



**2IEME COLLOQUE DE L'INFRASTRUCTURE INRAE GENOMICS**  
13, 14 et 15 novembre 2023



Hôtel Mercure Bordeaux Château Chartrons  
81 cours Saint-Louis  
33300 BORDEAUX  
France

[colloque-inraegenomics@inrae.fr](mailto:colloque-inraegenomics@inrae.fr)

# PROGRAMME

## LUNDI 13 NOVEMBRE

---

12H – 13H30

DEJEUNER (BUFFET)

14H – 15H30

PRESENTATION DE L'INFRASTRUCTURE INRAE GENOMICS ET DE L'EXPERTISE DES PLATEFORMES

15H30 – 16H

PAUSE

16H – 18H20

SESSION 1 - EXPLORATION DE LA DIVERSITE

Myriam Heuertz

*Plant mutations: slaying beautiful hypotheses by surprising evidence*

Aurélien Capitan

*Massive detection of cryptic recessive genetic defects in cattle mining millions of life histories*

Emilie Delpuech

*Séquençage de 400 génomes de bars (*Dicentrarchus labrax*) pour identifier des variants génétiques de résistance à des maladies*

Javier Belinchon-Moreno

*Nanopore adaptive sampling et caractérisation de la diversité génétique du NLRome/resistome chez le melon*

Stéphane Nicolas

*Genomic selection and Genome-wide association studies on DNA pools identifies promising maize landraces and genomic regions to develop next generation varieties*

19H30

DINER

## MARDI 14 NOVEMBRE

---

8H30 – 9H50

SESSION 2 - ASSEMBLAGE DE GENOMES, PANGENOMES ET ANALYSE DES SV (1)

Stéphanie Sidibe-Bocs

*Le séquençage haute-fidélité du génome complexe du vanillier apporte une preuve moléculaire de l'endoréplication partielle chez les plantes*

Alain Vignal

*Haplotypes et répétitions en tandem dans le génome de l'abeille à miel : où sont les centromères ?*

Ludovic Duvaux

*Testing pangenomic tools for structural variant detection in forest tree species*

9H50 – 10H20

PAUSE

10H20 - 12H

SESSION 3 - ASSEMBLAGE DE GENOMES, PANGENOMES ET ANALYSE DES SV (2)

Grégoire Aubert

*Séquençage du génome de la féverole : du statut d'espèce orpheline à l'ère de la pangénomique*

Véronique Decroocq

*De la génétique à la génomique chez les arbres fruitiers à noyaux ou comment repousser le champ des possibles chez des espèces orphelines, pérennes*

Thomas Faraut

*La variation de structure : dynamique, représentation et impact sur la fonction*

Jerome Gouzy

*ATLAS : Exploration et exploitation du catalogue d'allèles des espèces cultivées*

12H – 13H30

DEJEUNER

13H30 – 15H15

SESSION 4 - REGULATION DE L'EXPRESSION DES GENOMES

Pascal Martin

*Regulation of gene expression by the BORDER negative transcription elongation factors in *Arabidopsis thaliana**

Aurore Coissac

*Role of DNA methyltransferases of *Ralstonia pseudosolanacearum* (GMI1000 strain) in DNA methylation, virulence and growth*

Stéphane Maury

*Evolutionary and functional impact of epigenetic variations in forest trees facing climate change (EPITREE)*

Hervé Acloque

*INRAE contribution to the functional annotations of pig and chicken genomes as part of the European GENE-SWitCH project*

16H – 18H

VISITE DE LA CITE DU VIN

19H30

DINER

9H00 – 10H20

**SESSION 5 - METAGENOMIQUE**

Sylvie Combes

*Piloter l'installation du microbiote des animaux d'élevage : les technologies de séquençage pour explorer les communautés bactériennes*

Paola Fournier

*Vers une viticulture sans pesticide: analyse du microbiome de la vigne et des sols viticoles pour développer le biocontrôle microbien de l'oïdium et du mildiou*

Gaëlle Marliac

*Le microbiome du sol et la régulation d'une maladie du blé*

10H20 – 10H50 PAUSE

10H50 - 12H

**SESSION 6 - QUELS FUTURS POUR LA GENOMIQUE POUR LES AGROBIOSCIENCES ?**

*Retour sur l'enquête INRAE Genomics 2022*

*Table ronde / Discussion*

**FIN DU COLLOQUE**

# ACCES AU COLLOQUE

## LOCALISATION

### Hôtel Mercure Bordeaux Château Chartrons

81 cours Saint-Louis

33300 BORDEAUX

France

Coordonnées GPS: 44.85963, -0.56979

## TRANSPORTS

Gare Bordeaux Saint-Jean à 5 km

~30 min en transport en commun dont 10 min à pied en tram ligne C direction Gare de Blanquefort, arrêt Grand Parc

<https://www.infotbm.com/fr>

## INFORMATIONS UTILES

### ACCUEIL ET REMISE DES CLES

L'accueil et la remise des badges se feront dans le hall de l'hôtel à partir de 11h30.

La remise des clés des chambres se fera en fin d'après-midi, les bagages pourront être laissés en attendant dans la salle de conférence.

Le parking de l'hôtel est accessible aux participants.

### REPAS

Les repas et les pauses sont pris en charge du déjeuner du lundi 13 midi au petit déjeuner du mercredi 15 midi.

Le déjeuner du lundi 13 est servi sous forme d'un buffet accessible de 12h à 13h30.

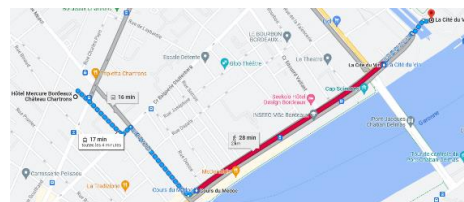
NB : Le déjeuner du mercredi 15 n'est pas inclus.

Les consommations hors repas et pauses sont à la charge du participant et sont à régler directement à l'hôtel.

### VISITE DE LA CITE DU VIN

La cité du Vin nous accueille le mardi 14 de 16h à 18h.

L'accès depuis l'hôtel se fera en transport en commun en tram ligne B direction Claveau, arrêt Cité du Vin



### CONTACT

[colloque-inraegenomics@inrae.fr](mailto:colloque-inraegenomics@inrae.fr)

## AVEC LA PARTICIPATION :

De la **Délégation aux Infrastructures Scientifiques Collectives d'INRAE (DISC)**

Des départements INRAE :

**Biologie et Amélioration des Plantes (BAP)**

**Ecologie et biodiversité des milieux forestiers, prairiaux et aquatiques (ECODIV)**

**Génétique Animale (GA)**

Des **plateformes INRAE GENOMICS**



**MERCI A TOUS :**  
**ORATEURS, ANIMATEURS ET**  
**PARTICIPANTS !**

# RESUMES DES PRESENTATIONS

## SESSION 1 - EXPLORATION DE LA DIVERSITE

**Myriam Heuertz**

***Plant mutations: slaying beautiful hypotheses by surprising evidence***

The Weismann theory states that hereditary traits are transmitted exclusively from the germline. The theory is valid in most animals where germline cells are set aside early in development. In plants, germline segregation is generally assumed to occur late in development, which leads to several predictions on the fate of somatic mutations occurring in plant tissues: mutations have generally low frequency in plant tissues; mutations at high frequency have a higher chance of intergenerational transmission; branching topology of the tree dictates mutation distribution; and, exposure to UV radiation increases mutagenesis. We produced a unique plant dataset of 60 high-coverage whole-genome sequences of two tropical tree species and identified 18,274 de novo somatic mutations, almost all at low frequency in tissues. We demonstrate that: 1) low-frequency mutations are transmitted to the next generation; 2) mutation phylogenies deviate from the branching topology of the tree; and 3) mutation rates and mutation spectra are not demonstrably affected by differences in UV exposure. Altogether, our results suggest far more complex links between plant growth, ageing, UV exposure, and mutation rates than commonly thought.

**Aurélien Capitan**

***Massive detection of cryptic recessive genetic defects in cattle mining millions of life histories***

We present a data-mining framework designed to detect recessive defects in livestock that have been previously missed due to a lack of specific signs, incomplete penetrance, or incomplete linkage disequilibrium. This approach leverages the massive data generated by genomic selection. Its basic principle is to compare the observed and expected numbers of homozygotes for sliding haplotypes in animals with different life histories. Within three cattle breeds, we report 33 new loci responsible for increased risk of juvenile mortality and present a series of validations based on large-scale genotyping, clinical examination, and functional studies for candidate variants affecting the NOA1, RFC5, and ITGB7 genes. In particular, we describe disorders associated with NOA1 and RFC5 mutations for the first time in vertebrates. The discovery of these many new defects will help to characterize the genetic basis of inbreeding depression, while their management will improve animal welfare and reduce losses to the industry.

**Emilie Delpuech**

***Séquençage de 400 génomes de bars (*Dicentrarchus labrax*) pour identifier des variants génétiques de résistance à des maladies***

Depuis la fin de ma thèse en 2021, j'ai mené mes recherches postdoctorales en tant que chercheuse junior au SYSAAF (Syndicat Avicole et Aquacole Français), puis dans l'UMR MARBEC de l'IFREMER à Palavas-les-Flots. Grâce à mes connaissances en génétique et en analyse de données de génome complet via des outils bioinformatiques, j'ai intégré le projet GeneSea dans lequel 400 génomes de bars (*Dicentrarchus labrax*) ont été obtenus, et qui nous ont permis d'identifier des gènes d'intérêts pour la résistance à deux maladies : la nodaviriose et la bactériose causée par *Vibrio harveyi*. Durant cette présentation, les méthodes d'exploration de données haut-débits et de GWAS vont être développées pour comprendre le schéma d'analyses mis en place dans ce projet de génétique des caractères. Parmi les QTL putatifs identifiés pour la résistance à la nodaviriose, les gènes IFI6/IFI27-like et ZDHHC14 sont positionnés à proximités. Pour une application pratique en sélection génétique, le SNP LG12\_8797936 identifié dans un QTL fortement associé à la résistance des bars et situé près du gène IFI6/IFI27-like présentant un fort potentiel pour une sélection assistée par marqueurs. En effet, nous avons pu montrer que les génotypes sensibles et résistants ont un effet similaire sur la survie à la nodaviriose dans des populations indépendantes de bars. La cartographie fine des régions d'intérêts associées à la résistance à des maladies, va permettre de mieux appréhender les phénomènes de résistances des populations sauvages face à l'émergence des pathogènes dans les milieux naturels directement liés à l'augmentation de la température des eaux.

**Javier Belinchon-Moreno** ***Nanopore adaptive sampling et caractérisation de la diversité génétique du NLRome/resistome chez le melon***

Les gènes de résistance de type Nucleotide-binding-site-leucine-rich-repeat (NLR) constituent la plus grande famille de gènes de résistance aux pathogènes et ravageurs chez les plantes, ainsi qu'une des plus diverses. En raison de cette grande diversité, la caractérisation de l'ensemble des gènes de résistance de type NLR d'une espèce (ou NLRome) représente un enjeu majeur pour l'agriculture durable, car elle faciliterait la création de variétés à très large spectre de résistance en maximisant l'utilisation du NLRome. Cependant, malgré les connaissances accumulées, le rôle spécifique de chaque NLR reste largement méconnu chez le melon, en particulier dans les résistances quantitatives. Cette méconnaissance est en partie due à un séquençage/assemblage/annotation qui n'ont pas été satisfaisants avec les technologies short reads du fait de la structure et de l'organisation de ces gènes. Ils sont organisés en clusters qui peuvent être complexes avec un grand nombre d'éléments répétés. Ce fait, associé à l'intra-structure répétitive des gènes NLR eux-mêmes (domaine leucine-rich-repeat), provoque de nombreux changements structurels majeurs dans les clusters de gènes NLR entraînant des polymorphismes présence-absence (PAV). Les méthodes de séquençage à lectures longues peuvent fournir des informations précieuses. Concrètement, le séquençage sélectif proposé par la technologie Nanopore (NAS) est une bonne approche ici, puisque le séquençage du génome entier n'est pas nécessaire. Il permet de réduire la quantité des données à gérer, notamment lorsque l'on travaille avec un grand nombre de

génotypes, tout en enrichissant des régions cibles par rapport à un séquençage standard. Aussi, il s'agit d'une solution facile et économique par rapport à d'autres méthodes de séquençage ciblées qui nécessitent la conception de sondes ainsi que des conditions d'hybridation très strictes. Notre projet poursuit la mise en oeuvre du NAS pour le séquençage et la caractérisation du NLRome chez le melon, une espèce végétale diploïde majoritairement homozygote à haute valeur économique, en utilisant ~150 accessions. Notre objectif final est de réaliser des études d'association génétique entre variabilité phénotypique et variabilité du NLRome sur les 150 accessions. Dans un premier temps, nous avons démontré la performance de l'approche NAS contre WGS pour le séquençage et l'assemblage des clusters NLR précédemment définis dans une variété de melon bien étudiée. En utilisant le kit 14/R10.4.1, un enrichissement moyen de 30X des régions cibles a été observé dans l'expérience NAS par rapport au WGS, permettant de meilleurs assemblages des clusters NLR. Nous avons également évalué la transférabilité de cette méthode à d'autres variétés, et le NLRome d'environ 150 variétés a déjà été séquençé et assemblé en multiplexant 10 accessions per flowcell. Des premiers essais de construction d'un graphe de NLRome ont été faites, en utilisant des outils comme cactus-minigraph et pgg. Dans les prochaines étapes du projet, nous prévoyons de trouver la méthode de construction la plus appropriée pour notre graphe de NLRome, et de l'utiliser comme référence pour faire les études d'association génétique.

**Stéphane Nicolas**

***Genomic selection and Genome-wide association studies on DNA pools identifies promising maize landraces and genomic regions to develop next generation varieties***

Landraces have a large diversity that could help to cope with climatic change and low input agriculture. A novel DNA pooling strategy was implemented to identify promising maize landraces and genomic regions to enlarge the genetic diversity of modern varieties. We assessed the genome-wide diversity of landraces at worldwide and regional levels by genotyping with SNP array the DNA pools of 156 American and European landraces as well as 264 French landraces that have been evaluated for different agronomic traits. They were compared to elite cultivars produced across 20th century, represented by 327 inbred lines. We identified several genomic regions involved in agronomic traits, environmental adaptation, notably in tolerance to abiotic and biotic stresses by either detecting selective footprints or by performing association studies with both environmental variables and agronomic traits. Allelic variation at Vgt3 locus in French landraces is both associated with flowering time variation and altitude in French landraces from Pyrenean Mountain suggesting that Vgt3 lead to adaptation to altitude in Pyrenean Mountain by shortening flowering time. Promising landraces were then identified by estimating their genomic contribution to a collection of 327 inbred lines. Most landraces do not have closely related lines while 10 landraces have a lot and cumulated half of the total contribution to inbred lines. We applied a genomic selection approach to identify promising landraces based on prediction of different agronomic traits. Our approach opens an avenue for the identification of promising landraces for pre-breeding.

## **SESSION 2 - ASSEMBLAGE DE GENOMES, PANGENOMES ET ANALYSE DES SV (1)**

**Stéphanie Sidibe-Bocs**

***Le séquençage haute-fidélité du génome complexe du vanillier apporte une preuve moléculaire de l'endoréplication partielle chez les plantes***

Vanilla planifolia, l'espèce cultivée pour la production d'un des arômes les plus appréciés au monde, est très sujette à la PE. Nos données cytogénétiques ont démontré que la taille de son génome diploïde était de 4,1 Gb, avec 16 paires de chromosomes, bien que des cellules En utilisant du séquençage PacBio HiFi, Illumina et des cartes optiques, nous avons assemblé, échafaudé et phasé en deux haplotypes, un génome diploïde de 3,4 Gb du cultivar 'CR0040'. Les fréquences atypiques L'annotation a révélé que 67 % des gènes étaient présents sur seulement 30 % du génome, qui est également très couvert, et a montré que la partie endorépliquée était riche en gènes. Au contraire, les régions à faible couverture (non-endorépliquées) étaient riches en éléments répétés et pauvres en gènes. En outre, l'assemblage présentait des variations haplotype-spéci Cet assemblage de haute qualité représente 80 % du génome estimé de V. planifolia. En plus d'apporter la première preuve moléculaire de la PE à l'échelle des chromosomes, il constitue une étape clé dans le décryptage de ce génome complexe. Cependant, un séquençage plus profond des parties non-endorépliquées, couplé à de l'Omni-C et à une carte génétique, permettra d'améliorer l'assemblage et l'ancrage; et conduira à une nouvelle version du génome phasé de V. planifolia Enfin, pour soutenir les projets post-génomiques, dont certains seront illustrés, nous avons développé le Vanilla Genome Hub, un portail web intégré et convivial qui permet un accès centralisé aux données génomiques haute-fidélité et aux autres données omiques, ainsi qu'une utilisation interopérable des outils bioinformatiques.

**Alain Vignal**

***Haplotypes et répétitions en tandem dans le génome de l'abeille à miel : où sont les centromères ?***

Les centromères sont le plus souvent composés d'ADN satellite, répétitions en tandem de séquences riches en AT et/ou d'éléments transposables. Malgré des études de cytogénétique et plusieurs assemblages de génomes de référence de haute qualité, basés sur les lectures longues, leur position n'est à ce jour pas clairement définie chez l'abeille à miel *Apis mellifera*.

Afin de tenter de résoudre cette question, nous avons tiré parti de deux jeux de données. Le premier est un ensemble de plus de 600 génomes de faux bourdons (mâles) haploïdes séquençés par la technologie Illumina, présentant l'intérêt d'être naturellement phasé et permettant d'estimer la diversité nucléotidique et le taux de recombinaison le long des chromosomes. Le deuxième est un génome de référence, AMelMel1.1, obtenu lui aussi à partir d'un seul faux bourdon, ce qui facilite grandement l'assemblage d'éléments répétitifs aussi complexes que les satellites alpha.

De manière surprenante, bien que plusieurs familles de répétitions en tandem aient été détectées avec succès dans AMelMel1.1, leur répartition en plusieurs régions chromosomiques ne correspond pas à ce qui est attendu pour des chromosomes monocentriques conventionnels.

Parmi ces répétitions en tandem, une famille majeure, détectée à plusieurs endroits distants de plusieurs Mb sur la plupart des chromosomes, est représentée par une séquence de 371 pb répétée jusqu'à 50 fois en tandem et formant ainsi des clusters couvrant jusqu'à 20 kb. Sa séquence consensus est composée de 28,5 % de GC, seulement légèrement inférieure à la moyenne du génome, qui est de 33 %. De plus, cette famille de répétitions ne semble pas être localisée dans des régions riches en AT ou dans des régions ayant des taux de recombinaison et une diversité nucléotidique spécifiquement faibles, caractéristiques de centromères.

Ces résultats soulèvent des questions sur la structure du génome de l'abeille mellifère, notamment la possibilité qu'il soit composé de chromosomes polycentriques ou holocentriques. Des études cytogénétiques supplémentaires pourraient aider à répondre à cette question.

**Ludovic Duvaux**

### **Testing pangenomic tools for structural variant detection in forest tree species**

Recent advances in the field of pangenomics - e.g. development of tools to build and manipulate pangenome graphs - now allow to detect and genotype structural variants (SVs) at a population scale for a reasonable cost. However, most pangenomics tools were developed with the human genome in mind and it is currently unclear to which extent they can be applied to non-model organisms with different genomic features (e.g. higher genetic diversity, different transposable element content, whole genome duplications and polyploidy) and less resolved genome assemblies. Here, we tested the robustness - to various genomic conditions deviating from the human model - of SV genotyping from short reads mapped on a pangenome graph using vg Giraffe. Using three datasets of previously assembled genomes from species with heterogeneous levels of genome complexity and genetic diversity (a fungi, an apricot tree and a European white oak), we tested the capacity of vg Giraffe to robustly and accurately map reads for individuals with high divergence to the pangenome graph. By simulating reads with increasing mutation rates, we found that vg Giraffe is able to accurately map reads diverging by up to 3% from the graph and that mapping quality remains above 30 for divergence up to 2% confirming that this tool is suitable for forest tree species with high genetic diversity. In further investigations, we will evaluate pangenome graph building tools like minigraph and PGGB and their impacts on pangenome quality and SV genotyping. We will estimate the minimal depth of sequencing required for accurate SV genotyping using either approach and check the SV genotyping accuracy of vg Giraffe. The pipeline and tests used for these analyses are available on gitlab as Snakemake/Singularity workflows.

## **SESSION 3 - ASSEMBLAGE DE GENOMES, PANGENOMES ET ANALYSE DES SV (2)**

**Grégoire Aubert**

### **Séquençage du génome de la féverole : du statut d'espèce orpheline à l'ère de la pangénomique**

La féverole (*Vicia faba*) est une légumineuse à graines cultivée pour l'alimentation humaine ou animale, et ses graines riches en protéines et sa capacité à fixer l'azote atmosphérique en font une culture importante dans les systèmes de culture agroécologiques. Cette culture est cependant confrontée à de nombreux challenges liés au changement climatique et à la diminution de l'utilisation des produits phytosanitaires. Peu d'outils génomiques étaient disponibles pour son amélioration variétale jusqu'à une époque récente mais les nouvelles techniques de séquençage long-fragment couplées à celles des cartes de contact chromosomiques ont permis de lever le verrou du séquençage d'un génome de grande taille (13Gb). Une séquence de référence a été produite pour Hedin/2, ainsi qu'une séquence à plus faible couverture pour un autre cultivar, Tiffany. Ceci a permis d'émettre des hypothèses pour expliquer la grande taille du génome observée pour l'espèce. L'obtention de cette séquence a permis également de développer des outils de génotypage performants, et leur efficacité a été démontrée pour identifier les gènes responsables de la couleur du hile et le déterminisme de la taille de la graine. Ceci ouvre la porte à l'identification du déterminisme génétique de nombreux caractères d'intérêt. Plusieurs projets de séquençage d'autres accessions sont en cours et permettront de développer les approches pangénomiques afin de mieux comprendre les variations qui sous-tendent la diversité de l'espèce.

**Véronique Decroocq**

### ***De la génétique à la génomique chez les arbres fruitiers à noyaux ou comment repousser le champ des possibles chez des espèces orphelines, pérennes***

De par le Monde, une large part de la production fruitière provient d'exploitations agricoles de petite structure, voire de jardins ou de collecte dans les forêts. Néanmoins, cette production revêt une importance socio-économique non-négligeable, notamment dans les pays non-industrialisés. Les fruits représentent à la fois un complément substantiel de ressources mais également ont une valeur nutritionnelle et diététique importante. Comme pour la plupart des espèces fruitières, les programmes de sélection des arbres fruitiers à noyaux (abricotier, pêcher, prunier, cerisier, amandier) doivent à la fois assurer une production saine et constante tout en prenant en compte un environnement en évolution constante (maladies, ravageurs, méthodes culturales). Actuellement, les défis les plus critiques de la filière sont le contrôle des maladies endémiques en vergers (sharka, oidium, monilia etc.), les pertes de production liées à des besoins en froid non-assouvis ou à des gels printaniers post-floraison et enfin la combinaison de ces deux caractères avec une qualité nutritionnelle du fruit préservée. Malheureusement ces caractères agronomiques sont principalement quantitatifs et donc difficilement manipulables dans le cadre de l'amélioration variétale. Les outils de la génomique tels les gènes ouvrent de nouvelles perspectives pour les sélectionneurs, optimisant et accélérant la sélection d'idéotypes, cumulant chacun de ces caractères. A l'UMR BFP, nous combinons les diverses techniques de séquençage haut-débit avec la variabilité génétique et



phénotypique de l'espèce cultivée mais également sauvage afin d'affiner nos connaissances sur le potentiel inexploité des espèces apparentées aux espèces cultivées. Ceci nous a permis de définir des collections de ressources génétiques inégalées au monde, de rationaliser leur gestion et leur conservation et de développer des outils d'analyse de la diversité intra et interspécifique, du nucléotide au variant structural de plus ou moins grande taille. Au-delà de cet aspect descriptif, nous proposons également d'évaluer le rôle de cette diversité génétique dans l'adaptation des espèces fruitières pérennes à plus ou moins long terme. Il est attendu qu'une meilleure connaissance des déterminants génétiques impliqués dans le processus d'adaptation de ces espèces pérennes devrait contribuer de manière significative à la sélection des arbres fruitiers dans un contexte de changement climatique et de modification des pratiques agricoles (agroécologie, réduction des intrants, diversification et innovation pour des marchés de niche, etc.).

**Thomas Faraut**

### ***La variation de structure : dynamique, représentation et impact sur la fonction***

Une variation de structure est définie comme une différence entre génomes qui porte sur au moins 50 nucléotides consécutifs. On comprend aisément que ce seuil de 50 nucléotides n'a pas de fondement biologique. Il résulte de l'approche courante de l'analyse des génomes basée essentiellement sur la technologie de séquençage courte lecture. Les lectures obtenues, d'une taille de l'ordre de 100 nucléotides, ne permettent pas d'identifier simplement les variations au-delà d'une certaine taille. Ce problème disparaît lorsque l'on utilise des longues lectures. Il est alors possible d'étudier simultanément, sans recourir à des méthodes de détection différentes, tout le continuum de variation, de l'indel de quelques nucléotides aux grandes variations de plusieurs kilobases. Grâce aux longues lectures produites dans le cadre du projet SeqOccin, nous avons pu étudier les dynamiques respectives des insertions et délétions sur tout le spectre de taille. Nous montrerons comment l'identification de l'allèle ancestral pour chaque variant permet de différencier les deux types de mutations; insertions et délétions. Cette démarche de reconstruction de l'allèle ancestral nous a permis, en particulier, de mettre en évidence un mécanisme de recombinaison non-homologue dans l'évolution d'une famille d'éléments transposables. On sait depuis peu que ces familles d'éléments transposables jouent un rôle dans l'organisation tridimensionnelle du noyau. C'est d'autant plus remarquable que ces familles, et en particulier les familles actives, diffèrent entre espèces. Nous aborderons, dans la dernière partie de l'exposé, le cas particulier du bovin et les familles d'éléments transposables associées à l'organisation tridimensionnelle du noyau dans cette espèce.

**Jerome Gouzy**

### ***ATLAS : Exploration et exploitation du catalogue d'allèles des espèces cultivées***

Grâce au soutien financier de l'Institut Carnot Plant2Pro l'équipe bioinformatique du LIPME développe l'environnement d'analyse bioinformatique ATLAS. Ce projet vise à décrire le plus précisément possible la diversité des régions fonctionnelles des génomes des espèces cultivées en exploitant les données génomiques des centaines de génotypes disponibles pour ces espèces. A travers un ATLAS d'allèles, la caractérisation plus exhaustive et précise des séquences des allèles des gènes de l'espèce cultivée offre de nombreuses possibilités pour mieux exploiter la diversité. Ainsi, l'identification des séquences d'haplotypes permet de définir les blocks haplotypiques, d'identifier les génotypes introgressés ou les génotypes avec de nouveaux allèles de gènes d'intérêt. Les alignements multiples de séquences d'haplotypes permettent le design de marqueurs moléculaires avec une connaissance très précise de leur portabilité à l'échelle de l'espèce et les nouveaux allèles peuvent être annotés afin de mesurer l'impact des polymorphismes sur les protéines. L'entraînement de modèles basés sur l'apprentissage automatique permet d'identifier les gènes majeurs impliqués dans un trait mais aussi de prédire la valeur de trait pour les génomes séquencés non encore caractérisés. Nous présenterons les premiers résultats de notre approche à travers les ATLAS d'allèles de Tournesol et de Soja construits chacun avec plus de 1000 génomes.

## **SESSION 4 - REGULATION DE L'EXPRESSION DES GENOMES**

**Pascal Martin**

### ***Regulation of gene expression by the BORDER negative transcription elongation factors in Arabidopsis thaliana***

Much progress has been made in our understanding of how transcription elongation is regulated and how it affects gene expression, RNA maturation and chromatin organization. Like nematodes and yeast, plants lack the negative elongation factor (NELF) and do not exhibit widespread promoter-proximal pausing of RNA polymerase II (RNAPII) after transcription initiation. In this work, we characterized the genome-wide function of a new family of negative transcription elongation factors in Arabidopsis thaliana. We named the 3 members BorDeRs (BDR1, 2 and 3) because they are enriched, together with RNAPII, at evolutionarily-conserved, nucleosome-depleted regions located at the borders of a large fraction of genes across the genome. RNA-seq studies defined the repertoires of genes with modified expression levels in bdr triple mutant plants (bdr1,2,3), revealing impacts on flowering time and defense responses. Unexpectedly, we also found that, in bdr1,2,3 plants, a large fraction of genes appeared downregulated as a result of transcriptional interferences. In these mutants, defective transcriptional termination and 3' RNAPII pausing at the upstream tandem gene generates cis interference with transcriptional initiation at the promoter of downstream genes, thus inducing their downregulation. Altogether, our data illustrate how BDR negative elongation factors have opposite effects on gene expression by influencing different steps of the transcription cycle and strikingly play a crucial role in limiting transcriptional interferences at closely-spaced tandem genes.

**Aurore Coissac**

***Role of DNA methyltransferases of *Ralstonia pseudosolanacearum* (GMI1000 strain) in DNA methylation, virulence and growth***

Epigenetic modifications are heritable changes in gene expression that occur without changes in the DNA sequence. They are involved in numerous phenotypic traits in bacteria. The most common form of epigenetic modification is DNA methylation. In bacteria, DNA methylation is incorporated by DNA methyltransferases (MTases) on an adenine at the sixth position (6mA), or on cytosine at fourth position (4mC) or fifth position (5mC). *Ralstonia solanacearum* is a soil born proteobacteria, responsible for bacterial wilt disease on more than 250 plant species. SMRT sequencing of the *R. solanacearum* GMI1000 strain identified seven MTases, one associated with a restriction enzyme. The others MTase are six orphan MTases. SMRT-seq also identified methylation motifs targeted by two MTases; the 6mA modification of GTWWAC motif catalyzed by the *damB* MTase and the 5mC modification of YGCCGGCR motif catalyzed by the *dcmB* MTase. Simple and double deleted mutants of MTases were generated in the GMI1000 strain. SMRT-seq confirmed that the GTWWAC motifs were not methylated in the *damB* deleted mutant and EM-seq confirmed that the YGCCGGCR motifs were not methylated in the *dcmB* deleted mutant. Simple deletion didn't affect the growth and the virulence of the bacteria but three double deleted mutants were strongly affected. Unfortunately, no new motif could be identified. RNA-seq analysis revealed the activation of the SOS response in two double deleted mutant suggesting an impact of MTase deletion on DNA damages.

**Stéphane Maury**

***Evolutionary and functional impact of epigenetic variations in forest trees facing climate change (EPITREE)***

Forest die-off is reported all around the world due to heat and drought stress episodes. As fixed and long living organisms subjected to repeated environmental stresses, trees have developed mechanisms enabling them to cope with fluctuating environmental conditions. We proposed epigenetics as a hub of integration in meristematic cells linking developmental and hormonal responses with changing environment. This was based on our studies in the shoot apical meristem of poplar under water deficit conditions (Lafon-Placette et al., 2018; Le Gac et al., 2018; Sow et al., 2018a; Le Gac et al., 2019; Maury et al., 2019; Vigneaud et Maury, 2020; Sow et al., 2021; Vigneaud et al., 2023). Our objective is now to evaluate how DNA methylation can significantly participate to plasticity and adaptation in natural populations of trees facing climate change to test potential uses for trees breeders and forest managers (Sow et al., 2018b; Amaral et al., 2020; Kakoulidou et al., 2021). Here, I will focus on the strategy (Lesur et al, in prep) and methodology (Dugé de Bernonville et al., 2022; Agius et al., 2023) to study DNA methylation in plants. This study is part of the EPITREE ANR project and EPICATCH COST action aiming at defining the role of epigenetic mechanisms for crop adaptation to climate change.

**Hervé Acloque**

***INRAE contribution to the functional annotations of pig and chicken genomes as part of the European GENE-SWitCH project***

The global Functional Annotation of Farm Animal Genomes initiative (FAANG) aims to improve animal breeding by improving genomic prediction by integrating functional genomics information. The GENE-SWitCH project has produced extensive functional annotations for seven tissues at three developmental stages (early and late organogenesis and newborn) for both chickens and pigs. These datasets have been processed and analysed to produce both tissue and time-point specific transcript, gene, and regulatory annotation.

This presentation will highlight the aims and the main results of the project with a particular focus on those resulting from the contribution of INRAE to the functional annotation of pig and chicken genomes.

**SESSION 5 - METAGENOMIQUE**

**Sylvie Combes**

***Piloter l'installation du microbiote des animaux d'élevage : les technologies de séquençage pour explorer les communautés bactériennes***

Les relations symbiotiques qui s'établissent au début de la vie entre la communauté microbienne qui colonise l'intestin et l'hôte sont déterminantes pour la construction et la préservation de la santé du jeune mammifère. Le microbiote intestinal participe à la métabolisation de certains aliments et à la production de composés essentiels. Il joue un rôle dans la défense directe et médiée par l'hôte contre les agents pathogènes. Son action sur la différenciation et la maturation du système immunitaire muqueux et systémique ont également été mise en évidence. Chez les animaux d'élevage, optimiser la séquence d'implantation et le fonctionnement du microbiote intestinal est une stratégie prometteuse pour préserver la santé et limiter ainsi le recours aux intrants médicamenteux. Dans ce cadre, les technologies de séquençage ont permis des avancées remarquables pour (i) la caractérisation taxonomique du microbiote intestinal, (ii) la compréhension des facteurs structurant ces communautés, (iii) et leur fonctionnement. Les avancées futures en technologies de séquençage, mêlées à des approches fonctionnelles, permettront de progresser sur la connaissance des mécanismes permettant l'union sélective d'un hôte et de son microbiote, clé adaptative de l'holobionte.

**Paola Fournier**

***Vers une viticulture sans pesticide: analyse du microbiome de la vigne et des sols viticoles pour développer le biocontrôle microbien de l'oïdium et du mildiou***

Les vignes sont colonisées par une myriade de micro-organismes qui interagissent entre eux et avec leur hôte. L'ensemble de ces micro-organismes constitue le microbiote de la vigne et comprend des bactéries, des champignons et des oomycètes. Certains sont des phytopathogènes comme l'oomycète *Plasmopara viticola*, qui provoque le mildiou de la vigne. Lorsque ce pathogène arrive sur la feuille de vigne, il entre en contact avec le microbiome résident, qui peut agir comme une barrière à sa croissance. Notre objectif est d'identifier les membres de ce microbiote qui régulent naturellement le mildiou. Afin de cibler au mieux ces taxons microbiens bénéfiques, nous avons adopté une approche de caractérisation des communautés microbiennes par metabarcoding à l'échelle du tissu foliaire (sain versus infecté) et de la parcelle (résistante versus sensible). Grâce à des analyses d'abondances différentielles, nous avons identifié des candidats microbiens potentiels pour le biocontrôle du mildiou de la vigne. Cette approche est prometteuse pour mettre en évidence des consortiums microbiens d'intérêt, dont la dynamique et les interactions devront ensuite être étudiées expérimentalement pour développer le biocontrôle microbien du mildiou et faire évoluer la viticulture vers une viticulture sans pesticides.

**Gaëlle Marliac**

***Le microbiome du sol et la régulation d'une maladie du blé***