

## **DUDes: a top-down taxonomic profiler for metagenomics samples**

*Piro, Vitor C., Lindner, Martin S., Renard, Bernhard Y.*

*Research Group Bioinformatics (NG4), Robert Koch Institute, Nordufer 20, 13353 Berlin, Germany*

The fast increase of complete genome sequences available on public databases has allowed better predictions of the microbial content from sequenced environmental and clinical samples. The identification of species and their quantification are common tasks in metagenomics and pathogen detection studies. The most recent techniques are built on mapping the sequenced reads against a reference database (e.g., whole genomes, marker genes, proteins) and performing further analysis. Although these methods proved to be useful in many scenarios, there is still room for improvement in species and strain level detections, mainly for low abundant organisms. We propose a new method: DUDes, a reference-based taxonomic profiler that introduces a novel top-down approach to analyze metagenomic NGS samples. Differently from the main taxonomic profiling tools that base their predictions estimating abundances in the sample, DUDes does not account for abundances directly as a mean of identification. Our method identifies possible candidates by comparing the strength of the read mapping in each node of the taxonomic tree in an iterative manner. Instead of using the lowest common ancestor (LCA), a commonly used bottom-up approach to solve ambiguities in identifications, we propose a new approach: the deepest uncommon descendent (DUD). Differently from the LCA method that solves ambiguous identifications by going back one taxonomic level to the lowest common ancestor, the DUD approach starts at the root node and tries to go for deeper taxonomic levels, even when ambiguities are found. That way it is possible to have less conservative identifications in higher taxonomic levels. Besides, when the provided data does not allow a specific identification on higher levels, the method can identify a set of probable candidates. Permutation tests are performed to estimate p-values between nodes and to identify the presence of them on each level. We showed in experiments that DUDes works for single and multiple organisms and can identify low abundant taxonomic groups with high precision.