Metaprogramme
DIGIT-BIO

INRAE

Metaprogramme DIGIT-BIO : Digital biology to explore and predict living organisms in their environment

Scientifics networks and projects funded by DIGIT-BIO

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

TABLE OF CONTENTS

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

Directors
Hervé Monod
Carole Caranta

Project manager
Marjorie Domergue

Steering Committee
Hugues Berry (INRIA)
Julien Chiquet
Pauline Ezanno
Anne Goelzer
Olivier Hamant
Fabien Joudan
Matthieu Jules
Gabriel Krouk (CNRS)
Marie-Laure Martin-Magniette
Christele Robert-Granie
Leopoldo Sanchez-Rodriguez
Thierry Simonneau
Masoomeh Taghipoor
Isabelle Maillet (National Metaprogramme Coordination unit)

Metaprogram DIGIT-BIO

Digital biology to explore and predict living organisms in their environment

The quantitative and qualitative boom of data in biology, combined with the development of new tools for processing and analysing these data, is revolutionising research in the life sciences. This development opens up new perspectives for better understanding the functioning of biological systems and predicting their behaviour.

The life sciences have undergone a fundamental paradigm shift at the end of the 20th century, with living organisms being considered as dynamic, complex and evolving systems whose overall behavior cannot be deduced from the properties of their individual components.

Systems biology now seeks to integrate different levels of information in order to understand the functioning of a biological system and predict its behaviour, via the use of methods and computer tools for modelling and simulating biological processes. The specific properties and dynamic interactions between the components of these systems are being formalised and it makes it possible to observe emerging properties and to integrate them at different spatial and temporal scales.

© sheep : Jose Llamas - Unsplash, landscape : freestocks - Unsplash, organoid : Kjparjeter - Freepik, montage : Comscience

These *in silico* approaches now benefit from new technologies, including the massive integration of data and knowledge, intensive computing, new models and meta-models, machine learning algorithms, etc. These modelling and simulation methods have opened up unprecedented possibilities for better understanding biological processes, predicting their responses to different stresses and, more broadly, for designing and better driving these systems.

The metaprogramme was launched in 2021 and its aim is to support research at the interface between computational / engineering sciences and life sciences (biology, physics, chemistry or environmental sciences), in order to:

- **Understand the functioning and predict the behaviour of biological systems**

Anticipate the impact of stresses on these systems, reason out their management and develop levers for action. In the medium term, the ambition is to develop a small number of projects for *in silico* monitoring of biological systems, based on the concept of the "digital twin".

DIGIT-BIO in figures (2022)

- A community of **over 900 members** participating in the scientific activities (projects, events) of the metaprogramme
- **14 projects** and **5 consortia** funded since 2021, involving more than 200 researchers and engineers
- **9 interdisciplinary doctoral theses cofunded**

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

The metaprogram is structured around 4 research axes

DIGIT-BIO addresses the behaviour of biological systems **from the molecular scale to that of the organism and the population** within their surrounding environments (biotic, abiotic, practices and management methods). At the moment, it does not address larger scales, for example populations or species interactions within ecosystems.

Axis 1: Understand

© kjpargeter, CS - Freepik

Understanding biological processes, their regulation and how these processes interact or cooperate.

This axis concerns all levels of organisation of living organisms: from the molecule to the organism and the population.

The aim is to **describe, understand and model biological systems**, to establish links within and between biological scales, by integrating systemic effects, such as stochastic or feedback, as determinants of the dynamics and evolution of the system.

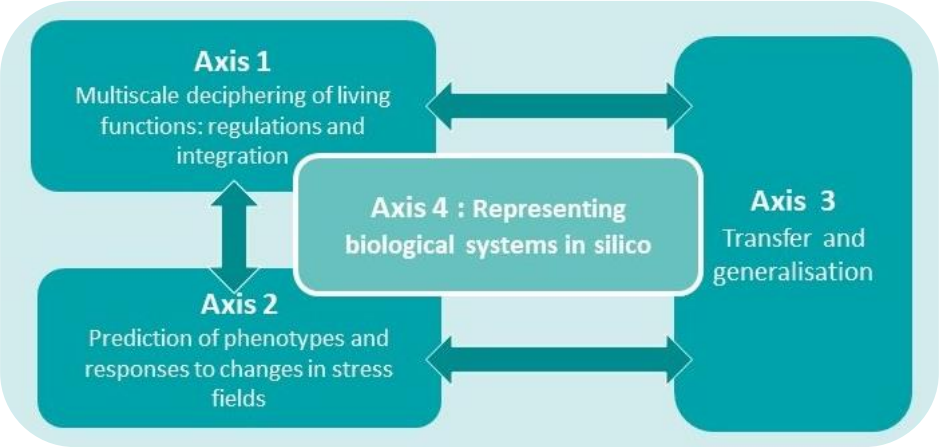
Axis 2: Predict

© INRAE

Predicting phenotypes *from the cell to the individual and the population,* ***their functionalities*** *and* ***responses to changes in stress fields*** *(biotic and abiotic environment, management methods, practices).*

One of the challenges of digital biology is to develop, compare and improve learning methods, adapting them to integrate multi-source data (omics, sensors, environment, data from participatory approaches). The modelling of biological and physiological processes to develop predictive approaches and the simulation of complex biological systems are also major challenges. In particular, the robustness of the models built in axis 1 must be tested by subjecting them to fluctuating conditions (internal or external).

Biological systems
from the cell to the individual and the population in their environment(s)



Axis 3: Transfer

© rawpixel.com - Freepik

Generalise and transfer the results and knowledge acquired towards scales, organisms, species or systems*, that have been less studied or only partially observed.*

This axis also aims to **develop more robust comparative approaches** that better evaluate the generic scope of the data.

Axis 4: Toward the digital twin?

Digital biology offers the possibility of experimenting and monitoring biological systems in silico, based on their computer representations and their regular updates from data collected in real time.

The concept of "**digital twin**", a true *in silico* copy of its real counterpart, is a promising tool for monitoring and steering systems, which the metaprogramme aims to explore.

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

ALGO-ROOT

EXPLORATORY PROJECT

2021-2023

Modeling decision algorithms for root development in heterogeneous environments

Coordination
Sandrine Ruffel
UMR BPMP
sandrine.ruffel@inrae.fr

Key words
Phenotype modelling
Decision algorithms
Root development
Nutritional heterogeneity of environments

INRAE units involved
IPSiM

Partnerships
Simons Center for Quantitative Biology, Laboratoire du Cold Spring Harbor NY - United States

Institut Alan Turing - United Kingdom

Context and challenges

To survive, plants must take up water and many nutrients from the soil. These resources are unevenly distributed and plants must explore the soil to find them. This exploration requires the extension of roots, which is a development that comes at a cost for the plant. To minimise resource expenditure while maximising nutrient acquisition, decisions about where to explore and when to forage probably need to be optimised. How do plants manage this trade-off?

One way to study this question is to present plants with choices and examine their behaviour. For this purpose, the so-called two-armed (or one-armed) bandit problem provides an interesting mathematical framework because it allows us to determine the decision algorithms underlying decision making when faced with two competing choices with different (but unknown) rewards. For plants, the dilemma will be between exploiting low or medium quality but immediately available, resources or exploring new parts of the soil where better quality resources may (or may not) reside. The general problem of optimisation between exploitation and exploration has already been studied in various fields, including psychology and economics, where quantitative frameworks have been well described to evaluate the advantages and disadvantages of different forage methods. However, this framework has not yet been used to understand and predict plant behaviour.

Goals

The objective of the ALGOROOT project is to discover the algorithms (reflecting a succession of decisions) that plant roots use to search for nutrients in heterogeneous environments.

The project's approach is four-fold:

1. Identify the algorithmic basis of the branching used by plant root systems to explore the soil;
2. Develop mathematical models to predict how roots "decide" between exploiting an available resource or exploring new territory in the hope of finding a better resource;
3. Evaluate how search strategies and decision-making algorithms are genetically encoded;
4. Compare and contrast root search algorithms with those used in other fields (e.g. chemotaxis, infotaxis, random walks) and test whether lessons from plant biology can be applied to computer science.

The ALGOROOT project is an interdisciplinary project that integrates theory and experimentation to solve a fundamental biological problem.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Plant biology and breeding	IPSiM	Expertise provided: Plant physiology and development
External partners		**Expertises**
Simons Center for Quantitative Biology / Cold Spring Harbor Laboratory, NY (United States)		Theoretical computer science, machine learning, systems biology
Alan Turing Institute (United Kingdom)		Biology and modelling

Reference

- **Ruffel, S., Krouk, G., Ristova, D., Shasha, D., Birnbaum, K.D., and Coruzzi, G.M. (2011).** Nitrogen economics of root foraging: transitive closure of the nitrate-cytokinin relay and distinct systemic signaling for N supply vs. demand. Proc. Natl. Acad. Sci. U.S.A. 108, 18524-18529.

DINAMIC

EXPLORATORY PROJECT

2021-2023

Analysing biological networks of mixed-type data with copula models

Coordination
Andrea Rau
UMR GABI
andrea.rau@inrae.fr

Key words
Mixed type data
Network plasticity
Copulas
Differential network analysis

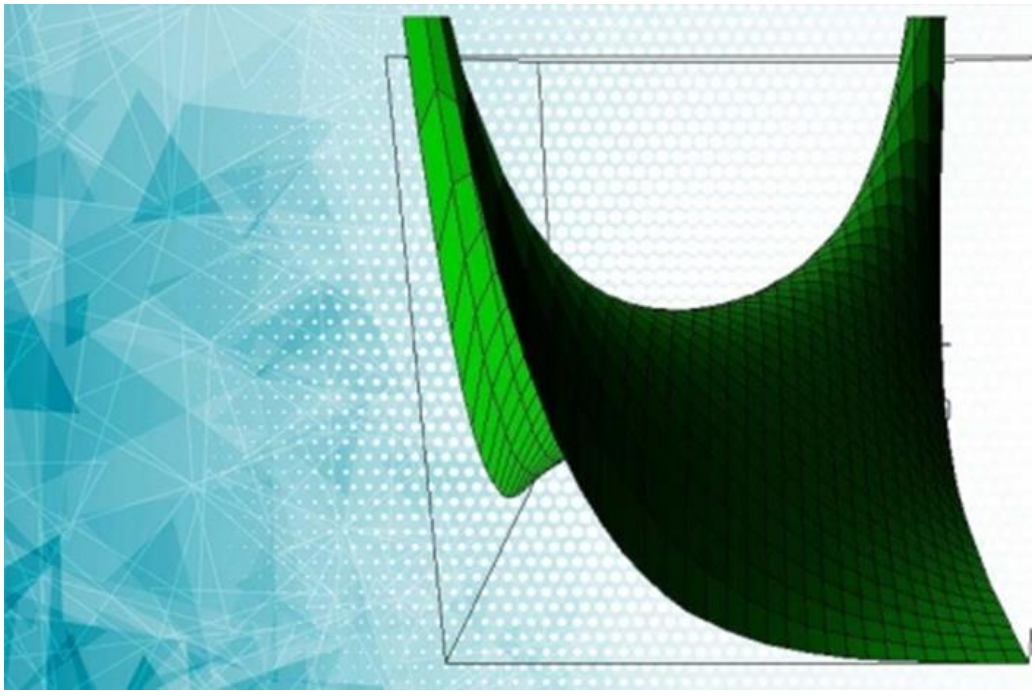
INRAE units involved
GABI
Transfrontalière BioEcoAgro
GQE-Le Moulon
MaIAGE
BREED
NutriNeurO

Partnerships
Athens University of Economics and Business - Greece

Context and challenges

Integrative biology is based on the study of complex biological networks. Understanding the plasticity of biological interaction networks due to phenotypic, environmental or interventional variability is an important challenge in fields as diverse as genomics or human nutrition. Such studies often include comparisons between contrasting groups, including variables of various natures (continuous, counts, binary, etc.). These so-called "mixed-type" data can be difficult to analyse in a unified way. While multivariate probabilistic models provide a robust framework for inferring interrelationships among continuous variables, an analogous model for mixed-type data has yet to be defined.

A particularly promising but as-yet unexplored approach for this purpose is the use of parametric copula models, which can be used to couple variables of disparate natures. The development of such a model in a computationally efficient graphical form thus represents an open methodological challenge for the inference of generic networks from mixed-type data.



© Charthur - Wikipedia & Starline - freepik

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

The DINAMIC project aims to develop and implement an innovative and widely applicable multivariate framework based on copulas and random pairwise likelihood (Mazo et al., 2021) for the differential analysis of mixed-type networks.

These methodological developments will be based on a succession of three applications covering several research themes at INRAE:

- cognitive health networks in seniors following the introduction of nutritional supplements;
- phenotypic networks in response to thermal stress in maize lines structured according to their genetic proximity;
- multi-omic networks in sperm from groups of bulls with contrasting fertility.

Each application will motivate a distinct facet of our approach, highlighting the added value of our interdisciplinary collaboration. To combine theoretically sound and computationally efficient statistical developments with relevant modelling assumptions aligned with the underlying biology, the DINAMIC project relies on a continuous cycle of interactions between methodologists and domain-specific experts.

Our multivariate mixed-type network model will represent a new approach to digital biology, with the potential to generate new insights into network plasticity in a wide variety of scientific disciplines.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Plant biology and breeding	GABI	Biostatistics
	Transfrontalière BIoEcoAgro	Quantitative genetics, plant genomics
	GQE-Le Moulon	Omic analysis
Mathematics, computer and data sciences, digital technologies	MaIAGE	Statistics, mathematics
Animal physiology and livestock systems	BREED	Animal genomics
Human nutrition and food safety	NutriNeurO	Human nutrition
External partners		**Expertises**
Athens University of Economics and Business (Greece)		Statistics and methodology

Reference

- **Mazo, G., Karlis, D., and Rau, A. (2021)** A randomized pairwise likelihood method for complex statistical inferences. Under review. ⟨hal-03126621⟩

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

GenIALearn

EXPLORATORY PROJECT

2021-2023

Application of machine learning and deep learning to improve animal genomic selection

Coordination
Eric Barrey
UMR GABI
eric.barrey@inrae.fr
Didier Boichard
UMR GABI
didier.boichard@inrae.fr

Key words
Genomics
Gene interactions
Statistical learning
Machine learning
Deep learning

INRAE units involved
GABI
MIA Paris Saclay

Partnerships
UEVE Université Paris-Saclay

Context and challenges

The development of genomic selection - and other "omics" analyses such as metagenomics, transcriptomics, metabolomics and proteomics - now makes it possible to characterise animals using thousands of measurements. This massive data is integrated into models to predict production traits with the highest possible degree of accuracy.

The most commonly used models in genomic prediction (additive genetic model such as GBLUP) are very efficient in predicting the genetic value of animals on a few genetically correlated traits. On the other hand, this type of model does not allow the integration of a very large number of heterogeneous measurements, nor does it predict many output traits without knowing their genetic correlations. Moreover, this model is limited in its ability to take into account the many non-linear interactions that occur between regions of the genome or environmental factors.

In order to overcome these obstacles, we propose statistical learning (machine learning) and deep learning methods, derived from AI, to process both additive genetic information and non-linear genetic information present in massive genotyping data.

© kjpargeter, CS - freepik

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

The GenIALearn project proposes to evaluate the performance of statistical and deep learning methods for the joint prediction of multiple complex traits, by integrating massive genotyping data. Two main families of methods will be compared altogether and versus the reference method GBLUP:

- on the one hand, ensemble learning methods (random forests, gradient boosting), coupled with a learning step to represent the input data, in order to propose reference prediction levels;
- on the other hand, deep learning methods of different architectures (neural networks), coupled with learning step on massive data base, which should produce predictive models adapted for animal genomic selection.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Animal genetics	GABI	Fine phenotyping of complex traits, multi-omics (genotyping, transcriptomics, metagenomics, metabolomics), genetic values evaluation and complex multi-trait predictions.
Mathematics, computer and data sciences, digital technologies	MIA Paris Saclay	Modelling, statistical learning, machine learning, large and heterogeneous data, application to life sciences.
External partners		**Expertises**
UEVE Université Paris-Saclay	UBISC	Neural network construction methods and deep learning, Applications for transcriptomic and image analysis

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

IMMO

EXPLORATORY PROJECT

2021-2023

Visualising fish oocytes using AI and 3D imaging

Coordination
Violette Thermes
LPGP
violette.thermes@inrae.fr
Romain Yvinec
UMR PRC
romain.yvinec@inrae.fr

Key words
Fertility
Oogenesis
3D imaging
Deep learning
Structured population dynamics

INRAE units involved
LPGP
PRC

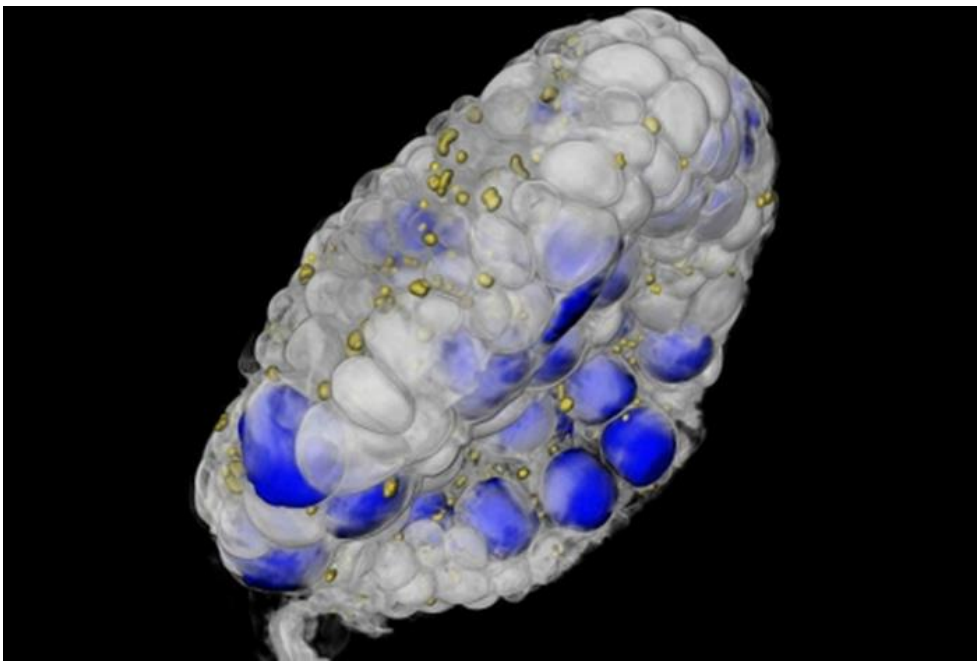
Partnerships
Inria

Context and challenges

In the natural environment as well as in fish farming, the process of formation and maturation of female gametes (oogenesis) is essential for reproductive success.

In multiple-spawning fishes, oogenesis involves anatomical structures in permanent renewal, the ovarian follicles, which accompany the development of the gametes until spawning. Despite the identification of numerous regulatory mechanisms of oogenesis in model fish (e.g. medaka, zebrafish), we still have an incomplete and mainly qualitative view of this dynamic process. In particular, major questions remain unanswered:

- Is there a detectable ageing of ovarian function?
- What are the key controls on ovarian follicles at different stages of maturity and to what extent is the follicle population self-controlled?



© Manon Thomas, Manon Lesage and Violette Thermes

Goals

The IMMO project proposes to exploit new 3D imaging and Artificial Intelligence (AI) methods to visualise and enumerate the completeness of oocytes in fish ovaries at different ages, in order to comprehensively and quantitatively describe the entire population of ovarian follicles and oocytes.

These data will be used to validate a mathematical model describing follicular dynamics and their controls over the life span of the fish, which will reveal information not available from the data alone. The model simulations will reproduce the different types of disturbances affecting the proper functioning of oogenesis.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Animal physiology and livestock systems	LPGP	Reproductive biology, imaging and image analysis
	PRC	Mathematical analysis, intracellular signalling, pharmacology
External partners		**Expertises**
Inria	Équipe projet MUSCA	Mathematical modelling/ Reproductive physiology, Mathematical analysis

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

MIRRORS

EXPLORATORY PROJECT

2021-2023

Predicting the response of plants exposed to chronic thermal stress

Coordination
Sophie Brunel Muguet
UMR EVA
sophie.brunel-muguet@inrae.fr

Key words
Repeated thermal stress
Plant performance
Acclimatisation
Modelling
Data mining

INRAE units involved
EVA
AGAP Institut
ISPA

Partnerships
Lorraine Research Laboratory in Computer Science and its Applications

Context and challenges

Climate change is characterised not only by variable and extreme intensities of the main climatic factors but also by an increased frequency of extreme events, such as heat waves, which are highly detrimental to field crop yields and harvest quality.

In this context, improving predictions of plant performances under repeated heat stress scenarii is a major challenge. The MIRRORS project is based on the hypothesis that the effect of a succession of stressful events is not equivalent to the sum of the individual effects of each event. Indeed, when plants have been exposed by an initial stress, their responses to subsequent stresses can be determined by this prior event as a consequence of a "memory effect" (which can be either penalising or beneficial).

In order to improve predictions of plant performance in repeated stress situations that are expected to occur more frequenlty, the MIRRORS project proposes methods and tools for generic predictions of the response of plants subjected to repeated thermal stress in particular.

© gpointstudio - freepik

Goals

We propose the following approaches, based on existing data sets for rapeseed and sorghum[1]:

1. Analyse the non-additive nature of the effects of heat stress events using complete datasets (climatic variables and plant performance criteria).
2. Identify agro-climatic indicators or specific thermal sequences related to the memory of heat stress. We will identify particular thermal scenarii, with recurrent patterns, and associate them with the observed plant performances (grain yield and seedquality criteria).
3. Then, two complementary approaches to modelling the effects of repeated thermal stresses will be developed, on both rapeseed and sorghum:
    - "concept-driven" based on the implementation of predictive ecophysiological models parameterised on these two species, in order to take these memory effects into account, and
    - "data-driven" guided by data mining methods with no mechanistic *a priori*.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Agronomy and environmental sciences for agroecosystems	EVA	Ecophysiology, agronomy, modelling
	AGAP Institut	Ecophysiology, statistical analysis and data mining
	ISPA	Ecophysiology, modelling, biogeochemistry
External partners		**Expertises**
Lorraine Research Laboratory in Computer Science and its Applications		Knowledge discovery, modelling

[1] in the field and in the greenhouse

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

PEERSIM

EXPLORATORY PROJECT

2021-2023

Predicting plant response to combined stresses

(CO_2 and Heat)

Coordination
Guillem Rigaill
UMR IPS2
guillem.rigaill@inrae.fr
Etienne Delannoy
UMR IPS2
etienne.delannoy @inrae.fr

Key words
Multi-Stress
Experiment design
Multi-omics
Integration
Plant Biology

INRAE units involved
IPS2
MIA Paris Saclay
MIAT

Context and challenges

Plants are constantly threatened by biotic and abiotic stresses, especially in the current context of climate change. The complexity of the stress response involves different levels of biological organisation, from genomes to metabolites. The study of multiple stresses shows that the impact of combined stresses is different from the sum of the impacts of individual stresses. How then can the impact of combined stresses be predicted by knowing only the impact of single stresses?

This conclusion is based on studies comparing lists of differential genes/metabolites subjected to individual stresses with the same lists subjected to combined stresses. However, these analyses are based on too few biological replicates (typically 3 in RNA-seq), which are insufficient to produce a robust and meaningful analysis, and only identify about 20% of the genes that are differentially expressed under the two stress conditions. This may partly explain the discrepancies observed between single and multiple stress conditions.

© Etienne Delannoy

Goals

The Peersim project proposes to effectively re-evaluate the prediction of combined stresses from individual stresses, by conducting an experiment combining 2 stresses: CO_2 and heat, with numerous replicates (~20).

Beyond the biological relevance of this dataset in the context of climate change, the project will allow progress on three essential points for the study of plant response to combined stresses:

- Effectively quantify the extent to which the impact of combined stresses is different from the sum of the impacts of individual stresses, and assess the possibility of predicting the actors of the response to combined stresses and their interactions.
- Develop and propose meaningful experimental designs.
- Develop and evaluate recent analysis and integration methodologies.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Plant biology and breeding	IPS2	Bioinformatics, Biostatistics, Segmentation, Chloroplast biology, Transcriptomics, Metabolism and plant adaptation to climate change
Mathematics, computer and data sciences, digital technologies	MIA Paris Saclay	Statistics and gene network inference
	MIAT	Data integration, network analysis

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

Physioscope

EXPLORATORY PROJECT

2021-2023

A new tool for exploring the multi-regulator and multi-scale network controlling plant architecture

Coordination
Jessica BERTHELOOT
UMR IRHS
jessica.bertheloot@inrae.fr
Frédéric BOUDON
UMR AGAP
fredric.boudon@cirad.fr

Key words
Multi-regulator network
Emergent behaviour
Virtual plant
Online interactive simulation interface

INRAE units involved
IRHS
AGAP Institut

Partnerships
Inria-ENS Lyon
CNRS

Context and challenges

To maintain the agronomic performance of plants in increasingly stressful environments, it is necessary to have a systemic vision of their adaptation mechanisms, particularly their architectural development, i.e. the initiation and development of new organs.

The mechanisms involved in this development are complex. They involve multiple regulators of different types (hormones, nutrients), controlled by different processes and at different scales (local, remote). Numerical models have proven to be effective tools for understanding some of these complex regulations, as they simulate non-intuitive behaviour induced by this complexity. They make it possible to test regulation hypotheses in experiments comparing the behaviour of a real and a virtual plant. Today, there are digital tools for simulating virtual plants, such as the L-Py platform.

However, their effective use for the virtual exploration of regulatory networks at the plant scale requires, on the one hand, facilitating their user-friendliness and interactivity with biologists and, on the other hand, improving the dialogue between biologists and modellers, who work at different scales (mechanisms vs. plant behaviour).

© kjpargeter - FREEPIK (modified)

Goals

The Physioccope project aims to provide an efficient tool for smooth and collaborative interaction between biologists working at different scales through a virtual plant model. The tool will be developed specifically to understand how light regulates bud outgrowth, but it is designed to be used more broadly. The project's objectives are threefold:

1. The integration of the mechanistic network controlling bud outgrowth along an axis into a virtual plant, coded in L-Py;
2. The development of an intuitive tool for interaction and visualisation of this network via the virtual plant, based on the coupling between L-Py and the MorphoNet browser (dedicated to the interaction with morphodynamic structures);
3. The identification of new hypotheses on the bud outgrowth regulation network by comparing the behaviour of the plant between virtual and real experiments.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Agronomy and environmental sciences for agroecosystems	IRHS	Expertise provided: Modelling, at the interface between physiology and ecophysiology, of mechanisms regulating plant architecture
Plant biology and breeding	AGAP Institut	Expertise provided: Simulation of architecture and functioning. Distributed computing
External partners		**Expertises**
Inria-ENS Lyon	RDP	Modelling of plants and molecular networks
CNRS	LIRMM	Data Science for Biology, Interaction and Visualisation of Models

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

PlantRBA

EXPLORATORY PROJECT

2021-2023

Coordination
Anne Goelzer
UMR MAIAGE
anne.goelzer@inrae.fr
Olivier Loudet
IJPB
olivier.loudet@inrae.fr

Key words
High-throughput phenotyping
Combined abiotic stress
Genotype-phenotype relationship
Resource allocation
Constraint-based models
Plant systems biology

INRAE units involved
MaIAGE

Partnerships
IJPB

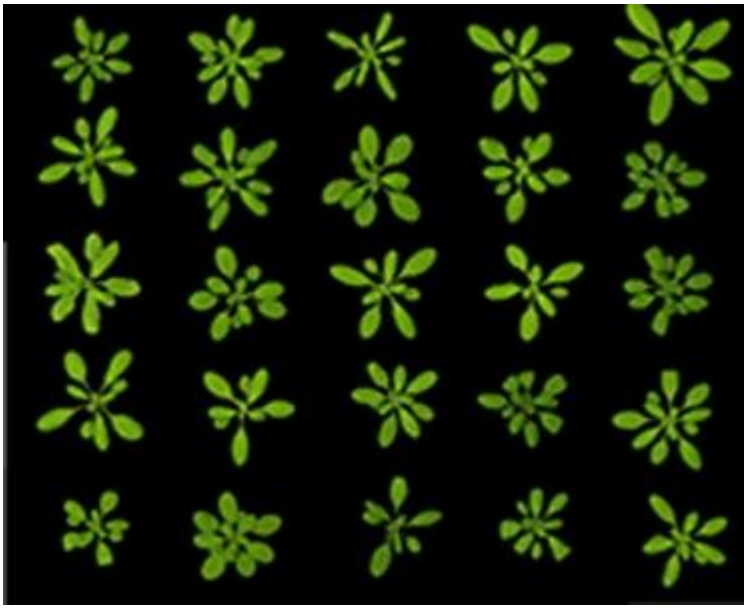
Predicting plant phenotypes under combined stress

Context and challenges

Climate change, the scarcity of certain natural resources and the need to reduce agricultural inputs have increased the number and diversity of situations that agronomists need to understand.

They need plant models with extensive predictive capability and capable of taking into account complex environmental conditions, where different constraints (stresses) come into play at the same time.

Well-established plant models at the individual level, such as the ecophysiological models they usually use, generally fail to cope with such realistic conditions. Indeed, the cellular scales, i.e. the scales where adaptation occurs, are poorly described in these models. The challenge of this project is therefore to refine the description of cellular and sub-cellular scales in plant modelling (and more generally in the modelling of multicellular organisms) and thus better link the genotype and phenotype of an organism.



© INRAE

Goals

This project aims to develop, calibrate and experimentally validate a mathematical model predicting the behaviour of the *Arabidopsis thaliana* plant under abiotic constraints (limited water and/or nitrogen availability). This model is based on the parsimonious distribution of resources between the different biological functions of the plant and thus reconciles the smallest scales (genes) with the phenotype.

The project combines cutting-edge mathematical models in plant modelling with state-of-the-art experimental techniques designed to grow plants under the most robust environmental conditions, on the Phenoscope platform, to generate very high-quality biological data for model calibration and validation.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Mathematics, computer and data sciences, digital technologies	MaIAGE	Modelling, systems biology, omics data analysis and integration, bioinformatics
External partners		**Expertises**
IJPB Institut Jean-Pierre Bourgin		Phenotyping, physiology, bioinformatics, genetics

PrionDif

EXPLORATORY PROJECT

2021-2023

Prion diseases: modelling the of dissemination and neuroinvasion

Coordination

Human Rezaei
UMR VIM
human.rezaie@inrae.fr

Key words

Prion
Neurodegeneration
Autocatalytic process
Diffusion reaction
Prion strain

INRAE units involved

VIM
IHAP

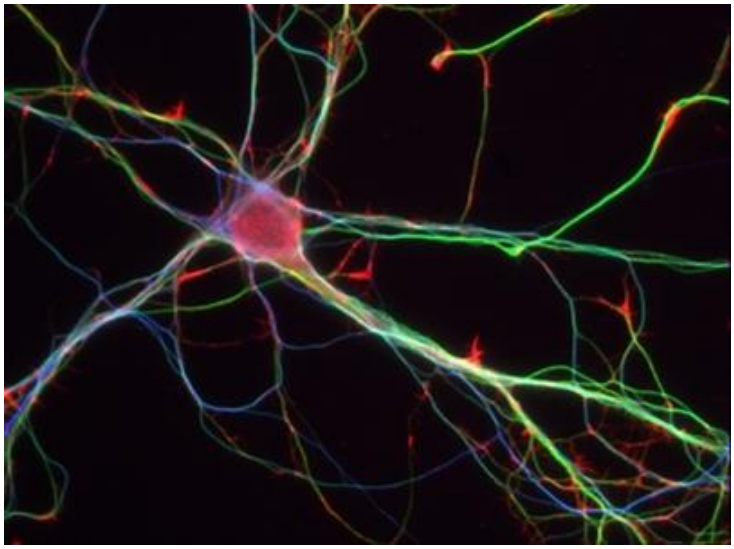
Partnerships

Inria

Context and challenges

The prion paradigm unifies a number of age-related, devastating neurodegenerative pathologies caused by autocatalytic protein misfolding and aggregation. In the prion paradigm framework, host-encoded monomeric proteins are converted into misfolded aggregated assemblies, which serve as a template for further autocatalytic recruitment and conversion in the brain. Since the late 2000s, the prion paradigm has been extended to other neurodegenerative diseases due to protein misfolding such as Alzheimer's and Parkinson's disease.

In mammalian prion diseases, also known as Transmissible Spongiform Encephalopathies (TSE), prion assemblies (PrPSc), formed from the cellular prion protein (PrPC), contain all the structural information necessary to their replication and their specific stereotyped disease phenotype in the infected host. In TSE, multiple PrPSc conformational variants exist. They define the prion strains and dictate specific physiopathological patterns such as region-specific PrPSc deposits in the same host species. Although self-replicative processes provide a mechanistic framework for the prion paradigm, to date there is no mechanistic link between prion replication, the neuroinvasion process and the strain-specific neuropathological pattern.



© Inserm-L.Peris

Goals

The PrionDif project seeks to develop a multi-scale mechanistic model accounting for the spatiotemporal dynamic of prion spreading within the brain by integrating experimental observations with an effective model of prion replication which takes into account the dynamicity of PrPSc assemblies. By integrating the spatio-temporal mapping of the spread of prion replicative centres with the prion replication/dissemination model, we aim to build a synthetic

multi-scale model of prion structural diversification and lesional propagation. This open-access model will allow us to investigate which parameters of the prion replication process specific to each strain dictate the progression of the disease and the apparition of strain specific PrPSc deposition patterns.

Ultimately, this synthetic approach will allow the identification of key processes to enable therapeutic advances and promote early diagnosis.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Animal health	VIM	Macro-Assembly Pathology and Prion Diseases (MAP²) team expertise: molecular biophysics & biochemistry, non-equilibrium kinetics and modelling, stochastic process, Gillespie-type approach, retro-synthetic approach, characterisation of prion assemblies, patterning and prion strains, spatial-temporal evolution of different prion assembly subspecies
	IHAP	Pathogenesis of transmissible spongiform encephalopathies team: physiopathology of prions, tractography, systemic and tissue dissemination of prions, typing of prion strains
External partners		**Expertises**
Inria	Équipe projet Dracula	Modelling of prion diseases: mathematical modelling of reactions under diffusion controls, data integration, synthetic biology, control theory, optimisation, predictive approach

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

BovMovie2Pred

SCIENTIFIC NETWORK

2021-2022

Early categorisation of bovine embryos to boost IVF success

Coordination
Alline Reis
BREED
alline.reis@vet-alfort.fr

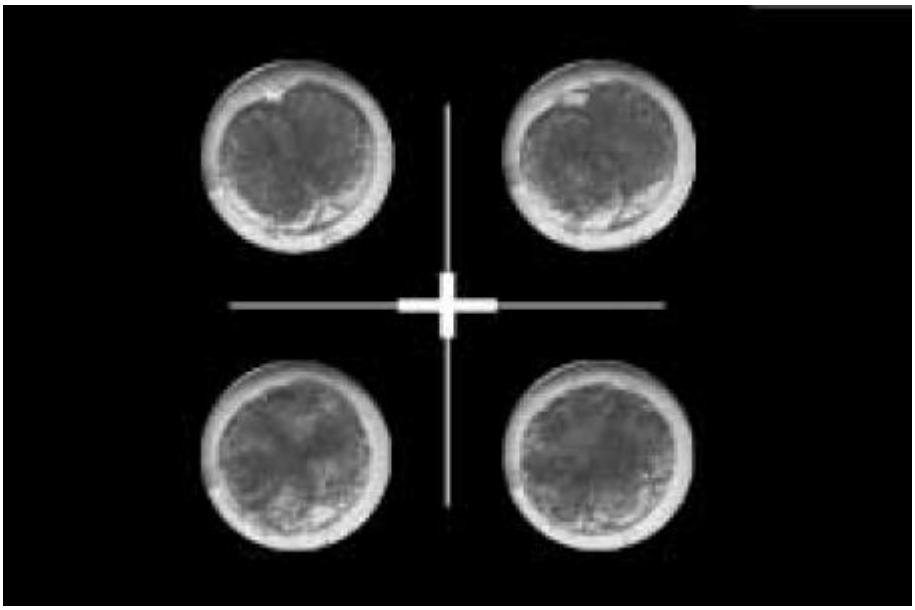
Key words
Deep learning
Statistical learning
Video
Developmental biology
In vitro fertilisation

INRAE units involved
MaIAGE
MIA Paris Saclay
BREED

Partnerships
Inria

Context and challenges

A major issue in in vitro fertilisation (IVF) is the selection of the "best" embryo, i.e. the one most likely to implant in the uterus. Currently, in cattle, the success rate of IVF and embryo transfer does not exceed 30% of viable births. The selection of embryos (from oocytes collected in vivo or post mortem and then fertilised) is based on a classification at D7 after fertilisation. One of the keys to increasing IVF performance is to optimise this selection as early as possible.



© INRAE / Alain Trubuil and Alline Reis

Goals

The objective of the BovMovie2Pred consortium is to propose solutions to assist in the selection of bovine embryos in order to significantly increase the percentage of viable births from in vitro produced embryos.

The aim is to optimise the selection of embryos as early as possible by exploiting their morphokinetic history, from fertilisation to the day of transfer. This history is traced from annotated videos. However, expert annotations of videos have the double disadvantage of being laborious to carry out and having a subjective element.

In order to overcome these constraints, the BovMovie2Pred consortium proposes to organise one or more data challenges within the framework of the RAMP (Rapid Analytics and Model Prototyping) platform of the DATA-IA Convergences Institute. These challenges will bring together the skills of experts on AI issues as well as those of students or PhD students in this field. The expertise of the consortium, coupled with existing annotation work, will

make it possible at the end of the project to provide researchers in developmental biology with a classification methodology requiring as little video annotation as possible.

Research units involved and partners

INRAE scientific division	INRAE research units	Expertises
Mathematics, computer and data sciences, digital technologies	MaIAGE	Video analysis
	MIA Paris Saclay	Statistical learning
Animal physiology and livestock systems	BREED	Developmental biology
External partners		**Expertises**
Inria	Équipe projet SERPICO	Video analysis
	DATA-IA	Data challenge platform

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

IFM2A2

SCIENTIFIC NETWORK

2021-2022

Building a new approach to integrate the functioning of apical meristems in the dynamic modelling of plant aerial architecture

Coordination

Jean-Louis Durand
URP3F
jean-louis.durand@inrae.fr
Jessica Bertheloot
IRHS
jessica.bertheloot@inrae.fr
Christophe Godin
Inria
christophe.godin@inria.fr

Key words

Aerial morphogenesis
Apical meristem
Self-centred model
Plant architecture
Physical constraints
Genetic variability

INRAE units involved

URP3F
IRHS
LEPSE
PIAF
IJPB
AGAP Institut
MIAT

Partnerships

Inria

Context and challenges

The use of a greater diversity of plant species to optimise natural resources has highlighted the need for a better understanding of the dynamics of plant stands. Competition for light between individuals is a key phenomenon in these dynamics. This is why the simulation of aerial architecture is essential. This is essentially determined by the functioning of the apical meristems of the different axes of the plant, which includes the production of apexes (branching), the production of leaves and the elongation of the axes, up to the formation of fruits and their positioning in the stand.

Recent work by INRAE and INRIA on this topic has been the subject of high-impact publications. However, a better understanding of the determinants of aerial morphogenesis in response to the environment requires a new and multidisciplinary approach, in order to take into account different levels of scale, from the gene to the stand.

© WangXINa - freepik

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

The IFM2A2 consortium proposes to bring together in a sustainable manner the different scientific communities that are currently working separately on simulating the functioning of apical meristems at different scales, operating in different INRAE departments (BAP, MAthNum and AgroEoSystem) in close interaction with INRIA.

In order to achieve such syntheses, which allow science to move from the subcellular or cellular scales to those of the organ and the plant in the stand, it is essential to build a space for sharing and exchanging information between scientific communities from different disciplines, thanks to effective communication tools.

With a view to bringing these different communities together, the consortium calls for:

- The organisation of an international seminar (28-30 November 2022 in Poitiers);
- The joint writing of a multi-scale synthesis journal;
- The construction of a sustainable animation network, in order to accelerate the production of models integrating the different scales.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Agronomy and environmental sciences for agroecosystems	URP3F	Plant ecophysiology: conducts a programme of architectural modelling of plants in ecophysiology
	IRHS	Plant biology: studies the genetic and environmental control of meristem branching activity
	LEPSE	Dynamic simulation of aerial morphogenesis. Individual-centred modelling of plant-environmental stress interaction
	PIAF	Study of the biomechanical determinants of plant growth
Plant biology and breeding	IJPB	Physiology of the primary wall. Develops plant biology programmes on primary wall synthesis and branching
	AGAP Institut	Simulation mathématique de la morphogenèse des apex
Mathematics, computer and data sciences, digital technologies	MIAT	Mathematical simulation of apex morphogenesis.
External partners		**Expertises**
Inria	Équipe projet MOSAIC	Mathematics and plant biology: modelling of plant morphogenesis and apical meristem function

References

- **Azpeitia, E., Tichtinsky, G., Le Masson, M., Serrano-Mislata, A., Lucas, J., Gregis, V., Gimenez, C., Prunet, G., Farcot, E., Kater, M.,M., Bradley, D., Madueño, F., Godin, C. & Parcy, F. (2021).** Cauliflower fractal forms arise from perturbations of floral gene networks. Science, 373(6551), 192-197.
- **X., Barillot, R., Chambon, C., Fournier, C., Combes, D., Pradal, C., & Andrieu, B. (2019).** Investigation of complex canopies with a functional–structural plant model as exemplified by leaf inclination effect on the functioning of pure and mixed stands of wheat during grain filling. Annals of Botany, 123(4), 727-742.
- **Haas, K. T., Wightman, R., Meyerowitz, E. M., & Peaucelle, A. (2020).** Pectin homogalacturonan nanofilament expansion drives morphogenesis in plant epidermal cells. Science, 367(6481), 1003-1007.
- **Rameau, C., Bertheloot, J., Leduc, N., Andrieu, B., Foucher, F., and Sakr, S. (2015).** Multiple pathways regulate shoot branching. Frontiers in Plant Science 5, 741.
- **Vernoux, T., Besnard, F., & Godin, C. (2021).** What shoots can teach about theories of plant form. Nature Plants, 7(6), 716-724.

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

PhenoDyn

SCIENTIFIC NETWORK

2021-2022

Bringing together INRAE expertise for the prediction of dynamic phenotypes

Coordination
Nicolas Verzelen,
UMR MISTEA
nicolas.verzelen@inrae.fr

Key words
High throughput phenotyping
Growth curves
Statistics for functional data: growth model
Genetic-environmental interaction model

INRAE units involved
MISTEA
MIAT
MaIAGE
GQE-Le Moulon
GenPhyse

Context and challenges

In response to the multiple challenges of climate change and multi-performance agriculture, the sciences of breeding and plants are faced with the challenge of selecting breeds or varieties on the basis of increasingly complex phenotypes. Examples include plant growth curves in the face of water stress, microbial community growth in the face of nutrient restriction, and weight gain dynamics in animal husbandry. Thanks to the popularisation of sensor technologies and the emergence of digital agriculture, INRAE researchers now have access to medium and even high-speed growth data.

This is the case, for example, thanks to the PHENOME-EMPHASIS plant phnomics infrastructure in plant science, to automated distribution methods for concentrated feeds (DAC) in animal husbandry, or to real-time monitoring methods for bacterial communities. Nevertheless, the analysis and prediction of these phenotypes and, ultimately, their use in selection schemes, raise many challenges, linked to the noisy nature of the data and their highly complex structure (response in the form of curves linked to environmental covariates).

Currently, different INRAE teams contribute to these challenges, but in a relatively individual way. This segregation is as much related to the diversity of the species studied (microbial, plant or animal), to the diversity of the biological aspects (complex phenotypes versus genetics) as to the diversity of the statistical approaches used.

Goals

The objective of the PhenoDyn consortium is to bring together statisticians and geneticists from the institute who are interested in the prediction of complex dynamic phenotypes, in order to compare the approaches used and decompartmentalise the fields of application. As such, the consortium includes members of four INRAE departments (MathNum, GA, AgroEcosystem and Plant Biology and Breeding). Initially, the consortium will draw up an inventory of the various methodological contributions to the study of dynamic phenotypes, which currently include a wide range of approaches: semi- or non-parametric methods of functional statistics, Markovian dynamics models, non-linear models based on more refined dynamic modelling, etc.

This first step will make it possible to enhance INRAE's contributions in the field, but also to guide future users - including private partners - on the methodological choices to be adopted according to the identified objective: measurement of a genetic index (e.g. heritability of the dynamic phenotype), prediction of the phenotype at future times, identification of characteristic phenotypic profiles, etc. Then, we will try to compare approaches, in two stages:

- extracting some key indicators from the complex phenotype
- plugging these indicators into a multidimensional GXE model with more integrative approaches, aiming to simultaneously model the dynamics in a complex integrated model, coupling dynamic and GXE interaction aspects.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Mathematics, computer and data sciences, digital technologies	MISTEA	Analysis of high-throughput phenotyping data
	MIAT	Statistical learning
	MaIAGE	Mixed models, GXE analysis models
Plant biology and breeding	GQE-Le Moulon	Statistical methods in quantitative genetics
Animal genetics	GenPhyse	Phenotyping and animal genetics DAC data analysis

Deep-Phenomic

EXPLORATORY PROJECT

2022-2024

Améliorer les performances de sélection chez les bovins laitiers grâce à la sélection phénomique

Coordination
Pascal Croiseau
UMR GABI
pascal.croiseau@inrae.fr

Key words
Phenomic selection
Genomic selection
Deep learning
Functional regression

INRAE units involved
GABI
MIA Paris Saclay
GQE-Le Moulon
AGAP Institut

Partnerships
Elliance

Context and challenges

In plant and animal genetics, selection programmes aim to identify individuals whose perfor-mance (yield, resistance to disease or environmental stress) meets previously defined criteria. This selection requires the acquisition of data, in the field or in breeding, which can be costly or time-consuming.

Since the 2000s, breeding programmes have used performance predictions to complement data on non-evaluated individuals. These predictions are based on information from the genome of the individuals: genotyping data. This strategy, known as genomic selection, has significantly increased the efficiency of breeding programmes for many animal and plant species and has become a ref-erence method in genetic improvement.

However, genomic selection has one drawback: the need to have genotyping data, which in some cases is too expensive to obtain (e.g. for field crop species for which thousands of candidates are produced each year, or for orphan species for which no efficient genotyping tool exists).

Phenomenal selection: a promising new alternative?

One alternative is to use phenomenal selection, recently introduced by Rincent et al. (2018), which consists of making performance predictions from phenomenal data obtained by spectroscopy, ra-ther than from genomic data. Spectroscopy has the advantage of being inexpensive, non-destructive, and already routinely implemented, both in breeding programmes for many plant species (to assess product quality) and in some animal species, notably in milk improvement pro-grammes.

The prediction performances obtained for different study cases are similar to those obtained with genomic prediction models. This very recent method has never yet been evaluated in an animal model and needs to be more widely tested and optimised.

© jcomp - Freepik

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

The Deep-Phenomic project proposes a first application of phenomenal selection to an animal model: the method will be tested in dairy cattle, in a large-scale system (several tens of thousands of animals with mid-infrared spectra on milk, of which approximately 8,000 are genotyped).

The results of the phenomenal predictions will be compared with those of a classical genomic evaluation.

The project also plans to optimise the exploitation of spectral data with functional methods on the one hand and neural networks on the other:

- functional analysis will be specifically tested in a multi-environment context, where the prediction of unobserved spectra could increase the accuracy of phenomenal prediction.
- Neural networks will be used to test the interest of artificial intelligence methods in the context of phenomenal selection, thanks to the very broad scope of the experiment.

If successful, this work could have important implications for dairy cattle improvement, and would constitute a proof of concept for many other animal and plant species.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Animal genetics	GABI	Genomic evaluation; bovine genetics
Mathematics, computer and data sciences, digital technologies	MIA Paris Saclay	Statistical learning, Artificial Intelligence
Plant biology and breeding	GQE-Le Moulon	Quantitative Genetics, Phenomic Selection, Cereals
	AGAP Institut	Quantitative Genetics, Phenomic Selection, perennial plants
External partners		**Expertises**
Elliance		Knowledge of bovine genomic evaluations

References

- **Rincent R, Charpentier J-P, Faivre-Rampant P, Paux E, Le Gouis J, Bastien C, Segura V (2018)** Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar. G3, 8(12), doi: https://doi.org/10.1534/g3.118.200760

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

Epipredict

EXPLORATORY PROJECT

2022-2024

Integrative Epigenetics to predict the adaptive capacities of pests

Coordination
Nadia Ponts
UR MycSA
nadia.ponts@inrae.fr
Gael Le Trionnaire
UMR IGEPP
gael.le-trionnaire@inrae.fr
David Causeur
Institut Agro -Rennes Angers
david.causeur@agrocampus-ouest.fr

Key words
Environmental stress
Acclimation
Association models
High-throughput functional data
Stochastic de-pendence

INRAE units involved
MycSA
IGEPP

Partnerships
Institut Agro – Rennes Angers

Context and challenges

Today, agriculture faces many challenges, including to avoid the development of certain pathogens resulting from the reduction in the use of inputs with a view to sustainable agriculture as well as the effects of climate change.

In this context, many questions arise in the short term about the adaptive capacities of these bio-aggressors. Will an insect pest resist the next heat wave? Or will it instead be greatly affected by rising temperatures and cease to be a threat ?

The EPIPREDICT project proposes to answer these questions for two examples of pests with clonal reproduction and remarkable resilience :

- the pea aphid Acyrthosiphon pisum, which causes a wide range of damage on various leguminous plants and has remarkable phenotypic plasticity in response to its biotic and abiotic environment;
- the mycotoxin-producing filamentous fungus Fusarium graminearum, which is responsible for disastrous episodes of Fusarium head blight in wheat throughout the world, and which also displays a formidable capacity for adaptation.

© Mathias70 - Pixabay

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

Epigenetic variations are heritable modifications of the expression of a genome that do not affect its sequence. Under environmental constraints, on short-time scales, the implementation of epigenet-ic modifications appears to be an efficient way for organisms to express new heritable phenotypes in order to ensure their survival and continue to develop. This epigenetic code is studied using high-throughput sequencing approaches, generating large volumes of data of a heterogeneous nature for which current analysis methods provide a limited understanding.

The EPIPREDICT project proposes to develop innovative statistical and mathematical approaches in order to identify in these data the elements that allow the description of variations in the expres-sion of genes (in particular those responsible for the virulence and aggressiveness of pathogens and pests), taking into account the spatial characteristics of genomes.

Ultimately, decoding how genes are expressed in response to the environment could provide a deci-sion support model for developing resilient and economically viable agro-ecosystems.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Plant health and environment	MycSA	Fungal functional genomics and epigenomics, bioinformatics
	IGEPP	Functional genomics and epigenomics of insect pests
External partners		**Expertises**
Institut Agro – Rennes Angers	Irmar	Functional data, high dimensional dependence, co-expression networks, computational statistics, software tool development

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

IMAGO

EXPLORATORY PROJECT

2022-2024

Exploring the function of hormone receptor signalling pathways in mammals

Coordination
Frédéric Jean-Alphonse
UMR PRC
frederic.jean-alphonse@inrae.fr
Béatrice Laroche
UMR MAIAGE
beatrice.laroche@inrae.fr

Key words
GPCR
Cell signalling
Intracellular imaging
Dynamic model
Complex system

INRAE units involved
PRC
MaIAGE

Partnerships
Inria
Imperial College London

Context and challenges

G protein-coupled receptors (GPCRs) play a key role in cellular communication in mammals. Among these, the membrane receptors for the gonadotropic hormones - luteinizing hormone (RLHCG) and follicle stimulating hormone (FSH) - are essential for reproduction.

The binding of each hormone to its receptor ultimately results in an adapted biological response through the transduction of several intracellular signalling pathways. These signalling pathways are relatively well described individually, but their organisation into networks is complex. Indeed, the biochemical reactions that make up these pathways are difficult to capture, as they are not only kinetically regulated, but also constrained in intracellular space. To date, the dynamics of signalling pathways are only described in a simplified manner: their interactions, spatial-temporal organisation and the intensity of signals remain poorly accessible simultaneously.

In order to understand how the cell decodes this complexity of intracellular signals and then produces graduated physiological regulations, it is essential to have a detailed analysis of the dynamics of the signalling networks and their organisation. That is where the IMAGO project comes in. It proposes to build models of the spatio-temporal organisation of signalling pathways associated with gonadotropin receptors.

© pikisuperstar - Freepik

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

The IMAGO project proposes to explore the complexity of the spatial-temporal organisation of signalling pathways and to develop dynamic models to understand their functioning at the cellular and molecular levels.

Firstly, the project plans to interrogate several distinct signalling pathways (cAMP, PKA, ERK and Ca2+) simultaneously, using fluorescence microscopy and biosensor approaches selectively addressed in various cellular compartments (e.g. nucleus, mitochondria, plasma membrane, endosomes, endoplasmic reticulum, etc.). The objective is to reveal the mechanisms of localisation of signalling pathways according to the site of activation of the receptor, the kinetics of activation and the spatial-temporal dynamics of biochemical reactions.

To do this, the IMAGO project will rely on the generation of data from biosensors of signalling pathways and fluorescence microscopy approaches.

These data will allow the following:

1. To perform a quantitative multiplexed analysis of the compartmentalisation of different signalling pathways and receptor trafficking
2. To develop a dynamic model of the signalling networks and receptor traffic

In the long term, this research into receptors of interest in reproductive physiology could lead to a review of traditional pharmacological approaches, which essentially target receptors located at the plasma membrane. This project will also provide new knowledge to the signalling network community.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Animal physiology and livestock systems	PRC	Quantitative cell imaging, real-time signalling, design and analysis of spatial-temporal models, deterministic and stochastic modelling
Mathematics, computer and data sciences, digital technologies	MaIAGE	Dynamic modelling (EDO, EDP), data analysis and estimation
External partners		**Expertises**
Inria	Équipe projet SERPICO	Microscopy image processing for intracellular transport analysis
	Équipe projet MUSCA	Dynamic modelling, endocrinology
Imperial College London		Traffic analysis and recycling of RLHCG and RFSH

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

MIDIIVEC

EXPLORATORY PROJECT

2022-2024

Coordination
Gaël Beaunée
UMR BIOEPAR
gael.beaunee@inrae.fr
Vincent Raquin
UMR IVPC
vincent.raquin@univ-lyon1.fr

INRAE units involved
BIOEPAR
IVPC
MaIAGE

Partnerships
Inria

New modelling approaches to anticipate vector-borne disease transmission

Context and challenges

Emerging arboviruses (e.g. Zika, West Nile virus) represent a global threat to human and veterinary public health. Mostly of zoonotic origin, these viruses are transmitted to vertebrate hosts by arthropod vectors, such as mosquitoes or ticks.

Transmitted by mosquitoes of the genus Aedes and Culex, the Rift Valley fever virus (RVFV) is endemic in Africa. However, its area of incidence is gradually expanding (Arabian Peninsula, Mayotte) with imported human cases reported in mainland France and China, making RVFV research a priority for the WHO and WOAH .

Arbovirus transmission is a dynamic, multi-scale process where small-scale individual infection dynamics can impact large-scale inter-population circulation, under the influence of several (a)biotic factors. At the vector scale, the ability of a mosquito to get infected then subsequently transmit an arbovirus is referred as vector competence, which depends notably on vector and virus genotype as well as temperature. Vector competence is characterized by three major steps :

1. Viral infection of the vector's gut following a blood meal on a viremic host
2. Dissemination of the virus from the gut into the circulatory system of the vector
3. Infection of the saliva, which conditions virus transmission to a new host during the next bite

At each barrier, infection can be stopped. However, each state of the vector (infected (I), disseminated (D) or infectious (T)) is irreversible, as the virus is not eliminated by vector's defences.

In epidemiological modelling on a population scale, vector competence is mostly studied as a qualitative phenotype (a vector is classified as competent or not), thereby ignoring the dynamic aspect of intra-vector viral infection (IVD) and its high potential epidemiological impact.

At epidemic scale, the distribution (in the mosquito population) of the time to reach the infectious state can have a major role on the epidemiological dynamics and the impact of biotic (genotype & viral dose) and abiotic (temperature) factors on IVD remains poorly characterized. Finally, the impact of IVD variability on large-scale vector transmission remains unknown. Characterising IVD and its (a)biotic determinants is therefore a major biological challenge.

The MIDIIVEC project aims to fill this knowledge gap in order to better anticipate and control the circulation of vector-borne diseases.

© Horror by Numbers - Unsplash

Goals

By mobilising an integrative and interdisciplinary approach linking experimental and numerical biology, the MIDIIVEC project intends to develop new models of IVD in order to better characterise its inter-individual heterogeneity. This will require the removal of several methodological barriers, both in mathematical modelling (in order to integrate IVD into multi-scale epidemiological models), in inference (to take into account an observational model in addition to the mechanistic model) and on issues of identifiability (i.e. to determine whether the available data allow the parameters to be estimated and with what bias and precision).

More precisely, the methodology is broken down into four steps:

1. Co-construction of mechanistic models of IVD with virologists to incorporate biological hypotheses of interest
2. Estimation of key parameters of these models to characterise the inter-individual heterogeneity of IVD
3. Co-construction of reasoned experimental designs to guide future experiments
4. Comparison of several modelling approaches at the vector scale to guide the integration of IVD in future epide-miological models on a larger scale

The ultimate goal is to propose new approaches for modelling IVD, in order to better understand its impact on arbovirus transmission.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Animal health	BIOEPAR	Mathematical modelling in epidemiology, stochastic simulations and inference
	IVPC	Entomology, Virology
Mathematics, computer and data sciences, digital technologies	MaIAGE	Stochastic modelling, inference (particle filtering)
External partners		**Expertises**
Inria	Équipe projet RAPSODI	Deterministic models (PDE), numerical analysis, optimisation

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

TEMPLATE

EXPLORATORY PROJECT

2022-2024

Simulating plant-pathogen interactions to better understand plant immune responses

Coordination
Frédérick Garcia
UMR MIAT
frederick.garcia@inrae.fr
Adelin Barbacci
LIPME
adelin.barbacci@inrae.fr

Key words
Modelling & simulation
DEVS
Multi-scale
Dynamic signalling
Plant pathology

INRAE units involved
MIAT
LIPME
IPSIM

Partnerships
Olivier Navaud (freelance consultant)

Context and challenges

Understanding how plants defend themselves against pathogens is a major challenge for moving towards an agriculture that uses fewer pesticides.

The immune response of plants differs from that of animals in that all plant cells are immunocompetent, i.e. they all have the same capacity to respond to the attack of a pathogen. This specificity has two im-portant consequences: (i) the regulation of immunity is an important determinant of the resistance phe-notype and (ii) the plant immune response is highly spatially structured, as the position of the cells is sta-ble.

Faced with attacks by necrotrophic fungi (including Sclerotinia sclerotiorum, which causes white rot), plants predominantly mount a form of immune response called Quantitative Disease Resistance (QDR). At present, QDR is mainly studied at the cellular level, without really taking into account the spatial and tem-poral dimensions of this resistance.

Recent work shows that plant-pathogen interactions are closely linked to the spatial and temporal charac-teristics of the entities and processes involved.

To better understand QDR, it is therefore necessary to integrate the dynamics of environmental percep-tion, signalling and transduction from the sub-cellular to the organ or whole-plant scale.

For this, we propose to use modelling and computer simulation techniques of plant tissues at the cellular and multi-cellular scales. Although the interest of this approach is widely recognised for studying complex systems, involving the interaction of a very large number of entities in a network, it is still innovative in plant biology.

© Jymm

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

The TEMPLATE project aims to implement a dynamic model of plant-fungus interaction, in order to test several hypotheses on the establishment of the immune response in time and space by simulation. This model will attempt to represent the progression of a colony of mycelium of the pathogenic fungus Sclero-tinia sclerotiorum in a leaf of the model plant A. thaliana.

The biological question at the heart of this project will be that of the formation of patterns of immune response localised in time and space and associated with the modulation of the level of resistance linked to transcriptomic reprogramming.

We propose to use Discrete Event System Specification (DEVS), a formalism which is known for its repro-ducibility and modularity and allows a multi-formalism modelling approach.

In order to fully support biologists in their experiments, we will aim to develop a computational frame-work that allows interactive modelling and simulation bringing the experimenter, biological subject and digital model together.

This project therefore aims both for a better understanding of the immune response of plants and for the development of a new methodology for interactive simulation in the field of biology.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Mathematics, computer and data sciences, digital technologies	MIAT	Computer science, modelling, simulation, bioinformatics
Plant health and environment	LIPME	Modelling, plant pathology, molecular biology
	IPSIM	Cellular signalling, plant physiology, imaging
External partners		**Expertises**
Olivier Navaud (freelance consultant)		Bibliographic synthesis, creation of a pathway signalling/plant immunity database

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

Behind the Count'her

SCIENTIFIC NETWORK BEHIND THE COUNT HER

2022-2023

Modelling the heritability of traits from count data

Coordination
Bertrand Servin
bertrand.servin@inrae.fr

Key words
Count data
Poisson-lognormal model
Heritability
Statistics
Genetics

INRAE units involved
MIA Paris Saclay
MaIAGE
GenPhySE

Partnerships
University of Paris

Context and challenges

Understanding the genetic determinism of a trait, i.e. the set of genes involved in the development and expression of this trait, is a major challenge for better understanding biological processes and supporting genetic improvement programmes.

The recent development of genotyping and massive sequencing tools, which allow the rapid sequencing of several thousand to millions of DNA or RNA molecules simultaneously, has considerably increased the power of experimental devices in this field, leading to new challenges in analysing massive sequencing data.

In this context, one of the emerging issues is the analysis of data that correspond to a number of observed events (count data). Indeed, the analysis of this type of data by hierarchical generalized linear models is notoriously difficult, in particular when it comes to estimating the heritability of traits.

The Behind the Count'her consortium proposes to use recent statistical developments to adapt a model for estimating genetic parameters. It will be based on two cases of application in the field of breeding: the distribution of recombinations along the genome and the diversity of the ruminal meta-genome.

Goals

Consortium partners have recently proposed a flexible statistical model for count data, based on the Poisson-lognormal distribution, which allows complex effects to be modelled and estimated in reasonable time: covariance structure, clustering and dimension reduction, network inference, etc. However, adapting this model to the context of inferring genetic parameters requires specific developments.

The Behind the Count'her consortium therefore proposes to bring together teams from quantitative genetics and statistics to jointly develop new statistical models for the analysis of count data.

The developments will be based on two study contexts that allow a wide range of applications to be covered:

- The modelling of data on the distribution of crossings on the genome during meiosis (for which the genes involved and their variations remain unknown in many species).
- Analysis of the diversity of the meta-genome between individuals and use of these results to measure the effect of the microbiota on other traits.

The ambitions of the Behind the Count'her consortium are both methodological and genetic. The consortium will both remove methodological barriers to the efficient exploitation of count data and also provide new knowledge on the genetic effects of recombination phenotypes and microbiota.

Research units involved and partners

INRAE scientific division	INRAE research units	Expertise
Mathematics, computer and data sciences, digital technologies	MIA Paris Saclay	Statistical Modelling
	MaIAGE	Statistical Modelling, Evolutionary Genomics
Animal genetics	GenPhySE	Metagenomics, Genetics, Statistics
External partners		**Expertise**
University of Paris	LPSM	Statistical Modelling

Reference

- **J. Chiquet, M. Mariadassou and S. Robin:** The Poisson-lognormal model as a versatile framework for the joint analysis of species abundances, Frontiers in Ecology and Evolution, 2021
- **S. Fresco, C. Marie-Etancelin, A. Meynadier, G. Martinez-Boggio:** Variation in Rumen Bacteria of Lacaune Dairy Ewes From One Week to the Next, Frontiers in Microbiology, 2022
- **H. Vassilief, M. Id Bella, D. Hazard, F. Tortereau, T. Faraut, S.E. Johnston, B. Servin:** Sex differences in recombination maps are associated with differential hotspot usage in Sheep. Proceedings of the World Congress on Genetics Applied to Livestock Production, Rotterdam, 2022

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

MIMS

SCIENTIFIC NETWORK MIMS

2022-2023

Cross Methodological Insights for Multi-source Data Integration

Coordination

Mohamed Hanafi
StatSC
mohamed.hanafi@oniris-nantes.fr
Jean Michel Roger
UMR ITAP
jean-michel.roger@inrae.fr

INRAE units involved

StatSC
BIA
QuaPA
SPO
LBE
ITAP
MaIAGE
CSGA Centre des Sciences du Goût et de l'Alimentation
UNH Unité Nutrition Humaine
PhAN
LABERCA
Micalis
Prose
BioForA
LBLGC
AGAP Institut
SELMET

Partnerships

Faculté des Sciences, Paris
INRIA
University of Genève
University of Toulouse
ANSES
CNAM
University of Paris-Saclay
University of Montpellier
ADLIN

Context and challenges

In biology, as in other scientific fields, the integration of multi-source data is more relevant than ever. Indeed, the data collected are increasingly complex and their volume is growing, due to the development of analytical platforms, imaging techniques, the rise of omics data, etc.

This context has stimulated the search for new methods allowing the joint analysis of several data sets (structured data, multi-block, multi-channel) in many fields, such as:

- Machine Learning, where several approaches are considered for the processing of multi-source data (matrix factorisation, probabilistic approach).
- Chemometrics, where different methods are proposed to establish a chemical mapping of samples using several analytical techniques (generalisations of canonical analysis, NIPALS algorithm and tensor decompositions)
- Bioinformatics, where integrative methodological approaches allow the most complete picture possible of the dynamics of molecular systems to be drawn.

In order to contribute to meeting the challenge of analysing and exploiting these multi-source data from an exploratory, but also predictive perspective, it is essential to bring together different viewpoints, practices and paradigms in order to reconcile these different approaches. It is also necessary to encourage collaboration between "method generators" and "data generators" in the various application fields.

This is the challenge that the MIMS consortium proposes to take up, by bringing together an interdisciplinary community working on approaches to the analysis and integration of multi-source data.

Logo MIMS

Metaprogramme
DIGIT-BIO

Contact : digitbio@inrae.fr
www.inrae.fr/digitbio/

French Wine and vine Institut

Goals

MIMS is a multidisciplinary consortium gathering more than 60 researchers, whose objective is to examine the analysis and exploitation of multi-source data, both in an exploratory and predictive perspective.

This consortium brings together multidisciplinary skills: information processing, biological sciences and analytics. The implementation of this multi-disciplinarity and its management will be based on the sharing of data, practices and methods between the partners, with the aim of formalising a scientific project to meet a common challenge: the opti-mal analysis of multi-source data for exploratory and predictive purposes.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Sciences for food, bioproducts and waste engineering	USC StatSC	Sensometry, Chemometrics, Statistics, Multispectral imaging
	BIA	Chemometrics, computer science
	QuaPA	Volatolomics, MRI Chemometrics, Data Analysis, Image Analysis, System & Data Management
	SPO	Chemometrics
	LBE	Biostatistics, machine learning
Mathematics, computer and data sciences, digital technologies	ITAP	Chemometrics
	MAIAGE	Mathematical statistics, applied statistics, bioinformatics
Human nutrition and food safety	CSGA Centre des Sciences du Goût et de l'Alimentation	Chemometrics
	UNH Unité Nutrition Humaine	Bioinformatics, metabolomics, chemometrics
	PhAN	Perinatal nutrition and metabolic diseases, Bioinformatics, Data analysis, metagenomics and metabolomics
	LABERCA	Metabolomics, Chemometrics, Expology, Epidemiology
Microbiology and the food chain	Micalis	Biologist, Microbiota, Data Analysis
	Prose	
Ecology and biodiversity of forest, grassland and freshwater environments	BioForA	Quantitative Genetics, Modelling
	LBLGC	Physiology
Plant biology and breeding	AGAP Institut	Quantitative genetics, Genomics, Biochemistry, Evolutionary genetics, Selection, Ecophysiologist, Biostatistics, Bioinformatics
Animal physiology and livestock systems	SELMET	Biometrics, Chemometrics, Machine Learning, Agronomy
External partners		**Expertises**
Faculté des Sciences, Paris	Centre Boreli	Unsupervised learning, Statistics, Graph networks, Bioinformatics
INRIA	Équipe projet LORIA	Knowledge Discovery, Life Sciences

University of Genève	Sciences Analytiques	Metabolomics, Chemometrics
University of Toulouse	Institut de mathématique de Toulouse	Statistics, Multi-omics data analysis and integration
ANSES	Laboratoire de Ploufragan-Plouzané	Statistics, multi-block methods Epidemiology
CNAM	EPN6 - Mathématiques et Statistique	Analysis of complex heterogeneous data, Clusterwise methods, High dimensional classification
University of Paris-Saclay	Signaux et Statistique	Multi-block data analysis, tensor analysis (high dimensional), Structural equation models
University of Montpellier	Institut Montpelliérain Alexander Grothendieck	Supervised component models, classification
ADLIN	ADLIN	Finance, Strategy, Multi-omics, Bioinformatics, Transcriptomics, Visualisation
French Wine and vine Institut	IFV	Chemometrics, Analytical Chemistry

References

- Bersanelli, M, Mosca E, Remondini D. et al. Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics 17, S15 (2016). https://doi.org/10.1186/s12859-015-0857-9
- **Boccard J, Schvartz D, Codesido S, Hanafi M, Gagnebin Y, Ponte B, Jourdan F, Rudaz S. (2021).** Gaining insights into metabolic networks using chemometrics and bioinformatics: chronic kidney disease as a clinical model. *Frontiers in Molecular Biosciences* 8, 682559.
- **Boutalbi R, Labiod L, Nadif M. (2021):** Implicit consensus clustering from multiple graphs. D*ata Min. Knowl. Discov.* 35(6): 2313-2340
- **Hanafi M, Kiers H. A. L. (2006).** Analysis of K sets of data, with differential emphasis on agreement between and within sets. *Computational Statistics and Data Analysis.* (51), 3, 1491-1508.
- Eicher T, Kinnebrew G, Patt A, Spencer K, Ying K, Ma Q, Machiraju R, Mathé AEA. Metabolomics and Multi-Omics Integration: A Survey of Computational Methods and Resources. Metabolites. 2020 May 15;10(5):202. doi: 10.3390/metabo10050202. PMID: 32429287; PMCID: PMC7281435.

Metaprogramme DIGIT-BIO
Contact : digitbio@inrae.fr

www.inrae.fr/digitbio

National Research Institute for Agriculture, Food and Environment

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INRAE

Graphic design : INRAE // Layout: Plume & Sciences // Cover photos: Kjparjeter - Freepik, montage: Comscience / Etienne Delannoy