



HAL
open science

Réflexion prospective interdisciplinaire. Approches prédictives pour la biologie et l'écologie. Rapport de synthèse.

Carole Caranta, Hervé Monod, Hugues Berry, Michaël Chelle, Michel Génard, Fabien Jourdan, Christian Lannou, Emmanuelle Maguin, Françoise Médale, Sylvie Oddou-Muratorio, et al.

► To cite this version:

Carole Caranta, Hervé Monod, Hugues Berry, Michaël Chelle, Michel Génard, et al.. Réflexion prospective interdisciplinaire. Approches prédictives pour la biologie et l'écologie. Rapport de synthèse.. [0] 2019, 40 pp. hal-02791153v1

HAL Id: hal-02791153

<https://hal.inrae.fr/hal-02791153v1>

Submitted on 5 Jun 2020 (v1), last revised 24 Nov 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC0 - Public Domain Dedication 4.0 International License



RÉFLEXION PROSPECTIVE INTERDISCIPLINAIRE

Approches prédictives pour la biologie et l'écologie

Rapport de synthèse

Pour citer ce rapport :

Caranta C., Monod H., Berry H., Chelle M., Génard M., Jourdan F., Lannou C., Maguin E., Médale F., Oddou-Muratorio S., Rogel-Gaillard C., Traas J. 2019. Réflexion prospective interdisciplinaire Approches Prédicatives pour la Biologie et l'Écologie. Rapport de synthèse. 40pp.

Contributeurs

Coordinatrice

- Carole Caranta, INRA, Cheffe du département Biologie et Amélioration des Plantes (BAP)

Comité de pilotage

- Hugues Berry, INRIA
- Michaël Chelle, DTN, INRA, département Environnement et Agronomie (EA)
- Michel Génard, INRA, département EA
- Fabien Jourdan, INRA, département Alimentation Humaine (AlimH)
- Christian Lannou, INRA, Chef du département Santé des Plantes et Environnement (SPE)
- Emmanuelle Maguin, INRA, département Microbiologie et Chaîne Alimentaire (MICA)
- Françoise Médale, INRA, Cheffe du département Physiologie Animale et Systèmes d'Élevage (PHASE)
- Hervé Monod, INRA, Chef du département Mathématiques et Informatique Appliquées (MIA)
- Sylvie Oddou-Muratorio, INRA, département Ecologie des Forêts, Prairies et milieux Aquatiques (EFPA)
- Claire Rogel-Gaillard, INRA, département Génétique Animale (GA)
- Jan Traas, INRA, département BAP

Composition des groupes de travail

Régulations et compromis métaboliques

Animateurs : Françoise Médale (PHASE), Yves Gibon (BAP)

Représentants : Sophie Barbe (CEPIA), Frédérique Clément (INRIA), Zhanwu Dai (EA), Hélène Fouillet (AlimH), Stéphane Genin (SPE), Anne Goelzer (MIA), Florence Gondret (PHASE), Hervé Guillou (SA), Stéphanie Heux (MICA), Gabriel Krouk (CNRS), Jean-Charles Leple (EFPA), Tatiana Zerjal (GA).

Morphogenèse et développement des organismes

Animateurs : Philippe Andrey (BAP), Jan Traas (BAP, COPIL)

Représentants : Olivier Hamant (BAP), Bruno Moulià (EA), Frédérique Pitel (GA), Alline Reis (PHASE), Caroline Teyssier (EFPA), Alain Trubuil (MIA.)

Modélisation et prédiction des phénotypes

Animateurs : Michel Génard (EA, COPIL), Claire Rogel-Gaillard (GA, COPIL)

Représentants : Pierre Casadebaig (EA), Marie-Christine Champomier (MICA), Timothée Flutre (BAP), Hélène Gilbert (GA), Sandrine Lagarrigue (GA), Fabrice Laurent (SA), Brigitte Mangin (MIA), Pierre Martre (EA), Nicolas Parisey (SPE), Laurence Puillet (PHASE), Bénédicte Quilot-Turion (BAP), Stéphane Robin (MIA).

Holobiontes, connaissance et modélisation des interactions entre systèmes microbiens et hôtes

Animatrices : Béatrice Laroche (MIA), Emmanuelle Maguin (MICA, COPIL)

Représentants : Romain Barnard (EA), Matthieu Barret (SPE), Julia Buitink (BAP), Joël Dore (MICA), Jordi Estelle Fabrellas (GA), Sonia Lamande (SA), Emmanuelle Le Chatelier (MICA), Rafael Munoz-Tamayo (PHASE), Thierry Thomas-Danguin (CEPIA), Corinne Vacher (EFPA).

Toxicologie et écotoxicologie prédictive

Animateurs : Fabien Jourdan (AlimH, COPIL), Jeanne-Marie Membre (MICA)

Représentants : Marie-Agnès Coutellec (EFPA), Jean-Pierre CRAVEDI (AlimH), Juliette Dibie (MIA), Erwan Engel (CEPIA), Florence Forget (MICA), Sophie Langouet-Prigent (INSERM), Claire Nedellec (MIA), Isabelle Oswald (SA), Anne Siegel (CNRS PHASE), Jessica Tressou (MIA), Julien VIGNARD (SA).

Dimension prédictive de la démo-génétique appliquée aux changements globaux et aux invasions biologiques

Animateurs : Christian Lannou (SPE, COPIL), Sylvie Oddou-Muratorio (EFPA, COPIL)

Représentants : Xavier Bailly (SA), Jérôme Coville (MIA), Isabelle Litrico-Chiarelli (BAP), Gaéтан Louarn (EA), Ludovic Mailleret (EA), Nathalie Peyrard (MIA), Christelle Robinet (EFPA), Joëlle Ronfort (BAP), Bertrand Servin (GA), Samuel Soubeyrand (SPE).

Analyses et modèles multi-échelles et multi-systèmes

Animateurs : Hugues Berry (INRIA, COPIL), Hervé Monod (MIA, COPIL)

Représentants : Bruno Andrieu (EA), Pauline Ezanno (SA), Robert Faivre (MIA), Nicolas Friggens (PHASE), Vincent Fromion (MIA), Florence Jaffrezic (GA), David Legland (CEPIA), Marie-Laure Magniette (BAP), Nathalie Perrot (CEPIA), Jérôme Salse (BAP), Patrick Taillandier (MIA), Alain Trubuil (MIA), Alain Viari (INRIA), Romain Yvinec (PHASE).

Avec nos remerciements à Evelyne BARBIN, assistante de direction du département BAP pour son appui pendant la durée de la prospective et son aide pour la mise en forme du rapport.

Sommaire

1 – INTRODUCTION	- 7 -
1.1 – Approches prédictives : définition	- 7 -
1.2 – Contexte et enjeux du renouvellement des sciences du vivant par les approches prédictives -	7 -
1.3 – 2003-2017 à l’Inra : De la modélisation aux approches prédictives	- 9 -
1.4 – Prospective 2017 : mission	- 10 -
1.5 – Méthode de travail et structure du document	- 10 -
2 – DEFIS SCIENTIFIQUES	- 13 -
2.1 – Périmètre, enjeux et priorités des 7 thèmes	- 13 -
2.1.1 – Régulations et compromis métaboliques.....	- 13 -
2.1.2 – Morphogénèse et développement des organismes	- 16 -
2.1.3 – Modélisation et prédiction des phénotypes	- 19 -
2.1.4 – Holobiontes : connaissances et modélisation des interactions entre systèmes microbiens et hôtes	- 22 -
2.1.5 – Toxicologie et écotoxicologie prédictive.....	- 25 -
2.1.6 – Dimensions prédictives de la démo-génétique appliquée aux changements globaux et aux invasions biologiques	- 28 -
2.1.7 – Analyses et modèles multi-échelles et multi-systèmes	- 31 -
2.2 – Interfaces entre thèmes	- 35 -
3 – RECOMMANDATIONS	- 36 -
Axe 1 – Fédérer la communauté scientifique pour développer l’interdisciplinarité	- 36 -
Axe 2 – Soutenir des projets de recherche innovants	- 37 -
Axe 3 – Mettre en place les outils, les méthodes et les technologies répondant aux enjeux de la modélisation multi-échelle et multi-système	- 37 -
Axe 4 – Mettre en place des actions conjointes avec des partenaires académiques nationaux	- 38 -
Axe 5 – Accroître le rayonnement et l’attractivité de l’Inra aux échelles nationale, européenne et internationale	- 38 -
Axe 6 – Monter en compétence	- 39 -
Axe 7 – Établir et mettre en œuvre une stratégie de partenariat socio-économique	- 39 -
Axe 8 – Alimenter et partager un cadre éthique et juridique en lien avec les approches prédictives pour l’évaluation et la gestion	- 39 -
4 – MISE EN PERSPECTIVE	- 40 -

1.1 – Approches prédictives : définition

Parler d'approches prédictives aujourd'hui, c'est se fixer des ambitions scientifiques fortes et conjointes en termes de compréhension, d'anticipation et de capacité d'action sur des systèmes qui sont à la fois complexes, dynamiques et incertains. Il s'agit en effet de mesurer le degré de compréhension de tels systèmes, par la capacité à en prédire de façon fiable le comportement (à la fois qualitatif et quantitatif) dans des environnements divers et variés, possiblement différents, voire en rupture avec les environnements dans lesquels ils ont été observés. L'intérêt porté aux approches prédictives est consubstantiel d'une approche de plus en plus interdisciplinaire de la science, mêlant sciences du vivant et sciences formelles pour des objectifs partagés. Il est étroitement lié à l'explosion des données (« big data ») et aux différentes activités que la transition numérique bouleverse en profondeur : acquisition et partage de données, modélisation et simulation, algorithmique et méthodes d'apprentissage. Ce tournant a été rendu possible par la mise en place d'infrastructures informatiques (« data centre ») permettant d'accroître la puissance de calcul et donc d'augmenter le pouvoir prédictif des algorithmes (e.g., « deep learning »). En termes de finalités, les approches prédictives s'inscrivent dans des objectifs de compréhension fine des systèmes biologiques et écologiques, mais aussi de prédiction de la réaction de ces systèmes aux changements, ou encore de gestion, d'optimisation voire de conception de systèmes biologiques adaptés à des objectifs fixés. En termes de méthodes, deux grandes classes de démarches se distinguent, se complètent et s'hybrident parfois : d'une part celles dites « pilotées par les données » (« data-driven ») dont l'objectif est de prédire des variables sur de nouveaux individus par des techniques d'apprentissage ; d'autre part celles dites « pilotées par les concepts » (« concept-driven ») qui s'attachent à l'établissement et à l'analyse d'un modèle basé sur des lois physiques, physico-chimiques ou biologiques et ont une visée explicative plus affirmée.

1.2 – Contexte et enjeux du renouvellement des sciences du vivant par les approches prédictives

Les sciences du vivant sont incontestablement en pleine transformation, transformation provoquée par les « big data ». En effet, l'explosion quantitative des données en biologie, permise par le développement sans précédent des technologies d'acquisition à toutes les échelles (données omiques, phénotypiques, écologiques, climatiques, ...) et par l'essor des sciences participatives, permet d'aborder des questionnements de plus en plus complexes sur le fonctionnement du vivant. Par exemple, la biologie des systèmes s'appuie sur l'acquisition de données aussi exhaustives que possible concernant le système étudié.

En lien direct, l'utilisation de méthodes sophistiquées de représentation des connaissances, d'analyse des données, de modélisation et de simulation, est devenue essentielle pour l'exploitation des données. Ainsi, depuis une vingtaine d'années, en complément aux approches classiquement utilisées en biologie, des développements mobilisant les mathématiques, les sciences de l'information, la physique ou encore la chimie ont fait naître de nouvelles communautés de recherche associant les sciences du vivant et une ou plusieurs de ces disciplines.

De façon connexe, les sciences du vivant ont également connu ces dernières années un changement de paradigme dans le sens où les systèmes biologiques sont aujourd'hui considérés comme des systèmes dynamiques complexes pouvant et devant être étudiés en tant que tels. En effet, il n'est pas possible de déduire simplement le comportement global des systèmes biologiques à partir des propriétés de leurs composantes. Les approches des systèmes complexes offrent un cadre de travail pour comprendre et prédire leur comportement par la mobilisation systématique de méthodes mathématiques et informatiques et d'outils de modélisation, d'inférence (calibration, évaluation) et de simulation. Grâce à la formalisation des interactions et des propriétés particulières des éléments constitutifs des systèmes complexes, de multiples propriétés émergentes peuvent être observées et intégrées entre différentes échelles.

Plus récemment, l'essor de l'apprentissage automatique et de l'intelligence artificielle permet de revisiter les approches prédictives et promet de modifier en profondeur la façon dont elles sont conduites en interface entre l'homme et la machine, en s'appuyant sur des corpus de données numériques, textuelles et d'imagerie plus vastes et sur des algorithmes plus autonomes pour extraire et synthétiser de la connaissance à partir de tels corpus. Ainsi le rapport Villani¹ parle de « *rupture* » et de « *passage progressif d'une logique de programmation à une logique d'apprentissage* ». Le programme de recherche MINDS annoncé récemment par Mines ParisTech évoque de son côté l'ambition de faire la « *synthèse entre simulation numérique et intelligence artificielle* ».

Il s'agira cependant de rester vigilant et conscient des potentielles limites (qu'il s'agit encore d'analyser) de ces approches basées sur la force brute des données (e.g., « effet lampadaire » mentionné dans le rapport Villani).

Dans ce contexte, les enjeux pour l'Inra se situent à trois niveaux :

- **Des enjeux de recherche et d'exploration de fronts de science** pour répondre à de nombreux défis stratégiques pour l'Institut. Même délimitées au périmètre de la biologie et de l'écologie, les approches prédictives irriguent de façon exponentielle l'ensemble de nos recherches et par voie de conséquence contribuent à une multiplicité d'enjeux tels que la prédiction et l'optimisation des phénotypes dans des environnements divers et fluctuants, la modélisation des interactions biotiques pour une vision et une gestion agroécologique des agroécosystèmes, l'identification de biomarqueurs et de marqueurs génétiques de caractères d'intérêts pour des interventions adaptées, la prédiction de l'exposition et des effets de composés ou microorganismes toxiques pour la santé et l'environnement, la prédiction des impacts écologiques des changements globaux et des invasions biologiques, l'optimisation de processus biotechnologiques, etc...
- **Des enjeux concernant l'exploitation des données massives** pour répondre aux défis décrits ci-dessus. Ces enjeux incluent notamment des problématiques de représentation des connaissances, d'extraction de l'information pertinente des résultats expérimentaux, d'intégration de données de natures différentes (e.g., données génomiques, phénotypiques, écologiques, climatiques, ...), d'analyse par des méthodes d'apprentissage et de modélisation. Ces enjeux sont étroitement liés à l'organisation des systèmes d'information, de stockage et de calcul (infrastructures matérielles), ainsi qu'à la planification et à l'évolution des méthodes d'acquisition de données.

¹ https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

- **Des enjeux de fédération, d'animation et de soutien à une nouvelle communauté scientifique interdisciplinaire** sur les approches prédictives en biologie et écologie dans un objectif de rayonnement de l'Inra aux échelles nationale, européenne et internationale. Dans ce cadre, il s'agit bien de décloisonner les communautés s'intéressant à des échelles d'organisation du vivant, des objets ou systèmes, et des thèmes scientifiques différents.

1.3 – 2003-2017 à l'Inra : De la modélisation aux approches prédictives

En lien avec la montée en puissance des approches de modélisation dans la presque totalité des thèmes scientifiques de l'Inra, des initiatives internes et collectives ont été initiées dès les années 2000. Ainsi de 2003 à 2005, une mission inter-départements, coordonnée par B. Goffinet² analyse le rôle de la modélisation et la place de l'Inra dans sept champs de recherche concernés par la modélisation et conduit à des propositions d'organisation et d'animation à l'échelle de l'Institut. Peu après, en lien avec l'essor des données en génomique et post-génomique, le rapport « Biologie Intégrative Végétale » (2005) contribue à la mise en place du programme fédérateur de recherches en biologie intégrative animale, végétale et microbienne (agroBI, 2006-2008), à une intensification et une formalisation des collaborations avec l'Inria (2008) et à la mise en place d'une « cellule bioinformatique » transversale aux départements et chargée d'animer, de coordonner et d'être force de proposition sur cette thématique (2009-2016). Dans le document d'orientation Inra 2010-2020³, les approches prédictives pour la biologie figurent déjà comme l'un des deux chantiers scientifiques interdisciplinaires prioritaires. Dans ce cadre, une première réflexion coordonnée par F. Garcia explicitement consacrée aux approches prédictives est conduite en 2012-2013 et propose une synthèse de l'offre et des besoins en matière de formation, ainsi que des propositions relatives à l'animation et au soutien du dispositif. Les réflexions sur le sujet sont également marquées pendant cette période par un focus du chantier scientifique « Agro-écologie » sur les données et la modélisation⁴. En parallèle à ces initiatives, le conseil scientifique de l'Inra s'est emparé du sujet à plusieurs reprises. En 2015, il émet un avis recommandant de poursuivre la réflexion en insistant sur son caractère interdisciplinaire et sur l'importance d'aboutir à une proposition de plan d'action, tout en précisant les questions prioritaires à traiter.

Au-delà de ces actions, dont les livrables et impacts n'ont d'ailleurs pas toujours été évalués, il ressort très clairement de l'analyse des schémas stratégiques des départements (SSD) 2016-2020 et des documents de premier bilan des méta-programmes (Février 2018) que les approches prédictives en biologie et écologie sont prises en compte dans la stratégie scientifique de l'Institut. Ainsi, l'ensemble des départements considère de façon explicite les approches prédictives en biologie et/ou écologie soit comme un enjeu structurant en tant que tel dans leurs SSD, soit comme une approche à développer en alimentation de priorités des SSD.

² http://www7.inra.fr/mia/doc/rapport_modelisation_VO.pdf

³ <http://inra.dam.front.pad.brainsonic.com/ressources/afile/224599-e325e-resource-document-d-orientation-2010-2020.html>

⁴ Garcia F., Gascuel-Oudou C., Soussana J.F. 2014. *Colloque*. Nouveaux défis de la modélisation : l'agro-écologie, Synthèse, INRA, 49 p

1.4 – Prospective 2017 : mission

Dans ce contexte et celui du plan d'action associé au document d'orientation #Inra2025⁵, les approches prédictives pour la biologie et l'écologie ont été choisies comme sujet pour l'un des quatre chantiers de prospective scientifique interdisciplinaire lancés en 2017. L'ambition est d'éclairer l'Institut sur les futurs fronts de science et d'enrichir le choix de nos orientations en proposant des actions qui favorisent le développement et le rayonnement international de nos recherches et contribuent à répondre à de grands défis sociétaux. Une déclinaison opérationnelle est attendue.

Le périmètre de la prospective couvre trois grands domaines identifiés dans l'objectif #OpenScience-3 :

- La biologie des systèmes et l'intégration des mécanismes de l'échelle moléculaire à celle de l'organisme dans son environnement biotique et abiotique ;
- La dynamique de populations et de communautés à l'échelle d'un écosystème ;
- L'exploitation de données massives dans les domaines pertinents.

1.5 – Méthode de travail et structure du document

Une réflexion approfondie sur les approches prédictives pour la biologie et l'écologie ne pouvait faire l'économie d'un état des lieux sur les différents systèmes en biologie et écologie modélisés à l'Inra, les approches mobilisées, les objectifs et les finalités des projets développés (cet état des lieux constitue un document de travail interne). En s'appuyant sur un premier retour par les départements, le comité de pilotage (COFIL) a élaboré une synthèse. Les modèles y sont classés selon l'échelle (ou les échelles) appréhendée(s), et selon une distinction entre modèles empiriques et modèles mécanistes. Cette typologie est forcément imparfaite mais les échanges générés ont été utiles pour cartographier les projets menés à l'Inra et établir « la ligne de base » de la prospective. La distinction entre modèles empiriques et modèles mécanistes a été choisie dès le départ car elle correspond à deux visions distinctes et *a priori* complémentaires de la biologie prédictive, l'une reposant sur la "force brute" de la science des données, l'autre sur une compréhension plus fine des mécanismes (bio)physiques. Ce choix n'a évidemment pas empêché de porter aussi la réflexion sur les enjeux qui se jouent aux interfaces entre ces deux visions, en lien avec le rôle essentiel des données dans un contexte d'évolutions technologiques et méthodologiques évoqué plus haut.

Sept thèmes de travail ont ensuite été identifiés collectivement en réponse à des enjeux cognitifs et/ou finalisés stratégiques pour l'Inra, en complémentarité avec le travail conduit dans le cadre des autres prospectives, ces thèmes ayant vocation à prendre de l'ampleur ou affirmer notre leadership dans les années à venir. Ils ont fait l'objet d'un travail approfondi par des groupes de travail (GT) pluridisciplinaires qui ont mobilisé environ 90 scientifiques de 11 des 13 départements de recherche de l'Inra, ainsi que quelques scientifiques de l'Inria, du CNRS et de l'INSERM. Ces GT ont été pilotés par des membres du COFIL.

Compte tenu de l'ampleur du sujet, le choix a été fait de traiter séparément ces 7 thèmes (correspondant souvent à des échelles d'organisation du vivant) tout en étant bien conscients des nombreuses interfaces, certaines étant elles-mêmes porteuses de fronts de science et d'enjeux d'intégration. En dernière étape, ces interfaces ont fait l'objet d'une réflexion.

⁵ <https://inra-dam-front-resources-cdn.brainsonic.com/ressources/afile/367569-9d913-resource-document-d-orientation-inra-2025.pdf>

Les 3 premiers thèmes concernent les approches prédictives à l'échelle de l'organisme et infra, l'organisme étant bien entendu considéré dans son environnement. Les 3 thèmes suivants concernent les approches prédictives à l'échelle d'un (éco) systèmes. Le dernier thème, plus méthodologique, est transversal et ambitionne de répondre à des défis partagés par les 6 thèmes précédents ainsi qu'à leurs interfaces, dans les domaines de l'analyse et de la modélisation.

Thème 1 : Régulations et compromis métaboliques. Il s'agit de décrire et hiérarchiser les acteurs du réseau métabolique et leurs dynamiques en réponse à l'environnement, de prédire les réponses phénotypiques et identifier des biomarqueurs, et d'optimiser ou créer de nouvelles voies métaboliques.

Thème 2 : Morphogénèse et développement des organismes. Une des questions fondamentales de la biologie est de comprendre comment des individus avec des structures et formes spécifiques se développent. Dans ce contexte, deux questions centrales se posent : quels sont les liens entre la régulation moléculaire, le comportement cellulaire et le développement de l'individu multicellulaire et comment ces liens sont-ils modulés par l'environnement ?

Thème 3 : Modélisation et prédiction des phénotypes. Dans un contexte de fortes attentes liées aux changements globaux, il est nécessaire de progresser dans nos capacités de prédiction des phénotypes d'intérêt, aussi bien au fil des générations qu'au cours de la vie, y compris très précocement, dans l'objectif notamment d'améliorer et/ou renouveler les outils pour la sélection et plus largement l'aide à la décision.

Thème 4 : Holobiontes, connaissances et modélisation des interactions entre systèmes microbiens et hôtes. Il s'agit de proposer une vision systémique de l'hôte et de son microbiote, de la modéliser pour développer un holobionte virtuel permettant de prédire les leviers d'action, les effets et l'adaptation de ces systèmes aux conditions environnementales pour une diversité d'applications (agroécologie, santé animale et santé humaine, efficacité de production, en particulier).

Thème 5 : Toxicologie et écotoxicologie prédictive. L'objectif est de mieux comprendre et prédire les risques en évaluant plus précisément l'exposition aux composés ou microorganismes en lien avec les pratiques de culture et d'élevage, lors de la transformation ou la consommation, ainsi que les effets de cette exposition sur la santé humaine, animale et sur l'environnement.

Thème 6 : Dimension prédictive de la démo-génétique appliquée aux changements globaux et aux invasions biologiques. Les changements globaux (climat, échanges, occupation des sols, prélèvement des ressources naturelles, etc.) ont des conséquences multiples sur la distribution des espèces et le fonctionnement des écosystèmes. Ce thème focalise sur la biologie des populations, et sur les approches permettant de prédire les impacts écologiques des changements globaux et des invasions biologiques.

Thème 7 : Analyses et modèles multi-échelles et multi-systèmes. Pour représenter le comportement de systèmes dynamiques complexes rencontrés en biologie et écologie au sens large, la prise en compte de plusieurs niveaux d'organisation et de plusieurs échelles de

temps et d'espace doit être intégrée dans les méthodes de modélisation, de calcul, de simulation et d'apprentissage. Le contexte de l'explosion des données met également au cœur des enjeux les problématiques de représentation des données, de processus et connaissances multi-échelles, d'interopérabilité, de planification et de visualisation.

Interfaces entre thèmes : Au-delà du caractère transversal du thème 7, des interfaces entre thèmes porteuses de défis scientifiques en biologie et écologie prédictive ont également été identifiées. De façon générale, ces interfaces répondent à des enjeux « d'intégration verticale » des organismes vivants mais également à des enjeux « d'intégration horizontale », les organismes étant dans ce contexte un élément parmi d'autres du système.

Les attendus pour chaque groupe de travail ont été :

- La précision du périmètre du thème ainsi que les enjeux cognitifs et finalisés ;
- Les avancées scientifiques majeures, quelques « success stories » des équipes de l'Institut, et les priorités scientifiques et méthodologiques pour les années à venir ;
- Le dispositif Inra : compétences, animations ainsi que les outils et infrastructures ;
- Les principaux partenariats académiques en cours et à développer aux échelles nationale, européenne et internationale ainsi que des éléments de vision concernant le potentiel de partenariat socio-économique.

Un séminaire regroupant l'ensemble des participants à la prospective a été organisé les 27 et 28 Février 2018 dont les objectifs ont été de :

- Faire un point d'étape sur l'avancement en mobilisant l'ensemble des acteurs ;
- Partager les réflexions des thèmes, identifier les questions en suspens, les liens/interfaces/recouvrements entre les thèmes, les points de blocages éventuels ;
- Faire émerger/consolider des problématiques/questions transversales ou aux interfaces entre les thèmes ;
- Définir le travail à réaliser pour suivre la feuille de route et fournir les livrables de la prospective.

Principales étapes de la prospective

- ⇒ 13/03/2017 : Lettre de mission
- ⇒ 10/05/2017 : Réunion de lancement avec le COPIL
- ⇒ Mai à Octobre 2017 : Travail du COPIL : cartographie, identification des thèmes et constitution des groupes de travail (GT), en interaction avec les départements
- ⇒ Octobre 2017 à Février 2018 : Travail des thèmes par les GT
- ⇒ 27-28/02/2018 : Séminaire de bilan à mi-parcours entre le COPIL et l'ensemble des membres des GT
- ⇒ Février à Mai : Travail des thèmes par les GT
- ⇒ Mai à Décembre 2018 : Travail du COPIL sur la finalisation du rapport des GT, le plan d'actions et la structure du rapport final
- ⇒ Février 2019 : Rapport final

La suite de ce rapport est divisée en deux parties.

La première partie « Défis Scientifiques » propose une synthèse des périmètres, enjeux, priorités scientifiques et verrous méthodologiques pour chacun des thèmes travaillés dans le cadre de la prospective. Les enjeux d'intégration et d'interfaces entre les thèmes sont également précisés. Cette synthèse a été réalisée à partir d'un document d'une 100aines de pages issu des travaux des différents groupes (disponible pour l'interne sur demande).

La seconde partie « Plan d'action » propose des recommandations pour favoriser le développement et le rayonnement des recherches de l'Inra dans le domaine des approches prédictives en biologie et écologie.

2 – DEFIS SCIENTIFIQUES

Le choix a été fait d'intégrer tout ce qui est indissociable des approches prédictives, y compris certains points concernant des verrous méthodologiques qui en dépassent le cadre strict, par exemple concernant l'acquisition des données.

2.1 – Périmètre, enjeux et priorités des 7 thèmes

2.1.1 – Régulations et compromis métaboliques

Le métabolisme (ensemble des processus biochimiques de transformation de matière et d'énergie par les organismes vivants) est extrêmement complexe avec ses milliers de réactions organisées en voies fortement interconnectées et souvent cycliques. De multiples signaux (hormones, facteurs de croissance ...) sont continuellement intégrés pour orchestrer la réponse métabolique aux modifications aiguës ou chroniques de l'environnement. Le métabolisme est avant tout cellulaire mais, chez les organismes multicellulaires, les échelles du tissu, de l'organe et de l'organisme entier doivent être prises en compte en raison de la fréquente spécialisation métabolique des cellules et des échanges de signaux et de métabolites entre cellules, tissus et organes. Les échelles d'études s'étendent donc du moléculaire (collecte et transduction de signaux, expression des gènes, relations structure-fonction des protéines enzymatiques) et supramoléculaire (interactions protéines-substrats) à l'organisme entier (notion de flux). Ces mécanismes permettent la plasticité phénotypique mais peuvent aussi conduire à des dérives métaboliques. L'association ou la dissociation des molécules sous l'effet des signaux assurent un transport rapide de l'information, mais l'organisation spatiale et temporelle inclut diverses échelles de temps de réponse selon les différents éléments cellulaires considérés. Il est donc crucial de considérer cette organisation temporelle face aux perturbations qui conduisent à la mise en place de compromis permettant la flexibilité métabolique.

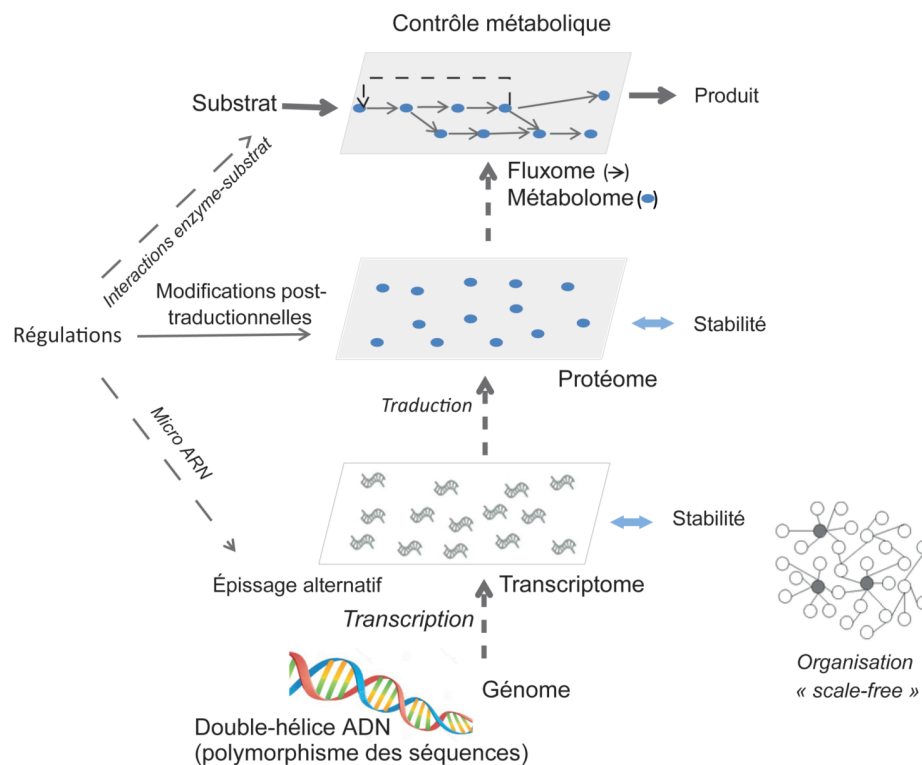


Figure 1 : Organisation du réseau biologique décrivant le métabolisme cellulaire ayant pour finalité la genèse de produits à partir de substrats (d'après⁶). L'organisation à l'échelle des transcrits est dite "scale-free" car la plupart des molécules ne partagent que peu d'arêtes entre elles, tandis qu'un petit nombre d'entre elles est à l'inverse très connecté et relie ainsi le reste des sommets du réseau.

Enjeux

- **Comprendre les mécanismes et leurs dynamiques.** Il s'agit de décrire et hiérarchiser les acteurs du réseau métabolique et définir les principaux éléments fonctionnels de régulation. L'ambition est de développer, pour divers organismes complexes, la reconstruction de réseaux métaboliques comme réalisé pour une cellule bactérienne⁷ et très récemment chez l'humain⁸. La finalité est de mieux appréhender la modulation des régulateurs clés par l'environnement et son influence sur les propriétés du métabolisme, la mise en place d'éventuels compromis et, par extrapolation, la réponse cellulaire, tissulaire ou de l'organisme aux fluctuations de l'environnement.
- **Prédire les réponses phénotypiques et identifier des biomarqueurs.** L'extension des possibilités de phénotypage au niveau de la cellule unique ou d'un ensemble hétérogène de cellules au sein d'un tissu particulier permet, sous réserve d'appliquer des méthodes de prédiction prenant en compte une grande masse d'informations (méthodes d'apprentissage automatisé⁹), de définir des molécules prédictives de caractères d'intérêt

⁶ Taghipoor M., Lemosquet S., Van Milgen J., Siegel A., Sauvant D., Gondret F. 2016. Modélisation de la flexibilité métabolique : vers une meilleure compréhension des capacités adaptatives de l'animal. *Prod. Anim* : 29 (3), 201-216

⁷ Karr et al., 2012. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*, 150:389-40

⁸ Brunk E. et al., 2018. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol.*, 36(3):272-281. doi: 10.1038/nbt.4072

⁹ LeCun Y., Bengio Y., and Hinton G. 2015. Deep learning. *Nature* 521, 436-444

globaux et complexes. Ces biomarqueurs permettent de progresser dans la connaissance de la variabilité interindividuelle et ouvrent des perspectives d'interventions adaptées.

- **Optimiser ou créer des voies métaboliques.** L'ambition est le façonnage sur mesure de catalyseurs enzymatiques permettant la production efficace de composés d'intérêt dans des systèmes cellulaires. Elle nécessite de repérer les activités recherchées par des approches de (méta)génomique fonctionnelle, d'identifier les catalyseurs à l'aide d'algorithmes de rétrosynthèse, de comprendre les mécanismes de la cascade catalytique grâce à la modélisation moléculaire multi-échelle et le design computationnel de protéines. Des assemblages d'enzymes peuvent ensuite être créés par modélisation métabolique *in silico* pour sélectionner les plus performants. Enfin, les possibles verrous à lever pour optimiser le fonctionnement des nouveaux réseaux métaboliques au sein de la cellule hôte peuvent être identifiés par des analyses fonctionnelles (métabolomique pour la quantité de substrats et de produits et fluxomique pour la vitesse des réactions).

Priorités scientifiques

- **Identifier les interactions entre acteurs.** Comprendre la façon dont les acteurs du métabolisme agissent les uns par rapport aux autres dans leur ensemble doit être l'objet d'une approche systémique afin de déduire les éléments clés de contrôle. Prendre en compte les combinatoires de signaux (multi-stress) est également nécessaire pour prédire la reprogrammation des réseaux d'acteurs qui permet la flexibilité métabolique mais conduit aussi parfois à des dérives du système.
- **Passer du statique au dynamique.** La notion de flux d'échanges entre entités du réseau (machineries moléculaires, RNAs, protéines, métabolites, etc.), de réponse à un besoin ou une perturbation à différents niveaux d'organisation implique d'être capable d'observer et d'intégrer des phénomènes biologiques ayant lieu à des échelles de temps très différentes. Par exemple, alors que les protéines sont synthétisées en quelques minutes, nombreuses sont celles qui peuvent avoir une durée de vie de plusieurs semaines et dont la concentration peut varier en fonction de certains signaux.
- **Intégrer les échelles.** L'intégration doit être abordée du point de vue phénoménologique (corrélation entre échelles) pour répondre à l'enjeu de prédiction des réponses phénotypiques et d'identification de biomarqueurs. Elle doit également être abordée du point de vue mécaniste par la compréhension des montées en échelle et des rétroactions. L'intégration doit prendre en compte les dynamiques cellulaires et les compromis entre fonctions pour faire le lien entre l'échelle de la cellule individuelle et l'échelle macroscopique (réponse physiologique). Un exemple est le compromis entre croissance et multiplication des cellules. Le volume cellulaire sous l'action de l'accumulation des produits du métabolisme est un des facteurs déclenchant des bascules dans les mécanismes du développement permettant de coupler des échelles de temps relativement courtes (régulations métaboliques) avec des échelles de temps plus longues (développement tissulaire). Enfin, l'hétérogénéité tissulaire doit être prise en compte, le défi étant de reconstruire des réseaux métaboliques qui soient cellules/tissus spécifiques (e.g., en toxicologie).

Verrous méthodologiques

- **Concernant les données,** un premier défi est l'acquisition de données quantitatives à des points de mesures correspondant aux transitions métaboliques. Un deuxième défi est de compléter les données acquises sur un tissu hétérogène ou un ensemble de cellules en culture, par des données sur une seule cellule (single cell) et à l'échelle subcellulaire. Un troisième défi concerne l'annotation. En effet, alors que l'annotation des transcriptomes

et protéomes est aujourd'hui largement automatisée, elle reste un défi majeur en métabolomique. Au niveau national, l'infrastructure MetaboHUB¹⁰ dynamise le développement d'outils d'aide à l'annotation, un effort qui devra être maintenu. Enfin pour l'apprentissage automatique, le verrou majeur actuel est le manque de données temporelles sur des temps très courts, après par exemple, une perturbation du système.

➤ **Concernant les modèles**, les défis s'articulent autour :

- Du développement et de la simulation de modèles multi-échelles utilisant différents formalismes. L'objectif est d'établir les phénotypes microscopiques ou macroscopiques intervenant dans le maintien ou la modification des propriétés du réseau métabolique étudié en intégrant les différents niveaux d'échelle de l'information biologique pour comprendre l'origine des mécanismes de compensation.
- De la rétro-ingénierie pour concevoir de nouvelles voies métaboliques et usines cellulaires. L'objectif est de développer de nouvelles méthodes pour concevoir de nouvelles voies métaboliques, et optimiser le fonctionnement de ces voies dans le contexte de l'organisme de production. Des avancées méthodologiques sont ainsi nécessaires (i) pour proposer des approches robustes de design computationnel d'enzymes dotées de propriétés adaptées et/ou de nouvelles voies métaboliques et (ii) pour permettre l'optimisation conjointe du bioprocédé et de l'organisme de production intégrant le(s) nouvelle(s) voie(s) métabolique(s).
- Du transfert des modèles cellulaires développés sur des organismes "modèles" aux organismes d'intérêt agronomique. Dans la majorité des cas, les modèles et plus généralement les méthodes de modélisation, quelles qu'elles soient, sont mises au point et validées sur des organismes "modèles", comme la bactérie *Escherichia coli* ou encore la plante *Arabidopsis thaliana*. Le transfert et la mise en œuvre de ces méthodes sur des espèces non modèles constitue un enjeu central pour l'Inra. C'est une démarche traditionnelle pour l'annotation des génomes. Néanmoins, dans le cas de la modélisation, l'enjeu va bien au-delà car le modèle doit être représentatif du fonctionnement de l'objet biologique modélisé.

2.1.2 – Morphogénèse et développement des organismes

Une des questions fondamentales de la biologie est de comprendre comment des individus avec des structures et formes spécifiques se développent. Dans ce contexte, deux questions centrales se posent : (i) quels sont les liens entre la régulation moléculaire, le comportement cellulaire et le développement de l'individu multicellulaire et (ii) comment ces liens sont-ils modulés par l'environnement ?

Au cours des dernières décennies, beaucoup d'efforts ont été concentrés sur la régulation moléculaire du développement : séquençage des génomes, approches génétiques, profils d'expression de milliers de gènes. Malgré des progrès spectaculaires, la complexité de ces données demeure un obstacle majeur à la compréhension des mécanismes qui relient la régulation génétique et le développement de l'architecture des plantes et des animaux. Par ailleurs, le paradigme du génome comme programme rigide qui orchestre le développement a atteint ses limites¹¹. Il est de plus en plus évident qu'un processus complexe comme le

¹⁰ <http://www.metabohub.fr>

¹¹ Ross and Arkin. 2009. Complex systems: From chemistry to systems biology. PNAS 106 : 6433-6434

Boyle et al. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169 : 1177-1186

développement ne peut être compris qu'en tenant compte de l'ensemble des échelles : des molécules interagissent pour former des cellules, les cellules forment des tissus, qui s'assemblent en organes et finalement en organismes entiers. Chaque échelle est caractérisée par des aspects spécifiques. A l'échelle subcellulaire, par exemple, les réseaux d'interaction moléculaire dominant, tandis qu'au niveau des tissus des interactions physiques et signalisations chimiques contrôlent les directions et vitesses de croissance ou la migration cellulaire.

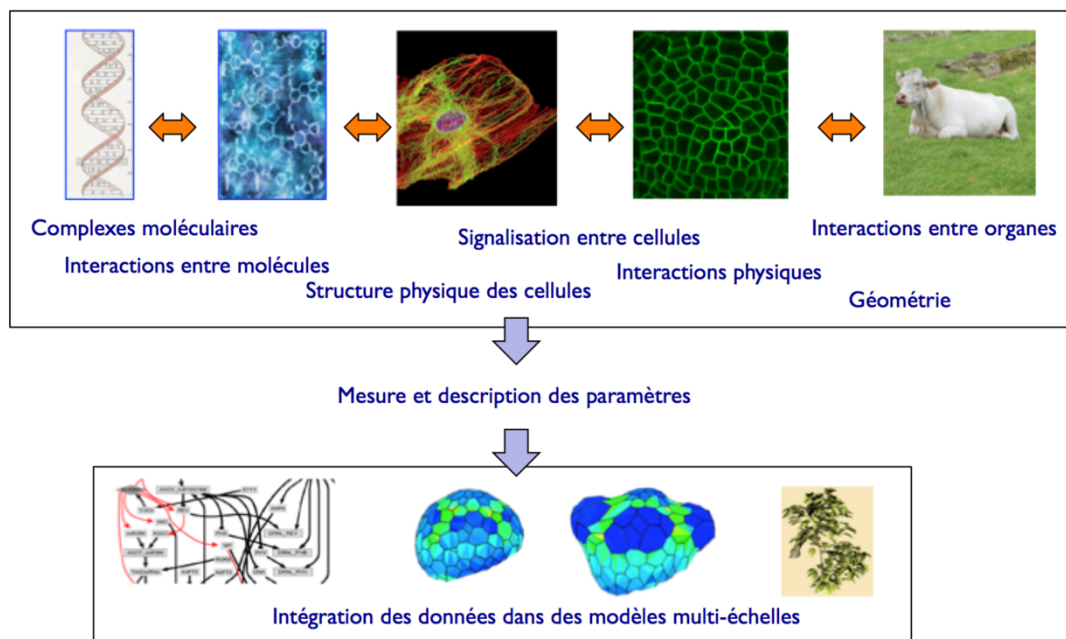


Figure 2 : Analyse multi-échelle des systèmes vivants. Le paradigme classique considère le programme génétique comme déterminant dominant du phénotype. Actuellement on considère les systèmes vivants comme des systèmes complexes, qu'il faut analyser à des multiples échelles. Chaque échelle se caractérise par des processus et paramètres spécifiques. Les données peuvent être intégrées dans des modèles qui permettent de formuler des hypothèses complexes. Un des enjeux est de produire des modèles mécanistes qui couvrent plusieurs échelles et un ensemble de paramètres.

Enjeux et priorités

Plusieurs projets sont en cours à l'Inra dans le domaine animal et végétal¹². Ces projets, souvent très fondamentaux, concernent des organismes modèles et des espèces d'intérêt agronomique. Ils ambitionnent de produire des modèles quantitatifs du développement de l'organisme entier, capables d'exprimer des propriétés biochimiques et biophysiques avec, au besoin, une résolution subcellulaire, et de reproduire une dynamique *in silico*. A terme, l'objectif est de produire des modèles robustes capables de prédire la forme et l'architecture d'un grand nombre d'organismes et la reproductibilité de leur développement dans des environnements divers et fluctuants.

¹² Verger S. et al. 2018. A tension-adhesion feedback loop in plant epidermis. *Elife*. 23;7. pii: e34460. doi: 10.7554/eLife.34460; Eloy C. et al. 2017. Wind loads and competition for light sculpt trees into self-similar structures. *Nat Commun*.8(1):1014. doi: 10.1038/s41467-017-00995-6; Robles et al. 2017. Does maternal size, nutrition and metabolic status affect offspring production traits in domestic species? *Animal Reproduction* DOI: 10.21451/1984-3143-AR996

- **Les mécanismes qui gouvernent la morphogénèse chez les plantes.** L'objectif principal est d'obtenir une compréhension fine des mécanismes contrôlant la forme et l'architecture. Les questions de recherche prioritaire visent à aborder la morphogénèse sous plusieurs angles :
 - (i) Du gène à la forme. Quel est le lien entre la régulation génétique et la géométrie/architecture de la plante ? Comment l'action moléculaire est-elle traduite en changement de géométrie au niveau des tissus et organes ? Pour répondre à ces questions des approches quantitatives à plusieurs échelles sont nécessaires. Il s'agit en plus d'intégrer des propriétés de nature très différente : biochimique, géométrique, biophysique.
 - (ii) Identification et rôle des rétroactions (« feedbacks »). La morphogénèse n'est pas juste le résultat d'une régulation moléculaire qui dicte la croissance des cellules et les changements aux échelles cellulaires et tissulaires agissent à leur tour sur le réseau de gènes. Dans ce contexte, la compréhension des mécanismes et des implications des rétroactions est fondamentale.
 - (iii) Reproductibilité et plasticité des formes. Comment les organes savent quand s'arrêter de grandir ? Qu'est ce qui définit la robustesse des phénotypes dans différents environnements ? Réciproquement, en quoi la plasticité des phénotypes permet l'adaptation ?

- **Le lien entre le développement précoce et la construction des phénotypes tardifs.** L'environnement influence le développement des organismes, et les événements intervenus au cours de l'embryogénèse précoce ont un impact important sur les phénotypes adultes. De plus, les futurs gamètes se différencient au cours de ces phases précoces et toute altération de leur différenciation pourra avoir des effets sur les générations suivantes. Dans ce contexte, l'objectif est d'améliorer le niveau de prédiction des phénotypes tardifs à partir de génotypes et phénotypes précoces, dans des environnements divers et fluctuants. Il s'agit également d'identifier des fenêtres d'action pendant lesquelles il va être possible d'appliquer différents stimuli pour moduler l'expression du génome et "programmer" les phénotypes à long terme, voire sur plusieurs générations. Nos connaissances des mécanismes impliqués dans la définition du phénotype tardif restent peu comprises et il apparaît pertinent de proposer une approche descriptive, sans *a priori*. Le but est d'établir de fortes corrélations entre développement précoce et phénotypes tardifs, formalisées dans des modèles statistiques. Ces modèles devraient déjà mener, notamment, à des améliorations de la prédiction des valeurs génétiques. Ces analyses permettront dans un deuxième temps une meilleure compréhension et modélisation des mécanismes impliqués. Ce thème, bien que centré sur les processus de développement, rejoint les problématiques et priorités du thème sur la prédiction des phénotypes.

Verrous méthodologiques

- **Les défis méthodologiques sur l'acquisition de données quantitatives et multi-échelles du développement** concernent :
 - L'analyse de la croissance en 3D à une résolution cellulaire de l'ensemble des tissus. Des méthodes de segmentation et d'identification de lignages cellulaires doivent être améliorées et des méthodes d'observation *in vivo* des tissus profonds restent à mettre au point, notamment pour les organes épais ou opaques (e.g., nanotomographie RX en synchrotron) ;

- Les outils pour mesurer les propriétés physiques qui sont prometteurs mais restent encore limités ;
 - La prise en compte des dynamiques des gradients moléculaires dans un système déformable et la combinaison avec la cinématique de croissance ;
 - La caractérisation omique à l'échelle de la cellule-unique et l'organisation spatiale du génome au cours du développement, et les méthodes de mesures à haut débit (4D) de la croissance de l'individu qui sont encore limitées à des petits animaux ou plantes.
- **Le traitement des données brutes issues de l'imagerie.** Les techniques d'analyse des images n'ont pas encore atteint le degré de maturité des analyses « omiques ». L'extraction des informations pertinentes pour la caractérisation et la modélisation de la morphogenèse ainsi que pour la prédiction requiert la mise en œuvre de chaînes de traitement et analyse d'images par lesquelles les objets d'intérêt (e.g., les cellules) sont détectés et quantifiés au travers d'ensembles de paramètres. Ainsi, il s'agit de (i) renforcer la robustesse des pipelines algorithmiques pour le filtrage, le recalage et la segmentation des images ainsi que pour la reconstruction des lignages cellulaires, dans la perspective de leur automatisation et de leur déploiement systématique sur de grands jeux de données et (ii) de dépasser le cadre des descripteurs classiquement utilisés en analyse d'images avec l'introduction de nouveaux descripteurs spatio-temporels à l'échelle supracellulaire.
- **En termes de modélisation,** le défi reste la production de modèles quantitatifs et mécanistes du développement de l'organisme entier, capables d'exprimer des propriétés biochimiques et biophysiques avec, au besoin, une résolution subcellulaire, et de reproduire une dynamique *in silico*. Il s'agit d'un défi majeur, puisque beaucoup d'outils informatiques nécessaires à la conception de ce type de modèles multi-échelles sont encore à développer.

2.1.3 – Modélisation et prédiction des phénotypes

Dans un contexte de fortes attentes liées aux changements globaux, il apparaît indispensable de progresser dans nos capacités de prédiction des phénotypes d'intérêt, aussi bien au fil des générations qu'au cours de la vie, y compris très précocement, afin d'améliorer et/ou renouveler les outils pour la sélection et plus largement l'aide à la décision.

On distingue deux grandes catégories de phénotypes : les phénotypes finaux et les phénotypes intermédiaires, qui sont des sources d'informations explicatives ou des proxys utiles pour prédire les phénotypes finaux. En effet, les déterminants des phénotypes finaux sont souvent complexes et il est informatif de les décomposer en phénotypes intermédiaires pour mieux comprendre les mécanismes sous-jacents, puis de combiner les phénotypes intermédiaires de façon pertinente et opérationnelle pour *in fine* prédire le caractère d'intérêt. Des phénotypes finaux correspondant à des objectifs de sélection identifiés comme majeurs dans le contexte actuel de changement global sont par exemple : (i) le niveau de production et la qualité des produits, avec la notion de régularité dans le temps, (ii) la santé, la compétence immunitaire et la résistance aux agents pathogènes (et bien-être chez les animaux) en lien avec la réduction des intrants comme les produits phytosanitaires chez les plantes, les anti-infectieux et hormones chez les animaux, (iii) l'adaptation à des environnements variables et sous contraintes (e.g., résilience face aux stress de température, à la sécheresse, à des régimes alimentaires sub-optimaux), ou encore, (iv) l'impact environnemental (e.g., émission de méthane pour les animaux ou la consommation en eau pour les plantes).

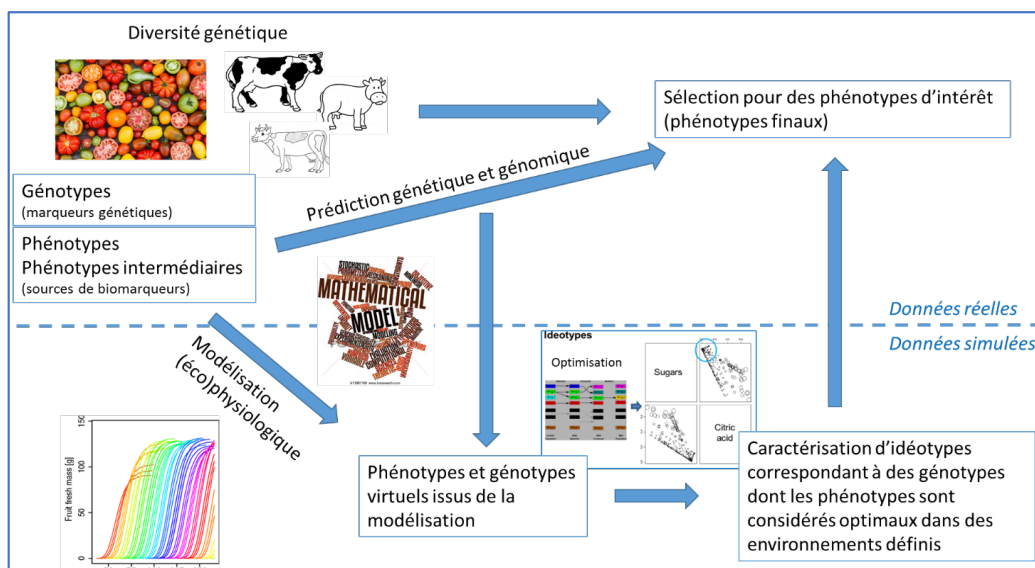


Figure 3 : Complémentarité des approches basées sur des données réelles (génotypes, phénotypes) et simulées (modèles, optimisations) pour sélectionner des phénotypes d'intérêt.

Enjeux

L'enjeu est de coupler les données de phénotypage de plus en plus massives et diversifiées avec les informations de génotypes (et épigénotypes) et d'environnement (y compris écosystèmes microbiens commensaux), afin de comprendre, modéliser et prédire la construction et la plasticité des phénotypes d'intérêt (figure 3). On s'attachera notamment à identifier les principaux processus impliqués (régulation génique, voies métaboliques importantes pour le phénotype, ou flux physiques à l'échelle de l'organisme) et leur contrôle par le génome, l'environnement et les pratiques d'élevage ou de culture. La modélisation (éco)physiologique permettra d'intégrer ces connaissances en vue de la prédiction des phénotypes dans un contexte environnemental et de pratiques donné¹³. Enfin, apporter de l'information biologique dans les modèles de prédiction génétique pour améliorer la précision, comme par exemple en sélection génomique, est également un enjeu important qui requiert le développement d'annotations fonctionnelles des génomes¹⁴.

D'un point de vue plus finalisé, les attendus principaux sont (i) l'identification de marqueurs génétiques (et épigénétiques) prédictifs des caractères d'intérêt dans des environnements variés, (ii) l'identification de nouveaux biomarqueurs utilisables comme des prédicteurs phénotypiques, et (iii) la conception et la sélection d'idéotypes ou génotypes optimaux¹⁵ adaptés aux conditions environnementales, aux modes de culture ou d'élevage, et aux produits attendus (rendement, qualité, faible sensibilité aux maladies, etc.).

¹³ Dingkuhn et al. 2017 Crop-model assisted phenomics and genome-wide association study for climate adaptation of indica rice. *J Exp Bot.*, 68, 4369-4388. doi: 10.1093/jxb/erx249

¹⁴ MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, Schrooten C, Hayes BJ, Goddard ME. 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics.* 17:144

Giuffra E, Tuggle CK, The FAANG Consortium. 2018. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annu Rev Anim Biosci.* doi: 10.1146/annurev-animal-020518-114913

¹⁵ Picheny V, Casadebaig P, Trépos R, Faivre R, Da Silva D, Vincourt P, Costes E. 2017. Using numerical plant models and phenotypic correlation space to design achievable ideotypes. *Plant, Cell & Environment*, 40, 1926-1939. doi: 10.1111/pce.13001

Priorités scientifiques

- **Décomposer les phénotypes d'intérêt pour exploiter des données multi-échelles.** Avec les récentes avancées scientifiques, il est aujourd'hui possible de générer des données à toutes les échelles biologiques, de la molécule à l'écosystème. Parallèlement, les nouveaux défis que doivent relever les systèmes de production impliquent d'être capable de prédire des caractères de plus en plus complexes (e.g., résilience, robustesse, efficacité). Une priorité est de décomposer le phénotype d'intérêt en phénotypes intermédiaires afin de mieux comprendre les mécanismes sous-jacents et de combiner ces derniers de façon pertinente et opérationnelle pour prédire le caractère d'intérêt.
- **Prédiction génétique et effet des interactions.** Prédire 100% de la valeur génotypique additive de génotypes distincts n'est pas atteint actuellement et, dans ce cadre, trois types de leviers doivent être examinés : (i) les données nécessaires pour bien estimer, puis prédire le phénomène cible, (ii) les modèles génétiques, qui actuellement ne permettent pas de tenir compte de tous les facteurs identifiés, et (iii) les conditions environnementales dans lesquelles la contribution de ces facteurs à la variabilité phénotypique est importante pour préciser la prédiction génétique. Une autre priorité concerne la prédiction de la performance d'une population hétérogène, en fonction de facteurs de variations divers. En effet, le concept phare de la sélection actuelle vise à optimiser la performance de la population par l'optimisation des performances individuelles des différents génotypes pris séparément. Aujourd'hui, il s'agit également optimiser la performance d'un "agro-écosystème" c'est-à-dire d'un ensemble de génotypes pris conjointement, et donc avec leurs interactions (sociales, de compétition, de symbiose, etc...), dans le temps et dans l'espace de réalisation des performances.
- **Prédiction et optimisation des phénotypes en lien avec la diversité des environnements et des systèmes.** Ces priorités concernent, d'une part la prédiction du phénotype de l'individu sous contrôles génétique et environnemental et d'autre part, la conception des idéotypes ou génotypes optimaux et le développement des approches de sélection assistée par modélisation pour tenter de les obtenir. Les idéotypes sont définis comme des génotypes ayant un phénotype optimal en termes de qualité, rendement, résistance aux bioagresseurs, etc., dans un environnement de culture ou d'élevage donné. Les modèles (empiriques ou mécanistes) peuvent être utilisés pour simuler le comportement de millions de génotypes virtuels définis par les valeurs qu'ils prennent pour les paramètres génétiques du modèle et des méthodes de recherche opérationnelle peuvent être utilisées pour retenir les idéotypes intéressants. La limite actuelle de cette approche est le manque de réalisme des idéotypes qui ne peuvent pas être obtenus parce qu'ils enfreignent des contraintes physiologiques (processus liés) et génétiques (pléiotropie, épistasie et liaisons physiques entre gènes).

Verrous méthodologiques

- **La diversification et l'automatisation des phénotypages** avec le développement de méthodes d'observation (télémétrie), d'analyses non destructives/invasives *in vivo*, performantes pour des mesures longitudinales, en unités expérimentales et sur le terrain.

Dans ce domaine, l'infrastructure Emphasis-France¹⁶ (et son prolongement à l'Europe) constitue un appui déterminant dans le domaine végétal.

- **Le développement de méthodes de phénotypage assistées par modèle.** Dans un grand nombre de situations de phénotypage, les traits intéressants tout comme les variables de l'environnement, ne sont pas toujours directement mesurables, mais doivent être estimés *via* un modèle. Par exemple, on mesure facilement la transpiration d'un organe, mais pas le trait contrôlé génétiquement (conductance de la surface de cet organe à l'eau) qui peut être estimé avec un modèle à partir des données de transpiration.
- **Le développement du phénotypage *in vitro* et *ex vivo*** (remplacement des animaux, règle des 3R), avec la mise en place de banques d'organoïdes issus d'individus de génotypes connus, de technologies « organ on chip » utilisant la microfluidique ou encore, la mise en place de criblage phénotypique *in vitro/ex vivo* qui prédise des phénotypes *in vivo*.
- **Le couplage des modèles en génétique statistique et des modèles de simulation.** Peu de modèles de simulation prennent en compte les polymorphismes génétiques.
- **Le développement des liens entre les modèles de prédiction des phénotypes et les modèles de décision des actes techniques** avec pour champ d'application l'agriculture de précision ou l'agriculture numérique, mais aussi l'agroécologie.

2.1.4 – Holobiontes : connaissances et modélisation des interactions entre systèmes microbiens et hôtes

Il est désormais acquis qu'un animal ou une plante ne peut se développer pleinement ou même parfois vivre sans les multiples microorganismes qui l'habitent. De cette constatation est né le concept d'«holobionte », qui désigne une entité formée d'un macroorganisme hôte et des écosystèmes microbiens associés : procaryotes (bactéries et archées), eucaryotes (protistes, champignons, etc.), bactériophages et virus. Le concept d'holobionte a été et reste encore sujet à débat¹⁷. Des exemples frappants¹⁸ ont conduit à une première vision de l'holobionte restreinte aux relations stables et héritables entre un hôte et ses microorganismes symbiotiques (symbiontes). Selon ce point de vue, l'hologénome, qui est l'ensemble des génomes des organismes qui constituent l'holobionte, est donc transmis verticalement à la descendance et constitue une unité de sélection. Une version élargie¹⁹ a été proposée, selon laquelle « les symbiontes microbiens peuvent être permanents ou non permanents, transmis verticalement ou horizontalement et selon le contexte, être nuisibles, inoffensifs ou utiles ». La figure 4 illustre bien cette définition qui englobe la définition précédente. C'est cette définition élargie qui est adoptée ci-après.

¹⁶ <https://www.phenome-emphasis.fr>

¹⁷ Moran NA, Sloan DB. 2015. The Hologenome Concept: Helpful or Hollow? PLoS Biol. 2015 Dec 4; 13(12):e1002311. doi: 10.1371

¹⁸ Selosse MA. 2016. Au-delà de l'organisme : l'holobionte. Pour la Science N°469, Nov. 2016

¹⁹ Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF, Bosch TC, Cryan JF, Gilbert SF, Goodnight CJ, Lloyd EA, Sapp J, Vandenkoornhuysen P, Zilber-Rosenberg I, Rosenberg E, Bordenstein SR. 2016. Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes. mSystems. 2016 Mar 29;1(2). pii: e00028-16. eCollection 2016

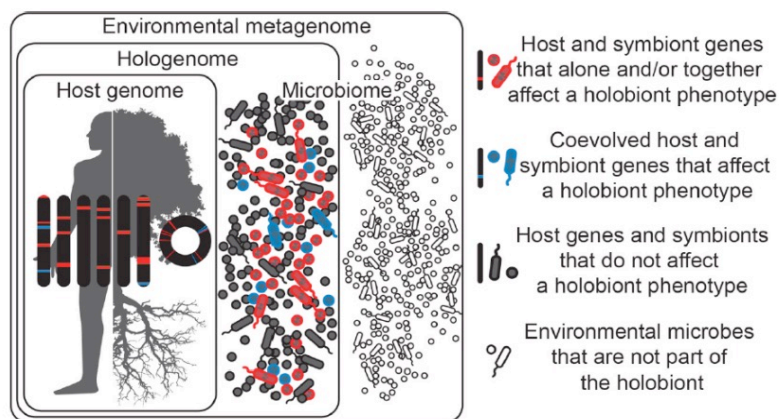


Figure 4: Le concept d'holobionte (Theis et al. 2016). *Holobionts are entities comprised of the host and all of its symbiotic microbes, including those which affect the holobiont's phenotype and have coevolved with the host (blue), those which affect the holobiont's phenotype but have not coevolved with the host (red), and those which do not affect the holobiont's phenotype at all (gray). Microbes may be transmitted vertically or horizontally, may be acquired from the environment, and can be constant or inconstant in the host. Therefore, holobiont phenotypes can change in time and space as microbes come into and out of the holobiont. Microbes in the environment are not part of the holobiont (white). Hologenomes then encompass the genomes of the host and all of its microbes at any given time point, with individual genomes and genes falling into the same three functional categories of blue, red, and gray. Holobionts and hologenomes are entities, whereas coevolution or the evolution of host-symbiont interactions are processes.*

Enjeux

- **Développer une vision systémique de l'hôte et son microbiote.** Le concept d'holobionte permet de dépasser les limites de celui d'organisme, en introduisant une vision de l'hôte et de ses symbiotes microbiens comme parties d'un système, dans lequel les interactions comptent autant que les constituants. Il mobilise, sans les opposer, des connaissances et des outils issus de l'écologie, la génétique et la physiologie, tout en rendant nécessaire le fait de les revisiter pour formuler un nouveau cadre conceptuel, des modèles ainsi que des outils expérimentaux, mathématiques et informatiques adaptés.
- **Comprendre et gérer les interactions hôte-microbiote.** Une approche "système" de l'hôte et de ses microbiotes en interaction est indispensable pour parvenir à une meilleure compréhension et une gestion des interactions hôte-microbiote dans l'environnement et dans des contextes agronomiques (cultures, élevages) ainsi qu'en nutrition et santé humaine. Elle offre ainsi des leviers pour répondre à des enjeux majeurs. Citons à titre d'exemples des enjeux environnementaux, tels que l'impact et la maîtrise des interactions entre les plantes ou les arbres et leurs microbiotes dans le cycle des nutriments, dans la production primaire, dans la régulation du climat, ou celui des émissions de méthane par les ruminants ; ou encore des enjeux d'agriculture et d'alimentation saines et durables, au sein desquels la compréhension et la maîtrise des holobiontes pourraient contribuer à la diminution des intrants utilisés pour la performance, la santé et la nutrition.

Priorités scientifiques

- **Comprendre le fonctionnement du microbiote et le modéliser.** Dans le cas de l'homme ou de l'animal, le microbiote (notamment intestinal) a été qualifié d'organe oublié. Pour conserver cette analogie en l'étendant aux plantes et à leurs cortèges microbiens, cette priorité concerne l'acquisition et la formalisation de connaissances sur le fonctionnement de ces organes oubliés que sont les écosystèmes microbiens complexes en interaction

avec leurs hôtes, qu'ils soient fermés ou semi-fermés c'est-à-dire peu exposés aux microorganismes environnementaux (par exemple, le microbiote intestinal hors période périnatale) ou ouverts, systèmes qui sont en constante interaction avec les microorganismes environnementaux (par exemple, la phyllosphère des plantes ou le microbiote de la peau).

- **Comprendre le fonctionnement des holobiontes et développer des holobiontes virtuels.** Cette priorité concerne la compréhension des interactions entre hôtes et microbiotes et les liens avec le fonctionnement et le phénotype de l'holobionte (e.g., efficacité alimentaire, performance, santé et résistance aux maladies). Il s'agira de connaître les traits du microbiote qui impactent l'hôte et vice-versa tout au long de la vie, d'identifier les boucles de rétroaction entre l'hôte et son microbiote et de mieux comprendre les phénomènes de coévolution, symbiose et d'évolution sous contraintes environnementales. Un enjeu important sera de dépasser les corrélations pour établir les liens de causalité entre état de la symbiose microbiote-hôte et phénotype. Outre l'interface mathématiques-biologie nécessaire à cette priorité, l'objectif sera de produire des holobiontes virtuels permettant de mieux identifier les leviers d'action (relevant de l'hôte, du microbiote, de leurs environnements) pour la performance et la santé, et leurs effets potentiels.
- **Évaluer et prédire l'état de l'holobionte.** L'objectif est d'évaluer l'état, la qualité de la symbiose au cours du temps et de façon rapide au moyen de biomarqueurs informatifs et pertinents. Il s'agira de surveiller le fonctionnement « optimal » de l'holobionte dans la durée et de prédire précocement les changements qui pourraient conduire à son dysfonctionnement. Cette priorité implique d'être à même de définir ce qu'est un fonctionnement normal, optimal ou dysfonctionnel. L'une des questions dans ce domaine est de déterminer la puissance prédictive et diagnostique de biomarqueurs de l'hôte, du microbiote et de leurs combinaisons.
- **Identifier les leviers et restaurer ou modifier le fonctionnement des holobiontes.** Le développement des connaissances sur le fonctionnement des holobiontes et le développement d'outils diagnostiques et pronostiques devraient conduire à identifier précocement les holobiontes à risque. Au-delà de cette identification, se pose la question de la préservation voire de la restauration d'une symbiose fonctionnelle et promotrice de performance, santé, etc. Cette priorité vise à développer des approches prédictives d'exploration et de simulation pour aider à concevoir des méthodes d'intervention et prédire celles qui seront les plus efficaces en prenant en compte l'ensemble de l'holobionte et ses externalités.

Verrous méthodologiques

- **Concernant les données**, il est crucial d'améliorer la qualité (en termes de précision, finesse de résolution) des données disponibles ainsi que les méthodes expérimentales de caractérisation et manipulation de la structure et des fonctions de communautés microbiennes complexes (microbiotes), et encore plus les microbiotes avec leurs hôtes. En parallèle, il faudra développer des méthodes de traitements bioinformatiques et statistiques pour l'analyse de données de très grande dimension et l'intégration de données hétérogènes. Une difficulté supplémentaire est liée au caractère non stationnaire, inhomogène et peu maîtrisable des communautés microbiennes qui rend délicat la capacité à reproduire les expérimentations. Enfin se pose la question de la formalisation des données, de leur comparaison entre systèmes et études.
- Dans le domaine des **mathématiques et de la modélisation** quelques-uns des verrous identifiés sont : (i) le changement d'échelle entre la modélisation d'un seul génome vers

celle d'une communauté de génomes (hologénome) intégrant sa variabilité aussi bien du point de vue des méthodes que de l'implémentation, (ii) la prise en compte des aspects dynamiques des interactions au sein de l'holobionte et de l'évolution de l'holobionte au cours du temps, alors que la plupart des méthodes d'analyse de données dans le domaine ignorent cet aspect, et (iii) la capacité à prédire l'effet d'interventions au sein de l'holobionte (par exemple modification de microbiote) ou *via* son environnement (par exemple l'alimentation) en intégrant la diversité de réponses des holobiontes.

2.1.5 – Toxicologie et écotoxicologie prédictive

Les études de toxicologie (au sens large englobant la microbiologie) consistent à caractériser le danger que des substances, ou groupes de substances, ou microorganismes présentent pour l'homme. L'écotoxicologie vise à estimer l'impact écologique des substances libérées par les activités humaines dans l'environnement. Les limites de ces démarches sont nombreuses : (i) pertinence des études animales pour évaluer des risques pour des populations humaines hétérogènes et le plus souvent exposées à des concentrations beaucoup plus faibles d'agents présumés toxiques ; (ii) études coûteuses et longues impliquant un déficit, voire une absence de données pour plusieurs milliers de substances auxquelles nous sommes susceptibles d'être exposés ; (iii) les expérimentations animales requièrent un nombre élevé d'individus et se heurtent aux recommandations européennes et à la pression sociétale ; (iv) la plupart des tests fournissent peu d'information sur les mécanismes d'action, essentiels à la compréhension des différences interspécifiques et interindividuelles de toxicité ; (v) incapacité des tests standardisés actuels à évaluer les effets écotoxicologiques aux niveaux d'organisations biologiques dont la protection est visée par les instances réglementaires ; (vi) manque de connaissances scientifiques sur de nombreuses espèces constitutives des écosystèmes d'intérêt et les biais liés à l'extrapolation des réponses produites par un organisme modèle à des espèces phylogénétiquement distantes ; (vii) complexité des processus éco-évolutifs en condition d'exposition chronique, qui dépendent notamment du nombre de générations exposées et de la sensibilité des espèces concernées.

Parallèlement aux études réglementaires, des travaux académiques portant sur les mécanismes d'actions d'un nombre limité de substances toxiques sont publiés. Ils reposent pour beaucoup d'entre eux sur des approches globales à haut débit (transcriptomique et métabolomique en particulier), et ont parfois recours à des modèles cellulaires. Cependant, ces travaux sont rarement pris en compte dans les démarches d'évaluation du risque en raison de protocoles inadaptés, de méthodes non normalisées, d'absence de démarche assurance-qualité vérifiable, de difficultés d'accès aux données brutes ou encore du manque de référentiels ou documents guides validés qui puissent permettre de tirer parti de ces données.

Enjeux

La toxicologie doit donc évoluer pour répondre aux défis de réduction des expérimentations animales, de la présence croissante de nouvelles molécules présentant une toxicité potentielle et donc au besoin d'accélérer le processus d'évaluation du risque, et également de la prise en compte et prédiction de la complexité de l'exposition dans sa diversité, dans son espace, dans sa durée, dans sa temporalité, dans ses cibles physiques et biologiques et dans sa quantité. Concernant l'écotoxicologie, les enjeux sont liés à la pression exercée par les activités humaines sur les écosystèmes naturels et la biodiversité, dont l'amplitude est sans précédent. La menace est particulièrement importante pour les écosystèmes aquatiques qui sont soumis à une contamination chimique très diversifiée et en perpétuelle augmentation,

et qui représentent souvent le réceptacle final de nombreuses substances. Le risque estimé pour ces milieux est significatif à l'échelle aussi bien locale que continentale. Ce problème concerne en particulier la question des effets biologiques à long-terme, branche actuellement émergente de l'écotoxicologie. L'objectif des approches prédictives en toxicologie et écotoxicologie sera donc d'employer des méthodes d'analyse de données afin de créer des modèles pouvant être utilisés pour prédire les effets toxiques de composés ou microorganismes sur la santé humaine, animale et sur l'environnement (figure 5).

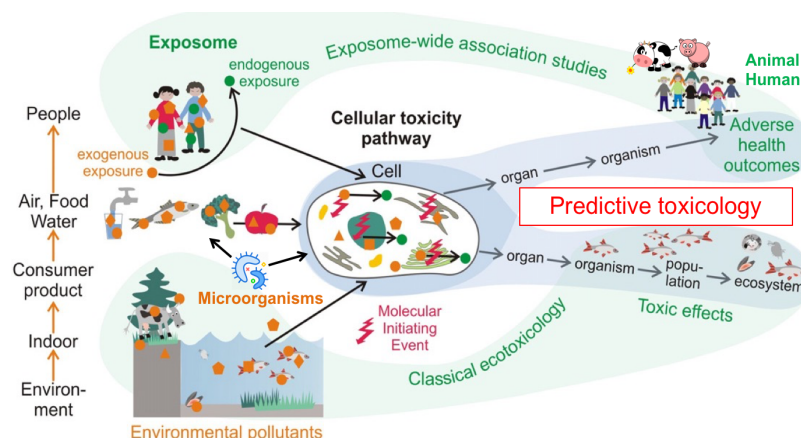


Figure 5 : Contexte et niveaux d'échelles de l'(éco)toxicologie prédictive (adapté de ²⁰)

Priorités scientifiques

- **Prédire l'exposition en lien avec les pratiques (culture et élevage).** L'objectif est de développer l'approche prédictive pour faire le lien entre pratiques agricoles (incluant antibiotiques) et toxicologie/microbiologie et ainsi aborder de nouvelles questions de toxicologie/microbiologie en lien avec le réchauffement climatique et les modifications des pratiques vers des systèmes durables. Dans ce cadre, il s'agit aussi de développer des approches d'analyse globale des risques et des analyses multi-critères.
- **Prédire l'exposition lors de la transformation et de la consommation.** Il s'agit par exemple d'intégrer les technologies « omiques » dans les modèles d'exposition au danger microbien, de prédire l'exposition aux composés chimiques néoformés et aux migrants, d'intégrer les modulations d'exposition aux contaminants chimiques induites par le système digestif ou encore d'étudier les flux de gènes entre bactéries issues de l'alimentation et le microbiote du tube digestif.
- **Prédire les effets, en particulier le lien entre modes d'actions et effets sur la santé.** Les progrès constants en matière de phénotypage et la diversité des modèles biologiques disponibles aujourd'hui en laboratoire rendent possible la compréhension moléculaire des voies toxicologiques pouvant être activées par des xénobiotiques. Les approches omiques permettent de caractériser le phénotype moléculaire associé à une exposition d'un système biologique à ces xénobiotiques. Le défi est dorénavant de prédire si cette réponse peut correspondre à celle observée lors d'un désordre physiopathologique et donc d'établir le lien, par des approches bioinformatiques, entre le toxome et les signatures obtenues au cours d'un processus pathologique chez l'homme²¹.

²⁰ Escher et al. 2017. Environment International 99: 97-106

²¹ Groh K.J., Carvalho R.N., Chipman J.K., Denslow N.D., Halder M., Murphy C.A., et al. 2015. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. Chemosphere, 120, 764-77

- **Comprendre l'effet des contaminants sur le métabolisme.** Des études *in vitro* et *in vivo* ont montré que certains composés chimiques peuvent avoir un impact sur le métabolisme endogène, notamment énergétique et lipidique²². Ces perturbations métaboliques peuvent avoir un impact à long terme sur la santé et peuvent entraîner des syndromes métaboliques. Le défi majeur est de faire un lien entre la modélisation des modulations du métabolisme et la complexité des modes d'exposition (mélanges, fenêtre d'exposition, faibles doses, chronicité ...). L'autre limite des recherches en toxicologie humaine est l'impossibilité de faire des études interventionnelles pour valider des hypothèses et analyser des mécanismes. Il est donc nécessaire d'avancer en parallèle sur l'étude de l'exposome et du métabolisme endogène.

Verrous méthodologiques

- **Intégrer les données omiques dans les modèles prédictifs.** La priorité méthodologique en termes de modélisation est de « contourner » la difficulté liée à l'intégration de données omiques dans les modèles d'exposition : grand jeu de données, données de différentes natures, données incertaines et variables. Si à cela, on ajoute la dimension temporelle, on se heurte à un double-problème de modélisation. Il faut aller vers les modèles de prédiction pour relier l'intervention ("risk mitigation strategy") à l'exposition²³.
- **Prédire les effets tox/écotox en utilisant des réseaux de connaissances** en vue de construire et formaliser une base de connaissances intégratives. Cette base de connaissance pourra par exemple être modélisée sous forme de graphe de connaissances "knowledge graph". Pour répondre à ces questions à partir des graphes des connaissances, il s'agit de développer des algorithmes et adapter des méthodes existantes dans d'autres domaines pour qu'elles prennent en compte la spécificité des questions de toxicologie²⁴. Ces graphes de connaissances incluront notamment la notion « d'Adverse Outcome Pathway » qui formalise l'enchaînement de événements moléculaires et physiologiques connectant un composé chimique à un effet sur la santé.
- **Faire évoluer les modèles de prédiction des risques et de l'exposition en y intégrant la variabilité.** La variabilité biologique renvoie aux différences réelles entre individus d'une population. Elle ne disparaît pas avec plus de mesures ou plus de connaissance. L'existence de variabilité entraîne au moins deux défis méthodologiques : (i) les modèles devront être probabilistes, et même plus précisément de 2nd ordre pour propager de façon indépendante variabilité et incertitude, et (ii) il sera nécessaire de développer des modèles doses-réponses qui répondent à l'ensemble des individus et non pas à un individu fictif moyen.
- **Développer des nouveaux paradigmes de modélisation intégrative.** Le défi lié à la prédiction de l'exposition en lien avec les pratiques agricoles et d'élevage est multi-systèmes : végétal, animal, humain, et même au sein de chaque système, pourra être multi-disciplines, par exemple sécurité des aliments vue sous l'angle microbiologique et toxicologique. Ce défi multi-systèmes va entraîner un défi méthodologique en termes de

²² Cabaton N. J., Canlet C., Wadia P. R., Tremblay-Franco M., Gautier R., Molina J., Zalko D. 2013. Effects of Low Doses of Bisphenol A on the Metabolome of Perinatally Exposed CD-1 Mice. *Environmental Health Perspectives*, 121(5), 586–93. <https://doi.org/10.1289/ehp.1205588>

²³ den Betsen H.M.W., Amézquita A., Bover-Cid S., Dagnas S., Ellouze M., Guillou S., Nychas G., O'Mahony C., Pérez-Rodríguez F., Membré J.M. 2018. Next generation of microbiological risk assessment: Potential of omics data for exposure assessment. *Int. J. Food Microbiol.* 287, 18-27 ; Haddad N., Johnson N., Kathariou S., Metris A., Phister T., Pielat A., Tassou C., Wells-Bennik M.H.J., Zwietering M.H. 2018. Next generation microbiological risk assessment-Potential of omics data for hazard characterisation. *Int. J. Food Microbiol*

²⁴ Ernst P., Siu A., Weikum G. 2015. KnowLife : a versatile approach for constructing a large knowledge graph for biomedical sciences. *BMC Bioinformatics*.

modélisation qui peut se décliner en trois points: (i) concevoir un modèle qui englobe l'ensemble de la problématique en vue *in fine* de prédire voire piloter les pratiques agricoles et d'élevage, il sera nécessaire de développer de nouveaux paradigmes ; (ii) mettre en place le modèle mathématique pour répondre aux questions biologiques avec un bon dosage qualité du modèle/ temps et difficulté sachant qu'il y aura de l'incertitude et de la variabilité, l'analyse de la propagation de l'incertitude dans des modèles multi-systèmes étant un défi en soi, et (iii) résoudre le système d'équation, inférer le modèle, l'alimenter par des données et de la connaissance.

2.1.6 – Dimensions prédictives de la démo-génétique appliquée aux changements globaux et aux invasions biologiques

Les changements globaux (climat, échanges, occupation des sols, taux de prélèvement des ressources naturelles, etc...) ont des conséquences multiples sur la distribution des espèces et le fonctionnement des écosystèmes. Le terme prédire recouvre un double sens : d'une part produire des postulats quant aux mécanismes influençant la dynamique des populations, en identifiant les paramètres pertinents et leurs effets (« prediction ») et d'autre part prévoir (« forecast ») l'évolution d'une population dans le futur. Les modèles établis dans le premier cas servent généralement à nourrir les méthodes du second. Ainsi, les deux sens correspondent à des pratiques complémentaires, nécessaires pour proposer des stratégies de gestion.

Les objets biologiques sont des espèces animales et végétales pouvant être (i) d'intérêt agronomique ou appartenant à leur cortège de bioagresseurs et d'auxiliaires, (ii) majeures dans des écosystèmes naturels ou peu anthropisés, soit parce qu'elles ont un rôle structurant dans l'agroécosystème soit parce qu'elles constituent (arbres, poissons, grands mammifères) ou affectent (insectes ravageurs, grands mammifères, micro-organismes pathogènes ou symbiotiques) une ressource exploitée, (iii) invasives, affectant potentiellement les 2 catégories d'espèces listées ci-dessus, avec des conséquences possibles à l'échelle des écosystèmes, ou encore (iv) des espèces modèles, particulièrement utiles au progrès académique (arabette, luzerne, souris, homme) et à la base de la biologie translationnelle.

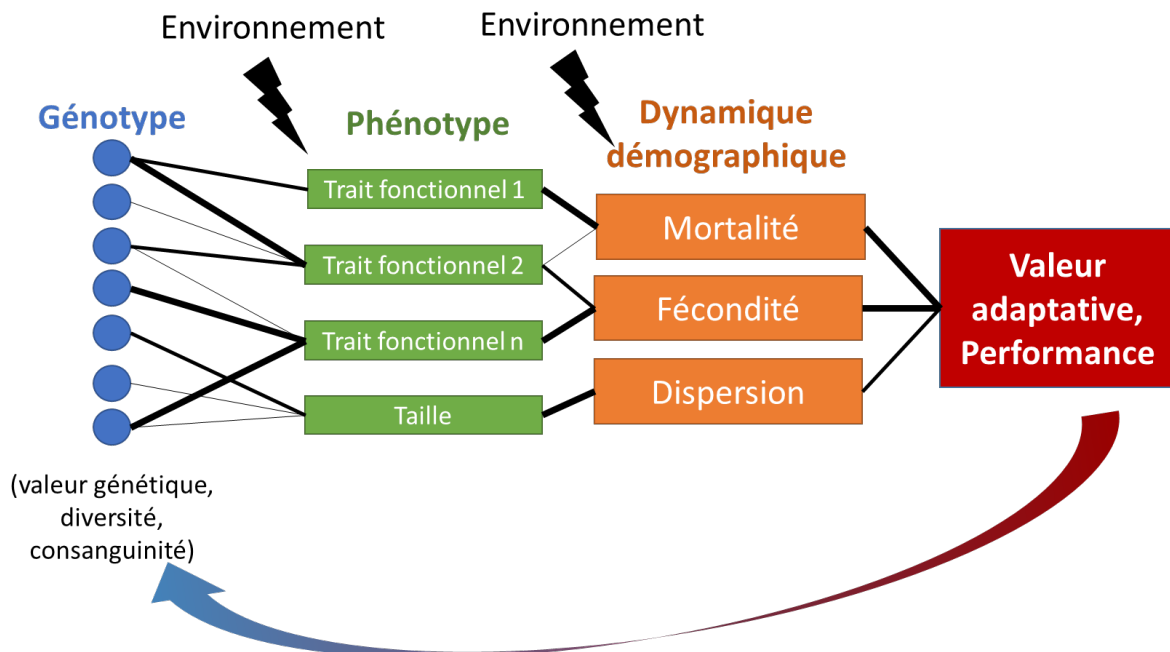


Figure 6 : Représentation schématique montrant comment les génotypes interagissent avec l'environnement pour produire les traits phénotypiques qui influencent, en relation avec l'environnement, les taux démographiques, et, de façon ultime, la valeur adaptative ou performance. En retour, les variations de valeur adaptative ou de performance entre individus vont jouer sur la composition génotypique à la génération suivante. Adapté de Coulson et al. (2006)²⁵.

Enjeux

- **Prédire les capacités d'adaptation des populations face aux changements globaux.** Les mécanismes par lesquels les populations peuvent s'adapter aux variations de leur environnement incluent la migration, la plasticité phénotypique, et l'adaptation génétique²⁶. L'enjeu est de comprendre comment ces différents mécanismes, non mutuellement exclusifs, affectent la dynamique démographique de populations soumises à des perturbations de l'environnement biotique. Il s'agit d'intégrer les processus rétroactifs entre génotype, phénotype et valeur adaptative, d'en caractériser les impacts démographiques respectifs et de déterminer ainsi les propriétés des systèmes en termes de résistance, résilience ou vulnérabilité²⁷. L'enjeu appliqué associé est d'identifier les pratiques de gestion favorisant l'adaptation face aux changements globaux. Ces pratiques peuvent s'appuyer sur des mécanismes naturels ou être assistées par l'homme. D'une manière générale, il s'agit d'améliorer les capacités de prédiction des performances et de construire des schémas de gestion des populations en se basant sur la connaissance de

²⁵ Coulson T., Benton T. G., Lundberg P., Dall S. R. X., Kendall B. E. 2006. Putting evolutionary biology back in the ecological theatre: a demographic framework mapping genes to communities. *Evolutionary Ecology Research*, 8(7), 1155–1171

²⁶ Alberto F. J., Aitken S. N., Alía R., González-Martínez S. C., Hänninen H., Kremer A., Savolainen, O. 2013. Potential for evolutionary responses to climate change - evidence from tree populations. *Global Change Biology*, 19(6), 1645–1661. doi:10.1111/gcb.12181

²⁷ Coulson T., Benton T. G., Lundberg P., Dall S. R. X., Kendall B. E. 2006. Putting evolutionary biology back in the ecological theatre: a demographic framework mapping genes to communities. *Evolutionary Ecology Research*, 8(7), 1155–1171

leurs propriétés génétiques et des concepts de la biologie des populations²⁸. Cela inclut par exemple la conservation des ressources génétiques des espèces d'intérêt, en limitant les risques de dérive génétique ou d'accumulation de caractères défavorables et en maintenant une large diversité génétique.

- **Comprendre les invasions et la recrudescence de bioagresseurs et améliorer le contrôle.** Pour être en mesure de proposer une gestion adaptée et raisonnée, il s'agit d'identifier les sources de bioagresseurs exotiques, connaître les caractéristiques biologiques de la population d'origine et leurs facteurs de régulation et retracer la dynamique d'invasion. En parallèle, il faut identifier le rôle de différents facteurs qui peuvent expliquer la recrudescence de bioagresseurs natifs. Lorsqu'une espèce invasive arrive sur un nouveau territoire, il est souvent plus facile de la contrôler au tout début de l'invasion et pour cela la mise au point d'outils de détection et de contrôle sont précieux. Ensuite, différentes mesures de réduction des risques peuvent être mises en place. L'enjeu finalisé est alors d'optimiser cette gestion et donc d'estimer son efficacité selon de multiples facteurs, et de quantifier les incertitudes associées²⁹.

Priorités scientifiques

- **Comprendre, prédire et gérer la réponse des populations aux changements globaux.** Anticiper et prédire l'impact des changements globaux sur le fonctionnement et la distribution des espèces et des écosystèmes est fondamental pour guider leur gestion. Différents verrous rendent cette anticipation difficile, notamment les multiples incertitudes sur les changements futurs (scénarios climatiques, usage des sols) et la complexité des systèmes écologiques (mécanismes et dynamique de réponse des populations). Cette priorité s'articule autour de 3 axes complémentaires : (i) comprendre et intégrer les mécanismes de la réponse des populations avec une meilleure caractérisation du triptyque génotype, phénotype et environnement tirant partie des analyses haut-débit ; (ii) prédire les rythmes de l'adaptation et les points de basculement en lien avec la non-linéarité de l'impact des changements environnementaux sur les écosystèmes et identifier les points de basculement liés aux niveaux de plasticité phénotypique ou de diversité génétique adaptative ; (iii) adapter les pratiques de gestion pour maintenir les performances et les services écosystémiques, avec l'enjeu de la prise en compte explicite des paramètres et scénarios de gestion dans les modèles de prédiction.
- **Anticiper et accompagner la gestion des bioagresseurs invasifs.** Avoir des capacités d'anticipation avant et au cours d'une crise sanitaire permet de mieux la gérer. Ainsi, prédire l'efficacité de mesures adaptatives de surveillance, de contrôle et plus généralement de gestion dans le cadre des crises sanitaires constitue une priorité. Elle reste cependant difficile à mettre en pratique car peu de données et peu de financements sont disponibles avant la crise sanitaire et durant ses premiers stades. Anticiper et accompagner la gestion des crises requiert de prévoir les zones à risque significatif d'introduction, de prévoir l'expansion future d'un bio-agresseur déjà introduit, de prévoir l'impact de nouvelles directives de surveillance, de prévoir l'effet de mesures de contrôle (sur l'expansion du bio-agresseur et sur le paysage socio-économique), autant d'objectifs

²⁸ Soubeyrand S, de Jerphanion P, Martin O, Saussac M, Manceau C, Hendrikx P, Lannou C. 2018. Inferring pathogen dynamics from temporal count data: the emergence of *Xylella fastidiosa* in France is probably not recent. *New Phytologist*, doi: 10.1111/nph.15177

²⁹ Zhan J., Thrall P. H., Papaix J., Xie L., Burdon, J. J. 2015. Playing on a Pathogen's Weakness: Using Evolution to Guide Sustainable Plant Disease Control Strategies. *Annual Review of Phytopathology*, 53(1), 19–43. doi:10.1146/annurev-phyto-080614-120040

qui nécessitent à la fois des recherches à portée générique et des recherches déclinées selon chaque cas d'étude. Cette priorité inclut les trois points suivants : (i) l'accès aux données de surveillance et leur incorporation dans les approches de modélisation, (ii) l'incorporation de variables explicatives massives pour déterminer les facteurs de risque et les situations à risque et ainsi renforcer le pouvoir prédictif des approches de prévision, et (iii) la prise en compte de la dimension multi-acteur dans la gestion de la crise et les prévisions (gestion des intérêts antagonistes).

Verrous méthodologiques

- **Favoriser l'acquisition et l'exploitation des données répétées dans le temps et dans l'espace.** L'essor des technologies conduit aujourd'hui à un changement important non seulement dans l'ampleur des données collectées, mais également dans leur provenance et leur qualité. Ainsi, une grande part des données est amenée à être collectée en dehors des dispositifs expérimentaux. C'est une opportunité d'obtenir des échantillons de grande taille mais cela nécessite des développements méthodologiques spécifiques pour prendre en compte leur hétérogénéité.
- **Favoriser le développement de modèles de prédiction génériques, intégrant la dimension spatio-temporelle et adaptés aux données massives.** Concernant la gestion des bioagresseurs, les crises sanitaires sont des situations d'urgence et sont causées par des organismes non-modèles. Il est donc essentiel de disposer d'approches pouvant être rapidement appliquées ou adaptées à différents organismes nuisibles, en s'appuyant sur l'expertise dans les premiers temps de la crise. Concernant l'impact des changements globaux, un modèle simple de type boîte noire ajusté sur des données très contextualisées peut être efficace en première approche en termes de prédiction. Mais pour être génériques et utilisables dans d'autres situations, d'autres modèles doivent être développés qui représentent explicitement les mécanismes physiologiques, démographiques et génétiques sous-jacents.
- **Validation.** La validation est souvent un problème à cause du manque de données, or il existe une très forte demande des porteurs d'enjeux pour la mise en œuvre d'expérimentations de validation. Co-concevoir avec eux ces tests de validation, ou mettre en œuvre des études pilotes permettrait de tester à posteriori les prédictions des modèles. Concernant l'évaluation de l'efficacité des méthodes de gestion, une meilleure interaction entre les personnes qui appliquent les mesures de gestion et les modélisateurs est indispensable, que ce soit pour bien intégrer les préconisations ou pour améliorer les modèles concernant la prise en compte des pratiques en retour.

2.1.7 – Analyses et modèles multi-échelles et multi-systèmes

Pour répondre aux défis associés aux approches prédictives, le développement de la modélisation pour représenter, étudier et simuler le comportement de systèmes dynamiques et stochastiques complexes est un point focal, qui soulève des questions à la fois scientifiques et organisationnelles. Les problématiques propres aux approches prédictives recouvrent en particulier : (i) les situations pour lesquelles la complexité et les emboîtements entre processus rendent nécessaires la prise en compte de plusieurs niveaux d'organisation et de plusieurs échelles de temps et d'espace et (ii) le cas des « systèmes de systèmes », où il s'agit de coupler les modèles de plusieurs sous-systèmes en interaction.

Les réflexions des 6 thèmes précédents décrivent de nombreux exemples de fronts de science impliquant l'analyse et la modélisation de tels systèmes, justifiant ainsi l'intérêt d'une réflexion transversale sur les enjeux méthodologiques des approches prédictives, centrée sur l'intégration multi-échelles et multi-systèmes (voir aussi Inria, 2018³⁰).

Les méthodes génériques répondant aux objectifs de modélisation les plus intégratifs incluent des approches « data-driven » (e.g., intégration de données omiques) et « concept-driven » basées sur les couplages de modèles ou sur des approches de modélisation systémiques (e.g., resource-balance analysis, systèmes de systèmes). Ces approches connaissent chacune leurs propres développements théoriques et méthodologiques dans les sciences mathématiques et informatiques³¹, et sont aussi appelées de plus en plus à se croiser et à s'hybrider³². Les évolutions portent aussi bien sur l'intégration plus fine dans les équations d'interactions et rétroactions entre entités à plusieurs échelles de temps et d'espace, que sur le développement continu des méthodes d'apprentissage, de calcul, d'exploration numérique et d'optimisation. Les développements actuels cherchent également à mieux intégrer la stochasticité et les incertitudes inhérentes aux domaines de la biologie et de l'écologie³³. Ils soulèvent la question des données à collecter et des méthodes inférentielles à mobiliser, ainsi que l'organisation collective et les infrastructures à développer pour optimiser les interactions entre chercheurs, chercheuses, et modélisateurs, modélisatrices travaillant à des échelles différentes, ainsi que ceux et celles intéressé.e.s à intégrer ces échelles³⁴.

Enjeux

Les enjeux vont de la compréhension fine du système global jusqu'à sa prédiction et, dans certains cas, à son contrôle et à son optimisation par exemple dans une démarche d'ingénierie des systèmes biologiques ou de gestion d'un agro-écosystème. Les impacts attendus sont très importants dans tous les domaines de l'Inra, de la santé humaine (nutrition, maladie métabolique, etc.) à l'agronomie (intégration systémique des plantes et animaux) et plus généralement aux agro-sciences et écologie (obtention de modèles prédictifs ayant des domaines de validité accrus – multi-stress). Les développements de l'approche systémique aux échelles subcellulaire et cellulaire ouvrent la voie au développement d'une réelle ingénierie des systèmes biologiques, dont le développement aura des conséquences directes en particulier dans le champ des biotechnologies et spécifiquement en biologie de synthèse à travers par exemple les possibilités qu'elle offre dans le design des « cell factories ».

L'enjeu en termes de modélisation est de construire des cadres formels permettant de déduire les liens entre des effets microscopiques et des effets macroscopiques selon un enchaînement de processus opérant à plusieurs niveaux d'organisation (de la molécule à l'individu, voire à la population), en intégrant des effets d'interaction et de rétroaction, tant au sein de l'objet étudié qu'avec son environnement, et ce, à plusieurs échelles temporelles et spatiales. À terme, il s'agit d'aller vers des modèles fins et intégrés des organismes ou des communautés, mêlant harmonieusement modèles (bio)mécaniques, (bio)physiques et (bio)chimiques avec des modèles systémiques.

³⁰ Inria, Direction scientifique, 2018. Plan stratégique scientifique 2018-2022. Inria : Rocquencourt

³¹ D. Chavalarias, P. Bourguin, E. Perrier et al. 2009. French Roadmap for complex systems 2008-2009, second issue by the French National Network for Complex Systems. <hal-00392486>. Mise à jour 2017 : <https://roadmap.iscpif.fr/>

³² F. Mazzocchi. 2015. Could Big Data be the end of theory in science ? EMBO Reports, 16, 1250:1255

³³ N. Mouquet, Y. Lagadeuc et al. 2015. Predictive ecology in a changing world. Journal of Applied Ecology, 2015 : 1-18

³⁴ I. Yanai et E. Chmielnicki. 2017. Computational biologists : moving to the driver's seat. Genome Biology, 18:223

Priorités scientifiques

➤ **Les défis « individu-centré »**

L'intégration verticale aux échelles subcellulaire et cellulaire, i.e. l'intégration progressive des échelles de temps et d'espace croissantes, se trouve à la croisée des efforts et résultats attachés à la biologie moléculaire et cellulaire, aux (nouvelles) technologies d'observation du vivant et de la modélisation du vivant dans un sens large. Les dix dernières années ont été marquées par d'incroyables progrès dans l'intégration verticale des cellules procaryotes, jusqu'à la mise à disposition emblématique de simulateurs de cellules bactériennes³⁵. L'intégration verticale des cellules *per se* porte deux défis couplés qui sont la gestion des données, des informations et des connaissances et le transfert et la généralisation des résultats obtenus sur des espèces modèles vers les autres espèces. En ce qui concerne le premier défi, c'est la question même de l'interprétation du déluge de données hétérogènes et leur intégration avec la connaissance connue *via* des représentations formelles qui pose des problèmes liés à la nature spécifique des organismes vivants. Il s'agit de fournir à l'ensemble de la communauté biologique les outils, les méthodes et les infrastructures leur permettant de produire et d'exploiter au mieux les potentiels offerts par les nouvelles technologies d'observation du vivant et la modélisation. L'autre enjeu majeur de la modélisation (au sens large) dans ce contexte est de permettre de transférer plus efficacement et rapidement les résultats obtenus sur les organismes modèles vers les organismes non modèles, permettant ainsi aux communautés de travailler ensemble plus efficacement.

L'intégration verticale des organismes vivants et des communautés recouvre l'ensemble des questions nécessitant l'intégration d'au moins une échelle supérieure à celle de la cellule (e.g., une ou plusieurs cellules dans leur environnement spécifique, des communautés de cellules plus ou moins organisées ou encore l'organisme/la communauté au sein de son environnement). Dans ce cadre, il s'agit de développer des modèles systémiques permettant de relier les connaissances physiologiques à l'échelle de l'individu aux connaissances acquises à l'échelle des cellules. C'est typiquement le phénotypage haut débit, les « omiques » et la possibilité d'inclure de façon plus ou moins fine les échelles infra et cellulaire (voir l'axe précédent) qui renouvellent le cadre. Enfin, le développement de tels modèles a des impacts directs et importants sur la priorité décrite ci-après.

La recherche de relations de type « cause à effet » dans des données multi-sources ou entre des grandeurs biologiques et physiques a toujours constitué une activité importante dans le champ biologique. Cet axe partage avec les précédents les difficultés liées à la manipulation des données et informations biologiques et soulève donc lui aussi des questions se rapportant à l'organisation et à la gestion des données et de l'information attachées aux systèmes biologiques. L'un des défis concerne la recherche de liens entre génotype(s) et phénotype(s). Ce défi est déjà central par exemple dans la recherche des acteurs moléculaires et des gènes associés responsables de certains types de cancer. Il est important de développer des approches qui non seulement s'appuient sur des flux de données multi-sources et souvent massifs, mais aussi intègrent les progrès réalisés dans la modélisation du vivant à toutes les échelles. Cela conduit à développer des méthodes dédiées aptes à identifier des relations

³⁵ Karr J. R. et al. 2012. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2), 389-401

significatives dans des données hétérogènes sur la base de l'exploitation de modèles systémiques.

➤ **Les défis « métapopulations- » et « environnement-centrés » où les individus sont des acteurs parmi d'autres**

Cela concerne les défis attachés à l'intégration du vivant au sein de l'environnement et les questions de recherche associées. Se posent les questions de la représentation adéquate dans « l'environnement » des entités biologiques et de leurs interactions et évolutions en allant de la compréhension à l'aide à la décision dans l'incertain (utilisation de modèles pour agir). La mise en place de modèles dynamiques structurés ou multi-échelles (en espace et temps) de très grande dimension pose des questions spécifiques liées à leur calibration (plans d'expérience, inférence des paramètres, assimilation de données, etc.), leur réduction et leur simplification, leur exploration numérique mais aussi des questions liées à leur robustesse, à la propagation d'incertitude, aux tests statistiques, et plus globalement à leur validation. Un des enjeux actuels est celui de la réconciliation/agrégation de modèles de différents types. Ces derniers représentent des phénomènes de différentes natures, couvrant des échelles de modélisation différentes (spatiales et temporelles). L'exemple type de ce défi est la construction d'un modèle intégré, à l'échelle d'une région, qui couple un modèle d'évolution climatique, hydraulique, ainsi que les composantes « vivantes » de cette région comme par exemple un modèle de couverts forestiers.

Verrous méthodologiques

- **Promouvoir des outils et méthodes favorisant l'inter-opérabilité.** Il s'agit de fournir à l'ensemble de la communauté scientifique les outils, méthodes et infrastructures permettant de produire et d'exploiter au mieux les potentiels offerts par les nouvelles technologies d'observation du vivant et la modélisation à différentes échelles pour de multiples systèmes. Cela conduit à contribuer à des actions autour de (i) la standardisation des données, connaissances (e.g., processus) et modèles (langages appropriés pour chaque grand domaine), (ii) l'annotation, l'assistance à l'annotation (interface homme-machine, apprentissage) et la visualisation de jeux de données multi-échelles de grande volumétrie et (iii) la construction et l'utilisation d'ontologies pour représenter les systèmes biologiques mais aussi organiser les données de tout type (e.g., Henry et al. 2017³⁶).
- **Acquérir les données adéquates pour coupler les échelles, maximiser les chances d'observer des événements mêmefurtifs.** La modélisation prédictive ne peut se concevoir sans données pour l'élaboration des hypothèses de modélisation ou la validation du modèle. Ainsi, un défi concerne la qualité et la richesse informative des données brutes et des premiers traitements sur ces données, en particulier (i) l'identification et la combinaison des technologies les mieux adaptées pour traiter le multi-échelle (imagerie 3D+temps, capteurs, single-cell, données grandes échelles...) et (ii) la fusion de données obtenues à différentes échelles et par des modes d'observation différents, comme dans le cas de l'imagerie corrélative. Un second défi relève de la planification expérimentale (e.g., Widder et al., 2016³⁷), par exemple (i) l'optimisation de

³⁶ V. J. Henry, A. Goelzer, A. Ferré, S. Fischer, M. Dinh, V. Loux, C. Froidevaux, et V. Fromion. 2017. The Bacterial Interlocked process ONtology (BiPON) : a systemic multi-scale unified representation of biological processes in prokaryotes. *Journal of Biomedical Semantics*, 8:53

³⁷ S. Widder et al. 2016. Challenges in microbial ecology : building predictive understanding of community function and dynamics. *The ISME Journal*, 10, 2557-2568

réseaux de capteurs dans le temps et dans l'espace, pour l'acquisition de séries temporelles d'observations multi-échelles ou (ii) la planification adaptative pilotée par des méthodes d'assimilation de données et la simulation.

- **Intégrer, simplifier, simuler et analyser formellement plus facilement.** La modélisation des systèmes est au cœur des développements actuels des modèles de cellules ou d'organismes. Il s'agit de développer une réelle ingénierie des systèmes biologiques, combinant les outils et moyens technologiques d'observation et de modification du vivant avec ceux des sciences formelles et quantitatives. Dans le cadre multi-échelles ou multi-systèmes les efforts doivent porter sur 7 points en particulier: (i) développer et affiner les modèles d'intégration verticale des cellules (procaryotes et eucaryotes), en y incluant des processus essentiels au fonctionnement de la cellule, (ii) conduire des efforts similaires sur l'intégration verticale des organismes vivants en y intégrant explicitement le rôle central des cellules, (iii) développer des méthodes d'inférence exploitant les connaissances intégrées dans des modèles systémiques pour rechercher plus efficacement des « relations significatives » dans des données hétérogènes (possiblement massives), (iv) développer des méthodes d'apprentissage adaptées au multi-échelle (classification supervisée et non supervisée, construction de sous-modèles corrélatifs (boîtes noires) pour des modèles globaux et optimisation en grande dimension), (v) poursuivre des travaux théoriques sur la réduction de modèles et la modélisation multi-échelle, (vi) développer les méthodes de couplage de modèles stochastiques, et (vii) développer l'exploration du comportement de modèles complexes par des méthodes basées sur l'analyse, le calcul et la simulation.

2.2 – Interfaces entre thèmes

Au-delà du caractère transversal du thème précédent, des interfaces entre thèmes porteuses de défis scientifiques en biologie prédictive ont également été identifiées. De façon générale, ces interfaces répondent à des enjeux « d'intégration verticale » des organismes vivants mais également des enjeux « d'intégration horizontale », les organismes étant dans ce contexte un élément parmi d'autres du système. « L'intensité » de ces interfaces est schématisée dans le tableau ci-dessous, et quelques exemples sont :

- Les questions d'intégration fonctionnelle du métabolisme et de couplage entre modèles métaboliques et autres modèles à différentes échelles (thème 1) apparaissent stratégiques, que cela soit dans les processus de développement (thème 2) et plus largement de prédiction des phénotypes (thème 3), ou encore dans le cadre du dialogue entre l'hôte et son microbiote (thème 4).
- La prédiction et l'évaluation du risque en toxicologie (thème 5) peuvent s'appuyer sur une description fine du métabolisme cellulaire (thème 1) comme sur la prise en compte de l'évolution et du potentiel adaptatif de populations à l'échelle de l'individu hôte (thème 4) ou, pour l'écotoxicologie, à l'échelle de populations (thème 6).

- La prédiction des phénotypes (thème 3) et de leur variabilité dans les populations est une étape clé pour comprendre leur dynamique adaptative sous l'effet du changement global ou lors des processus d'invasion (thème 6).

	Métabolisme	Morphogénèse	Phénotypes	Toxicologie	Holobionte
Morphogénèse					
Phénotypes					
Toxicologie					
Holobionte					
Populations					

Rouge : +++, Orange : ++, Jaune : +

3 – RECOMMANDATIONS

Ces recommandations se nourrissent des réflexions des groupes de travail et du COPIL de la prospective. Elles sont organisées autour **de 8 axes ayant pour objectif d'accompagner la mise en œuvre des priorités scientifiques et plus largement de soutenir la stratégie de l'Inra dans le domaine des approches prédictives en biologie et écologie**. Il est évident que certaines recommandations, en particulier en lien avec la gestion et l'analyse des données, ont vocation à irriguer au-delà du périmètre biologie et écologie de la prospective.

Axe 1 – Fédérer la communauté scientifique pour développer l'interdisciplinarité

L'objectif est de mettre en place un cadre d'animation **coordonné à l'échelle de l'Institut** pour fédérer une communauté scientifique à même de développer les recherches interdisciplinaires nécessaires à la mise en œuvre des approches prédictives pour la biologie et l'écologie (articulant sciences de l'observation et sciences formelles dès la conception des projets, mais aussi les sciences humaines et sociales quand nécessaire) **en décroissant les thèmes scientifiques, les objets et échelles d'étude**. Le principe de « cycles d'animation », plutôt que des animations ponctuelles et non connectées entre elles, s'inscrivant sur une durée prédéterminée apparaît pertinent pour accompagner et fédérer les communautés. Ces animations scientifiques seront ouvertes à nos partenaires ESR, en particulier dans le cadre d'actions de programmation conjointes avec d'autres organismes.

Il s'agira également de **capitaliser sur des réseaux d'animation existants** en interne ou au national (e.g., GDR du CNRS, les JOBIM³⁸, animations de l'IFB). Plus récemment, à l'échelle régionale, des dispositifs et programmes de recherche avec des ambitions en biologie prédictive ont été mis en place. Les Instituts Convergences DataIA³⁹, Digit@Ag⁴⁰, Inception⁴¹,

³⁸ JOBIM : Journées Ouvertes en Biologie, Informatique et Mathématique organisées sous l'égide de la Société Française de Bioinformatique, <https://www.sfbi.fr/content/jobim-2018>

³⁹ DataIA, sciences des données, intelligence et société, <https://dataia.eu>

⁴⁰ DigitAg, agriculture numérique, <https://www.hdigitag.fr/fr/>

⁴¹ Inception, methods for integrative biology for health issues, <https://www.inception-program.fr/en>

Cland⁴² ont tous des ambitions d'améliorer la capacité prédictive des modèles développés. Il en est de même pour les Labex NUMEV ou LMH⁴³, pour une part au moins de leurs objectifs. Une troisième proposition porte sur **l'adaptation et le développement des formes de travail collaboratif** dédiées à l'analyse et à la modélisation, selon le principe des hackathons. Il pourra s'agir de collectifs de maths/info qui se regroupent sur plusieurs jours pour développer des solutions de modélisation sur un objet/thème donné, ou encore de regroupements de maths/info ayant des compétences/spécialités différentes avec des biologistes fournissant des données pour répondre à une question biologique. Par exemple, dans le domaine de la modélisation des réseaux métaboliques, les « jamborees » consistent à regrouper sur un ou deux jours des experts du métabolisme d'un organisme afin de compiler leurs connaissances pour créer un modèle de haute qualité⁴⁴.

Axe 2 – Soutenir des projets de recherche innovants

La première initiative de l'Institut dans le domaine remonte au programme fédérateur de recherches en biologie intégrative animale, végétale et microbienne (AgroBI, 2006-2008) qui a permis le soutien à des projets de recherche et également le recrutement d'attachés scientifiques contractuels. Depuis de nombreuses initiatives ont été développées dans les départements et également au sein des méta-programmes. Un autre exemple concerne l'IMSV (Institut de Modélisation des Systèmes Vivants, 2013-2016), coordonné par l'Inra, qui s'est construit dans le cadre de l'Idex Paris-Saclay. Rassemblant une dizaine de partenaires institutionnels, il a permis de faire travailler ensemble plusieurs communautés de modélisateurs, autour des questions portées par l'intégration aux échelles subcellulaire et cellulaire (la bactérie servant d'exemple) et multi-échelles (la plante servant d'exemple). Afin d'amplifier la dynamique et de développer l'interdisciplinarité aux frontières des sciences du vivant et des sciences formelles, il est proposé **d'inscrire explicitement, et lorsque pertinent, les approches prédictives pour la biologie et l'écologie comme axe de programmation des méta-programmes de seconde génération.**

Axe 3 – Mettre en place les outils, les méthodes et les technologies répondant aux enjeux de la modélisation multi-échelle et multi-système

Il s'agit de réunir les conditions pour répondre aux enjeux de la modélisation multi-échelle et multi-système avec l'objectif **de mettre en système la production des données, les analyses et la modélisation et les e-infrastructures.** Cet axe s'articule autour de 4 enjeux stratégiques : (i) Assurer l'accès à des dispositifs d'acquisition de données de l'Inra et de partenaires, (ii) développer des systèmes d'information reposant sur les approches FAIR et sur des ontologies cohérentes entre données et processus modélisés, (iii) adapter, développer et mettre en œuvre des méthodes de modélisation, de statistique, d'apprentissage automatique et d'intelligence artificielle, sur de grands corpus de données et (iv) favoriser la généricité et la diffusion des méthodes par le développement et l'appropriation de produits logiciels génériques. Un effort de structuration important est nécessaire pour tirer le meilleur parti des ressources de l'Institut, en identifiant les ressources et outils collectifs à renforcer, en assurant

⁴² Cland, changement climatique et usage des terres, <https://cland.lsce.ipsl.fr>

⁴³ <http://www.lirmm.fr/numev/>; <https://www.fondation-hadamard.fr/fr/mathematiques-et-sciences-du-vivant/labex-lmh-mathematiques-et-sciences-du-vivant>

⁴⁴ Thiele, J. and Palsson, B.Ø. 2010. Reconstruction annotation jamborees: a community approach to systems biology. Mol. Syst. Biol., 6, 361

leurs synergies et complémentarités en tant qu'e-infrastructures de recherche et en s'ouvrant à des partenariats ambitieux. Il est évident que cette recommandation alimente plus largement plusieurs objectifs de la priorité OpenScience du document d'orientation Inra 2025.

Axe 4 – Mettre en place des actions conjointes avec des partenaires académiques nationaux

Les partenariats académiques de l'Institut sont riches dans le domaine des sciences du vivant. La valeur ajoutée paraît évidente et de portée large en particulier avec l'Inria et le CNRS (INSMI, l'INS2I, l'INSB et l'INEE⁴⁵), mais d'autres collaborations plus formelles sur certaines thématiques mériteraient également d'être instruites.

Les actions conjointes Inra-Inria bénéficient déjà d'une convention cadre entre les deux instituts, qu'il faudra renouveler en veillant à renforcer les collaborations entre équipes sur les enjeux de biologie et écologie prédictive. Concernant le CNRS, une réflexion visant à identifier des sujets porteurs de collaborations nouvelles et à forte valeur ajoutée a été initiée par les Directions Générales des deux Instituts (séminaire conjoint du 23 janvier 2019) et devra être poursuivie et opérationnalisée. Au-delà de collaborations existantes (par exemple avec le CIRAD), certaines plus spécifiques de thématiques mériteraient d'être développées : avec l'ANSES et l'EFSA sur les questions de toxicologie et écotoxicologie prédictive et la gestion des bioagresseurs invasifs, avec l'INSERM en lien avec son programme microbiote, avec l'Institut Pasteur dans le domaine de la biologie synthétique et finalement avec l'École Centrale de Paris dans le domaine de l'optimisation d'idéotypes (collaborations existantes sur la modélisation des plantes).

Axe 5 – Accroître le rayonnement et l'attractivité de l'Inra aux échelles nationale, européenne et internationale

Le thème des approches prédictives en sciences du vivant est en plein **essor au niveau national comme international** : nombre croissant de publications, naissance de journaux dédiés (e.g., International Journal of Integrative Biology), développement d'instituts dédiés dans certains pays (e.g., Institute of Integrative Biology, ETH Zurich ; Center for Plant Integrative Biology, Nottingham, UK ; Genome Institute of Singapore...). **L'enjeu de positionnement et de compétitivité dans les domaines de recherche de l'Institut est indéniable.** La recommandation est de capitaliser sur le leadership de nos équipes et sur les réflexions conduites dans le cadre de cette prospective pour accroître le rayonnement de l'Institut par (i) des actions de communications sur les « success stories » de nos équipes, (ii) des accords de collaboration avec des partenaires étrangers stratégiques (des actions dans ce cadre ont déjà été initiées e.g., LIA en projet avec le CNRS et le Sainsbury Laboratory de Cambridge sur la morphogénèse des plantes, ou encore le LIA Genetic Improvement of Indian Cattle and buffaloes avec le BAIF en Inde) et, (iii) la valorisation de ce travail de prospective dans le cadre de workshop thématiques associant des scientifiques reconnus mondialement dans le domaine.

⁴⁵ INSMI : Institut national des mathématiques et de leurs interactions ; INS2I : Institut des sciences de l'information et de leurs interactions ; INSB : Institut des sciences biologiques ; INEE : Institut Écologie et Environnement

Axe 6 – Monter en compétence

Le domaine de la formation et de la montée en compétences est un vaste sujet qui en interne a fait l'objet de recommandations dans le cadre du rapport de F. Garcia et al. (2013)⁴⁶, et a également constitué un axe de travail (biologie massive et développement des approches de modélisation) de la gestion prévisionnelle des emplois et compétences de l'Institut. A l'échelle nationale, un élément marquant de ces dernières années est la mise en place de dispositifs et programmes de recherche avec des ambitions en biologie prédictive et une part importante donnée aux actions de formations (Labex, Instituts Convergences, EUR, Infrastructure, etc.). Dans ce contexte très riche, nos recommandations ciblent les **actions de formation** en capitalisant sur les dynamiques à l'œuvre en particulier aux échelles régionales et avec nos partenaires ESR, et en mobilisant des nouveaux outils de formation (e.g., open online course, pages web avec tutoriaux et présentations powerpoint sonorisées). **L'attractivité de l'Institut** pour des jeunes scientifiques est un élément déterminant de notre positionnement dans le domaine aussi bien aux échelles nationale qu'internationale. Dans ce cadre, nous proposons d'augmenter significativement le nombre de thèse cofinancées par l'Inra et également de créer des packages (financement de base, thèses et si possible post-doc) permettant d'environner pendant 4 ans de jeunes chercheurs recrutés dans le domaine.

Axe 7 – Établir et mettre en œuvre une stratégie de partenariat socio-économique

En lien direct avec le processus de « mise en données » du monde, les approches prédictives constituent un domaine en plein essor pour le développement des partenariats socio-économiques. Si la dynamique est clairement à l'œuvre en marketing, ou encore en santé, elle démarre dans nos domaines de biologie prédictive. En génétique animale, l'avènement de la sélection génomique dans les années 2000 donne une idée du potentiel d'innovation de rupture apporté par l'introduction de nouvelles méthodes prédictives. Plus récemment, des exemples emblématiques ont été développés en épidémiologie ou dans le cadre de l'agriculture numérique. Au-delà de ces quelques succès, le potentiel reste très important pour exploiter les informations obtenues grâce à de nouvelles technologies d'observation et de calcul, et à de nouvelles méthodes de modélisation et d'apprentissage. Une réflexion a été initiée pour chacun des thèmes mais elle doit être approfondie par **des analyses stratégiques de partenariat**.

Axe 8 – Alimenter et partager un cadre éthique et juridique en lien avec les approches prédictives pour l'évaluation et la gestion

Ce sujet mérite incontestablement d'être approfondi ; nous en posons ici quelques bases. **Le développement des approches prédictives avec des objectifs d'évaluation et de gestion doit s'accompagner d'une « éthique de la prédiction »**. Il y a incontestablement des risques liés à l'usage des prédictions par les porteurs d'enjeux avec par exemple de mauvaises interprétations dues à la méconnaissance de la notion d'incertitude. D'un point de vue juridique, il apparaît également nécessaire de considérer la notion de responsabilité(s) en

⁴⁶ Garcia F. et al. 2013. Note de synthèse du groupe de travail « Chantier Biologie Prédictive », INRA, 11 p.

termes de prédiction. En effet, si les prédictions assorties de trop d'incertitude conduisent à de mauvais choix de gestion, comment se protéger de l'accusation de « fausse prédiction » ? Finalement, l'importance d'associer les porteurs d'enjeux dans des démarches de co-construction de scénario de gestion, en amont du développement des modèles pour la prédiction, est soulignée. Il serait pertinent d'analyser les critères (scientifiques, politiques, stratégiques...) sur lesquels se basent les décideurs pour prendre en compte ou non les résultats d'un modèle. La validation est souvent un problème à cause du manque de données, or il existe une très forte demande des porteurs d'enjeux pour la mise en œuvre d'expérimentations de validation. Co-concevoir avec eux ces tests de validation, ou mettre en œuvre des études pilotes permettrait de tester a posteriori les prédictions des modèles.

4 – MISE EN PERSPECTIVE

L'objectif de cette prospective était double : (i) éclairer l'Institut sur les futurs fronts de science pour certains thèmes du vaste champ des approches prédictives en biologie et écologie, thèmes identifiés comme stratégiques par son comité de pilotage et (ii) alimenter nos orientations en proposant des actions favorisant le développement et le rayonnement international de nos recherches dans le domaine et contribuant à répondre aux grands défis sociétaux. **Il s'agira notamment de mettre en œuvre ces réflexions et recommandations dans le cadre des futurs schémas stratégiques des départements, dans la construction des nouveaux méta-programmes ainsi que dans le cadre du plan stratégique INRAE 2030.** Cette mise en œuvre a d'ores et déjà été initiée. Ainsi, par exemple, le nouveau méta-programme HOLOFLUX (Holobiontes et flux microbiens au sein des systèmes agri-alimentaires⁴⁷) a été alimenté par les travaux sur le thème « Holobionte » et le programme interdisciplinaire sur la santé des plantes mobilise les travaux sur le thème « dimension prédictive de la démogénétique appliquée aux changements globaux et aux invasions biologiques ». Un nouveau méta-programme centré sur les approches intégratives et prédictives de la molécule à l'organisme est également en réflexion, en mobilisant les travaux sur les thèmes « régulations et compromis métaboliques », « morphogénèse et développement des organismes » et « modélisation et prédiction des phénotypes ».

⁴⁷ <https://intranet.inra.fr/metaprogrammes/d-infos-MP/Les-differents-MP/Les-MP-lances-depuis-2019/HOLOFLUX>

