> METAFLUX

A method for predicting metabolic fluxes by integrating proteomic data into a genome-scale constraint-based metabolic model

Maëla Sémery - PhD Student (3rd year)

Team BASE, UMR GQE-Le Moulon, IDEEV, Gif-sur-Yvette

PhD Supervisors

- Christine DILLMANN
- Mélisande BLEIN-NICOLAS

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

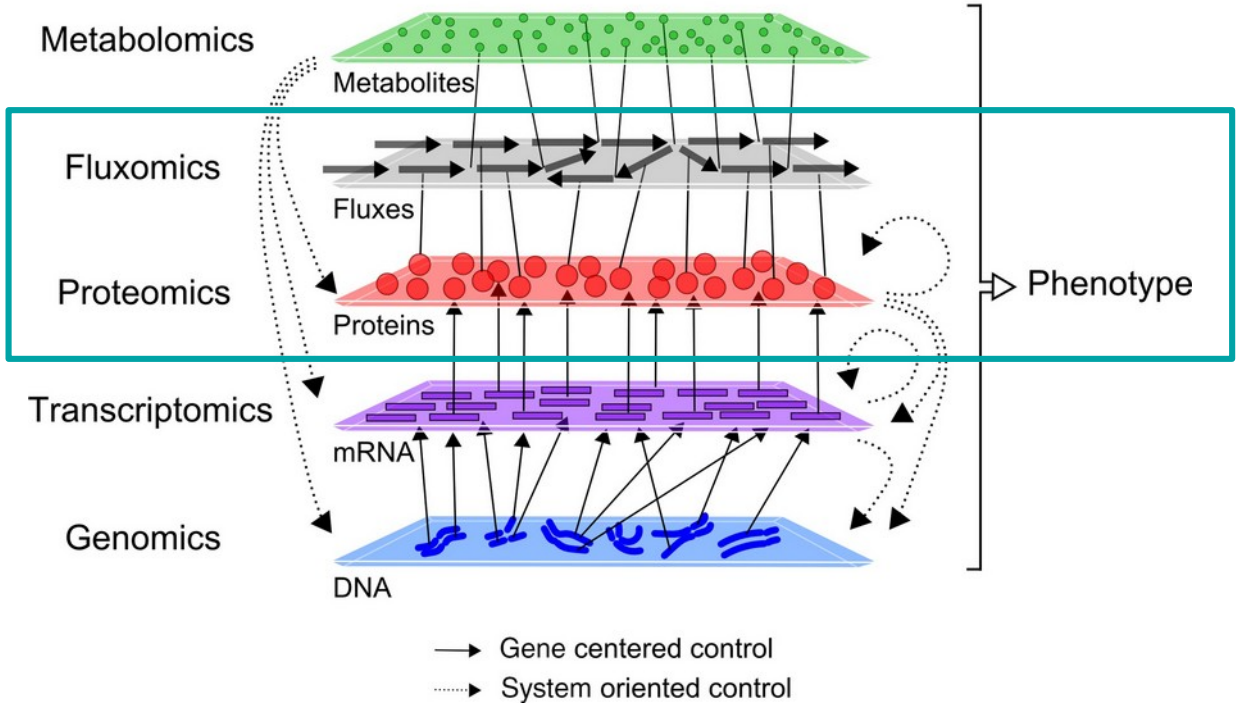
INRAE

université PARIS-SACLAY

AgroParisTech
Talents d'une planète soutenable

Génétique Quantitative et Évolution Le Moulon

How can we improve the understanding of response mechanisms to different environmental conditions?



Different omics scales to describe phenotype
(Peyraud *et al.* (2017) ; doi: 10.1111/tpj.13429)

Well-Watered (WW) | Water Deficit (WD)

Adapted from Prasanna *et al.* (2021)
doi: 10.1007/s00122-021-03773-7

→ Use estimated metabolic fluxes to describe the phenotype

What are the datasets used to develop the METAFLUX method?

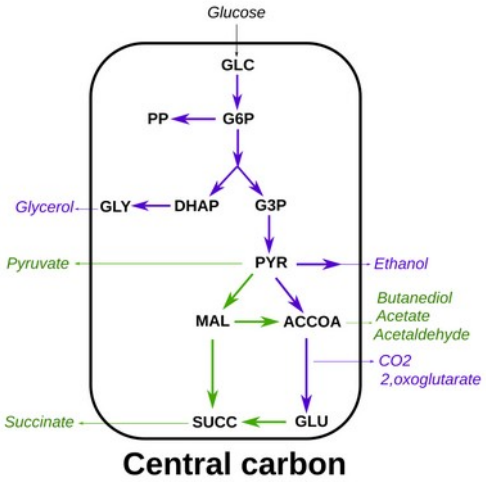
Yeast

Petrizzelli *et al.* (2021)
(doi: 10.1371/journal.pcbi.1009157)

Samples: 127 strains x temperature (18/26°C)
Experimental measure: - Protein abundances
- Flux measure on CO2_t

GSMM: 70 reactions of a simplify central carbon metabolism *(adapted from Celton et al. (2012), doi: 10.1016/j.ymben.2012.03.008)*

Coverage of the GSMM by proteomic data: *47%*



Central carbon

What are the datasets used to develop the METAFLUX method?

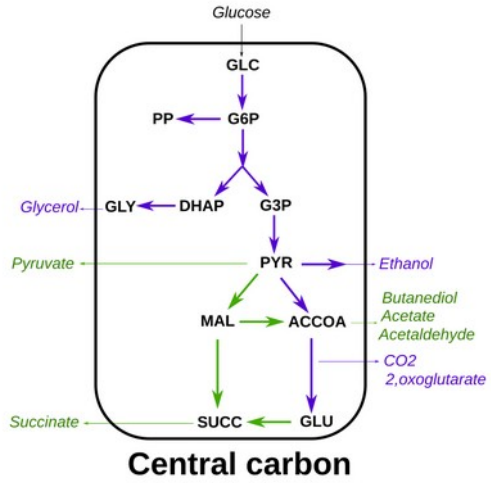
Yeast

Petrizzelli *et al.* (2021)
(doi: 10.1371/journal.pcbi.1009157)

Samples: 127 strains x temperature (18/26°C)
Experimental measure: - Protein abundances
- Flux measure on CO2_t

GSMM: 70 reactions of a simplify central carbon metabolism *(adapted from Celton et al. (2012), doi: 10.1016/j.ymben.2012.03.008)*

Coverage of the GSMM by proteomic data: *47%*



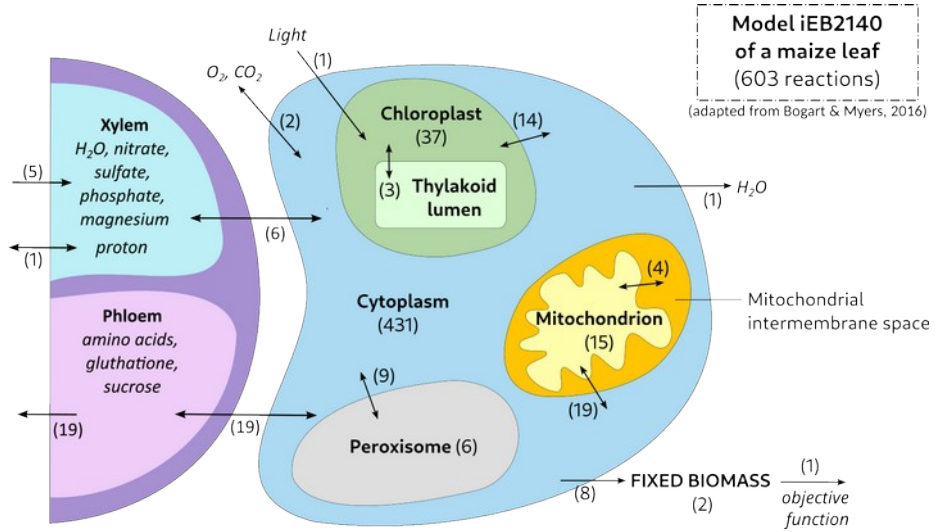
Maize

Negro *et al.* (2019)
(doi: 10.1186/s12870-019-1926-4)

Samples: 254 hybrids x 2 watering conditions (Well-watered / Water Deficit)
Experimental measure: - Protein abundances
- No flux measure

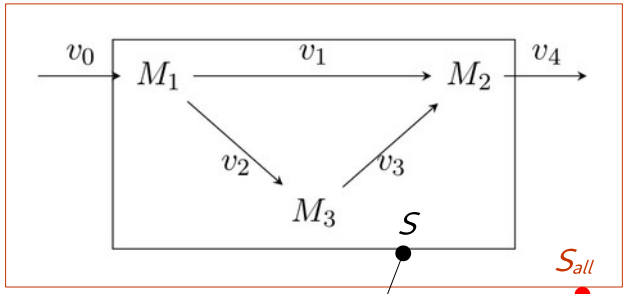
GSMM: 603 reactions of a simplify primary metabolism of maize leaf *(adapted from Bogart & Myers (2016), doi: 10.1371/journal.pone.0151722)*

Coverage of the GSMM by proteomic data: *44%*



Description of the METAFLUX method

Constraint-based model (CBM)



Internal reactions

Internal & exchange reactions

Stoichiometry matrix (S)

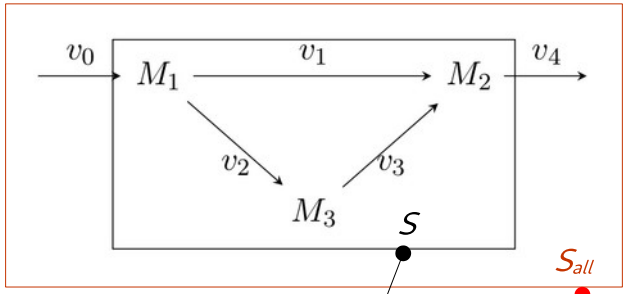
(R: reactions)

$$S_{all} = \begin{array}{c} \begin{array}{ccccc} R0 & R1 & R2 & R3 & R4 \end{array} \\ \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix} \end{array} \begin{array}{l} M1 \\ M2 \\ M3 \end{array}$$

(M : metabolites)

Description of the METAFLUX method

Constraint-based model (CBM)



Internal reactions

Internal & exchange reactions

Stoichiometry matrix (S)

(R: reactions)

$$S_{all} = \begin{array}{c} \begin{array}{ccccc} R0 & R1 & R2 & R3 & R4 \end{array} \\ \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix} \end{array} \begin{array}{l} M1 \\ M2 \\ M3 \end{array}$$

(M : metabolites)

At steady state (*no accumulation or degradation of metabolites inside the system)* :

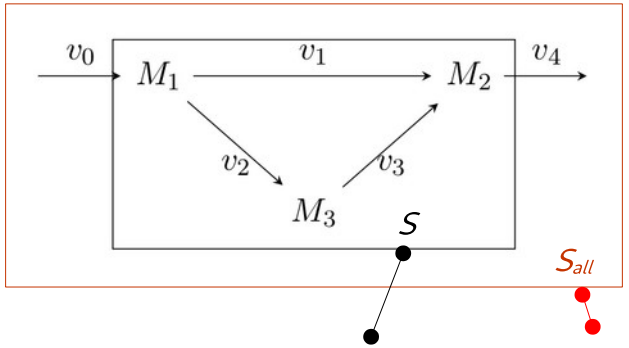
$$\dot{m} = 0 \Rightarrow S\,v = 0$$

$$S_{all}\,v = \dot{m}\,, with\; v^{min} \le v \le v^{max}$$

(adapted from Petrizzelli *et al.*, 2021, doi: 10.1371/journal.pcbi.1009157)

Description of the METAFLUX method

Constraint-based model (CBM)



Stoichiometry matrix (S)

(R: reactions)

$$S_{all} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix}$$

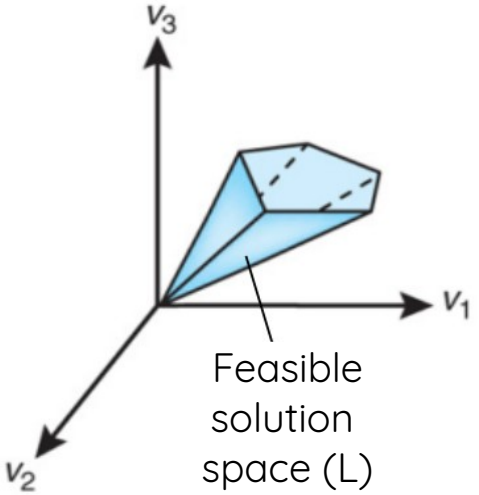
R0 R1 R2 R3 R4

M1 M2 M3 (M : metabolites)

At steady state (*no accumulation or degradation of metabolites inside the system)* :

$$\dot{m} = 0 \Rightarrow S\,v = 0$$

$$S_{all}\,v = \dot{m}\,, with\ v^{min} \le v \le v^{max}$$



More unknowns than equation so it exists an infinite number of solution to solve Sv = 0 !

$$L \equiv \{\nu \in \mathbb{R}^N | S\nu = 0, \nu^{inf} \le \nu \le \nu^{sup}\}$$

N = *reactions*
S = *stoichiometry matrix*
v = *flux*

→ Use proteomic data to identify the best solution in the feasible solution space

(adapted from Petrizzelli *et al.*, 2021, doi: 10.1371/journal.pcbi.1009157)

Description of the METAFLUX method (step-by-step)

Stoichiometry matrix (S)
Boundaries (v^{inf} & v^{sup})
Constraint on exchange reactions (e.g. CO_2)

Compute parameters of the posterior distribution of the feasible solution space (L)
(Expectation Propagation algorithm) (Braunstein et al., 2017, doi: 10.1038/ncomms14915)

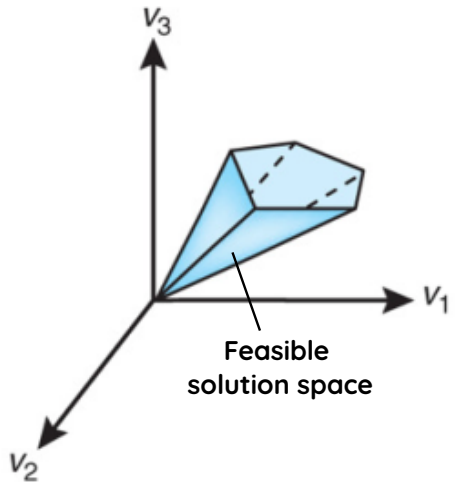
→ **pm** = vector of posterior means
→ **pv** = vector of posterior variances
→ **pcov** = posterior variance/covariance matrix

Example on reaction G3P_13DPG
(glyceraldehyde-3-phosphate dehydrogenase)

G3P + NAD ⇌ 13DPG +NADH

v^{inf} ——— v^{sup}
-1000 0 1000

13DPG + NADH → G3P + NAD → 13DPG + NADH



→ The feasible solution space is constraint by the boundaries of each reaction and the additional constraint on exchange reactions

(adapted from Petrizzelli *et al.*, 2021, doi: 10.1371/journal.pcbi.1009157)

Description of the METAFLUX method (step-by-step)

Stoichiometry matrix (S)
Boundaries (v^{inf} & v^{sup})
Constraint on exchange reactions (e.g. CO_2)

Compute parameters of the posterior distribution of the feasible solution space (L)
(Expectation Propagation algorithm) (Braunstein et al., 2017, doi: 10.1038/ncomms14915)

→ **pm** = vector of posterior means
→ **pv** = vector of posterior variances
→ **pcov** = posterior variance/covariance matrix

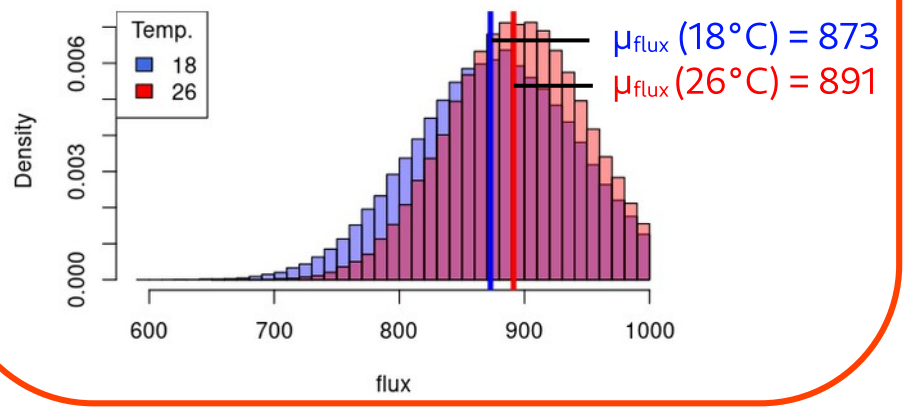
Random draw in the feasible solution space

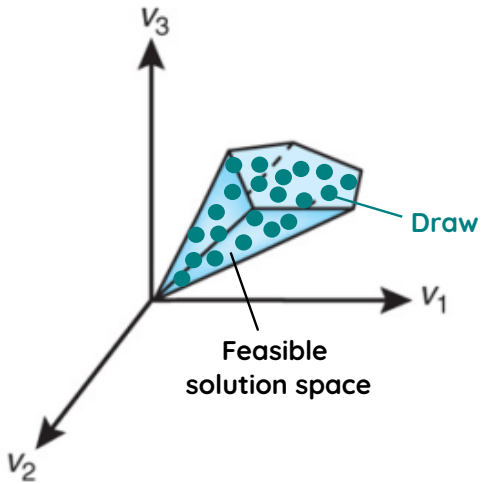
→ μ_{flux} = vector of flux means
→ σ_{flux} = vector of flux variances
→ w_d = probability of draw d

G3P + NAD ⇌ 13DPG +NADH

v^{inf} -1000 — 0 — 1000 v^{sup}

13DPG + NADH → G3P + NAD → 13DPG + NADH





(adapted from Petrizzelli *et al.*, 2021, doi: 10.1371/journal.pcbi.1009157)

Description of the METAFLUX method (step-by-step)

Stoichiometry matrix (S)
Boundaries (v^{inf} & v^{sup})
Constraint on exchange reactions (e.g. CO_2)

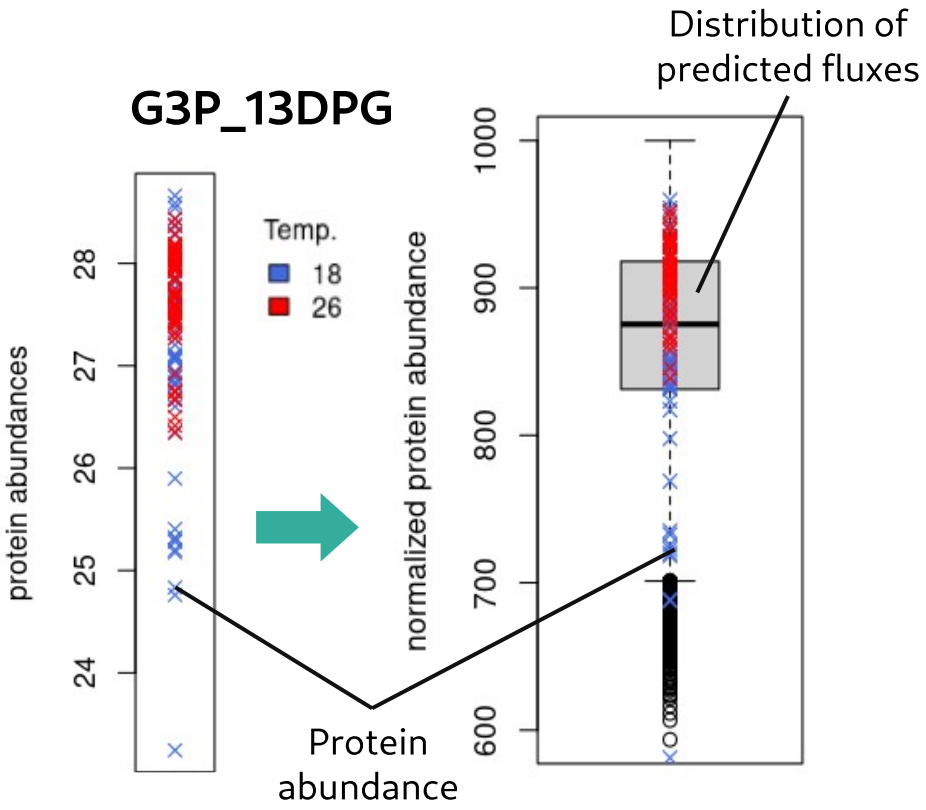
Compute parameters of the posterior distribution of the feasible solution space (L)
(Expectation Propagation algorithm) (Braunstein et al., 2017, doi: 10.1038/ncomms14915)

→ **pm** = vector of posterior means
→ **pv** = vector of posterior variances
→ **pcov** = posterior variance/covariance matrix

Random draw in the feasible solution space

→ μ_{flux} = vector of flux means
→ σ_{flux} = vector of flux variances
→ w_d = probability of draw d

Normalized protein abundances at metabolic fluxes scale (E'_{kr})



(adapted from Petrizzelli *et al.*, 2021, doi: 10.1371/journal.pcbi.1009157)

How normalize protein abundances?

1) Center reduce each protein abundance

$$\tilde{E}_{kr} = \frac{E_{kr} - \mu_{prot,r}}{\sigma_{prot,r}}$$

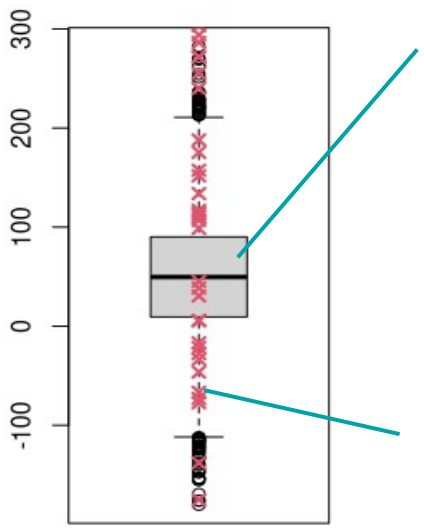
2) Partitioning each flux distribution

$$F_r^- = \{v_{rd} : v_{rd} < 0\}, with\ \mu_{flux,r}^-, \sigma_{flux,r}^-$$
$$F_r^+ = \{v_{rd} : v_{rd} \geq 0\}, with\ \mu_{flux,r}^+, \sigma_{flux,r}^+$$

3) Affine projection onto flux scale (α=10)

$$E'^-_{kr} = \alpha . \tilde{E}_{kr} . \sigma_{flux,r}^- + \mu_{flux,r}^-$$
$$E'^+_{kr} = \alpha . \tilde{E}_{kr} . \sigma_{flux,r}^+ + \mu_{flux,r}^+$$

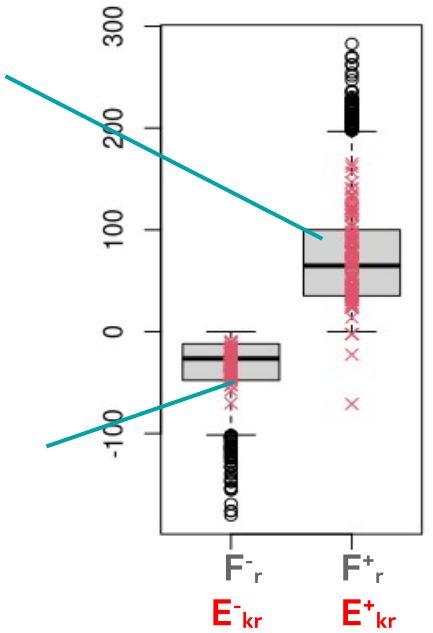
Normalization of protein abundance **without partitioning the flux distribution**



Distribution of the predicted fluxes

Normalized protein abundances

Normalization of protein abundance **with partitioning the flux distribution**



E_{kr} = log10 protein limiting abundances for sample *k*, reaction *r*
$\tilde{E}_{kr}$ = centered reduced protein abundance for sample *k*, reaction *r*
$\mu_{prot,r}$, $\sigma_{prot,r}$ = mean and standard deviation of E_r across *K samples*
$\mu^{\pm}_{flux,r}$, $\sigma^{\pm}_{flux,r}$ = mean and standard deviation of the positive/negative flux sub-distribution for reaction *r*
$F^{\pm}_r$ = positive/negative flux distribution for reaction r
$E'^{\pm}_{kr}$ = normalized protein abundances for positive/negative flux distribution for sample, reaction *r*

Description of the METAFLUX method (step-by-step)

Stoichiometry matrix (S)

Boundaries (v^{inf} & v^{sup})

Constraint on exchange reactions (e.g. CO_2)

Compute parameters of the posterior distribution of the feasible solution space (L)
(Expectation Propagation algorithm) (Braunstein et al., 2017, doi: 10.1038/ncomms14915)

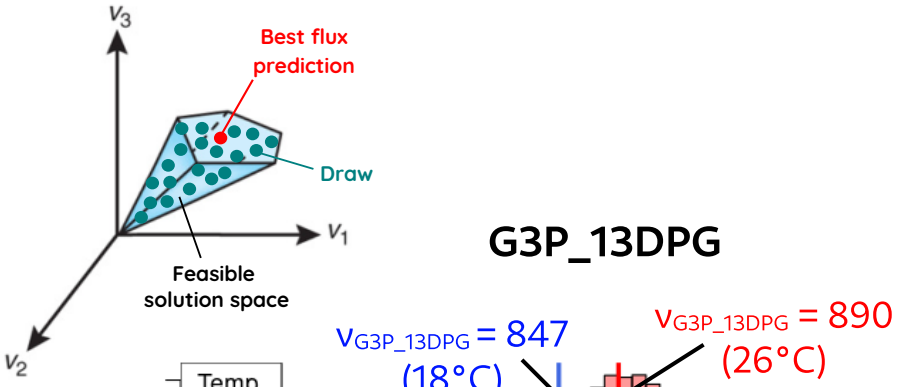
→ **pm** = vector of posterior means
→ **pv** = vector of posterior variances
→ **pcov** = posterior variance/covariance matrix

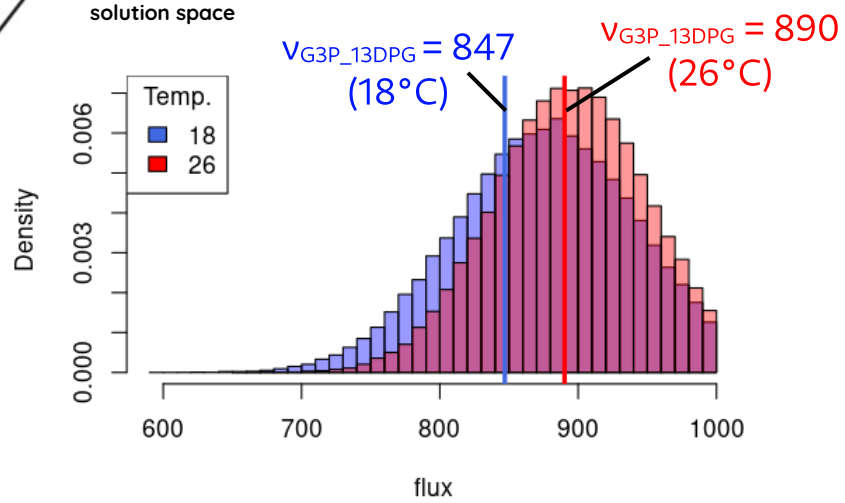
Random draw in the feasible solution space

→ μ_{flux} = vector of flux means
→ σ_{flux} = vector of flux variances
→ w_d = probability of draw d

Normalized protein abundances at metabolic fluxes scale (E'_{kr})

Select the best solution according to the data (v_d, w_d)
minimize the Zscore (suitability for the proteomic data) &
maximize the weigh (suitability for the metabolic model)





(adapted from Petrizzelli *et al.*, 2021, doi: 10.1371/journal.pcbi.1009157)

How choose the best draw according the proteomic data (low Zscore) and the metabolic model (high log density) ?

1) Calculate a Z score *(suitability for proteomic data)*

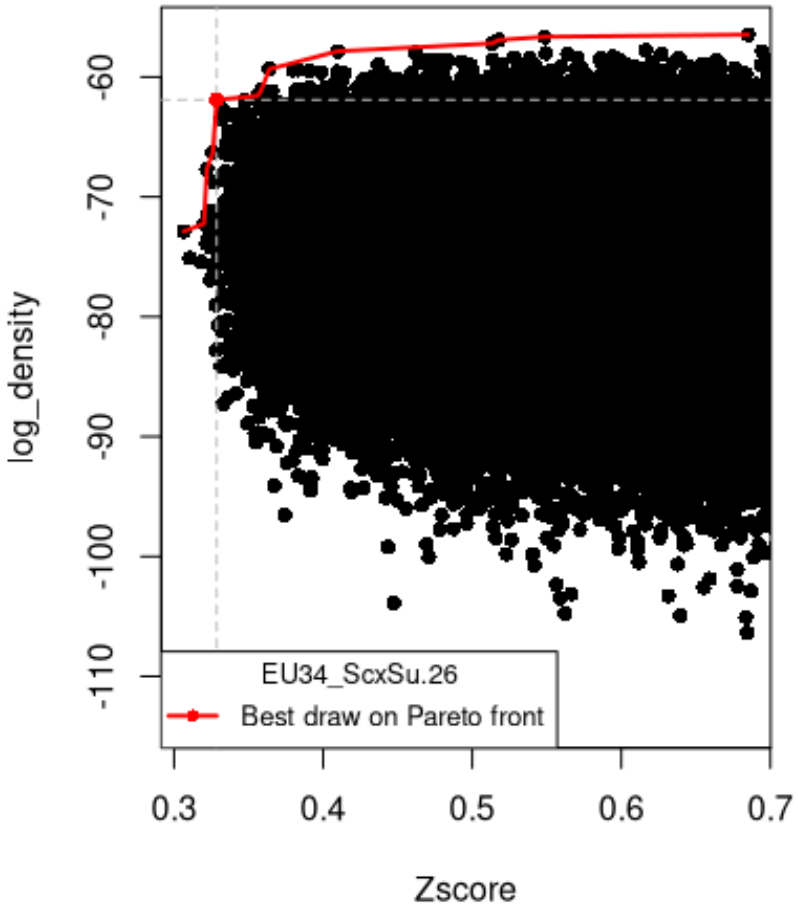
$$Z_{ikd} = \frac{\sqrt{\sum_{r=1}^{R_{obs}} \left(v_{rd} - E'^{s(r)}_{kr}\right)^2}}{\sqrt{\sum_{r=1}^{R_{obs}} \sigma^2_{flux,r}}} \qquad E'^{s(r)}_{kr} = \begin{cases} E'^{-}_{kr} \; if \; v_{rd} < 0 \\ E'^{+}_{kr} \; if \; v_{rd} \geqslant 0 \end{cases}$$

2) Define the Pareto optimum (points on the Pareto front) that minimize the Z score and maximize the log density *(proxy of probability to draw d in the feasible solution space)*

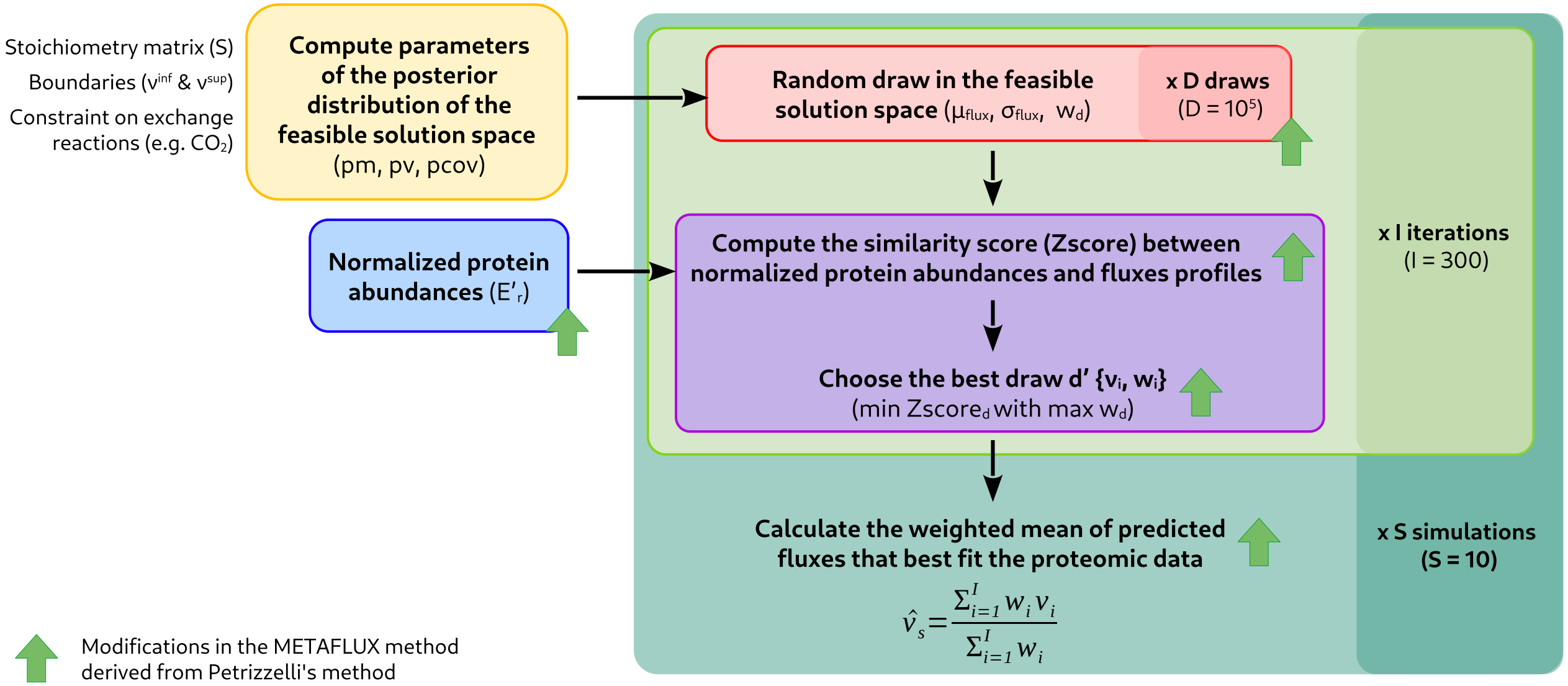
3) Calculate the slope between each point on the Pareto front

4) Calculate the maximum budget for Z score

→ Select the highest Pareto optimum with significant slopes that meets the maximum budget for Z score



Description of the METAFLUX method (overview)

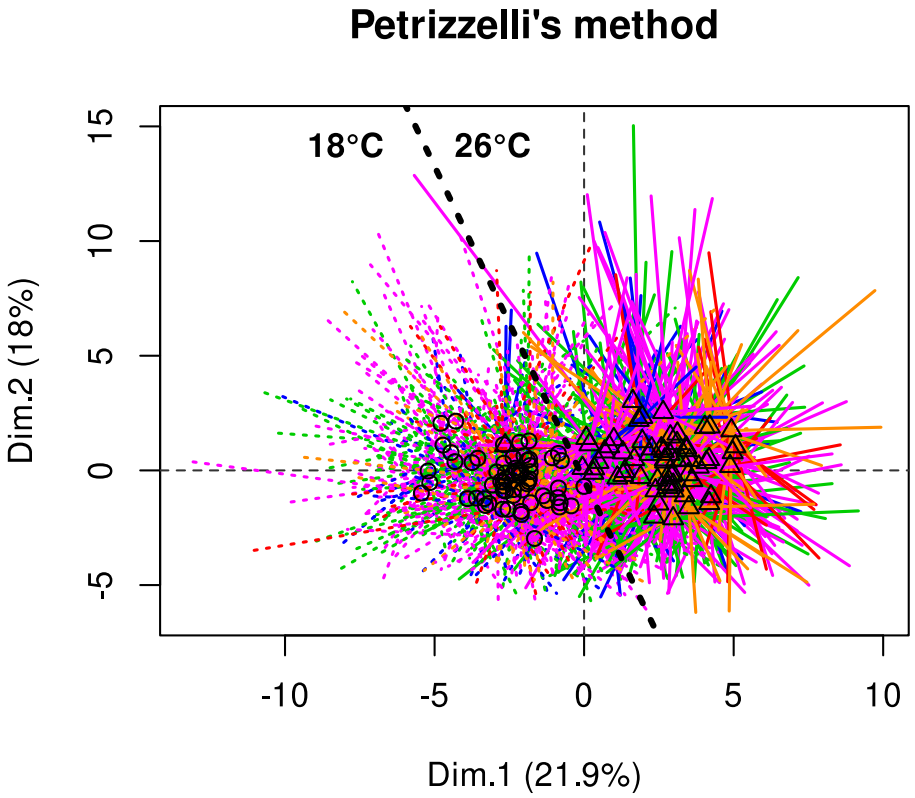


Modifications in the METAFLUX method derived from Petrizzelli's method

(adapted from Petrizzelli *et al.*, 2021, doi: 10.1371/journal.pcbi.1009157)

Results on yeast dataset

- Improvements of the original method -



Type of strain

- Sc
- ScxSc
- ScxSu
- Su
- SuxSu

Sc: *S.cerevisiae*
Su: *S.uvarum*

Temperature

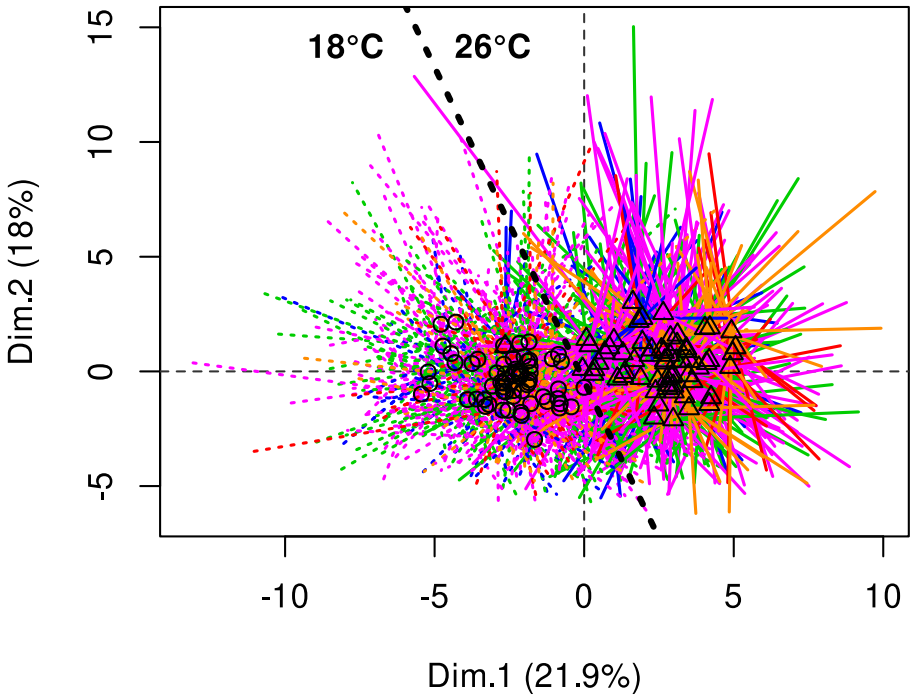
- ○ 18
- △ 26

- **Weak separation of sample by temperature**
- **Weak repeatability between simulations**

Results on yeast dataset

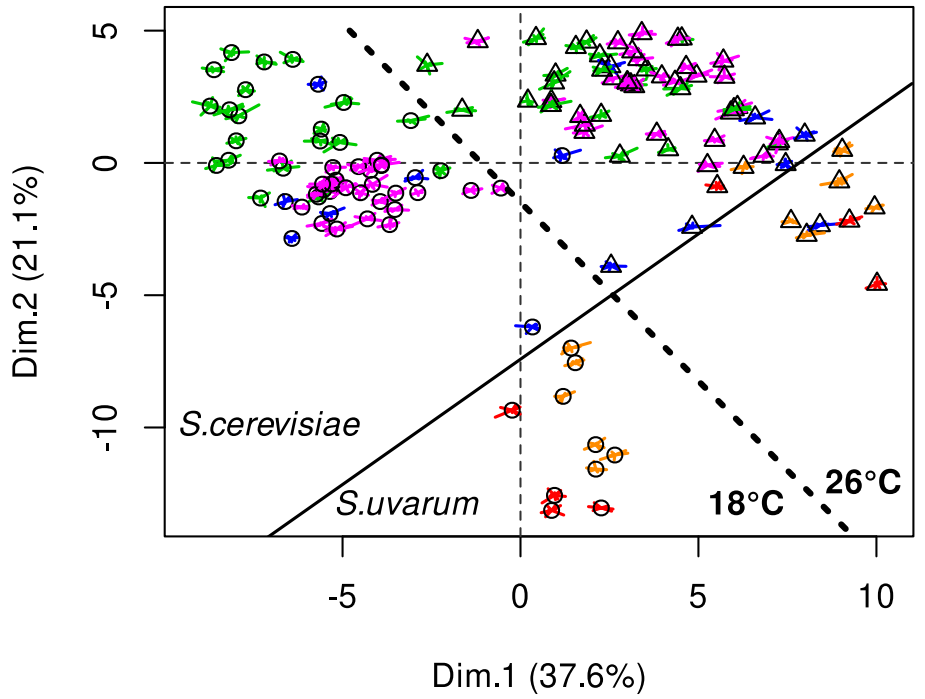
- Improvements of the original method -

Petrizzelli's method

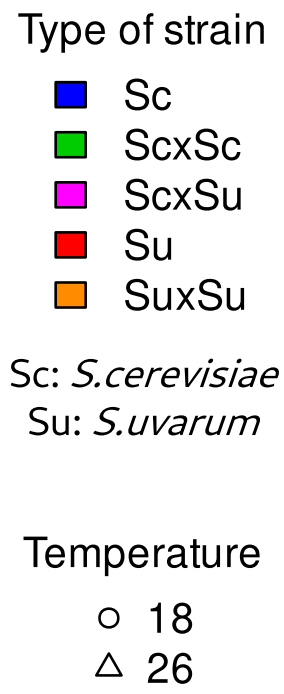


- Weak separation of sample by temperature
- Weak repeatability between simulations

METAFLUX method



- Well separation of sample by temperature and by type of strain
- High repeatability between simulations



Results on yeast dataset

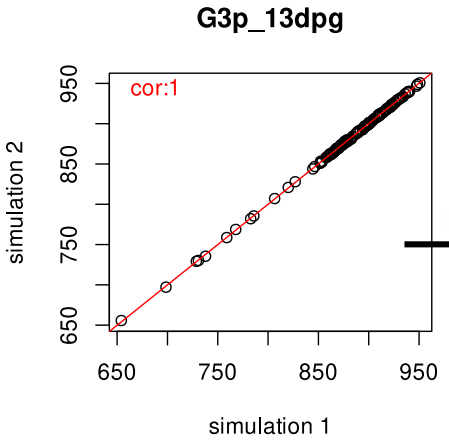
- Repeatability between simulations -

▶ **The reaction with proteomic data (*) are very well predicted (➔)**

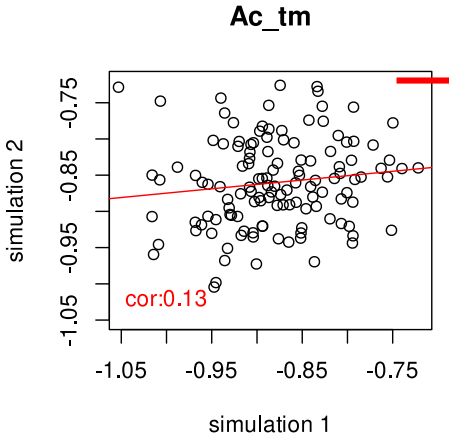
▶ **The coverage of the CBM by high correlated predicted fluxes (≥ 0.6) is higher than the one with proteomic data**

Coverage of the CBM

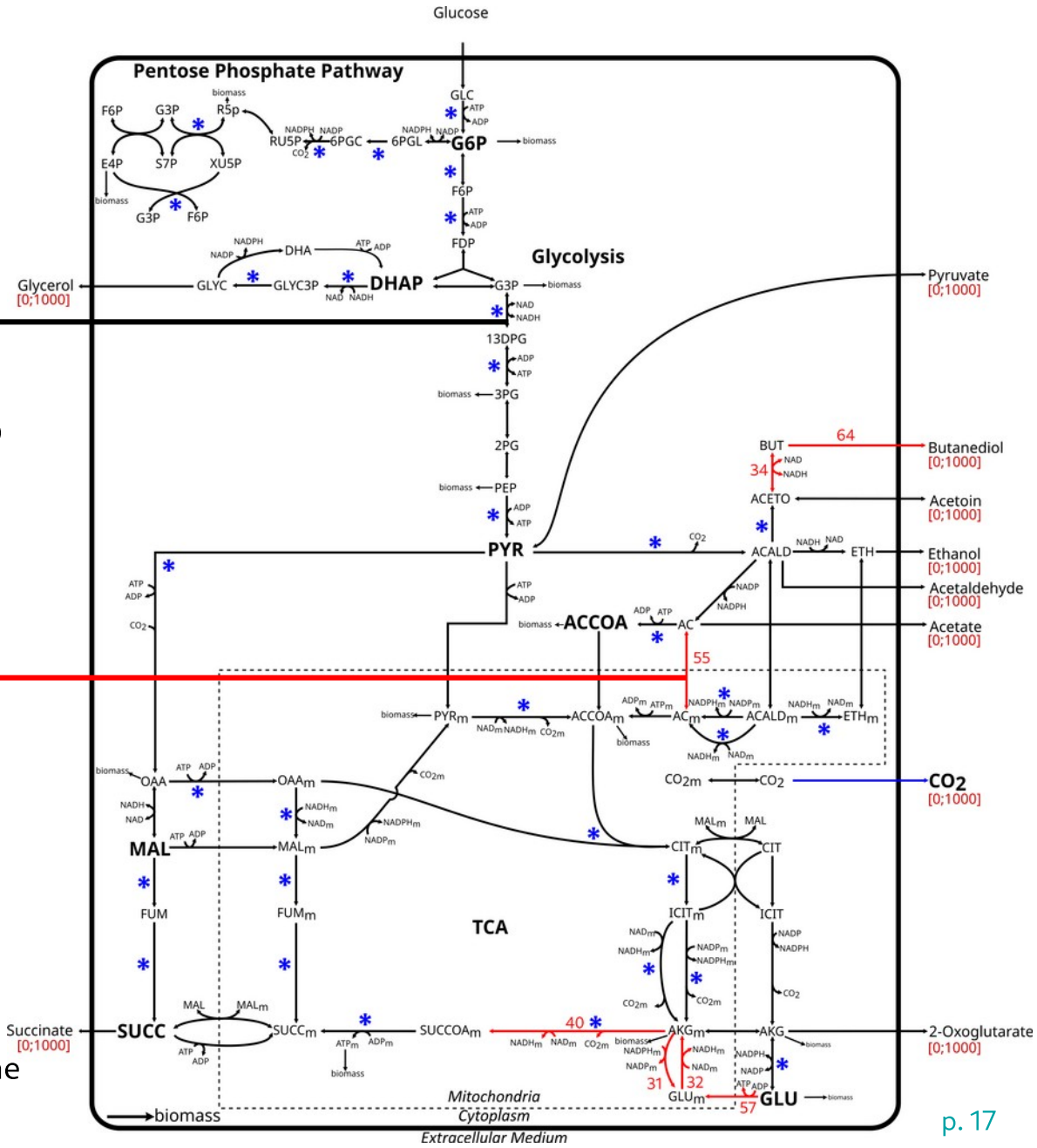
- by proteomic data : 33/70 (47%)
- by predicted fluxes : 63/70 (90%)



Strong correlation
= High repeatability

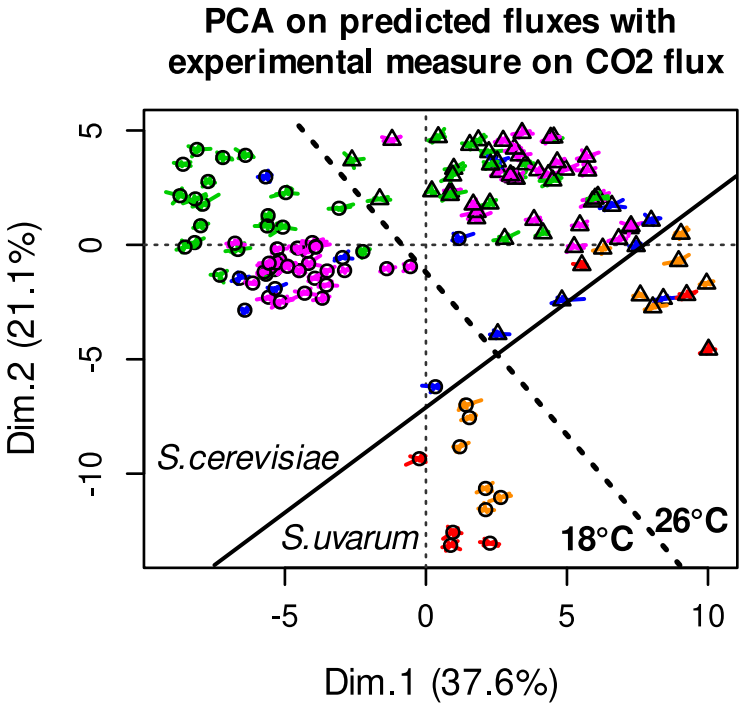


No correlation
= Unable to interpret the predicted flux



Results on yeast dataset

- Can we predict correctly the fluxes without experimental measure on CO_2? -

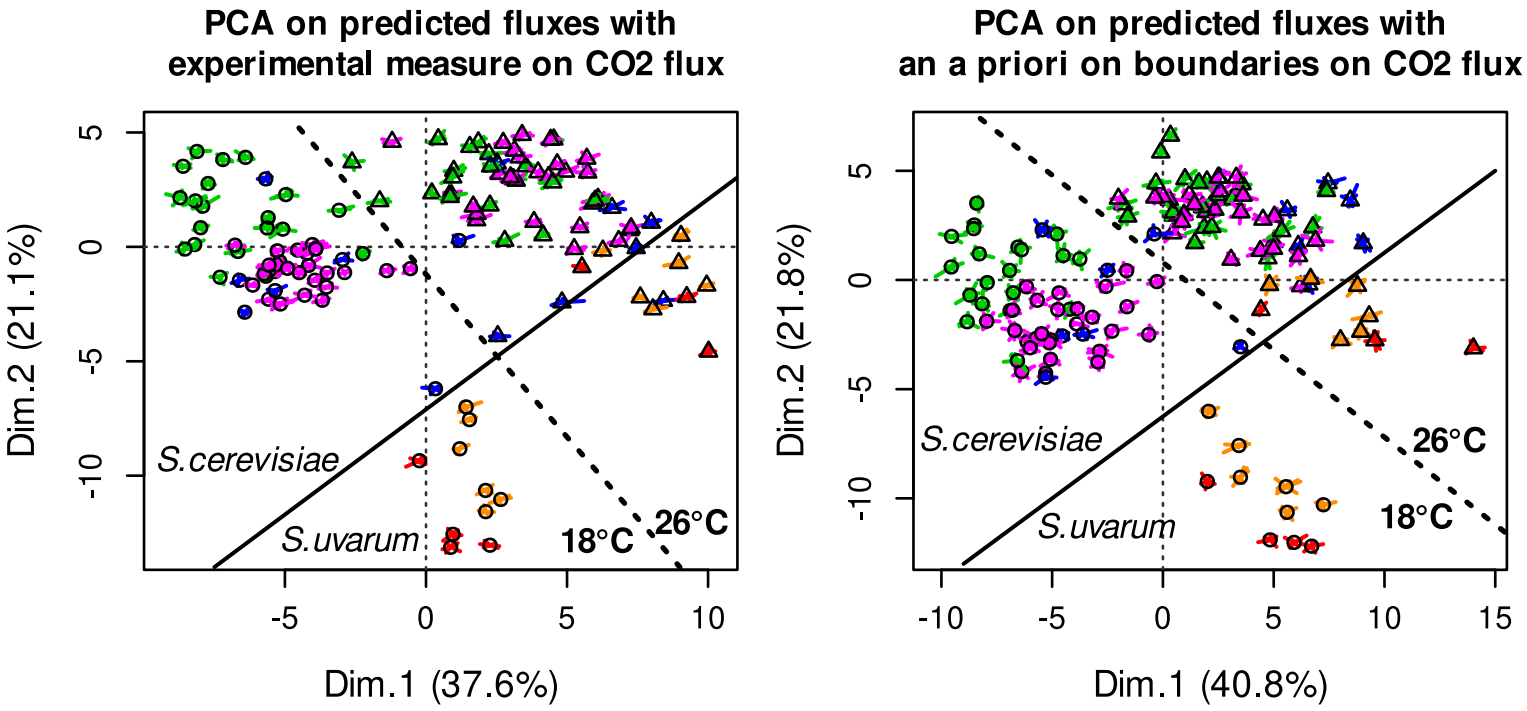


Type of strain
- Sc
- ScxSc
- ScxSu
- Su
- SuxSu

Temperature
- 18
- 26

Results on yeast dataset

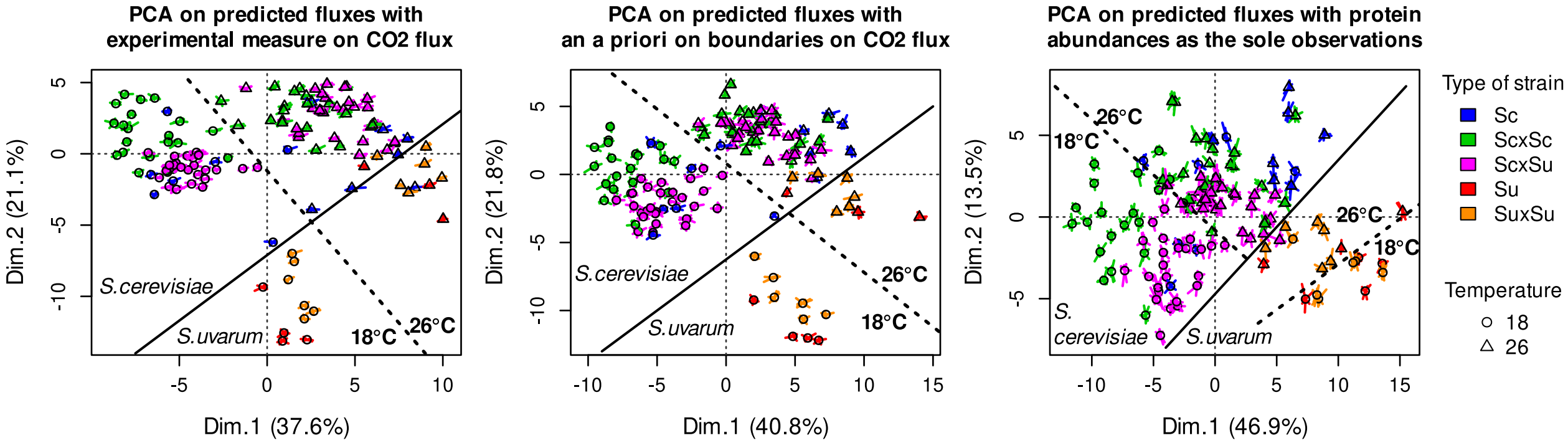
- Can we predict correctly the fluxes without experimental measure on CO_2? -



Fixed boundaries on CO2_t :
18°C [150 ; 360]
26°C [250 ; 610]

Results on yeast dataset

- Can we predict correctly the fluxes without experimental measure on CO_2? -

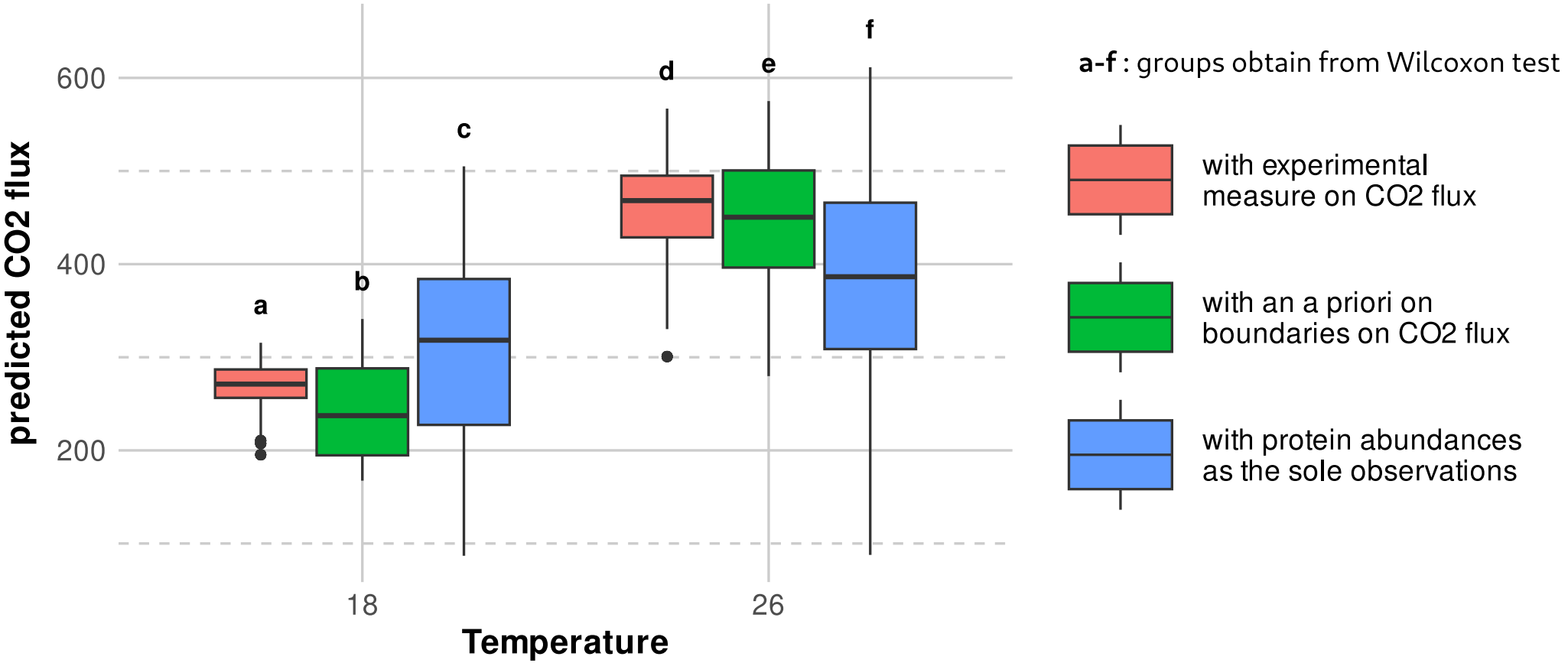


Fixed boundaries on CO2_t :
18°C [150 ; 360]
26°C [250 ; 610]

▶ The METAFLUX method is efficient to predict metabolic fluxes and separate sample per type of strain and temperature even without experimental measure

Results on yeast dataset

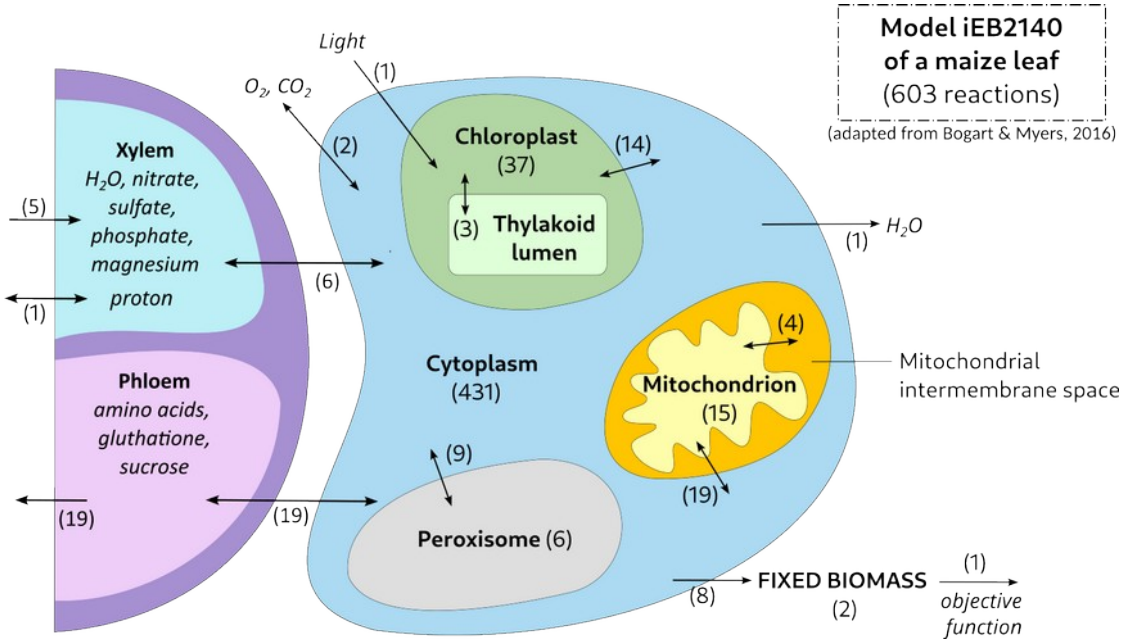
- Can we predict correctly the CO_2 flux without experimental measure ? -



Without additional information, the CO2 flux is not correctly predicted,
but prior information on the the boundaries of CO2 improve the prediction

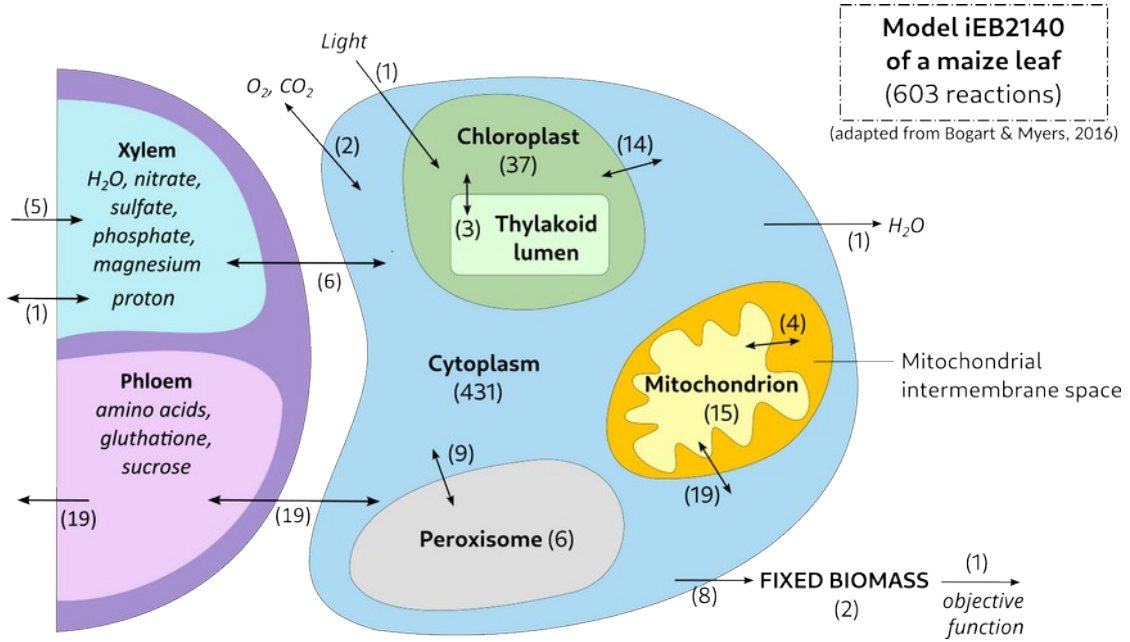
Preliminary results on maize

- **Less information are available on maize**
  - no experimental values to constraint fluxes
  - Less information to complete the GPR (Gene-Protein Reaction rule)
- **A changing in the scale**
  - 70 reactions (yeast) → 603 reactions (maize)

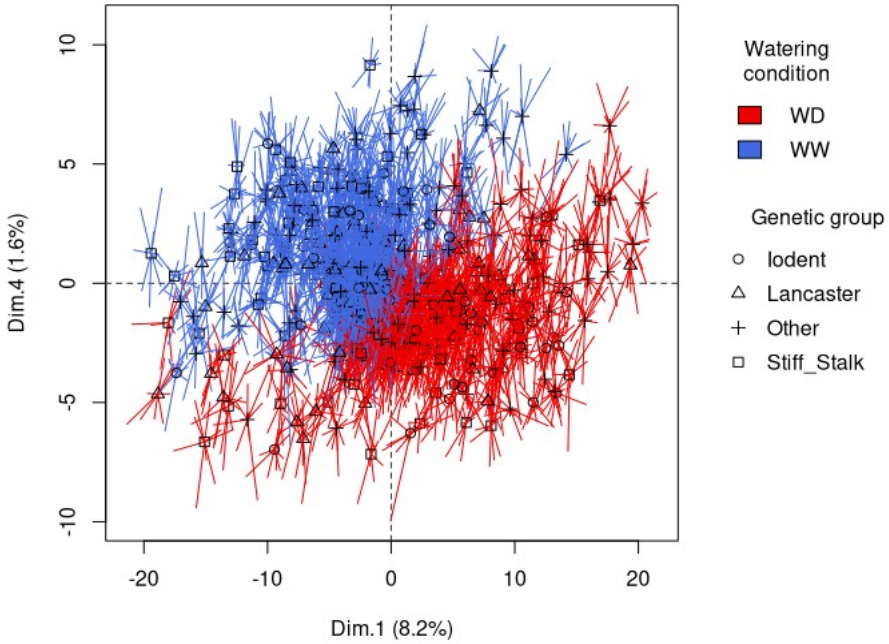


Preliminary results on maize

- **Less information are available on maize**
  - no experimental values to constraint fluxes
  - Less information to complete the GPR (Gene-Protein Reaction rule)
- **A changing in the scale**
  - 70 reactions (yeast) → 603 reactions (maize)



- **Preliminary result on maize**
  - Relatively good separation of samples per watering condition
    ↳ axis 1 and 4 contribute the most to group sample per condition



Conclusions on yeast

- **Efficient method to predict metabolic fluxes in different environmental condition**
- **Highly repeatable for most reactions**
- **Efficient to predict fluxes despite incomplete coverage by proteomic data**
- **Efficient to predict fluxes even without experimental measure on exchange flux provided some prior information on the boundaries**

Perspectives on maize

- Format maize data (GPR)
- Run flux prediction with optimum parameters
- Analyze and interpret the genetic variation and GxE at the flux level

Thanks to…

Christine DILLMANN - Université Paris-Saclay, INRAE, AgroParisTech, GQE-Le Moulon

Mélisande BLEIN-NICOLAS - Université Paris-Saclay, INRAE, AgroParisTech, GQE-Le Moulon
- Université Paris-Saclay, INRAE, AgroParisTech, CNRS, EMR 9005 GevAD

Sylvain PRIGENT - Université Bordeaux : UMR1332 BFP,
- Bordeaux Metabolome : MetaboHUB, PHENOME-EMPHASIS

Marianyela S. PETRIZZELLI - Sanofi

Dominique DE VIENNE - Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE-Le Moulon

Funding

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
Liberté Égalité Fraternité

CARNOT Plant2Pro