# TAXAPRO, A Streamlined Pipeline to Analyze Shotgun Metagenomes

**Authors :** Sofia Sehli, Zainab El Ouafi, Casey Eddington, Soumaya Jbara, Kasambula Arthur Shem, Islam El Jaddaoui, Ayorinde Afolayan, Olaitan I. Awe, Allissa Dillman, Hassan Ghazal

**Abstract :** The ability to promptly sequence whole genomes at a relatively low cost has revolutionized the way we study the microbiome. Microbiologists are no longer limited to studying what can be grown in a laboratory and instead are given the opportunity to rapidly identify the makeup of microbial communities in a wide variety of environments. Analyzing whole genome sequencing (WGS) data is a complex process that involves multiple moving parts and might be rather unintuitive for scientists that don't typically work with this type of data. Thus, to help lower the barrier for less-computationally inclined individuals, TAXAPRO was developed at the first Omics Codeathon held virtually by the African Society for Bioinformatics and Computational Biology (ASBCB) in June 2021. TAXAPRO is an advanced metagenomics pipeline that accurately assembles organelle genomes from whole-genome sequencing data. TAXAPRO seamlessly combines WGS analysis tools to create a pipeline that automatically processes raw WGS data and presents organism abundance information in both a tabular and graphical format. TAXAPRO was evaluated using COVID-19 patient gut microbiome data. Analysis performed by TAXAPRO demonstrated a high abundance of Clostridia and Bacteroidia genera and a low abundance of Proteobacteria genera relative to others in the gut microbiome of patients hospitalized with COVID-19, consistent with the original findings derived using a different analysis methodology. This provides crucial evidence that the TAXAPRO workflow dispenses reliable organism abundance information overnight without the hassle of performing the analysis manually.

**Keywords :** metagenomics, shotgun metagenomic sequence analysis, COVID-19, pipeline, bioinformatics
**Conference Title :** ICACSB 2022 : International Conference on Applications of Computational Structural Bioinformatics
**Conference Location :** Istanbul, Türkiye
**Conference Dates :** May 05-06, 2022