# Computational models of crop plant microbial biodiversity

*David James Sherman, Inria*
*Pleiade – joint Inria-INRAE project team*
*MISTIC – PEPR Agroécologie numérique*

Journées INRAE–Inria 2023-07-05

https://project.inria.fr/mistic

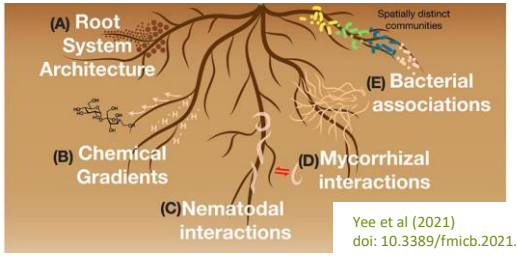
# Agroecology: mobilize the biological processes themselves

## Natural partnership between INRAE & Inria

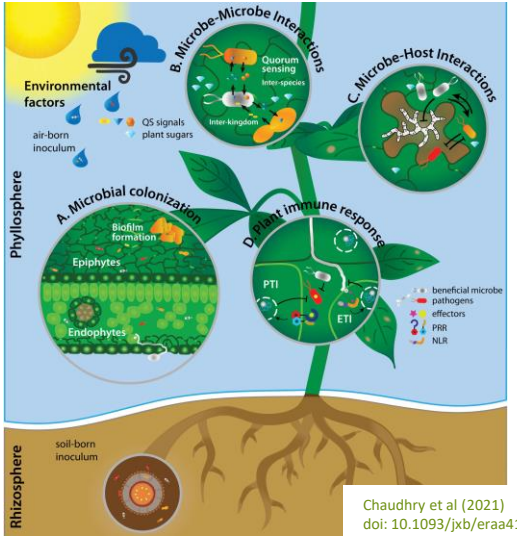
MISTIC: Agnotobiotic **microbial communities** in crop plants

- Beneficial and deleterious effects
- Naturally occurring — property of the land, *le terroir*
- Complex interactions

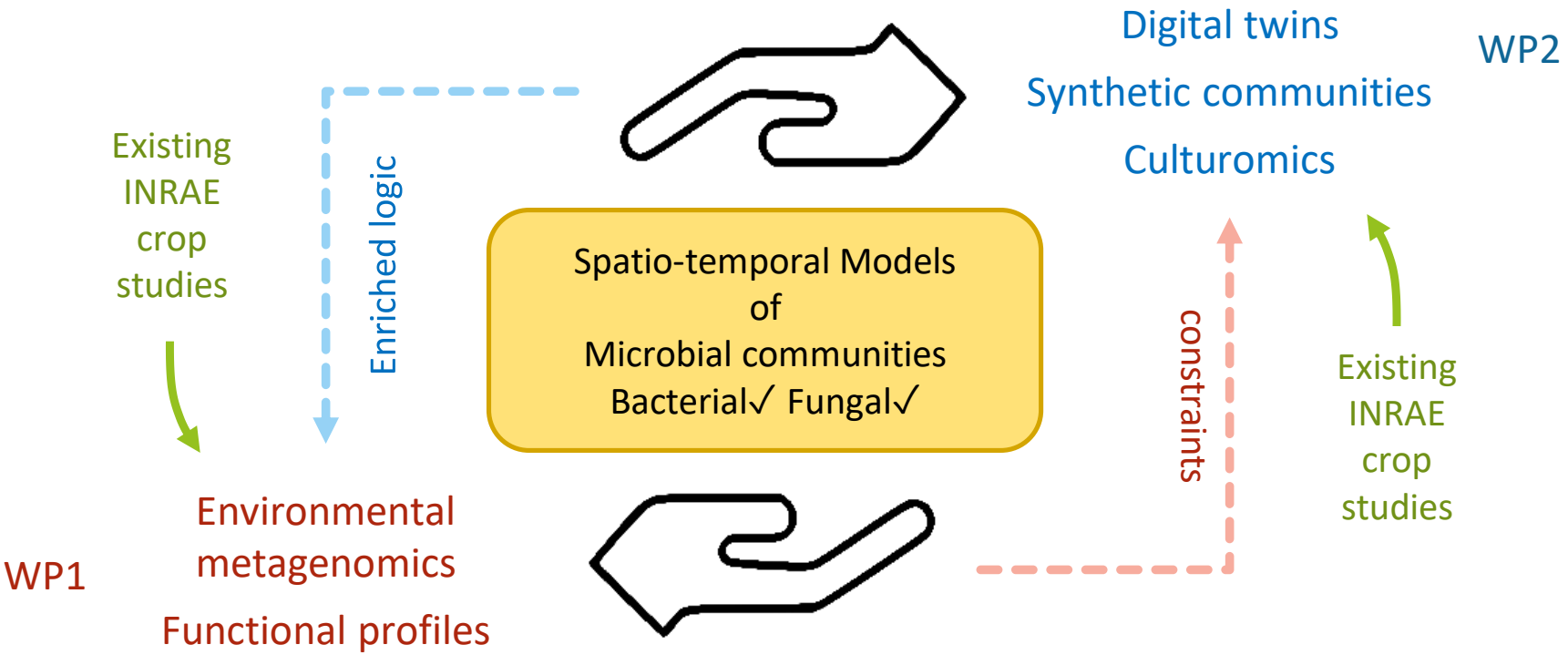
Key role in adaptation to environmental stresses including climate change




Yee et al (2021)
doi: 10.3389/fmicb.2021.625752




Chaudhry et al (2021)
doi: 10.1093/jxb/eraa417

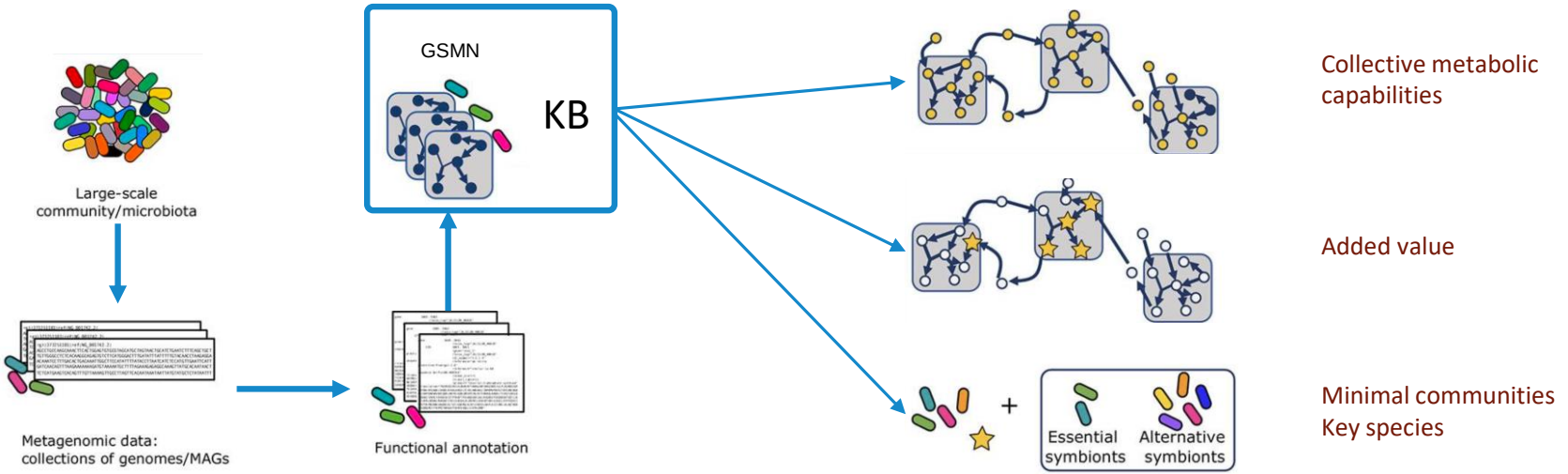

Digital twins
Synthetic communities
Culturomics
WP2
Existing
INRAE
crop
studies
Enriched logic
Spatio-temporal Models
of
Microbial communities
Bacterial✓ Fungal✓
constraints
Existing
INRAE
crop
studies
Environmental
metagenomics
Functional profiles
WP1

# Community-scale Metabolism is Hard

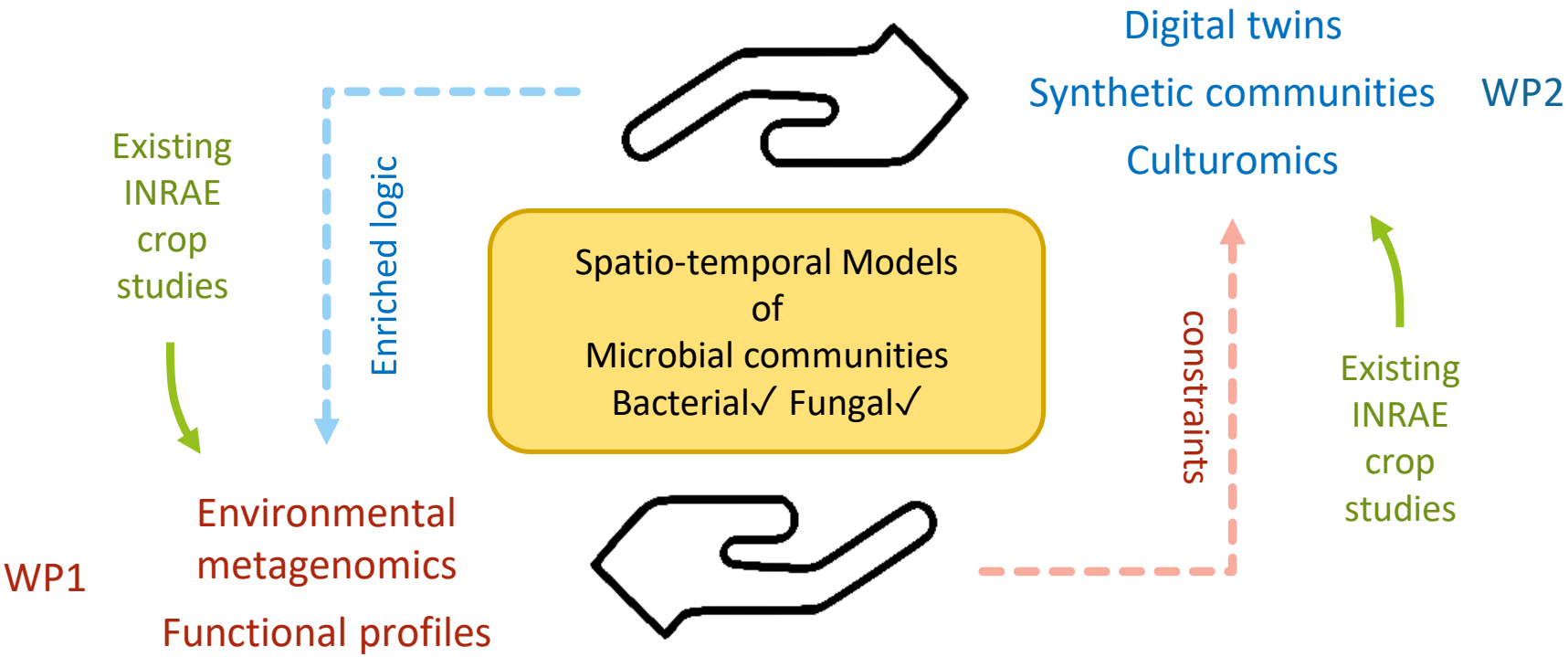
- Natural communities have $10^2$–$10^3$ interacting species
- Bag-of-enzyme data is hard to acquire
- Formal correctness is hard to audit

## *Machine reasoning*

- Large knowledge base of metabolic facts
- Inference using auditable rules
- Hybrid ASP + LP solving

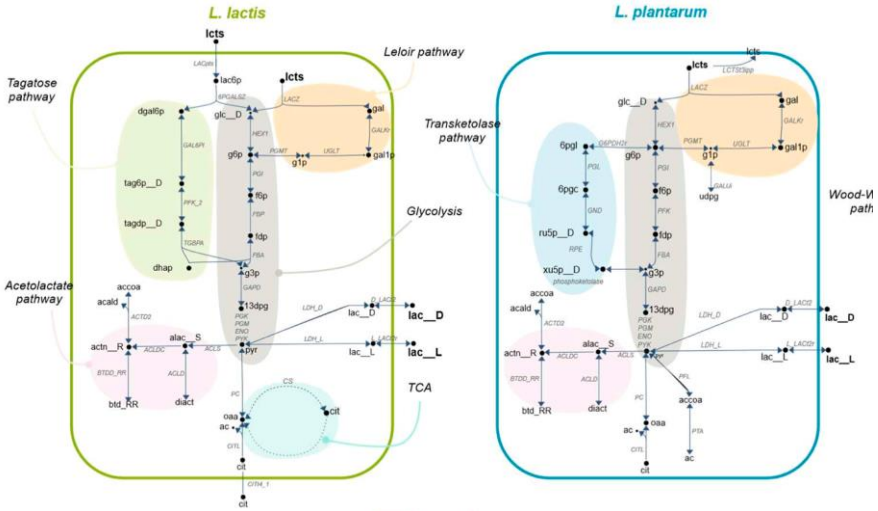



Frioux et al (2019) doi:10.1017/S1471068418000455
Belcour et al (2020) doi:10.7554/eLife.61968
Frioux et al (2020) doi:10.1016/j.csbj.2020. 06.028
Frioux et al (2023) doi:10.1016/j.chom.2023.05.024


Digital twins
Synthetic communities
Culturomics
WP2
Existing INRAE crop studies
Enriched logic
Spatio-temporal Models of Microbial communities Bacterial✓ Fungal✓
constraints
Existing INRAE crop studies
WP1
Environmental metagenomics
Functional profiles

# Digital Twins are Hard

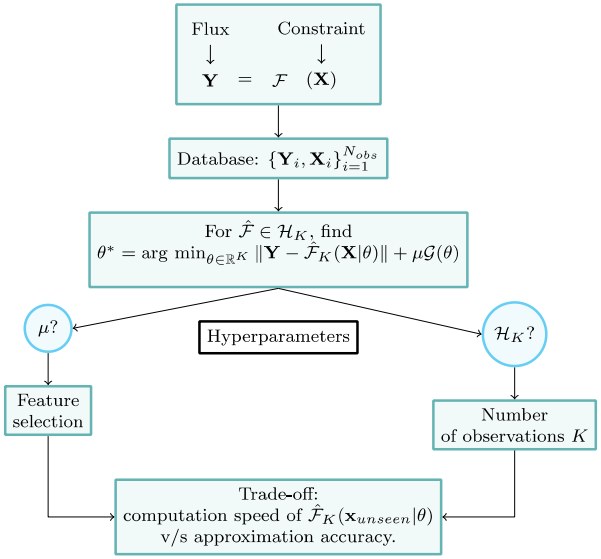
- Ex.: > 8500 person-hours by trained scientists and technicians
- Simulation – hypothesis – experimentation – biological interpretation – model refinement – repeat**!**
- How to accelerate simulation?




Lecomte et al submitted

## *Machine learning of surrogate models*

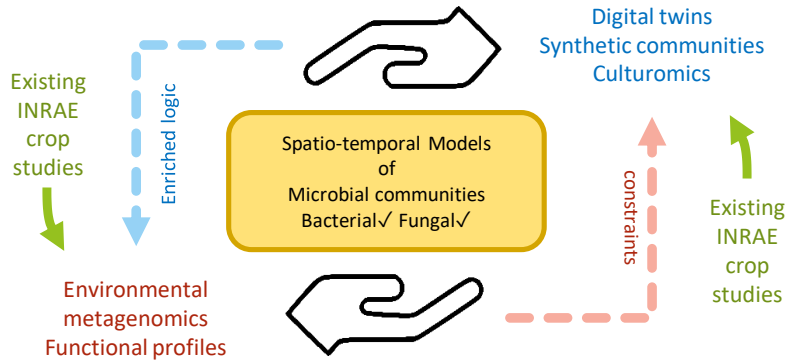
- Learn accelerated model from database of simulations
- Replace computationally intensive simulation step




Labarthe et al (2019) doi:10.1016/j.jtbi.2018.12.009
Kamari (2019) arXiv:1905.13695
Frioux et al (2023) in press

## Methodological outcomes

- Numerical and discrete modeling
- Supervised, unsupervised, and statistical learning
- Explainable model-agnostic AI
  - Counterfactuals
  - Anchoring conditions
  - Additive contributions/benefits
- Scaling up

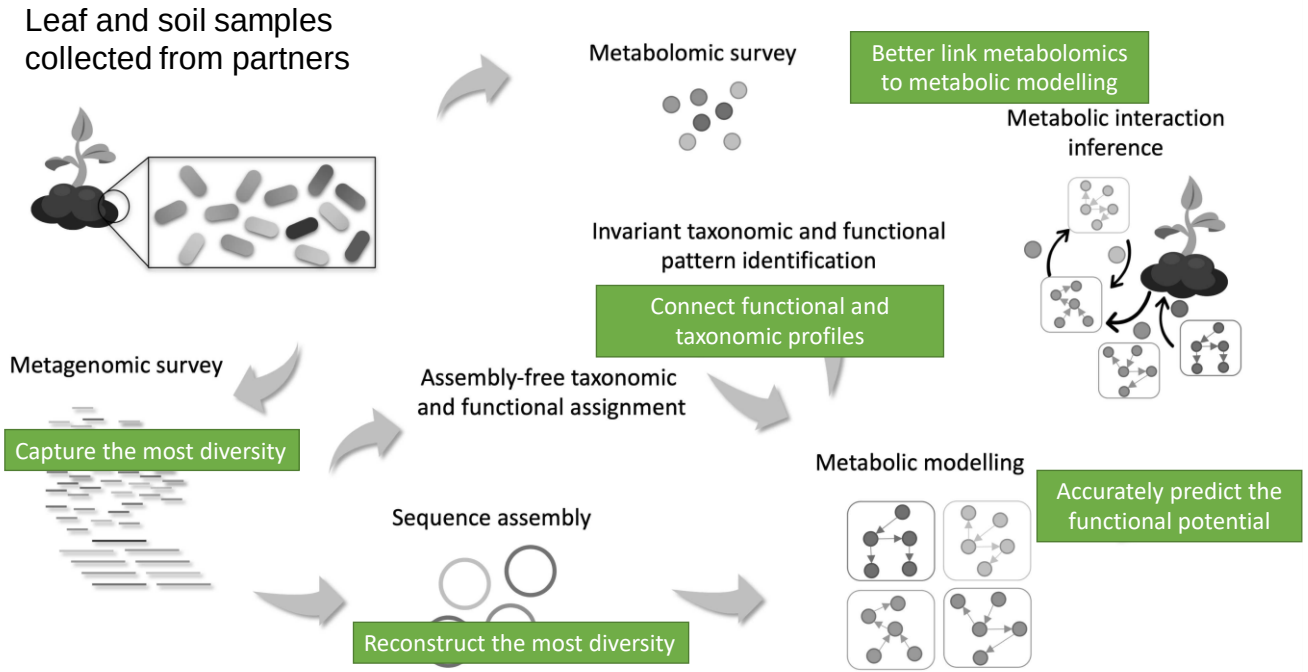



## Data outcomes

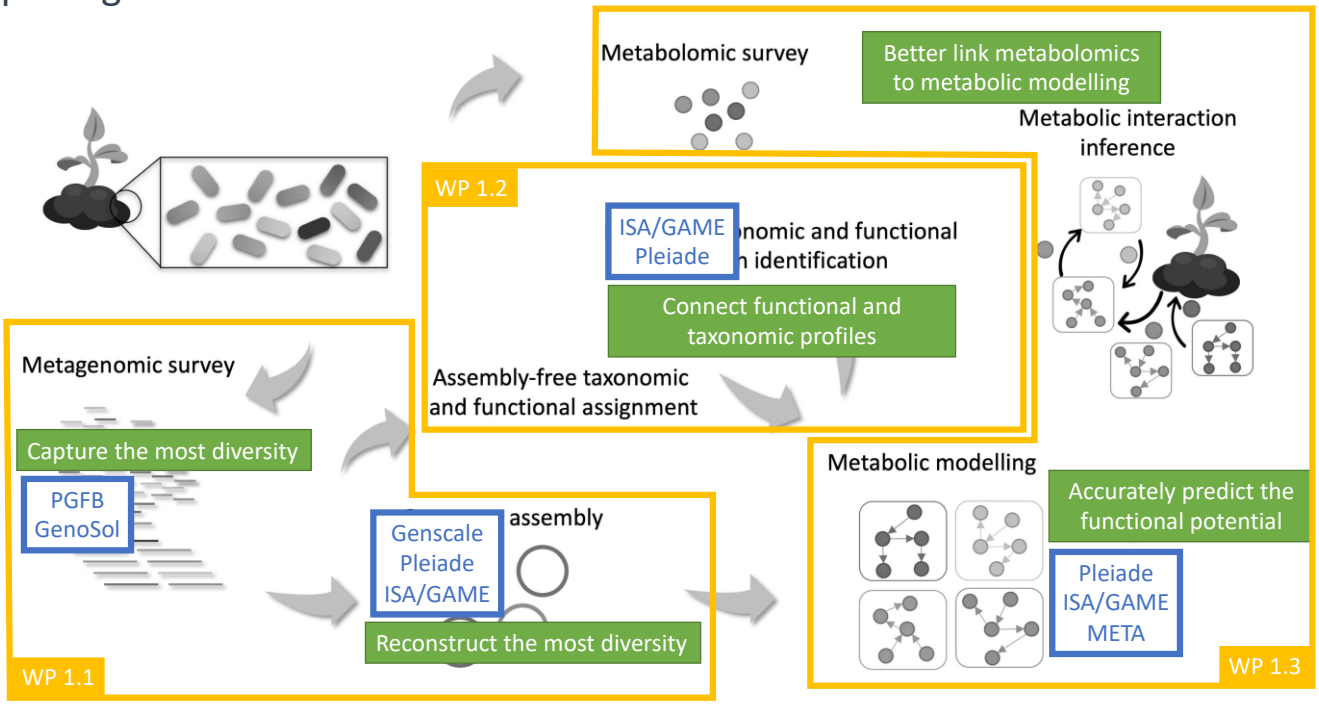
- Investment in data acquisition for training
- Novel reusable resource
- FAIR principles, RO Crate dissemination

## Experimental systems

- Open field crops: grapevine *Vitis vinifera*
- Horticultural systems : fruits, vegetables


Leaf and soil samples collected from partners
Metabolomic survey
Better link metabolomics to metabolic modelling
Metabolic interaction inference
Invariant taxonomic and functional pattern identification
Connect functional and taxonomic profiles
Metagenomic survey
Capture the most diversity
Assembly-free taxonomic and functional assignment
Metabolic modelling
Accurately predict the functional potential
Sequence assembly
Reconstruct the most diversity

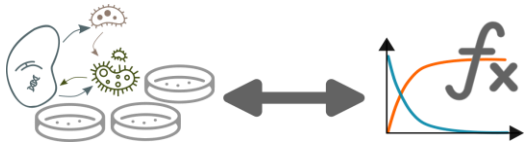
## Work packages



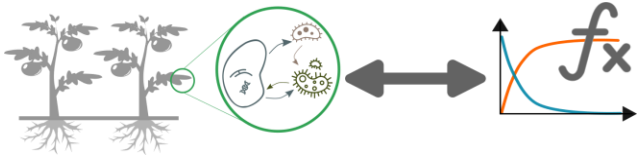
*Data*

WP 2.1 **Experimental** screening of **controled** pathobiont — SAVE

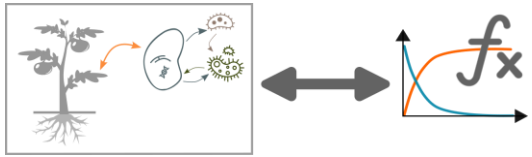

WP 2.2 **Digital twin** : modeling controled pathobiont — Pleiade ISA



WP 2.3 Function based models for **complex communities** — Pleiade ISA



WP 2.4 Multiscale model **host-pathobiome interactions** — ISA



*Pathobiome model*

vitae

Cap Phyto



Inria

# Open science, Open data

## Principles

> FAIR
> Plan S

## Data management plan

> https://dmp.opidor.fr/plans/19965
> Data lake & model mesh on *pleiadès k8s*
> Developer support *gitlab.inria.fr*
> Trustworthy data repository *data.gouv.fr*
> Metadata capture

## Computed results for method validation

> Dissemination RO-Crate in data.gouv.fr
> Coordination with V-3 BReIF

## Software dissemination

> Open-source license CECILL or equivalent
> Open-source audits using FOSSA
> Software documentation in HAL
> Protection of know-how EU 316/2014 respecting PEPR and institutional policies

# Organization

## Partnership INRAE-Inria

> Coordination Pierre Abad, David Sherman

> PhDs co-direction multi-site

## Stakeholder Council

## Institutional strategy

> Relations within PEPR Agroécologie
BReIF, AgroDiv, Holobiont

> PPR Cultiver et Protéger Autrement
Cap Zéro Phyto, VITAE, SUCSEED

> PEPR Nouvelles sélections variétales

> PEPR Système Alimentaire, le Microbiote et la Santé

## Common resources

> Biological resources
Bio-banking: INRAE Dijon, Bordeaux

> Data resources
Kubernetès infrastructure: Bordeaux & DSI-SP

> Computational resources
Kubernetès infrastructure: Bordeaux & DSI-SP

## Valorization

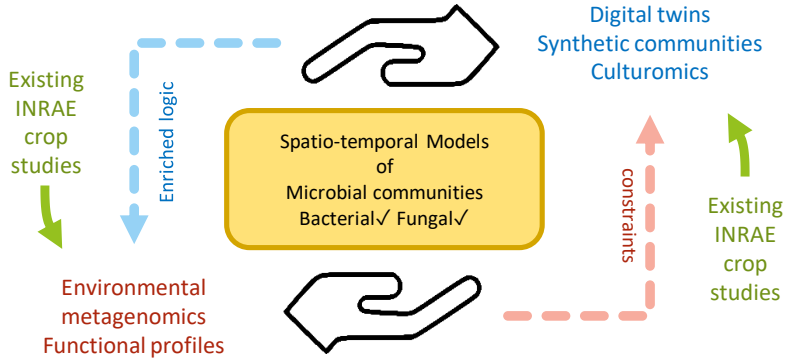
> Software valorization & APP

> Know-how including experimental protocols

> Commitment to data diffusion

## Participating teams

- Pleiade – Inria-INRAE : functional and taxonomic diversity ; dynamic numerical and discrete models
- SAVE – INRAE : vineyard health and agroecology
- Genscale – Inria : algorithms, tools, methods for (meta)genomic data
- META – INRAE : plant metabolism, (a)biotic stress
- PGTB – INRAE : genome-transcriptome platform
- GAME & M2P2 & BIG – INRAE Institut Sophia Agrobiotech : alternative solutions for sustainable management of agro-écosystèmes
- BIOCORE - Inria-INRAE : from control and dynamical systems theory to artificial ecosystems




## Contacts


- David Sherman david.sherman@inria.fr
- Pierre Abad pierre.abad@inrae.fr


https://project.inria.fr/mistic