

SeqOccln

Séquençage Occitanie Innovation

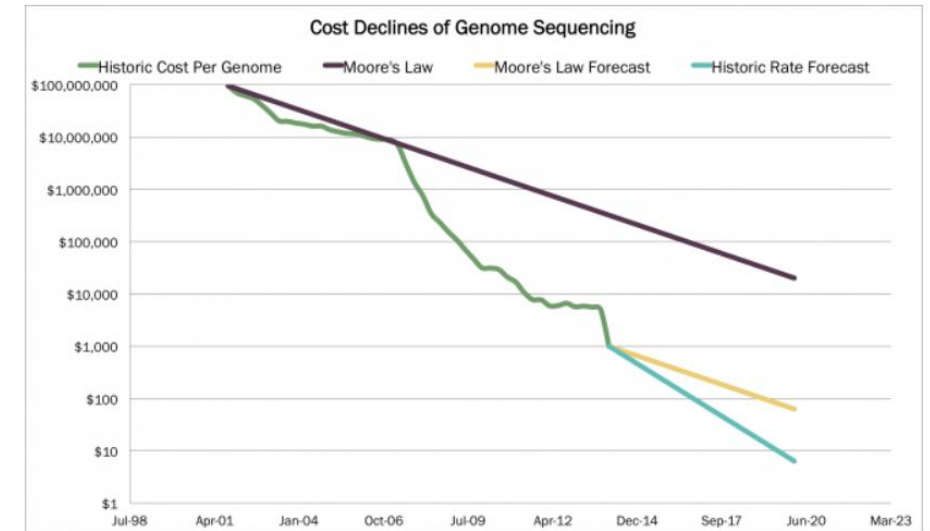
« Pour une meilleure connaissance des génomes »

Cécile Donnadieu, Christine Gaspin, Denis Milan
Journée d'ouverture 18 avril 2019

Le contexte du projet

Un domaine très évolutif

- Un appel à projet de la région Occitanie dans le cadre d'un programme de financement FEDER – FSE pour des projets de collaboration entre entreprises et structures de recherche mutualisées pour des projets d'innovations
- La génomique, une **source majeure d'innovation** pour la recherche publique comme pour les entreprises
- La nécessité de **développer davantage l'expertise bio-informatique** et bio statistique indispensable à ces analyses
- Un projet porté par **GeT-PlaGe** (Génome et Transcriptome, US1426), plateforme nationale stratégique INRA et membre du PIA **France Génomique** & par **Bioinfo Genotoul**, plateforme nationale stratégique INRA et membre du PIA **Institut Français de Bioinformatique**



Les objectifs

Acquisition d'expertises avancées sur l'utilisation des séquenceurs 3^e génération

- Acquérir une expertise avancée sur les **nouvelles technologies de séquençage « Longs fragments »** et « **Molécules uniques** » à haut débit disponibles :

- Potentiels comparés des technologies
- Identification des combinaisons de technologies à mettre en œuvre en fonction des objectifs



... à trois niveaux d'étude :

- Axe 1 Le **génom**e : Connaissance de la variabilité du génome (ponctuelle ou structurale)
- Axe 2 L'**épig**énom : Etudes des marques épigénétiques de régulation de l'expression du génome
- Axe 3 Les **métag**énom : Connaissance fine des communautés au sein d'un métagénom

- Savoir conseiller les équipes publiques et privées pour l'utilisation des nouvelles technologies en fonction des objectifs recherchés



Résultats attendus

Une expertise reconnue et une visibilité accrue des plateformes GeT-PlaGe et Bioinfo

- **Une expertise avancée sur l'utilisation des séquenceurs Longs fragments molécules uniques pouvant être proposée :**
 - Aux équipes académiques
 - ➔ Par **GeT-PlaGe & Genotoul Bioinfo** pour des projets de recherche (éventuellement avec partenariat Public Privé)
 - Aux entreprises privées (notamment sélection animale et végétale)
 - ➔ Par la Business Unit **GeT-IT** (INRA Transfert) pour des prestations de routine
- La **production de jeux de données de référence distribués en open data** à l'issue du projet, contribuant à **créer une communauté de développeurs bioinfo** autour des plateformes GeT-PlaGe et Bioinfo (notamment équipes bioinfo France Génomique et Institut Français de Bioinformatique)
- Une plus grande visibilité de GeT & Bioinfo dans le paysage national et international

Les partenaires

Un partenariat monté pour renforcer le tryptique vertueux : *Plateforme – Equipe recherche – Partenaires privés.*

- **Plateformes :**

- Plateforme **GeT-PlaGe** (Génome et Transcriptome) de Genotoul
- Plateforme **Bioinfo** de Genotoul



- **Equipes de recherches partenaires :**

- Unité **GenPhySE** (UMR1388, GA-PHASE, Toulouse)
- Unité **MIAT** (UR875, MIA, Toulouse)
- Unité **GABI** (UR1313, GA-PHASE, Jouy en Josas)
- Unité **GQE Le Moulon** (UMR0320, BAP, Versailles)



- **Partenaires privés : 15 signataires rassemblant 25 entreprises**

- **Sélection animale** : Apis Gene (bovin & caprin), CNBL (ovin), Alliance R&D (porc), SYSAAF (caille)
- **Sélection végétale** : Biogemma, Limagrain, RAGT, Euralis, Maisadour, Caussade Semences, KWS, Syngenta = **PIA Amaizing (maïs)**
- **Nutrition** : Lallemand, Adisseo
- **Start up développement automate** : Adelis

- **Une double volonté dans le montage du partenariat :**

- Embarquer des partenaires privés uniquement en association avec une équipe de recherche de l'INRA, pour aller plus loin dans l'analyse des données et renforcer le lien **Plateforme – Equipe Recherche INRA – Partenaires Privés**
- Par espèce : privilégier la mise en place d'une **dynamique nationale** autour des acteurs professionnels en **Occitanie**, éligibles pour des actions région. Exemple :
 - **Apis Gene** fédère toutes les principales entreprises de sélection bovine et caprine françaises : Evolution, Genes Diffusion, Auriva, Crealim, Elitest, Umotest, BGS, Capgenes
 - Alliance R&D fédère les partenaires de la sélection porcine : Axiom, Nucleus, Choice Genetics
 - Tous les partenaires privés **d'Amaizing** (sélection maïs) sont participants et présents en région

Quelques chiffres

Un financement Région/Europe essentiellement tourné vers du fonctionnement pour acquérir de l'expertise

- Un projet sur **3 ans** : 01/2019 – 12/2021
- Un budget total de **6 M€** :
 - **4.8 M€** FEDER / Région Occitanie
 - **0.77 M€** financement en cash des partenaires privés
 - **0.43 M€** apport INRA en personnels mis à disposition du projet
- Répartition du budget (porté en totalité par GeT) :
 - **1 M€** investissement matériels (dont 0.6 M€ stockage informatique sur PF Bioinfo)
 - **1.9 M€** personnel (1.57 M€ CDD et 0.43 M€ Ingénieurs INRA)
 - **3.1 M€** réactifs et overheads
- Engagement INRA :
 - Mise à disposition des locaux et infrastructures des PF GeT et Bioinfo
 - **4.95 ETP** au total (1.65 x 3) des PF comptés dans le projet
 - Participation des scientifiques des équipes de recherche à l'analyse des données mise à disposition de la recherche (toutes les données du projet)

Le projet plus en détail...

Pilotes : Christine Gaspin, Denis Milan, Cécile Donnadiou

Coordination technique : Carole Iampietro

Listes des participants, Missions et Responsabilités des participants

- **3 Porteurs en charge de :**
 - Donner du sens,
 - Déterminer les objectifs et faire les choix stratégiques,
 - Assurer le suivi et animer l'équipe projet
- **Des responsables d'axes en charge de :**
 - Décider des actions à mettre en œuvre pour atteindre les objectifs et livrables définis dans le projet - Identifier les verrous et proposer des évolutions pertinentes
 - Animer les réunions bimestrielles
 - Rédiger les bilans
- **1 coordinateur technique en charge de :**
 - Suivre et encadrer les biologistes,
 - S'assurer du respect des délais et des budgets
 - S'assurer du transfert d'informations entre les différents acteurs internes
- **Des encadrants en charge de :**
 - Recruter et encadrer les personnels CDD
 - Vérifier et valider le cahier de manips ou outils de suivi bioinfo
 - Identifier les verrous et proposer des évolutions pertinentes
- **Des Experts**
 - Apporter leur expertise technique et/ou scientifique

Listes des participants, Missions et Responsabilités des participants

Les scientifiques et Experts associés

Scientifiques et Experts associés			
Axe 1	Nathalie Iannuccelli	Biologiste	Genphyse
	Juliette Riquet	Scientifique	Genphyse
	Gwenola Tosser	Scientifique	Genphyse
	Didier Boichard	Scientifique	GABI
	Cécile Grohs	Biologiste	GABI
	Mekki boussaha	Bioinfo	GABI
	Clementine vitte	Biologiste	GQE_Le moulon
	Alain Charcosset	Scientifique	GQE_Le moulon
	Harry Belcram	Bioinfo	ABI_Le moulon
	Johann Joets	Bioinfo	ABI_Le moulon
	Cédric Cabau	Bioinfo	SIGENAE
	Philippe Bardou	Bioinfo	SIGENAE
	Céline Roques	Biologiste	GeT-PlaGe
	Axe 2	Frederique Pitel	Scientifique
Sophie Leroux		Biologiste	Genphyse
Nathalie Iannuccelli		Biologiste	Genphyse
Juliette Riquet		Scientifique	Genphyse
Julie Demars		Scientifique	Genphyse
Didier Boichard		Scientifique	GABI
Cécile Grohs		Biologiste	GABI
Axe 3	Sylvie Combes	Scientifique	Genphyse
Axe 4 Transversal	Marie-Stephane Trotard	Informaticienne	SIGENAE
	Partice Dehais	Bioinfo	SIGENAE

Axe Transversal

Gerald Salin & Céline Noirot

Encadrants : Gerald Salin, Carole Lampietro

2 IE Développeurs info Alexandra Korovina, Amandine Broha

1 IR Biologiste Erwan Denis

Traçabilité, Mise à disposition des données, Management des projets,

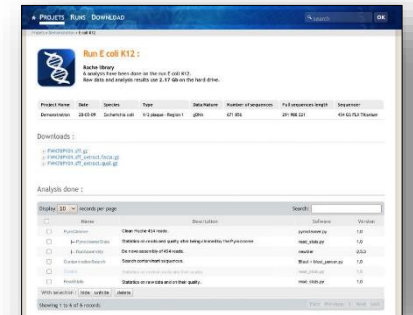


● Développement de l'Infrastructure

- Evolution du **Back office**: continuer le développement d'un LIMS de suivi des échantillons et des manipulations
- Evolution du **Front office**: développer (ou améliorer) l'interface de mise à disposition des données (NG6)
- Mise en place d'un plan de gestion des données dans le cadre de la valorisation des données et de l'Open data

● Installation d'une baie de stockage (600k€)

- Responsables Didier Laborie & Marie Stephane Trotard



Axe transversal

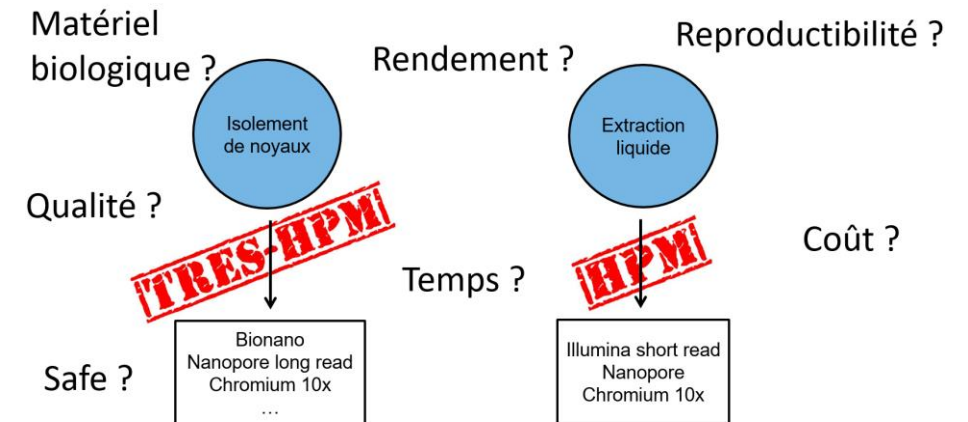
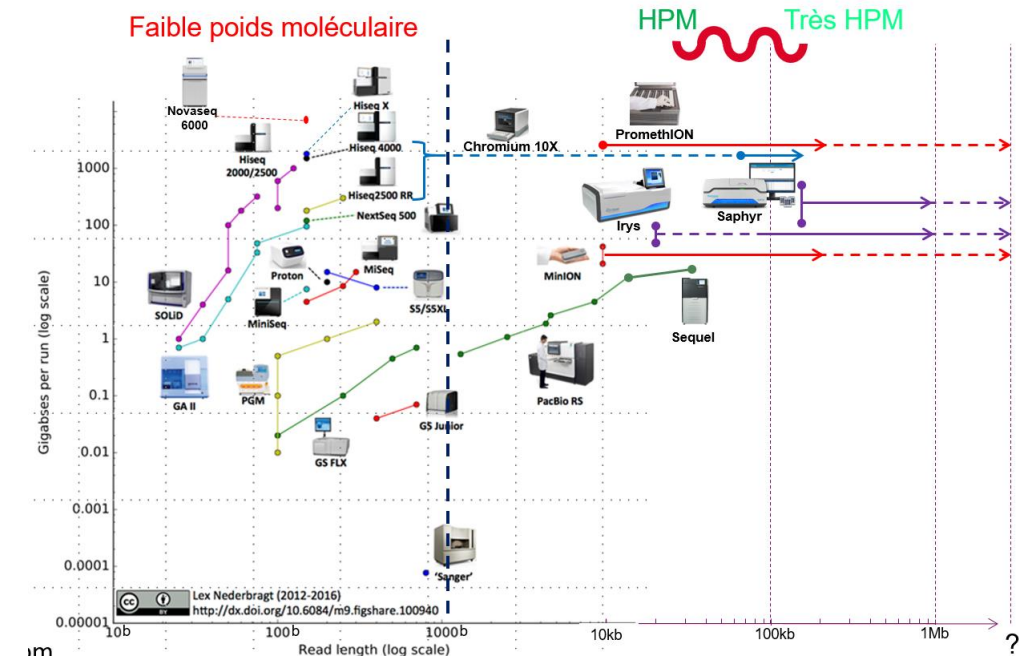
Extraction d'ADN de haut poids moléculaire

● Phase 1 (2019-2020)

- Développements pour réaliser des extraction d'ADN de haut poids moléculaire
 - Sang
 - Sperme
 - Tissus : oreilles, queue...
 - Echantillons environnementaux : fèces

● Phase 2 (2020-2021)

- Développements pour réaliser des extraction d'ARN pour de l'analyse Long Fragment



Axe 1 : Le Genome

Claire Kuchly & Christophe Klopp

Encadrants : Carole Lampietro, Christophe Klopp, Thomas Faraut, Mathias Zytnicki
1 IE bio Camille Eche, 1 IE Bioinfo (en attente de recrutement),
2 IR Developpeurs Bioinfo Sarah Djebali - Andreea Dréau

Axe 1 : Vers une meilleure connaissance de la variabilité des Genomes

Variations ponctuelles ou de structure

● Phase 1 (2019 – 2020)

- Identification de la meilleure combinaison de différentes technologies pour réaliser le meilleur assemblage possible en tenant compte du coût
 - L'assemblage "NRgene" comme référence – **150k€ for 3Gb genome**
 - Chromium ou mate pair short read
 - HI-C avec différents protocoles
 - Oxford Nanopore long reads
 - Pacific Bioscience long reads ?
 - Bionano - optical mapping ?
 - Autres technologies pas encore existantes
- Les 1^{er} tests sur le Bovin et le Maïs Les validations sur un plus grand nombre d'individu sur le Bovin, Maïs, Porc, la Chèvre et le Mouton
- Identification et/ou Développement de nouvelles méthodes pour l'assemblage de novo et l'identification de variants de structure

1 échantillon (3Gb) <10k€ de réactifs

NRGene SOLUTIONS

DeNovoMAGIC™

NRGene's first and most popular platform, DeNovoMAGIC™, can assemble virtually any genome, even the most challenging assembly within days of obtaining the raw sequencing data.

Parameter	An. tauschii (wheat ancestor)	Bovine	Maize
Fold coverage of short reads (PE & MP)	x200	x200	x180
Scaffold assembly N50 (L50)	11.4 Mbp (106)	38.9 Mbp (22)	35.5 Mbp (18)
Scaffold assembly N90 (L90)	2.27 Mbp (405)	8 Mbp (74)	11 Mbp (58)
Total assembly size	4.09 Gbp	2.71 Gbp	2.13 Gbp
Unfilled gaps	1.39%	1.77%	1.9%

Axe 1 : Vers une meilleure connaissance de la variabilité des Genomes

Variations ponctuelles ou de structure

● Phase 2 (2020-2021)

- Développement de nouvelles méthodes de Genotypage par Séquençage : pour une meilleure connaissance des génomes, pour des prix plus attractifs
 - Dépendant de la Phase 1
 - Analyse de Pan genome
 - Enrichissement...

● Phase 3 (2021-2022)

- Développement de nouvelles méthodes de Genotypage ciblé
 - En fonction des innovations du domaine

Axis 2 : L'Epigenome

Céline Vandecasteele & Christophe Klopp

Encadrants : Carole Iampietro – Christophe Klopp

1 IE bio Remy Serre, 1 IE bioinfo Paul Terzian

Axe 2: Etude des marques Epigénétiques dans la régulation de l'expression

Identification des marques de methylations

● Phase 1 (2019-2020)

- Développement de méthodes biologique et bio-informatique pour identifier les marques de methylation sur de l'ADN natif: 5mC, 6mA
- Premiers tests sur la caille et le porc

1 échantillon (1Gb) <2500€ réactifs

● Phase 2 (2020-2021)

- Développement de méthodes pour identifier les marques de methylation sur l'ARN

● Phase 3 (2021-2022)

- Développement de nouvelles méthodes d'analyse des marques de methylation ciblées
 - En fonction des innovations du domaine

Axe 3 : Le Métagénomome

Olivier Bouchez & Claire Hoede & Geraldine Pascal

Encadrants : Carole Iampietro, Claire Hoede, Geraldine Pascal

1 IE bio Adrien Castinel – 1 IE bioinfo Jean Mainguy

Axe 3 : Connaissance détaillée des communautés au sein d'un échantillon environnemental

Métagénome

- **Phase 1 :**
 - Développement d'une Approche ciblée amplicons longs fragments
 - **In silico : Identification de nouveaux marqueurs de bio-diversité**
 - Evaluation de l'approche long fragment pour l'analyse de régions cibles
- **Phase 2 :**
 - Evaluation de l'approche long fragment pour l'analyse **Whole metagenome**
 - Evaluation d'autres technologies pour l'amélioration de l'assemblage des génomes : HI-C, Chromium...

Axe 1, 2 et 3 Points-clefs (réunions, livrables)

- L1.1 Une revue d'axe bi trimestrielle
- L2.1 L'acquisition d'une expertise formalisée par la rédaction de protocoles à la fin de chaque développement
- L3.1 La publication des méthodologies développées à la fin de chaque phase (cf partie calendrier détaillé)
- L4.1 Un pipeline d'analyse de données à la fin de chaque phase qui sera incrémenté dans l'axe transversal
- L5.1 La publication des résultats scientifiques en lien avec les développements méthodologiques
- L6.1 La mise à disposition des partenaires les données au fur et à mesure de leur production
- L7.1 La mise à disposition de la communauté en open data

Axe Transversal Points-clefs (réunions, livrables)

- L8. Développement d'un automate d'extraction d'ADN HPM à partir de sang avec la société Picometrics
- L9. Document de synthèse permettant de déterminer le meilleur protocole d'extraction d'ADN et d'ARN HPM
- L10. Plan de gestion durable des données mise en place d'un nouveau processus
- L11. Nouvelle infrastructure de mise à disposition des données (incrémenté)
- L12. Nouveau système LIMS (incrémenté)
- L13. Nouveau système de suivi de projet et notamment des temps personnels

Gouvernance Points-clefs (réunions, livrables)

- Une réunion de lancement (kick of meeting)
- Une réunion annuelle
- Une réunion semestrielle (six mois après la réunion annuelle) : Revue de projet
- Un séminaire final
- Des actions de communication afin de faire connaître le projet et ses avancées

	2019		2020		2021		2022	
	1er semestre	2ème semestre	1er semestre	2ème semestre	1er semestre	2ème semestre	1er semestre	2ème semestre
L1.1	X	X	X	X	X	X	X	X
L2.1		X		X		X		X
L3.1		X		X		X		X
L4.1		X		X		X		X
L5.1						X	X	X
L6.1				X		X		
L7.1								X
L8						X		
L9			X	X	X	X		
L10	X							
L11			X	X	X	X	X	X
L12			X	X	X	X	X	X
L13			X					
Reunion lancement	X							
Reunion annuelle			X		X			
Reunion semestrielle		X		X		X		
Seminaire final							X	
Communication	X	X	X	X	X	X	X	X

- **Ce qui est financé:**
 - 12 CDD sur 3 ans (1,5M€)
 - 12 Ordinateurs pour les CDD
 - Réactifs (2,7M€) : 2M€ Axe1, 450k€ Axe2, 200k€ Axe3, 90k€ Extraction
 - Equipements (1M€) : 200X3 K€ Baie de stockage, 400k€ Equipements de préparation ou de validation des échantillons

- **Ce qui n'est pas financé (à prendre sur Over head 290k€ ou sur autres financements):**
 - L'organisation des journées annuelles,
 - Les déplacements, participations à des congrès, formations,
 - Les frais de publication,
 - Les acquisitions de nouvelles technologies de Séquençage,
 - Les manip que nous n'avons pas encore imaginées...

Comment voit-on que l'on progresse?

- **Des réunions bimestrielles par Axe - Responsable d'axe, CDD, Encadrants, Pilotes et Scientifique Référent**
 - Bilan de la période écoulée réalisé par l'équipe projet
 - Fixer les objectifs et s'assurer de leur suivi
 - Identification des verrous et Définition des nouvelles actions à mettre en œuvre
- **Des réunions bi mensuelles pour les porteurs**
 - Définition et Coordination de l'ensemble des actions (hors technique) à mettre en œuvre
 - Suivi des objectifs (hors technique)
 - Prendre des décisions si verrous majeurs, demande particulière, nouvelles orientations stratégiques...
- **Des réunions annuelles avec l'ensemble des membres du projet**
 - Bilan de l'année écoulée
 - Avancée du projet et des livrables
 - Décider des actions phares à mettre en œuvre

Remerciements

- Les équipes des SDAR du centre INRA
 - Gilles Bariteau, Claudine Derras, Estelle Laffont, ...
- Le Service Soutien à la recherche
 - Frederic Pileur, Marie-Laure Pouilles, Nolwenn Gourmelen, Benoît Mingam, ...
- Les Gestionnaires des unités MIAT et GeT-PlaGe
 - Anne Fleurbe, Dounia ben Salah, Melanie Delbancut, Fabienne Ayrignac, Nathalie Julliard

