

Data integration: Bayesian modeling of fluxes through metabolic pathways

C. Dillmann, P.Y. Bourguignon, L. Roman-Padilla

MetaColi ANR

AppliBugs 2014

Les années 90's



statistiques

GENETICS AND BREEDING

Estimation of Heterogeneous Variances Using Empirical Bayes Methods: Theoretical Considerations¹

D. GIANOLA, J. L. FOULLEY,² R. L. FERNANDO,³ G. R. HENDERSON,^{3,4} and K. A. WEGEL⁵

Copyright © 1992 by the Genetics Society of America

Prediction of Selection Response for Threshold Dichotomous Traits

Jean L. Foulley

sélection

APTITUDES MATERNELLES DES RACES LAITIÈRES, MIXTES, RUSTIQUES ET À VIANDERIE POUR LA PRODUCTION DE VELEUX DE BOUCHERIE EN FERME PAR CROISEMENT INDUSTRIEL

J. L. FOULLEY, F. MÉNÈSIER, J. GARLAND* — Département de Génétique animale, C. N. R. Z. 78250 Jouy en Josas (France), * Institut Technique de l'Élevage bovin, ITR, R. de Berry, 35012 Paris (France).

Genet. Sel. Evol. 98 (1986) 510-561

543

Another look at multiplicative models in quantitative genetics

Original article

Genet. Sel. Evol. 32 (2000) 187-203
© INRA, EDP Sciences

187

Original article

Genetic diversity of eleven European pig breeds



évolution



ENSAI, Rennes-DEA de Génétique Multifactorielle, Paris VII-XI-INA.PG

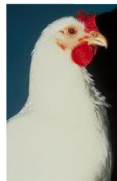
LE MODELE LINEAIRE MIXTE

24. Interprétation bayésienne

enseignement

Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds.
R Ciampolini, K Moazami-Goudarzi, D Vaiman, C Dillmann, E Mazzanti, J L Foulley, H Leveziel and D Cianci

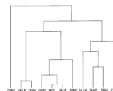
J ANIM SCI 1995, 73:3259-3268.



génétique quantitative

Long-term effects of selection based on the animal model BLUP in a finite population

E. Verrier¹, J. J. Colleau², J. L. Foulley³



The MetaColi project



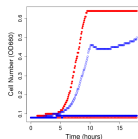
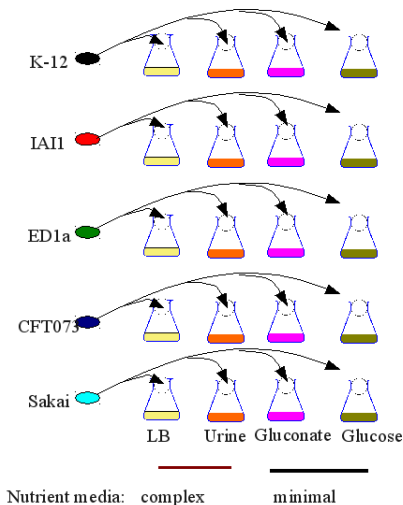
INSERM U722
Univ Paris 7
Denis Diderot



Gain a deeper understanding of the metabolic adaptation of a variety of *E. coli* strains to different lifestyles

- Quantitative description of metabolic diversity
- Integration of a broad range of heterogeneous experimental data into models of metabolism
- Prediction of bacterial growth and survival for various environmental conditions.

Experimental set-up

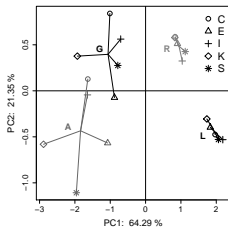


- **Life-History traits:** growth rate, lag-time, total proteins.
- **Protein abundance:** quantitative proteomics, 2D PAGE.
- **Enzyme specific activity:** enzyme assays.
- **Input and output fluxes:** metabolite assays.

Global patterns of metabolic diversity

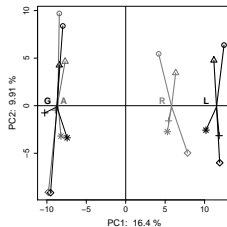
Life-History traits

Clear separation between the four culture media with G by E interactions



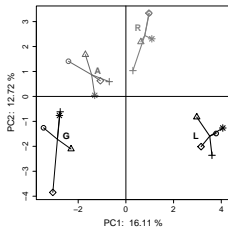
Proteom: All proteins

Strains ranked according to their phylogeny whatever the culture medium



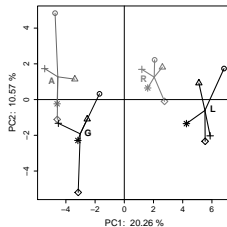
V_{max} : Central carbon met.

Good correlation between Protein abundance and Enzyme activity

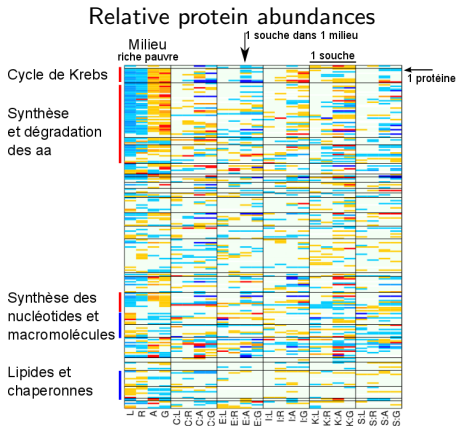


Proteom: Central Carbon met.

Similar to life-history traits but noticeable differences

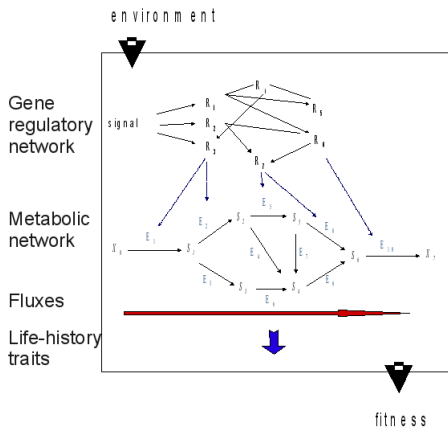


Phenotypic plasticity and proteomic tinkering



Phenotypic plasticity: nutrient resources determine average protein abundance within pathways. **G by E interactions:** differences between genotypes depend on nutrient resources.

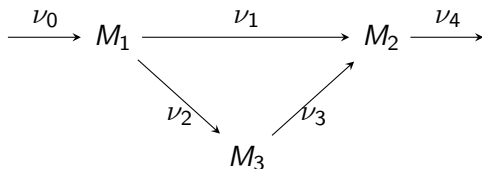
Metabolism



Metabolic network

- Links Metabolites M_i through enzymatic reactions
- Links reactions through shared metabolites
- = bipartite graph
- Fluxes = reactions rates ν_i ; determine cell growth dynamics.

Metabolic networks



Matter conservation

$$\dot{M}_1 = \nu_0 - 2\nu_1 - \nu_2$$

$$\dot{M}_2 = \nu_1 + \nu_3 - \nu_4$$

$$\dot{M}_3 = \nu_2 - \nu_3$$

$$\vec{\dot{M}} = S\vec{\nu}$$

Kinetics

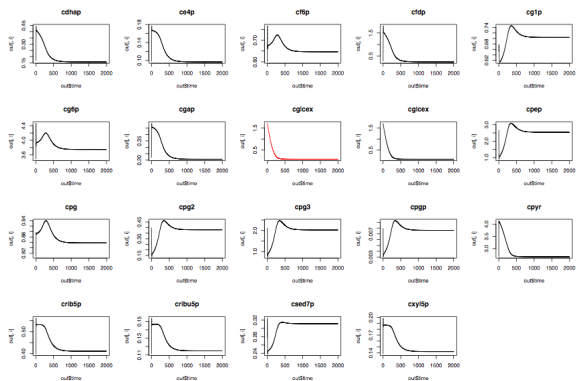
$$\nu_1 = s_{11}k_1 \frac{M_1 - M_2/K_{e2}}{K_{m1} + M_1 + M_2}$$

$$\nu_2 = s_{12}k_2 \frac{M_1 - M_3/K_{e3}}{K_{m1}(1 + M_2/K_{e2}) + M_1 + M_3}$$

...

S is the matrix of stoichiometry coefficients.

Stationary State



At stationary state, internal metabolite concentrations are constant.
Exchange rates are constant for external metabolites.

$$S\vec{v} = 0$$

Resolving steady-state equations

Flux Balance

- Stationary State
 $S\vec{\nu}^* = 0$
- Constraints (thermodynamics)
 $C\vec{\nu} \geq \vec{c}$
- Solutions space
 $\vec{\nu}^* = \vec{\nu}_0^* + Kq$

K is a base obtained from singular values decomposition of $S = \text{null space of } S$. $K'.K = I$, and $K.S = 0$.

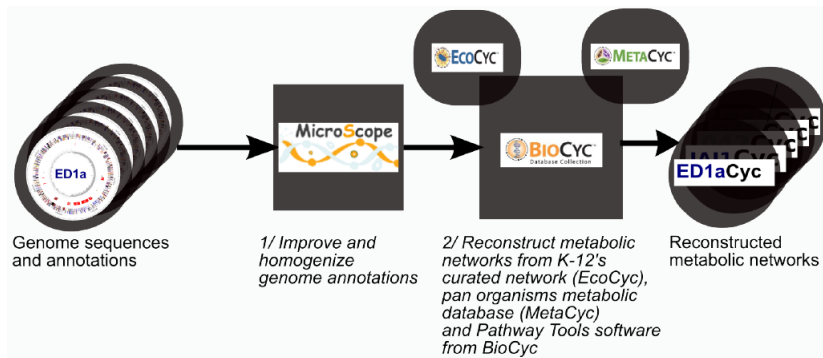
Kinetic models

$$\nu_1 = s_{11}\kappa_1 \frac{M_1 - M_2/K_{e2}}{K_{m1} + M_1 + M_2}$$

- Equilibrium constants
 K 's from literature.
- Enzymatic activities $\vec{\kappa}$
unknowns, related to protein abundances.
- Numerical resolution of ODE system
 $\nu^{kin} = g^*(\vec{\kappa})$

Automated reconstruction of detailed whole-genome metabolic networks

Gilles Vieira, Victor Sabarly et al, J. Bact, 2011



29 strains reconstructed using high quality genome sequences and annotations, and information propagation

Semi-automated reconstruction of metabolic models

Gille Vieira, Maxime Durot, François Lefevre (Genoscope)

● Data

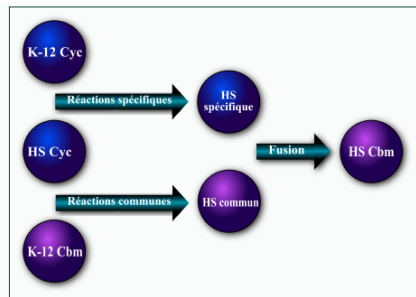
Metabolic networks ($K12_{cyc}$, $S88_{cyc}$, ...), Reference metabolic data-base, Reference model ($K12MG1655$)

● Tools

Networks manipulation platform (MicroCyc), CBM manipulation platform (Nemo Studio)

● Approach

- ① Find common reactions
- ② Find and create new reactions
- ③ Fuse and check the consistency of the two reaction sets

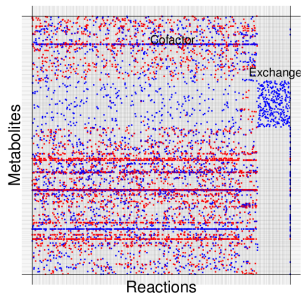


20 strains reconstructed

<http://www.genoscope.cns.fr/agc/metacoli/>

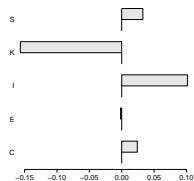
The *In silico* metabolic diversity is low

Flux Balance Analysis on reconstructed metabolic models

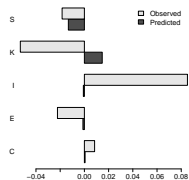


Example of stoichiometry matrix S
2356-2404 reactions. 1039-1718 metabolites

Relative growth on Glucose

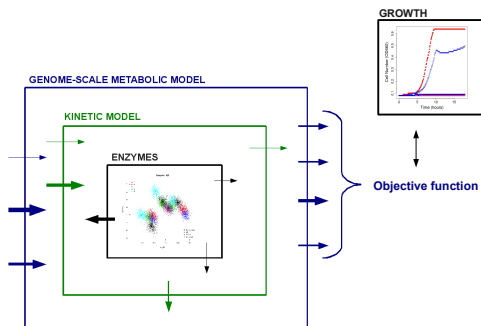


Relative growth on LB



FBA fails to predict the observed variation on growth rate

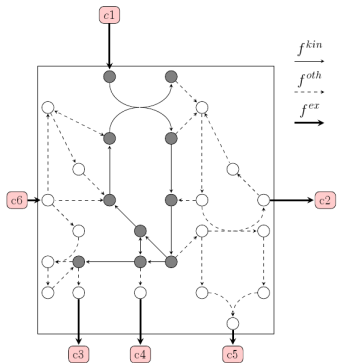
Integrating quantitative data into metabolic models



- ① Link the genome-scale metabolic model to a kinetic model from central carbon metabolism.
- ② Integrate quantitative data on protein abundances *via* the kinetic model.
- ③ Bayesian model to estimate posterior fluxes distributions given observed data.

Models coupling

$$\begin{bmatrix} \vec{M}^{kin} \\ \vec{M}^{co} \\ \vec{M}^{oth} \\ \vec{M}^{ex} \end{bmatrix} = \begin{bmatrix} S^{kin-kin} & S^{kin-oth} & 0 \\ S^{co-kin} & S^{co-oth} & 0 \\ 0 & S^{oth-oth} & 0 \\ S^{ex-kin} & S^{ex-oth} & S^{ex-ex} \end{bmatrix} \cdot \begin{bmatrix} \vec{v}^{kin} \\ \vec{v}^{oth} \\ \vec{v}^{ex} \end{bmatrix} \quad (1)$$



- Steady-state equations can be rewritten as :
 $S^{kin} \vec{v}^{kin} + S^{\bar{kin}} \vec{v}^{\bar{kin}} = 0$
- $\vec{\kappa}$ bears information on steady-state :
 $\vec{v}^{kin} = g^*(\vec{\kappa})$
- Conditional on $\vec{\kappa}$, the prior on steady-state condition for the genome-scale models changes :
 $S^{kin} g^*(\vec{\kappa}) + S^{\bar{kin}} \vec{v}^{\bar{kin}} = 0$

Models coupling

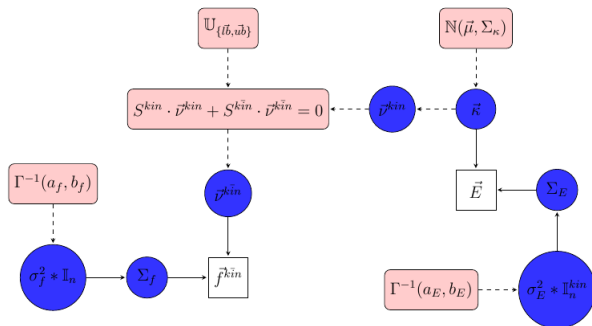


Figure 2: Description of the variation observed at one cell level. Elements inside blue circles refer to parameters of the model. White nodes represent (at least partially) observed data. Elements inside pink nodes refers to the prior information about the parameters.

Bayesian Model

Prior distributions and prior information

● Prior distributions

- i. $\vec{\kappa} \sim \mathbb{N}_{n^{kin}}(\vec{\mu}, \Sigma_{\kappa})$ ($\vec{\kappa} \geq \vec{0}$), $\{\vec{\mu}, \Sigma_{\kappa}\}$ are known.
- ii. $\sigma_E^2 \sim \Gamma^{-1}(a_E, b_E)$, $\{a_E, b_E\}$ given.
 $\sigma_f^2 \sim \Gamma^{-1}(a_f, b_f)$, $\{a_f, b_f\}$ given.
- iii. Fluxes \vec{v}^{kin} , $v_{lower,i}^{kin} \leq v_i^{kin} \leq v_{upper,i}^{kin} \Rightarrow \vec{v}^{kin} \sim \mathbb{U}_{\{\vec{l}b, \vec{u}b\}}$.
- iv. Combining information (i) and (iii), $\pi_{\kappa|\vec{v}^{kin}}$ is a mixed distribution between an multivariate Normal and a Uniform distribution.

Stationary steady-state

$$S^{kin} \cdot g^* + S^{\bar{kin}} \cdot \vec{v}^{\bar{kin}} = 0$$

$$\downarrow$$

$$\left\{ \vec{v}^{\bar{kin}} | \vec{\kappa} : S^{kin} \cdot g^* + S^{\bar{kin}} \cdot \vec{v}^{\bar{kin}} = 0 \right\}$$

$$\downarrow$$

$$\pi_{\nu|\kappa} \propto \mathbb{U}_{\{S^{kin} \cdot g^* + S^{\bar{kin}} \cdot \vec{v}^{\bar{kin}} = 0\}}$$

$$\mathbb{U}_{\{\cdot\}} = \begin{cases} 1 & \text{if } S^{kin} \cdot g^* + S^{\bar{kin}} \cdot \vec{v}^{\bar{kin}} = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Prior (joint) distribution

$$\pi_{prior}(\vec{\kappa}, \vec{v}, \sigma_E, \sigma_f) \propto \underbrace{\pi_{\kappa|\vec{v}^{kin}} * \pi_{\sigma_E} * \pi_{\sigma_f}}_{\text{prior dists.}} * \underbrace{\pi_{\nu|\kappa}}_{\text{steady-state}}$$

Bayesian Model

Likelihood function

$$\vec{E} \sim \mathbb{N}_{n^{kin}}(\vec{\kappa}, \Sigma_E) \quad \text{and} \quad \vec{f}^{\bar{kin}} \sim \mathbb{N}_{n^{\bar{kin}}}(\vec{\nu}^{\bar{kin}}, \sigma_f^2 * \mathbb{I}_{n^{\bar{kin}}}), \quad \Sigma_E = \sigma_E^2 * \mathbb{I}_{n^{kin}}$$

$$\vec{E} = \begin{pmatrix} \vec{E}_{obs} \\ \vec{E}_{nobs} \end{pmatrix} \quad \text{and} \quad \vec{f}^{\bar{kin}} = \begin{pmatrix} \vec{f}_{obs}^{\bar{kin}} \\ \vec{f}_{nobs}^{\bar{kin}} \end{pmatrix} \quad (n^{kin} = n_{obs}^{kin} + n_{nobs}^{kin} \quad \text{and} \quad n^{\bar{kin}} = n_{obs}^{\bar{kin}} + n_{nobs}^{\bar{kin}})$$

$$\vec{E}_i^t = (-, -, -, \underbrace{E_{4i}}_6, \underbrace{E_{5i}}_8, \underbrace{E_{6i}}_6, \underbrace{E_{7i}}_2, \dots, \underbrace{E_{13i}}_2, \underbrace{E_{14i}}_6, \dots) \quad \mathbf{A3: Independence across repetitions.}$$

$$l(\vec{\kappa}, \vec{\nu}^{\bar{kin}}, \sigma_E^2, \sigma_f^2 | \mathbf{E}, \mathbf{f}) \propto (\sigma_E^2)^{-R_1 n_{obs}^{kin}/2} * (\sigma_f^2)^{-R_2 n_{obs}^{\bar{kin}}/2}$$

$$* \exp \left\{ -\frac{1}{2\sigma_E^2} \left(\sum (E_{(1)} - \kappa_{E1})^2 + \sum (E_{(2)} - \kappa_{E2})^2 + \dots + \sum (E_{(n_{obs}^{kin})} - \kappa_{E n_{obs}^{kin}})^2 \right) \right\}$$

$$* \exp \left\{ -\frac{1}{2\sigma_f^2} \left(\sum (f_{(1)} - \nu_{f1}^{\bar{kin}})^2 + \sum (f_{(2)} - \nu_{f2}^{\bar{kin}})^2 + \sum (f_{(n_{obs}^{\bar{kin}})} - \nu_{fn_{obs}^{\bar{kin}}}^{\bar{kin}})^2 \right) \right\}.$$

$$R_1 = r_{E1} + r_{E2} + \dots + r_{E n_{obs}^{kin}} \quad \text{and} \quad R_2 = r_{f1} + r_{f2} + \dots + r_{fn_{obs}^{\bar{kin}}}.$$

Bayesian Model

Posterior distribution

$$\begin{aligned}
 \pi(\vec{\kappa}, \vec{\nu}^{\bar{kin}}, \sigma_E^2, \sigma_f^2 | \mathbf{E}, \mathbf{f}) &\propto l(\vec{\kappa}, \vec{\nu}^{\bar{kin}}, \sigma_E^2, \sigma_f^2 | \mathbf{E}, \mathbf{f}) * \pi_{prior}(\vec{\kappa}, \vec{\nu}, \sigma_E^2, \sigma_f^2) \\
 &\propto (\sigma_E^2)^{-(a_E+1+R_1 n_{obs}^{\bar{kin}}/2)} * e^{-b_E/\sigma_E^2} * (\sigma_f^2)^{-(a_f+1+R_2 n_{obs}^{\bar{kin}}/2)} * e^{-b_f/\sigma_f^2} \\
 &* \exp\left\{-\frac{1}{2}(\vec{\kappa} - \mu)^t \Sigma_{\kappa}^{-1}(\vec{\kappa} - \mu)\right\} * \mathbb{U}_{\{S^{kin} g^*(\kappa) + S^{\bar{kin}} \vec{\nu}^{\bar{kin}} = 0\} \cap \{\bar{ib}, \bar{ub}\}} \\
 &* \exp\left\{-\frac{1}{2\sigma_E^2} \left(\sum^{r_{E1}} (E_{(1)} - \kappa_{E1})^2 + \sum^{r_{E2}} (E_{(2)} - \kappa_{E2})^2 + \dots + \sum^{r_{E n_{obs}^{\bar{kin}}}} (E_{(n_{obs}^{\bar{kin}})} - \kappa_{E n_{obs}^{\bar{kin}}})^2 \right)\right\} \\
 &* \exp\left\{-\frac{1}{2\sigma_f^2} \left(\sum^{r_{f1}} (f_{(1)} - \nu_{f1}^{\bar{kin}})^2 + \sum^{r_{f2}} (f_{(2)} - \nu_{f2}^{\bar{kin}})^2 + \sum^{r_{f n_{obs}^{\bar{kin}}}} (f_{(n_{obs}^{\bar{kin}})} - \nu_{f n_{obs}^{\bar{kin}}}^{\bar{kin}})^2 \right)\right\}.
 \end{aligned}$$

Bayesian Model

Posterior distribution

$$\begin{aligned}
 \pi(\vec{\kappa}, \vec{\nu}^{\bar{kin}}, \sigma_E^2, \sigma_f^2 | \mathbf{E}, \mathbf{f}) &\propto l(\vec{\kappa}, \vec{\nu}^{\bar{kin}}, \sigma_E^2, \sigma_f^2 | \mathbf{E}, \mathbf{f}) * \pi_{prior}(\vec{\kappa}, \vec{\nu}, \sigma_E^2, \sigma_f^2) \\
 &\propto (\sigma_E^2)^{-(a_E+1+R_1 n_{obs}^{\bar{kin}}/2)} * e^{-b_E/\sigma_E^2} * (\sigma_f^2)^{-(a_f+1+R_2 n_{obs}^{\bar{kin}}/2)} * e^{-b_f/\sigma_f^2} \\
 &* \exp\left\{-\frac{1}{2}(\vec{\kappa} - \mu)^t \Sigma_{\kappa}^{-1}(\vec{\kappa} - \mu)\right\} * \mathbb{U}_{\{S^{kin} g^*(\kappa) + S^{\bar{kin}} \vec{\nu}^{\bar{kin}} = 0\} \cap \{\vec{lb}, \vec{ub}\}} \\
 &* \exp\left\{-\frac{1}{2\sigma_E^2} \left(\sum^{r_{E1}} (E_{(1)} - \kappa_{E1})^2 + \sum^{r_{E2}} (E_{(2)} - \kappa_{E2})^2 + \dots + \sum^{r_{E n_{obs}^{\bar{kin}}}} (E_{(n_{obs}^{\bar{kin}})} - \kappa_{E n_{obs}^{\bar{kin}}})^2 \right)\right\} \\
 &* \exp\left\{-\frac{1}{2\sigma_f^2} \left(\sum^{r_{f1}} (f_{(1)} - \nu_{f1}^{\bar{kin}})^2 + \sum^{r_{f2}} (f_{(2)} - \nu_{f2}^{\bar{kin}})^2 + \sum^{r_{f n_{obs}^{\bar{kin}}}} (f_{(n_{obs}^{\bar{kin}})} - \nu_{f n_{obs}^{\bar{kin}}}^{\bar{kin}})^2 \right)\right\}.
 \end{aligned}$$

Bayesian Model

Marginal posterior distributions

- $$\pi(\sigma_E^2 | a_E, b_E, \vec{\kappa}, \mathbf{E}) \sim \text{IG} \left(a_E + R_1 \frac{n_{obs}^{kin}}{2}, b_E + \frac{1}{2} \left(\sum^{rE1} (E_{(1)} - \kappa_{E1})^2 + \sum^{rE2} (E_{(2)} - \kappa_{E2})^2 + \dots + \sum^{rE n_{obs}^{kin}} (E_{(n_{obs}^{kin})} - \kappa_{E n_{obs}^{kin}})^2 \right) \right),$$

- $$\pi(\sigma_f^2 | a_f, b_f, \vec{\nu}^{kin}, \mathbf{f}) \sim \text{IG} \left(a_f + R_2 \frac{n_{obs}^{kin}}{2}, b_f + \frac{1}{2} \left(\sum^{rf1} (f_{(1)} - \nu_{f1}^{kin})^2 + \sum^{rf2} (f_{(2)} - \nu_{f2}^{kin})^2 + \sum^{r f n_{obs}^{kin}} (f_{(n_{obs}^{kin})} - \nu_{f n_{obs}^{kin}}^{kin})^2 \right) \right),$$

- $$\pi(\vec{\kappa} | \mu, \Sigma_{\kappa}, \sigma_E^2, \mathbf{E}) \sim \mathbb{N}_{n_{kin}}(\mu, \Sigma)$$

$$\mu_i = \begin{cases} \frac{r_{Ei} \bar{E}_{(i)} \sigma_{\kappa i}^2 + \mu_{\kappa i} \sigma_E^2}{r_{Ei} \sigma_{\kappa i}^2 + \sigma_E^2} & \text{if } E_i \text{ is observed,} \\ \mu_{i(nobs)} & \text{otherwise.} \end{cases}, \sigma_{ii}^2 = \begin{cases} \left(\frac{r_{Ei}}{\sigma_E^2} + \frac{1}{\sigma_{\kappa i}^2} \right)^{-1} & \text{if } E_i \text{ is observed,} \\ \sigma_{\kappa i(nobs)} & \text{otherwise.} \end{cases}$$

- $$\pi(\vec{\nu}^{kin} | \sigma_f^2, \vec{l}b, \vec{u}b, g^*(\vec{\kappa}), \mathbf{f})$$

$$\propto \exp \left\{ -\frac{1}{2\sigma_f^2} \left(\sum^{rf1} (f_{(1)} - \nu_{f1}^{kin})^2 + \sum^{rf2} (f_{(2)} - \nu_{f2}^{kin})^2 + \sum^{r f n_{obs}^{kin}} (f_{(n_{obs}^{kin})} - \nu_{f n_{obs}^{kin}}^{kin})^2 \right) \right\}$$

$$*\mathbb{U}_{\{S^{kin} g^*(\kappa) + S^{kin} \vec{\nu}^{kin} = 0\} \cap \{\vec{l}b, \vec{u}b\}}$$

Bayesian Model

Algorithm

- **Experimental values:** $\{\mathbf{E}, \mathbf{f}\}$.
- **Initial values:** $\{\bar{\kappa}, \bar{\nu}^{kin}, \sigma_E^2, \sigma_f^2\}$.
- **A priori knowledge:** $\{\Sigma_{\kappa}, \bar{\mu}, a_E, b_E, a_f, b_f\}$.

Integration of the kinetic model.

- 1 Let $k = 0, 1, \dots, K$. Initial values $\bar{\kappa}_{(0)} = \bar{\kappa}$ and $\bar{\nu}_{(0)}^{kin} = \bar{\nu}^{kin}$.
- 2 Generate $\sigma_{E(k+1)}^2$ and $\sigma_{f(k+1)}^2$.
- 3 Generate $\bar{\kappa}_{(k+1)}$.

Integration of metabolic data

- 4 Compute $\mathbf{g}^{*(k+1)}(\bar{\kappa}_{(k+1)})$.
- 5 Generate $\bar{\nu}_{(k+1)}^{kin}$ Hit&Run, with probability of acceptance

$$\alpha(\bar{\nu}_{(k)}^{kin}, \bar{\nu}_{(k+1)}^{kin}) = \min \left\{ \frac{\psi(\bar{\nu}_{(k+1)}^{kin}(obs))}{\psi(\bar{\nu}_{(k)}^{kin}(obs))}, 1 \right\} \text{ where}$$

$$\psi(\cdot) \equiv \exp \left\{ -\frac{1}{2\sigma_f^2} \left(\sum^r f_1(f_{(1)} - \nu_{f_1}^{kin})^2 + \dots \right) \right\}.$$

- 6 **Acceptance/Rejection probability of the whole vector** $\boldsymbol{\theta}_{(k+1)} = \{\sigma_{E(k+1)}^2, \sigma_{f(k+1)}^2, \bar{\kappa}_{(k+1)}, \bar{\nu}_{(k+1)}^{kin}\}$

$$\eta(\boldsymbol{\theta}_{(k)}, \boldsymbol{\theta}_{(k+1)}) = \min \left\{ \frac{\pi(\boldsymbol{\theta}_{(k+1)} | \{\mathbf{E}, \mathbf{f}\})}{\pi(\boldsymbol{\theta}_{(k)} | \{\mathbf{E}, \mathbf{f}\})}, 1 \right\} \text{ where } \pi(\cdot | \cdot) \text{ is the (joint) posterior distribution.}$$

- 7 Repeat steps (2-6) until have K iterations.

Merci !!!

Moulon, Univ Paris-Sud

Dominique de Vienne

Olivier Martin

Victor Sabarly

Thibault Nidelet

Thierry Balliau

Bichat, INSERM U722, Paris

Eric Denamur

Odile Bouvet

Cécile Aubron, Jeremy Glodt

Max Plank Institut, Leipzig

Pierre-Yves Bourguignon

Genoscope, CEA, Evry

Claudine Medigue

Maxime Durot

Gilles Vieira

François Lefevre

David Vallenet

Lizbeth Roman-Padilla

