# The development of a metagenomics module for the Nutritional Phenotype Database

**Robert Horlings[2], Martien Caspers[1], Jildau Bouwman[1], Kees van Bochove[1,3], Ben van Ommen[1,4], Roy Montijn[1], Bart Keijser[1]**

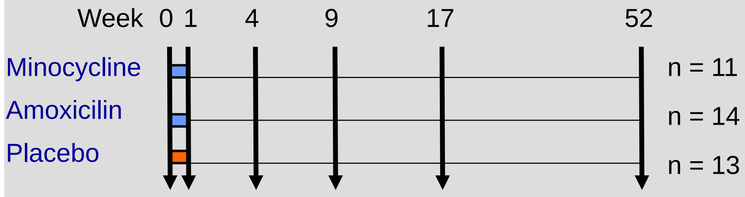
[1]TNO Microbiology and Systems Biology, Zeist, The Netherlands
[3] IsDat software development, Utrecht, The Netherlands,
[3] The Hyve, Utrecht, The Netherlands
[4] NuGO;
* Bart.Keijser@tno.nl

The human body provides an excellent niche for a diverse set of microorganisms that occupy various sites within the human body. Recent studies have revealed that the number of microbial cells within the human body exceeds that of the human host by a factor ten, and the number of microbial genes (the human microbiome) exceeds that of the human host by a factor 100. The human host has co-evolved with its indigenous microbiota, resulting in a deeply intertwined physiological relationship. In our efforts of gaining better understanding of the human physiology and the impact of food ingredients on human health, integration of human -omics data with those on the human microbiome is an absolute necessity. To facilitate this integration, we have expanded an available platform (Nutritional Phenotype Database - dbNP) for human systems biology study evaluations with a module for metagenomics data. The current version of the metagenomics module enables researchers to:

- Store metagenomic and contextual data from intervention studies
- Perform database searches based on study and microbiota characteristics
- Merge and export datasets for in silico meta-analysis

The NuGO network ensures broad European coverage and accessibility. The enhanced Nutritional phenotype database thus provides a systems biology study evaluation platform that enables multi-omics research on host and microbiota and facilitates collaborations between researchers at various locations

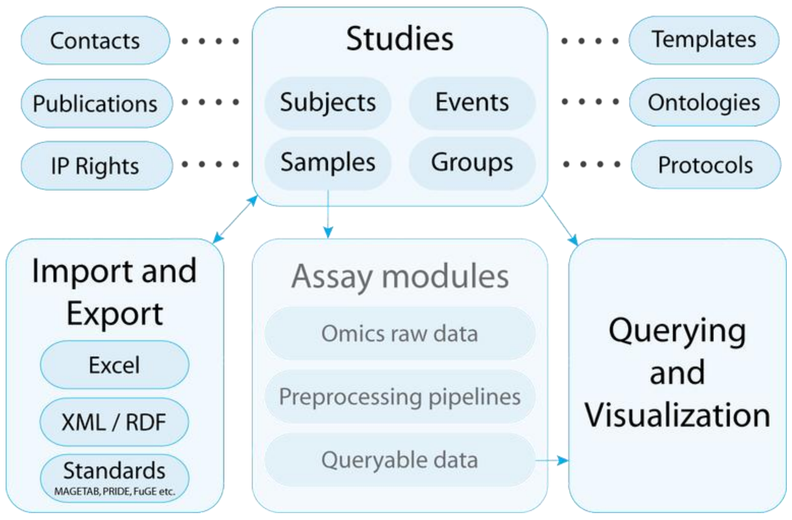
## Example study: Antibiotic intervention




Study design and volunteer phenotype

**1. Capturing study design**

The generic study capturing framework provides an ordered approach to document study set-up, volunteer phenotype and sample characteristics. In the documentation of study design, details are recorded of the subjects participating in the study, the events, the duration, and sampling. The use of templates ensures correct import of information. Importantly, queries can be used to retrieve information as well as associated study data for later analysis. Through assay modules, study data can be uploaded and linked to the study. The current version includes modules for metabolomic, transcriptome and clinical data. We have developed a novel module to enable storage of metagenomics data.

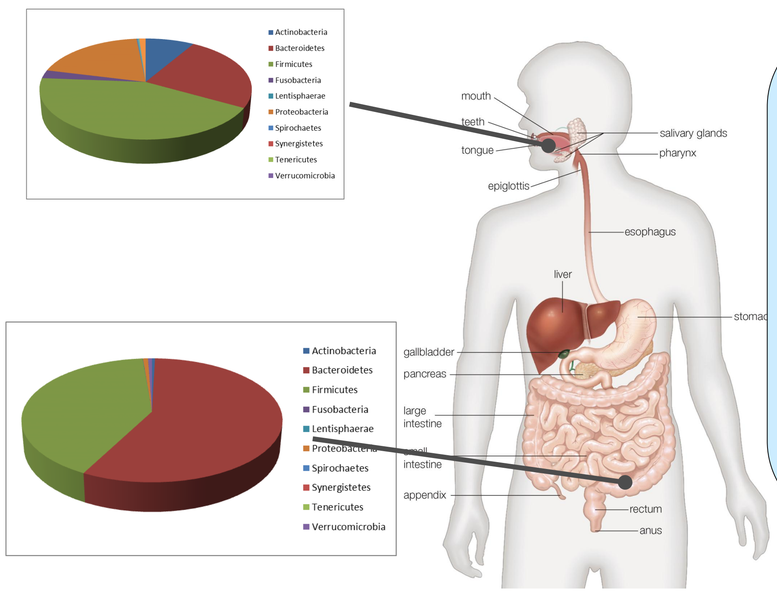



http://studies.dbnp.org

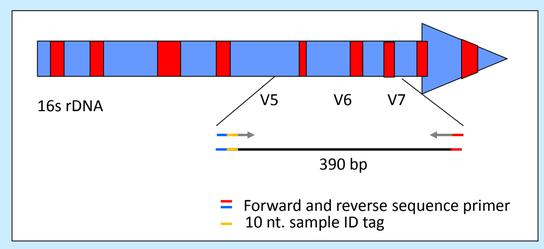
**2. Collecting and storing metagenomics data**

The indigenous microbiota composition can be evaluated through mass sequence analysis. In this approach, a hypervariable segment of the small ribosomal gene is amplified by PCR, introducing unique sample barcodes. Next, amplified DNA for the samples is pooled and sequenced. During data processing, sequence data is linked to the individual samples of origin and taxonomically classified.

The metagenomics module that we have developed enables researchers to upload metagenomic sequencing data along side with the taxonomic classification files, and to link these data with other study data.



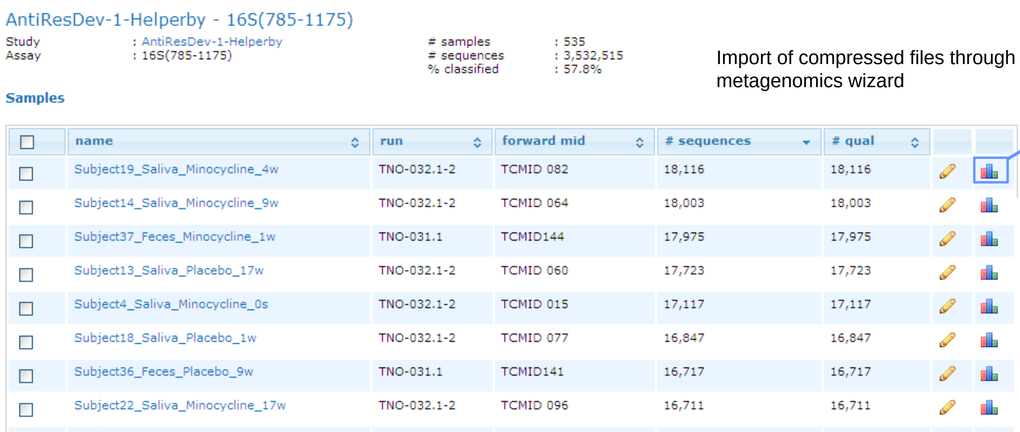

## Metagenomics Assay module



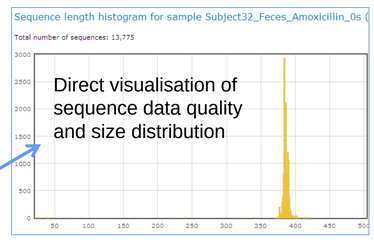

- Raw sequence data (.ssf)
- Primary data processing and quality filtering
- De-replicating sequence data according to sample ID *(.fa) (.qual)*
- Taxonomic classification file *(.tax)*
- Append metagenomics data with sample and study data

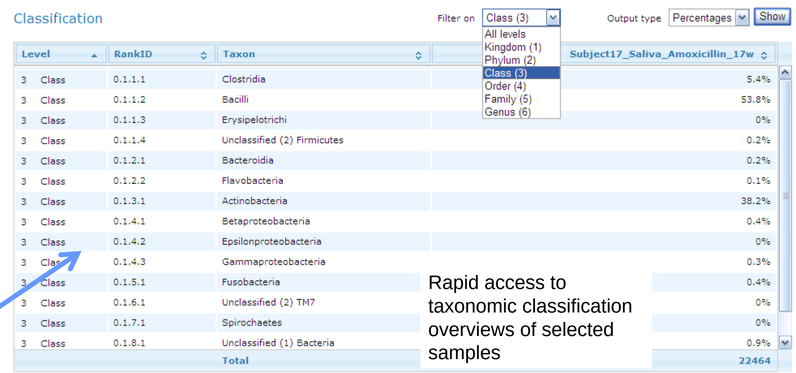
http://metagenomics.dbnp.org




Import of compressed files through metagenomics wizard



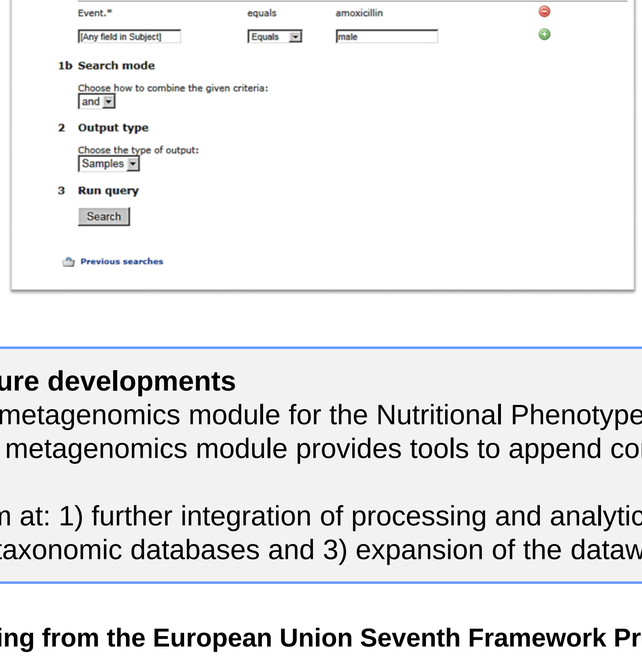



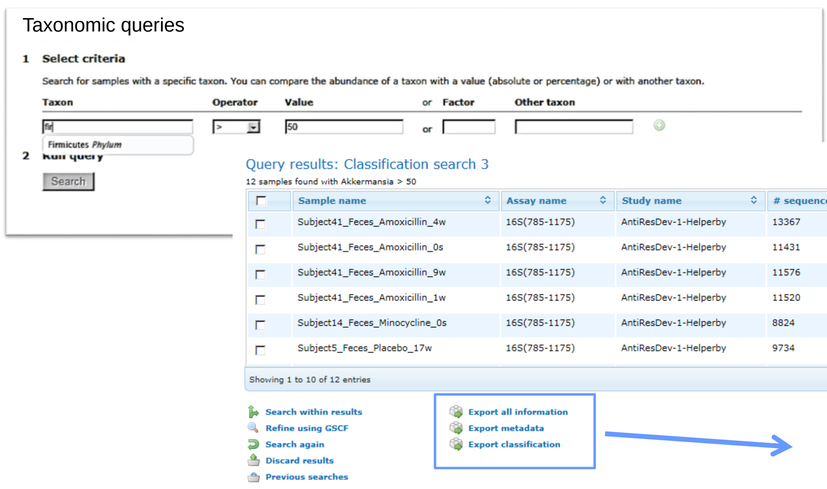

Rapid access to taxonomic classification overviews of selected samples

**3. Data visualization and retrieval**

The sequencing data quality can be inspected by direct visualization of sequence data quality and size distribution. Also the taxonomic distribution for selected samples can be displayed. Tailor made summaries of data can be exported to excel.

Through queries, data stored in the data warehouse system from multiple studies can be identified, combined and exported for analysis. This enables researchers to perform meta analysis of data from various studies. Complex queries can be build to identify sample data according to study design, clinical data, subject characteristics as well as taxonomic characteristics of the microbiome. Taxonomic queries can include abundance of selected taxonomic groups, but also co - occurrence or ratio's in abundance between taxonomic groups. The queries enable tailor made selection of data from multiple studies for meta analysis.




Direct export functions of selected data along with all meta data.

**4. Conclusions and future developments**

The development of the metagenomics module for the Nutritional Phenotype Database, expands its applicability towards host – microbiota-food interactions. The current metagenomics module provides tools to append community profiling data (16s, ITS and 18s) to the human nutrigenomics data stored in dbNP.

Future developments aim at: 1) further integration of processing and analytical tools with data warehousing capabilities, 2) Integration of microbial genetic, physiological as well as taxonomic databases and 3) expansion of the datawarehouse capabilities towards full metagenomic datamanagement.

**The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013), grant agreement n° 241446 (project ANTIRESDEV).**