

The development of a metagenomics module for the Nutritional Phenotype Database

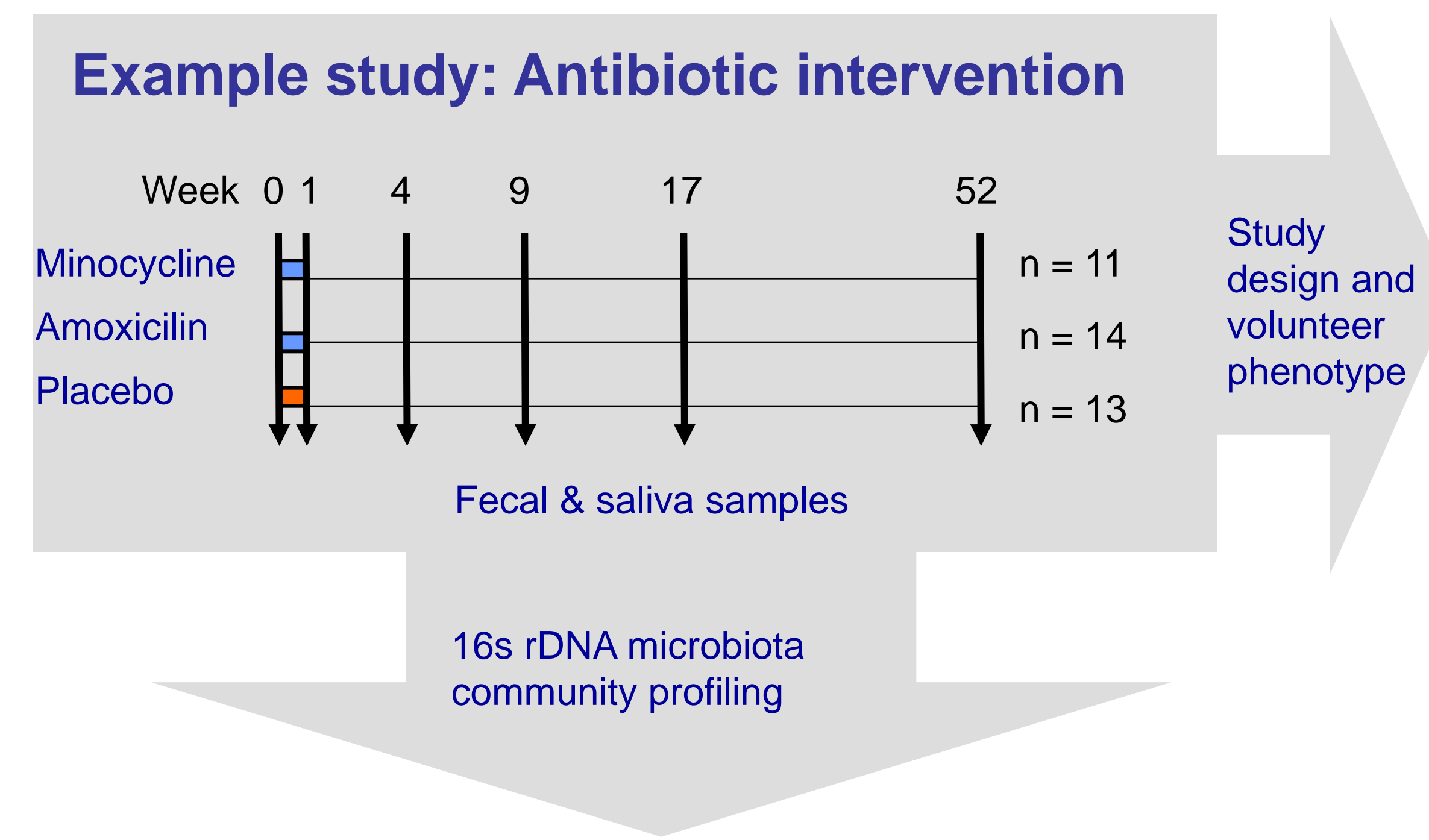
Robert Hurlings², Martien Caspers¹, Jildau Bouwman¹, Kees van Bochove^{1,3}, Ben van Ommen^{1,4}, Roy Montijn¹, Bart Keijser¹

¹TNO Microbiology and Systems Biology, Zeist, The Netherlands
²IsDat software development, Utrecht, The Netherlands,
³The Hyve, Utrecht, The Netherlands
⁴NuGO;
 * Bart.Keijser@tno.nl

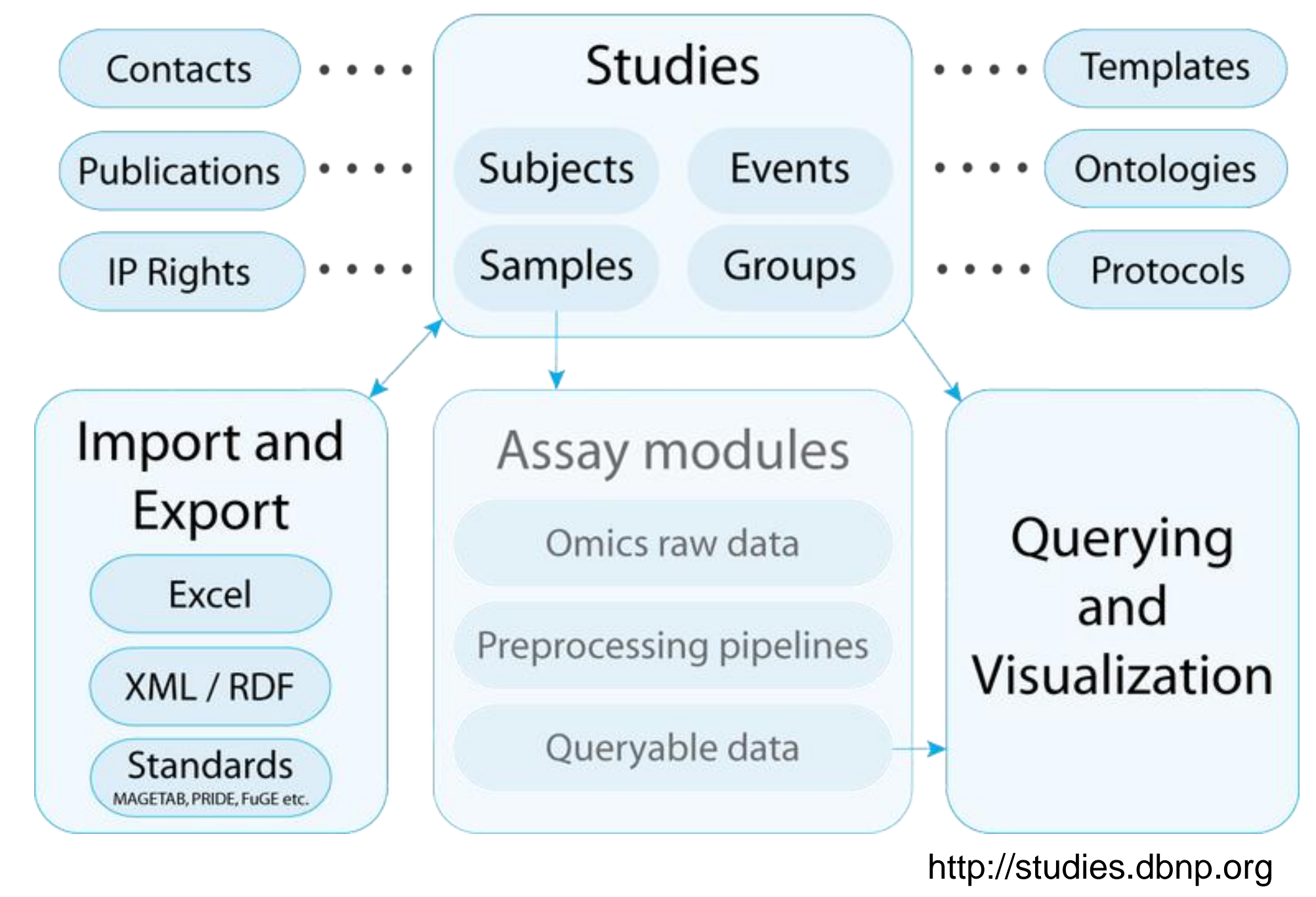
The human body provides an excellent niche for a diverse set of microorganisms that occupy various sites within the human body. Recent studies have revealed that the number of microbial cells within the human body exceeds that of the human host by a factor ten, and the number of microbial genes (the 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:

- o Store metagenomic and contextual data from intervention studies
- o Perform database searches based on study and microbiota characteristics
- o 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

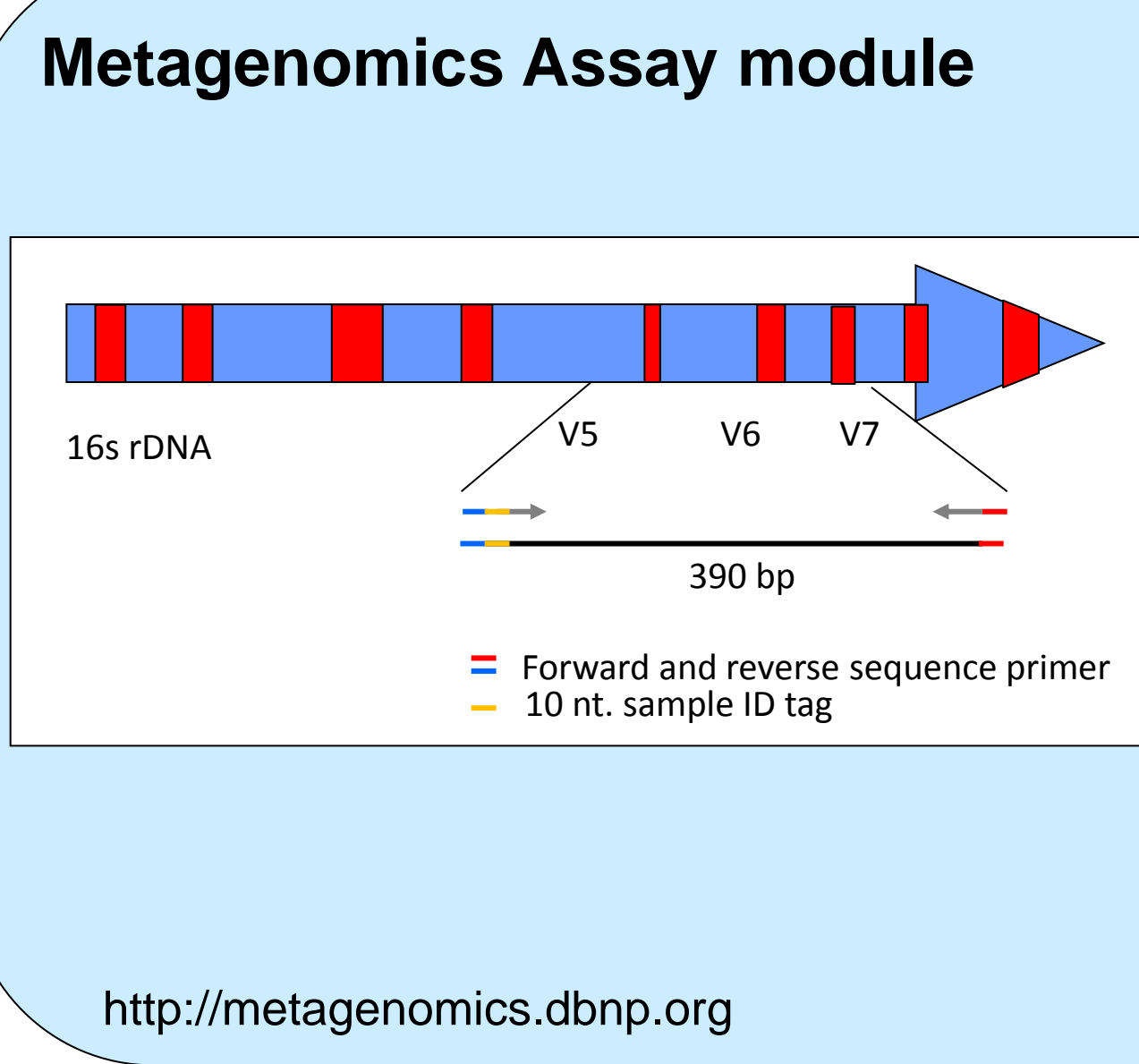
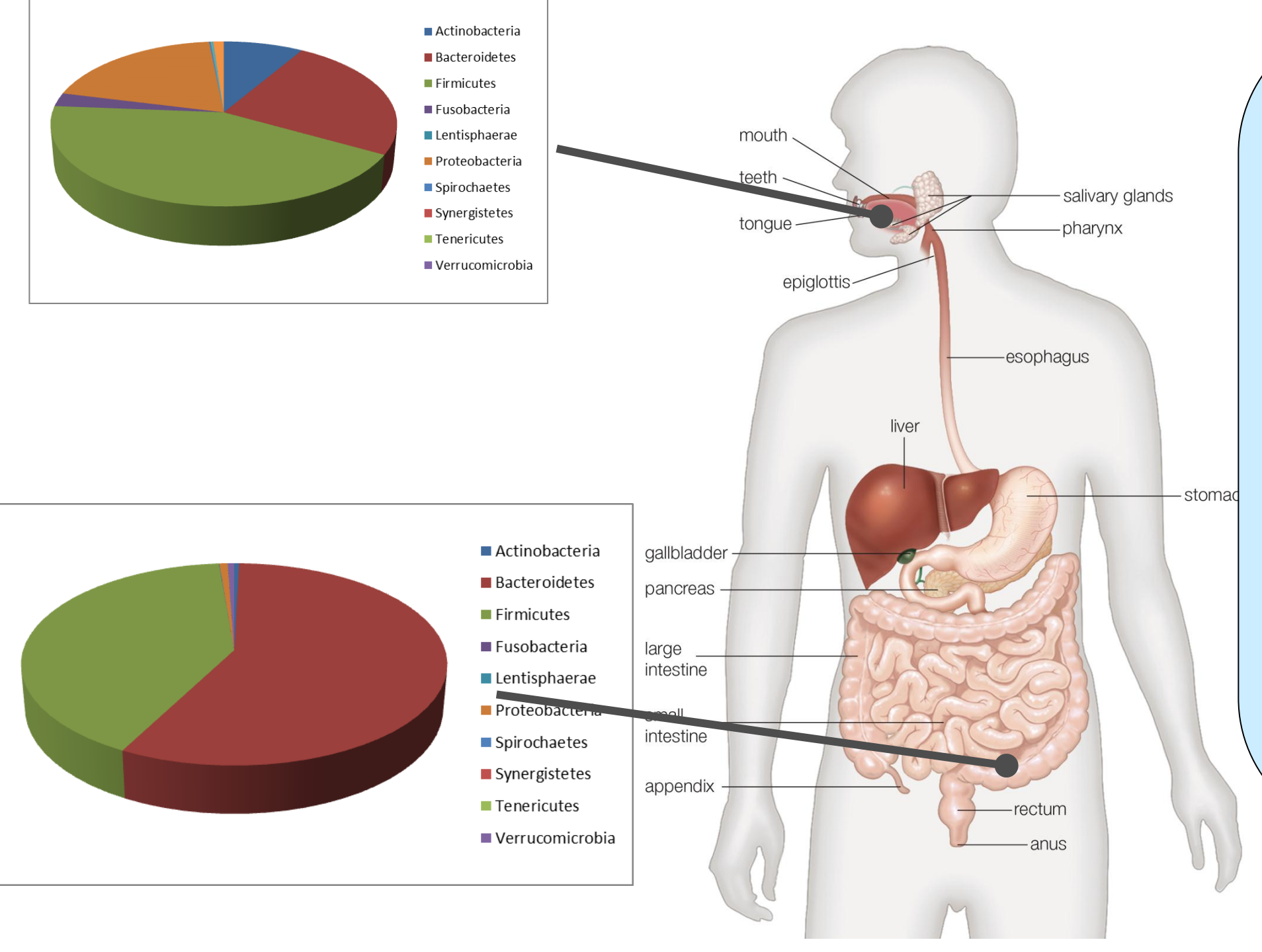


1. Capturing study design
 The generic study capturing framework provides an ordered approach to document study set-up, volunteer phenotype and sample characteristics. In the documentation of study design, details are recorded of the subjects participating in the study, the events, the duration, and 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.



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.



- 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

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.

AntiResDev-1-Helperby - 16S(785-1175)

Study Assay : AntiResDev-1-Helperby : 16S(785-1175)

samples : 535
 # sequences : 3,532,515
 % classified : 57.6%

Import of compressed files through metagenomics wizard

Direct visualisation of sequence data quality and size distribution

name	run	forward mid	# sequences	# qual
Subject19_Saliva_Minocycline_4w	TNO-032.1-2	TCMID 082	18,116	18,116
Subject14_Saliva_Minocycline_9w	TNO-032.1-2	TCMID 064	18,003	18,003
Subject37_Feces_Minocycline_1w	TNO-031.1	TCMID144	17,975	17,975
Subject13_Saliva_Placebo_17w	TNO-032.1-2	TCMID 060	17,723	17,723
Subject4_Saliva_Minocycline_0s	TNO-032.1-2	TCMID 015	17,117	17,117
Subject18_Saliva_Placebo_1w	TNO-032.1-2	TCMID 077	16,847	16,847
Subject36_Feces_Placebo_9w	TNO-031.1	TCMID141	16,717	16,717
Subject22_Saliva_Minocycline_17w	TNO-032.1-2	TCMID 096	16,711	16,711
Subject5_Saliva_Placebo_17w	TNO-032.1-2	TCMID 024	16,535	16,535
Subject12_Saliva_Minocycline_9w	TNO-032.1-2	TCMID 055	16,518	16,518

Classification

Level	RankID	Taxon	Percentage
3	0.1:1.1	Clostridia	5.4%
3	0.1:1.2	Bacilli	53.8%
3	0.1:1.3	Erysipelotrichi	0%
3	0.1:1.4	Unclassified (2) Firmicutes	0.2%
3	0.1:2.1	Bacteroidia	0.2%
3	0.1:2.2	Flavobacteria	0.1%
3	0.1:3.1	Actinobacteria	38.2%
3	0.1:4.1	Betaproteobacteria	0.4%
3	0.1:4.2	Epsilonproteobacteria	0%
3	0.1:4.3	Gammaproteobacteria	0.3%
3	0.1:5.1	Fusobacteria	0.4%
3	0.1:6.1	Unclassified (2) TM7	0%
3	0.1:7.1	Spirochaetes	0%
3	0.1:8.1	Unclassified (1) Bacteria	0.9%
Total			22464

Rapid access to taxonomic classification overviews of selected samples

Study database queries

1. Select criteria
 N.B. Comparing numerical values is done without taking into account the units. E.g. a weight of 1 kg equals 1 gram.
 Field: Event, Operator: equals, Value: amoxicillin, Run query

2. Output type
 Choose the type of output: Samples

3. Run query

Taxonomic queries

1. Select criteria
 Search for samples with a specific taxon. You can compare the abundance of a taxon with a value (absolute or percentage) or with another taxon.
 Taxon: Firmicutes:Phylum, Operator: >, Value: 50, Run query

Query results: Classification search 3
 12 samples found with Akkermansia > 30

Sample name	Assay name	Study name	# sequences	# Akkermansia
Subject41_Feces_Amoxicillin_4w	16S(785-1175)	AntiResDev-1-Helperby	13367	2534
Subject41_Feces_Amoxicillin_0s	16S(785-1175)	AntiResDev-1-Helperby	11431	114
Subject41_Feces_Amoxicillin_9w	16S(785-1175)	AntiResDev-1-Helperby	11576	103
Subject41_Feces_Amoxicillin_1w	16S(785-1175)	AntiResDev-1-Helperby	11520	174
Subject14_Feces_Minocycline_0s	16S(785-1175)	AntiResDev-1-Helperby	8824	121
Subject5_Feces_Placebo_17w	16S(785-1175)	AntiResDev-1-Helperby	9734	56

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.

