Generation and Analysis of Biolog Data for Metabolic Modeling

Heather Danhof, Firas Midani, James Collins
Laboratory of Robert Britton
Baylor College of Medicine, Houston

Overview:

- Introduction to what the Biolog system is
- Rationale for protocol modifications
- Example data and talking points
- Data workflow and analysis.

BCM
Baylor College of Medicine

Texas Children's Hospital

Alkek Center for Metagenomics and Microbiome Research

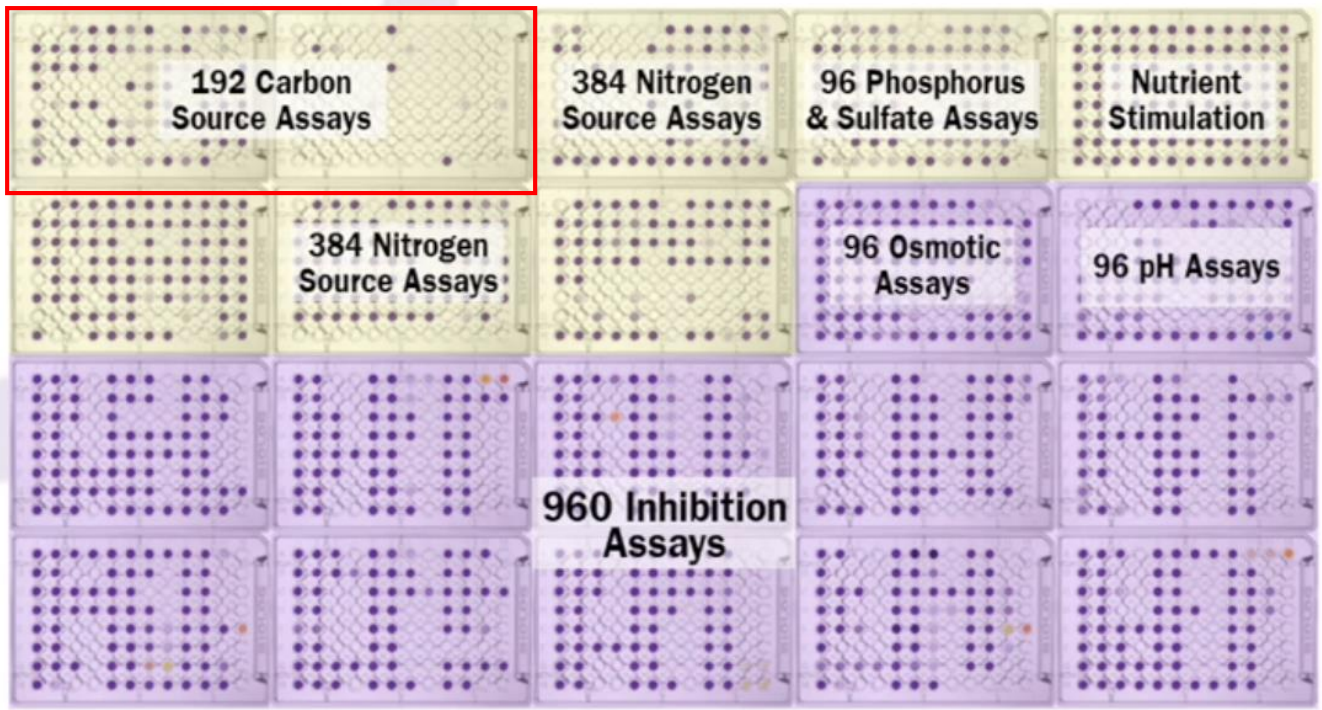
SYSTEMS BIOLOGY RESEARCH GROUP

Phenotype MicroArray Technology

Patented 96-well array system

Standardized comparisons across strains

Medium through put



BIOLOG

How the Biolog system works

- Utilizes proprietary reporter system of NADH production to reduce a tetrazolium dye and produce a color change (purple).
- The regents are designed to slow respiration in “weakly” positive cells and reduce any false positive signals.
- Uses a specialized OmniLog instrument to read and analyze data

https://www.toshindia.com/products/bacteria-yeast-and-fungi-identification-system

Using Biolog Phenotype Arrays

Benefits

- High throughput as marketed
- Convenient pre-arrayed compounds
- Standardized

Limitations

- Proprietary
  - No details on concentrations
  - No ability to change media components
- Reagents and dyes are very oxygen sensitive
- Special equipment is costly

Modifications to the Standard Biolog Protocol

- OD_{620} Absorbance to monitor growth directly.
  - Uses a standard plate reader inside the chamber.
    - Programmable readings
    - Control over chamber environment
  - Solves the issue of the oxygen sensitive dyes in the anaerobic chamber.

- Use completely defined minimal growth medium that has been optimized for *C. difficile*.

Experimental setup

Strains are grown overnight in BHIS

Diluted 1:10 in CDMM

Aliquoted in Biolog plates (100 uL/well)

Incubated in anaerobic plate reader, 37 degrees
    Shakes 30 seconds before reading every 10 minutes for ~16 hours

Each strain is done in duplicate, sometimes more*

Biolog data

Each plate has unique growth curves

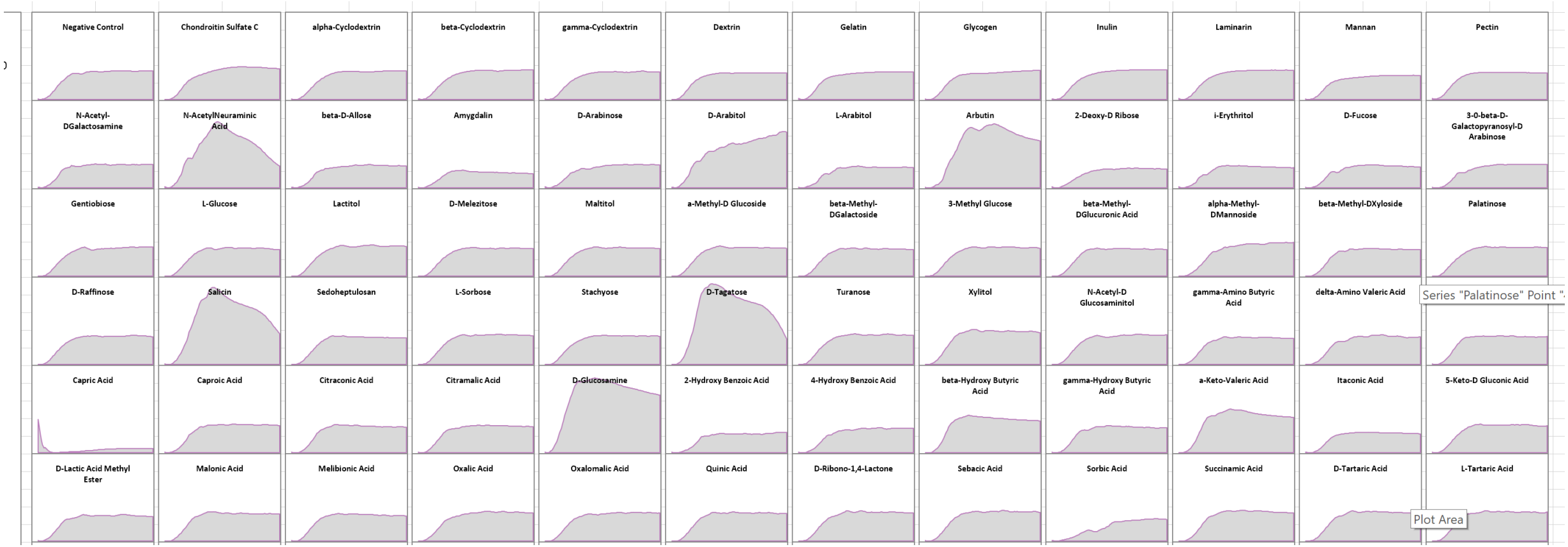
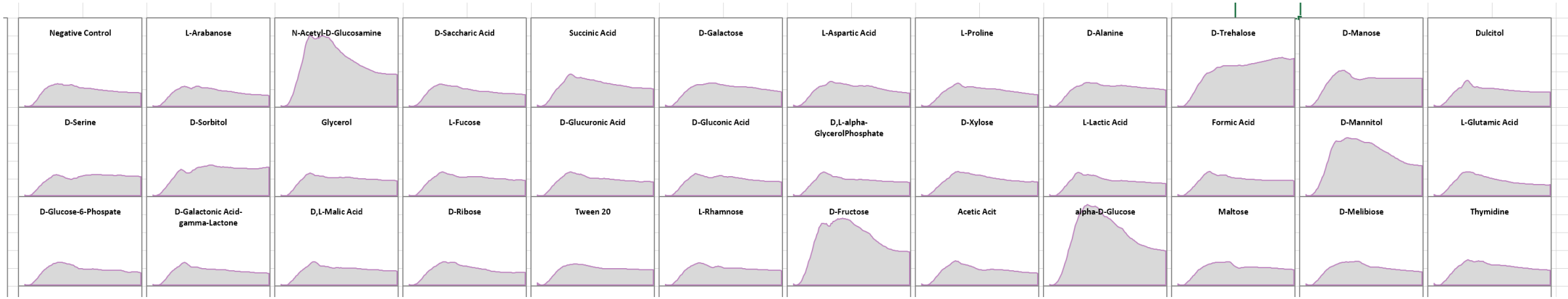
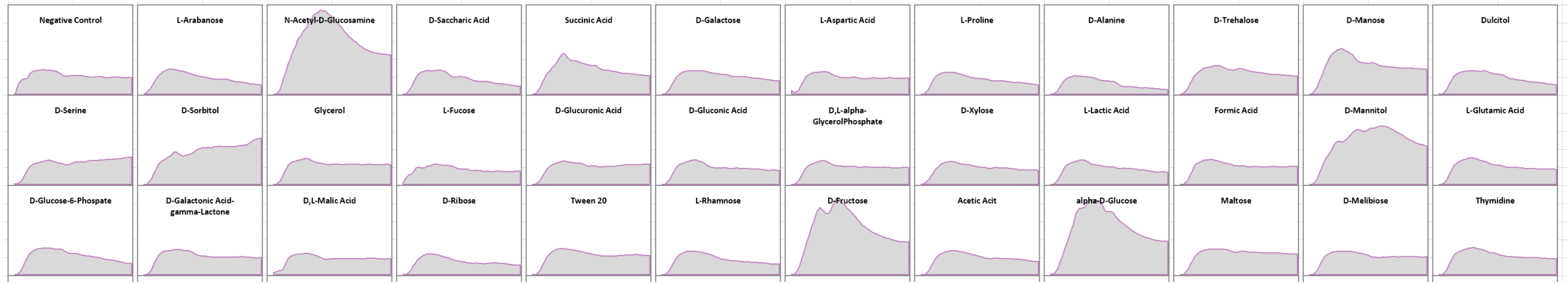
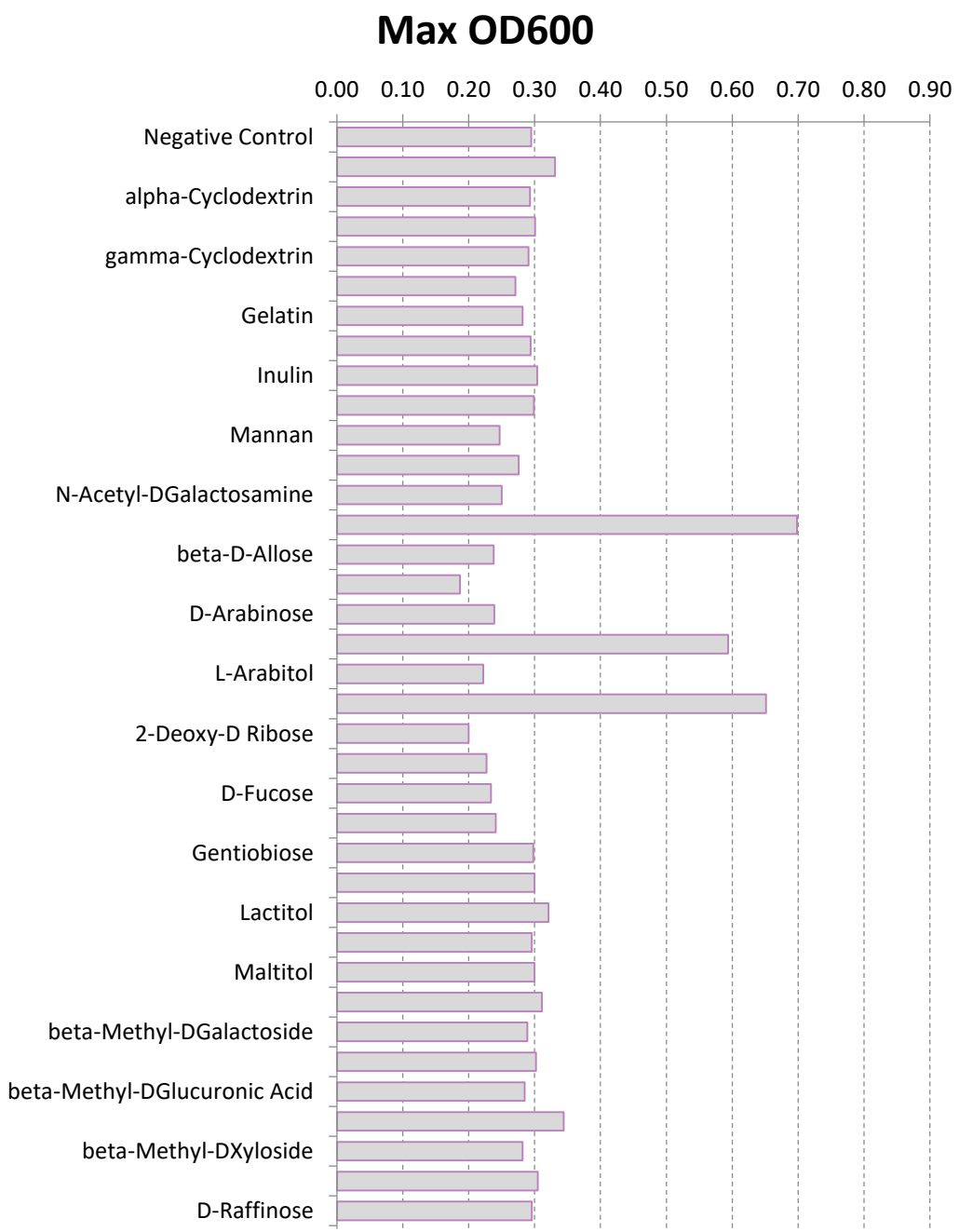
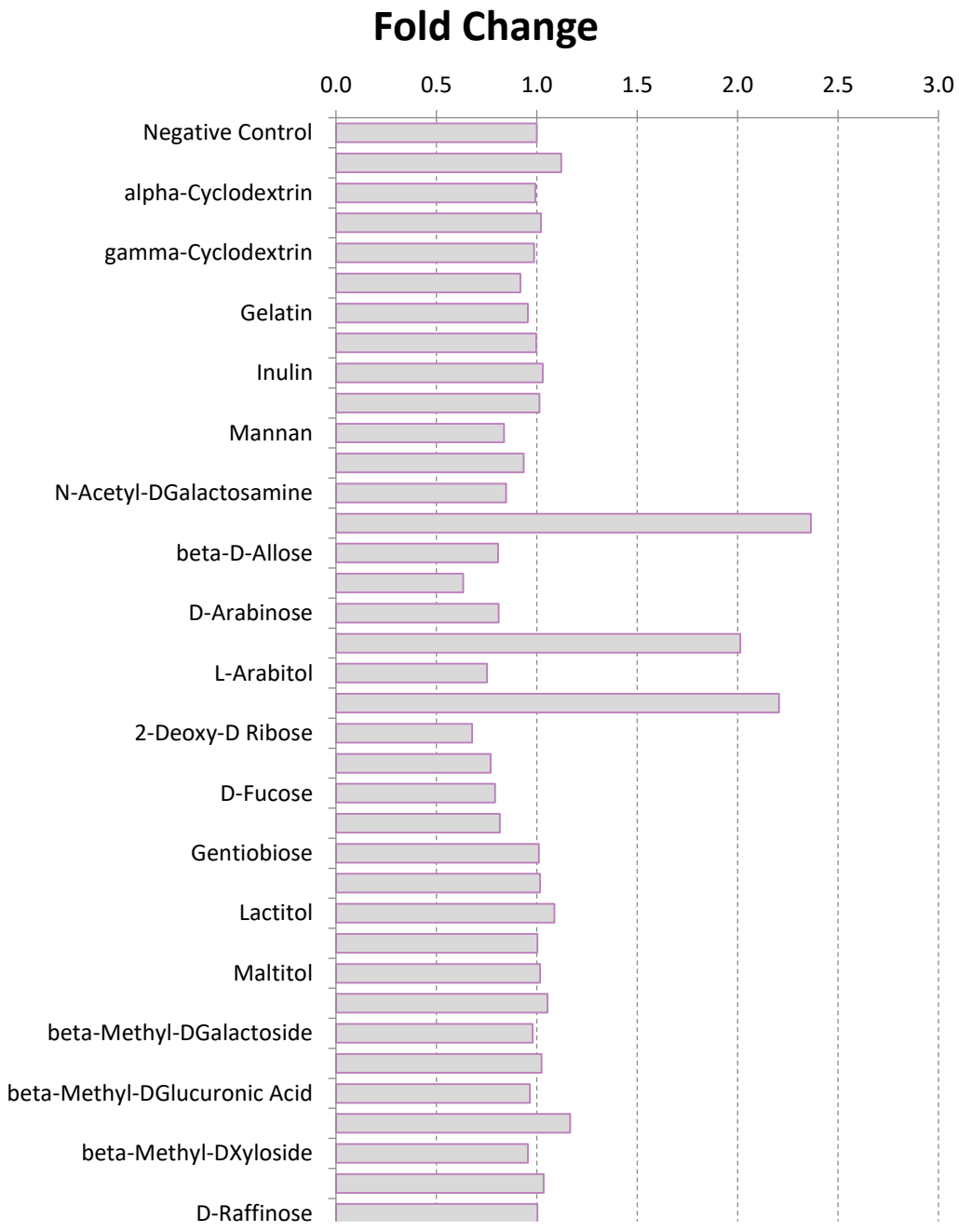


Plate to plate variability






Max OD600
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
Negative Control
alpha-Cyclodextrin
gamma-Cyclodextrin
Gelatin
Inulin
Mannan
N-Acetyl-DGalactosamine
beta-D-Allose
D-Arabinose
L-Arabitol
2-Deoxy-D Ribose
D-Fucose
Gentiobiose
Lactitol
Maltitol
beta-Methyl-DGalactoside
beta-Methyl-DGlucuronic Acid
beta-Methyl-DXyloside
D-Raffinose

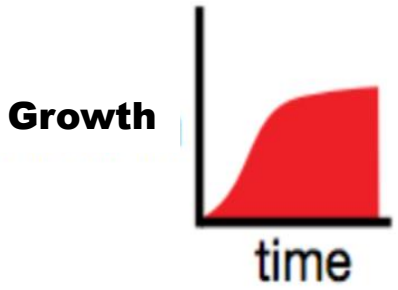

Fold Change
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Negative Control
alpha-Cyclodextrin
gamma-Cyclodextrin
Gelatin
Inulin
Mannan
N-Acetyl-DGalactosamine
beta-D-Allose
D-Arabinose
L-Arabitol
2-Deoxy-D Ribose
D-Fucose
Gentiobiose
Lactitol
Maltitol
beta-Methyl-DGalactoside
beta-Methyl-DGlucuronic Acid
beta-Methyl-DXyloside
D-Raffinose

Determining Growth and fold change

Negative Control

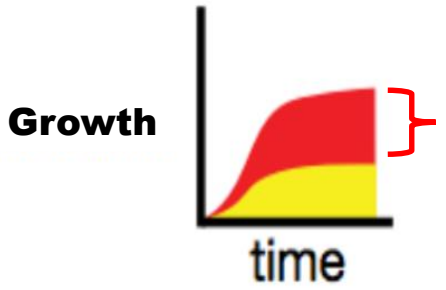
+

Experimental Carbon



=

Comparison



Fold change

$$\text{Fold change} = \frac{\text{Max Growth Xpt. Carbon}}{\text{Max Growth Neg. control}}$$

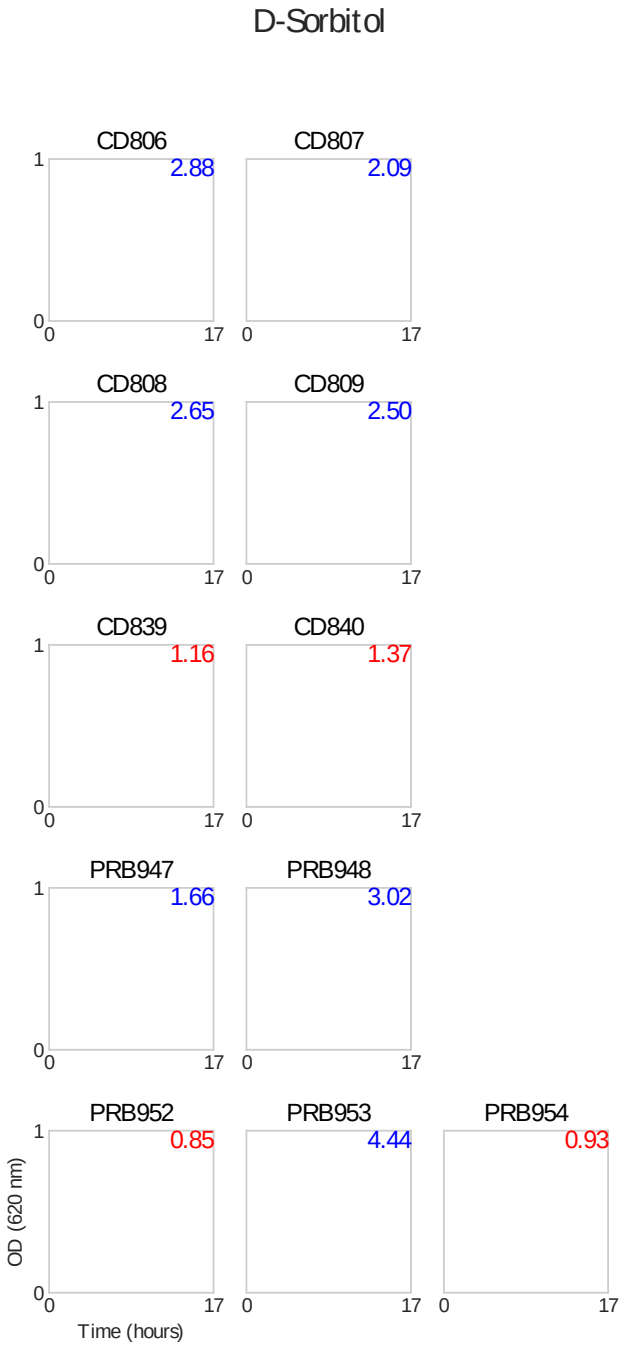
							First Test	Repeated	Repeated	Repeated	First Test	Repeated	Repeated	First Test	Repeated	Repeated	First test	Repeated					
	Growth Fold-Change	RT															Sym	Sym	Sym	Sym	Sym	Sym	S
		Strain					Savage	Savage	Savage	Savage	Britton	Britton	Britton	Britton	Britton	Britton	PRB802	PRB802	PRB803	PRB804	PRB805	PRB806	PR
Biolog Plate Summary		53	27	244	17	78	12	12	12	12	12	12	12	12	12	12	RT014	RT014	RT014	RT001	RT05	RT002	RT
	Xpt Compound	CD2048	CD2015	DL3110	M68	CD1015	CD630	CD630	CD630	CD630	CD630	CD630	CD630	CD630erm[S]	CD630erm[S]	CD630erm[S]	S0076	S0076	S0078	S0067	S0060	S0064	S0
PM1	L-Arabanose	0.8		1.2	0.8	0.6	1.2	0.8	1.2	1.1	0.8	1.1	1.1	0.9	0.9	0.9	0.7	1.2	0.8	1.1	0.9	0.8	
	N-Acetyl-D-Glucosamine	6.1	>1.5	3.4	3.5	2.3	4.2	1.6	2.4	4.1	4.3	4.1	4.0	3.2	3.6	3.2	4.4	2.8	3.3	4.0	4.1	2.9	
	D-Saccharic Acid	1.5		1.3	1.0	0.7	1.4	0.9	1.0	1.0	1.3	1.0	1.1	1.0	0.3	0.8	1.9	0.8	1.1	1.1	1.0	1.0	
	Succinic Acid	2.0		1.8	1.3	0.9	2.0	1.9	1.8	1.5	1.6	1.5	1.7	1.1	0.4	1.2	1.2	1.1	1.4	1.7	1.6	1.3	
	D-Galactose	1.2		1.3	1.0	0.6	0.9	0.7	1.1	0.9	0.9	0.9	1.1	0.8	0.4	1.0	0.9	1.5	1.0	1.1	1.1	0.9	
	L-Aspartic Acid	1.3		1.2	1.0	0.7	1.1	0.9	1.0	1.2	0.9	1.2	1.0	0.9	0.3	1.0	0.9	1.4	1.1	1.0	1.1	0.9	
	L-Proline	1.6		1.3	1.0	0.6	2.5	1.7	1.5	1.5	0.9	1.1	1.0	0.9	0.4	1.0	5.0	2.2	4.3	1.8	1.8	1.3	
	D-Alanine	1.2		0.9	0.8	0.5	0.8	0.9	1.0	1.1	0.8	1.1	1.0	0.9	0.3	1.0	5.1	1.7	0.9	1.0	1.1	0.8	
	D-Trehalose	1.0	>1.5	2.5	1.9	2.6	5.1	1.8	2.8	4.2	2.0	4.2	1.8	2.7	2.2	2.9	18.3	2.6	5.2	1.5	1.3	1.0	
	D-Manose	5.3	>1.5	3.3	3.0	2.2	3.9	1.9	2.6	4.6	4.3	4.6	3.1	2.7	2.4	2.0	6.2	2.3	1.8	3.1	2.9	2.5	
	Dulcitol	1.5		1.3	1.2	0.7	0.5	0.9	0.8	0.9	0.9	0.9	1.1	0.9	0.3	0.9	0.2	1.1	0.8	1.1	1.1	0.9	
	D-Serine	1.9		0.6	1.2	0.7	1.4	0.9	1.1	0.7	0.6	0.7	1.1	0.9	0.4	1.1	0.5	1.5	1.1	1.1	1.1	1.2	
	D-Sorbitol	5.6	>1.5	2.7	1.0	2.3	5.4	1.5	2.3	2.1	3.3	2.1	2.9	1.6	1.6	2.4	0.6	1.9	2.2	3.9	1.1	3.1	
	Glycerol	1.4		1.1	1.0	0.6	1.2	0.8	1.0	0.7	0.8	0.7	1.0	1.0	0.3	0.9	0.7	2.4	3.5	1.0	1.0	1.0	
	L-Fucose	1.5		1.0	1.0	0.6	1.3	0.7	1.0	0.6	0.7	0.6	1.1	0.8	0.3	0.8	0.6	0.8	1.0	1.1	1.0	0.9	
	D-Glucuronic Acid	1.2		1.0	1.0	0.6	1.3	0.7	1.0	0.6	0.8	0.6	1.0	0.9	0.3	0.9	0.6	2.0	1.0	1.0	1.0	0.9	
	D-Gluconic Acid	1.3		0.9	1.0	0.7	1.2	0.8	0.9	0.7	0.8	0.7	1.1	0.9	0.4	1.0	0.6	2.1	0.9	1.1	1.0	0.9	
	D,L-alpha-GlycerolPhosphate	1.3		1.1	1.0	0.7	1.3	0.7	0.9	0.7	0.8	0.7	1.0	0.9	0.5	1.0	0.8	0.9	1.2	1.0	1.0	0.8	
	D-Xylose	1.7		1.3	1.0	0.7	1.3	1.1	1.4	1.3	0.8	1.1	1.3	1.0	0.2	1.2	7.6	1.6	0.8	1.3	1.4	0.9	
	L-Lactic Acid	1.4		1.0	1.0	0.7	1.2	0.7	1.0	0.7	0.8	0.7	0.9	0.8	1.0	1.0	7.3	1.9	1.0	0.9	1.0	0.9	
	Formic Acid	1.5		1.2	0.9	0.7	1.2	0.7	1.0	0.9	0.6	0.9	1.1	0.9	0.4	1.0	5.2	1.9	1.1	1.1	1.1	0.9	
	D-Mannitol	5.8	>1.5	3.0	3.2	2.2	4.7	1.6	2.4	3.9	4.0	3.9	3.8	2.6	2.3	2.7	5.3	2.4	2.4	4.1	4.4	3.1	
	L-Glutamic Acid	1.5		1.1	1.1	0.7	1.2	0.8	0.7	0.7	0.6	0.7	1.0	0.9	0.3	1.0	0.3	1.3	0.9	1.0	1.1	1.0	
	D-Glucose-6-Phospate	1.1		0.6	1.0	0.7	1.1	0.7	0.8	0.7	0.6	0.7	1.0	0.9	0.5	1.0	1.2	0.9	1.0	1.0	1.1	1.0	
	D-Galactonic Acid-gamma-Lacton	1.3		0.8	0.9	0.6	1.2	0.8	0.9	0.6	0.5	0.6	1.0	0.8	0.7	0.9	0.9	1.4	0.9	1.0	1.0	0.9	
	D,L-Malic Acid	1.0		1.0	1.0	0.7	0.9	0.9	0.9	0.6	0.7	0.6	1.0	0.8	0.6	1.0	0.5	0.7	1.0	1.0	1.2	1.0	
	D-Ribose	1.3		1.9	0.9	0.7	1.7	1.6	1.6	1.7	2.0	1.7	1.8	1.0	0.3	1.1	2.5	2.0	1.0	1.2	0.9	0.9	
	Tween 20	1.3		0.9	0.9	0.7	1.0	0.6	0.9	0.5	0.9	0.5	1.0	0.7	1.0	0.8	0.5	0.9	0.9	1.0	1.0	1.2	
	L-Rhamnose	1.1		1.0	0.9	0.7	1.1	0.7	1.0	0.6	0.6	0.6	1.2	0.7	0.4	0.9	0.7	1.9	1.1	1.2	1.1	0.8	
	D-Fructose	6.5	> 1.5	3.3	3.7	2.4	4.2	1.6	2.6	4.4	4.3	4.4	4.3	3.0	2.2	3.3	0.4	2.5	3.1	4.3	4.5	3.1	
	Acetic Acit	1.0		1.3	0.9	0.7	0.7	0.7	0.9	0.7	0.6	0.7	1.0	0.8	0.8	0.9	5.8	1.6	1.0	1.0	1.1	0.9	

PM1 PM2 PM3

ss ESC to cancel Copy

Reproducibility

		Repeated	Repeated
	Sym	Sym	Sym
	PRB807	PRB807	PRB807
	RT002	**RT002**	**RT002**
Xpt Compound	**S0065**	**S0065**	**S0065**
L-Arabanose	**1.2**	**0.9**	**0.9**
N-Acetyl-D-Glucosamine	**3.1**	**3.2**	**2.8**
D-Saccharic Acid	**1.1**	**1.1**	**0.9**
Succinic Acid	**1.4**	**1.4**	**1.2**
D-Galactose	**1.0**	**1.1**	**0.9**
L-Aspartic Acid	**1.0**	**1.1**	**0.9**
L-Proline	**1.6**	**1.0**	**0.8**
D-Alanine	**1.0**	**1.0**	**0.8**
D-Trehalose	**1.3**	**2.3**	**1.9**
D-Manose	**3.2**	**2.4**	**2.0**
Dulcitol	**1.0**	**1.1**	**0.9**
D-Serine	**1.2**	**1.2**	**1.0**
D-Sorbitol	**3.6**	**2.3**	**1.9**
Glycerol	**1.0**	**1.1**	**1.0**
L-Fucose	**1.0**	**1.1**	**0.9**
D-Glucuronic Acid	**0.9**	**1.1**	**1.0**
D-Gluconic Acid	**1.1**	**1.1**	**0.9**
D,L-alpha-GlycerolPhosphate	**1.0**	**1.1**	**0.9**
D-Xylose	**1.2**	**1.2**	**1.3**
L-Lactic Acid	**1.0**	**1.1**	**0.9**
Formic Acid	**1.0**	**1.1**	**0.9**
D-Mannitol	**1.0**	**1.1**	**0.9**
L-Glutamic Acid	**3.3**	**2.8**	**2.2**
D-Glucose-6-Phospate	**1.1**	**1.1**	**0.9**
D-Galactonic Acid-gamma-Lactone	**1.1**	**1.1**	**0.9**
D,L-Malic Acid	**1.1**	**1.1**	**1.0**
D-Ribose	**1.1**	**1.1**	**0.9**
Tween 20	**1.1**	**1.2**	**1.3**
L-Rhamnose	**1.0**	**1.0**	**0.8**
D-Fructose	**1.0**	**1.0**	**0.9**
Acetic Acit	**3.4**	**3.3**	**2.7**
alpha-D-Glucose			



Main objectives of data workflow

Comparative analysis:

Main objectives of data workflow

Comparative analysis:

(i) Growth on same carbon substrate across different genotypes.

Main objectives of data workflow

Comparative analysis:

(i) Growth on same carbon substrate across different genotypes.

(ii) Growth of isolates from same genotype on different carbon substrates.

Main objectives of data workflow

Comparative analysis:

(i) Growth on same carbon substrate across different genotypes.

(ii) Growth of isolates from same genotype on different carbon substrates.

General questions:

1. *Does each isolate have a unique carbon substrate utilization profile?*

Main objectives of data workflow

Comparative analysis:

(i) Growth on same carbon substrate across different genotypes.

(ii) Growth of isolates from same genotype on different carbon substrates.

General questions:

1. *Does each isolate have a unique carbon substrate utilization profile?*
2. *Do isolates of the same genotype grow on the same carbon substrates?*

Main objectives of data workflow

Comparative analysis:

(i) Growth on same carbon substrate across different genotypes.

(ii) Growth of isolates from same genotype on different carbon substrates.

General questions:

1. *Does each isolate have a unique carbon substrate utilization profile?*
2. *Do isolates of the same genotype grow on the same carbon substrates?*
3. *Is there strong correspondence between genotype and phenotype?*

Main objectives of data workflow

Comparative analysis:

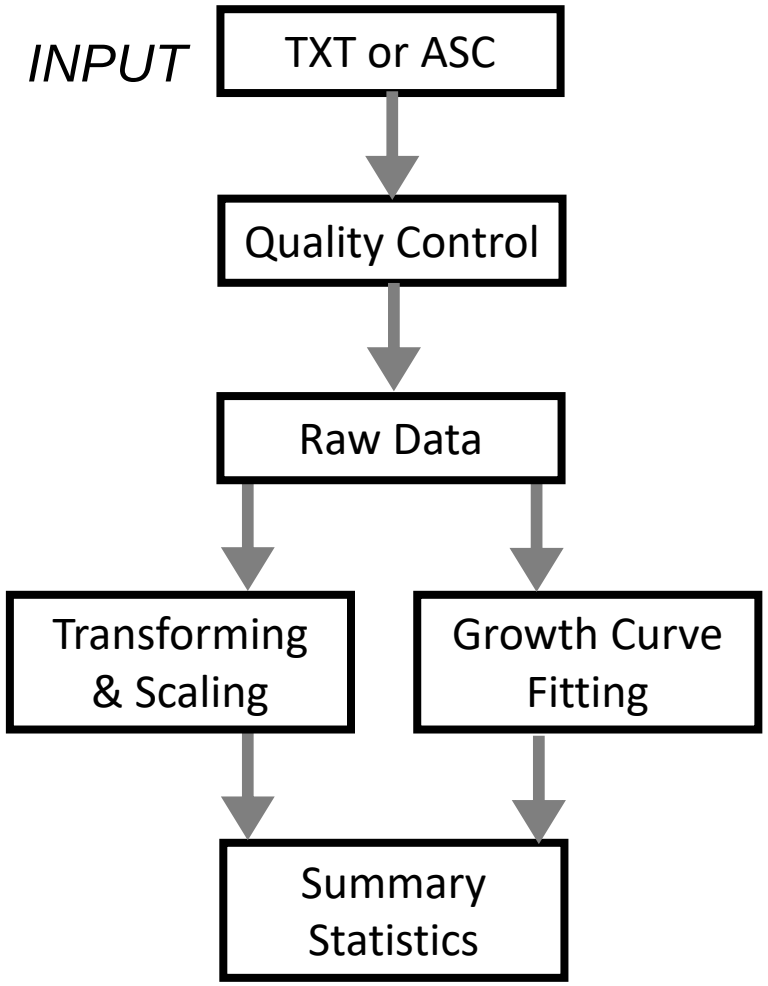
(i) Growth on same carbon substrate across different genotypes.

(ii) Growth of isolates from same genotype on different carbon substrates.

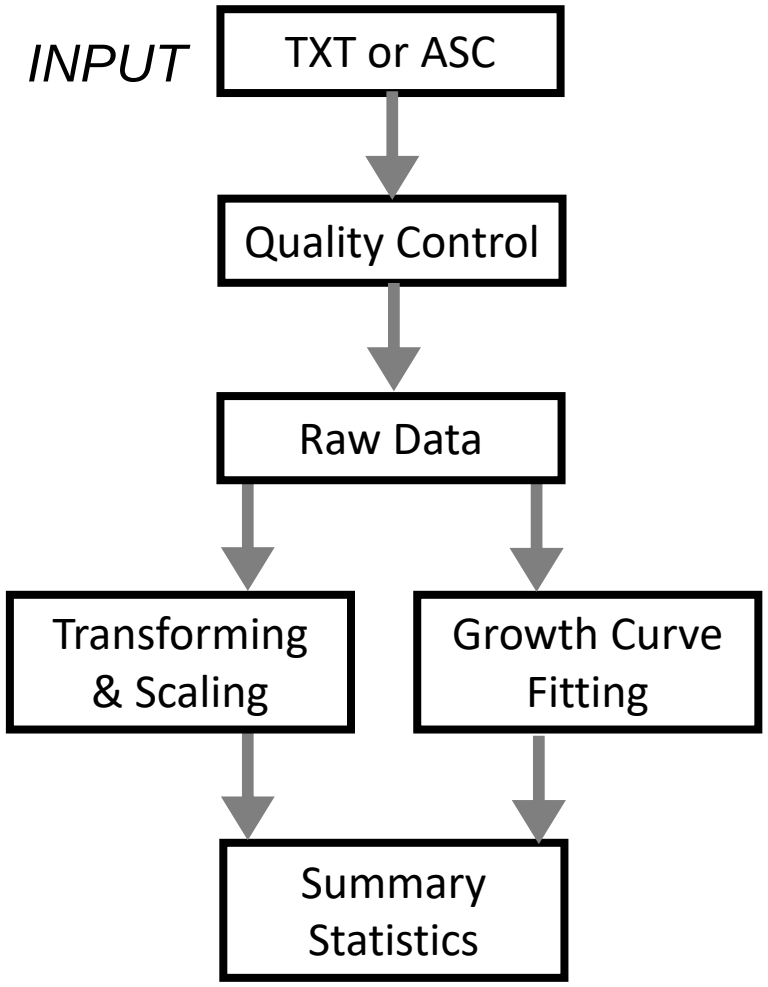
General questions:

1. *Does each isolate have a unique carbon substrate utilization profile?*
2. *Do isolates of the same genotype grow on the same carbon substrates?*
3. *Is there strong correspondence between genotype and phenotype?*
4. *Which substrates may be driving adaptation in C. difficile populations?*

Overview of data processing

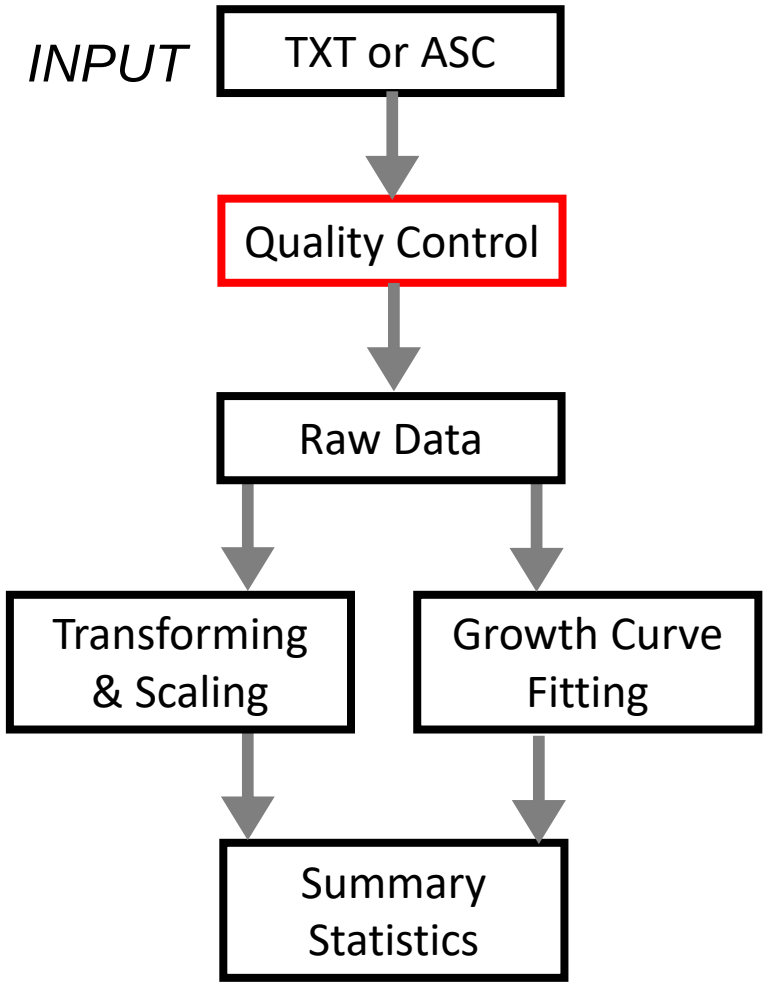


Overview of data processing

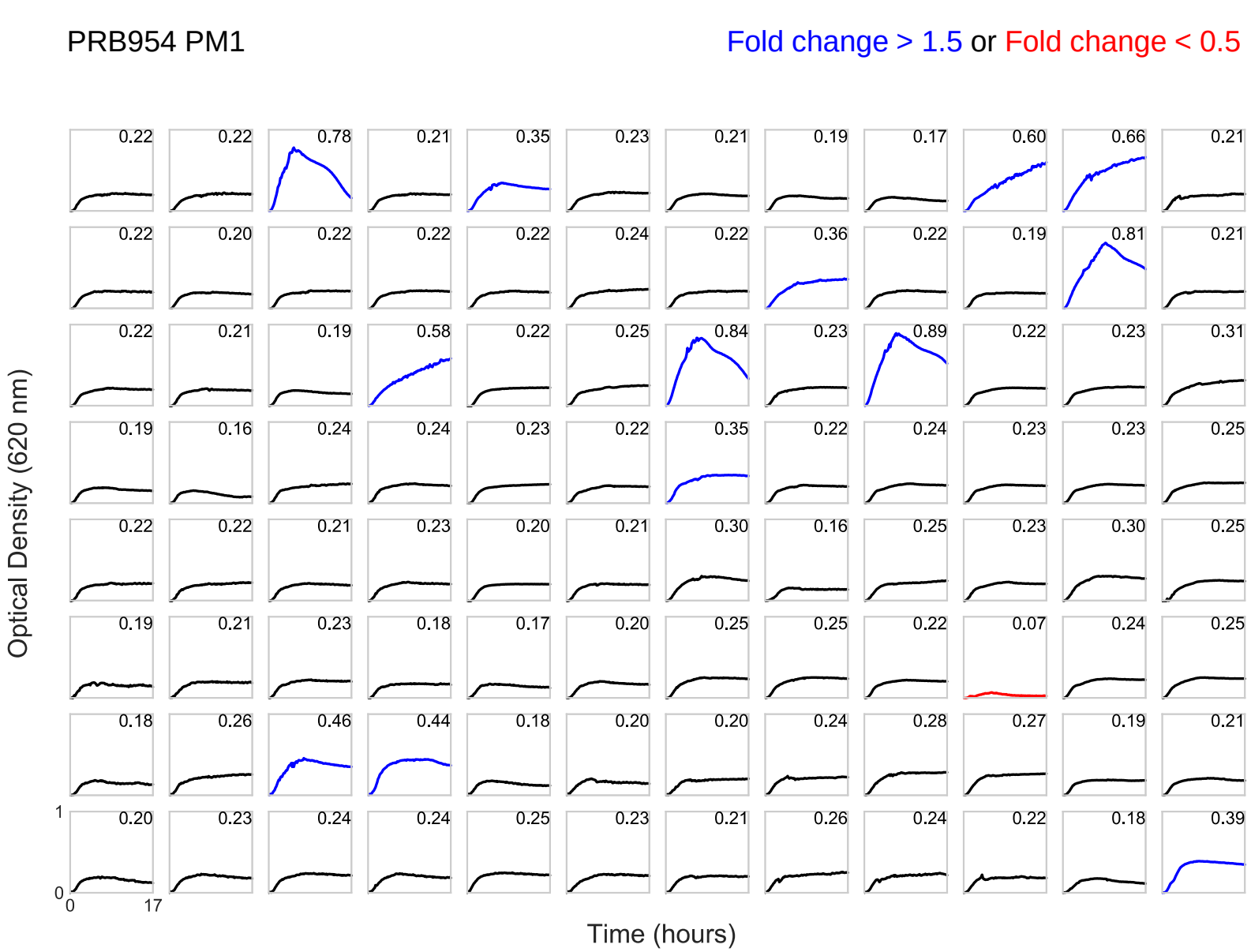


Simple implementation is possible with a scientific computing language (e.g. Python or R) and/or a spreadsheet (e.g. Microsoft Excel).

Overview of data processing



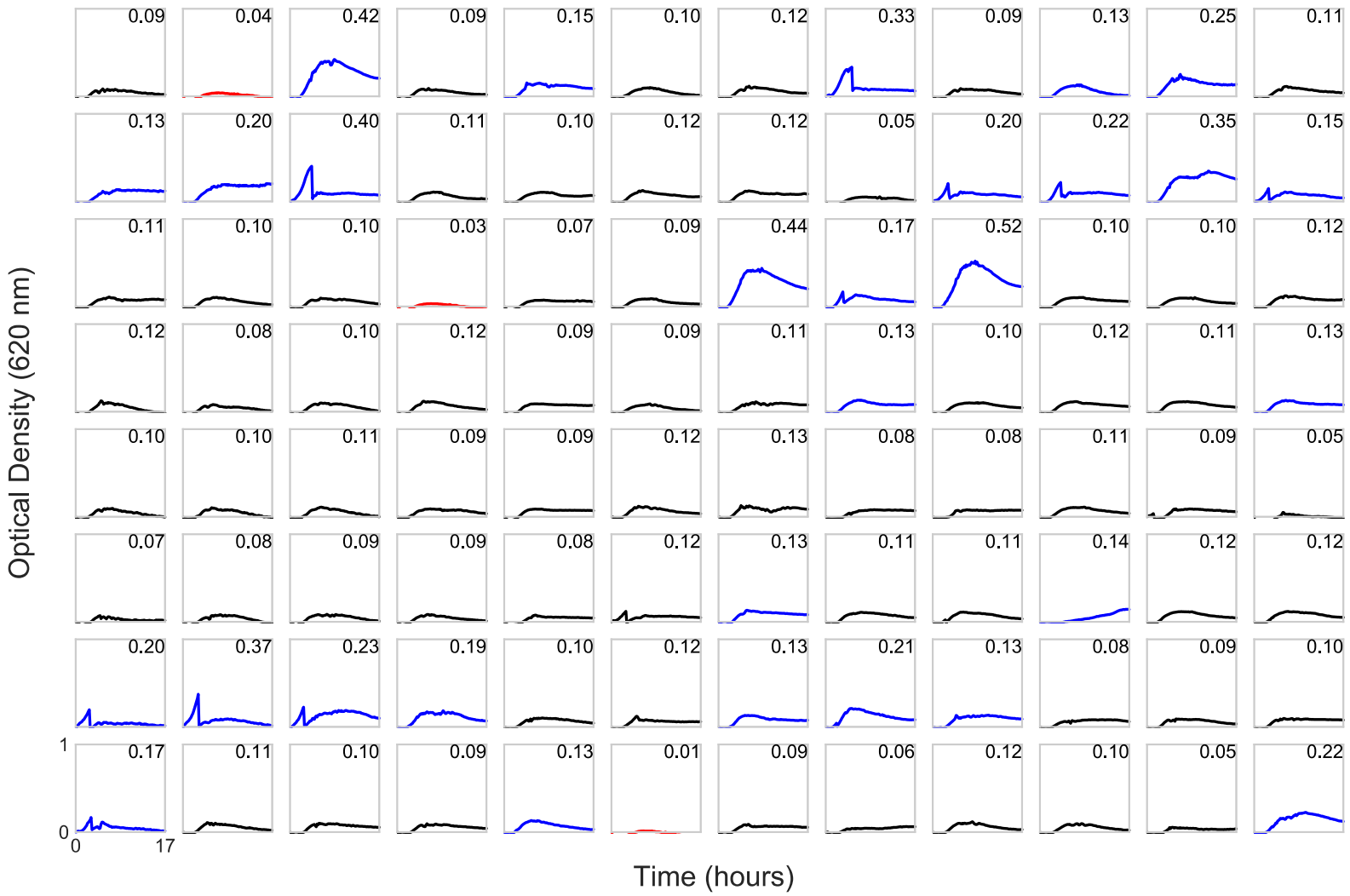
Biolog Phenotype Microarrays capture utilization of carbon substrates.



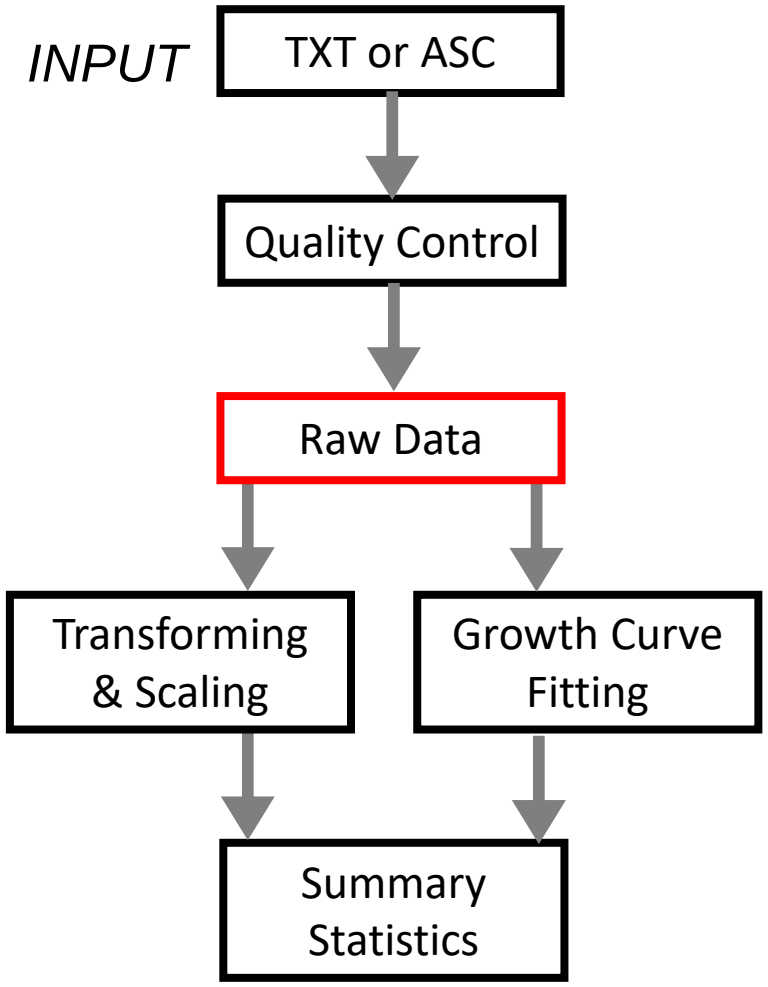
Manual and automated quality checks are necessary!

CD801 PM1

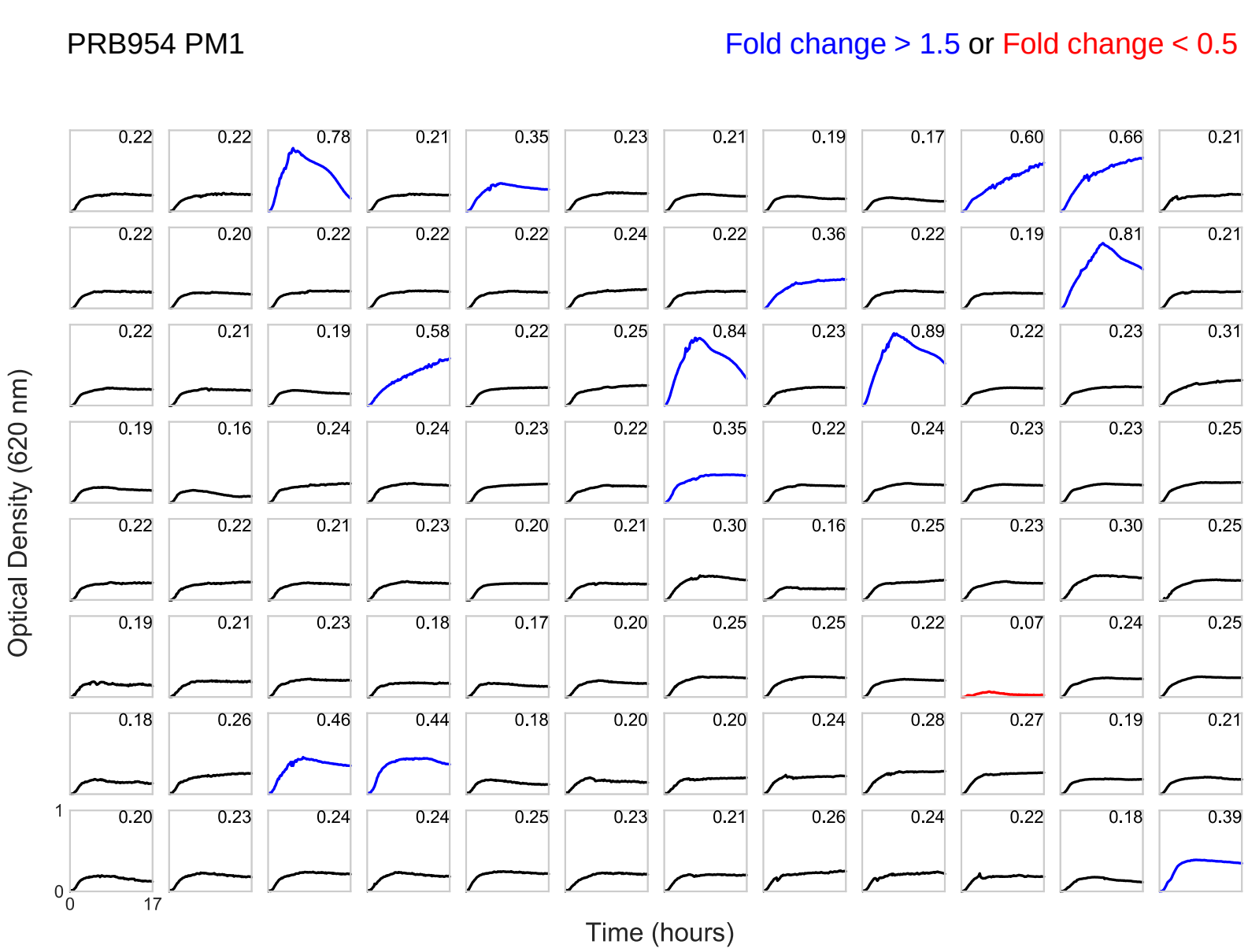
Fold change > 1.5 or Fold change < 0.5



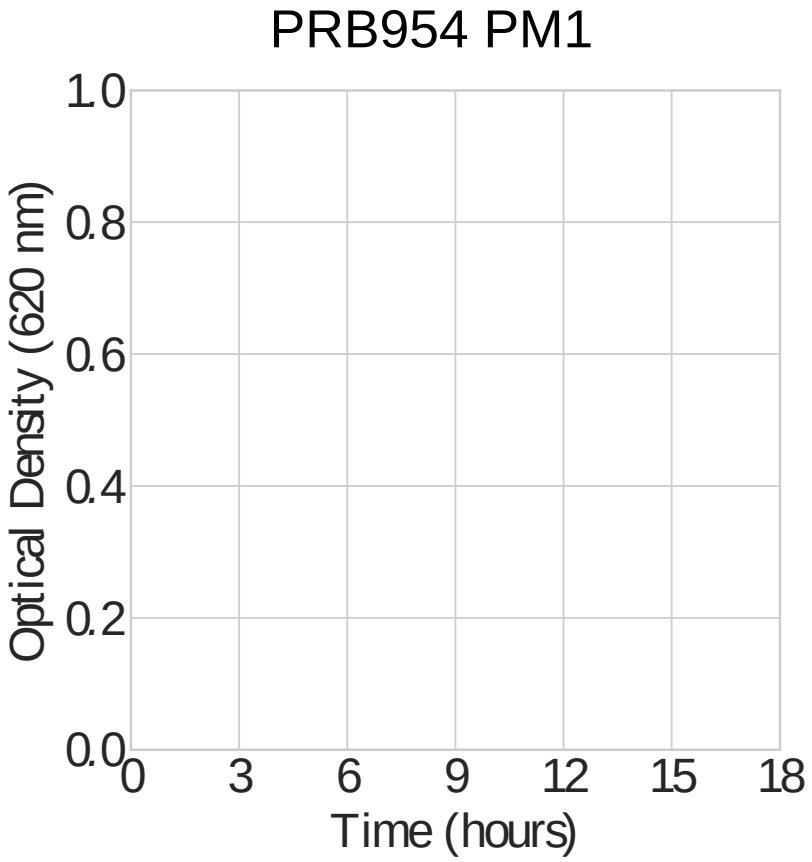
Overview of data processing



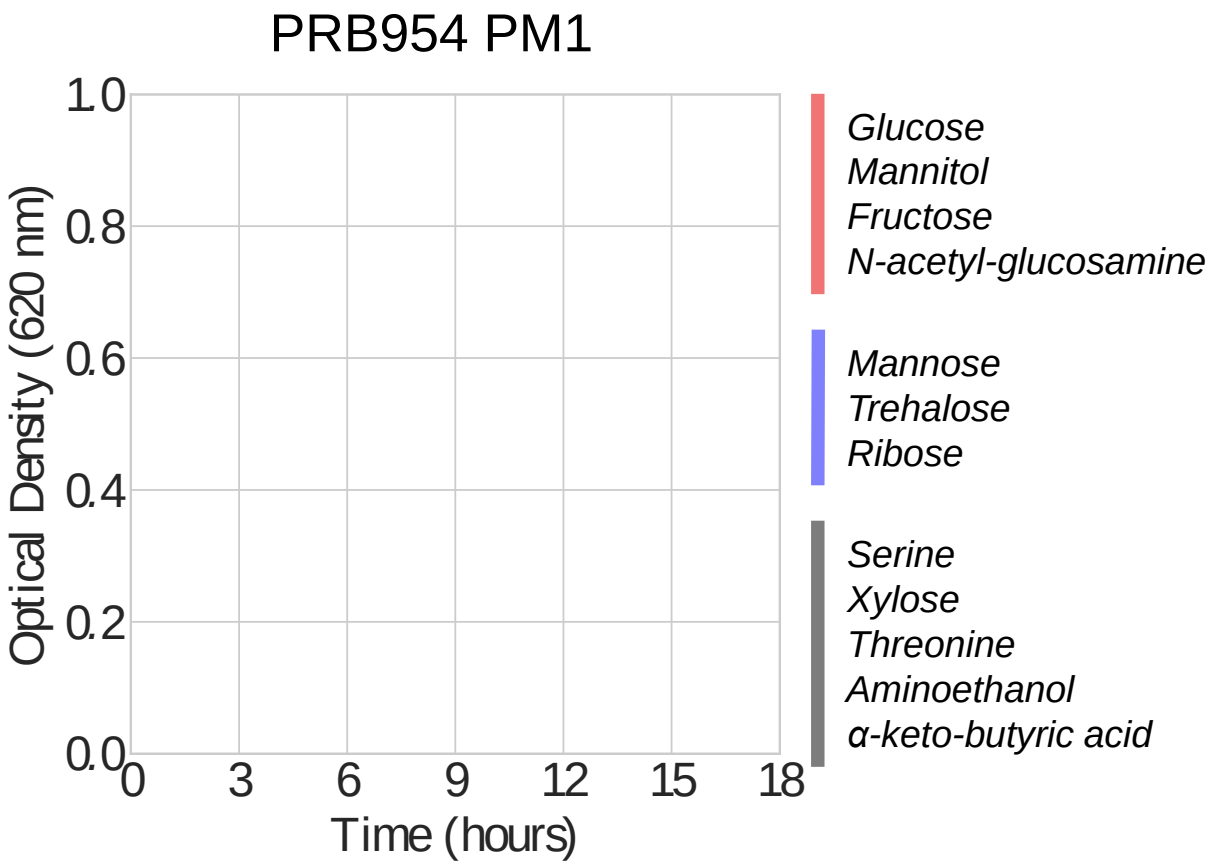
Biolog Phenotype Microarrays capture utilization of carbon substrates.



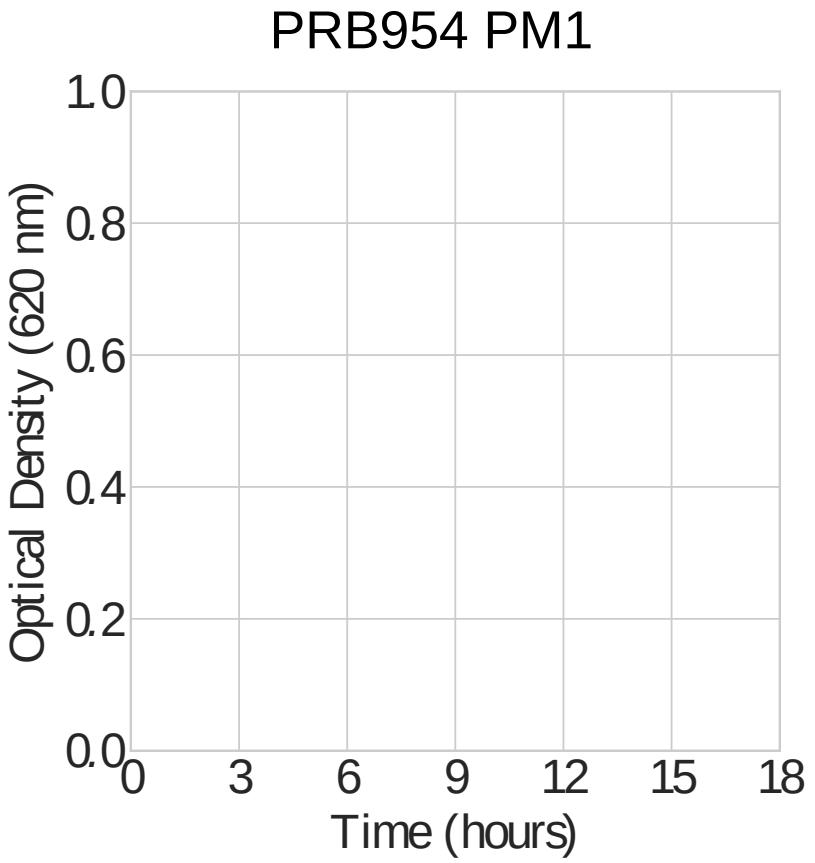
Clinical isolates exhibit varying growth dynamics on different carbon substrates

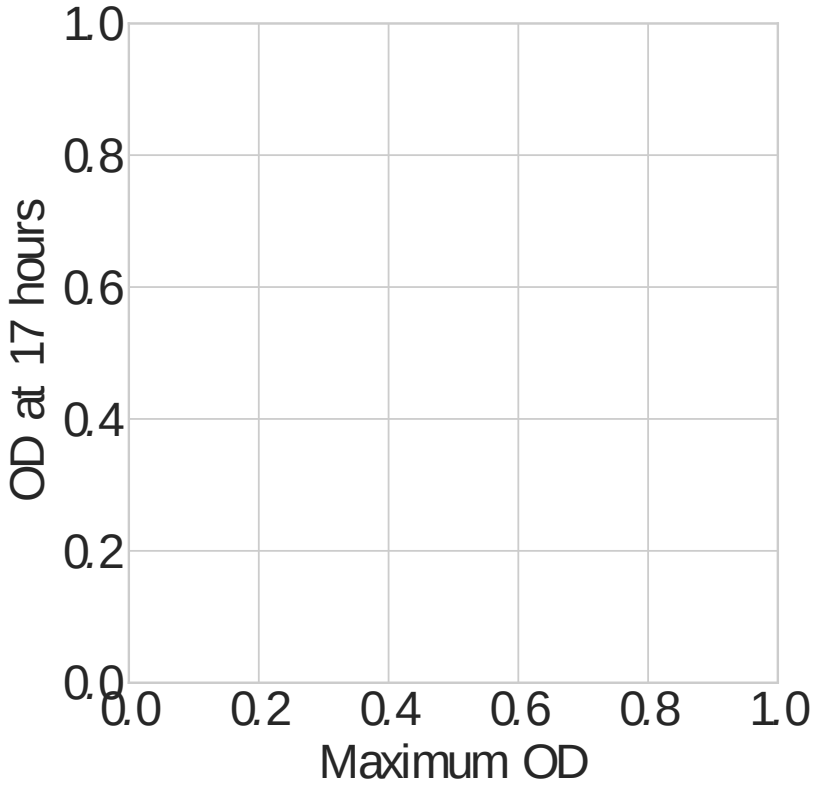


Clinical isolates exhibit varying growth dynamics on different carbon substrates

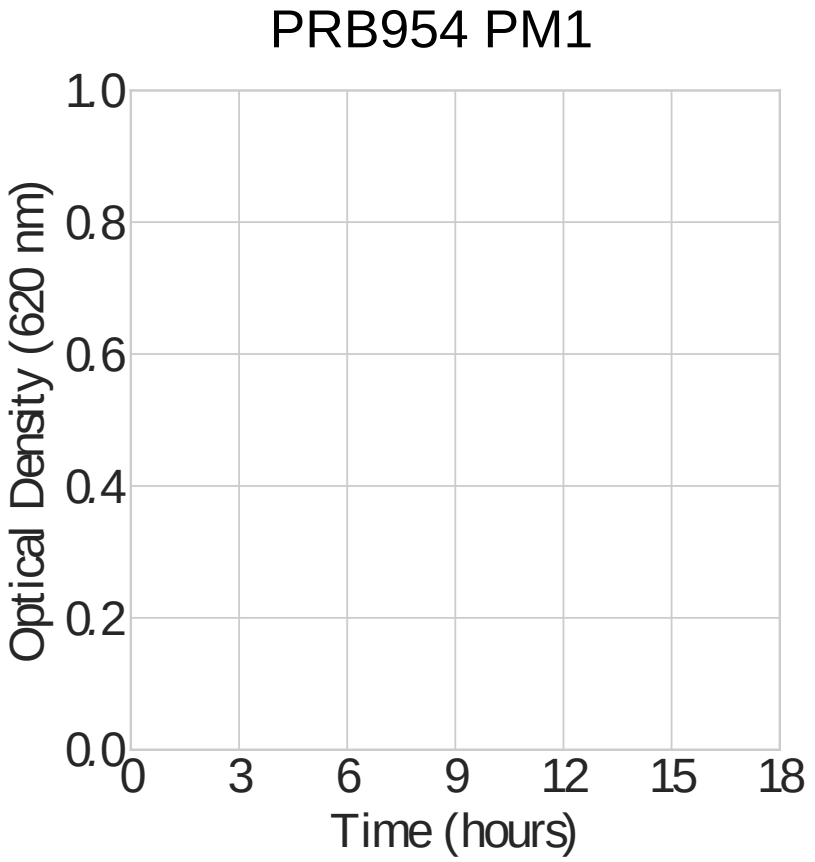


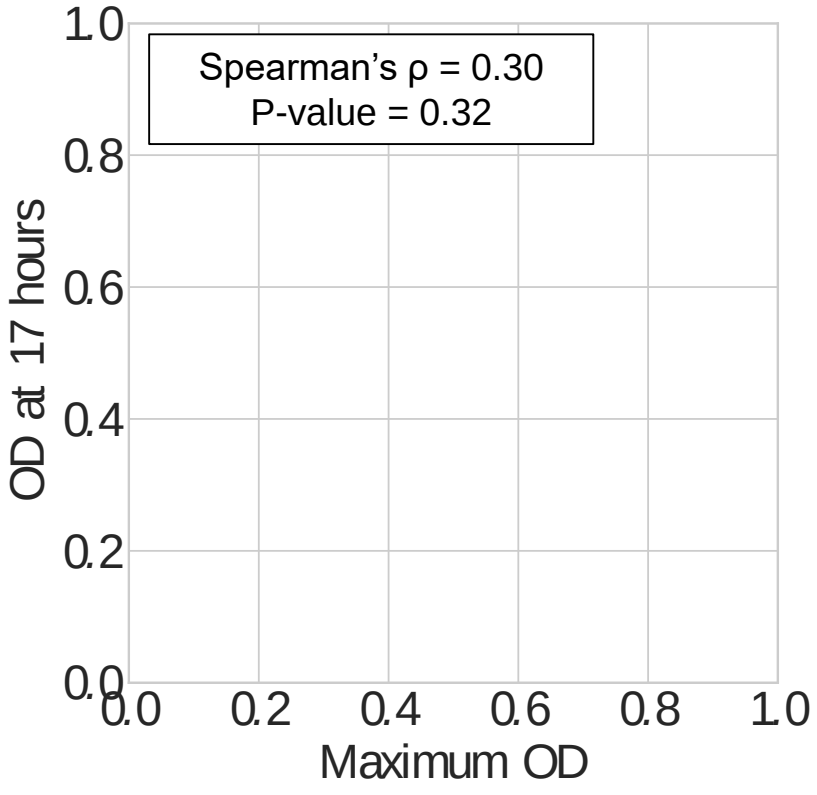
Final time points do not reflect the true relative fitness of a carbon substrate.





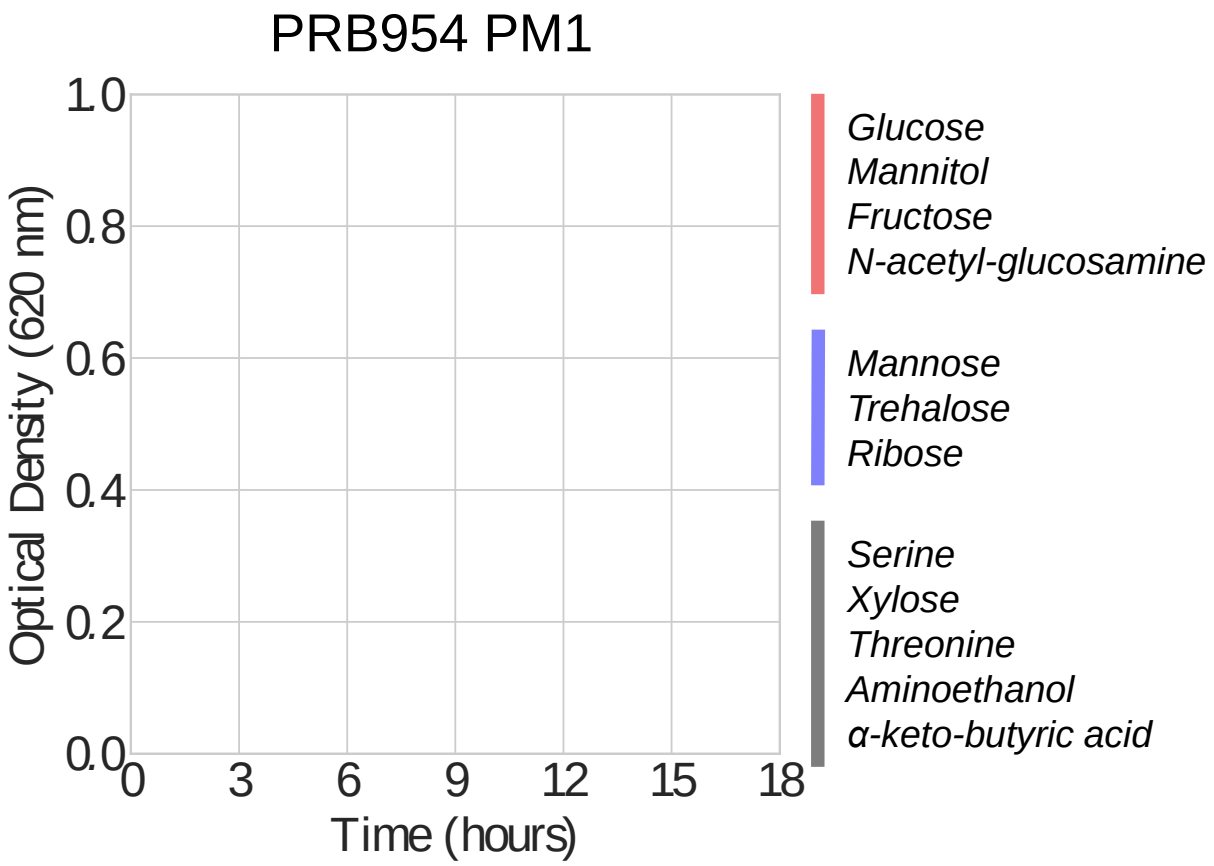
Final time points do not reflect the true relative fitness of a carbon substrate.



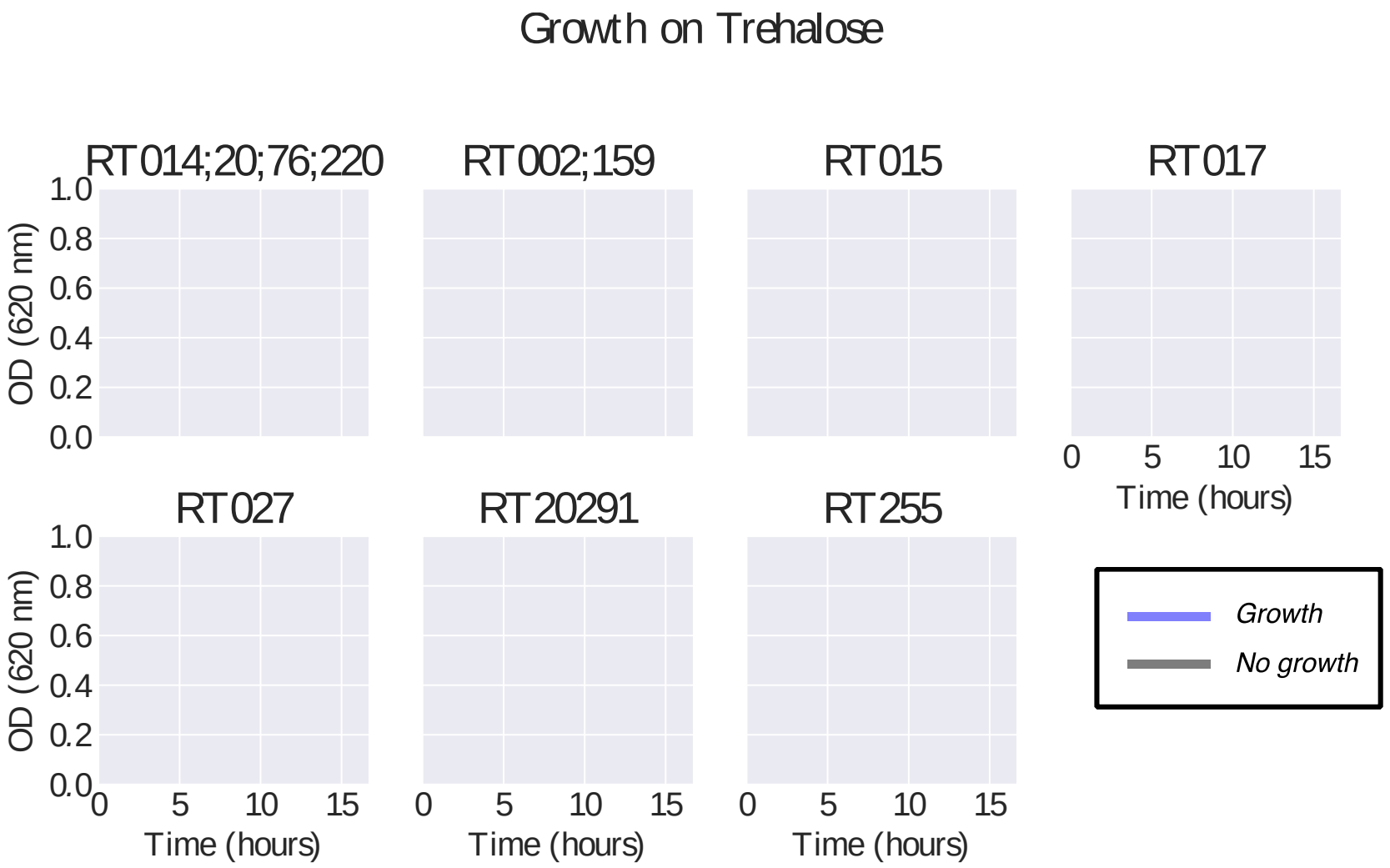


There is clear trade-off between assay throughput and biological insight!

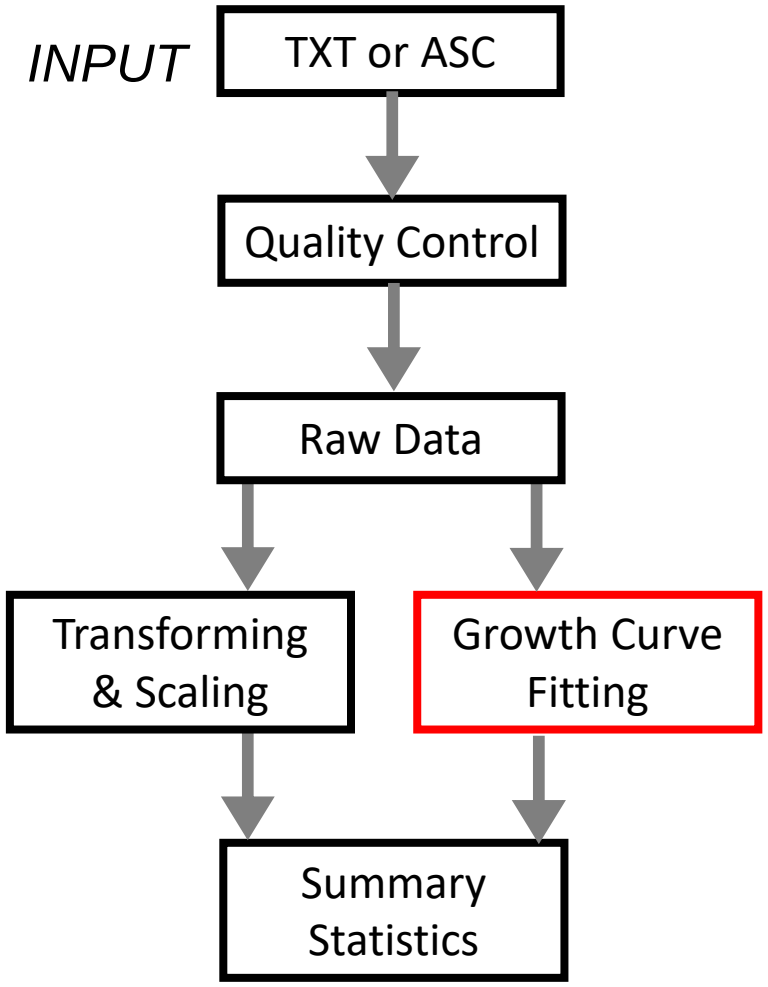
Clinical isolates exhibit varying growth dynamics on different carbon substrates



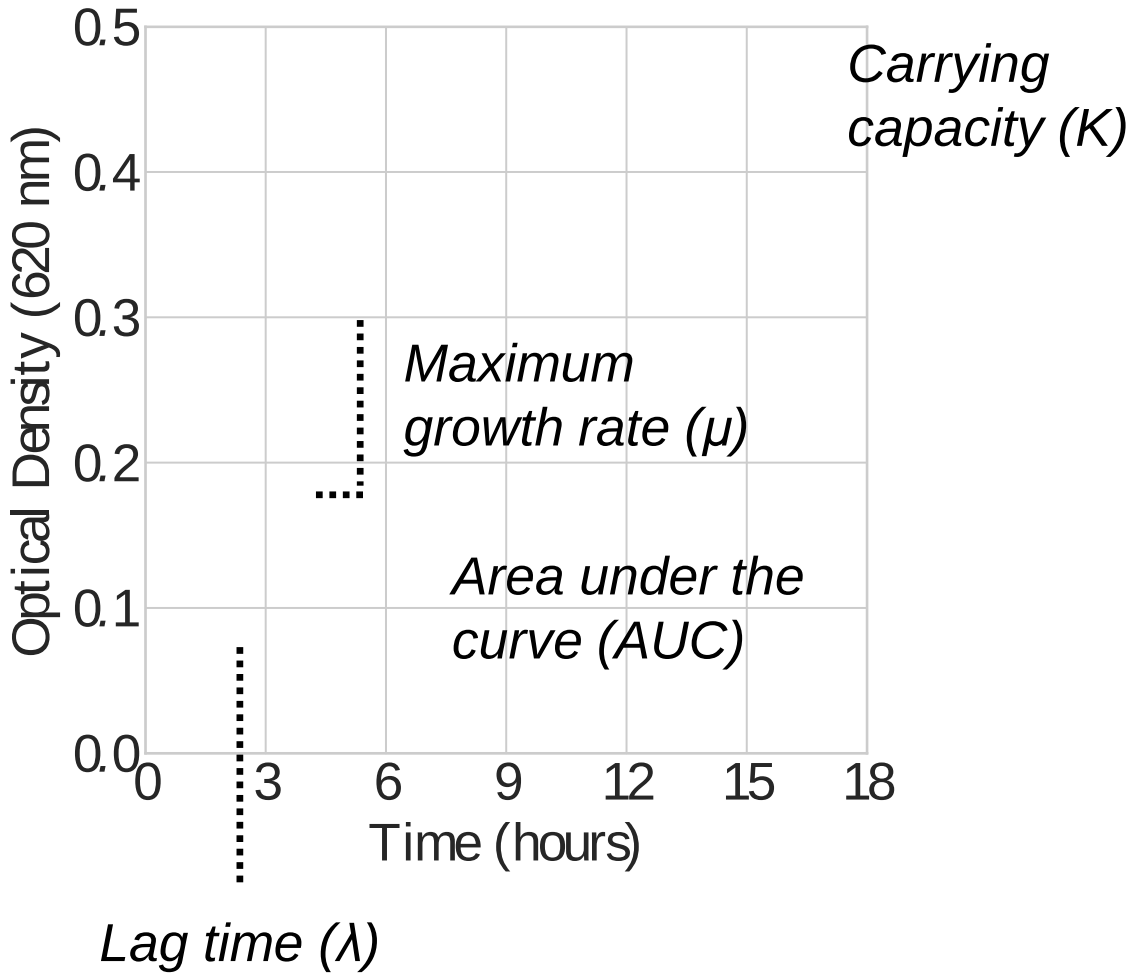
Varying growth dynamics occur also on the same carbon substrate



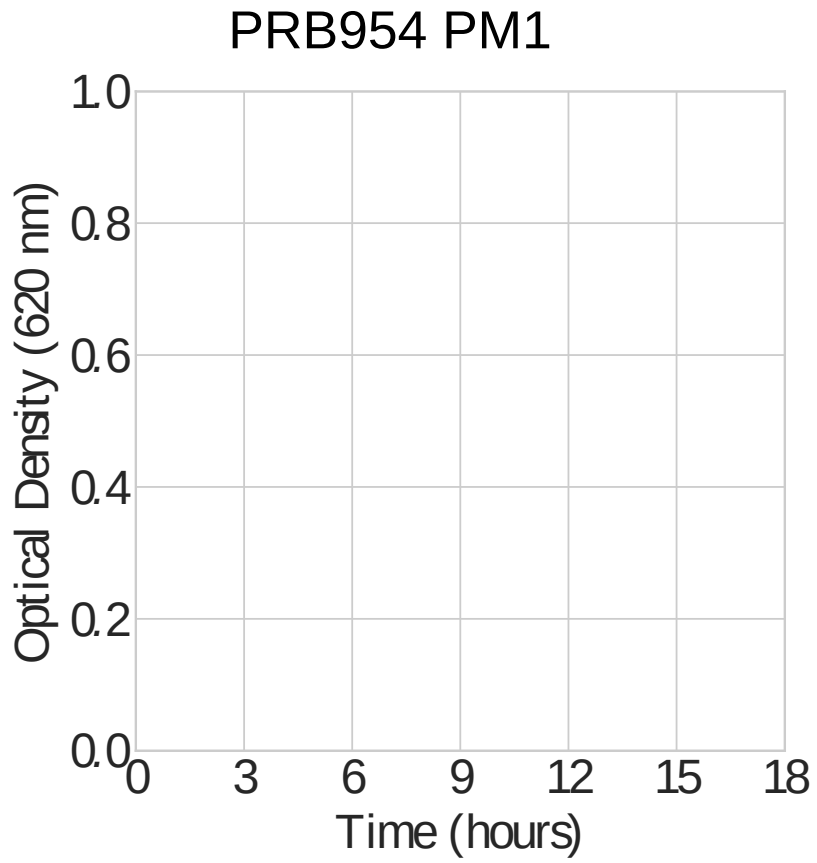
Overview of data processing



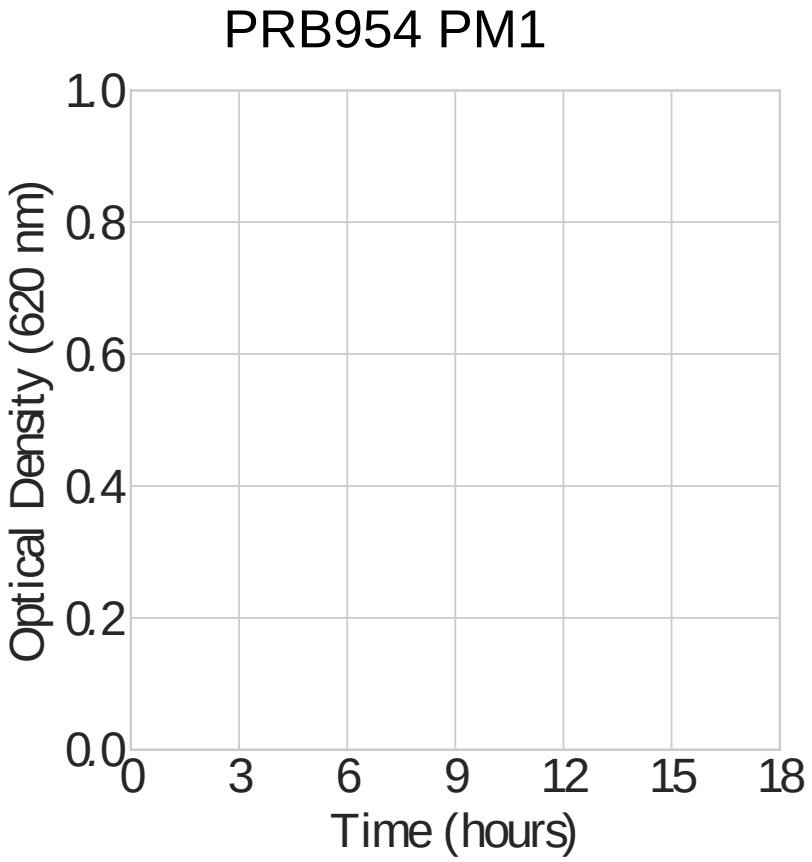
Growth curve dynamics are more informative then final OD measurement.

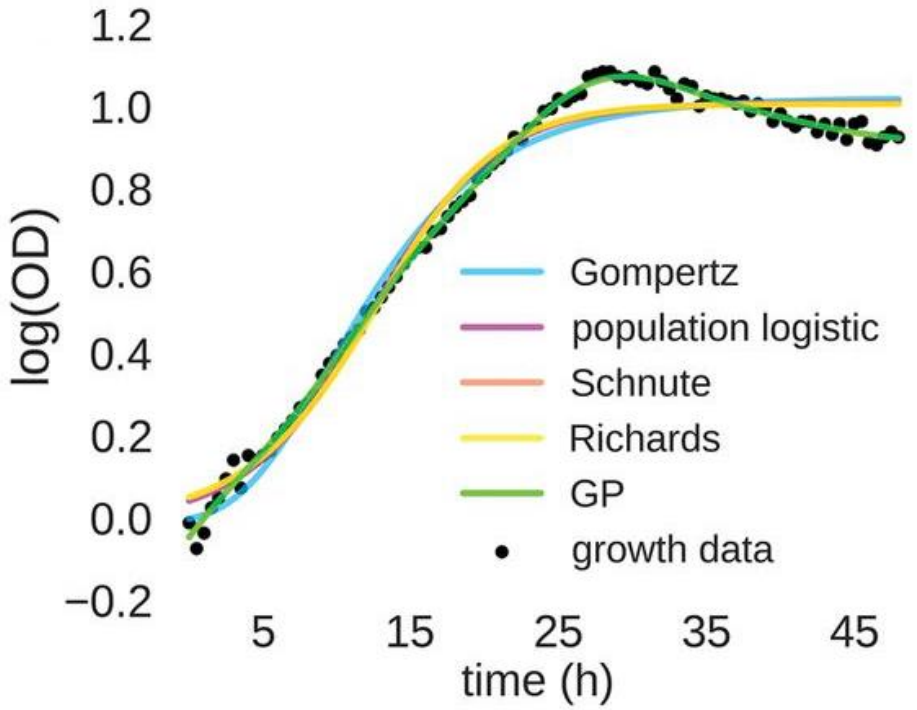


A non-parametric approach for growth curve fitting may be useful.



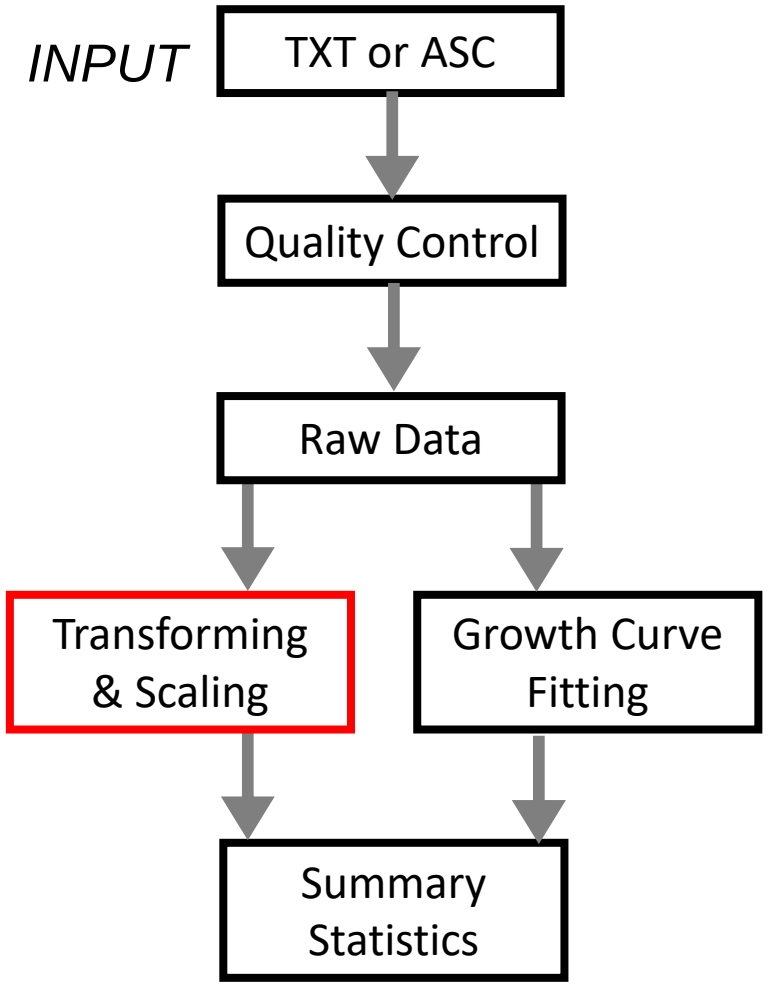
A non-parametric approach for growth curve fitting may be useful.



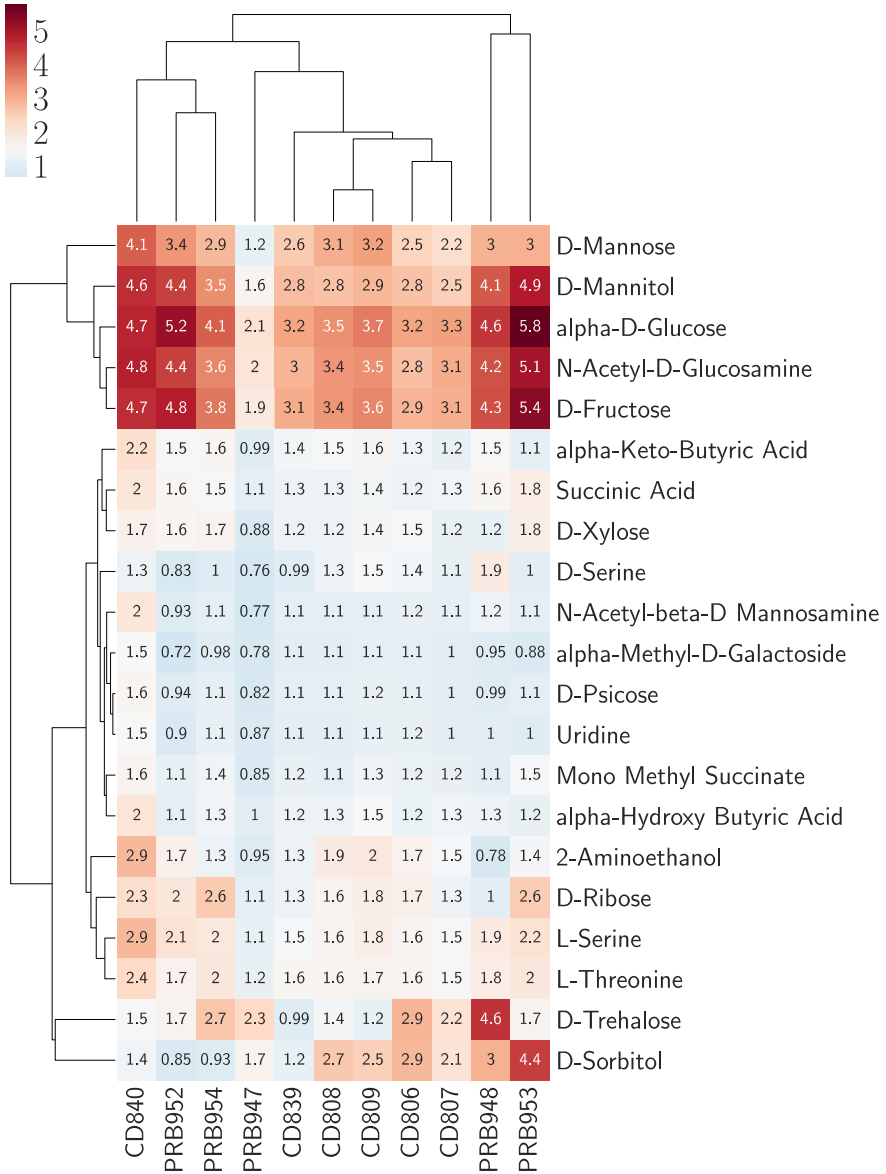


Tonner et al. *Genome Biology*. 2017.

Overview of data processing

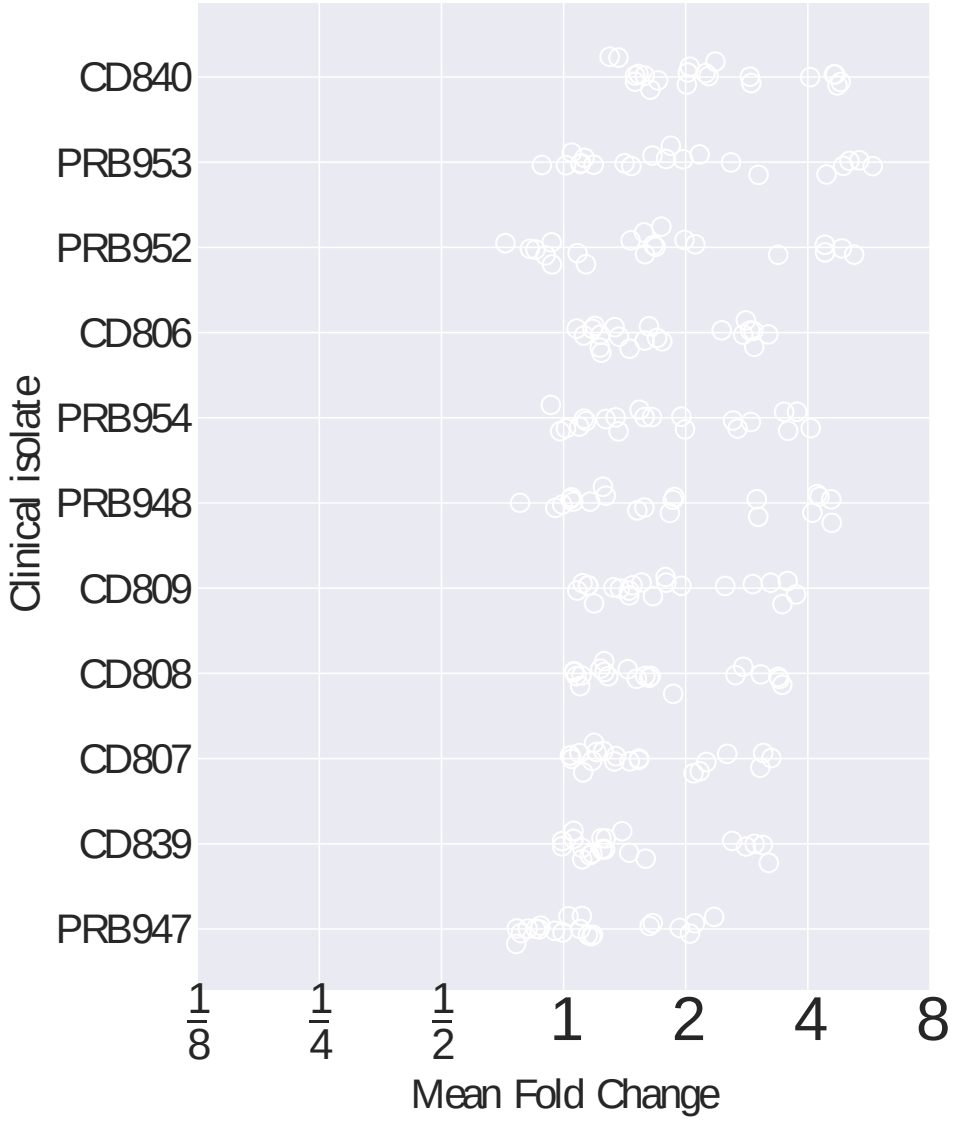


Baseline growth of different genotypes can vary and bias downstream analysis.



Subset of carbon substrates that are utilized by at least two isolates.

Baseline growth of different genotypes can vary and bias downstream analysis.



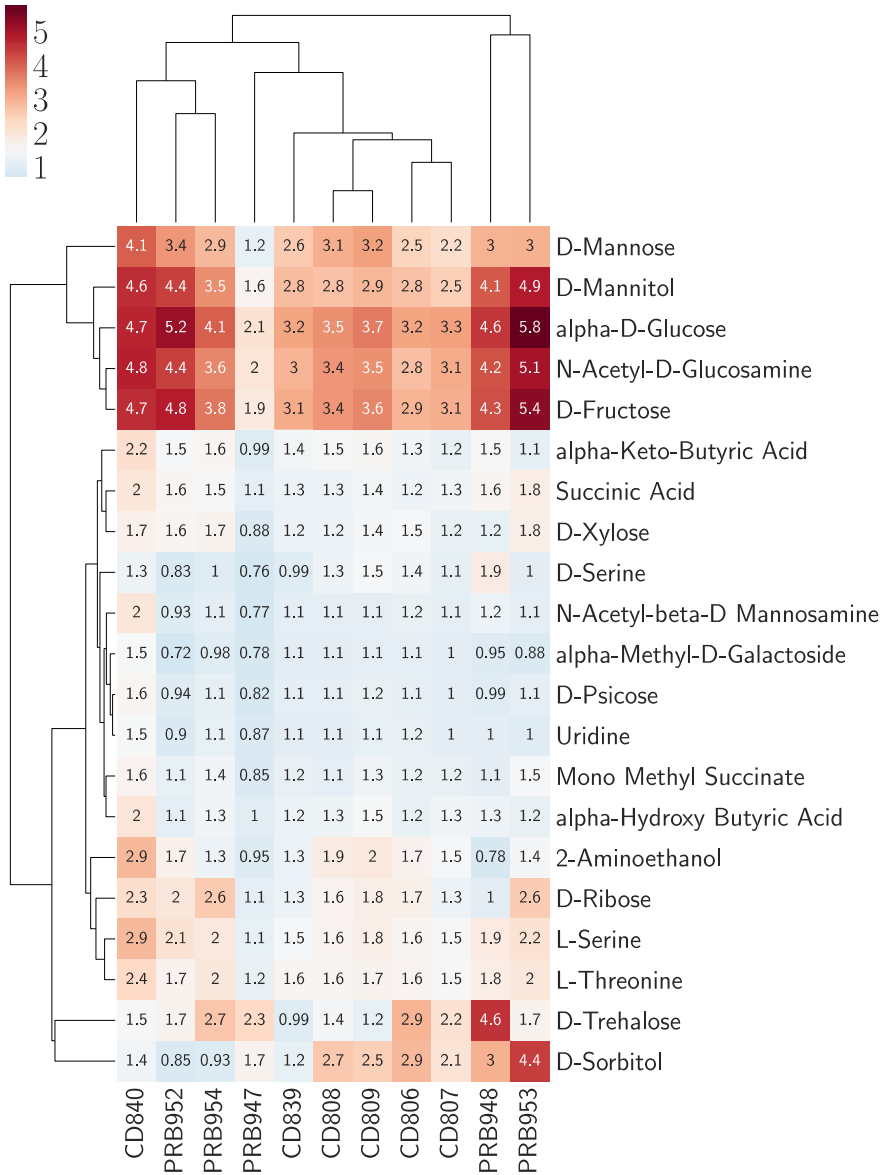
Subset of carbon substrates that are utilized by at least two isolates.

Instead of a fold change that can range between 0 and ∞, we can transform and scale the data for example into z-scores.

$$Z_c = \frac{(F_c - \mu_g)}{\sigma_g}$$

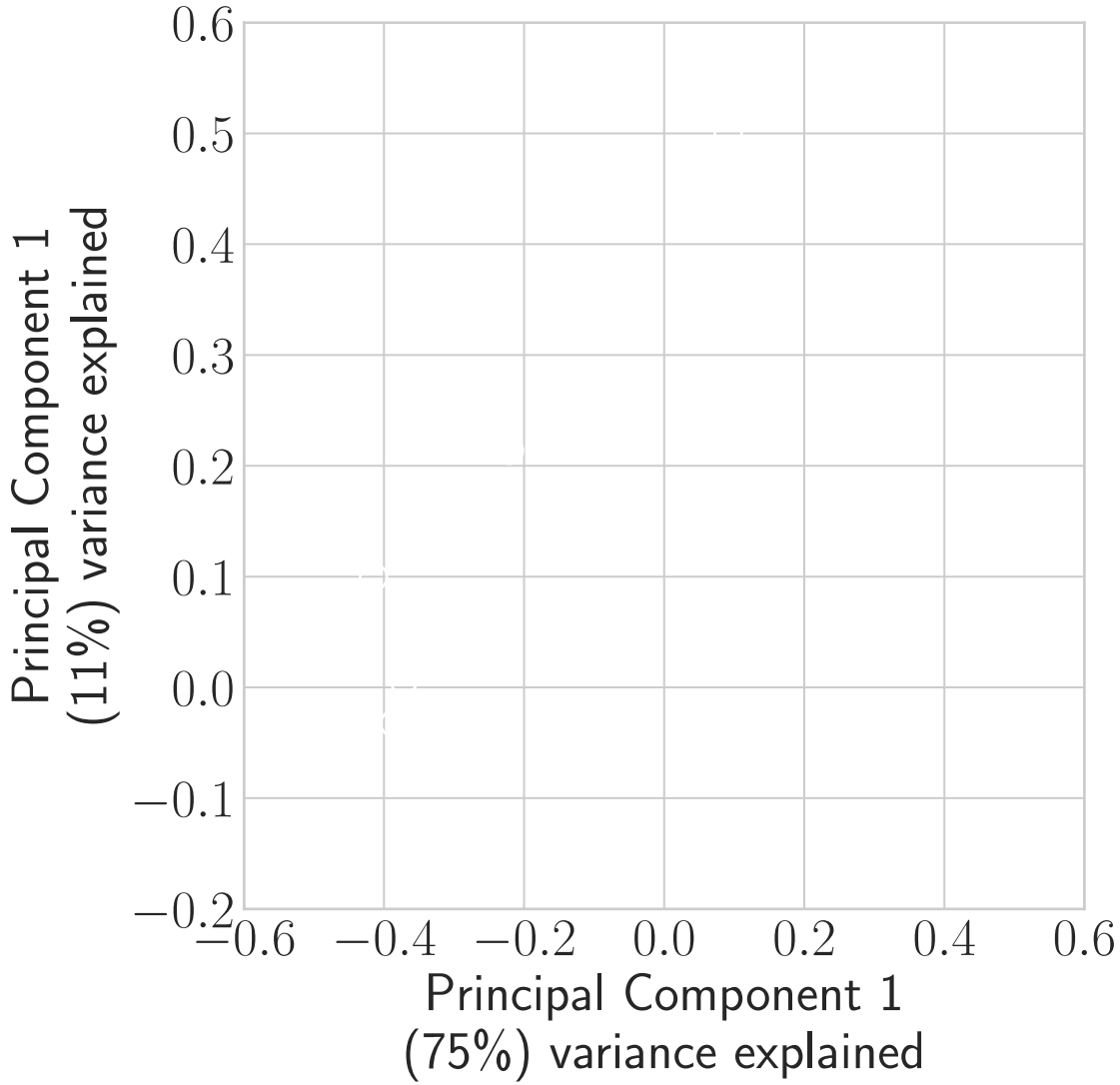
where F_c *is the fold change on a specific carbon, while* μ_g *and* σ_g *are the mean and standard deviation, respectively, of fold changes for a specific genotype.*

How well does phenotype correspond to genotype?

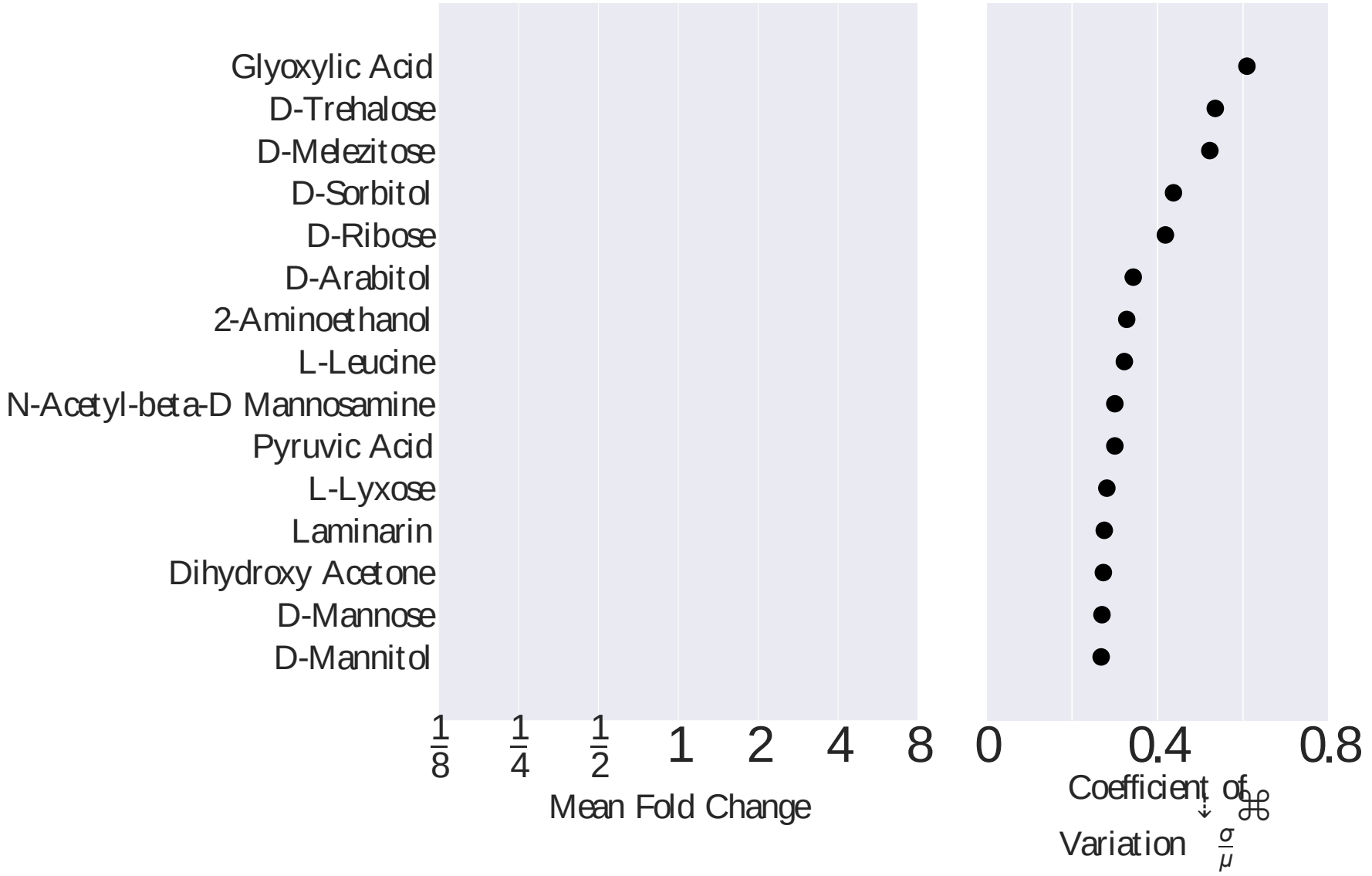


Subset of carbon substrates that are utilized by at least two isolates.

How well does phenotype correspond to genotype?

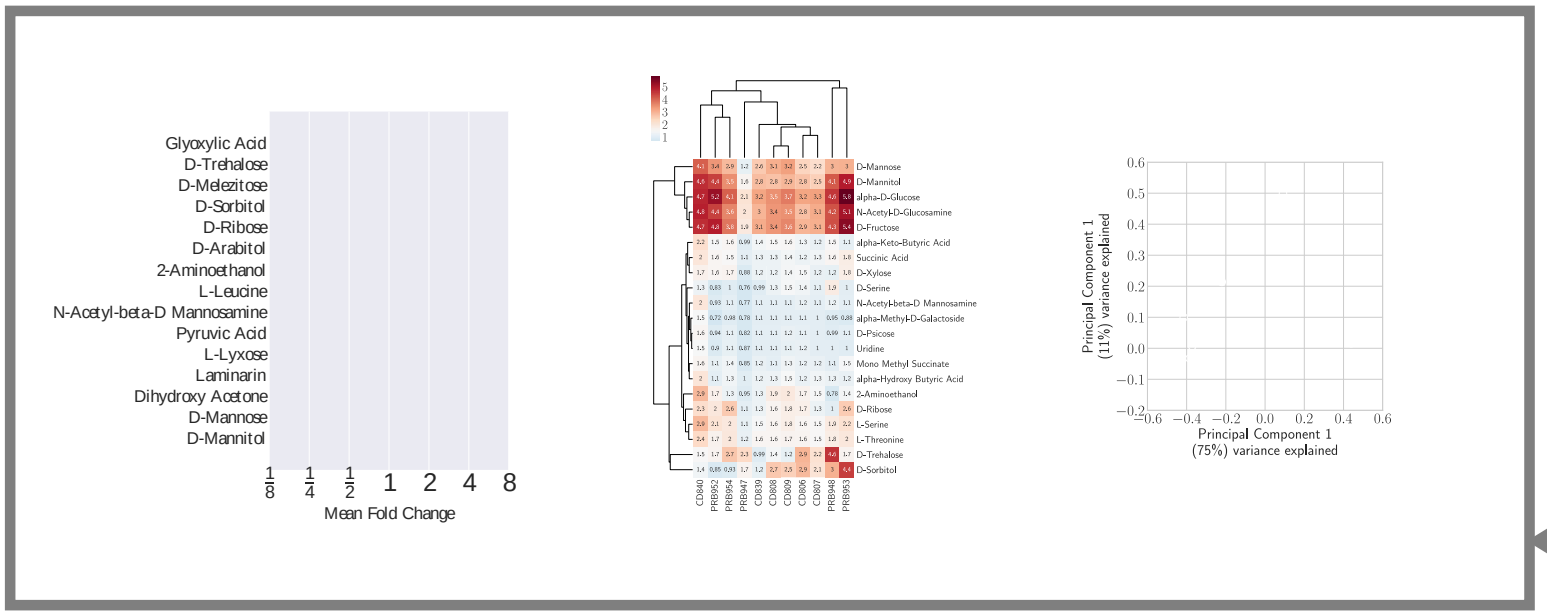


Which substrates may explain most of the variation between the different isolates?



Feedback is welcome on how improvements to data workflow can better address questions & objectives.

Many ways to analyze the data, but outcome strongly depends on the input and data manipulation.



INPUT

TXT or ASC

Quality Control

Raw Data

Transforming & Scaling

Growth Curve Fitting

Summary Statistics

Feedback is welcome on how improvements to data workflow can better address questions & objectives.

Comparative analysis:

(i) Growth on same carbon substrate across different genotypes.

(ii) Growth of isolates from same genotype on different carbon substrates.

General questions:

1. *Does each isolate have a unique carbon substrate utilization profile?*
2. *Do isolates of the same genotype grow on the same carbon substrates?*
3. *Is there strong correspondence between genotype and phenotype?*
4. *Which substrates may be driving adaptation in C. difficile populations?*