Co-inoculation of bacterial strains and data analysis
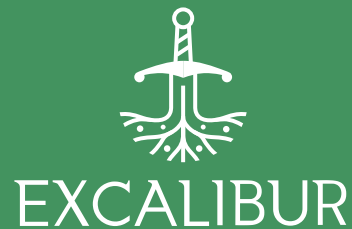
April 19th, 2023

Stefano Mocali, CREA

E-mail: stefano.mocali@crea.gov.it

EXCALIBUR Training Series:

Addressing microbial metabolic profile by means of Phenotype Microarray technology (BIOLOG)

EXCALIBUR

The genomic era

Addressing the functionality of genomes is one of the most important and challenging tasks of today's biology. This, together with the impressive improvements of sequencing tecnology and bioinformatic tools, have provided a huge enhancement of the number of genomic projects over the world.

However, even though such techniques are of great help, **the final expression of the genomic information is represented by the phenotype**

- Metabolic networks reconstruction
- From genomes to metabolomes
- High throughput genomics/metabolomics

www. genomesonline.com
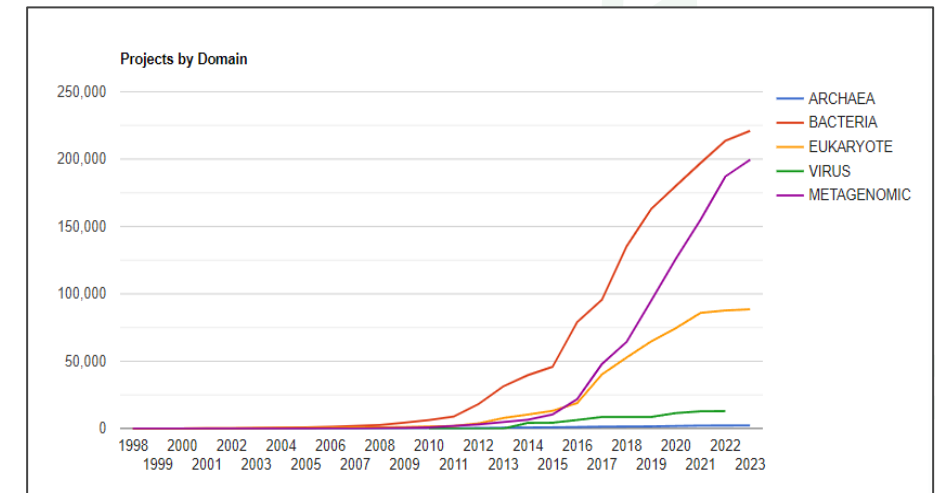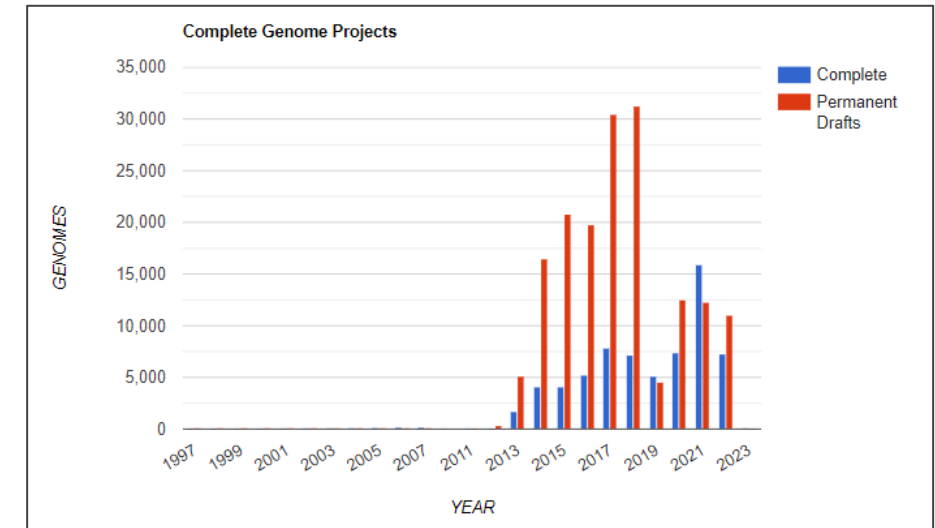
The phenomic era

The OmniLog™ platform, uses 96-well PM plates, composed of more than 2000 temperature-controlled conditions.

In each well the metabolism of the cell can be monitored using respiration as a reporter system

- Up to 20 plates (1920 wells)
- Up to 9 compound categories
- Time series data
- Many compounds on KEGG DB
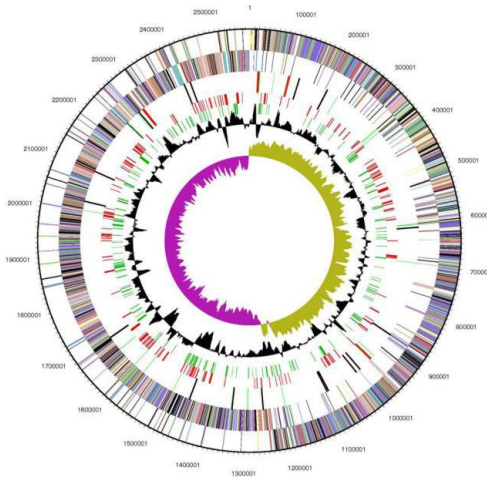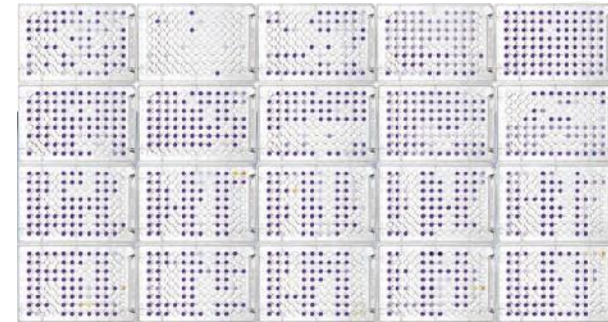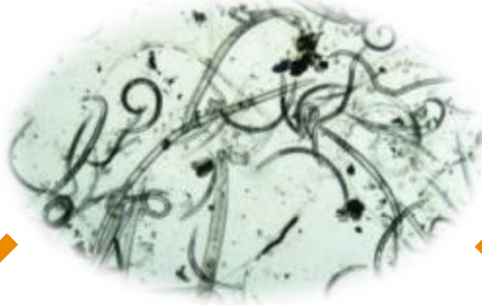- High throughput phenomics
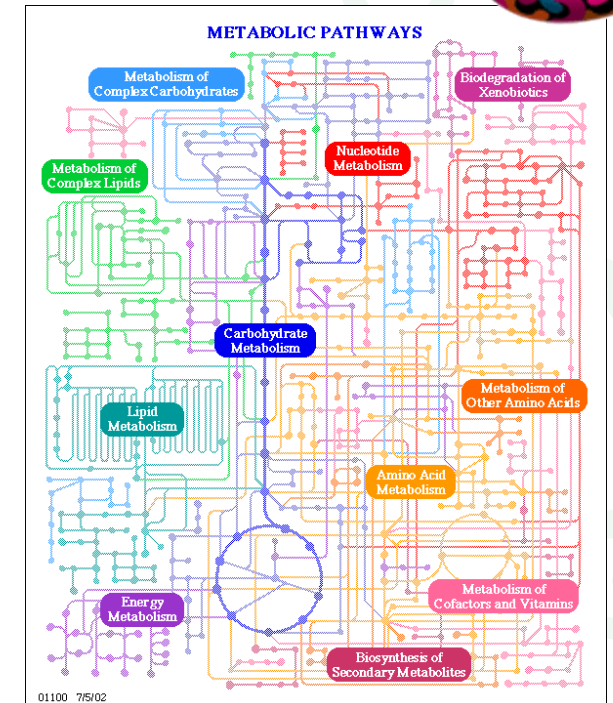
The 'omic' era

Genome and phenome data can be achieved from soil bacteria

Genome data analysis
- Genome map to KEGG
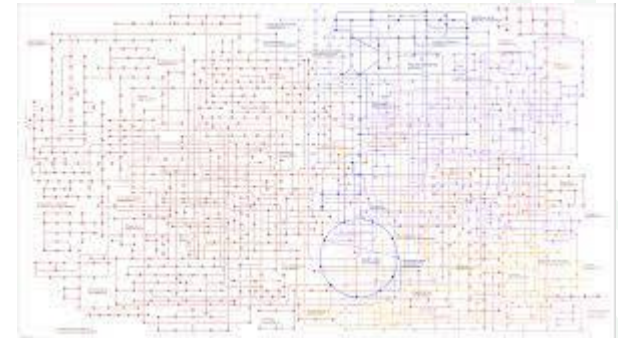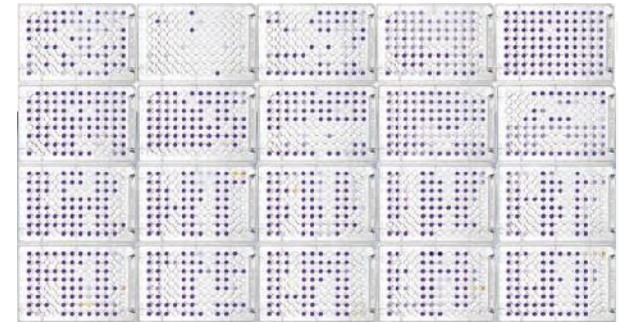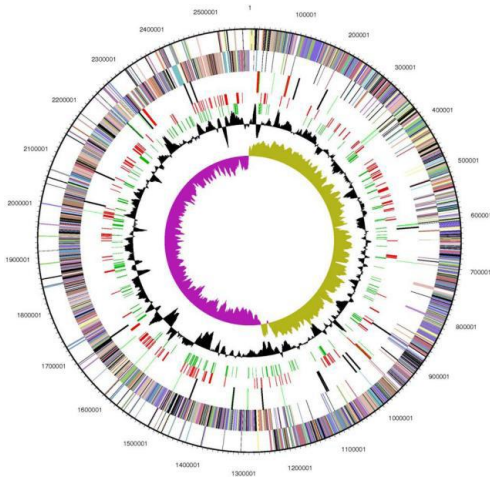- Pangenome prediction
- core
- accessory
- unique

Phenome data analysis
- Metabolic activity parameters
- Replica management
- Clear comparisons
- Clear visualizations
- Compounds map to KEGG

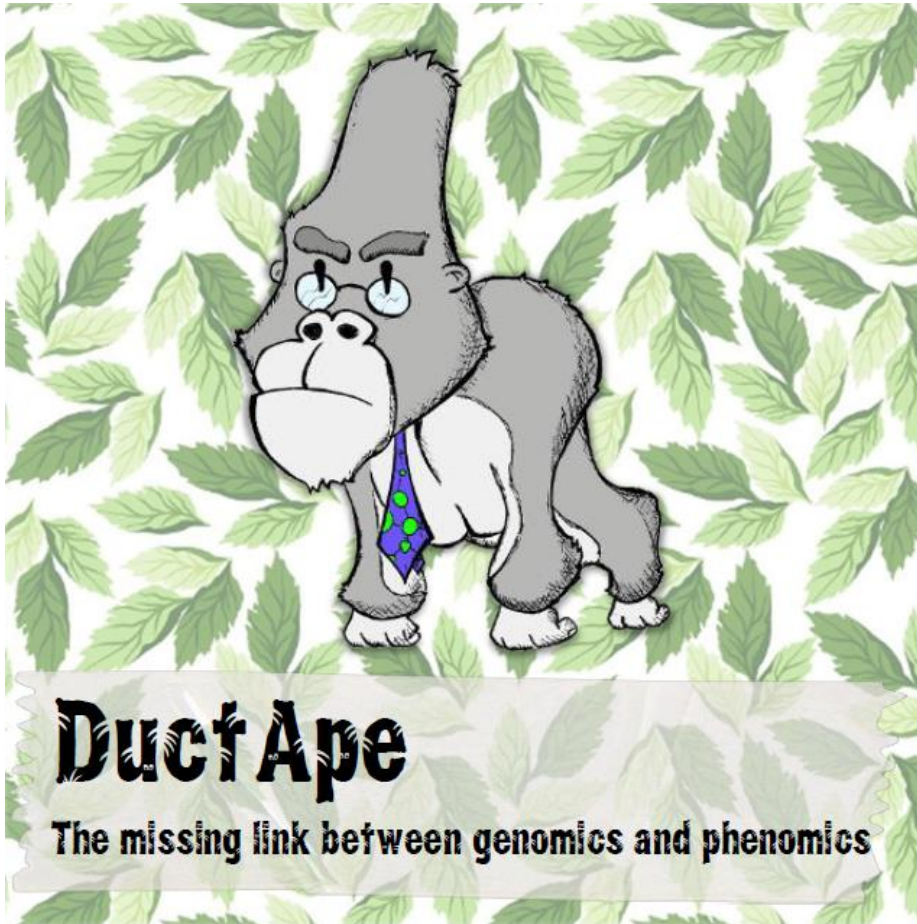**But…which genes may be responsible for such phenotypic variations?**

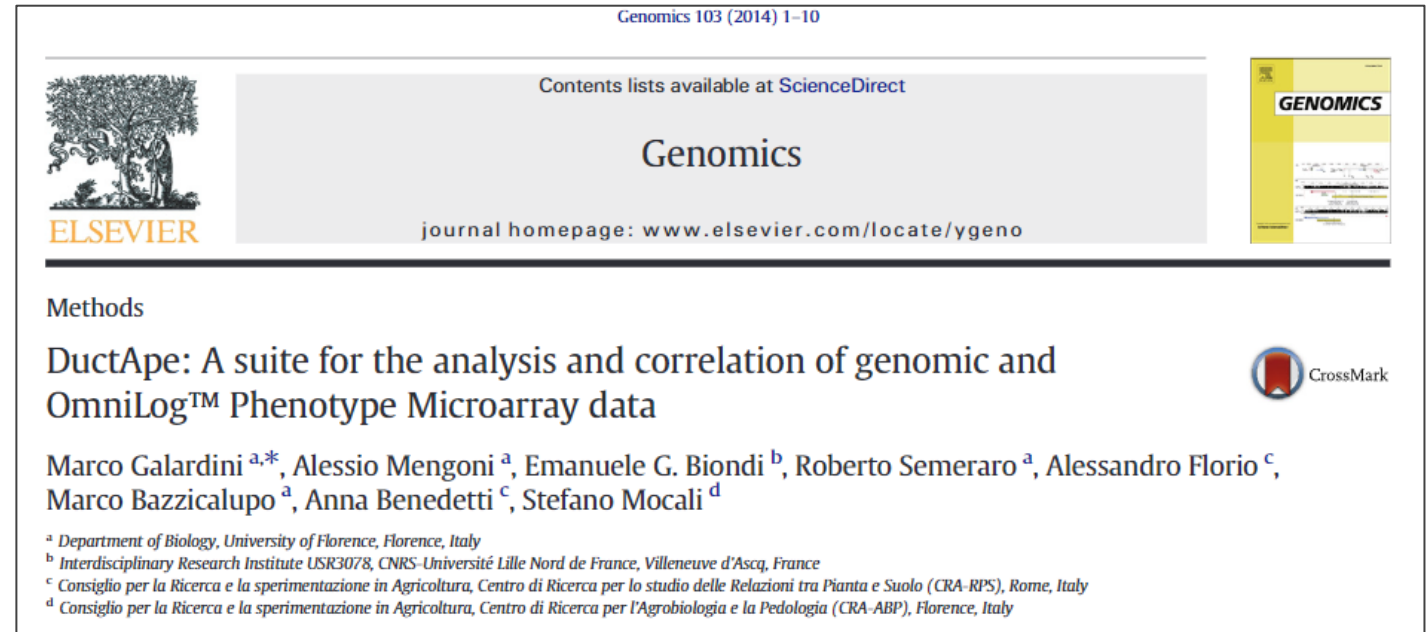**How to combine genomic and phenomic data?**

- All data in a single metabolic map
- Genetic basis for phenotypic differences

Ductape software

http://combogenomics.github.com/DuctApe

Methods

DuctApe: A suite for the analysis and correlation of genomic and OmniLog™ Phenotype Microarray data

Marco Galardini [a,*], Alessio Mengoni [a], Emanuele G. Biondi [b], Roberto Semeraro [a], Alessandro Florio [c], Marco Bazzicalupo [a], Anna Benedetti [c], Stefano Mocali [d]

[a] Department of Biology, University of Florence, Florence, Italy
[b] Interdisciplinary Research Institute USR3078, CNRS-Université Lille Nord de France, Villeneuve d'Ascq, France
[c] Consiglio per la Ricerca e la sperimentazione in Agricoltura, Centro di Ricerca per lo studio delle Relazioni tra Pianta e Suolo (CRA-RPS), Rome, Italy
[d] Consiglio per la Ricerca e la sperimentazione in Agricoltura, Centro di Ricerca per l'Agrobiologia e la Pedologia (CRA-ABP), Florence, Italy
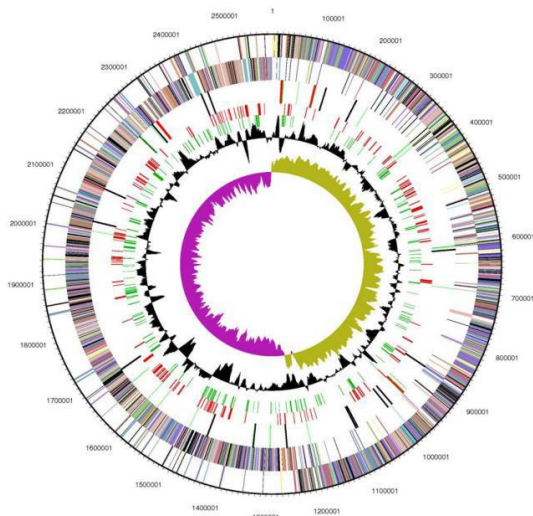
- Reconstruct metabolic maps from proteome sequences (derived from genomes)
- Interpret Phenotype Microarray (BIOLOG) curves
- Provide testable hypothesis for genotype-to-phenotype links
- Available as a command line tool (not in active development) for Unix systems (Linux/macOS) (on Windows you can use a VM)

Ductape software
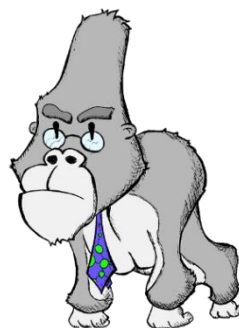
dape
- Generation of combined KEGG metabolic maps
- Metabolic network analysis
- Metabolic hotspots prediction
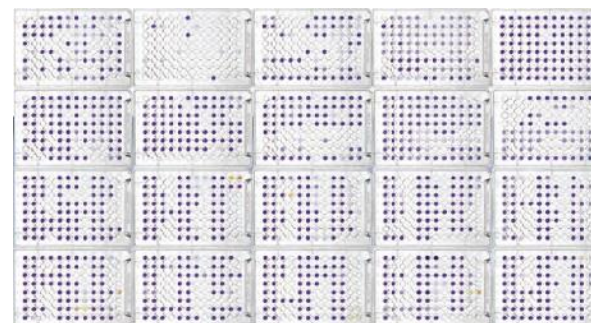
dgenome
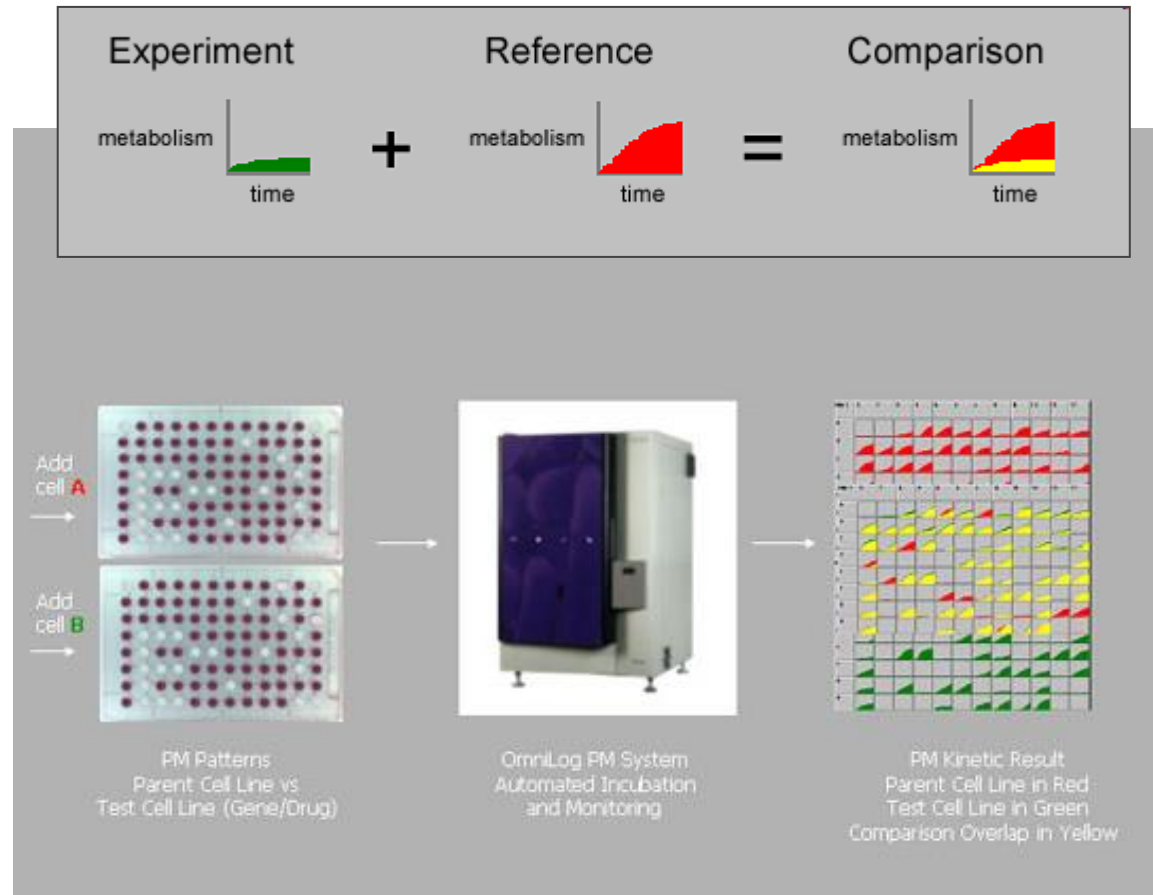- Genes are mapped to KEGG database
- PanGenome prediction (Blast-BBH)

dphenome
- Phenotype microarray data handling
- Classification of metabolic activity
- Compounds are mapped to KEGG DB

Comparison of phenotypes

To compare the phenotypes of two cell lines or strains, one is recorded as a **RED** tracing and one as a **GREEN** tracing. These graphs can then be overlaid by the bioinformatic software to detect differences.

Areas of overlap (i.e. no change) are colored **YELLOW**, whereas differences are highlighted as patches of red or green
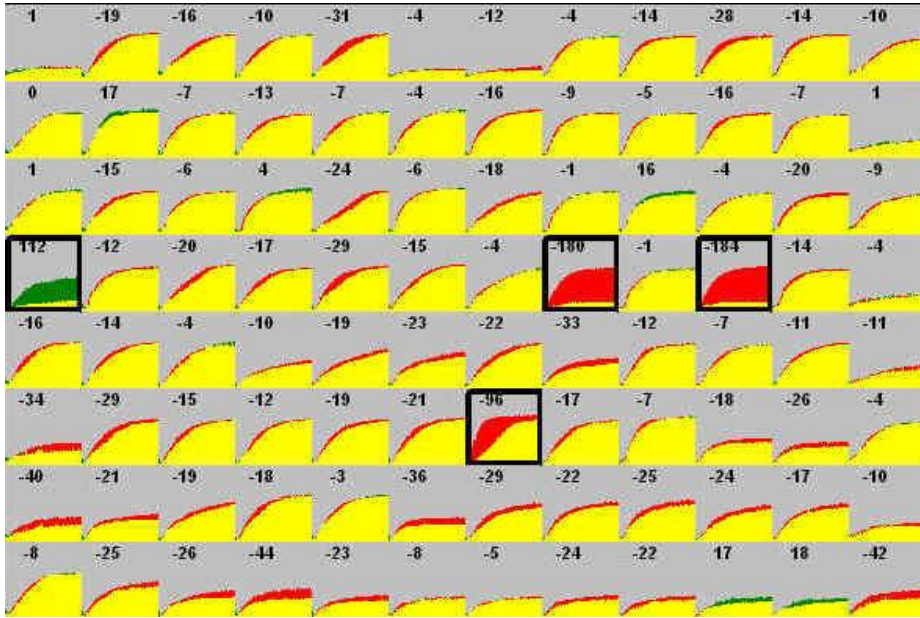
- Multiple strain comparison
- How to discriminate different activities?
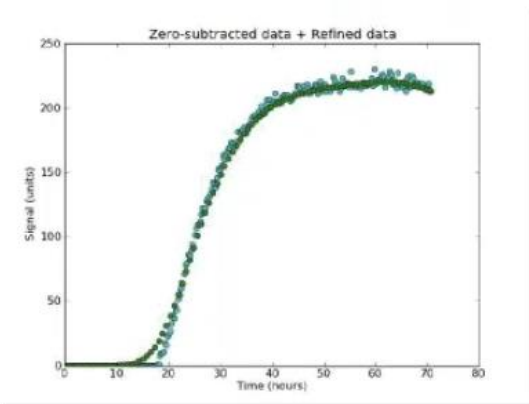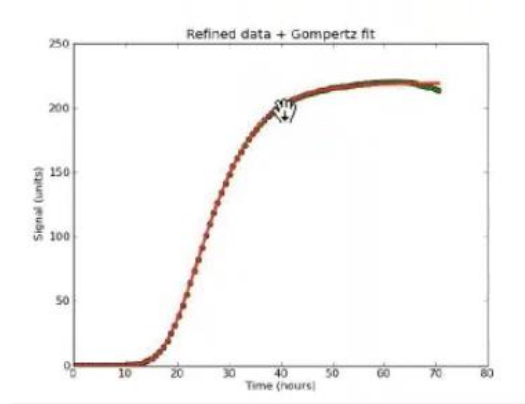- A single, summarized value is needed
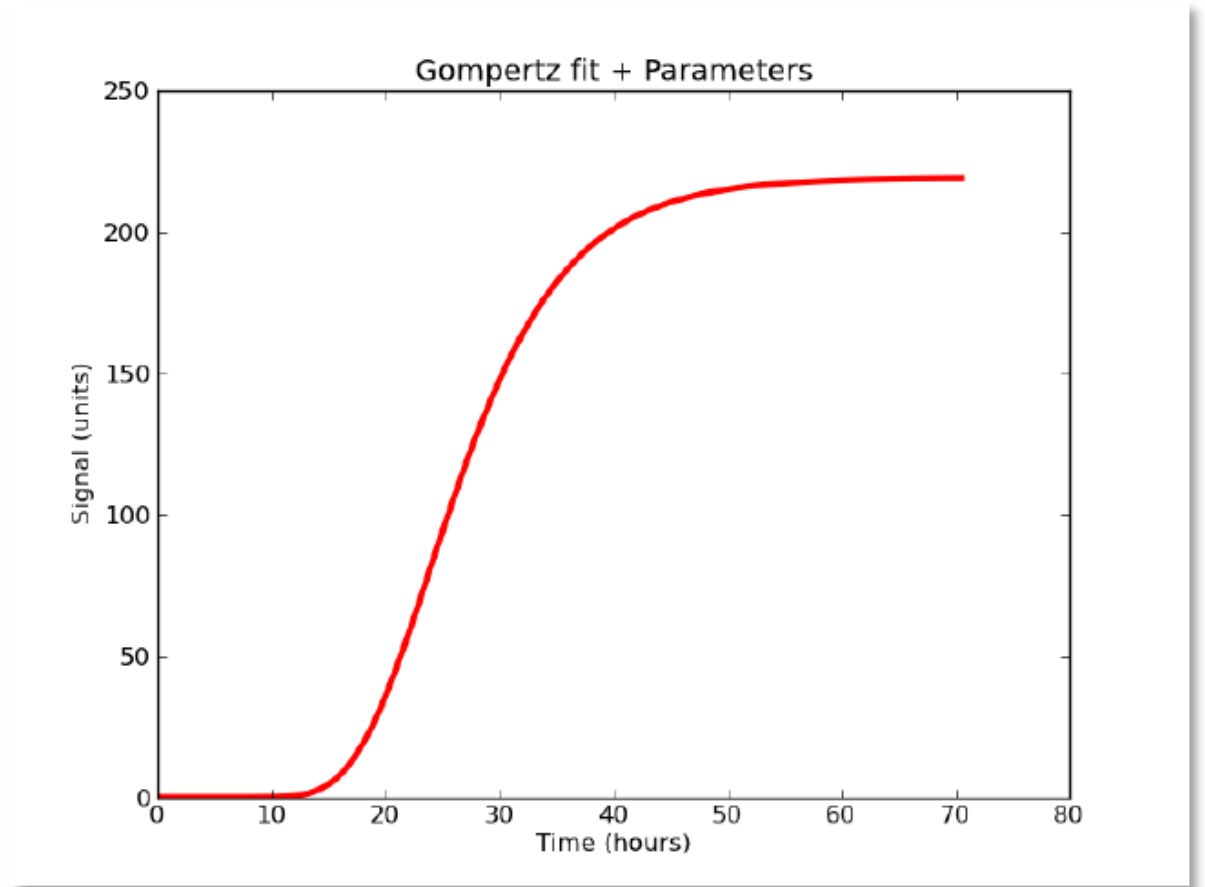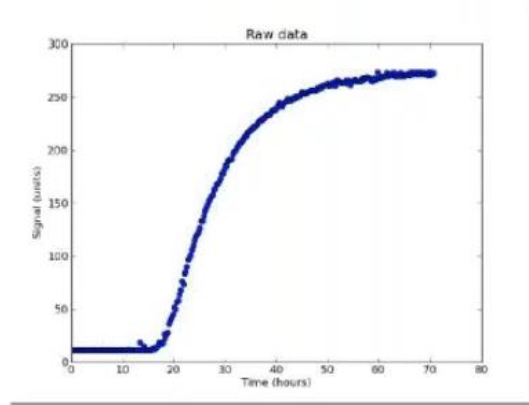
1. Parsing
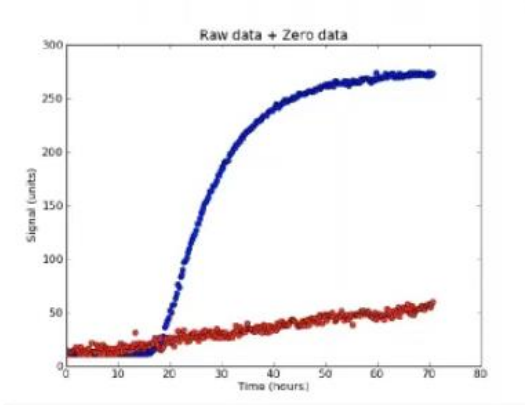
2. Control signal subtraction (optional)
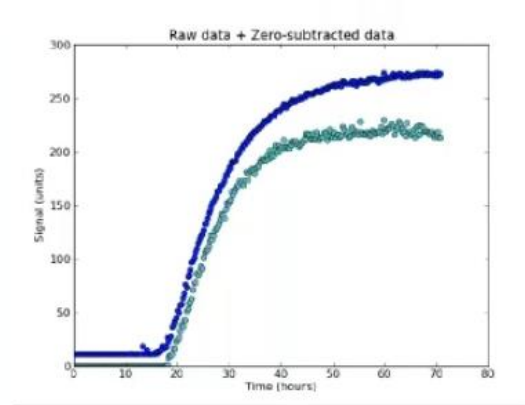
3. Signal refinement

4. Sigmoid fit

Modeling of the Bacterial Growth Curve

M. H. ZWIETERING,* I. JONGENBURGER, F. M. ROMBOUTS, AND K. VAN 'T RIET
Department of Food Science, Agricultural University Wageningen, P.O. Box 8129,
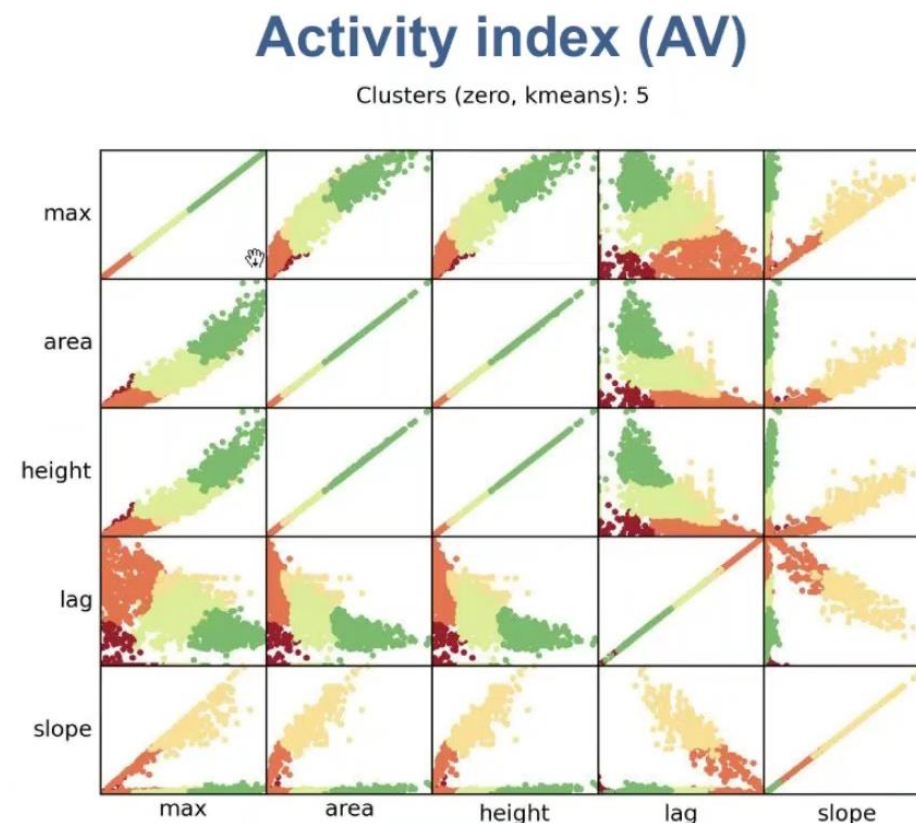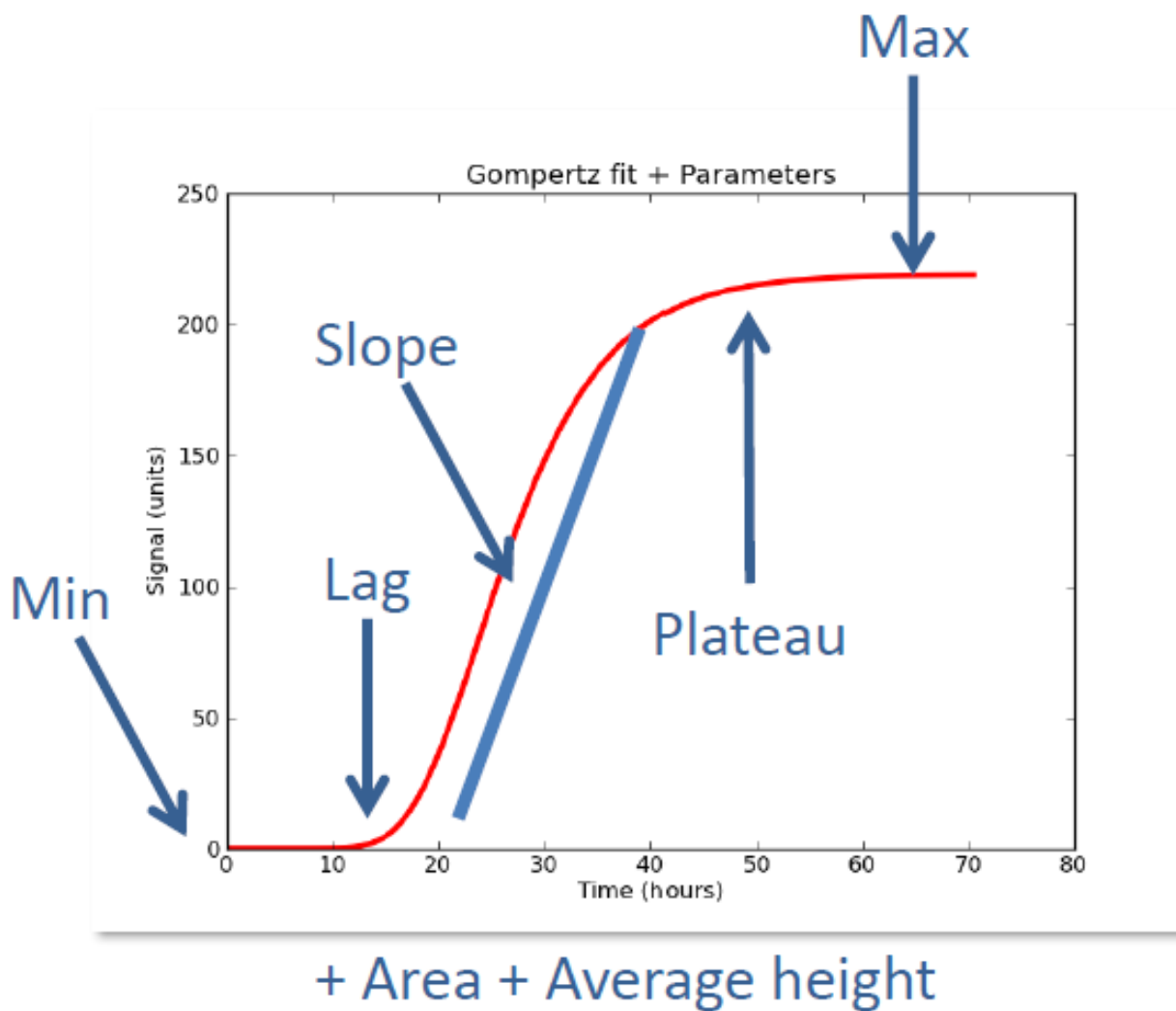6700 EV Wageningen, The Netherlands

Received 5 January 1990/Accepted 4 April 1990

Model	Equation
Logistic	$y = \dfrac{a}{[1 + \exp(b - cx)]}$
Gompertz	$y = a \cdot \exp[-\exp(b - cx)]$
Richards	$y = a\{1 + v \cdot \exp[k(\tau - x)]\}^{(-1/v)}$

10

Activity index (AV)

Activity index (AV)

Gompertz fit + Parameters

Max

Slope

Min

Lag

Plateau

+ Area + Average height

Activity index (AV)

Clusters (zero, kmeans): 5

max

area

height

lag

slope

max   area   height   lag   slope
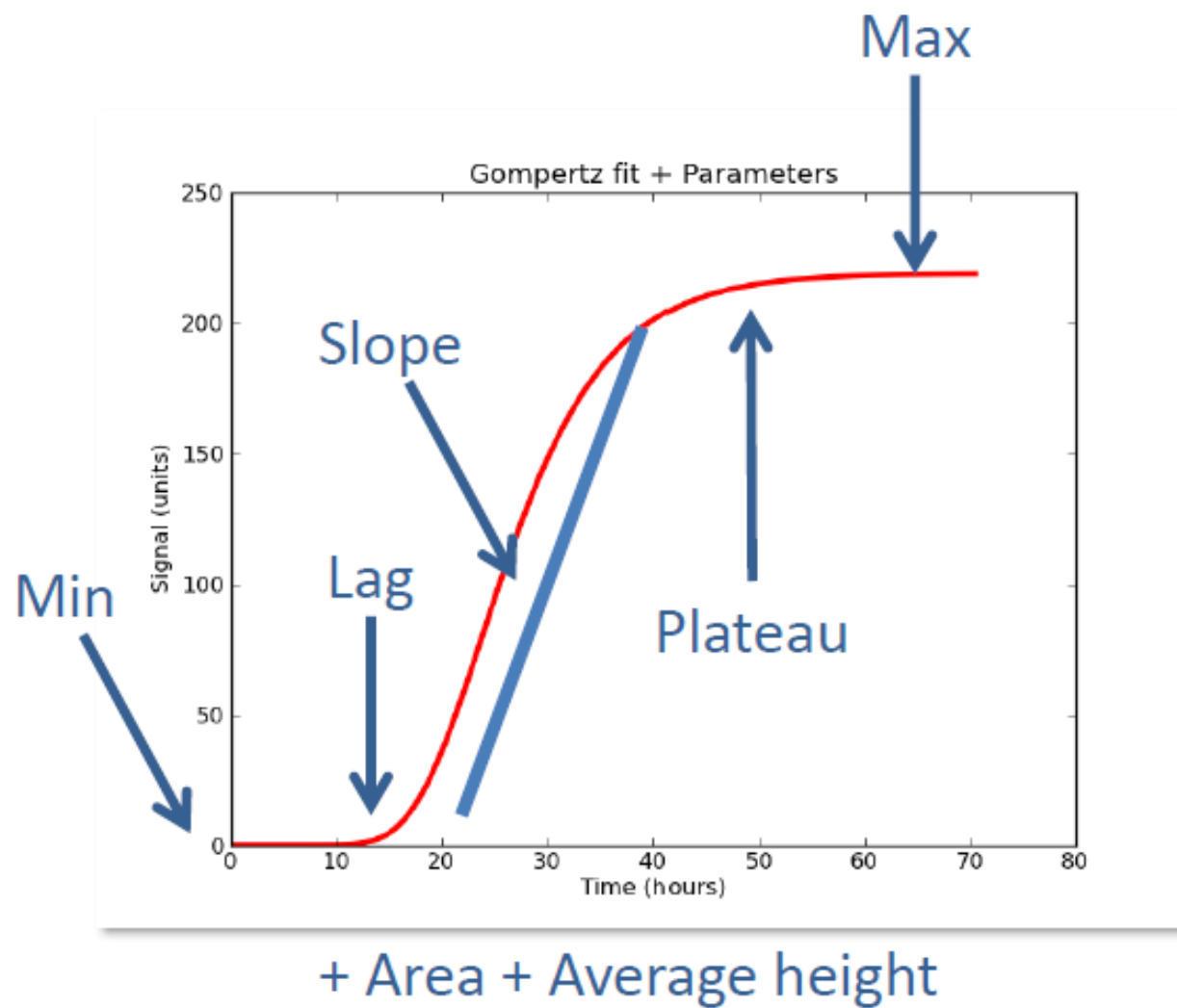
K-means clustering on 5 parameters
(5 clusters in this example, tunable by the user)

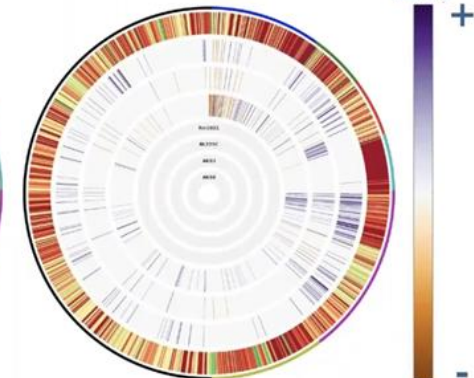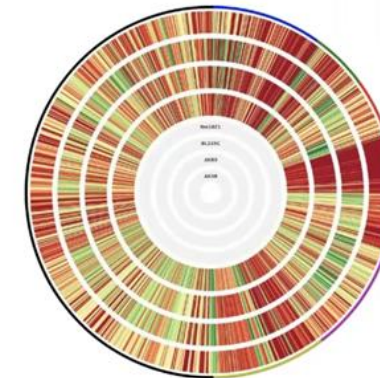**Fast:** from raw .csv files to AV in less than 5 minutes

EXCALIBUR

Plates AV heatmaps: phenotypic variability at a glance
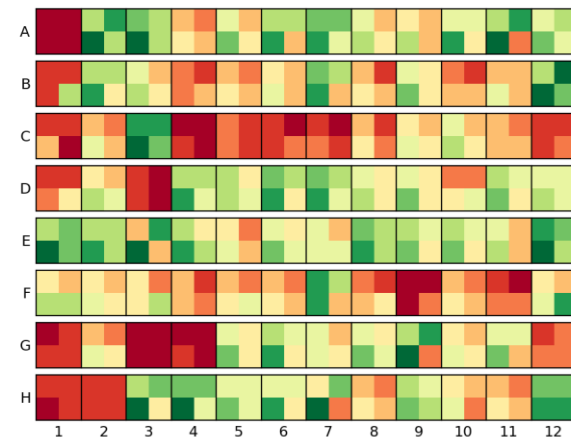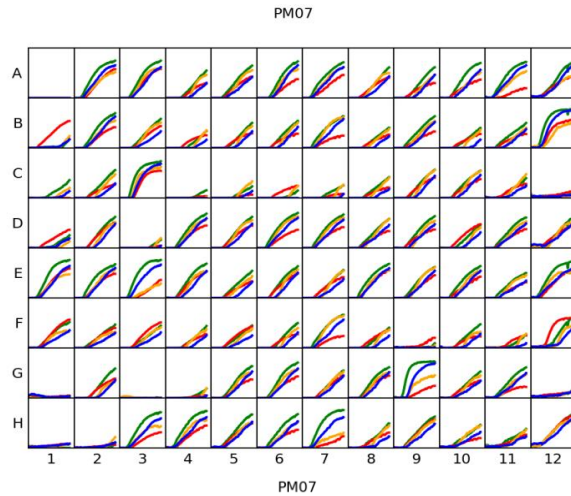
AV boxplots: overall strains comparison
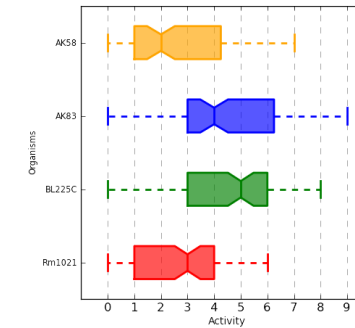
AV rings: overall strains comparison
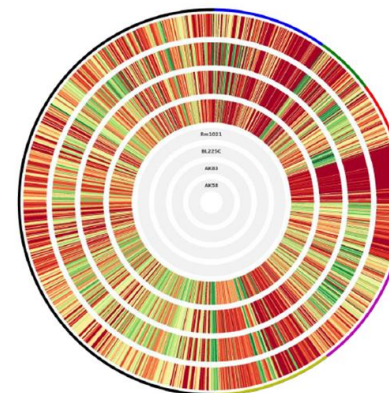
dape: single genome metabolic network

Nitrogen metabolism

Interactive metabolic maps (as web pages)
- Reactions copy number
- Compounds AV

EXCALIBUR

Highlights on differences of specific substrates/ conditions and the related metabolic pathways

Bacillus subtilis vs Paenibacillus polymixa (PM1,2,3 e 4)

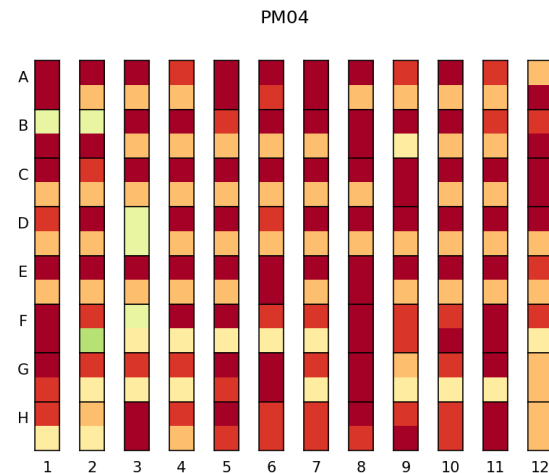# Case study n.1
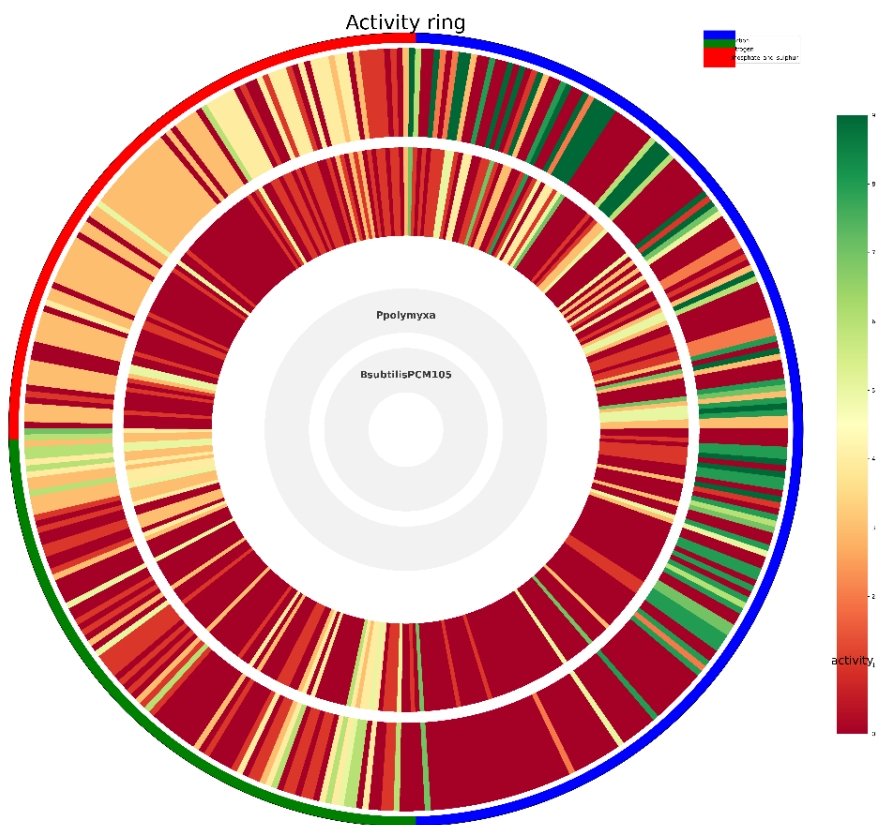
Bacillus subtilis vs Paenibacillus polymixa (PM1,2,3 e 4)

Bacillus subtilis vs Paenibacillus polymixa (PM1,2,3 e 4)

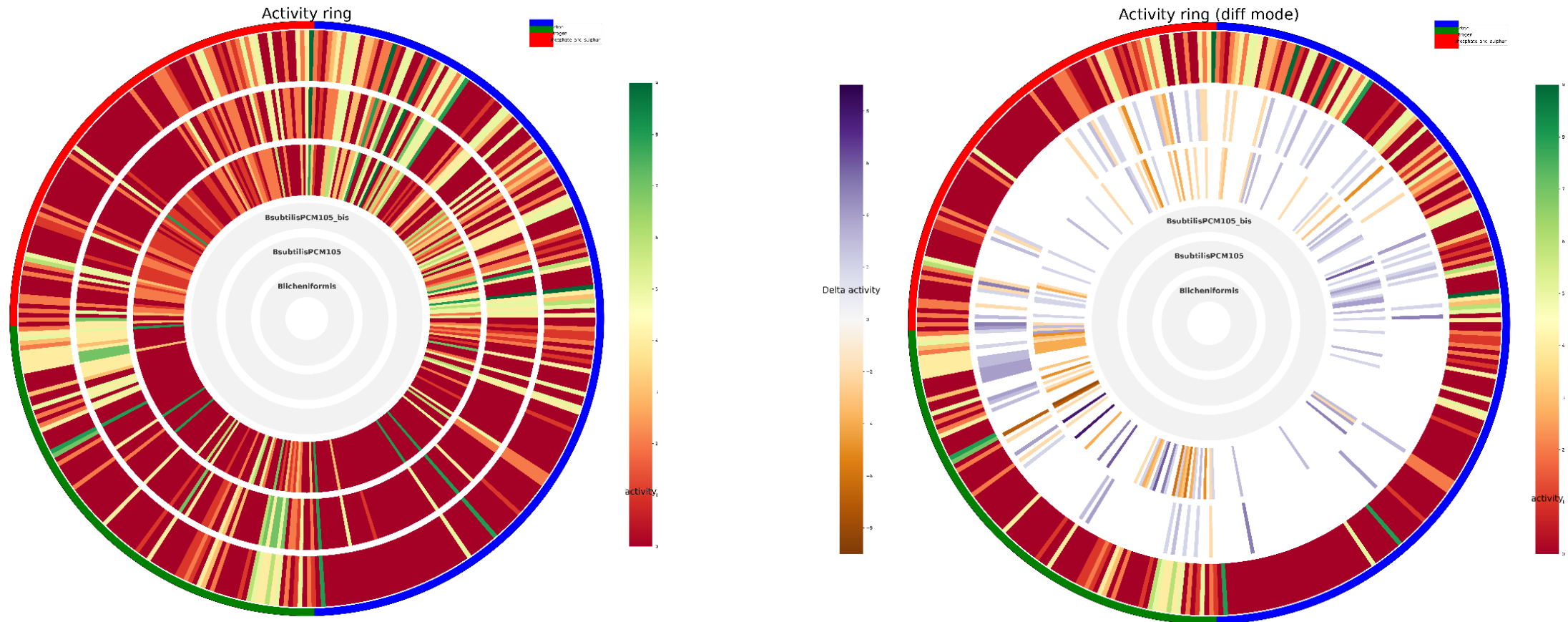| Pathway	Name	Category	BsubtilisPCM105		Ppolymyxa	
			Reactions	meanAV	Reactions	meanAV
path:map01240	Biosynthesis of cofactors	nitrogen	23,00	1,82	35,00	3,36
path:map01240	Biosynthesis of cofactors	carbon	23,00	1,75	35,00	1,88
path:map01240	Biosynthesis of cofactors	phosphate_and_sulphur	23,00	0,33	35,00	2,83
path:map00230	Purine metabolism	carbon	21,00	2,29	20,00	3,00
path:map00230	Purine metabolism	nitrogen	21,00	2,25	20,00	2,17
path:map00230	Purine metabolism	phosphate_and_sulphur	21,00	0,88	20,00	2,75
path:map01232	Nucleotide metabolism	carbon	15,00	3,50	16,00	5,50
path:map01232	Nucleotide metabolism	nitrogen	15,00	1,29	16,00	1,21
path:map01232	Nucleotide metabolism	phosphate_and_sulphur	15,00	0,00	16,00	3,00
path:map00061	Fatty acid biosynthesis	carbon	15,00	0,00	15,00	0,00
path:map00061	Fatty acid biosynthesis	phosphate_and_sulphur	15,00	NA	15,00	NA
path:map00061	Fatty acid biosynthesis	nitrogen	15,00	NA	15,00	NA
path:map00640	Propanoate metabolism	carbon	9,00	0,20	6,00	0,00
path:map00640	Propanoate metabolism	phosphate_and_sulphur	9,00	NA	6,00	NA
path:map00640	Propanoate metabolism	nitrogen	9,00	NA	6,00	NA
path:map00620	Pyruvate metabolism	carbon	8,00	2,83	6,00	3,67
path:map00620	Pyruvate metabolism	phosphate_and_sulphur	8,00	0,00	6,00	3,00
path:map00620	Pyruvate metabolism	nitrogen	8,00	NA	6,00	NA
path:map00240	Pyrimidine metabolism	carbon	7,00	3,25	10,00	3,75
path:map00330	Arginine and proline metabolism	nitrogen	7,00	3,11	4,00	1,67

KEGG Database: Nitrogen metabolism (map1240)
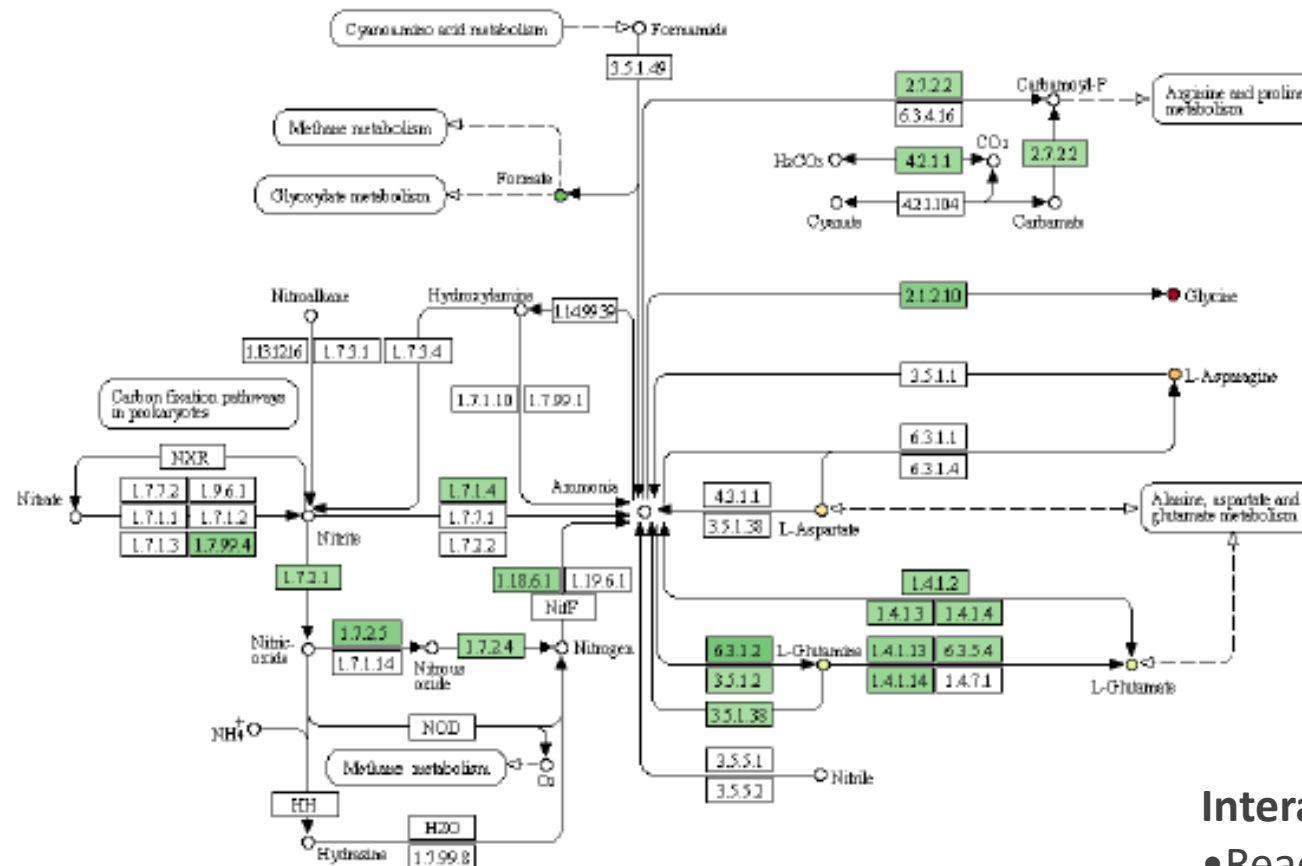
Interactive metabolic maps (as web pages)
- Reactions copy number
- Compounds AV

Bacillus subtilis and B.licheniformis (PM 1,2,3 e 4), single strains and co-inoculated
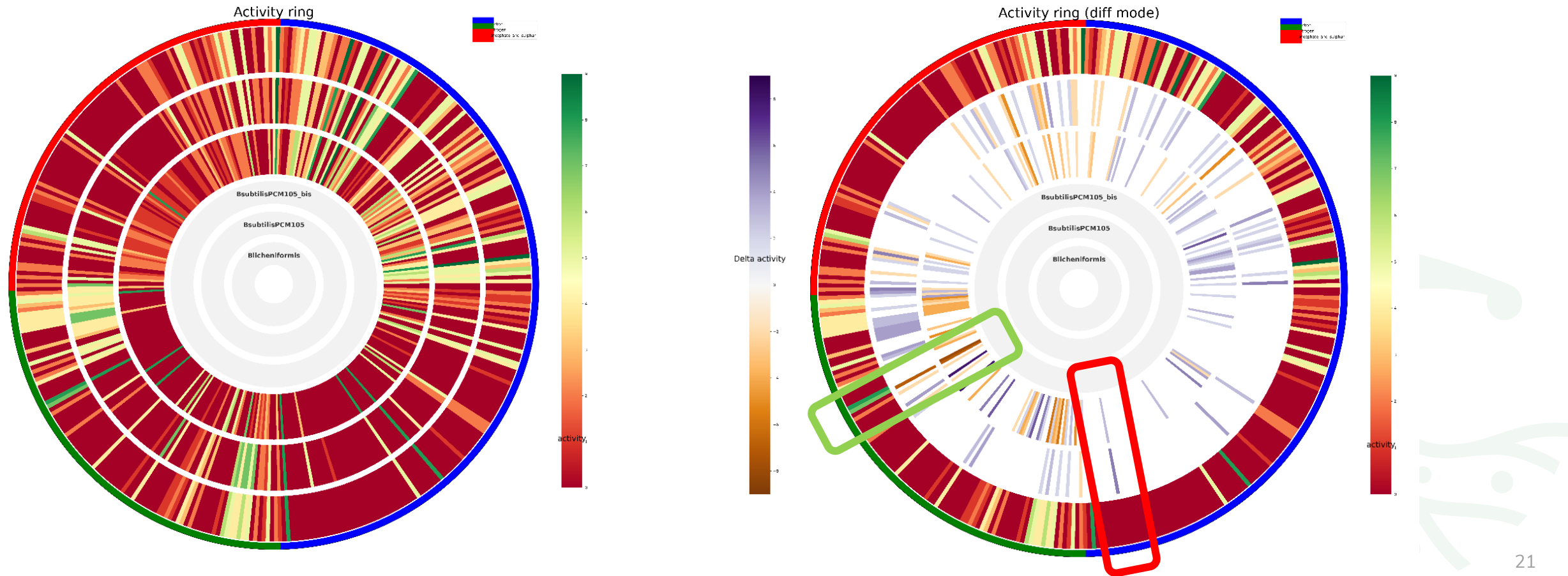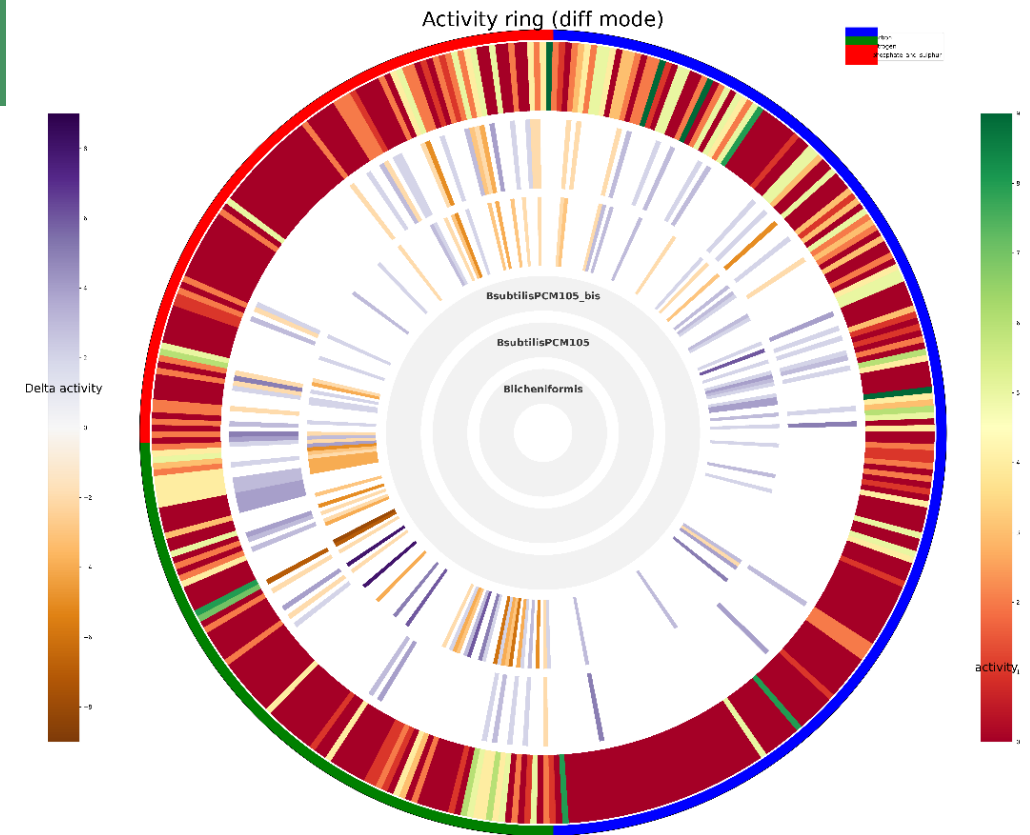
Bacillus subtilis and B.licheniformis (PM 1,2,3 e 4), single strains and co-inoculated

Case study n.2

Bacillus subtilis and B.licheniformis (PM 1,2,3 e 4)

From AV data (diff mode) you can easily identify the substrates that are metabolized (or not) in the co-inoculum as compared to the single strains

Activity ring (diff mode)

category	moa	co_id	Blicheniformis	BsubtilisPCM105	Mean activity	CoInoculum	Difference	ass(diff)
nitrogen	N-Source, other	C00242	0	0	0	7	7	7
phosphate & sulphur	S-Source, organic		0	0	0	4	4	4
carbon	C-Source, carbohydrate	C00124	0	0	0	2	2	2
carbon	C-Source, carbohydrate	C01019	0	0	0	2	2	2
carbon	C-Source, alcohol	C00466	0	0	0	2	2	2
nitrogen	N-Source, other	C03570	0	0	0	2	2	2
phosphate & sulphur	P-Source, organic		0	0	0	2	2	2
phosphate & sulphur	S-Source, organic	C06809	0	0	0	2	2	2
phosphate & sulphur	S-Source, organic	C11145	0	0	0	2	2	2

GC/MS measurements (Volatile Organic Compounds)

GC/MS measurements (Volatile Organic Compounds)

Conclusions

- The BIOLOG technology strongly support the genomics analysis in a complementary way

- Ductape software includes the tools to analyze genomes and phenomes as well as linking them with specific metabolic pathways

- More specifically, the definition of a AV and the visualization options included, offer the possibility to successfully and easily compare multiple strains together

- More specific details on data analysis will be presented in the next presentation

EXCALIBUR

Thanks for you attention