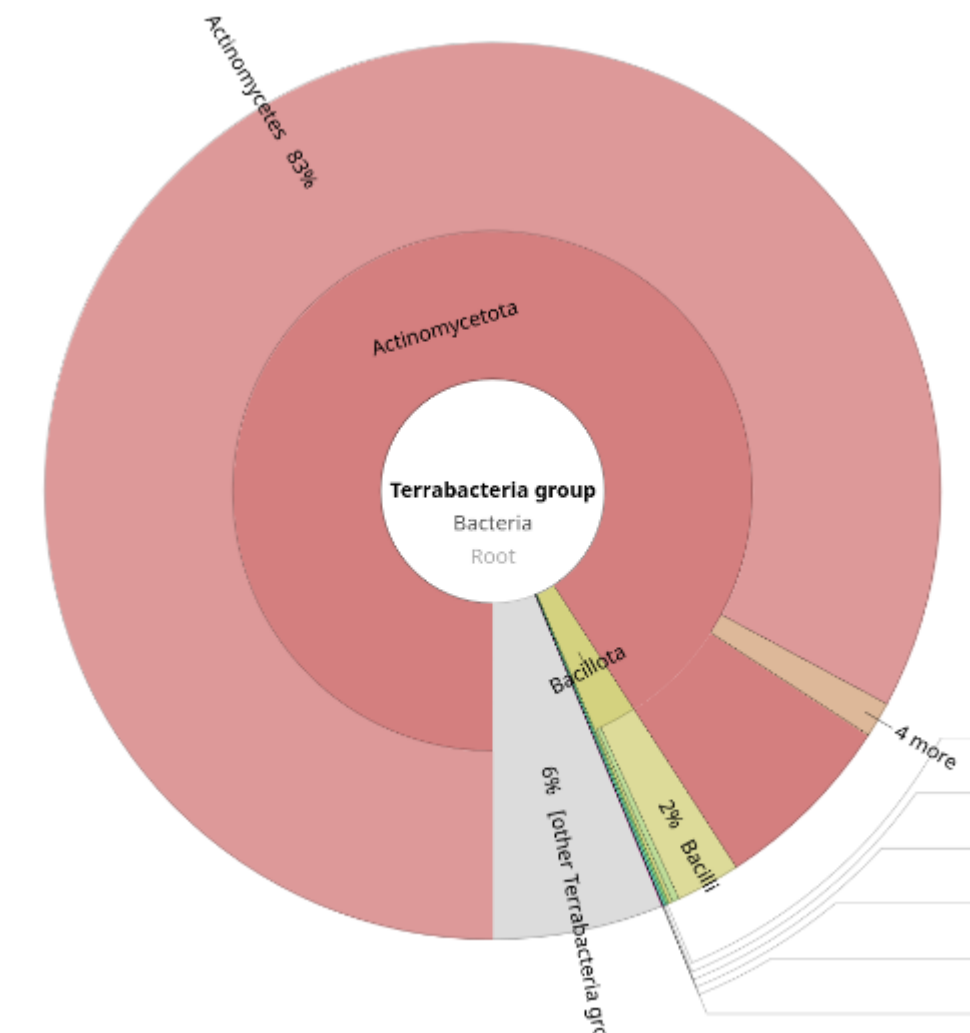
From the reads on contigs alignment:

- Aligned reads ratio : Aligned reads count / Total reads count
- Aligned bases ratio : Aligned bases count / Total reads length

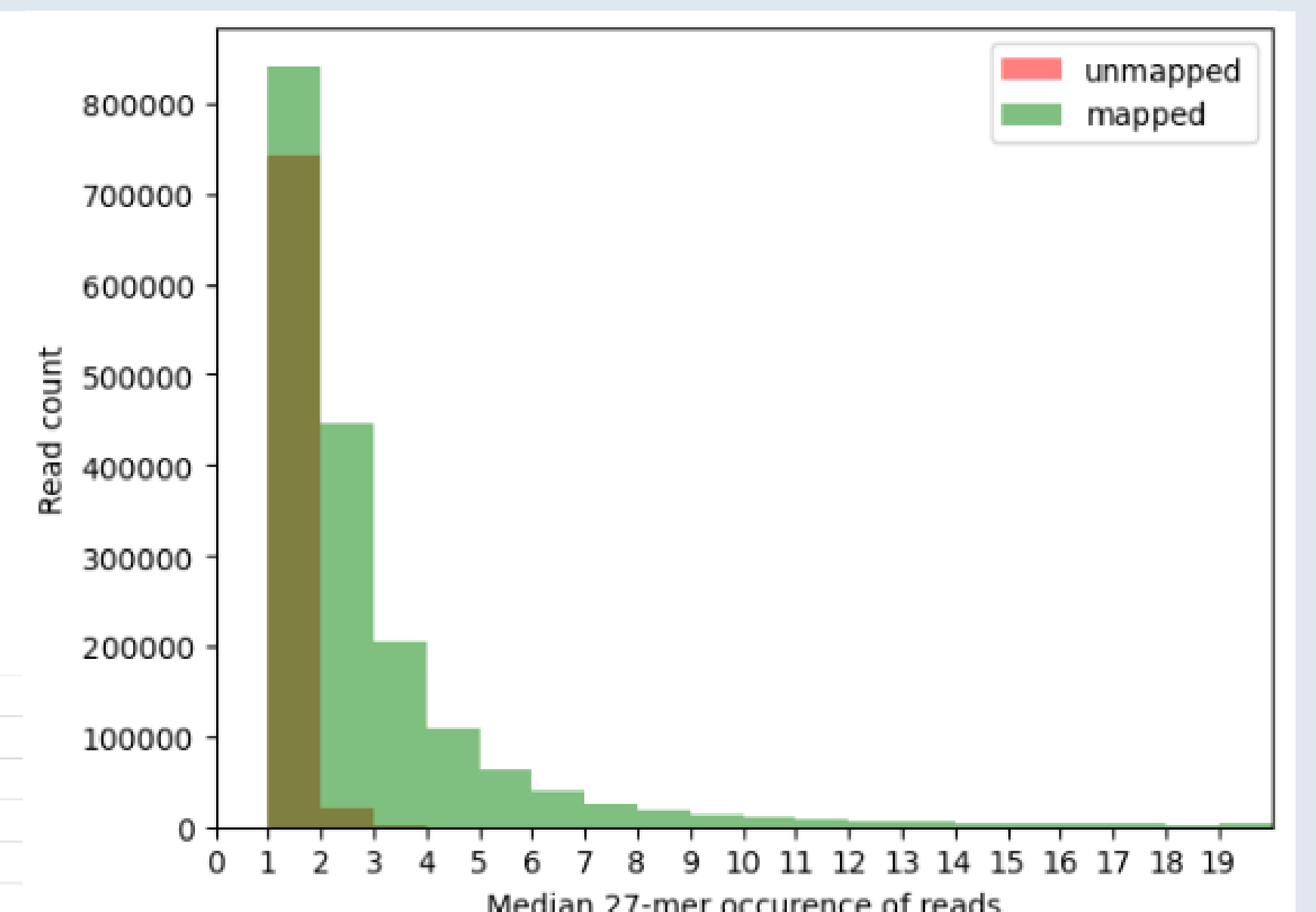
Combined with bin quality, estimate how much diversity is captured at each level of quality. Difference between aligned reads and bases ratio indicate partially aligned reads.

Mapler estimate divergence between aligned reads (participated in the assembly) and unaligned reads using:

- FastQC : read length, read quality
- Kraken 2 : taxonomic assignment
- KAT [3] : read abundance



Taxonomic assignment in unaligned reads of a metaMDBG rhizosphere assembly



Abundance distribution in a metaMDBG rhizosphere assembly

## Application

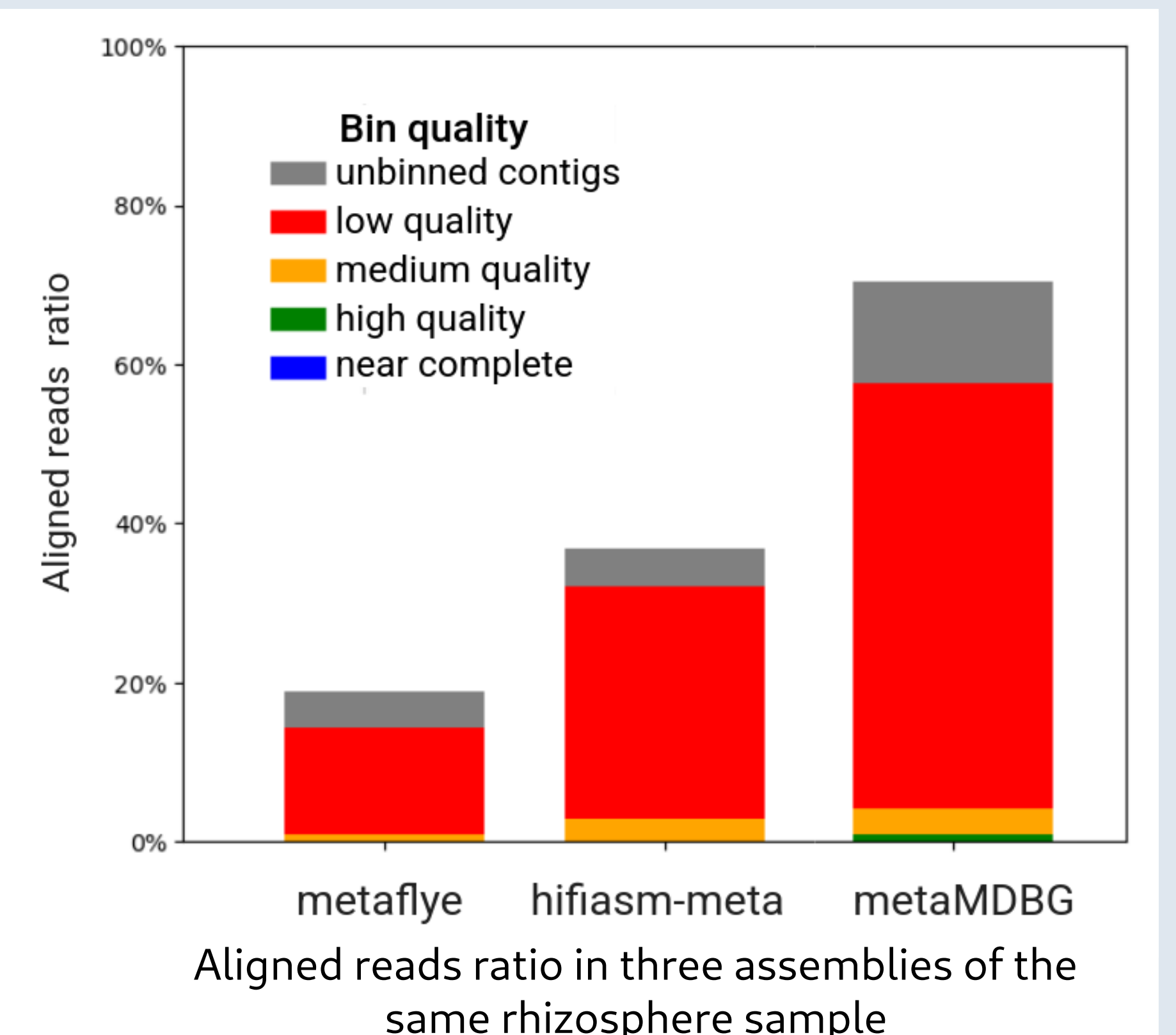
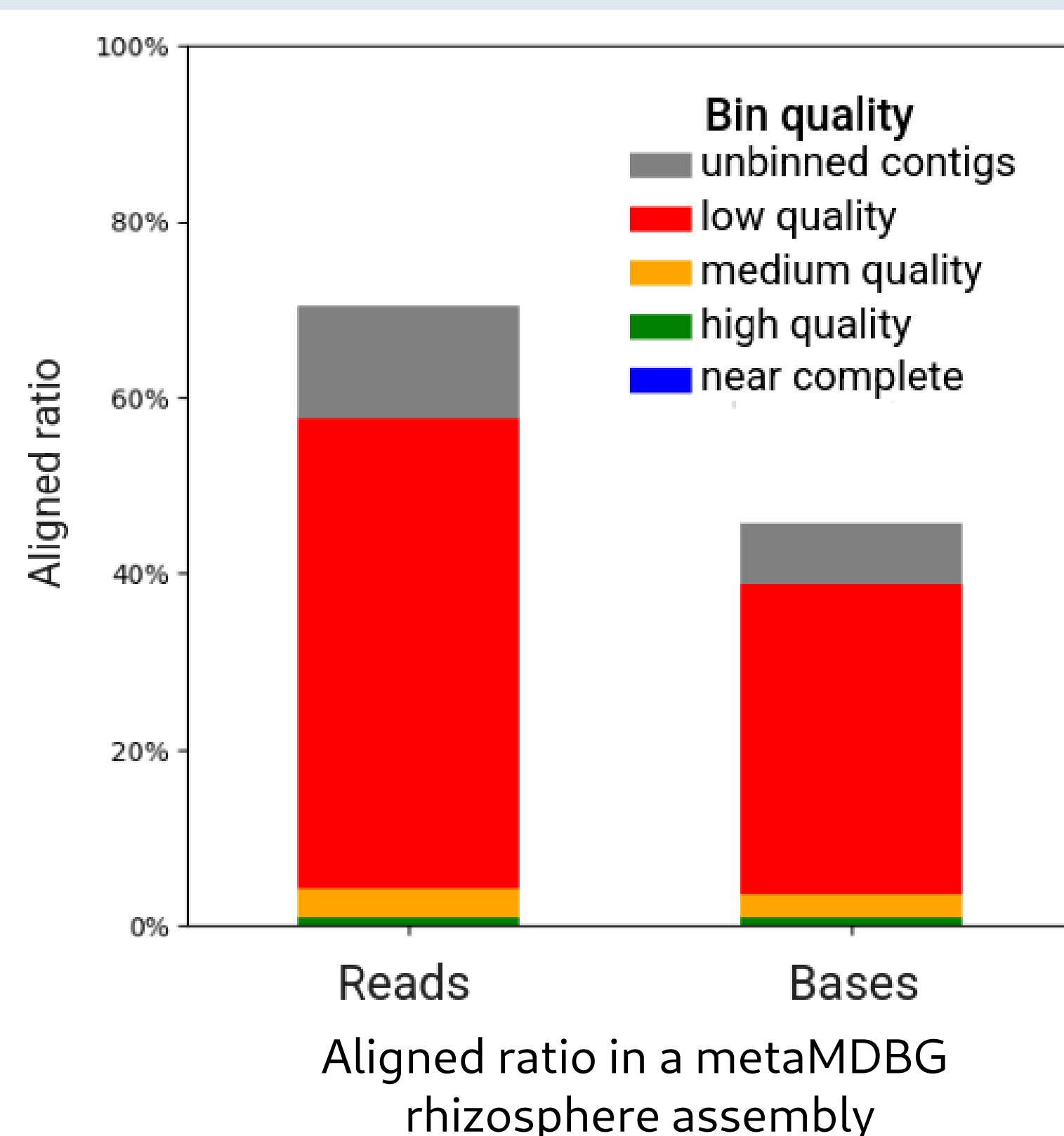
### Comparing assembly methods in taxonomically rich ecosystems

In highly taxonomically-rich ecosystems such as soil, most of the diversity is uncaptured:

- Several taxa are not assembled
- Most low abundance species are not assembled
- Most of what is assembled is only partially assembled

Still, some methods fare better than others, and metaMDBG currently capture a higher proportion of the diversity at all levels of quality than competing assemblers.

Mapler is also used in the elaboration of two upcoming papers, in link with the MISTIC project of the PEPR: an application note presenting the pipeline itself and a paper on the comparison of short and long reads in soil.



## Conclusion

Mapler is an easy-to-use evaluation pipeline that can be used to compare samples and assembly methods in taxonomically-rich ecosystems, by aggregating state-of-the-art tools and exploiting the alignment of reads on contigs.



MISTIC



[1] Hoysted G. et al., 2018, Curr. Op. in Pl. Bio  
 [2] Fierer N., 2017, Nat Rev Microbiol  
 [3] Mapleson D. et al., 2016, Bioinformatics