

### Introduction

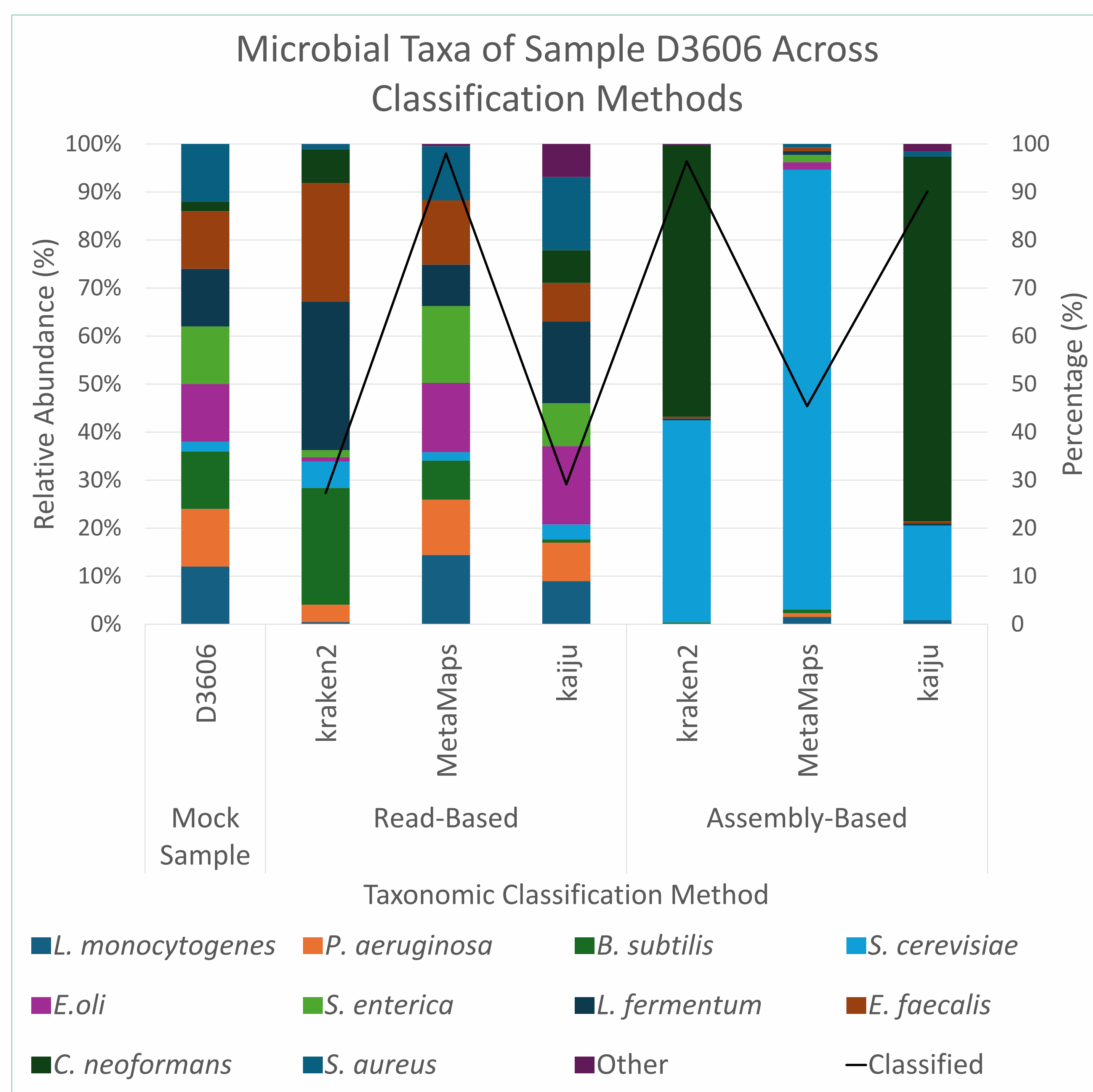
Climate change calls for innovative solutions to reduce waste and carbon emissions. Mixed Culture Fermentation (MCF) offers a sustainable approach by converting waste streams into valuable products like Volatile Fatty Acids (VFAs), which have industrial applications such as biodegradable plastics [1][2]. Optimizing VFA production is critical to improving efficiency and reducing costs, and this depends on the dynamics of mixed microbial communities.

Understanding microbial interactions and their relationship with operational conditions is key to controlling VFA production. However, **strain-level classification of microorganisms** remains challenging due to genetic similarities and big data. To address this, *StrainoScope* has been developed, a **bioinformatic pipeline** for strain-level taxonomic classification of microorganisms using metagenomic and big data, providing insights that can enhance the MCF of VFA.

### Results

After running *StrainoScope* the results revealed that the **read-based method** with the option MetaMaps is the most homogeneous with the mock sample. Moreover, it has the highest classified percentage, at around 98%.

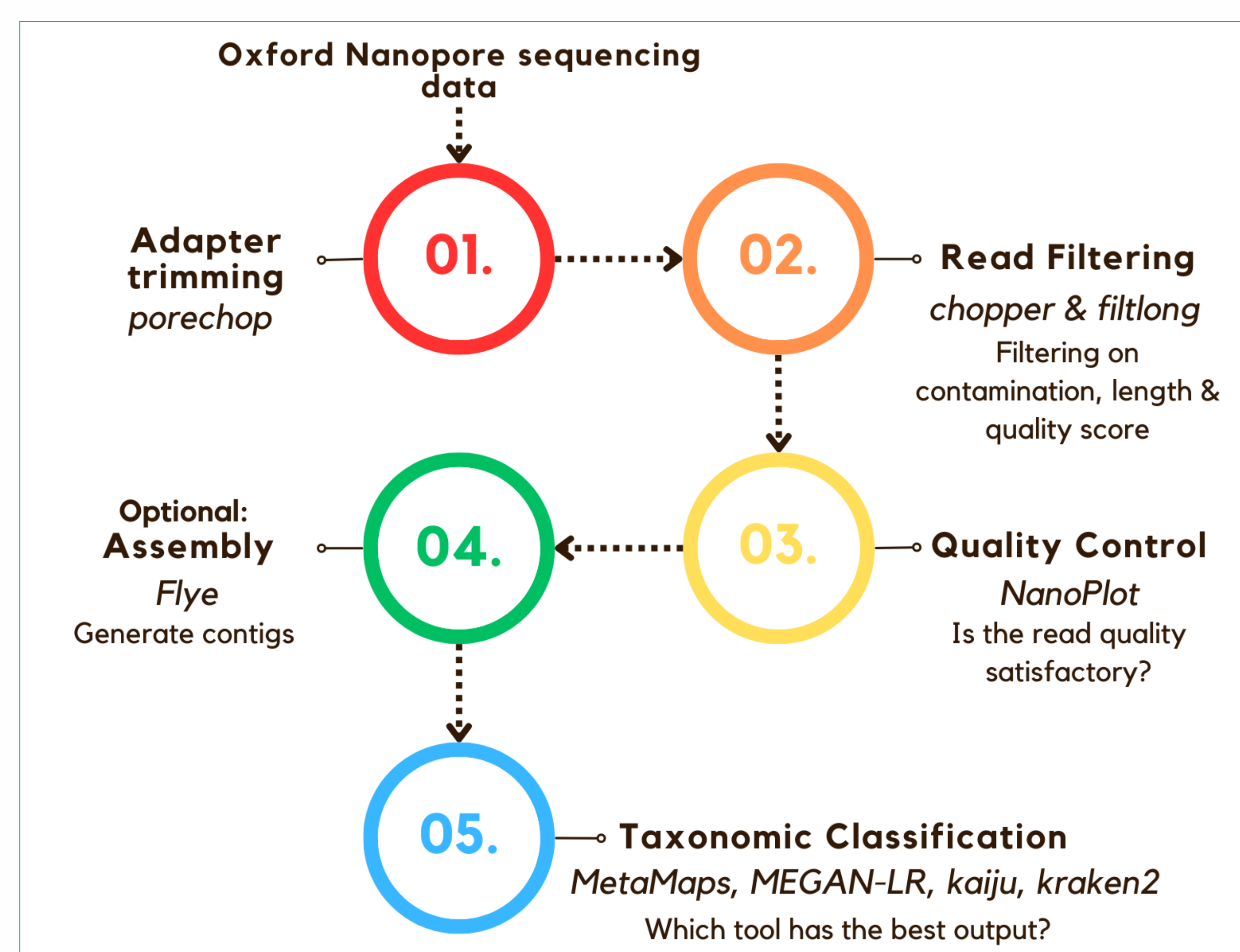
Other read-based methods diverge relatively more and not a lot of reads are classified. In comparison to the read-based methods, the assembly-based methods deviate a great quantity from the mock sample.



**Fig 2.** Comparison of microbial community composition in sample D3606 using various taxonomic classification methods (read- and assembly-based), with the tools kraken2, MetaMaps, and kaiju. Relative abundances of taxa are shown alongside the percentage of classified reads or contigs (black line).

### Material & Methods

The workflow of *StrainoScope*, a Python based pipeline, is illustrated in figure 1. The pipeline's input was a mock sample consisting of 10 species, with known relative abundances quantitatively distributed as either 14% or 2%. The pipeline was executed using various taxonomic classification methods, including both read-based and assembly-based approaches, with the computational tools MetaMaps, Kraken2, and Kaiju.



**Fig 1.** Workflow of the analysis with *StrainoScope* using Oxford Nanopore sequencing data. Key steps include pre-processing, quality control, assembly, and taxonomic classification.

### Discussion & Conclusion

*StrainoScope* delivers the most promising results using the **read-based method with MetaMaps**, as it closely aligns with the relative abundances of mock sample D3606. In addition to its **accuracy**, this method also achieves a high recall score, with nearly 98% of the reads successfully classified.

Other read-based methods show significantly lower recall, resulting in a substantial portion of the data being unclassified and causing an inaccurate representation of the species. On the other hand, assembly-based methods generally demonstrate higher recall but fail to provide correct relative abundances.

In conclusion, *StrainoScope* produces the most accurate results by running a read-based analysis with the taxonomic classification option MetaMaps. In future research, it is essential to test the pipeline on real data.

### References

- [1] Atasoy, M., et al. DOI: <https://doi.org/10.1016/j.biortech.2019.121889> Bioresource Technology 292(2019)
- [2] Hesselstine, C. W. DOI: <https://doi.org/10.17226/1939> Publisher: National Academies Press (US), Washington (DC), 1992; pp. 52-57

### Acknowledgements

I would like to express my sincere gratitude to my supervisors Bazante Sanders and Tim Verschuren.

This project is financially supported by the European Just Transition Fund via project FermiChem (JTFZ-00022).

Contact information: t.verschuren3@avans.nl