An advanced stochastic framework for the simulation of transgenerational hologenomic data

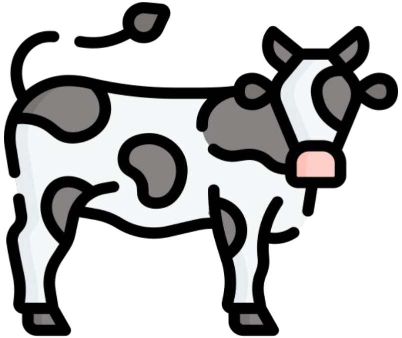
Solène Pety

Supervised by :
Andrea Rau (INRAE, GABI),
Mahendra Mariadassou (INRAE, MaIAGE),
Ingrid David (INRAE, GenPhySE)

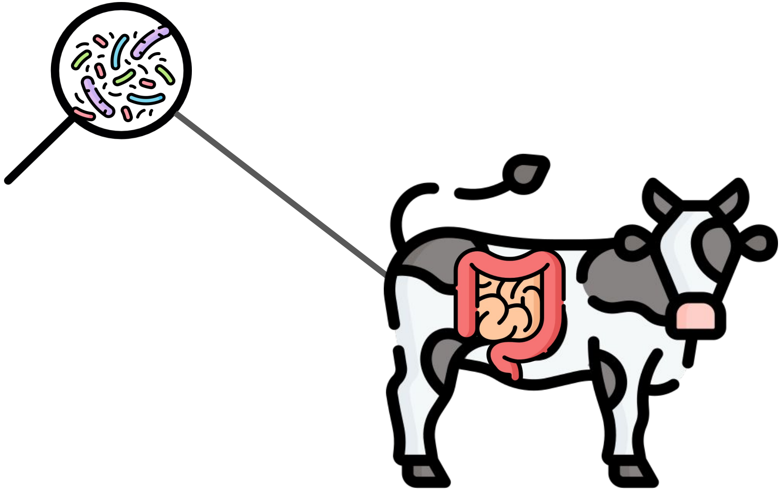
Funded by the **HOLOBIONTS** targeted project of the **PEPR Agroécologie et Numérique**, INRAE **HoloFlux** metaprogram label.

SMPGD presentation - 24/01/2025

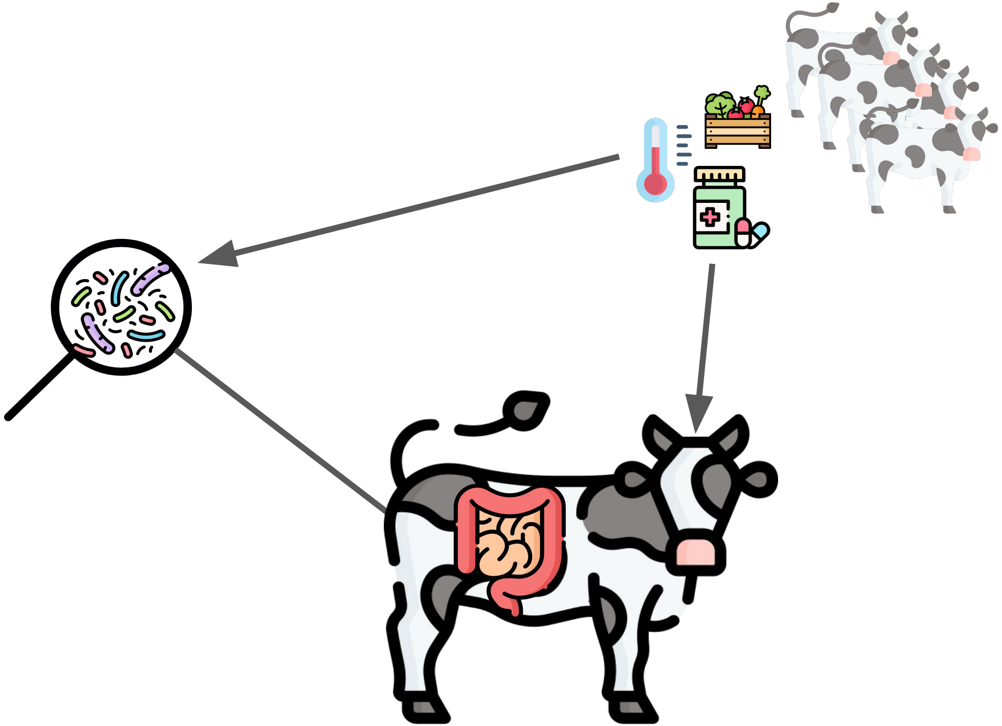
A complex entity formed by a host and its microbiome



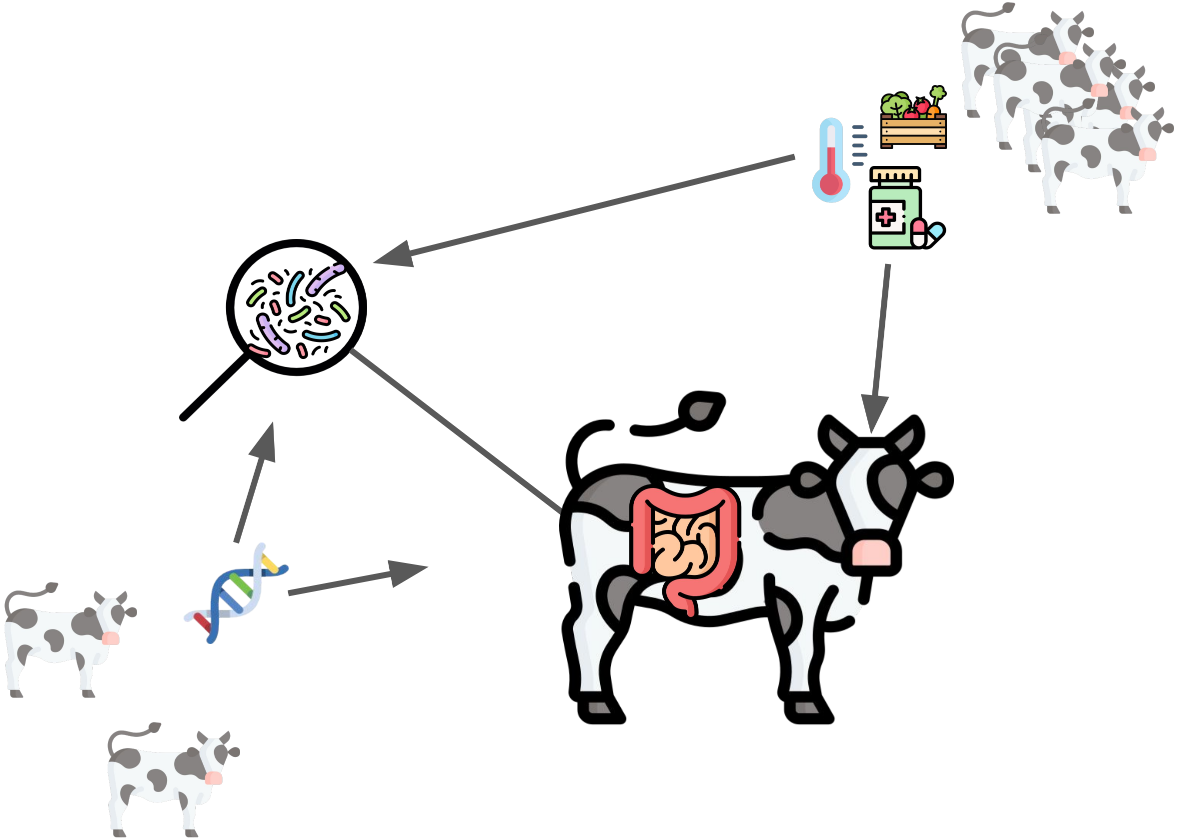
A complex entity formed by a host and its microbiome



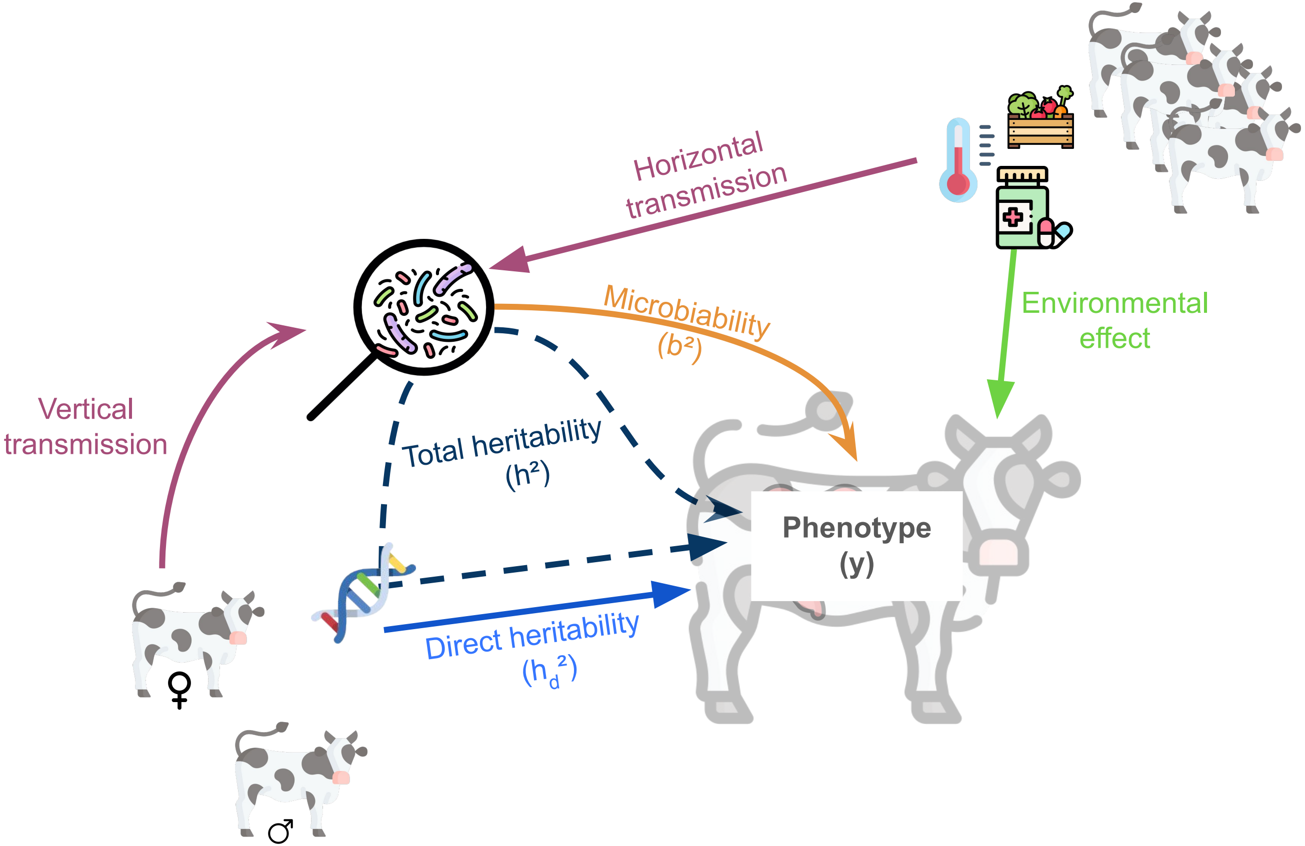
A complex entity formed by a host and its microbiome



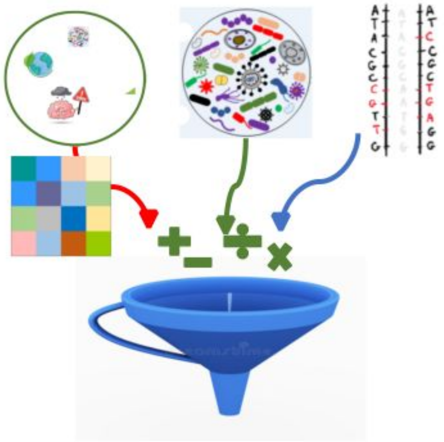
A complex entity formed by a host and its microbiome



A complex entity formed by a host and its microbiome



Theis et al. 2016

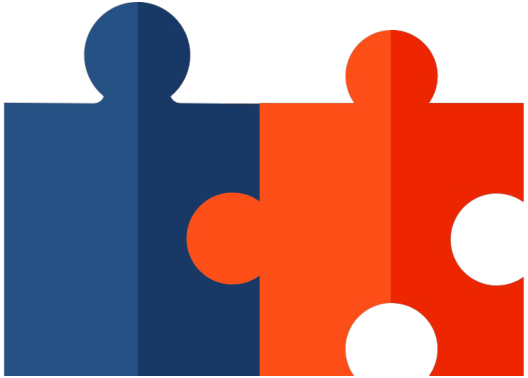


Could the integration of hologenomic variability improve the prediction of phenotypes and genetic values?

Complexity of integrating microbiome and genomic data

SIM

Complexity of integrating microbiome and genomic data



SIMULAT

Need to know the truth to be able to evaluate prediction models

Complexity of integrating microbiome and genomic data

Data collection is expensive and time consuming

SIMULATION

Need to know the truth to be able to evaluate prediction models

R **I**mplementation of a **T**ransgenerational **H**ologenomic **M**odel-based **S**imulator

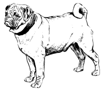
(1) Use **real genotype and microbiota** data to initialize simulation,

(2) Simulate **coupled genomic, microbiota, and phenotype** data,

(3) Identify reproducing animals using a **selection criterion**,

(4) Repeat over multiple generations to **obtain simulated transgenerational hologenomic data.**

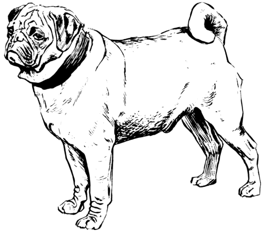
Our model



Genotypes

Number of individuals (n_{ind}),
No overlap between generations,
Sex assignment from sex ratio

Genotypes are simulated with MoBPS package

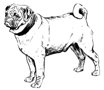


From real data

- Realistic linkage disequilibrium
- Realistic allelic frequency distribution

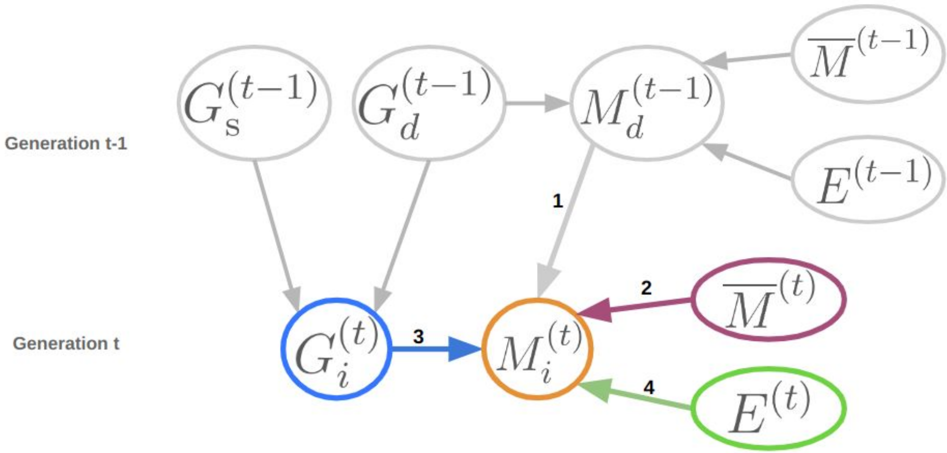
Pook et al. 2020

Our model



Genotypes

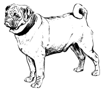
Individual number (n_{ind}),
No overlap between generation,
Sex assignment from sex ratio



Genotypes are simulated with MoBPS package

Simulated microbiomes account for complexity of microbiota architecture

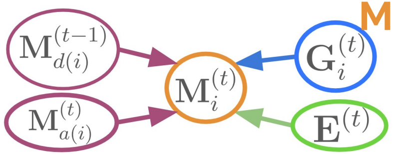
Our model



Genotypes

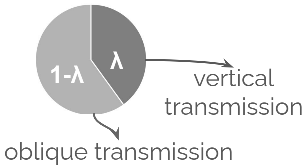
Individual number (n_{ind}),
No overlap between generation,
Sex assignment from sex ratio

Microbiomes

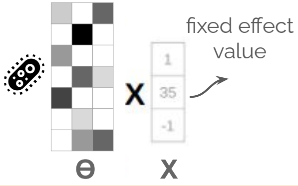


$$\mathrm{CLR}(\boldsymbol{M}_i^{(t)}) = \mathrm{CLR}(\lambda \boldsymbol{M}_{m(i)}^{(t-1)} + (1-\lambda)\boldsymbol{M}_{a(i)}^{(t)}) + \boldsymbol{\theta}^{(t)}(\boldsymbol{X}_i^{(t)})^T + \boldsymbol{\beta}\boldsymbol{G}_i^{(t)} + \boldsymbol{\epsilon}_i^{(t)}$$

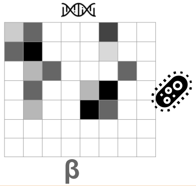
Transmissibility



Environmental effect



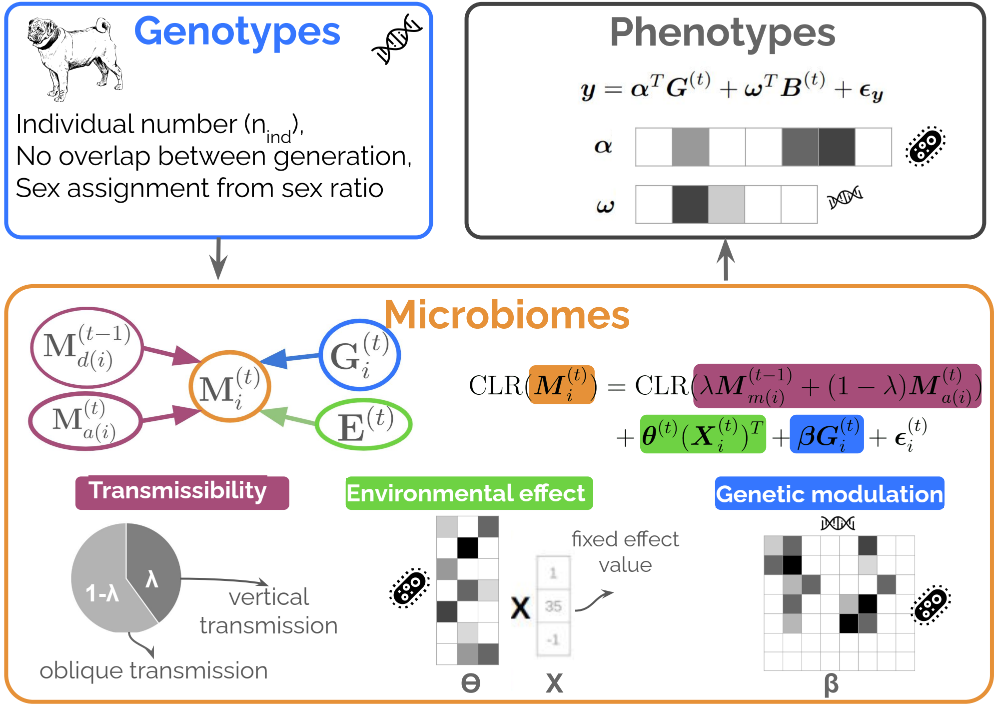
Genetic modulation



Genotypes are simulated with MoBPS package

Simulated microbiomes account for complexity of microbiota architecture

Our model



Genotypes

Individual number (n_{ind}),
No overlap between generation,
Sex assignment from sex ratio

Phenotypes

$$\boldsymbol{y} = \boldsymbol{\alpha}^T \boldsymbol{G}^{(t)} + \boldsymbol{\omega}^T \boldsymbol{B}^{(t)} + \epsilon_{\boldsymbol{y}}$$

Microbiomes

$$\mathrm{CLR}(\boldsymbol{M}_i^{(t)}) = \mathrm{CLR}(\lambda \boldsymbol{M}_{m(i)}^{(t-1)} + (1-\lambda)\boldsymbol{M}_{a(i)}^{(t)}) + \boldsymbol{\theta}^{(t)}(\boldsymbol{X}_i^{(t)})^T + \boldsymbol{\beta}\boldsymbol{G}_i^{(t)} + \epsilon_i^{(t)}$$

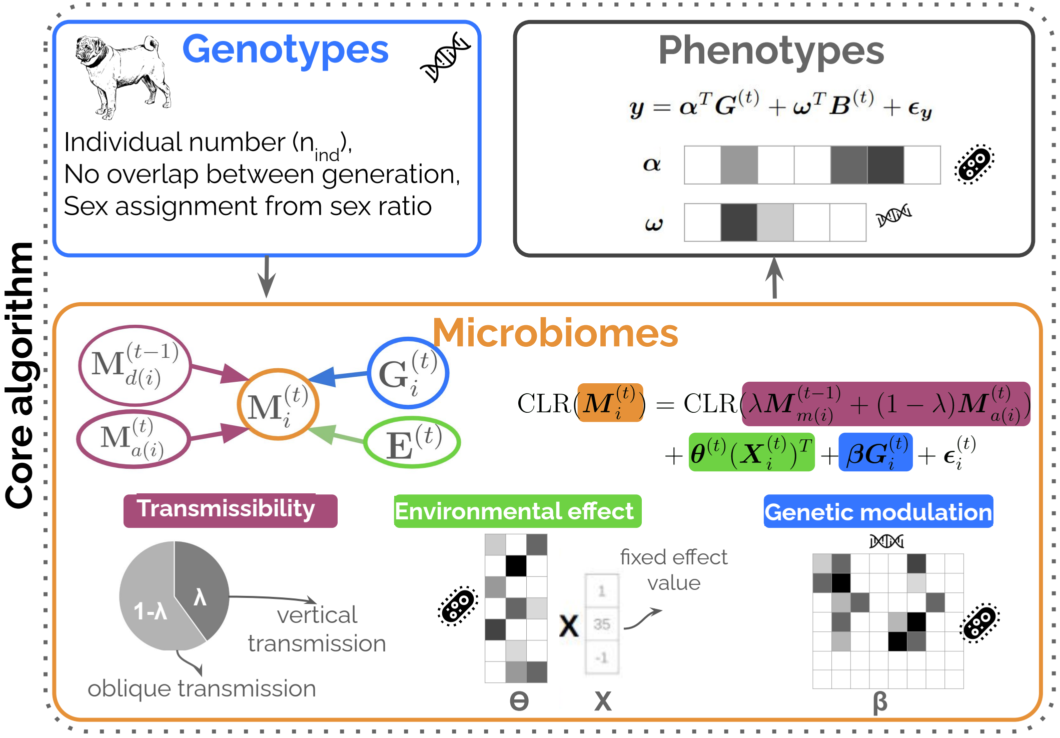
Genotypes are simulated with MoBPS package

Simulated microbiomes account for complexity of microbiota architecture

Phenotype values have a genetic and a microbiota component

Perez-Enciso et al. 2021

Our model



Genotypes

Individual number (n_{ind}),
No overlap between generation,
Sex assignment from sex ratio

Phenotypes

$$\boldsymbol{y} = \boldsymbol{\alpha}^T \boldsymbol{G}^{(t)} + \boldsymbol{\omega}^T \boldsymbol{B}^{(t)} + \boldsymbol{\epsilon_y}$$

Microbiomes

$$\mathrm{CLR}(\boldsymbol{M}_i^{(t)}) = \mathrm{CLR}(\lambda \boldsymbol{M}_{m(i)}^{(t-1)} + (1-\lambda)\boldsymbol{M}_{a(i)}^{(t)}) + \boldsymbol{\theta}^{(t)}(\boldsymbol{X}_i^{(t)})^T + \boldsymbol{\beta}\boldsymbol{G}_i^{(t)} + \boldsymbol{\epsilon}_i^{(t)}$$

Transmissibility

Environmental effect

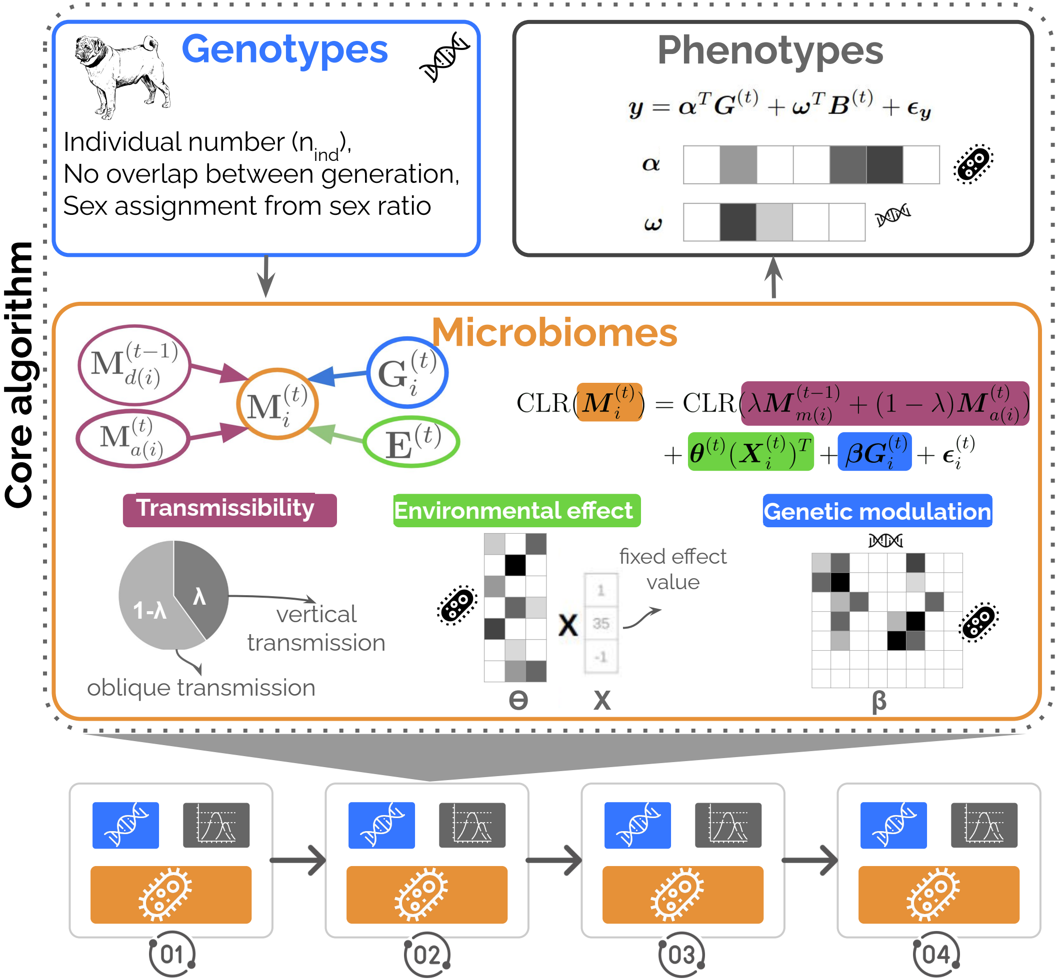
Genetic modulation

Genotypes are simulated with MoBPS package

Simulated microbiomes account for complexity of microbiota architecture

Phenotype values have a genetic and a microbiota component

Our model



$$y = \alpha^T G^{(t)} + \omega^T B^{(t)} + \epsilon_y$$

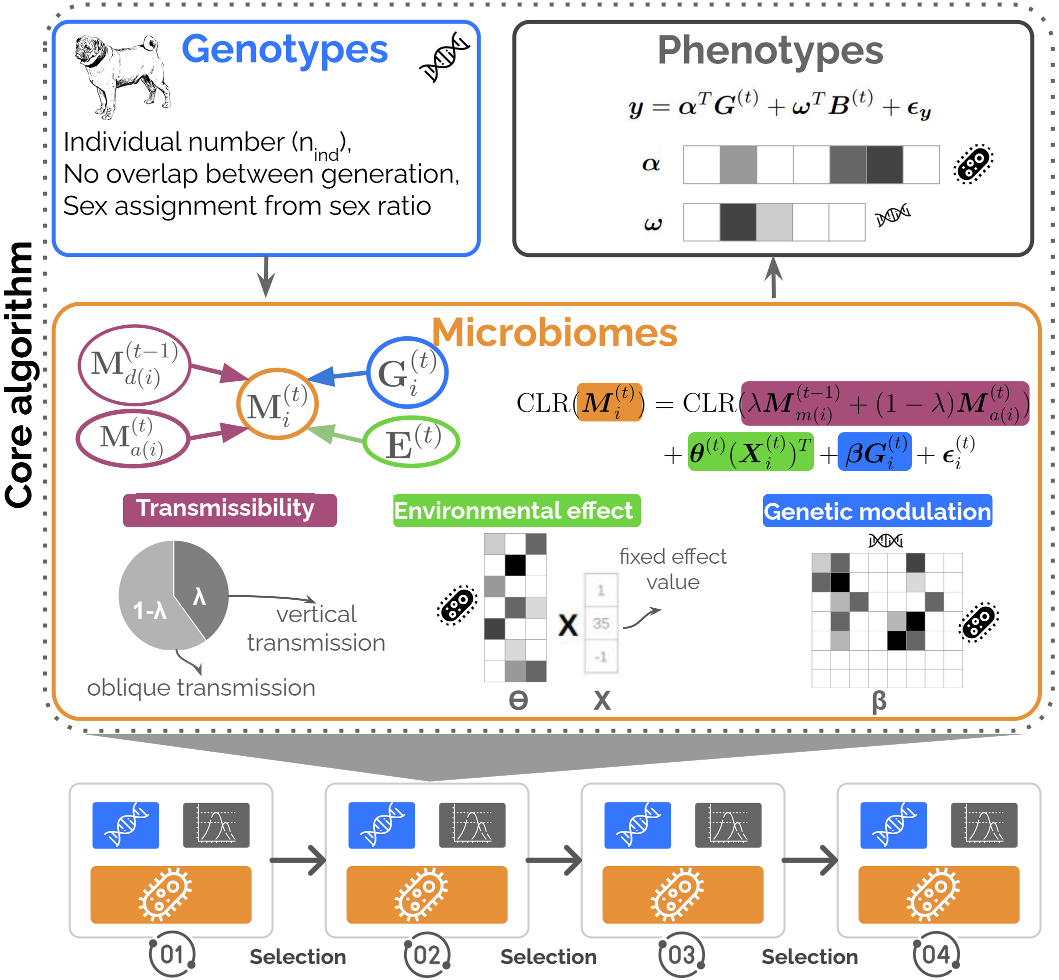
$$\mathrm{CLR}(M_i^{(t)}) = \mathrm{CLR}(\lambda M_{m(i)}^{(t-1)} + (1-\lambda) M_{a(i)}^{(t)}) + \theta^{(t)}(X_i^{(t)})^T + \beta G_i^{(t)} + \epsilon_i^{(t)}$$

Genotypes are simulated with MoBPS package

Simulated microbiomes account for complexity of microbiota architecture

Phenotype values have a genetic and a microbiota component

Our model



Core algorithm

Genotypes

Individual number (n_{ind}),
No overlap between generation,
Sex assignment from sex ratio

Phenotypes

$$y = \alpha^T G^{(t)} + \omega^T B^{(t)} + \epsilon_y$$

Microbiomes

$$\mathrm{CLR}(M_i^{(t)}) = \mathrm{CLR}(\lambda M_{m(i)}^{(t-1)} + (1-\lambda) M_{a(i)}^{(t)}) + \theta^{(t)} (X_i^{(t)})^T + \beta G_i^{(t)} + \epsilon_i^{(t)}$$

Transmissibility

Environmental effect

Genetic modulation

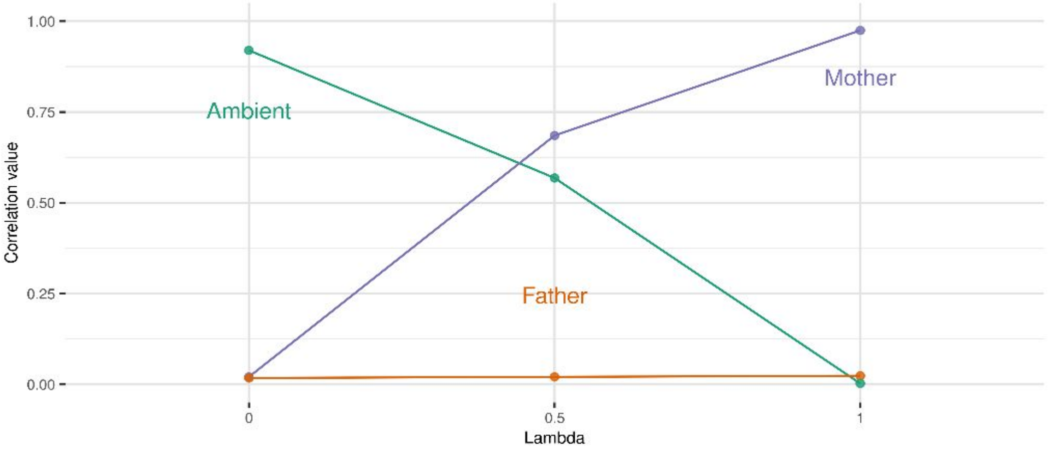
Genotypes are simulated with MoBPS package

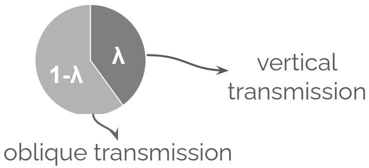
Simulated microbiomes account for complexity of microbiota architecture

Phenotype values have a genetic and a microbiota component

A selection step is performed between two generations according to a criterion (genetic or microbiota values)

RITHMS allows tuning of maternal and environmental contributions

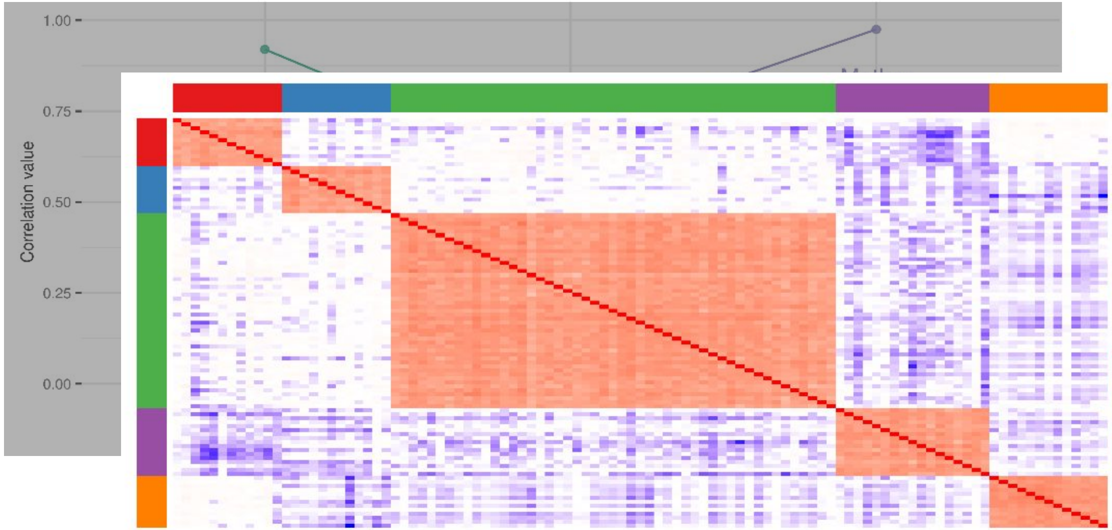




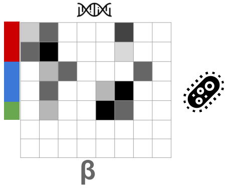
Correlation between α-diversity of an offspring and that of either its parents or the ambient microbiota for increasing λ

Results using base population made up of a subset of n=750 pigs from Déru et al. 2020

RITHMS allows the definition of complex taxa abundance dependency structures

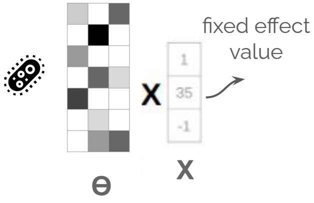


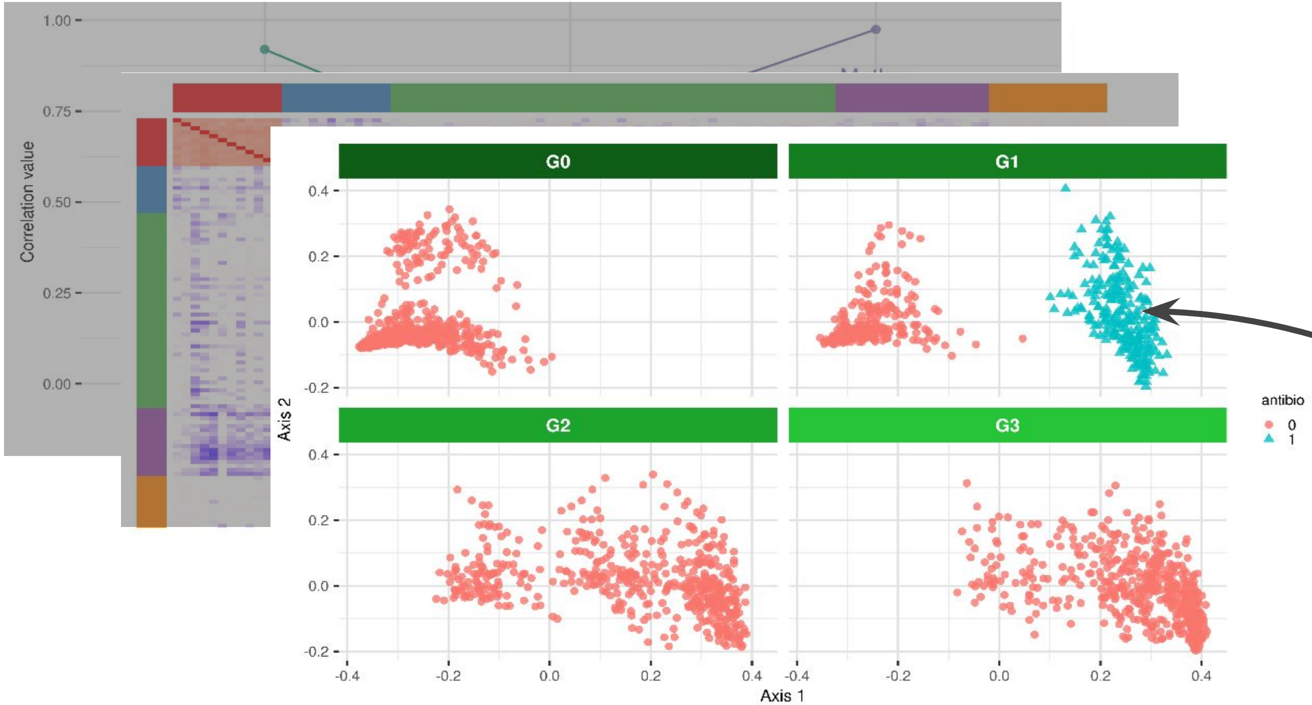
Pairwise correlation of abundances for taxa under genetic control, grouped into five clusters



Results using base population made up of a subset of n=750 pigs from Déru et al. 2020

RITHMS allows for the introduction of transient perturbations of the microbiota

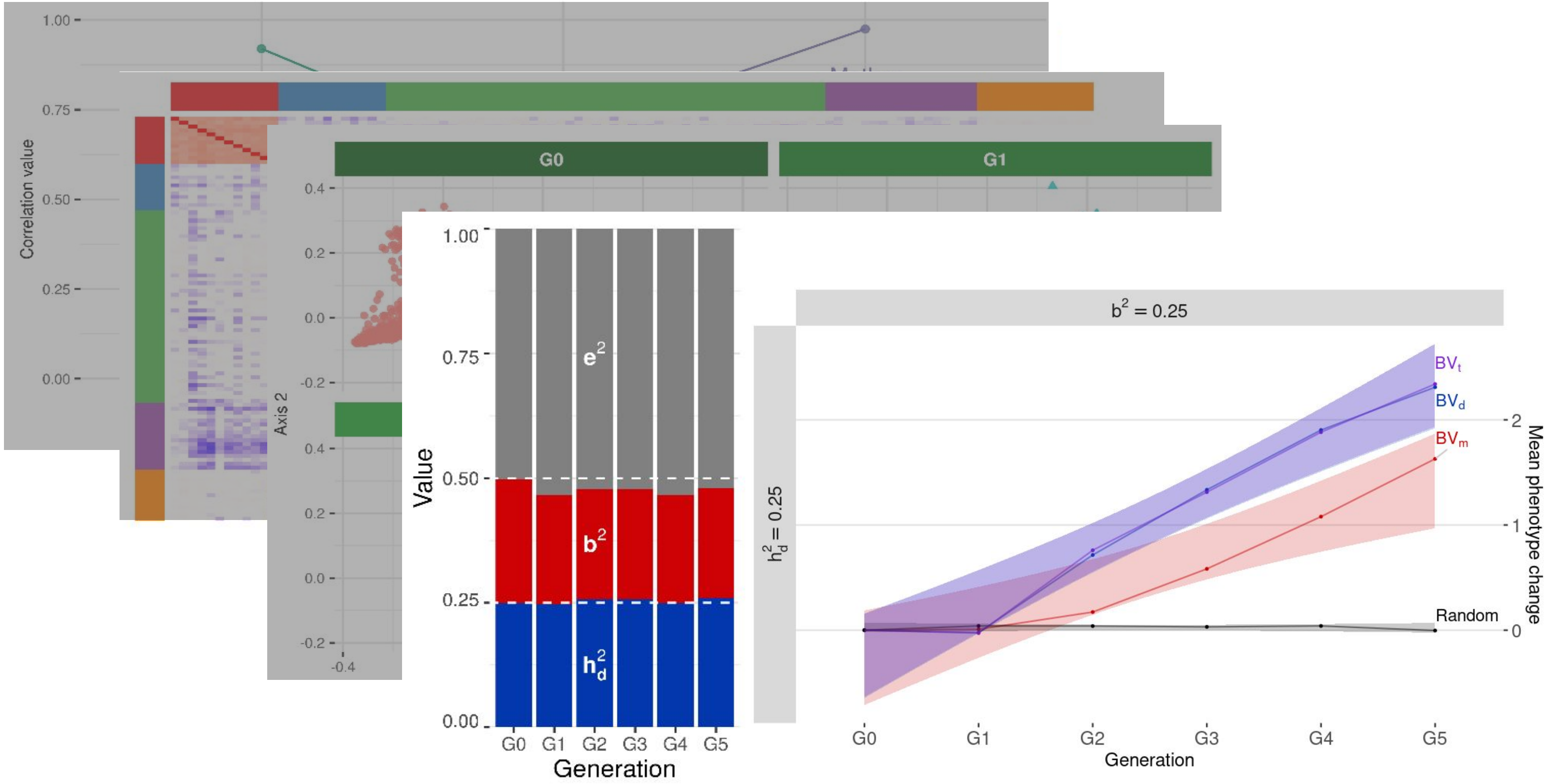




Multidimensional scaling of microbial abundances (Bray-Curtis distance) after transient perturbation

Results using base population made up of a subset of n=750 pigs from Déru et al. 2020

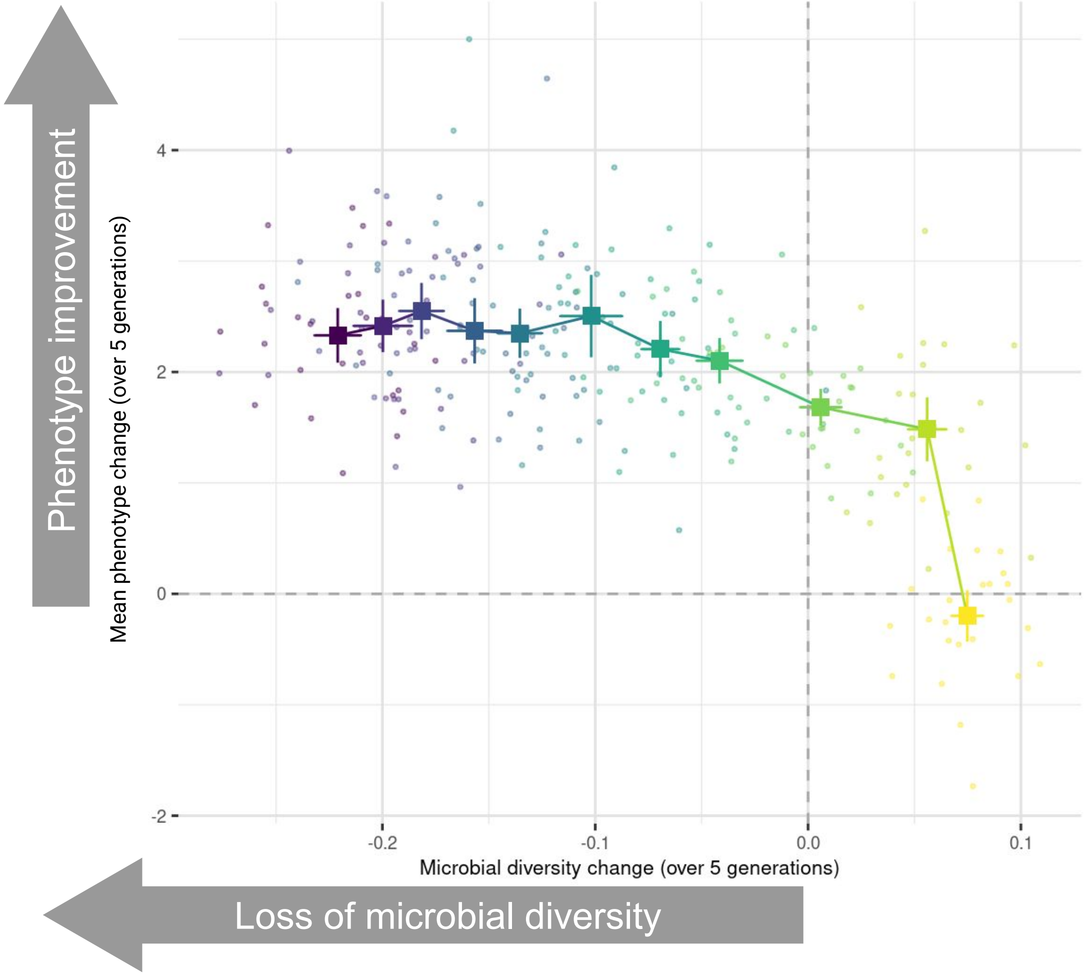
RITHMS allows a fine control of heritability, microbiability and selection schemes



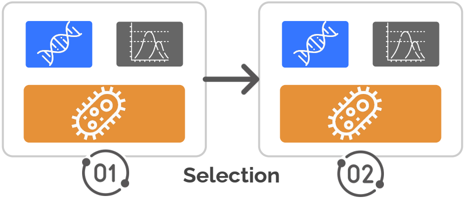
Observed direct heritability (h_d^2) and microbiability (b^2) without selection (target values $h_d^2 = b^2 = 0.25$)
Mean phenotype change across generations according to selection strategy

Results using base population made up of a subset of n=750 pigs from Déru et al. 2020

Example of a selection index based on a combination of trait and diversity metrics

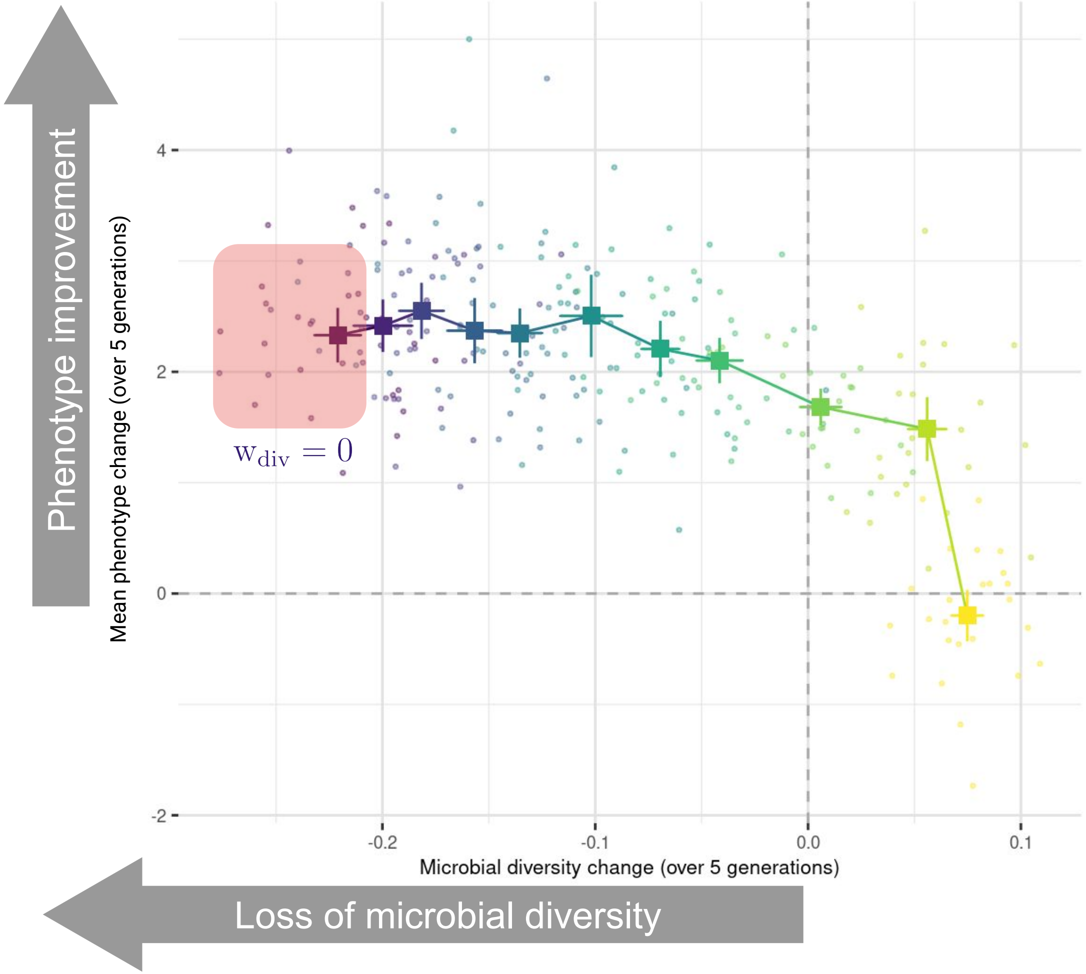


$$\mathrm{w}_{\mathrm{div}}.\boldsymbol{\delta}^{(t)} + (1 - \mathrm{w}_{\mathrm{div}}).\mathbf{BV}_t^{(t)}$$



30% of females
30% of males

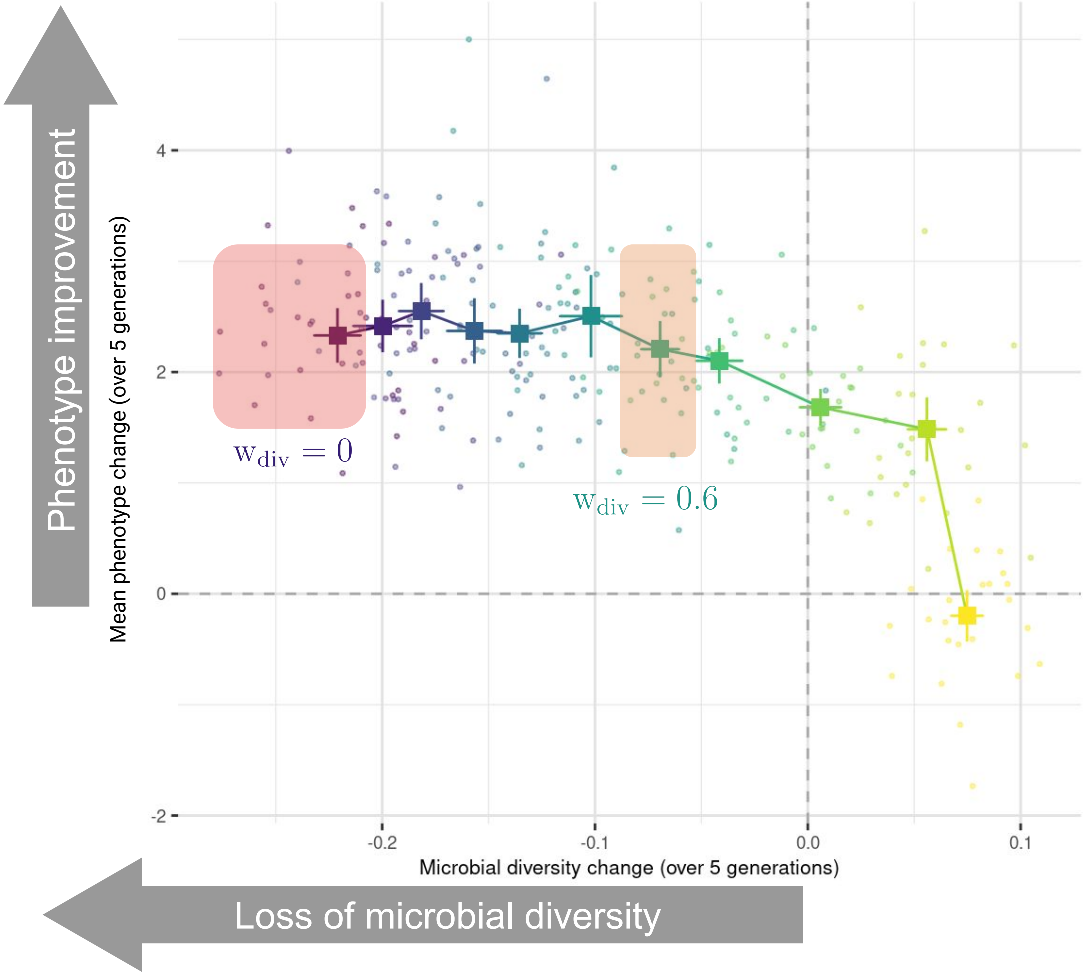
Example of a selection index based a combination of trait and diversity metrics



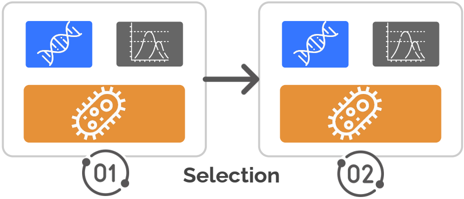
$$w_{div}.\boldsymbol{\delta}^{(t)} + (1 - w_{div}).\mathbf{BV}_t^{(t)}$$

30% of females
30% of males

Example of a selection index based a combination of trait and diversity metrics



$$w_{\text{div}}.\boldsymbol{\delta}^{(t)} + (1 - w_{\text{div}}).\mathbf{BV}_t^{(t)}$$



30% of females
30% of males

Conclusion and future work

Conclusion and future work

Use **real genotype and microbiota** data to initialize simulation, ✓

Conclusion and future work

Use **real genotype and microbiota** data to initialize simulation, ✓

Simulate **coupled genomic, microbiota, and phenotype** data, ✓

Conclusion and future work

Use **real genotype and microbiota** data to initialize simulation, ✓

Simulate **coupled genomic, microbiota, and phenotype** data, ✓

Identify reproducing animals using **selection criterion**, ✓

Conclusion and future work

Use **real genotype and microbiota** data to initialize simulation, ✓

Simulate **coupled genomic, microbiota, and phenotype** data, ✓

Identify reproducing animals using **selection criterion**, ✓

Repeat over multiple generations to **obtain simulated transgenerational hologenomic data** ✓

Conclusion and future work

Use **real genotype and microbiota** data to initialize simulation, ✓

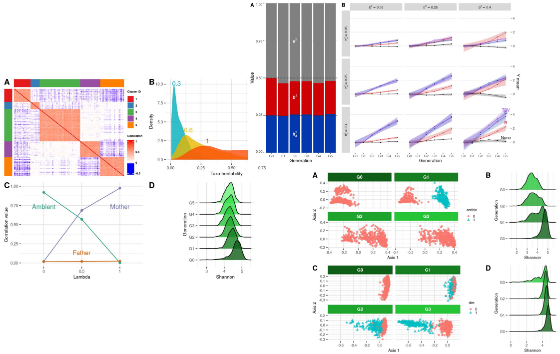
Simulate **coupled genomic, microbiota, and phenotype** data, ✓

Identify reproducing animals using **selection criterion**, ✓

Repeat over multiple generations to **obtain simulated transgenerational hologenomic data** ✓

Additional **developments** and **use cases** using a variety of hologenomic **datasets**

Described in a **preprint**



Conclusion and future work

Use **real genotype and microbiota** data to initialize simulation, ✓

Simulate **coupled genomic, microbiota, and phenotype** data, ✓

Identify reproducing animals using **selection criterion**, ✓

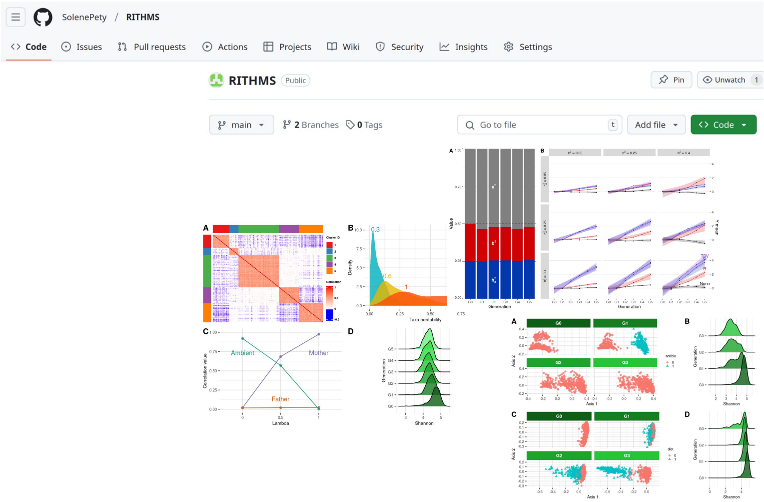
Repeat over multiple generations to **obtain simulated transgenerational hologenomic data** ✓

Additional **developments** and **use cases** using a variety of hologenomic **datasets**

Described in a **preprint**

The package is **available** on GitHub.

https://solenepety.github.io/RITHMS/



https://solenepety.github.io/RITHMS/

Funded by the **HOLOBIONTS** targeted project of the **PEPR Agroécologie et Numérique**, INRAE **HoloFlux** metaprogram label.

Thanks for your attention !

solene.pety@inrae.fr to reach me