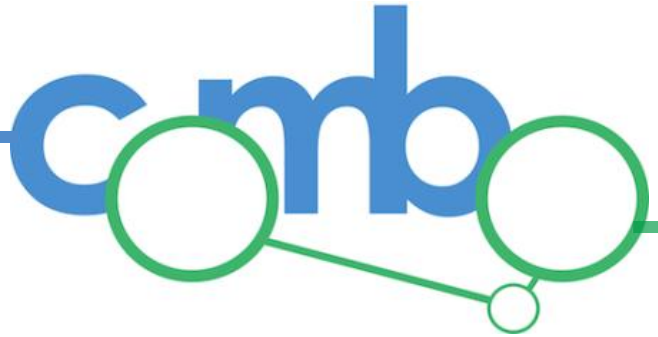
Marco Galardini
(@mgalactus)



DuctApe

a tool for the analysis and correlation of genomic and high throughput phenotypic Biolog data

University of Florence
Microbial genetics lab
Florence computational biology group

04/03/2013

Who we are

florence
computational biology group

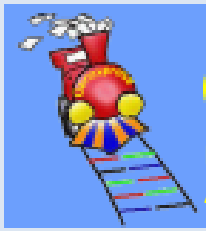
@combogenomics
combo.unifi@gmail.com
http://www.unifi.it/dbefcb

- Three bioinformatics groups from Unifi
- Est. 2011
- Microbiology (clinical, agronomical, ecological)
- Biological sequences information analysis
- Bioinformatics softwares development

- Italian Agricultural Research Council
- Soil and agricultural microbiology

CRA
CONSIGLIO PER LA RICERCA
E LA SPERIMENTAZIONE
IN AGRICOLTURA

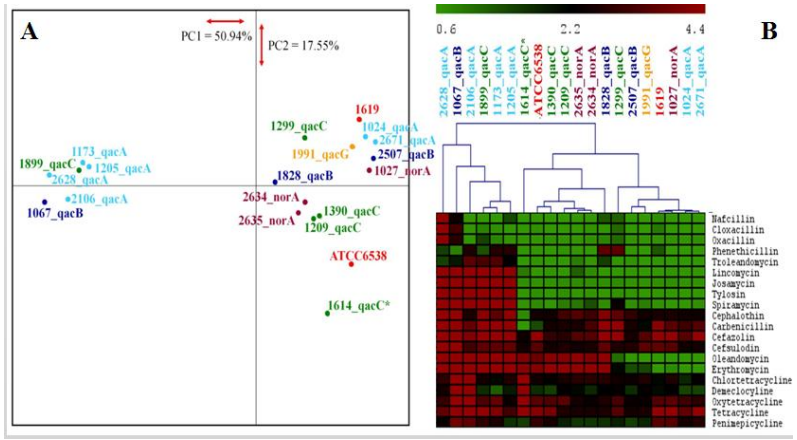
Other collaborations

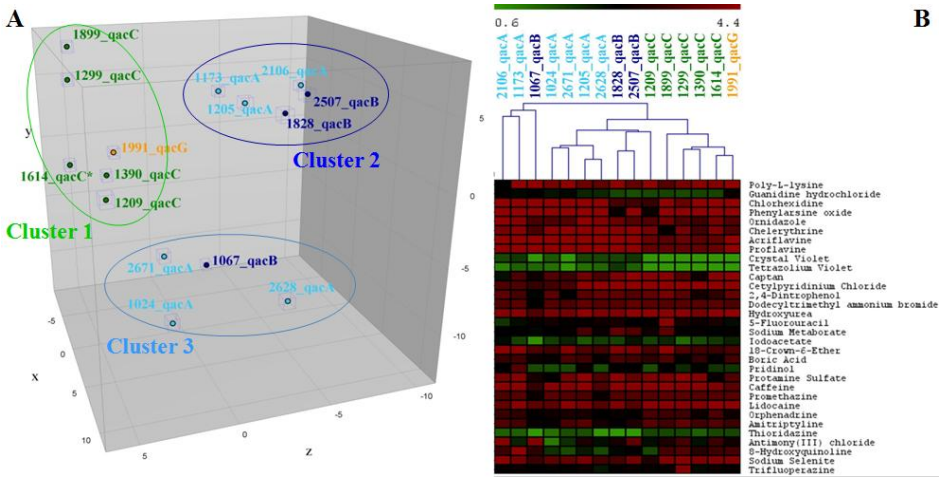


Genexpress

Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente

- Bacterial genomics and phenomics
- Phenotypic assays on chemical sensitivities





Florence Conference on Phenotype MicroArray Analysis of Microorganisms

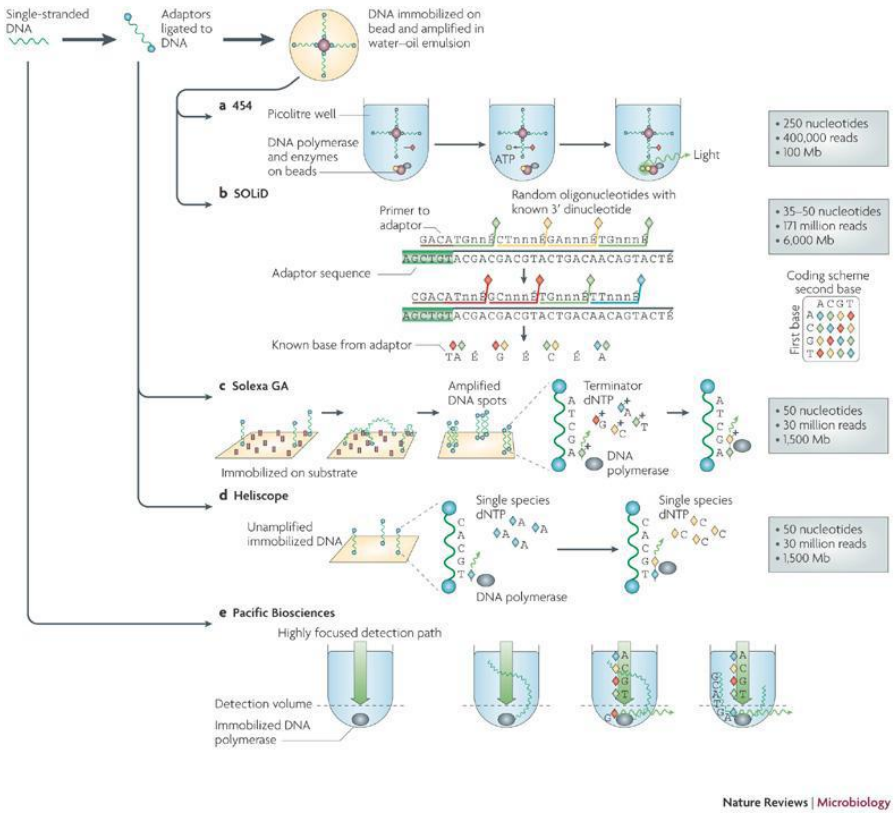
The Environment, Agriculture, and Human Health

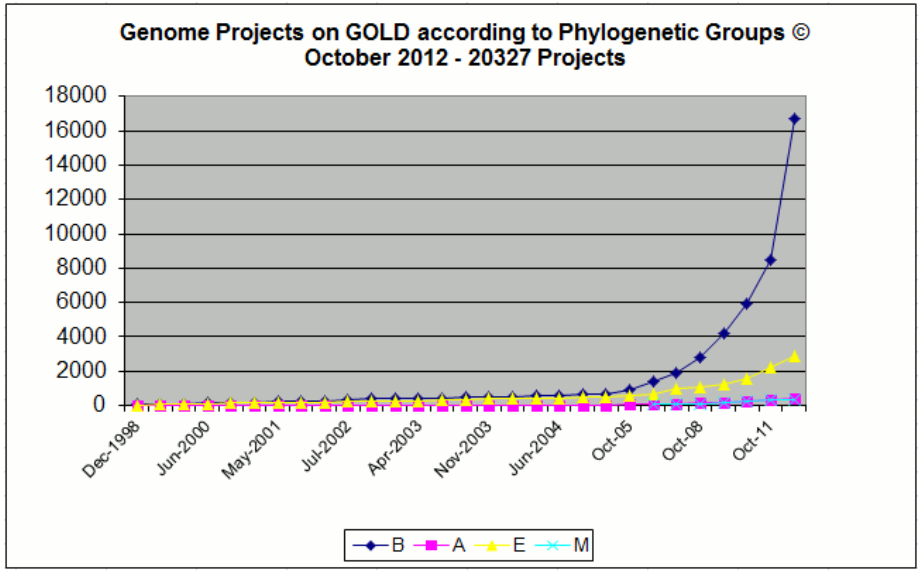
Polo Scientifico di Sesto Fiorentino, Florence, Italy, March 19-21, 2008

The wishing well

The genomics and phenomics era

The genomics era





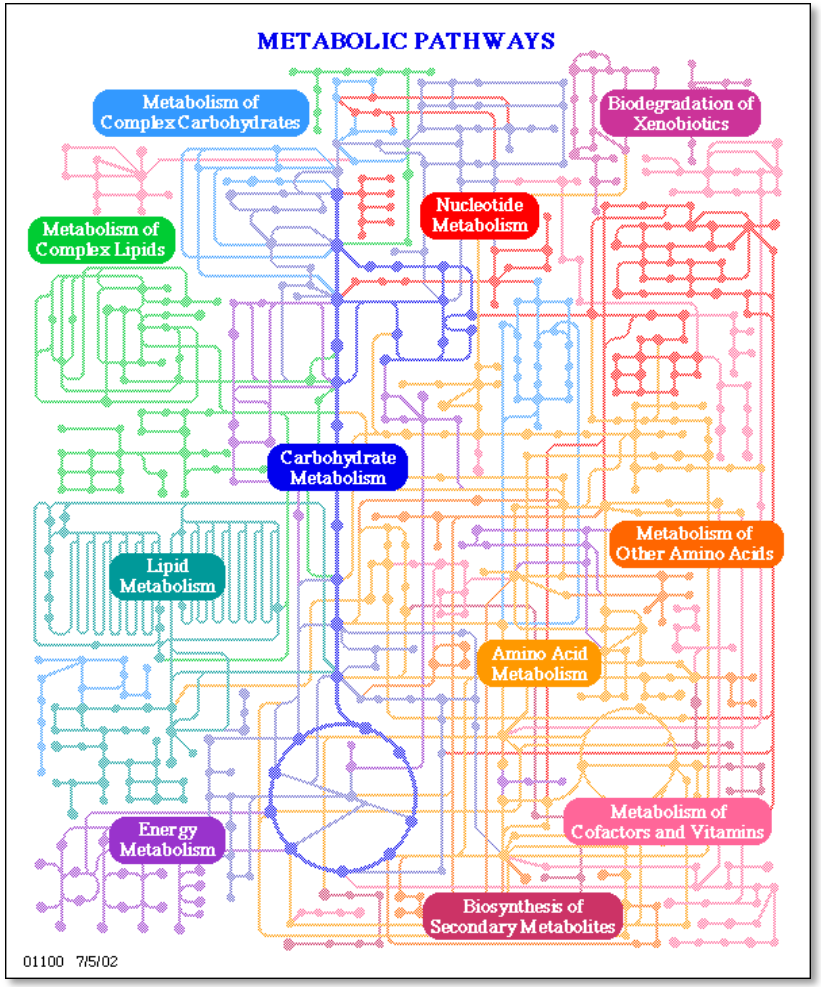
MacLean et al., 2009

genomesonline.com

The genomics era

- Metabolic networks reconstruction
- From genomes to metabolomes

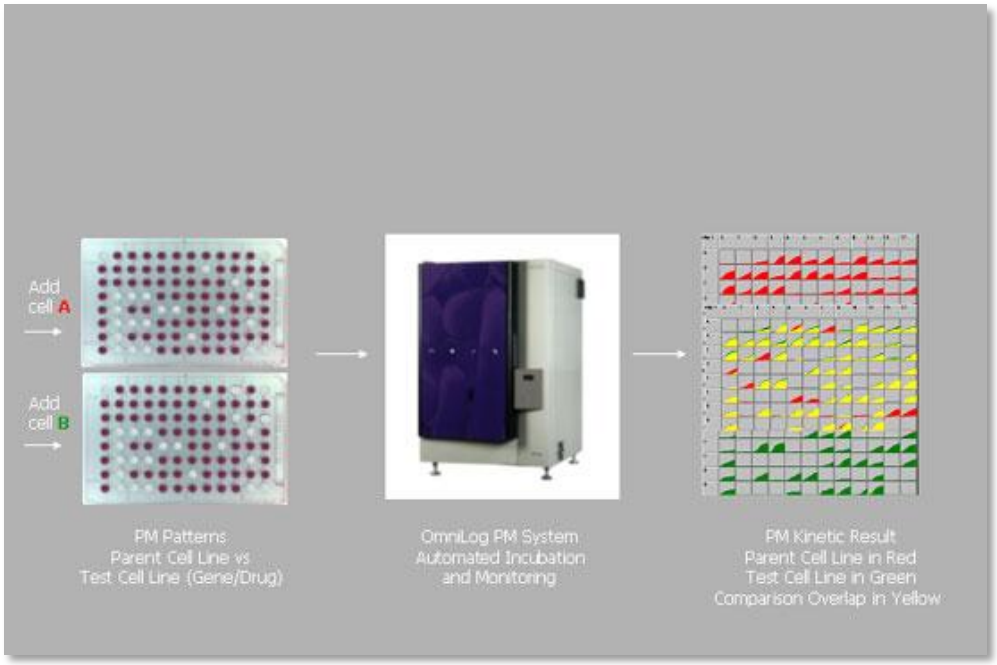
- **High throughput genomics/metabolomics**

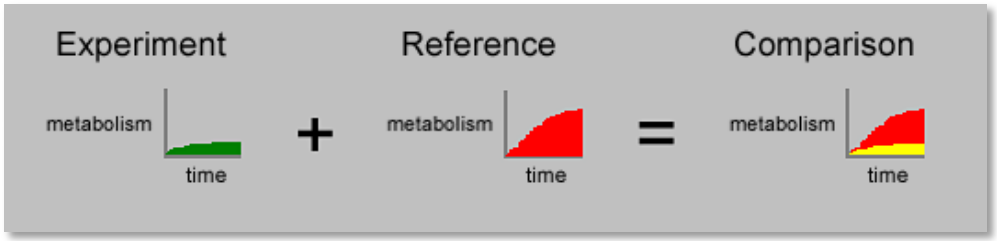


http://www.genome.jp/kegg/

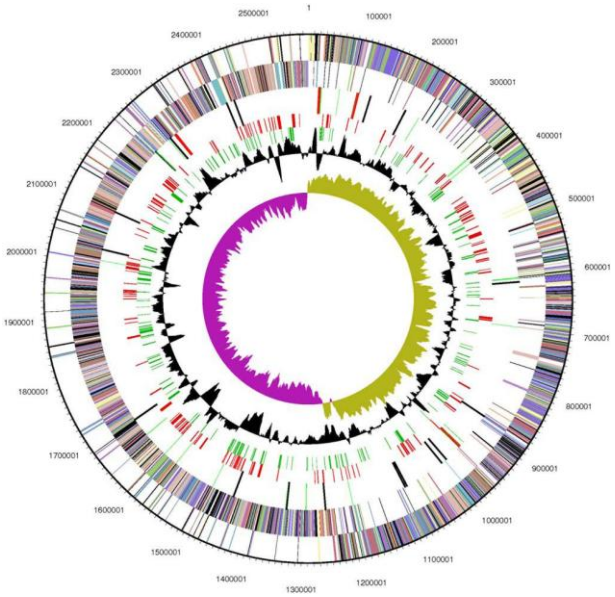
The phenomics era

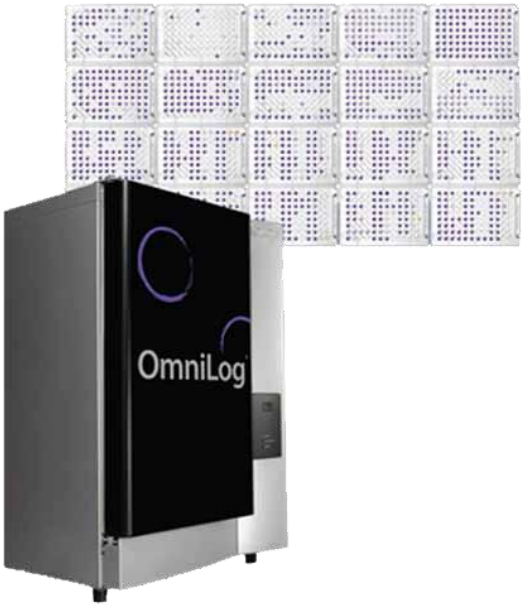
- Many compounds on KEGG DB
- **High throughput phenomics**





Introduction





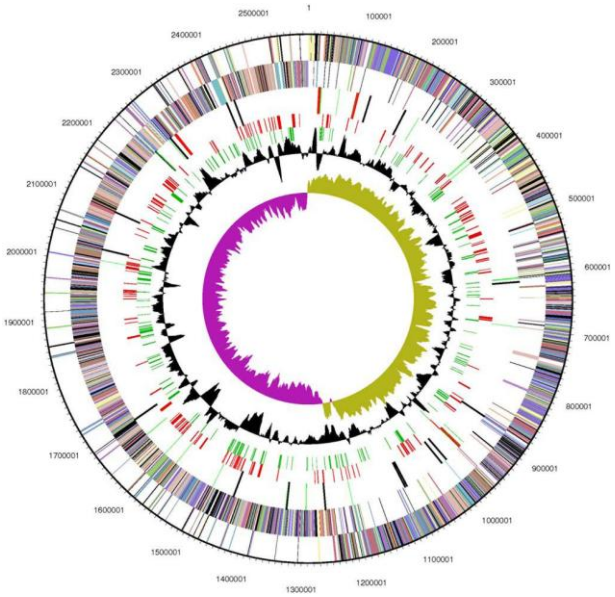
Genome data analysis

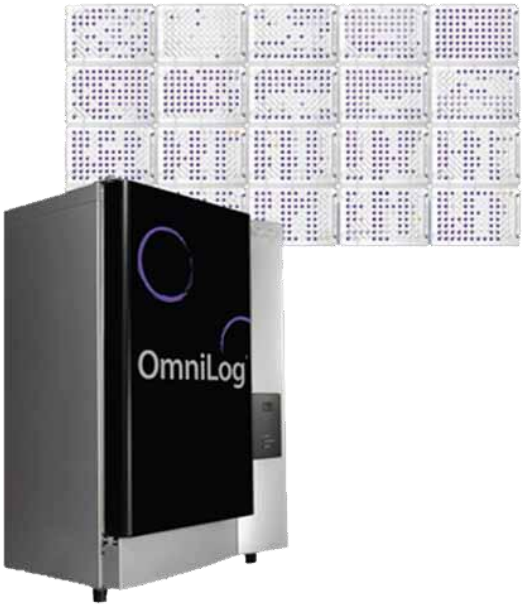
- Genome map to KEGG
- Pangenome prediction
  - core
  - accessory
  - unique

Phenome data analysis

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

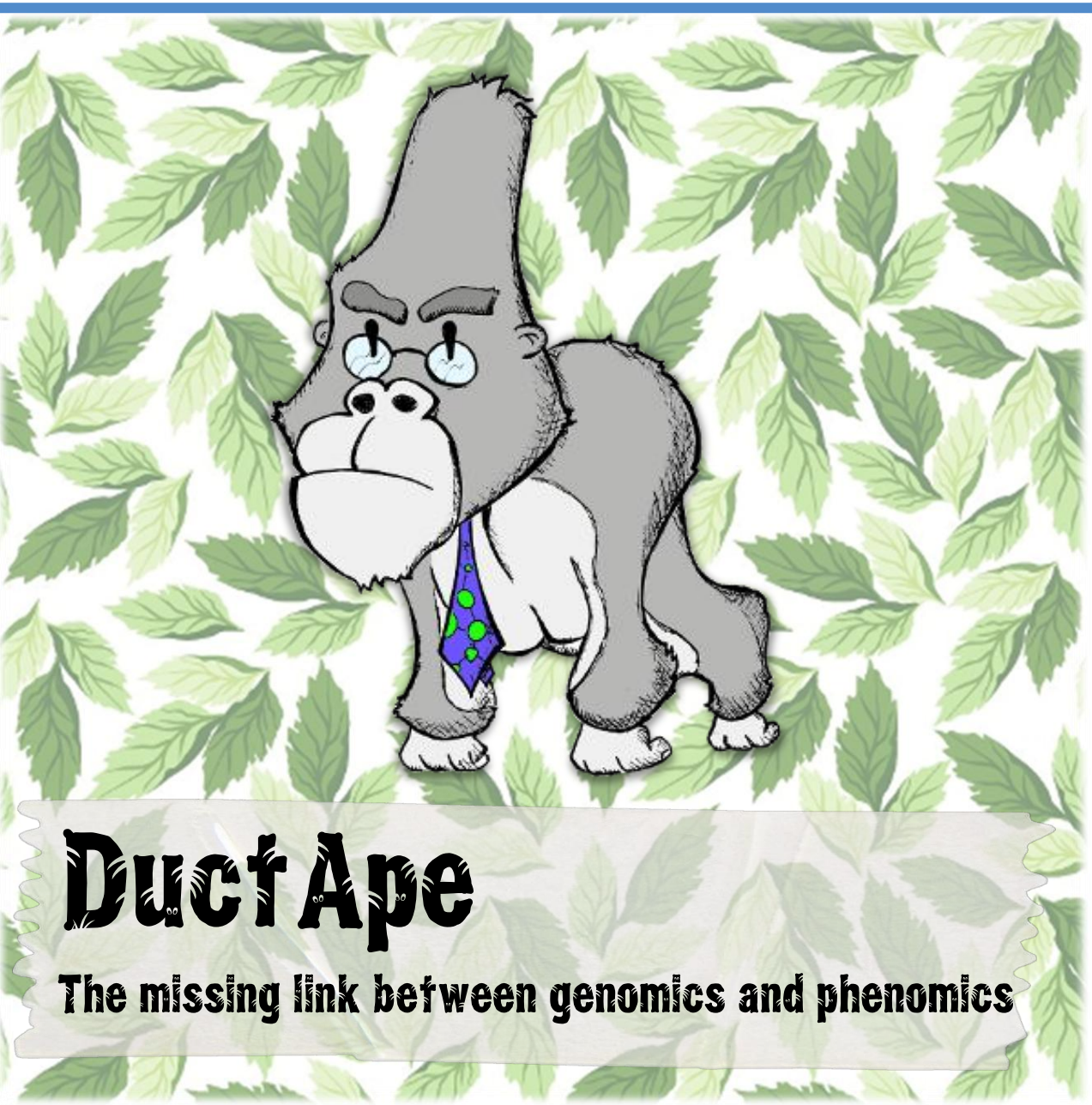
Introduction



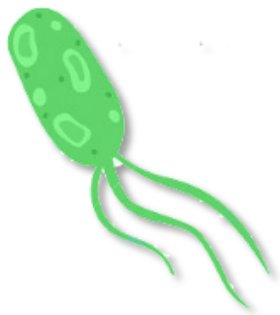


How to combine genomic and phenomic data?

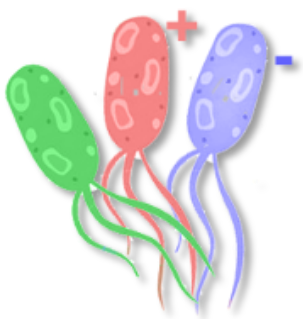
- All data in a single metabolic map
- Genetic basis for phenotypic differences


Duct Ape
The missing link between genomics and phenomics

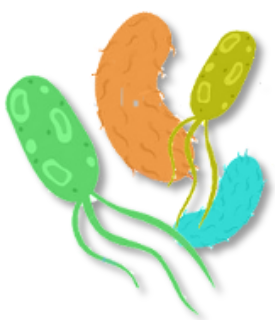
Three different experimental setups



Single strain(s)

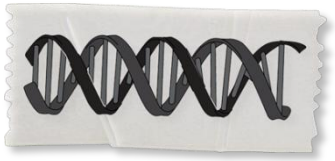


Mutant(s)
- Correlation of mutated genes / different phenotypes
- Deletion / insertion mutants



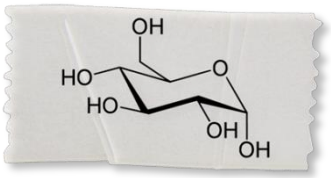
PanGenome
- Prediction of Core / Accessory / Unique genome
- Correlation between Dispensable genome and phenotypes

Three different modules



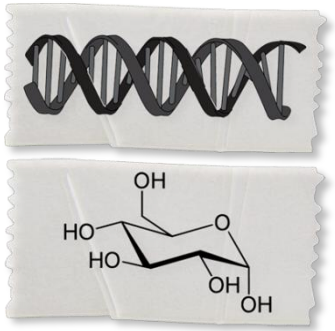
dgenome

- Genes are mapped to KEGG database
- PanGenome prediction (Blast-BBH)



dphenome

- Phenotype microarray data handling (sigmoid fit)
- Classification of metabolic activity (**Activity index**)
- Compounds are mapped to KEGG database

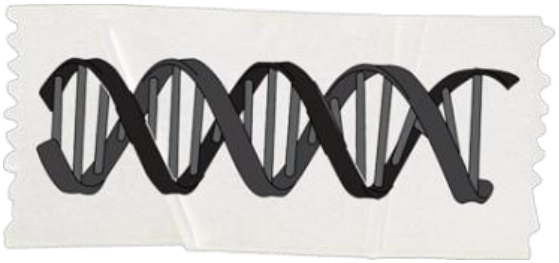


dape

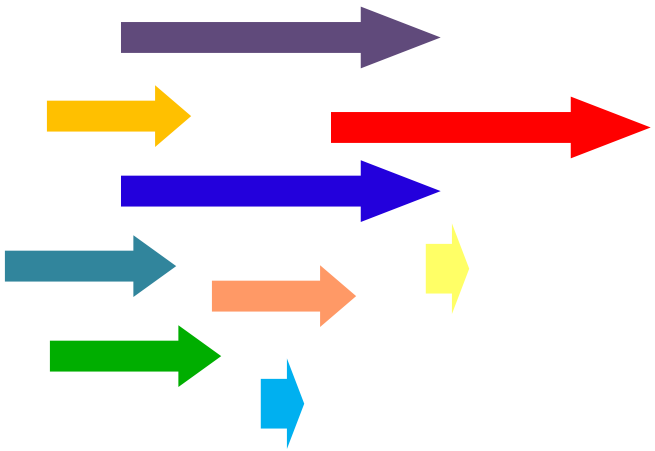
- Generation of combined KEGG metabolic maps
- Metabolic network analysis (through graph algorithms)
- Metabolic hotspots prediction

dgenome

Genomics made easy



Genome map to KEGG (1)

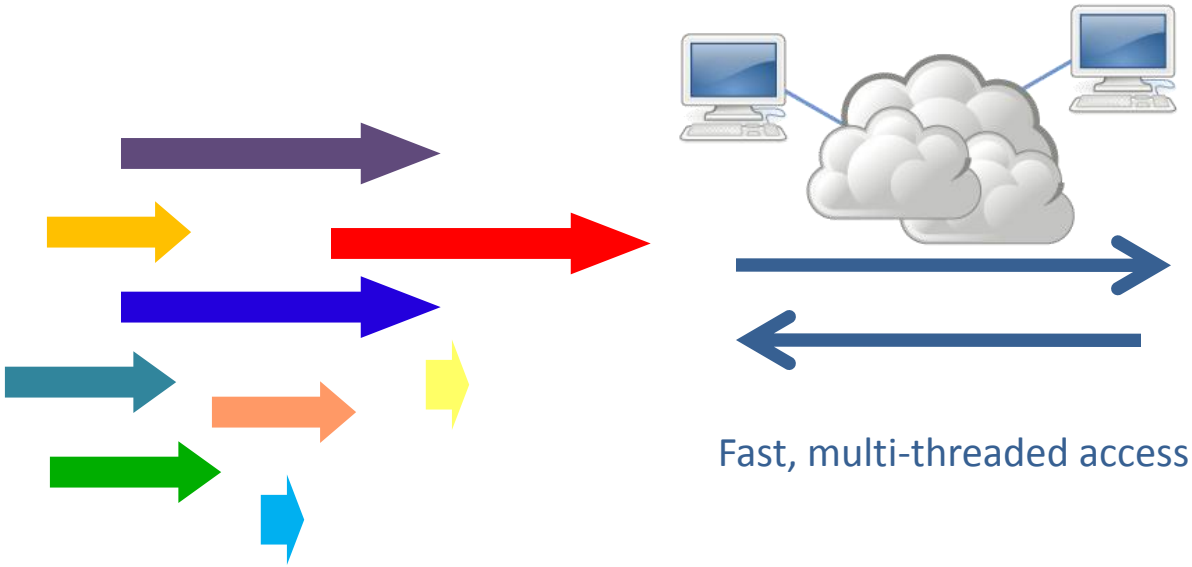


Blast BBH on a local KEGG database*

Blast BBH using KASS web-server**

*Since July 1th 2011, the access to KEGG FTP needs a $2000/$5000 licence
**Available for free, fast and reliable

Genome map to KEGG (2)



Fast, multi-threaded access

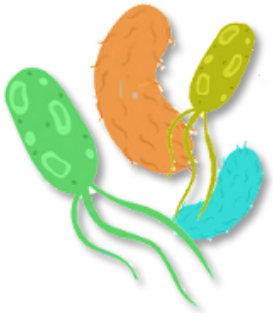
KEGG public API

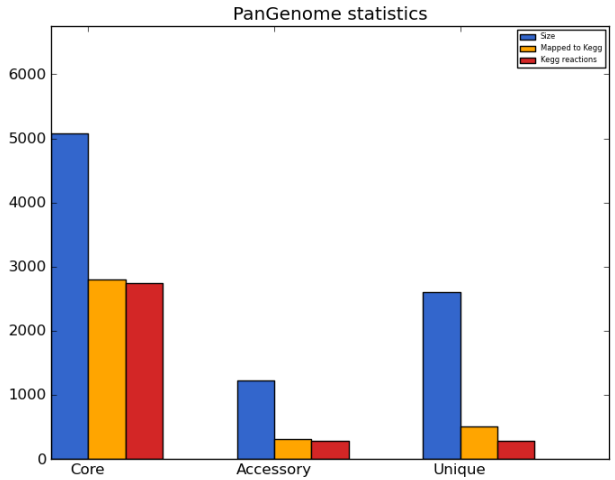
Detailed info on:

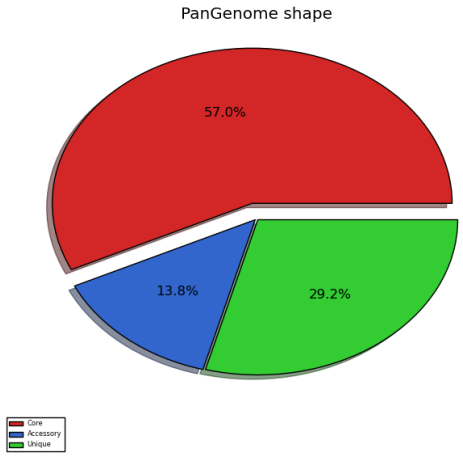
- Reactions
- Compunds
- Pathways

Pangenome prediction

- Many genomes
  - Serial BBH
  - User-defined PanGenome
    - Core Genome (conserved pathways)
    - Dispensable Genome (variable pathways)
      - Accessory Genome
      - Unique Genome

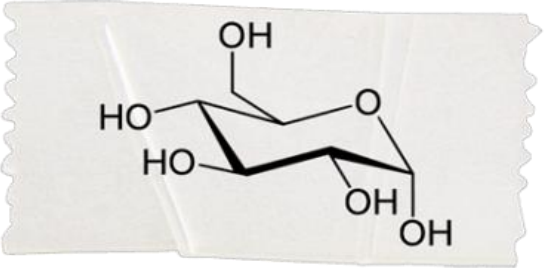




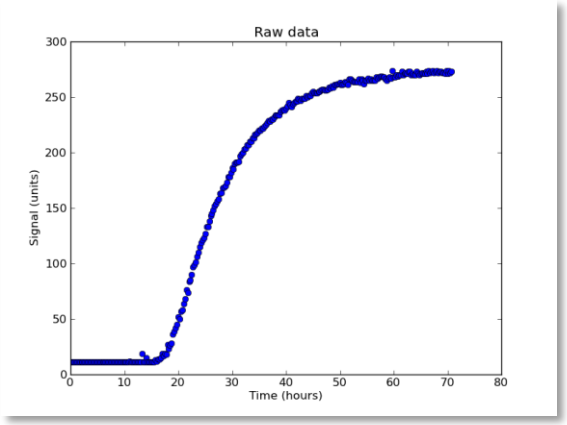


dphenome

Painless high-throughput phenomics

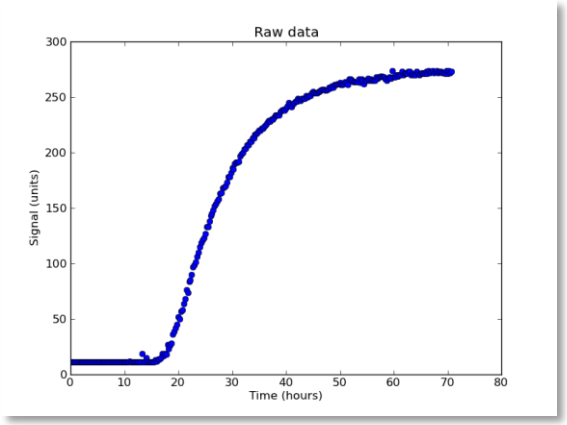


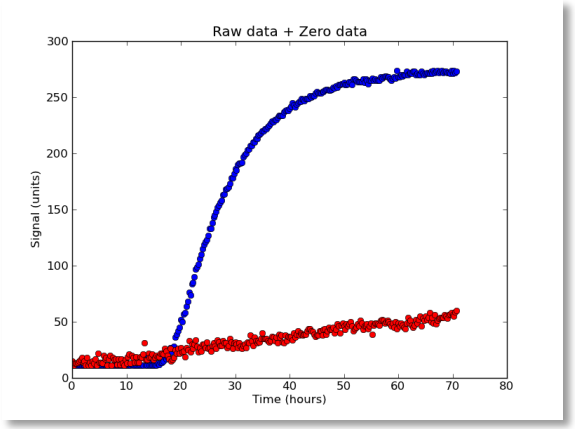
From raw data to phenotypic variability

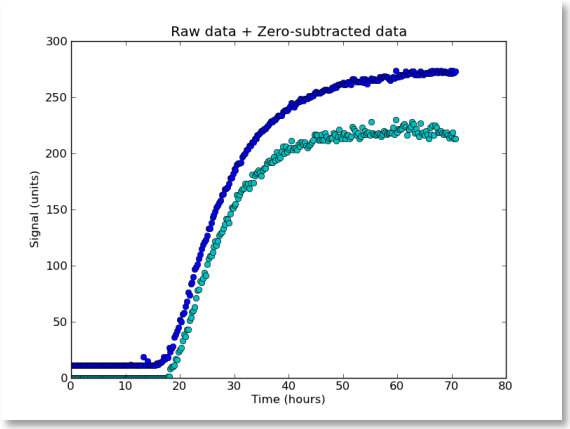


1. Parsing

From raw data to phenotypic variability



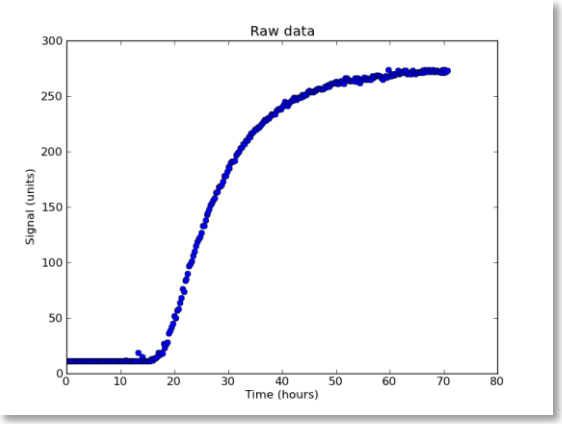




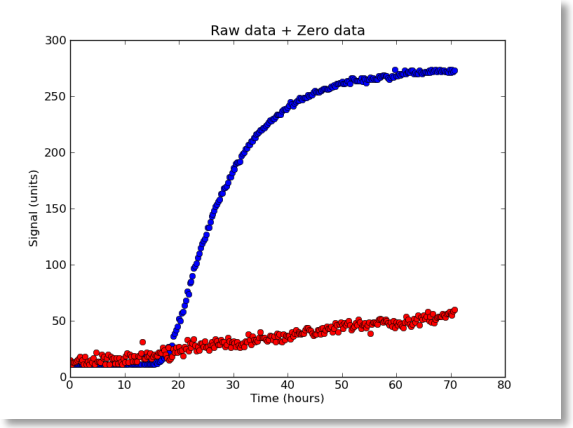
1. Parsing

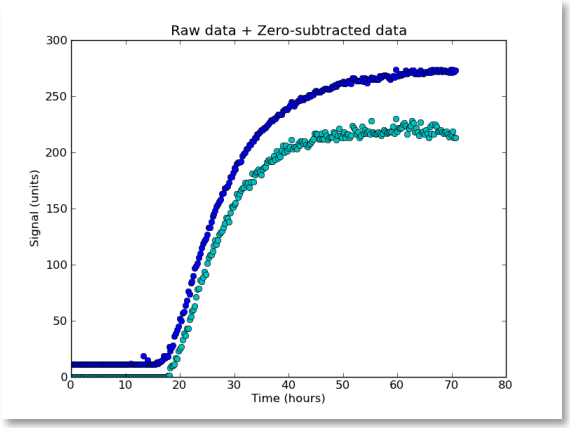
2. Control signal subtraction (optional)

From raw data to phenotypic variability



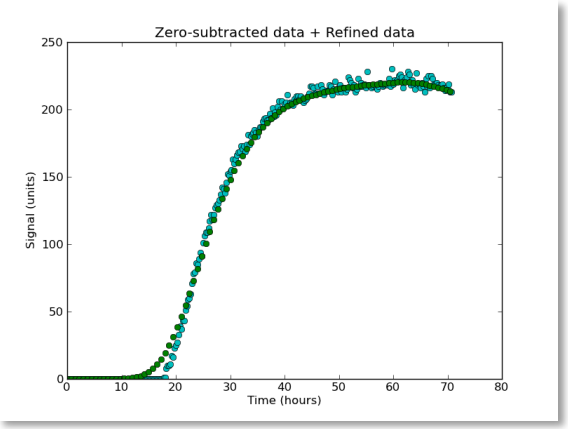
1. Parsing



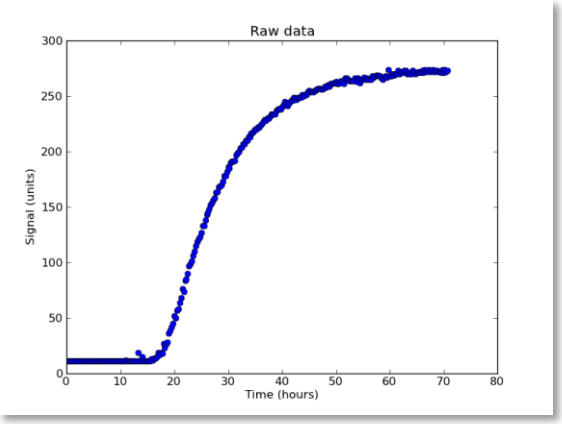


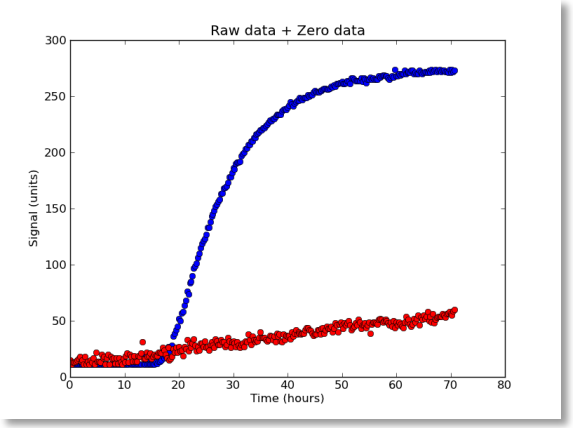
2. Control signal subtraction (optional)

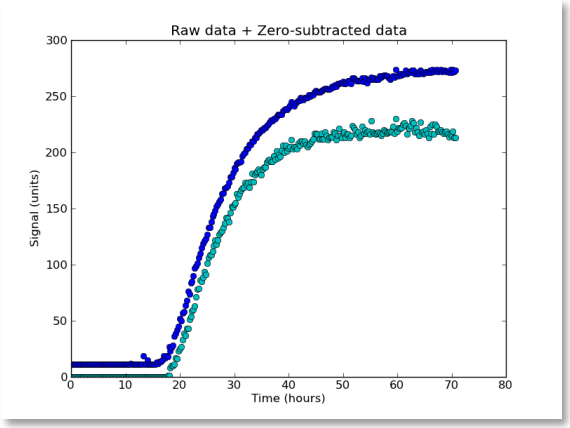
3. Signal refinement



From raw data to phenotypic variability





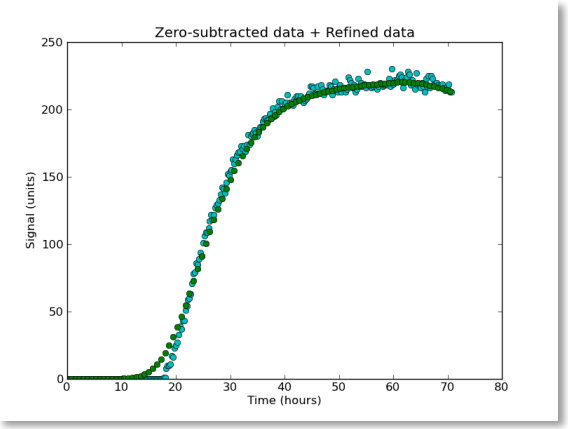


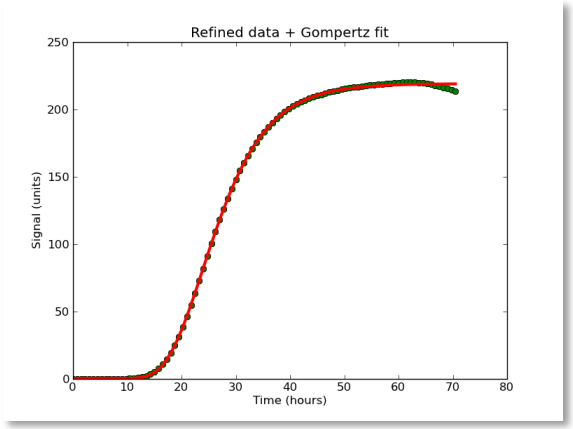
1. Parsing

2. Control signal subtraction (optional)

3. Signal refinement

4. Sigmoid fit





Modeling of the Bacterial Growth Curve

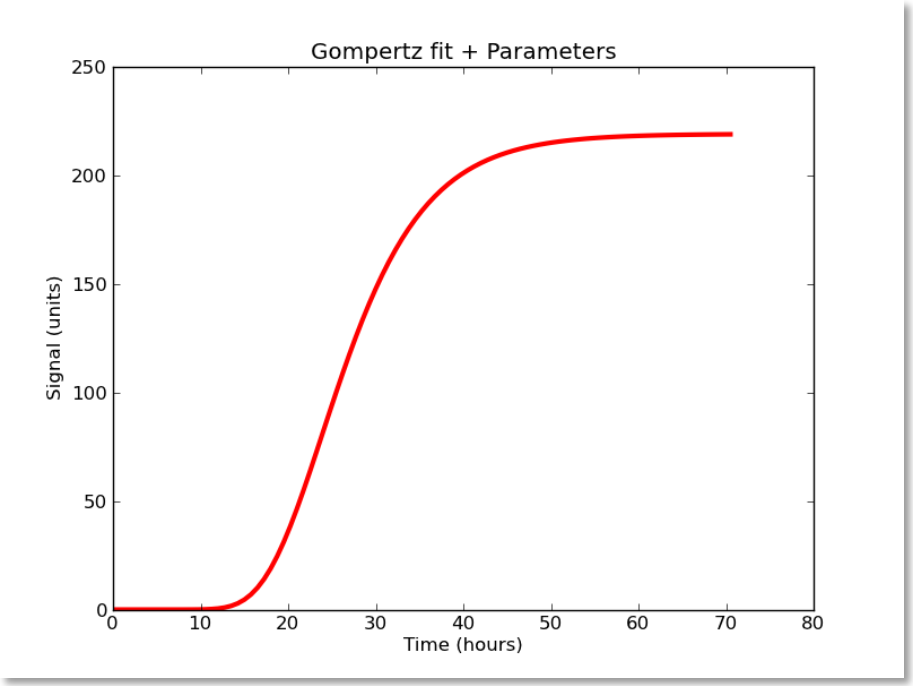
M. H. ZWIETERING,* I. JONGENBURGER, F. M. ROMBOUTS, AND K. VAN 'T RIET
Department of Food Science, Agricultural University Wageningen, P.O. Box 8129, 6700 EV Wageningen, The Netherlands

Received 5 January 1990/Accepted 4 April 1990

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

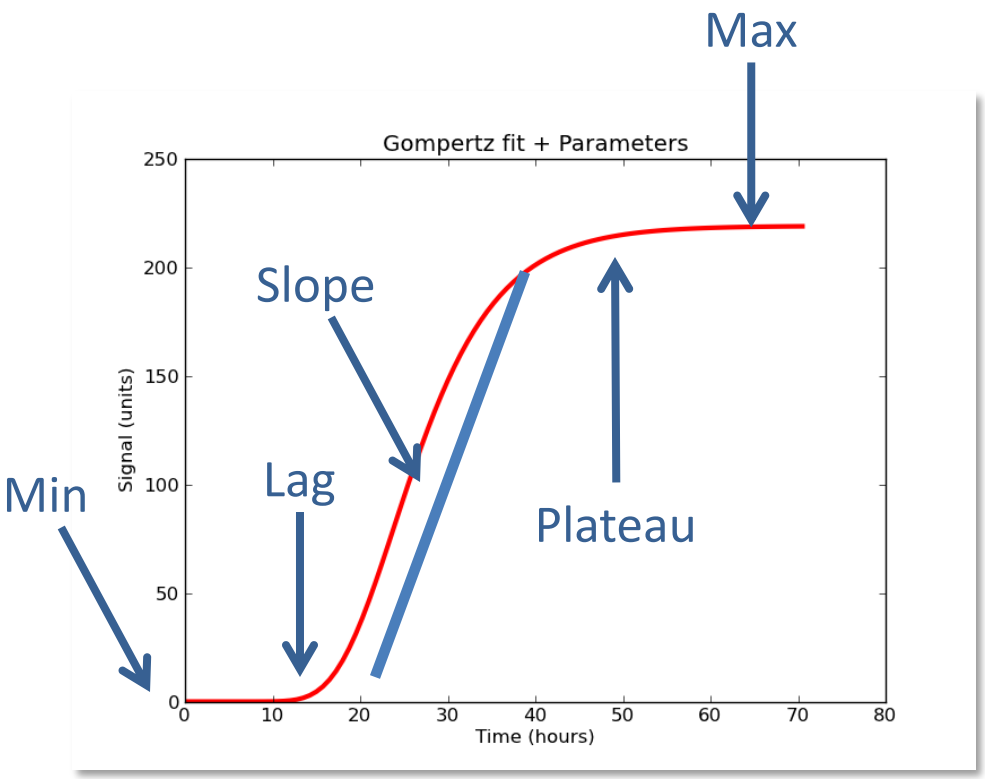
From raw data to phenotypic variability

5. Parameters extraction



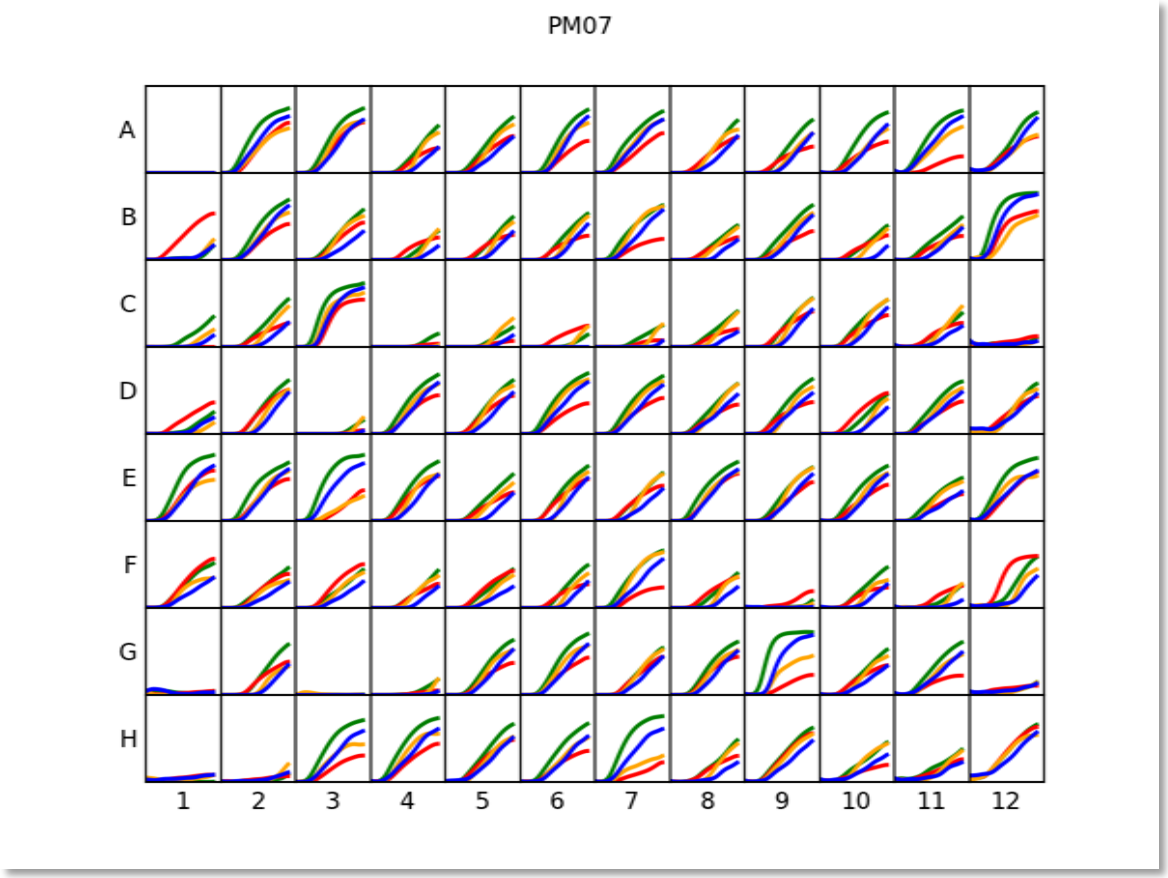
From raw data to phenotypic variability

5. Parameters extraction

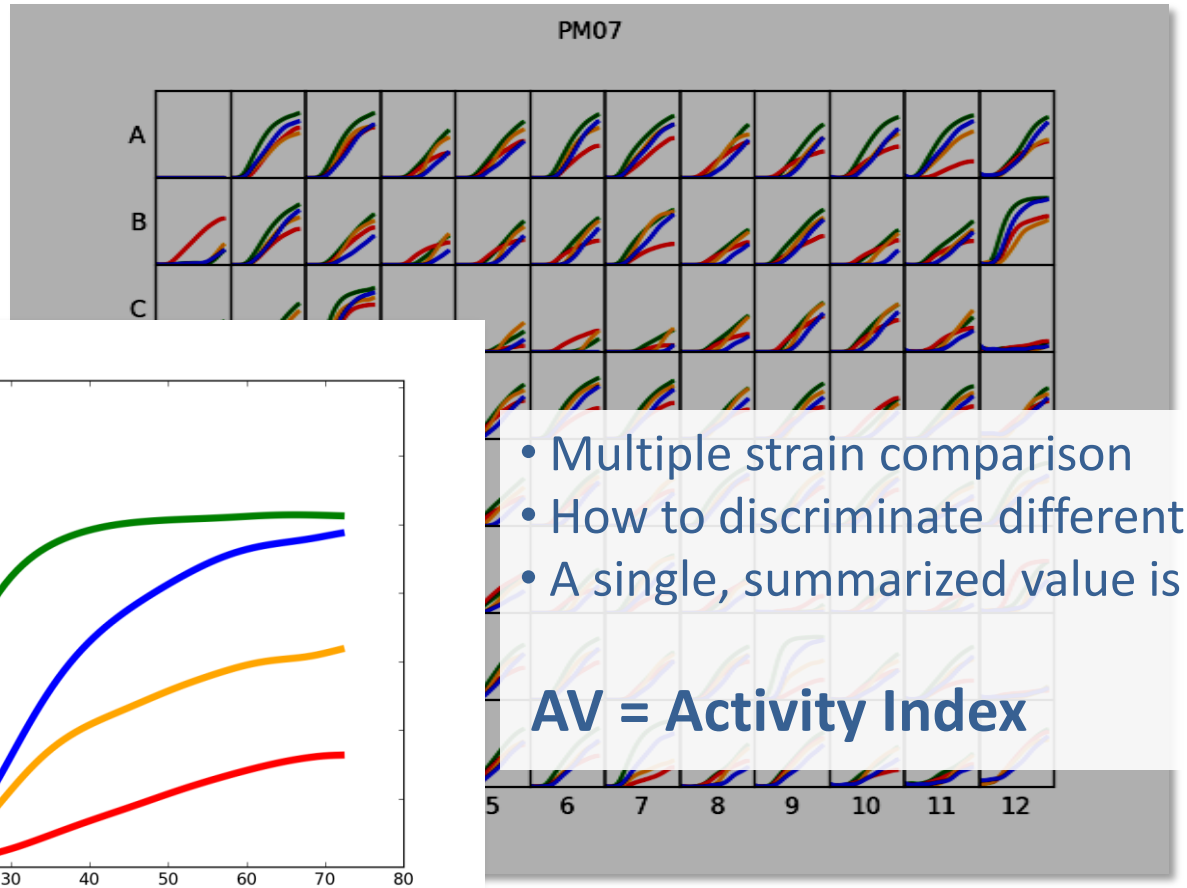


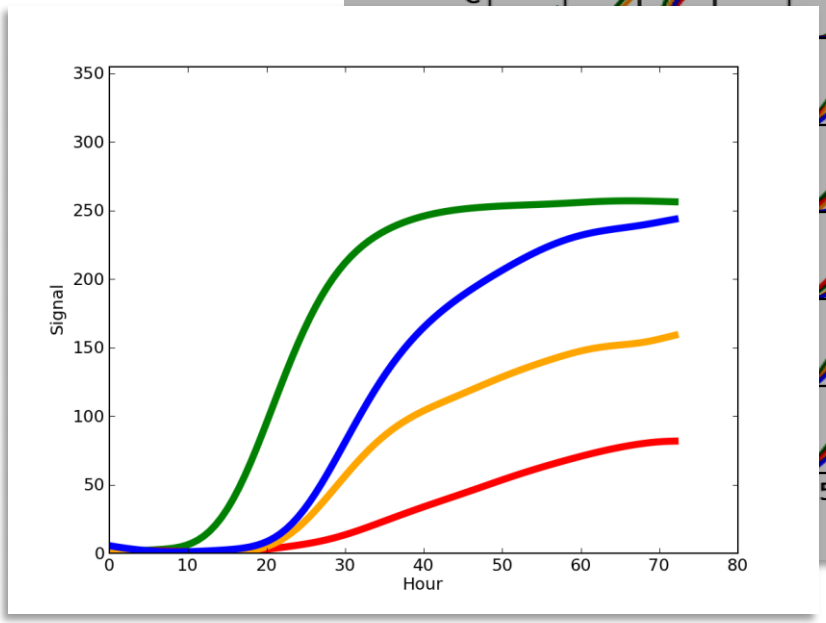
+ Area + Average height

Phenotypic variability at a glance



Phenotypic variability at a glance

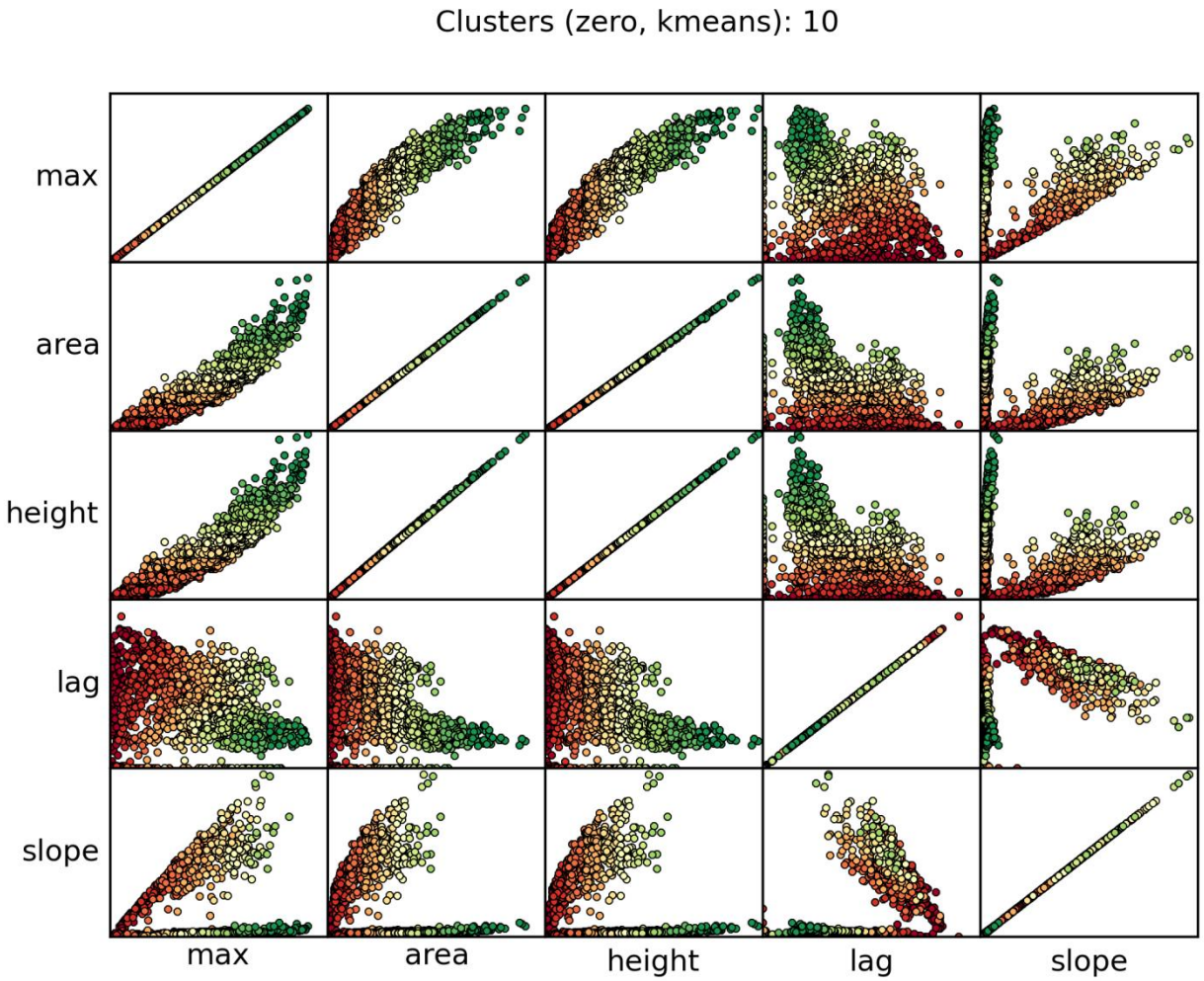




- Multiple strain comparison
- How to discriminate different activities?
- A single, summarized value is needed

AV = Activity Index

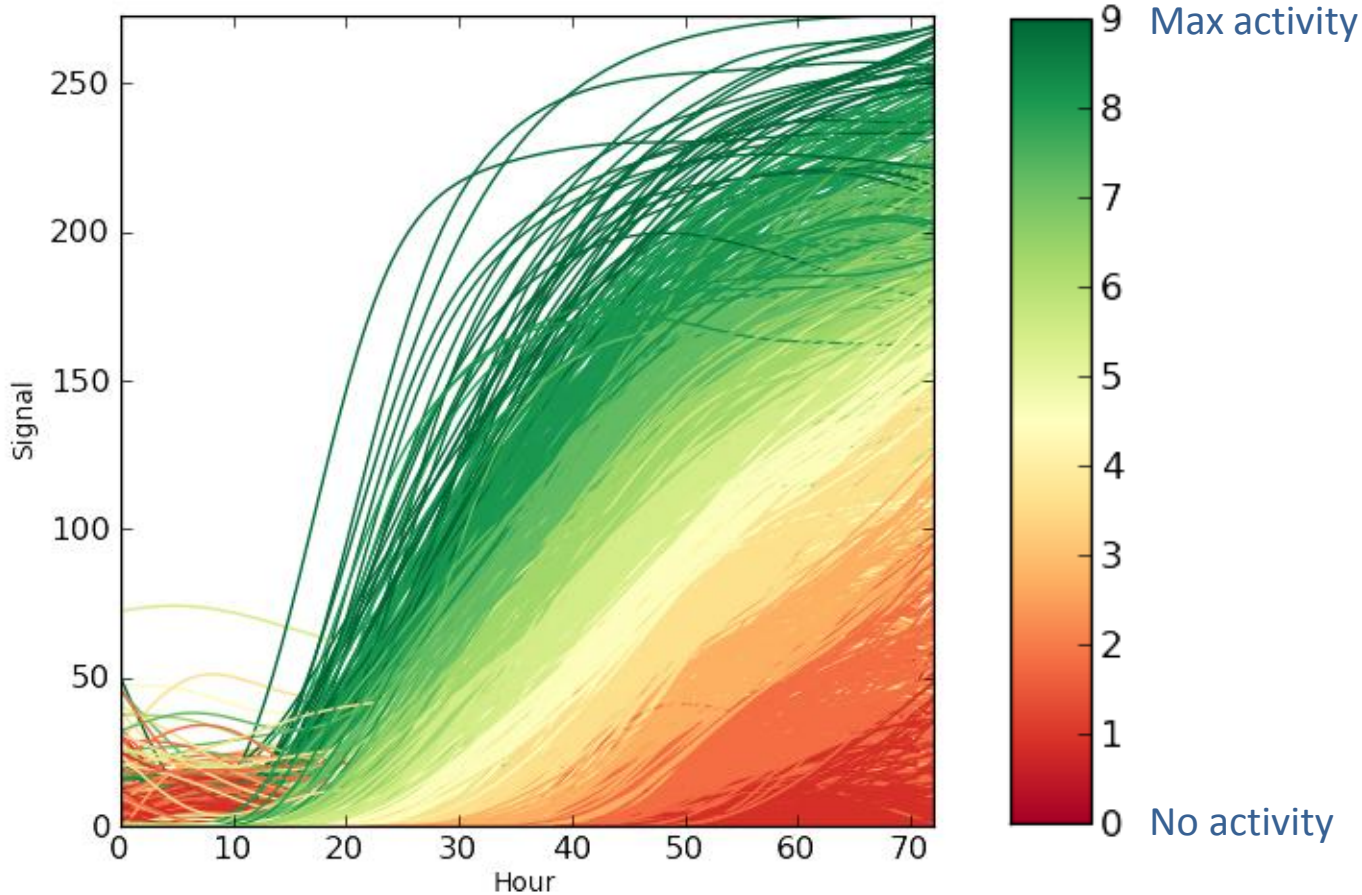
Activity index (AV)



K-means clustering on 5 parameters, with 10 clusters

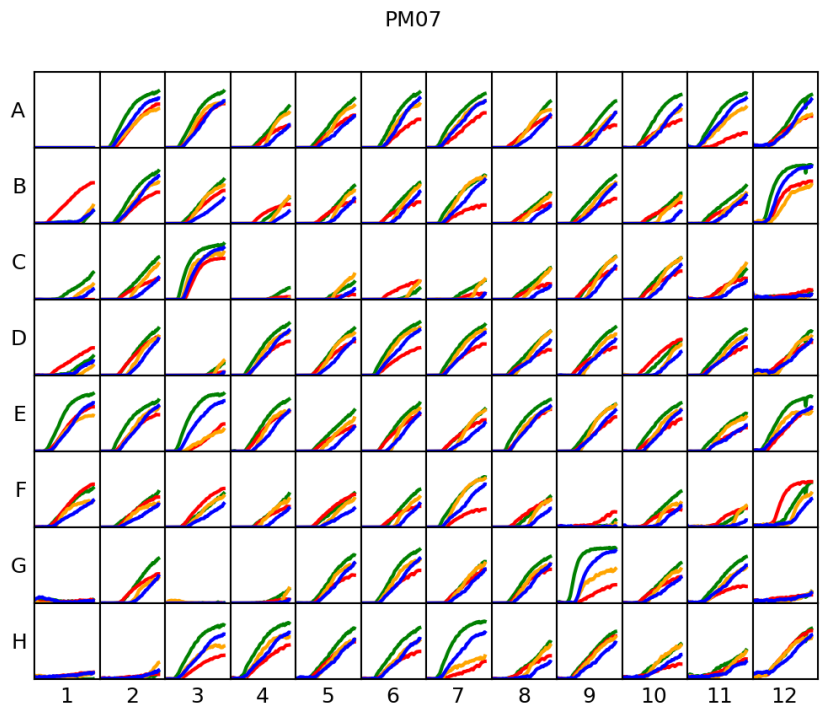
Fast: from raw .csv files to AV in less than 5 minutes

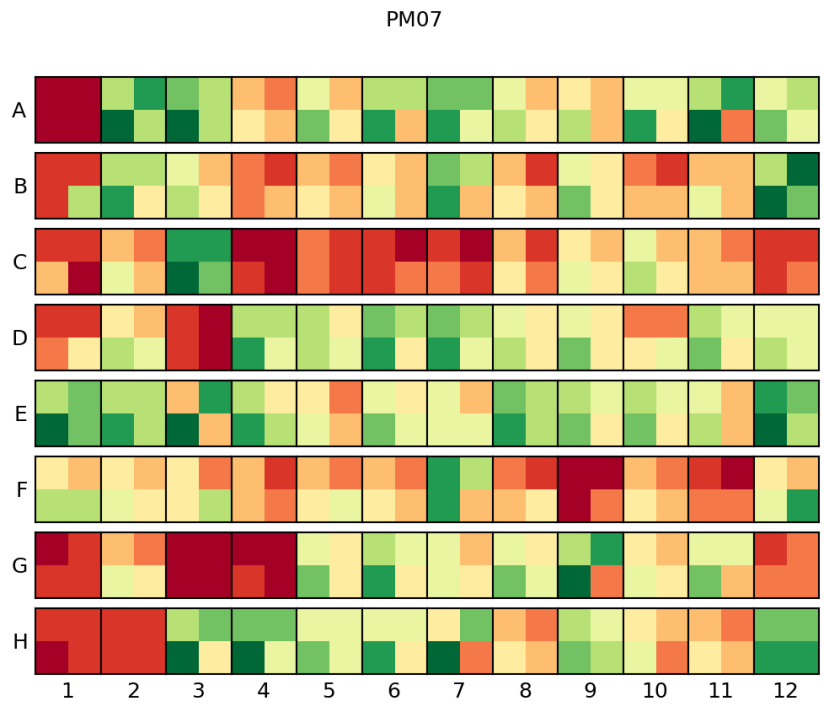
Activity index (AV)



- Easier comparisons
- More understandable metrics
- Different experiments comparison

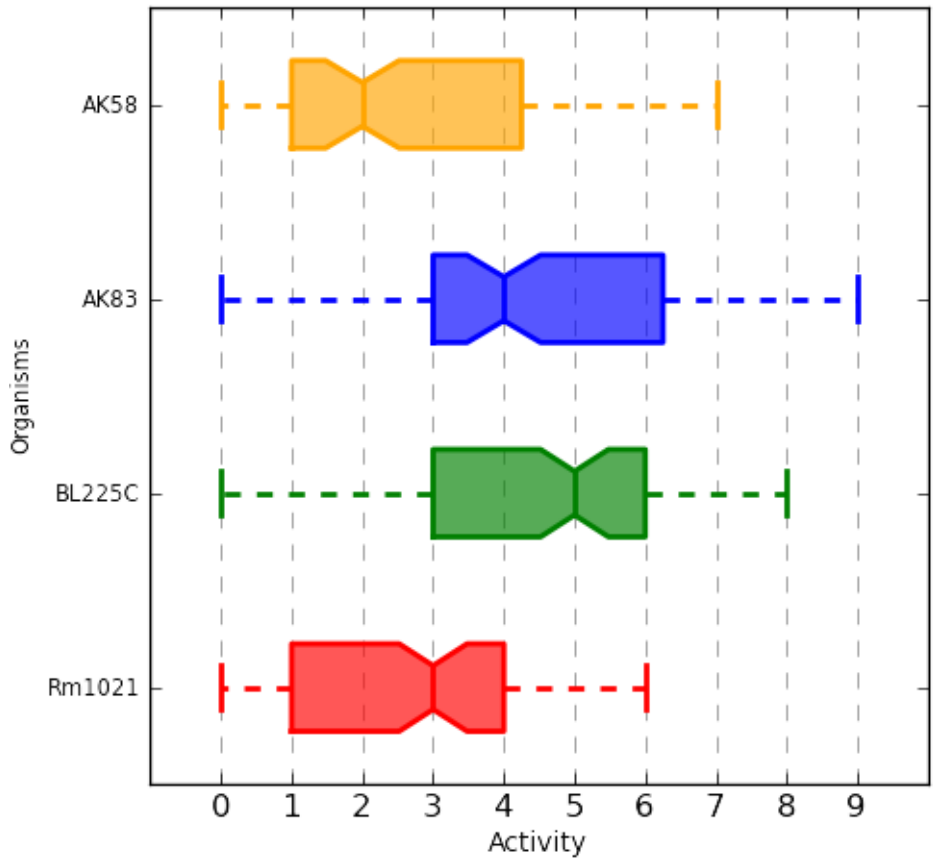
Activity index (AV)





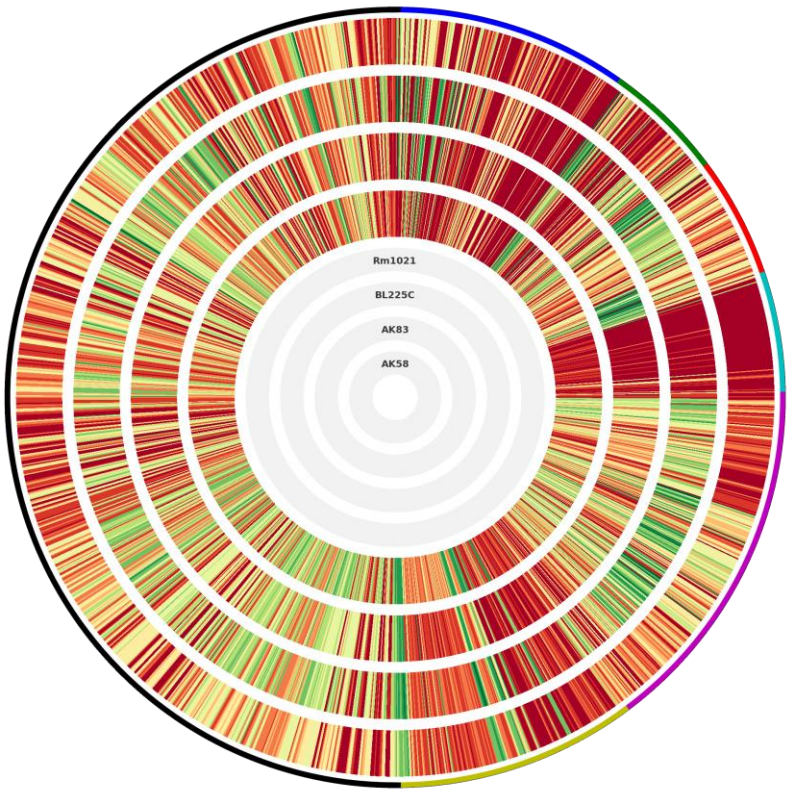
Plates heatmaps: phenotypic variability at a glance

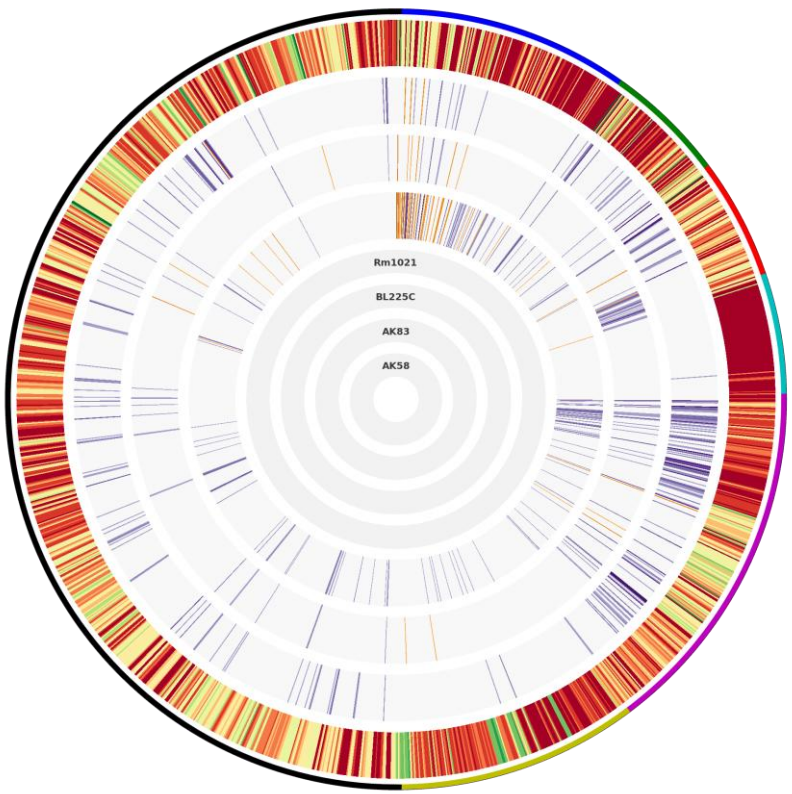
Activity index (AV)

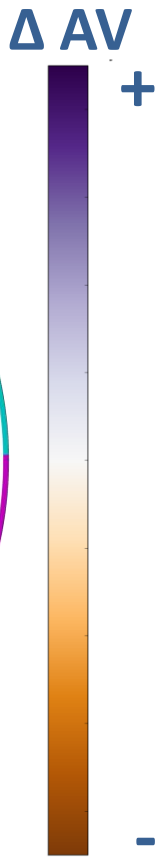


AV boxplots: overall strains comparison (also on single compounds categories)

Activity index (AV)

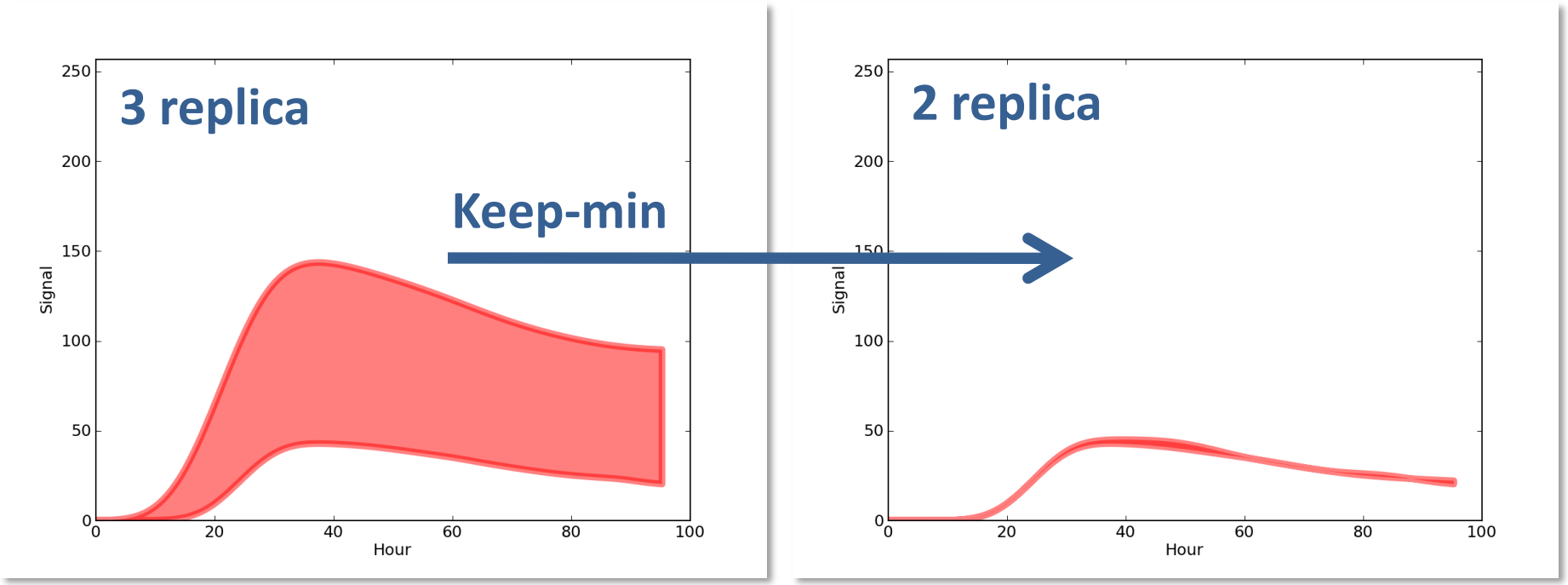






AV rings: overall strains comparison

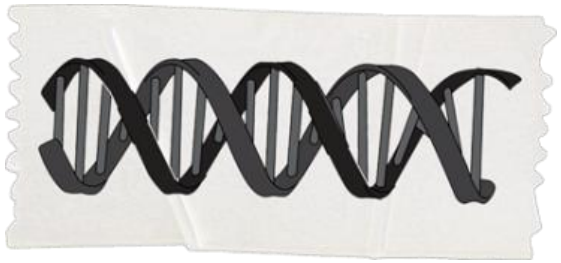
Activity index (AV)

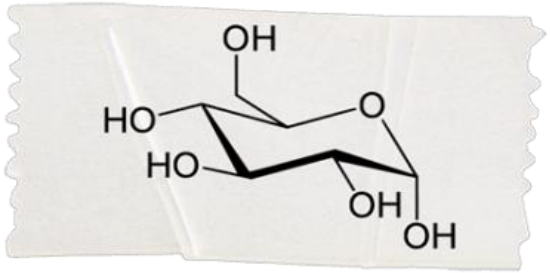


Replica management: discard inconsistent replica using the Δ AV

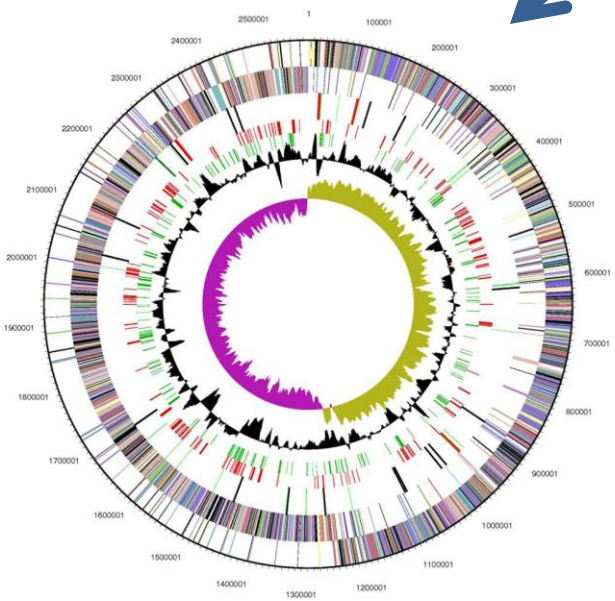
dape

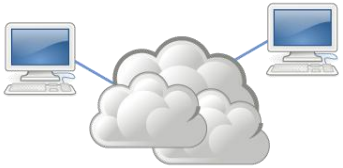
The missing link



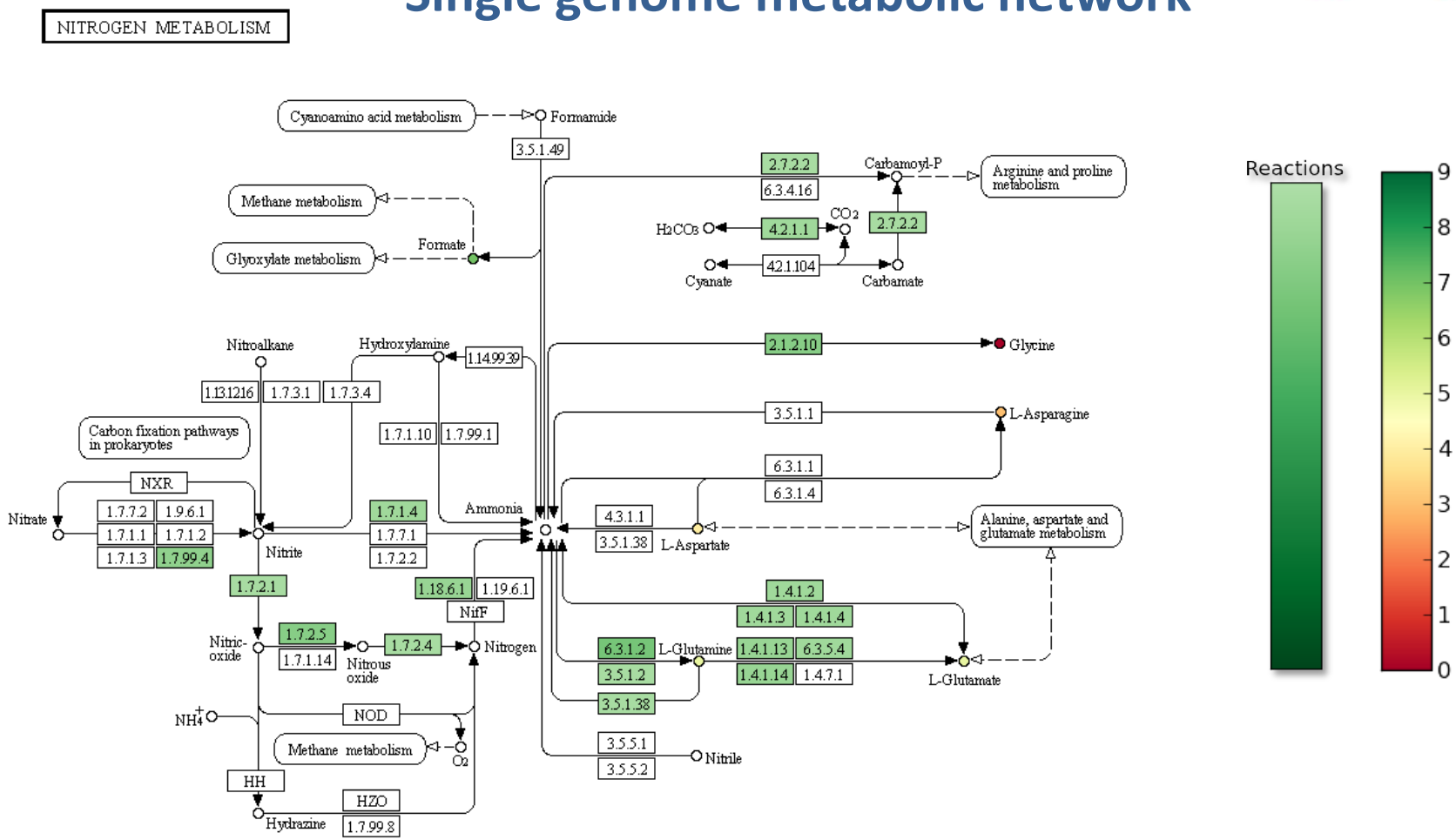


Whole metabolic network reconstruction





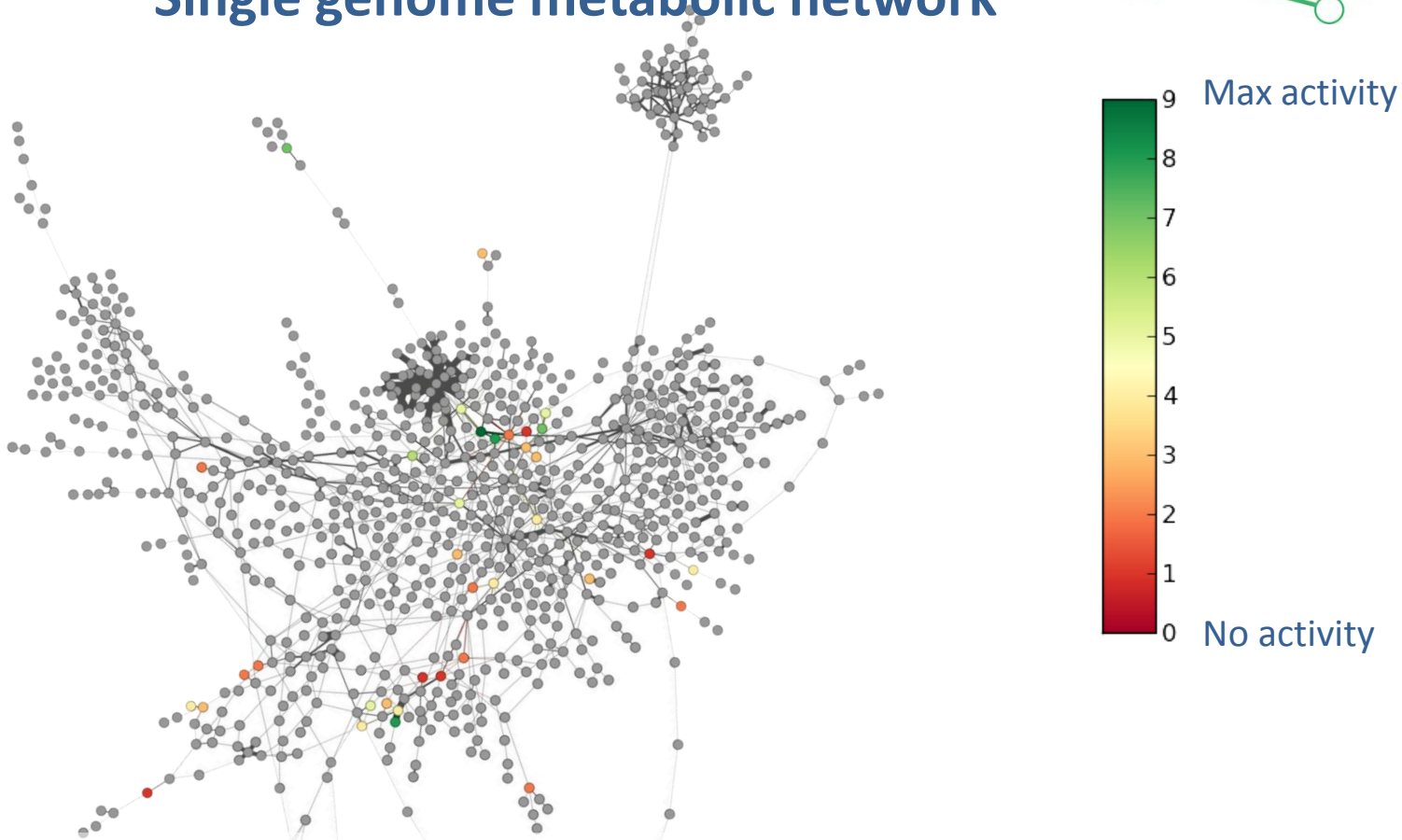
Single genome metabolic network



Interactive metabolic maps (as web pages)

- Reactions copy number
- Compounds AV

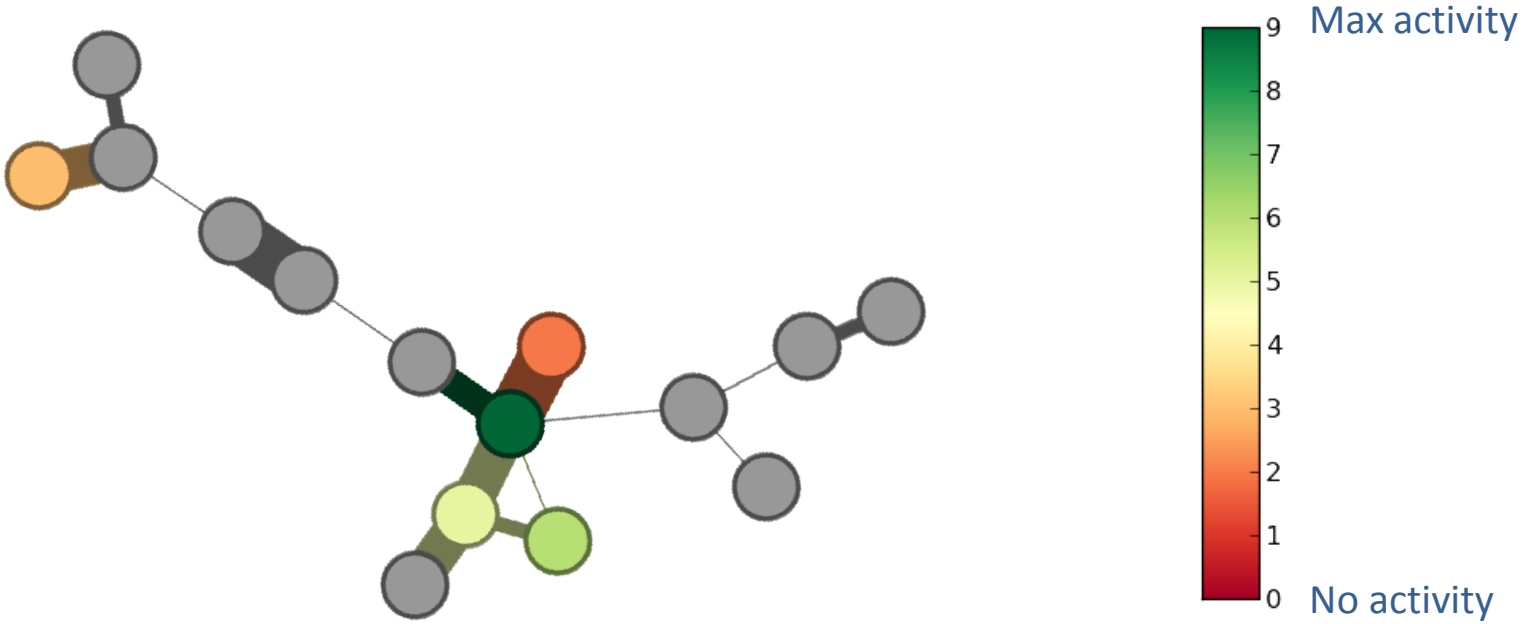
Single genome metabolic network



Interactive metabolic maps (as graph files)

- Can be used with graph analysis softwares (i.e. Gephi)
- Generation of tables with network statistics on single pathways

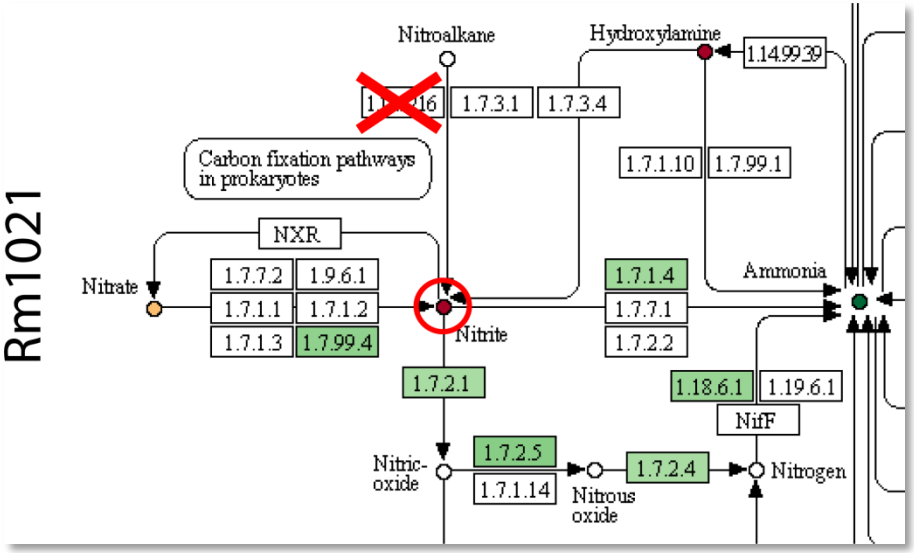
Single genome metabolic network



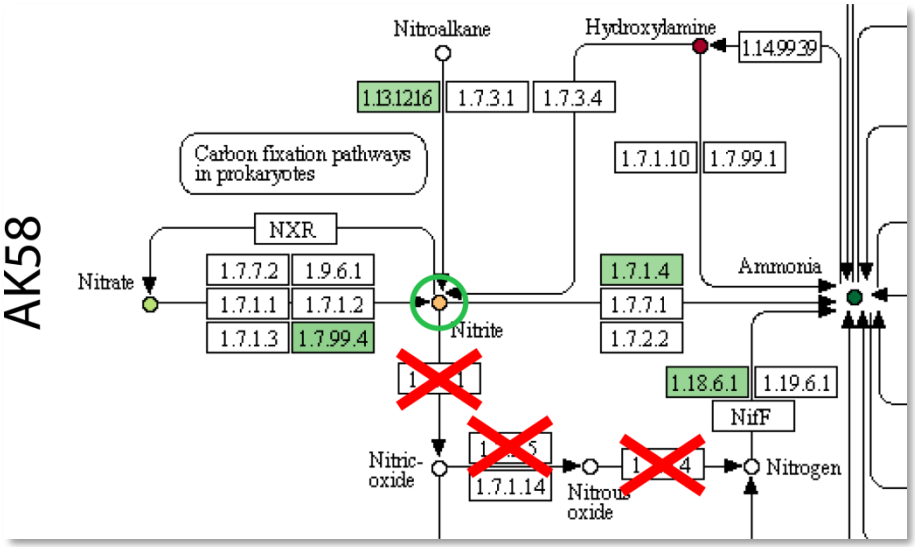
Interactive metabolic maps (as graph files)

- Can be used with graph analysis softwares (i.e. Gephi)
- Generation of tables with network statistics on single pathways

Rm1021



AK58

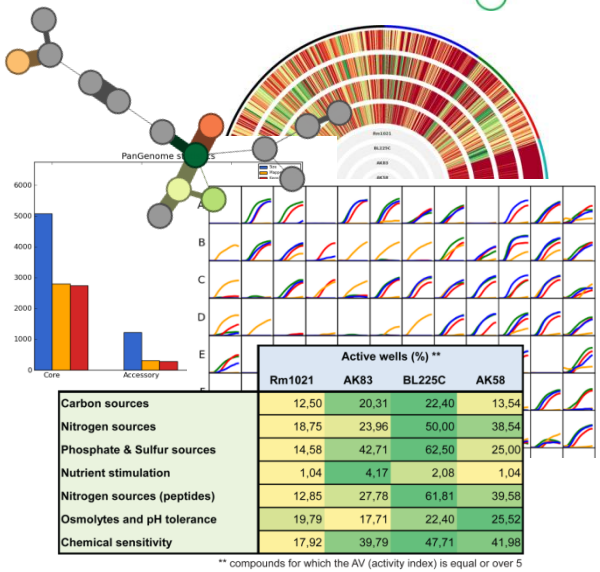


Metabolic network comparisons

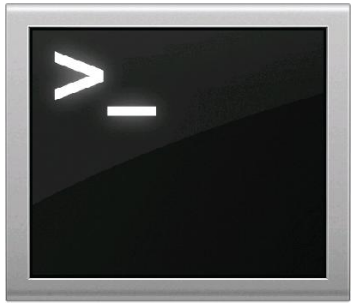
Under the hood

Technical features

Technical features



Inputs **DuctApe project file** **Outputs**



DuctApe comes as a UNIX command line program

- Clear, modular and expressive syntax
- A web interface is under development
- Next versions will be compatible with opm

Technical features

Language

python

Standing on the shoulders of giants

- Curve fitting
- Signal handling
- Clustering
- Sequence handling
- Plots
- Metabolic network (networkx)

SciPy

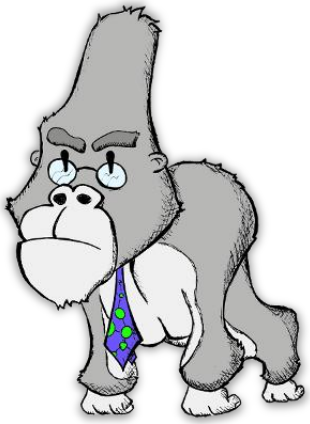
NumPy

scikits learn

BIOPYTHON

matplotlib

Under the hood



http://combogenomics.github.com/DuctApe

"combogenomics ductape"

ductape-users@googlegroups.com

@combogenomics

Acknowledgements

- **University of Florence**

Alessio Mengoni
Marco Bazzicalupo
Emanuela Marchi
Giulia Spini
Francesca Decorosi
Carlo Viti
Luciana Giovannetti

- **Biolog Inc.**

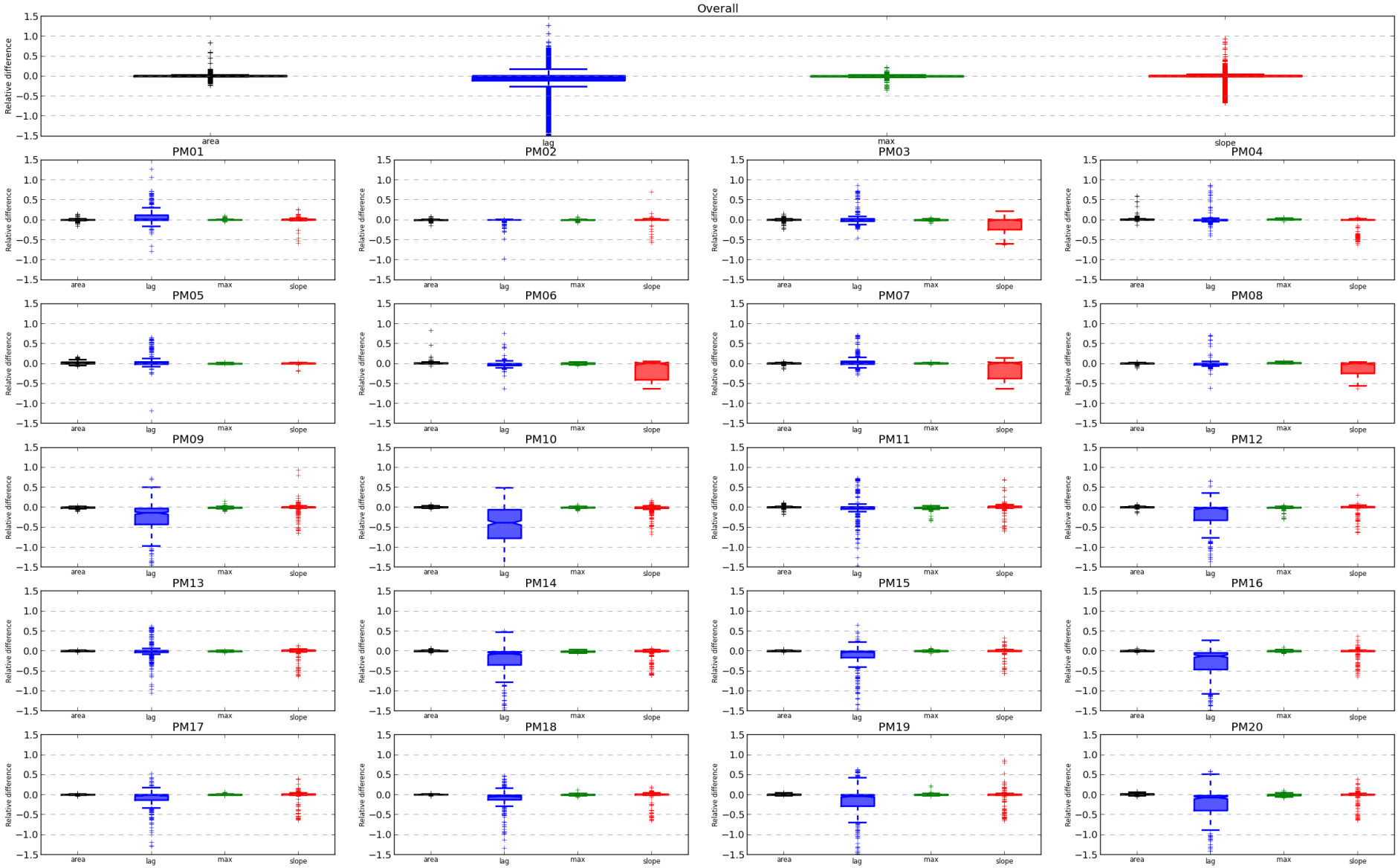
Barry Bochner

- **CRA**

Stefano Mocali
Alessandro Florio
Anna Benedetti

- **University of Lille**

Emanuele Biondi


Overall
PM01
PM02
PM03
PM04
PM05
PM06
PM07
PM08
PM09
PM10
PM11
PM12
PM13
PM14
PM15
PM16
PM17
PM18
PM19
PM20
Relative difference
area
lag
max
slope
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5