

## Sample origin prediction in forensics with use of targeted metagenomic sequencing

Paweł P Łabaj Małopolska Centre of Biotechnology, Jagiellonian University, Kraków, Poland

**Abstract:** Microbiome data have been successfully applied in forensics. However, studies by MetaSUB [1] and CAMDA [2] have shown that the full potential of metagenomics is yet to be unveiled. To allow the unconstrained analysis and interpretation the reference free approaches need to be used.

Here we characterize the composition of soil microbiome in Poland with MetaGraph [3] on Whole Metagenome Sequencing data. We have collected about 1000 samples throughout the year from multiple locations and then sequenced. Those constitute “positive class” for Poland while MetaSUB and EMP500 collections “negative class”. We have constructed corresponding graphs and extracted features in form of the unitigs, which are sampling-site specific. Those were used to design a targeted panel allowing to obtain metagenomic profiles of high uniqueness. The obtained from TMS sample profiles are first clustered by their similarity to each other and as embeddings are passed to the DNN classifier, which predicts the probability of the origin of the sample.

In the pilot we have tested samples from 6 Polish cities with classifier trained on fraction of WMS data. Still the results are very promising (accuracy >90%), indicating that application of k-mer based approaches will be a step towards new era of metagenomics in forensic applications.

1. Danko et al. A global metagenomic map of urban microbiomes and antimicrobial resistance (2021) *Cell* 184(13)
2. <https://camda.info>
3. Karasikov et al. Metagraph: Indexing and analyzing nucleotide archives at petabase-scale (2020) *BioRxiv*, <https://doi.org/10.1101/2020.10.01.322164>