



**HAL**  
open science

# Éléments transposables, dynamique des génomes et des épigénomes : un voyage au cœur de la variabilité structurale et fonctionnelle

Clémentine Vitte

## ► To cite this version:

Clémentine Vitte. Éléments transposables, dynamique des génomes et des épigénomes : un voyage au cœur de la variabilité structurale et fonctionnelle. Sciences du Vivant [q-bio]. Université Paris Saclay, 2024. <tel-04919781>

**HAL Id: tel-04919781**

**<https://hal.science/tel-04919781v1>**

Submitted on 29 Jan 2025

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 CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Éléments transposables, dynamique  
des génomes et des épigénomes :  
un voyage au cœur de la variabilité  
structurale et fonctionnelle

**Habilitation à diriger des recherches  
de l'Université Paris-Saclay**

**présentée et soutenue à Gif-sur-Yvette, le 12/12/2024,  
par**

**Clémentine VITTE**

**Composition du jury**

<b>Pascale LESAGE</b> DR CNRS, Université Paris-Cité, France	Rapportrice
<b>Josep CASACUBERTA</b> Group leader, CRAG, Espagne	Rapporteur
<b>Vincent CASTRIC</b> DR CNRS, Université de Lille, France	Rapporteur
<b>Séverine CHAMBEYRON</b> DR CNRS, Université de Montpellier, France	Examinatrice
<b>Cécile RAYNAUD</b> DR CNRS, Université Paris Saclay, France	Examinatrice
<b>Philippe GALLUSCI</b> Professeur, Université de Bordeaux, France	Examinateur



## Table des matières

<b>Remerciements .....</b>	<b>2</b>
<b>Curriculum vitae .....</b>	<b>6</b>
<b>Avant-propos.....</b>	<b>15</b>
<b>Activités de recherche .....</b>	<b>17</b>
<b>1- Introduction générale.....</b>	<b>17</b>
<b>2- Contribution des éléments transposables à la variation de la taille des génomes .....</b>	<b>20</b>
2.1 Introduction.....	20
2.2 Contexte scientifique.....	20
2.3 Dynamique du génome au sein du genre <i>Oryza</i> .....	21
2.4 Dynamique du génome du maïs.....	28
2.5 Eléments transposables et taille de génome au sein des Angiospermes .....	29
2.6 Structure des génomes et dynamique des éléments transposables chez les Asparagales (oignon et asperge) .....	30
2.7 Lien entre génomique structurale et génomique fonctionnelle.....	32
<b>3- Caractérisation de la variabilité structurale intra-spécifique et lien avec l'expression des gènes .....</b>	<b>34</b>
3.1 Contexte scientifique.....	34
3.2 Dynamique intra-spécifique de régions ciblées.....	35
3.3 Analyses de régions spécifiques d'un génotype .....	36
<b>4- Epigénome : des stress abiotiques à la spécificité tissulaire .....</b>	<b>43</b>
4.1 Contexte scientifique.....	43
4.2 Analyse de la méthylation du gène <i>VRNA-1</i> en réponse au froid par une approche ciblée .....	44
4.3 Analyse de la variation de méthylation de l'ADN entre lignées de maïs à l'échelle du génome par l'approche MeDIP-seq.....	46
4.4 Impact de l'environnement sur la méthylation du génome du maïs.....	46
4.5 Eléments <i>cis</i> -régulateurs impliqués dans la tissus-spécificité chez le maïs : contribution des éléments transposables.....	49
<b>5- Trajectoire et perspectives .....</b>	<b>52</b>
5.1 Contribution de la méthylation de l'ADN à l'adaptation du maïs : rôle des éléments transposables ? .....	53
5.2 Contribution des éléments transposables à la variabilité des éléments <i>cis</i> -régulateurs du maïs ..	54
5.3 Contribution des éléments <i>cis</i> -régulateurs à l'évolution phénotypique à court terme .....	55
5.4 Contribution des éléments transposables à l'évolution des espèces .....	56
<b>6- Faire vivre la science .....</b>	<b>58</b>
6.1 Former des étudiants .....	58
6.2 Diffuser les connaissances.....	59
6.3 Transmettre le goût de la science et de la recherche .....	61
6.4 Promouvoir des interactions entre chercheurs .....	61
6.5 Produire une science de qualité .....	62
6.6 Recruter les chercheurs de demain .....	63
<b>7- Regard sur mon parcours .....</b>	<b>64</b>
<b>Références bibliographiques.....</b>	<b>65</b>

# Remerciements

*merci* aux membres de mon jury d'avoir accepté d'expertiser mon travail malgré leur agenda déjà chargé. J'espère que la lecture de ce manuscrit et les discussions que nous aurons lors de la soutenance seront l'occasion de resserrer nos liens, et ouvriront des pistes de collaboration.

*merci* à tous les étudiants et CDD qui ont fait le choix de venir rejoindre mes projets, et qui les ont fait avancer. Ça a été un plaisir pour moi de travailler avec vous, et j'espère vous avoir transmis ma passion de la recherche, ainsi que des connaissances et compétences qui vous servent encore aujourd'hui.

*merci* aux personnels techniques du Moulon. Sans vous, rien de tout cela n'aurait été possible. Beaucoup de personnes se sont succédées et je n'ai pas la place de toutes les citer ici... Un grand *merci* à Agnès Rousselet, Anthony Venon et Harry Belcram pour votre ténacité face aux centaines d'extractions d'ARN du « genes atlas » Amaizing, à Martine Le Guilloux pour ton aide dans mes différents projets, et à Carine Remoué pour te lancer aujourd'hui dans l'aventure du SAM-seq. *merci* aussi à Valérie Combes et Delphine Madur pour votre présence jamais lointaine et vos conseils précieux, ainsi que pour votre aide dans le programme SeqOccin, et à Fabrice Dumas pour ton aide dans le programme Amaizing. *merci* à Valérie Lespinas pour l'organisation périlleuse des missions du « gene atlas » Amaizing, et à Sandrine Le Bihan pour la gestion de mes missions dans les méandres des outils de la suite Etamine, et de l'organisation des missions des membres de ce jury. *merci* à Nathalie Galic pour ton implication dans la mise en place des cultures des projets SATURN et EpiFlorE, et à Didier Tropée pour m'avoir initiée à la culture du blé. Enfin, *merci* à Adrien Falce et Benoît Johannet pour la mise en place des logiciels et la sécurisation de nos précieuses données.

*merci* aux plateformes de génomique pour leur soutien aux projets innovants et de grande ampleur, en particulier au CNRGV (Hélène Bergès et William Marande) qui a été présent tout au long du projet Amaizing, à GetPlage pour le projet SeqOccin (Carole Iampietro et tous les acteurs du projet) et à POPS (Ludivine Soubigou-Taconnat et Véronique Brunaud) pour votre implication dans le séquençage des ARN du « gene atlas » Amaizing.

*merci* aux plateformes de culture et de phénotypage, en particulier PhenoArch pour la culture en conditions hydriques contrôlées (Llorenç Cabrera-Bosquet), l'UE de Mons-en-chaussée pour la culture froide (Hélène Sellier), et l'UE de Saint-Martin de Hinx (Carine Palaffre) pour la culture en champ et les prélèvements complexes dans un endroit sans cuve d'azote.

*merci* à mes collègues de l'équipe GEvAD (anciennement GEAR puis DyGAP !) pour vos retours enthousiastes sur mes travaux tout au long de ces années. Un merci plus particulier à Karine Alix, Maud Tenailon, Johann Joets et Maud Fagny pour votre implication active dans mes projets et pour le co-encadrement d'étudiants.

**merci** à mes collègues du Moulon, pour les collaborations passées et futures, et pour la complémentarité des recherches qui y sont menées et qui procurent un environnement multifactoriel. Un merci particulier à Alain Charcosset, pour m'avoir donné l'opportunité de coordonner le WP3 du projet Amaizing, et à Isabelle Goldringer et Jérôme Enjalbert pour m'avoir permis de co-encadrer mon premier doctorant. Merci à Elodie Marchadier pour l'opportunité de cette nouvelle aventure autour du projet EpiFlore, et à Diala Abu-Awad pour notre coordination joyeuse de l'axe 'Evolution' de l'unité (et les soirées karaoke !).

**merci** à mes collègues du programme Amaizing qui m'ont fait découvrir leurs questions biologiques autour du maïs, et en particulier aux collègues du LEPSE (Olivier Turc, Claude Welcker, François Tardieu), de l'unité AgroImpact (Catherine Giauffret), du RGP (Peter Rogowsky) et de l'IJPB (Sylvie Coursol) pour m'avoir initiée au développement et à la physiologie du maïs et pour leurs yeux pétillants quand je leur ai proposé mes idées.

**merci** à mes collègues de Biogemma qui ont participé au WP3 du programme Amaizing. Un merci particulier à Jean-Philippe Pichon qui avait commencé cette aventure avec moi, et à Magalie Leveugle et Hervé Duborjal qui ont pris sa relève. Merci aussi à Jorge Duarte pour nos discussions sur les assemblages de séquences génomiques de maïs.

**merci** à Leandro Quadrana (IPS2), Hadi Quesneville et Johann Confais (URGI) de vous lancer à l'aventure d'une ANR avec moi. J'espère que SATURN nous fera atteindre les étoiles !

**merci** aux rapporteurs de mes projets, qui ont vu l'intérêt de mes idées et m'ont permis de les réaliser, et aux rapporteurs de mes articles, qui ont vu l'intérêt de mes travaux et m'ont permis de les améliorer.

**merci** à Denis Poinot, Hervé Merçot, Virginie Raynal, Deborah Charlesworth, Hervé Vaucheret et Frédéric Bocard pour avoir vu en moi une stagiaire d'intérêt, et pour m'avoir transmis les fondamentaux de la recherche en biologie, votre passion et votre rigueur, qui ne m'ont jamais quittée.

**merci** à Olivier Panaud et Jeff Bennetzen pour votre vision, votre confiance et la liberté que vous m'avez donnée pour commencer ma carrière de jeune chercheuse, pour nos discussions passionnées autour de la génomique et des éléments transposables, et pour votre soutien.

**merci** à tous les membres du GDR « Eléments transposables », avec qui j'ai toujours tant de plaisir à discuter de ces entités spéciales et qui nous sont si chères. Notre communauté est une richesse que nombreux pays nous envient, et c'est un honneur et une joie pour moi d'en faire partie !

**merci** à Eric Bonnivard, pour m'avoir transmis ta passion des éléments transposables dès ma première année d'université, et pour avoir accompagné mes débuts dans la recherche.

**merci** à Vincent Colot, pour savoir me rappeler mon potentiel à chaque occasion. Nos discussions sont souvent des électro-chocs, et l'existence de ce manuscrit leur doit sûrement beaucoup.

**merci** à Angélique Déleris pour m'avoir donné l'opportunité de contribuer à tes projets sur la régulation épigénétique des éléments transposables d'*A. thaliana*. Je n'avais jamais travaillé avec des copies aussi détériorées !

**merci** aux membres du GDR « Epigénétique en Ecologie et Evolution », pour les discussions passionnantes sur ce sujet et pour les opportunités d'animation scientifique et de co-rédaction de chapitre de livre.

**merci** aux membres du GDR « Epiplant » pour votre vision moléculaire qui donne une autre lumière à mes réflexions.

**merci** à mes parents et à ma soeur, pour m'avoir donné les possibilités de faire un métier que j'aime, et pour m'avoir accompagnée de votre amour tout au long de ce chemin.

**merci** à Johann pour ton soutien et ton amour inconditionnels, et bien plus encore. **merci** à mes deux petits curieux du monde et de la nature, qui savez si bien me rappeler vers la 'vraie vie' par vos câlins, vos sourires et vos dessins quand je passe un peu trop de temps à travailler.

A la mémoire de Darshan Brar, Ana Pontaroli et Peter Rogowsky, qui ont accompagné chacun une partie de ma carrière. **merci** pour votre générosité, votre vision et votre soutien.

A ma maman (1947-2018)

# Curriculum vitae

## Clémentine Vitte

Née le 23 août 1978 à Paris 14<sup>ème</sup>  
Pacsée, 2 enfants (nés en 2014 et 2017)

ORCID : 0000-0001-5908-0955  
<https://moulon.inrae.fr/en/personnes/cvritte/>

### ACTIVITES DE RECHERCHE

---

- 2008-actuel**    **Chargée de Recherche CNRS** (CR2 en 2008, CR1 en 2013, CRCN en 2017)  
GQE-Le Moulon, Gif-sur-Yvette, France. Quotité de travail : 100%
- 2007**    **Post-doctorat**, laboratoire Dynamique du Génome et Evolution (équipe H. Quesneville)  
Institut Jacques Monod, Paris, France  
*Impact des éléments transposables sur la formation de l'hétérochromatine chez Drosophila melanogaster*
- 2005-2006**    **Post-doctorat**, Bennetzen laboratory  
Université de Géorgie, Athens, Géorgie, Etats-Unis  
*Génomique comparée et éléments transposables*
- 2001-2004**    **Doctorat**, université Paris 6 (directeur de thèse : O. Panaud)  
Laboratoire Ecologie, Systématique et Evolution, Orsay, France  
*Éléments transposables et évolution du génome du riz (genre Oryza)*
- 2003**    **Visiteur international**, Wide Hybridization group (5 mois)  
International Rice Research institute, Los Baños, Philippines  
*Culture et validation phénotypique de variétés traditionnelles de riz ; transfert de la technique RDA pour comparer des variétés de riz distantes*
- 2001**    **Stage de DEA** (5 mois, sous la direction d'O. Panaud)  
Laboratoire Ecologie, Systématique et Evolution, Orsay, France  
*Etude des différenciations génomiques au sein du genre Oryza par la technique de clonage soustractif « Representational Difference Analysis » (RDA)*
- 2001**    **Stage de DEA** (4 mois, sous la direction d'H. Vaucheret),  
Laboratoire de Biologie Cellulaire (INRA), Versailles  
*Analyse de mutants d'Arabidopsis thaliana affectés dans le contrôle épigénétique*
- 2000**    **Stage de DEA** (4 mois, sous la direction de F. Boccard)  
Centre de Génétique moléculaire, Gif-sur-Yvette, France  
*Etude des interactions à longue distance au sein du chromosome d'Escherichia coli*
- 1999**    **Stage de recherche maîtrise** (2 mois ½, sous la direction de D. Charlesworth)  
Université d'Edimbourg, Edimbourg, Ecosse  
*Evolution des chromosomes sexuels chez Silene latifolia: analyse de la diversité de séquence du contrôle autosomal CCLS 37.1*
- 1998**    **Stage de recherche bénévole** (2 mois, sous la direction de V. Raynal)  
Institut National de Transfusion Sanguine, Paris, France  
*Diagnostic anténatal du facteur Rhésus par PCR quantitative*
- 1997**    **Stage de recherche bénévole** (2 mois, sous la direction d'H. Merçot)  
Institut Jacques Monod, Paris, France  
*Analyse de l'incompatibilité cytoplasmique due aux bactéries endocellulaires du genre Wolbachia chez Drosophila simulans*

## ACTIVITES D'ENSEIGNEMENT

---

<b>2024</b>	Cours. Master 2 GENEE (1h30/an)
<b>2022-actuel</b>	Cours. Master 2 EvoGEM (6h/an)
<b>2011-présent</b>	Cours. 3ème année AgroParisTech (4h/an)
<b>2008 et 2009</b>	Cours et travaux pratiques, école d'été « MEEG », station biologique de Roscoff (16h/an, en anglais).
<b>2008 et 2009</b>	Cours Master 2 « GCC », Université Paris 6 (4h/an)
<b>2001-2004</b>	Monitorat, université Paris XI. Travaux dirigés (30h/an) et travaux pratiques (34h/an), licence « Génétique et Microbiologie Générales »

## FORMATION ET DIPLOMES

---

2001-2004	<b>Thèse de doctorat</b> (bourse du ministère) <i>Eléments transposables et évolution du génome du riz (genre Oryza)</i> . Diplôme de l'Université Paris 6 obtenu le 11 octobre 2004. Mention très honorable. Allocataire Moniteur de recherche, Université Paris Sud
2000-2001	<b>DEA de Génétique</b> (Universités Paris 6/Paris 7/Paris XI/Versailles St Quentin). Mention Très Bien - Classée 3 <sup>e</sup> sur 14
1998-2001	<b>Magistère de Biologie/Biochimie</b> (Ecole Normale Supérieure de Paris)
1996-1998	<b>DEUG Sciences de la Vie</b> . Université Paris 7. Mention Bien

## ENCADREMENT ET FORMATION

---

### *Encadrement de doctorants :*

**2014-2018. Zeineb Achour** (bourse mission universitaire de Tunisie à Paris et projet Amaizing). Encadrement 100%. Actuellement CDI Saphetor, Caen

*Réponse du méthylome suite à l'exposition au froid chez une espèce à génome complexe : le maïs (Zea mays ssp. mays)*

**2010-2013 Abdul Rehman Kahn**. Co-encadrement 30%. Actuellement Tenured Associate Professor, Department of Biotechnology, COMSATS University Islamabad, Pakistan.

*Réponse à court terme des populations de blé européens aux conditions agro-climatiques contrastées : analyse génétique et première étape vers le développement de marqueurs épigénétiques dans les gènes de précocité VRN-A1*

### *Encadrement de post-doctorants et CDD :*

**Juin-Sept. 2021. Arnaud Desbiez-Piat**. Post-doctorant. Co-encadrement (50%). Actuellement post-doctorant au LEPSE (Montpellier)

*Investigation of the role of DNA methylation changes in maize adaptation*

**Jan. 2020-Juill. 2021. Kevin Baudry**. Post-doctorant. Co-encadrement (20%). Actuellement post-doctorant à l'Université de Cornell (Etats-Unis)

*Analyse transcriptionnelle des gènes 'core' et 'accessoires' chez le maïs*

**Jan. 2019-Mars 2020. Maud Fagny**. Post-doctorante. Encadrement 100%. Actuellement chargée de recherche INRAE à GQE-Le Moulon

*Analyse des réseaux de régulation tissus-spécifiques chez le maïs, et lien avec les éléments transposables*

**Nov. 2018-Nov. 2019. Jonathan Torel-Salat**. CDD IE. Encadrement 100%. Actuellement développeur informatique chez Sopra Steria

*Analyse des profils de méthylation de l'ADN de régions ciblées au cours du développement et dans la descendance*

**Sept. 2012-Août 2014 Sara Castelletti**. Encadrement 100%. Actuellement biotechnologiste à IGA Technology Services, Italie

*Analyse de la variation du méthylome entre lignées de maïs d'origines génétiques contrastées*

### *Encadrement de stagiaires:*

**2024 Mathieu Fressac.** M2 Plant Sciences, Université Paris-Saclay

*Study of cis-regulatory elements responding to water-deficit in B73 maize husk*

**2024 Gabrielle Minne.** M1 Plateforme GBMC, Université Paris-Saclay. Co-encadrement 35%

*Rôle de la méthylation des gènes dans l'adaptation du maïs européen*

**2023 Thémis Lemarchand.** M1 GENIOMHE, Université Paris-Saclay. Co-encadrement 70%.

*Etude de l'impact du déficit hydrique sur l'activation des éléments cis-régulateurs dans différents génotypes de maïs*

**2022 Klaus von Grafenstein.** M2 AMI2B Université Paris-Saclay. Encadrement 100%.

*PIF/Harbinger MITE transposable elements impact on Zea mays gene expression*

**2021 Abirami Soundiramourthy.** M2 AMI2B Université Paris-Saclay. Co-encadrement 50%.

*Émergence des sites de fixation aux facteurs de transcription dans les copies MITE PIF/Harbinger et impact sur l'expression génique chez Zea Mays*

**2014 Genserik Beydon.** M2 Biosciences végétales, Université de Toulouse. Co-encadrement 50%.

*Genomic analysis of the transposable element structural diversity among maize lines*

**2014 Iris Grosjean.** M1 Génomique et Productivité Végétale, Université Paris7-Denis Diderot. Co-encadrement 50%

*Relation entre variation structurale, méthylation et expression des gènes chez le maïs (Zea mays ssp. mays)*

**2012 Emeline Baptiste.** M2 Génomique et Productivité Végétale, Université Paris 7-Denis Diderot. Encadrement 100%.

*Comparaison de la diversité en éléments transposables au sein de différentes lignées de maïs (Zea mays ssp. mays)*

**2012 Aurélie Vaurijoux.** M2 Gènes, Cellules, Développement, Evolution, Université Paris Sud XI. Co-encadrement 40%.

*Caractérisation des gènes de type Dicer et Argonaute dans le genre Brassica.*

**2012 Sivasangari Nandy.** M2 Bioinformatique et Biostatistique, Université Paris-Sud. Co-encadrement 50%.

*Etude de l'impact de la taille d'insert des banques Illumina pairées sur la détection de variants structuraux et la qualité des assemblages de novo : exemple du maïs.*

**2010 Lilia Moulfi.** L3 Pro en Biotechnologie, Université de Cergy Pontoise). Encadrement 100%.

*Travaux préliminaires pour l'analyse du profil de méthylation de trois gènes impliqués dans le contrôle de la précocité floraison chez le maïs*

**2003 Marion Vankerssaver** (M1, université Pierre et Marie Curie). Co-encadrement 70%.

*Study of the accumulation of deletions in the rice LTR retrotransposon Houba sequences*

### *Comités de suivi de thèse :*

**Mai 2024 Inès Matrougui,** Université Paris-Saclay (Dir. C. Gilbert & S. Charlat, EGCE, Gif-sur-Yvette, France)

**Juin 2023, Sept. 2024 Mounia El Messaoudi,** Université Paris-Saclay (Dir. P. Baduel & V. Colot, IBENS, Paris, France)

**Oct. 2022, Oct. 2023, Sept. 2024 Mireia Bueno,** Université Paris-Saclay (Dir. L. Quadrana, IPS2, Gif-sur-Yvette, France)

**Jan. 2022, Jan. 2023 Somia Saidi,** Université Paris-Saclay (Dir. H. Quesneville & J. Confais, URGI, France)

**Oct. 2022 Christine Tranchant-Dubreuil,** université de Montpellier (Dir. F. Sabot & Y. Vigouroux, IRD, France)

**Sept. 2022 Abirami Soundiramourthy,** Université de Perpignan (Dir. O. Panaud & M. Mirouze, IRD, France)

**Juill. 2021, Juin 2022 Inès Ouedraogo,** Université de Montpellier (Dir. Daphné Autran, IRD, Montpellier, France)

**Juin 2020, Nov. 2021, Jan. 2022, Jan. 2023, Mars 2023 Wenbo Xu**, Université de Zürich (Dir. Anne Roulin, Univ. Zurich, Suisse)

**Sept. 2018, Sept. 2019 Arnaud Desbiez-Piat**, Université Paris-Saclay (Dir M. Tenaillon et C. Dillman, GQE-Le Moulon, Gif sur Yvette)

**Sept. 2019, Sept. 2020, Oct. 2021 Janan Gawra**, Université de Perpignan (Dir. C. Grunau & J. Vidal-Dupiol, IHPE, Montpellier-Perpignan)

**Mars 2016, Avril 2017 Ophélie Jouffroy**, Université Paris-Saclay (Dir. H. Quesneville & F. Maumus, URGI, Versailles)

**Oct. 2015, Nov. 2016 Victoire Baillet**, Université Paris Sciences et Lettres (Dir. V. Colot, IBENS, Paris)

**Oct. 2015 Anne-Laure Legac**, Université d'Orléans (Dir. S. Maury, LBLGC, Orléans)

**Nov. 2011, Déc. 2012 Paulina Martinez-Palacios**, Université Paris Sud (Dir. K. Alix & J. Joets, GQE-Le Moulon, Gif sur Yvette)

**Déc. 2012 Manon Richard**, Université Paris Sud (Dir. Valérie Geffroy, UMR 8618, Gif sur Yvette)

**Juill. 2011, Juill. 2013 Mélanie Thomas**, Université Clermont-Ferrand (Dir. Christophe Tatout, UMR CNRS 6293, INSERM U1103, Clermont-Ferrand)

**Juin 2009 Alexis Sarazin**, Université Paris Sud (Dir. V. Colot, UMR 8186, Paris)

**Oct. 2010 Véronique Sarilar**, Université Paris Sud (Dir. K. Alix, GQE-Le Moulon, Gif sur Yvette)

**Sept. 2009, Oct. 2010 Pauline Goubet**, Université de Lille (Dir. X. Vekemans, UMR 8016, Lille)

**Nov. 2009 Mathieu Piednoël**, Université Paris 6 (Dir. E. Bonnard, UMR 7138, Paris)

## **BOURSES ET FINANCEMENTS**

---

**2025-2029 : SATURN.** (ANR, France. 727 k€). Coordinatrice

**2023- 2026 : EDin** (ANR, France. 718 k€). Participante

**2021: COMET** (Trans-national Easigenomics call, production de données ATAC-seq). Coordinatrice, 4 partenaires, international

**2019-2022 : SeqOccin** (Occitanie region/private companies/INRAe, 6 M€). Co-coordinatrice de la partie relative au maïs

**2011-2021 : Amaizing** (ANR and France Agrimer, France. 9 M€). Coordinatrice du WP3, 2.5 M€, membre du comité exécutif

**2012-2015 : AdaptInWild** (ANR Bioadapt, France. 278 k€). Participante

**2011-2013 : CNV\_Maize** (ANR Génomique, France. 599 k€) Participante

**2008-2010 : INRA** projet innovant (France, 62 k€). Investigatrice principale

**2007 : Bourse individuelle de post-doctorat** 12 mois (Urgence FRM/Post-docs)

**2005 :Bourse individuelle de post-doctorat** 12 mois (Lavoisier Fellowship, ministère des affaires étrangères)

**Collaboratrice ANR JCJC (France):** 2023-2026: NETWITS (Maud Fagny); 2023-2026: DaTE Palm (Yann Bourgeois) ; 2021-2024: PLEASURE (Amandine Cornille)

## **EXPERTISE POUR DES JURYS**

---

### *Membre de jurys de thèse à l'international (rapporteuse)*

**2020 et 2021 Laurie Maistriaux**, Université de Louvain-la-Neuve, Belgique (Dir. F. Chaumont). Pré-soutenance privée avec les rapporteurs le 1<sup>er</sup> octobre 2020 et défense publique le 27 janvier 2021.  
*Genetic variability of the expression of plasma membrane aquaporins in maize leaves: from eQTLs to characterization of cis- and trans-acting regulatory factors*

### *Membre de jurys de thèse à l'international (examinatrice)*

**2023 Maria-Estefania Lopez**, université de Genève, Suisse (Dir. : Etienne bucher)  
*Mitotic and Meiotic Maintenance of Stress Memory in Woodland Strawberry*

**2012 Taciane Finatto**, codirection université fédérale de Pelotas, Brésil (Dir. A. Costa de Oliveira) et université de Perpignan (Dir. O. Panaud et N. Picault)

*Analyses transcriptomiques des gènes et des rétrotransposons à LTR chez le riz (Oryza sativa ssp. Japonica) en réponse à un stress ferrique*

#### *Membre de jurys de thèse en France (examinatrice)*

**2023 Flavia Pavan**, Université de Lille (Dir. V. Castric & S. Legrand)

*Origin and evolutionary trajectory of microRNA genes in Arabidopsis species*

**2022 Nguyet Tang**, Université de Montpellier (Dir. F. Sabot, IRD)

*Méthodes d'inférence des variations structurelles à l'échelle du génome dans les données séquençage basse profondeur à l'aide du graphe de pangéome*

**2021 Margot Berger**, Université de Bordeaux Bordeaux (Dir. P. Gallusci)

*Régulation épigénétique et mémoire des stress en lien avec le métabolisme du carbone chez la vigne*

**2021 Marie-Christine Carpentier**, Université de Perpignan (Dir. O. Panaud)

*Etude de l'activité transpositionnelle chez le genre Oryza à l'ère des nouvelles technologies de séquençage*

**2020 Adrien Perrin**, Université d'Angers (Dir. E. Bucher, IRHS)

*Transmission des marques épigénétiques lors de la multiplication sexuée et de la propagation asexuée chez le pommier (Malus domestica)*

**2019 Romain de Oliveira**, Université de Clermont Auvergne (Dir. E. Paux & F. Choulet)

*Variabilité, diversité et évolution du génome du blé*

**2018 Nicolas Daccord**, Université d'Angers (Dir. E. Bucher, IRHS)

*Analyse bioinformatique du génome et de l'épigénome du pommier*

**2018 Benjamin Liégard**, Université de Rennes (Dir. M. Manzaneres-Dauleux & M. Jubault, UMR IGEPP)

*Rôle des variations épigénétiques transgénérationnelles dans la résistance quantitative à la hernie chez Arabidopsis thaliana*

**2016 Emilie Debladis**, Université de Perpignan (Dir. O. Panaud & E. Lasserre, UMR LGDP)

*Etude de l'activité transpositionnelle en condition de stress chez le riz, Oryza sativa*

**2015 Aurélie Deremetz**, Université Paris-Saclay (Dir. N. Bouché, IJPB)

*Caractérisation fonctionnelle de la protéine SG1 et implication dans les processus de régulations épigénétiques chez Arabidopsis thaliana*

**2012 Audrey Courtial**, Université de Toulouse 3 (Dir. Y. Barrière & J. Grima-Pettenati)

*Vers l'identification de gènes contrôlant la dégradabilité de la paroi secondaire lignifiée chez le maïs à travers l'élucidation de QTLs à effets forts*

**2011 Grégory Carrier**, Université de Montpellier (Dir. P. This & J.-M. Boursiquot, UMT GénoVigne)

*Identification du polymorphisme moléculaire à l'origine de la variation clonale chez la vigne*

#### *Membre de jurys de thèse (membre invité)*

**2013 Abdul Rehman Khan**, Université Paris-Sud (Dir. I. Goldringer)

*Short term response of European wheat populations to contrasted agro-climatic conditions: a genetic analysis and first step towards development of epigenetic markers in earliness gene VRN-A1*

#### *Membre de jurys de recrutement*

**2024 Jury de l'école doctorale SEVAB**, Université Toulouse, France

**2021 Jury de l'école doctorale SEVE**, Université Paris-Saclay, France

**2021 Jury de recrutement maître de conférences**, Université de Rennes, France

**2019 Jury de recrutement maître de conférences**, Université d'Angers, France

**2019 Jury de l'école doctorale SEVE**, Université Paris-Saclay, France

**2015 Jury de recrutement CR2 INRA SPE**, France

## **PARTICIPATION A LA FORMATION DES CHERCHEURS**

---

**Ecole chercheurs « Génomique et diversité des caractères à déterminisme complexe ».** La Colle sur Loup, France. 15-20 mai 2011 : conférence sur le thème « Les éléments transposables: de leurs caractéristiques à leur impact sur le phénotype ».

**Ecole chercheurs « Epigénétique en Ecologie et Evolution ».** Saint Pierre d'Oléron, France. 18-22 mai 2015. Cours interactif « Interaction between the genome and the epigenome » et travaux pratiques « Bisulfite-seq methylome analysis on plant data ».

## **AUTRES EXPERTISES ET RESPONSABILITES**

---

**Expertise pour des projets de recherche** depuis 2009 (région, INRA, IRD, ANR, NSF)

**Editrice académique du journal Plos Genetics** depuis 2024

**Membre du comité éditorial du journal Mobile DNA** depuis 2023

**Editrice invitée pour PLoS Genetics** en 2023

**Co-éditrice d'un 'research topic' pour Frontiers** en 2023

**'Invited peer reviewer' pour 31 articles** de journaux à comité de lecture: Nature Plant, PLoS Genetics, The Plant Journal, New Phytologist, Journal of Experimental Botany, Annals of Botany, Molecular Biology and Evolution, Genome Biology and Evolution, etc.

**Membre du conseil de l'école doctorale SEVE** depuis 2023

**Membre du conseil scientifique du GDR2157** « Evolution des éléments transposables: des génomes aux populations » depuis 2009

**Membre du conseil scientifique du GDR** « Epigénétique en Ecologie et Evolution » de 2019 à 2023.

**Membre du comité de pilotage** du Réseau Thématique Pluridisciplinaire INEE "Epigénétique en Ecologie et Evolution" (RTP 3E) de 2013 à 2018

**Membre du GDR** « Epigénétique des plantes » (Epiplant) depuis 2019

**Membre du réseau international** « Maize chromatin network » depuis 2019 (fin en 2021).

**Membre élu au conseil scientifique de l'IDEEV depuis 2020**

**Co-organisatrice de congrès internationaux** : Workshop « Transposable Elements » au sein du congrès international « Plant and Animal Genomes » à San Diego, Etats-Unis depuis 2020 ; International Congress on Transposable Elements (ICTE) en 2012 et 2024 ; Symposium « Epigenetics in ecology and evolution » au sein du Congrès international de la SBE, Lyon, 2010

**Co-organisatrice de congrès nationaux** : Congrès National des Eléments Transposables (CNET), Gif-sur-Yvette, 2022 ; Colloque « Epigénétique en écologie et évolution », Gif sur Yvette, 2013 ; Congrès « The central dogma of molecular biology revisited: high-throughput sequencing and new concepts », Institut Curie, Paris, 2013 ; Colloque « Eléments Transposables, Virus et Domestication des Espèces », Lyon, 2010

**Co-organisatrice d'école chercheurs** : « Epigénétique en Ecologie et Evolution ». Saint Pierre d'Oléron, France. 18-22 mai 2015

## **VULGARISATION SCIENTIFIQUE**

---

**Presse écrite** : Interview par la revue 'Epsilon' pour le dossier « Transposons, les gènes qui boostent l'évolution », mai 2022 ; Participation à la rédaction de l'article « Concevoir et évaluer les maïs du futur » pour le journal Perspectives Agricoles, septembre 2020.

**Interventions en milieu scolaire** : Participation au projet « Chercheurs en herbe », intervention au lycée Assomption de Rennes (classe de terminales) sur le métier de chercheur et la génomique. Proposition de sujets pour la préparation des élèves au grand oral du baccalauréat. Intervention récurrente en milieu périscolaire (ALSH) : atelier « graines, domestication et diversité ».

## PRODUCTION SCIENTIFIQUE

---

### 27 publications dont 17 dans des revues à comité de lecture.

\* co-premier auteur, <sup>§</sup> corresponding author

Les noms des **post-doctorants** encadrés ou co-encadrés sont indiqués en rouge. Les noms des **doctorants** encadrés ou co-encadrés sont indiqués en vert.

### Articles de recherche publiés dans des journaux à comité de lecture

- B. Leduque, A. Edera, **C. Vitte**, L. Quadrana (2024). Simultaneous Profiling of Chromatin Accessibility and DNA Methylation in Complete Plant Genomes Using Long-Read Sequencing. NAR
- A. Desbiez-Piat, A. Ressayre, E. Marchadier, A. Noly, C. Remoué, **C. Vitte**, H. Belcram, A. Bourgeois, N. Galic, M. Le Guilloux, M.I. Tenaillon, C. Dillmann. Pervasive GxE interactions shape adaptive trajectories and the exploration of the phenotypic space in artificial selection experiments (2023). Genetics. Oct 12:iyad186. doi: 10.1093/genetics/iyad186. Epub ahead of print. PMID: 37824828.
- M. Fagny**, M. L. Kuijjer, M. Stam, J. Joets, O. Turc, J. Roziere, S. Pateyron, A. Venon, **C. Vitte**<sup>§</sup> (2021). Identification of key tissue-specific, biological processes by integrating enhancer information in maize gene regulatory networks. Front. Genet., <https://doi.org/10.3389/fgene.2020.606285>
- C. Mabire, J. Duarte, A. Darracq, A. Pirani, H. Rimbart, D. Madur, V. Combes, **C. Vitte**, S. Praud, N. Riviere, J. Joets, J.P. Pichon, S. Nicolas (2019). High throughput genotyping of structural variations in a complex plant genome using an original Affymetrix® Axiom® array. BMC Genomics, 1 (20) 848
- A. Darracq\*, **C. Vitte**\*, S. Nicolas, J. Duarte, J.P. Pichon, T. Mary-Huard, C. Chevalier, A. Berard, Le M.C. Paslier, P. Rogowsky, A. Charcosset, J. Joets (2018). Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. BMC Genomics, 19 (1) 119
- J. T. Brandenburg, T. Mary-Huard, G. Rigail, S. J. Hearne, H. Corti, J. Joets, **C. Vitte**, A. Charcosset, S. Nicolas, M. Tenaillon (2017). Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. PLoS Genet, 13 (3) e1006666.
- A. R. Khan**, J. Enjalbert, A.-C. Marsollier, A. Rousselet, I. Goldringer, **C. Vitte**<sup>§</sup> (2013). Vernalization treatment induces DNA site-specific hypermethylation at the *VERNALIZATION-A1 (VRN-A1)* locus in hexaploid winter wheat. BMC Plant Biology. 13:209.
- C. Vitte**, M. Estep, J. Leebens-Mack, J. Bennetzen (2013). Young, intact and nested retrotransposons are abundant in the onion and asparagus genomes. Annals of Botany. doi: 10.1093/aob/mct155
- J. Shim, O. Panaud, **C. Vitte**, M. Mendioro, D. S. Brar (2010). RDA derived *Oryza minuta*-specific clones to probe genomic conservation across *Oryza* and introgression into rice (*O. sativa* L.), Euphytica 176:269–279
- C. Vitte**, O. Panaud, H. Quesneville. LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss. (2007) BMC Genomics 8:218.
- R. Liu\*, **C. Vitte**\*, J. Ma, A.A. Mahama, T. Dhliwayo, M. Lee, J.L. Bennetzen (2007). A GeneTrek analysis of the maize genome. Proc. Nat. Acad. Sc. U.S.A 104:11844-49.
- C. Vitte**, C. Chaparro, H. Quesneville, O. Panaud (2007). *Spip* and *Squiq*, two novel rice non-autonomous LTR retro-element families related to *RIRE3* and *RIRE8*. Plant Science, 172(1):8-19.
- C. Vitte**, J.L. Bennetzen (2006). Analysis of Retrotransposon Structural Diversity Uncovers Properties and Propensities in Angiosperm Genome Evolution. Proc. Nat. Acad. Sc. U.S.A. 103(47):17638-43.
- C. Vitte**, T. Ishii, F. Lamy, D.S. Brar and O. Panaud (2004). Genomic paleontology provides evidence for two distinct origins of asian rice (*Oryza sativa*, L.). Mol. Gen. Genomics. 272(5):504-11.

- C. Vitte**<sup>§</sup>, O. Panaud (2003). Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol. Biol. Evol.* 20(4):528-40.
- O. Panaud, **C. Vitte**, J. Hivert, S. Muzlak, J.D. Talag, D.S. Brar and A. Sarr (2002). Genomic differentiation between rice (*Oryza sativa* L.) and foxtail millet (*Setaria italica* L. Beauv.) revealed through representational difference analysis. *Mol. Gen. Genomics.* 268(1):113-21.
- D.A. Filatov, V. Laporte, **C. Vitte** and D. Charlesworth (2001). DNA diversity in sex-linked and autosomal genes of the plant species *Silene latifolia* and *Silene dioica*. *Mol. Biol. Evol.* 18(8):1442-54.

#### Articles de recherche, preprints

- V. Hure, F. Piron-Prunier F., T. Yehouessi T., **C. Vitte**, A.E. Kornienko, G. Adam, M. Nordborg, A. Délérís. (2024). Alternative silencing states of Transposable Elements in Arabidopsis. *Biorxiv*  
doi: <https://doi.org/10.1101/2024.03.16.585326>
- Z. Achour, J. Joets, M. Leguilloux, H. Sellier, J.-P. Pichon, M. Leveugle, H. Duborjal, J. Caius, V. Brunaud, C. Paysant-Le Roux, T. Mary-Huard, C. Giauffret, C. Vitte (2019). Low temperature triggers genome-wide hypermethylation of transposable elements and centromeres in maize. *Biorxiv*  
<https://doi.org/10.1101/573915>.

#### Article de revue publiés dans des journaux à comité de lecture

- M. Mirouze, **C. Vitte** (2014) Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. *J Exp Bot.* 65(10):2801-12.
- C. Vitte**, M.-A. Fustier, K. Alix, M. Tenaillon (2014). The bright side of transposons in crop evolution. *Briefings in Functional Genomics.* 13(4):276-95.
- C. Vitte**, O. Panaud (2005). LTR retrotransposons and plant genome size: emergence of the increase/decrease model. *Cytogenetics and Genome Research.* 110(1-4):91-107. Article invité dans l'édition spéciale « Retrotransposons, Retrotransposition and Genome Evolution ».

#### Chapitres de livres (invités)

- C. Vitte**, S. Chambeyron and C. Vieira (2024). Epigénétique et Eléments Transposables. In Raconte-moi l'épigénétique, editors C. Grunau and S. Maury. ISTE, p.76-90 (sous presse)
- C. Vitte**, N. Nègre (2024). L'épigénétique : la même pour toutes les espèces? In Raconte-moi l'épigénétique, editors C. Grunau and S. Maury. ISTE, p.91-99 (sous presse)
- J. Joets, **C. Vitte**, A. Charcosset (2019). Draft assembly of the F2 European maize genome sequence and its comparison to the B73 genome sequence: a characterization of genotype-specific regions. In *The maize Genome*, editors Jeffrey Bennetzen, Professor. Sherry Flint-Garcia, Professor. Candice Hirsch and Dr. Roberto Tuberosa. Springer Nature, p.3-12
- C.M. Diez, **C. Vitte**, J. Ross-Ibarra, B.S. Gaut, M. Tenaillon (2013). Using next generation sequencing to investigate genome size variation and transposable element content. In M. A. Grandbastien and J. Casacuberta, *Plant transposons (Topics in current genetics)*, Springer-Verlag, Germany.
- J. Joets, **C. Vitte**, S. Pasternak, D. Ware (2011). Deciphering the multi-faceted aspects of the maize genome: new insights from sequencing projects. In J.-L. Prioul, *Advances in Maize (Taylor & Francis)*, the Society for Experimental Biology, London, UK. pp. 111-120.
- C. Vitte**, T. Peterson, J. Bennetzen (2011). Transposable elements and the evolving structure of the maize genome. In J.-L. Prioul, *Advances in Maize (Taylor & Francis)*, the Society for Experimental Biology, London, UK. pp. 2-36.
- P. SanMiguel, **C. Vitte** (2009). The LTR retrotransposons of maize. In J. Bennetzen and S. Hake, *The Maize Handbook – Volume II: Domestication, Genetics and Genomics (Springer, ISBN 0387778624, 9780387778624)*, pp.307–327.

### *Article éditoriaux*

G. Moissiard, M. Mirouze, C Carles, **C. Vitte**. Editorial: Plant epigenetics and chromatin dynamics - EPIPLANT 2021-2022. *Frontiers*.

### *Proceedings de congrès*

A. Ainouche, M. Betermier, M. Chandler, R. Cordaux, G. Cristofari, J.M. Deragon, P. Lesage, O. Panaud, H. Quesneville, C. Vaury, C. Vieira, **C. Vitte**. (2012) International Congress on Transposable Elements (ICTE) 2012 in Saint Malo and the sea of TE stories. *Mob. DNA*, 3 (1) 17.

## **CONFERENCES INVITEES**

---

### *Dans des congrès internationaux*

**2023 European Maize Meeting**, Bologne, Italie. 14-16 juin 2023.

**2021 Epiplant III**, Visioconference, 8-12 février 2021.

**2019 Plant and Animal Genomes XXVII**. San Diego, CA, Etats-Unis. 12-16 Janvier 2019.

**2019 Maize chromatin network annual meeting**. Saint Louis, MO, Etats-Unis., 13 mars 2019.

**2018 International plant epi/genetics symposium**, 29-31 octobre 2018, Angers, France. (Keynote talk).

**2016 International Conference of Ecological Sciences**, 24-27 Octobre 2016, Marseille, France.

**2015 XXIIIrd EUCARPIA Maize and Sorghum Conference**, Montpellier, 10-11 juin 2015.

**2013 International workshop on Plant Genome Dynamics**, Banyuls sur Mer, 12-13 sept. 2013.

**2009 9<sup>th</sup> International Plant Molecular Biology congress**. St Louis, Missouri, Etats-Unis. 25-30 octobre 2009.

**2003 7<sup>th</sup> International Plant Molecular Biology congress**. Barcelone, Espagne. 23-28 juin 2003.

### *Dans des congrès nationaux*

**2015 Workshop du GIS-BV** « Prise en compte des processus épigénétiques en amélioration variétale ». Paris, 9 novembre 2015.

**2010 Colloque** « Eléments Transposables, Virus et Domestication des Espèces ». Lyon. 11-12 mai 2010.

### *Dans des instituts internationaux*

**2022 Université de Louvain-la-Neuve**, Belgique

# Avant-propos

Pendant mon enfance, l'observation de la nature tenait une part importante de ma vie familiale, de la découverte des insectes, cloportes, oiseaux, hérissons et autres écureuils des parcs parisiens, à celle de l'estran et de la faune qui la compose, lors de nos promenades estivales. J'étais fascinée par la quantité de 'bestioles' qui vivait là, dans le sable, dans l'eau, sous les algues, sous les rochers, chacune ayant trouvé l'endroit qui lui convenait. J'étais aussi fascinée par la diversité des formes, des couleurs, des tailles des organismes que j'y ai découvert.

Le monde des mécanismes de la biologie, des molécules que j'ai rencontré au cours de ma scolarité m'a tout de suite passionnée par son mélange de complexité et de champs des possibles. Cette passion ne m'a jamais quittée, et fait encore briller mes yeux aujourd'hui, quand je sillonne les génomes et les épigénomes de mes analyses, tel un plongeur à la découverte des fonds océaniques, pour en comprendre le fonctionnement.

C'est à l'université que j'ai découvert la biologie végétale, la génétique moléculaire et l'évolution, trois domaines qui ont fortement influencés mes choix de parcours universitaire et professionnel. Mais, dans cette période universitaire, ce sont surtout les stages en laboratoire qui ont orienté mon parcours. A la fin de ma première année d'université, désireuse d'expérimenter si le travail en laboratoire me plairait, j'ai souhaité faire un stage bénévole en laboratoire pendant mes vacances d'été. Venant d'un milieu ouvrier je ne connaissais personne dans le monde de la recherche, et me voilà donc partie à frapper aux portes des laboratoires de l'Institut Jacques Monod à Paris. Je me souviendrai toujours de mes premiers pas dans un laboratoire. De l'odeur des milieux de culture, des sons des appareils, des discussions des gens dans les couloirs. Ce jour-là, j'ai su.

Ce premier stage a été le commencement d'une série d'étés en laboratoire, puis de stages inclus dans mon parcours universitaire. Ils se sont articulés autour de trois domaines qui m'ont très tôt intéressée, et qui ont façonné les thèmes de recherche que j'ai développés par la suite : la dynamique des génomes, la régulation de la chromatine, et l'évolution.

J'y ai acquis des connaissances et des compétences, mais aussi une façon de voir le monde un peu différemment, par la lentille de l'hypothèse, de l'expérimentation et de la rigueur scientifiques. Ces valeurs sont cruciales pour moi, et je les transmets tous les jours aux étudiants que j'encadre, car au-delà de la créativité que ce métier nécessite, la rigueur, l'autocritique et l'honnêteté sont indispensables à la génération de connaissances solides. Ces valeurs m'ont aussi guidée vers la médiation scientifique, parce qu'il me paraît indispensable de les transmettre aux élèves et futurs citoyens, pour construire un monde fondé sur la science plutôt que sur les croyances. C'est un travail ardu mais je crois qu'il en vaut la peine.

Un thème commun à cette HDR est donc mon intérêt profond pour la compréhension de l'interaction entre le génome et l'épigénome, et surtout, la manière dont ces deux entités peuvent évoluer, avec un accent particulier sur la contribution des éléments transposables, de leur dynamique et de leur régulation épigénétique dans ce processus.

J'ai choisi de retracer mon parcours à partir de mes années de thèse, qui ont marqué le début de ma carrière en génomique végétale. J'ai choisi une présentation assez chronologique de mon parcours, car, si les éléments transposables ont toujours été au centre de mes travaux de recherches, les questions que j'ai développées ont évolué en lien avec l'évolution des connaissances et des innovations technologiques. Si mes travaux de thèse et de stages post-

doctoraux sont présentés de façon chronologique, ceux que j'ai menés depuis mon recrutement au CNRS sont présentés sous deux grands thèmes, que j'ai menés en parallèle.

Après une introduction permettant de positionner mes travaux dans une perspective générale, je présenterai les projets que j'ai menés, le contexte de ces recherches, les résultats scientifiques obtenus, et les encadrements que j'y ai réalisés. Ils seront présentés en suivant 3 grandes thématiques :

- **Etude de la contribution des éléments transposables à la variation de la taille des génomes** (chapitre 2) – *génomique comparée : plusieurs espèces, un génotype par espèce*
- **Caractérisation de la variabilité structurale intra-spécifique et lien avec l'expression des gènes** (chapitre 3) – *génomique et transcriptomique comparée : une espèce, plusieurs génotypes*
- **Epigénome : des stress abiotiques à la spécificité tissulaire** (chapitre 4) – *génomique et épigénomique: une espèce, plusieurs échantillons contrastés*

Je ferai ensuite une présentation générale de mes projets en cours et à venir (chapitre 5), qui sont à la convergence de mes travaux passés. Je finirai par deux chapitres plus personnels présentant ma vision de l'encadrement et de la diffusion de la science (chapitre 6) et un regard sur mon parcours (chapitre 7).

# Activités de recherche

## 1- Introduction générale

Le génome, d'abord défini pour décrire l'ensemble des **gènes** codés par un organisme sur l'ensemble de ses **chromosomes**, est maintenant défini comme l'ensemble du matériel génétique, codant et non codant, d'un organisme sur l'ensemble de ses chromosomes. La génomique, qui a émergé depuis les années 1970, est la discipline (que certains voient comme un outil, mais qui est clairement une discipline) qui étudie le fonctionnement biologique – d'un organisme, d'un organe, d'une cellule – à l'échelle de son génome. Deux branches complémentaires la caractérisent : la **génomique structurale**, qui vise à décrypter le contenu d'un génome et son organisation, et la **génomique fonctionnelle**, qui vise à déterminer la fonction des gènes et à comprendre comment ils sont exprimés. La **génomique évolutive** se situe à l'interface de ces deux branches, et vise à caractériser les forces évolutives qui façonnent la variation génomique et son expression.

En 50 ans, les avancées en génomique ont fait évoluer notre conception du génome d'un espace codant des protéines, vers un espace portant une grande diversité de gènes (codants des protéines ou des ARN « non codants » régulateurs de l'expression des gènes) et de séquences intergéniques, parmi lesquelles des régions *cis*-régulatrices capables de réguler l'expression des gènes, même à plusieurs kilobases de distance. Ainsi, plus nous observons les génomes, plus notre conception des séquences qui sous-tendent le phénotype augmente, s'étend, et se diversifie. De façon frappante, chez le maïs, on observe une association plus fréquente des phénotypes avec les régions intergéniques qu'avec les régions codant des protéines (Wallace *et al.*, 2014), suggérant que l'évolution des phénotypes est plus souvent sous-tendue par des régions non identifiées comme géniques.

En parallèle, la comparaison de génomes de plus en plus nombreux nous indique que ceux-ci n'évoluent pas seulement par de petites mutations comme les SNP (« Small Nucleotide Polymorphisms ») ou les petits indels (« Insertion/deletion »), mais aussi *via* des variations structurales comme des inversions, des duplications ou des délétions de grande taille (plusieurs kilobases). Cette dynamique structurale, d'abord établie entre espèces, est aujourd'hui clairement aussi démontrée au sein des individus qui composent une même espèce. Ces variants peuvent avoir des effets phénotypiques forts, notamment dans le cas des duplications ou des délétions de tout ou partie de régions codantes ou de régions *cis*-régulatrices.

Parmi les différents éléments qui constituent les génomes, **les éléments transposables** sont clairement des entités à part et tiennent une place particulière dans la dynamique des génomes et leur évolution. D'abord décrits comme une curiosité sans importance, ils sont aujourd'hui considérés comme des moteurs puissants de la variabilité structurale et fonctionnelle des génomes, et de la variabilité phénotypique des individus (Fedoroff, 2012; Lisch, 2013).

**Entités génétiques complexes, ubiquitaires et parasitaires, les éléments transposables sont capables de se déplacer à l'intérieur des génomes** en utilisant la machinerie cellulaire de l'organisme dans lequel ils résident. Ils sont classiquement séparés en deux grandes classes en

fonction de l'intermédiaire moléculaire qu'ils utilisent comme matrice de transposition : ARN pour les rétrotransposons (classe I) et ADN pour les transposons (classe II). Ces deux classes sont ensuite divisées en superfamilles en fonction du nombre de coupures qu'ils induisent lors de leur insertion, des protéines qu'ils codent et des structures particulières qu'ils contiennent. Enfin, les familles contiennent des ensembles d'éléments proche en termes de séquence nucléotidique (Wicker *et al.*, 2007). Au-delà de cette classification générale, on distingue les éléments autonomes, qui portent les gènes codant les protéines nécessaires à leur mobilité, et les éléments non-autonomes, qui en sont dépourvus et utilisent pour leur transposition des facteurs protéiques codés par des éléments autonomes. De façon remarquable, la distribution en classes, superfamilles et familles diffère considérablement d'une espèce à l'autre, indiquant un potentiel de mobilisation différent selon les espèces (Vitte *et al.*, 2014; Wells and Feschotte, 2020). De plus, le génome de chaque espèce semble porter un ensemble de copies à divers stades de décomposition évolutive, et qui reflète les événements d'amplification des différents éléments à travers le temps.

**Capables de mouvement, les éléments transposables sont des mutateurs puissants** qui peuvent induire des effets phénotypiques. Si les premiers cas de phénotypes associés à des insertions d'éléments transposables correspondaient essentiellement à la disruption de la phase ouverte de lecture, générant la production d'une protéine non fonctionnelle, la majorité des insertions à effet phénotypiques découvertes depuis le début du XXIème siècle sont liées à une modification des régions régulatrices de l'expression des gènes, notamment chez les plantes cultivées (Vitte *et al.*, 2014). Ces effets peuvent être dus à une disruption de séquences *cis*-régulatrices existantes, mais aussi à l'apport de nouvelles séquences régulatrices à proximité d'un gène (Lisch, 2013). En effet, les éléments transposables ont la particularité d'être des entités génétiques à part entière, contenant notamment des promoteurs et d'autres séquences permettant de réguler l'expression des gènes qu'ils portent. En sautant, ils peuvent ainsi apporter localement une séquence régulatrice à proximité d'un gène et ainsi modifier son profil d'expression. En outre, leur capacité de séquences mobiles leur confère la propriété de pouvoir disséminer ces séquences à divers endroits du génome, générant potentiellement des réseaux de gènes dont l'expression est régulée de façon concertée.

**Les mécanismes épigénétiques permettent de réguler la transposition.** Une transposition trop importante pouvant engendrer une surcharge mutationnelle ayant de lourdes conséquences phénotypiques, les organismes ont développé des mécanismes épigénétiques permettant de limiter ou supprimer l'activité des éléments transposables (Liu *et al.*, 2022). Chez les plantes et les mammifères, ces mécanismes dépendent en grande partie de la méthylation de l'ADN. Chez les plantes, cette méthylation des séquences d'éléments transposables se produit dans les trois contextes de cytosine CG, CHG et CHH (où H peut être A, T ou C), et la maintenance de la méthylation sur chacun de ces contextes dépend de voies spécifiques. Ce système, qui permet un contrôle sélectif de l'activité des éléments transposables, est transmis à travers les divisions cellulaires et les générations (Slotkin and Martienssen, 2007). Du fait de cette régulation épigénétique, une fois insérées, la plupart des séquences d'éléments transposables deviennent inactives et accumulent des mutations au cours du temps.

**Des erreurs dans l'établissement ou la maintenance de ces marques génèrent des épiallèles,** qui peuvent être induites par des modifications génétiques, chimiques ou de l'environnement.

Elles affectent la séquence des éléments transposables eux-mêmes, mais peuvent aussi s'étendre localement à leurs séquences flanquantes. De tels épiallèles sont ainsi 'facilités' par la présence de l'élément transposable (Richards, 2006; Baduel and Colot, 2021). Chez les plantes, les séquences *cis*-régulatrices de l'expression des gènes étant généralement caractérisées par une faible méthylation (Crisp *et al.*, 2020), l'apparition d'épiallèles dans ces séquences peut avoir un impact phénotypique important.

**La dynamique des éléments transposables a un lien fort avec l'environnement.** Les stress abiotiques (par exemple sécheresse, salinité, températures extrêmes, rayonnement ultraviolet, etc.) peuvent induire l'activation transcriptionnelle de certains éléments transposables, et conduire, dans certains cas, à la transposition (Ito, 2022; Pulido and Casacuberta, 2023). Chez les plantes, de nombreux éléments transposables de classe I et II sont réactivés sur le plan transcriptionnel par ce type de stress. L'activation transcriptionnelle peut notamment se produire par la liaison de facteurs de transcription spécifiques réagissant au stress, sur des séquences régulatrices spécifiques localisées dans le promoteur de l'élément, souvent similaires à des motifs connus permettant l'activation de gènes en réponse au stress. Ces séquences particulières confèrent aux éléments transposables qui les portent la possibilité de modifier l'expression des gènes proximaux en réponse à l'environnement, comme l'illustrent plusieurs cas (Beguiristain *et al.*, 2001; Kimura *et al.*, 2001; Grandbastien *et al.*, 2005; Salazar *et al.*, 2007; Woodrow *et al.*, 2011).

**En conclusion, les éléments transposables constituent un écosystème complexe et génèrent, par leur mouvement, une fluidité du génome et de l'épigénome.** Une fois générés, les allèles et épiallèles produits par les éléments transposables sont soumis à la dérive et à la migration, mais aussi à la sélection, dont la force va dépendre de l'effet phénotypique engendré. Cette fluidité, bien plus élevée que celle observée au niveau des séquences géniques elles-mêmes, et pouvant avoir un effet plus subtil qu'un simple dysfonctionnement des protéines codées, pourrait être la clé d'une modulation fine du phénotype. Enfin, par leur effet concerté sur plusieurs gènes, les éléments transposables ont sans doute un rôle important dans la régulation des caractères complexes, qui est actuellement proposé, mais qui reste à valider de façon directe.

**Le maïs est un modèle de choix pour analyser la dynamique des éléments transposables et leur impact phénotypique.** Son génome, de 2.1 Gb, est rempli à plus de 85% par des séquences d'éléments transposables (Schnable *et al.*, 2009), qui présentent un fort polymorphisme entre individus (Wang and Dooner, 2006). Il contient de nombreux éléments *cis*-régulateurs proximaux et distaux (Ricci *et al.*, 2019). Ces derniers sont localisés en moyenne à 60kb des gènes et sont pour plus de 60% d'entre eux associés à des séquences d'éléments transposables. Plusieurs cas d'insertions d'éléments transposables ont été montrés comme acteurs de la modulation de l'expression de gènes localisés à plusieurs dizaines de kilobase, et sous-tendent des phénotypes majeurs impliqués dans la domestication (Studer *et al.*, 2011) ou l'adaptation de l'espèce (Yang *et al.*, 2013; Zhang *et al.*, 2018). Certains cas sont liés à une modification de la méthylation locale de l'ADN (Xu *et al.*, 2020). Domesticqué il y a 9000 ans à partir de la téosinte, le maïs est aujourd'hui cultivé dans le monde entier. Du fait de cette large diffusion en dehors de son centre d'origine, le maïs s'est adapté à de nombreux environnements. Ses ressources génétiques sont structurées en plusieurs groupes génétiques

qui reflètent cette diffusion et ces adaptations locales, et permettent l'analyse de la contribution des éléments transposables à l'évolution de l'espèce.

## 2- Contribution des éléments transposables à la variation de la taille des génomes

### 2.1 Introduction

Mes premiers travaux de recherche, d'abord en thèse au laboratoire Ecologie, Systématique et Evolution (Orsay) sous la direction d'Olivier Panaud, puis en post-doctorat au laboratoire du professeur Jeffrey Bennetzen à l'université de Géorgie aux Etats-Unis, ont porté sur la compréhension de l'origine des variations en taille de génome chez les plantes à fleurs, et en particulier la contribution des éléments transposables à ce processus. Je me suis intéressée à cette question sur plusieurs échelles de temps, en comparant des espèces issues d'un même genre (*Oryza*, 15 millions d'années), d'un même ordre (deux espèces d'Asparagales ayant divergé il y a 87 millions d'années) ou au sein de groupes plus larges comme celui des angiospermes (analyse d'espèces ayant divergé de 10 à 150 millions d'années). Mes travaux ont permis de proposer un nouveau modèle prenant en compte à la fois les mécanismes d'augmentation, mais aussi de diminution du nombre de bases ('increase/decrease model'), qui a permis une avancée conceptuelle importante.

### 2.2 Contexte scientifique

Dès la fin des années 1940, et avant même la découverte de la molécule d'ADN, Colette et Roger Vendrely observent que le contenu du noyau est invariable entre individus d'une même espèce, mais très variable entre les espèces. Ils proposent que l'étude systématique du contenu des noyaux à travers de nombreuses espèces devrait permettre des « suggestions intéressantes en ce qui concerne le problème de l'évolution » (Vendrely and Vendrely, 1950). A partir des années 1960, avec l'avancée apportée par les courbes de *Cot*, c'est non seulement la taille de ce contenu, mais également sa composition, qui peut être analysé. Ceci permet de montrer d'une part que les séquences uniques (donc la majorité des gènes) sont minoritaires dans les génomes, et d'autre part que la fraction 'moyennement répétée' (qui contient la majorité des éléments transposables) est variable entre les espèces (Flavell, 1986). Il apparaît ainsi que l'activité des éléments transposables pourrait être un moteur important des variations en taille de génome observées. Parmi ceux-ci, les rétrotransposons à LTR, par leur cycle de type 'copier-coller' ont la capacité d'envahir les génomes. Ce sont aussi les éléments les plus abondants des 'gros' génomes de plantes. Ces deux observations, combinées à l'analyse fine des tailles de génomes au sein de la famille des Poacées ont fait émerger le modèle 'increase only', dans lequel les génomes augmentent en taille progressivement par l'activité des rétrotransposons à LTR (Bennetzen and Kellogg, 1997). Dans ce modèle, les variations de taille de génome observées entre les espèces sont dues à des différences d'amplification des éléments qu'ils contiennent. Le début du séquençage de génomes dans les années 2000 et la possibilité de comparer les séquences de régions orthologues entre plusieurs espèces remet cependant ce modèle en cause. D'une part, l'observation de délétions plus fréquentes chez la drosophile (180 Mb) que chez le criquet (1,9 Gb), suggère que les différences en taille de génome pourraient être dues à des différences de taux de délétion

entre les espèces (Petrov *et al.*, 2000). D'autre part, l'observation de « solo-LTR », vraisemblablement issus de la recombinaison entre les deux LTR d'un même élément, suggère que la quantité d'ADN apportée par un rétrotransposon à LTR pourrait être, au moins partiellement, contre-balançée (Shirasu *et al.*, 2000). S'il est alors clair que les génomes peuvent à la fois augmenter (notamment par l'activité des éléments transposables, et en particulier celle des rétrotransposons à LTR chez les plantes) et diminuer (par délétion et recombinaison intra- ou inter- élément) en taille, la mise en place d'un nouveau modèle d'évolution de la taille des génomes requiert de caractériser l'intensité et la temporalité relative de ces deux processus.

### 2.3 Dynamique du génome au sein du genre *Oryza*

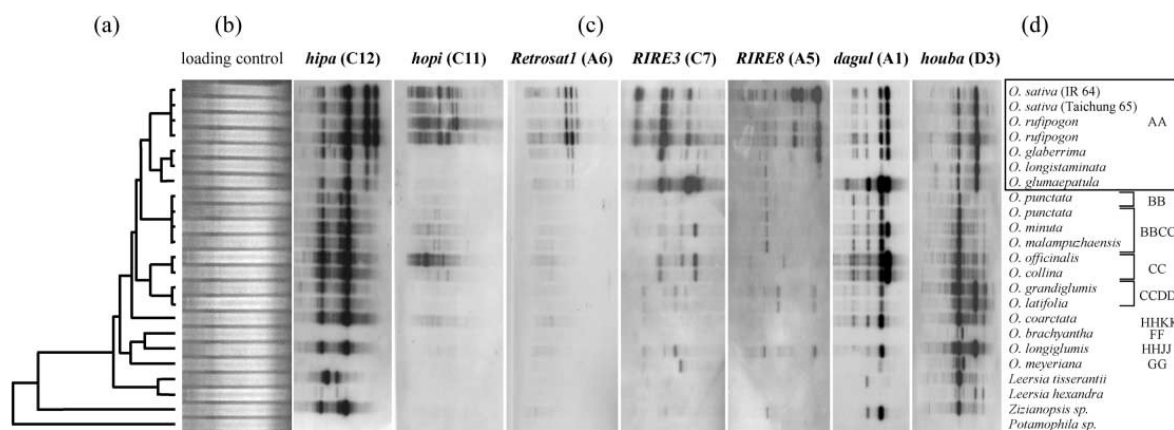
3<sup>ème</sup> stage de DEA et thèse au laboratoire Ecologie, Systématique et Evolution (2001-2004)

Séjour dans un autre laboratoire français (Institut Jacques Monod, 4 mois en 2004)

Séjour à l'International Rice - Research Institute aux Philippines (5 mois en 2003)

*Encadrement : Marion Vankerssaver (Master 1), Shunjung-Shim (thèse), Jayson Talag (technicien)*

Afin de répondre à cette question, et dans un contexte historique dans lequel le génome du riz n'était pas encore séquencé, j'ai d'abord cherché à caractériser des éléments transposables ayant participé aux différences de taille de génome observées au sein de la famille des Poacées. La technique de « Representational Difference Analysis » (ou RDA) avait été appliquée au laboratoire entre le riz et le millet, deux Poacées à petit génome (430 et 500 Mb, respectivement) et avait permis de cloner 11 fragments d'ADN présents dans le génome du riz et absent de celui du millet. En analysant la séquence alors disponible du génome du riz asiatique *Oryza sativa* (correspondant à environ 30% de la séquence totale), j'ai pu montrer que ces fragments sont localisés dans 7 éléments transposables, dont 4 qui n'avaient encore jamais été décrits, et que j'ai caractérisés. En les utilisant comme sondes pour l'hybridation par Southern blot sur les ADN des espèces du genre *Oryza* et de trois genres apparentés, j'ai montré que ces éléments ont eu une activité relativement récente et que plusieurs amplifications indépendantes de ces éléments transposables se sont vraisemblablement produites au sein du genre *Oryza* (Figure 1).



**Figure 1.** Analyse par 'Southern blot' des clones RDA dans les espèces du genre *Oryza*

a. Arbre phylogénétique (d'après Ge *et al.*, 1999), montrant les relations génétiques entre les espèces analysées (listées dans le panneau d). Le type de génome est indiqué en d lorsqu'il est disponible. b. Profil des ADN digérés après électrophorèse et avant transfert sur membrane. c. Autoradiogrammes obtenus avec les sept clones (indiqués entre parenthèses). Le nom de l'élément correspondant est indiqué en haut. Le sens de migration est de gauche à droite.

Ces résultats, réalisés au cours de mon stage de DEA et du début de ma thèse, ont fait l'objet de la publication :

O. Panaud, **C. Vitte**, J. Hivert, S. Muzlak, J.D. Talag, D.S. Brar and A. Sarr (2002). Genomic differentiation between rice (*Oryza sativa* L.) and foxtail millet (*Setaria italica* L. Beauv.) revealed through representational difference analysis. *Mol. Gen. Genomics*. 268(1):113-21.

Ces éléments, montrant une expansion récente, étaient de bons candidats pour tester l'existence d'une force de diminution et d'estimer la force relative des deux processus. J'ai ainsi analysé la dynamique d'amplification et de recombinaison pour les trois rétrotransposons à LTR de type *gypsy*. J'ai recherché les copies homologues de ces trois éléments dans la séquence génomique de l'espèce *O. sativa* afin de caractériser les copies complètes et les solo LTR. Pour chaque famille, l'analyse combinée des relations phénétiques entre les copies et de la divergence entre les LTR des copies complètes m'a permis de mettre en évidence des événements d'amplification successifs de copies « maitres » au cours des cinq derniers millions d'années et ayant eu lieu à des périodes significativement différentes. L'analyse des solo-LTR m'a ensuite permis de montrer que, pour les trois éléments, des événements de recombinaison ont eu lieu au cours de cette même période de façon concomitante avec l'amplification des copies « maitres ». Un exemple de dynamique d'amplification (celle de la famille Hopi) est présenté dans la Figure 2.

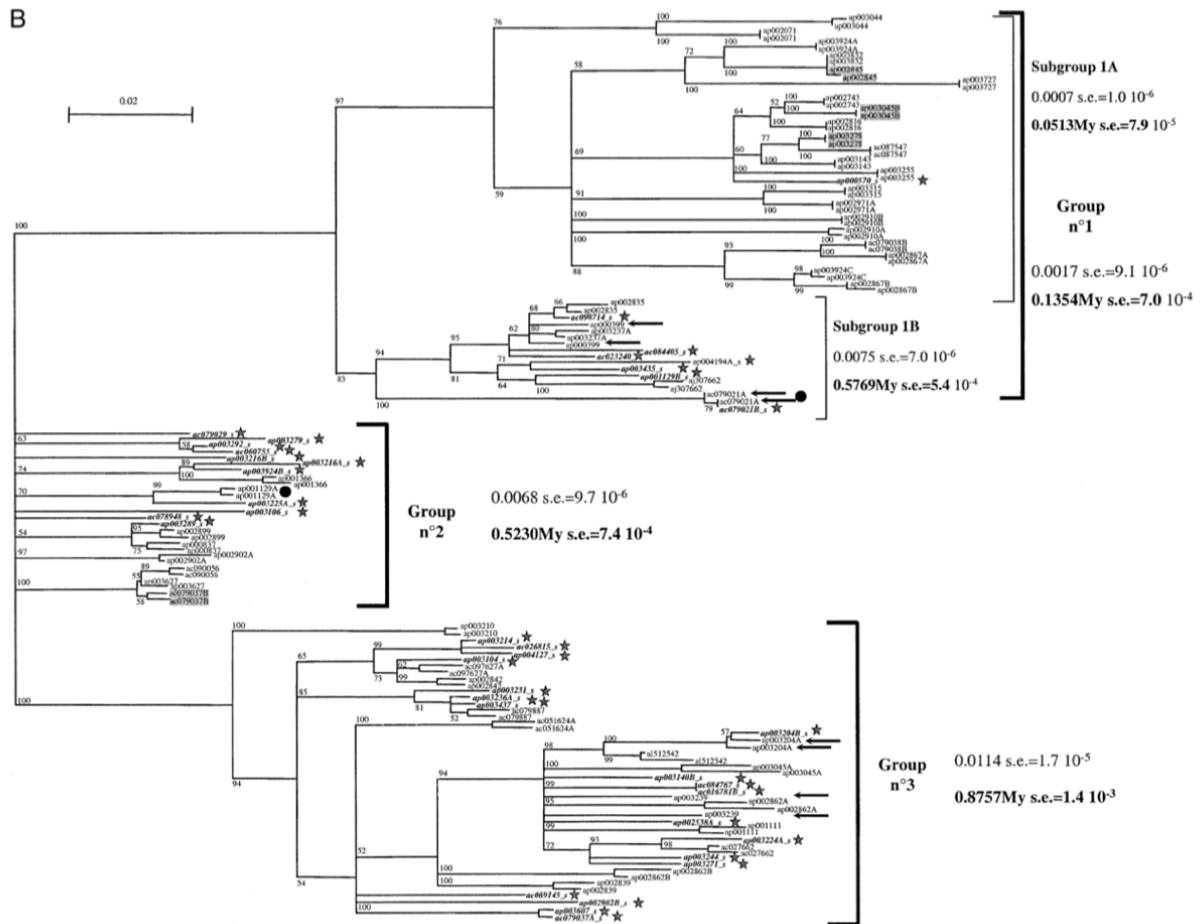
Ces résultats ont conduit à la publication ci-dessous, pour laquelle je suis autrice correspondante.

**C. Vitte**, O. Panaud (2003). Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol. Biol. Evol.* 20(4):528-40.

Ces résultats nous ont mené à proposer un modèle d'augmentation/diminution pour l'évolution du génome du riz, dans lequel les processus d'amplification et de recombinaison entraînent des variations de la taille du génome de façon concomitante. J'ai ainsi rédigé un article de synthèse présentant ce nouveau modèle ainsi qu'une synthèse des connaissances relatives à la contribution des éléments transposables à la dynamique de la taille des génomes chez les angiospermes. Il a été publié dans une édition spéciale de la revue *Cytogenetics and Genome Research* sur le thème « Retrotransposons, Retrotransposition and Genome Evolution » :

**C. Vitte**, O. Panaud (2005). LTR retrotransposons and plant genome size: emergence of the increase/decrease model. *Cytogenetics and Genome Research*. 110(1-4):91-107.

Si cette étude a apporté des éléments nouveaux et originaux sur la dynamique temporelle des processus d'amplification et d'élimination, elle est restée assez limitée quant à l'analyse de l'ampleur de ces deux mécanismes antagonistes, notamment du fait du caractère partiel de la séquence génomique du riz disponible au moment de l'étude, et de la prise en compte uniquement de la force de recombinaison entre LTR.



**Figure 2. Arbre 'neighbour-joining' des séquences LTR de l'élément *Hopi***

Les valeurs de 'bootstrap' supérieures à 50 sont indiquées, les nœuds présentant des valeurs inférieures à 50 ont été traités comme non résolus. Les solo-LTR sont indiqués par des étoiles. Pour chaque groupe, la moyenne et la variance de divergence des LTR sont indiquées sous le numéro du groupe ; le temps d'insertion estimé correspondant est indiqué en gras.

Une question majeure restée non élucidée était la possibilité, ou non, pour un génome de revenir à sa taille initiale par un processus de délétion. Or, la formation de solo-LTR, si elle permettait d'éliminer une grande part de la séquence insérée, ne permettait pas ce retour. Afin de répondre à cette question, il fallait également prendre en compte d'autres types de délétions, indépendantes de la recombinaison entre LTR. J'ai ainsi réalisé une étude approfondie de la dynamique d'amplification et de délétion sur un ensemble de 41 familles de rétrotransposons à LTR (16 de type *gypsy* et 25 de type *copia*), et en analysant conjointement copies complètes, solo-LTR, copies tronquées et petites délétions. Cette analyse, rendue possible par la parution de la séquence génomique quasi complète de l'espèce *O. sativa* en 2003, nécessitait aussi la mise en place d'outils d'analyse plus systématiques, que nous avons réalisée grâce à une collaboration avec Hadi Quesneville (à l'époque à l'Institut Jacques Monod, Paris). Un point critique de la caractérisation de l'intensité de la force d'élimination était de ne pas comptabiliser comme « copies tronquées » des séquences de petites tailles insérées dans la copie de référence, qui auraient fortement biaisé les résultats. Par une analyse fine des copies détectées, et après un gros travail de filtres afin d'éviter les faux-positifs, j'ai

travaillé avec un ensemble de 19 familles (10 de type *copia* et 9 de type *gypsy*) représentant un total de 5154 copies. J'ai pu mettre en évidence que ces familles ont connu des périodes d'amplifications massives et successives au cours des 5 derniers millions d'années, dont le nombre, l'amplitude et la date varie entre les familles, révélant une histoire d'amplification particulière pour chacune d'entre elles.

Par l'observation d'un nombre de LTR solitaires variable d'une famille à l'autre et corrélé positivement avec la taille des LTR, j'ai pu suggérer que la fréquence de la formation de solo-LTR est un processus dépendant de la famille. Enfin, l'analyse conjointe de la taille cumulée de ces copies lors de l'insertion (52,3 Mb) et de leur taille cumulée actuelle (20,2 Mb) m'a permis d'estimer une perte de 32,1 Mb (61,4%) depuis leur insertion, indiquant une force importante de la délétion. L'estimation de ce taux de délétion et des âges détectées pour chaque groupe amplifié m'a permis de prédire que la demi-vie des séquences de rétrotransposons LTR est d'environ 19 millions d'années chez le riz.

Ce travail, que j'ai terminé en parallèle de mon premier séjour post-doctoral, a fait l'objet de la publication :

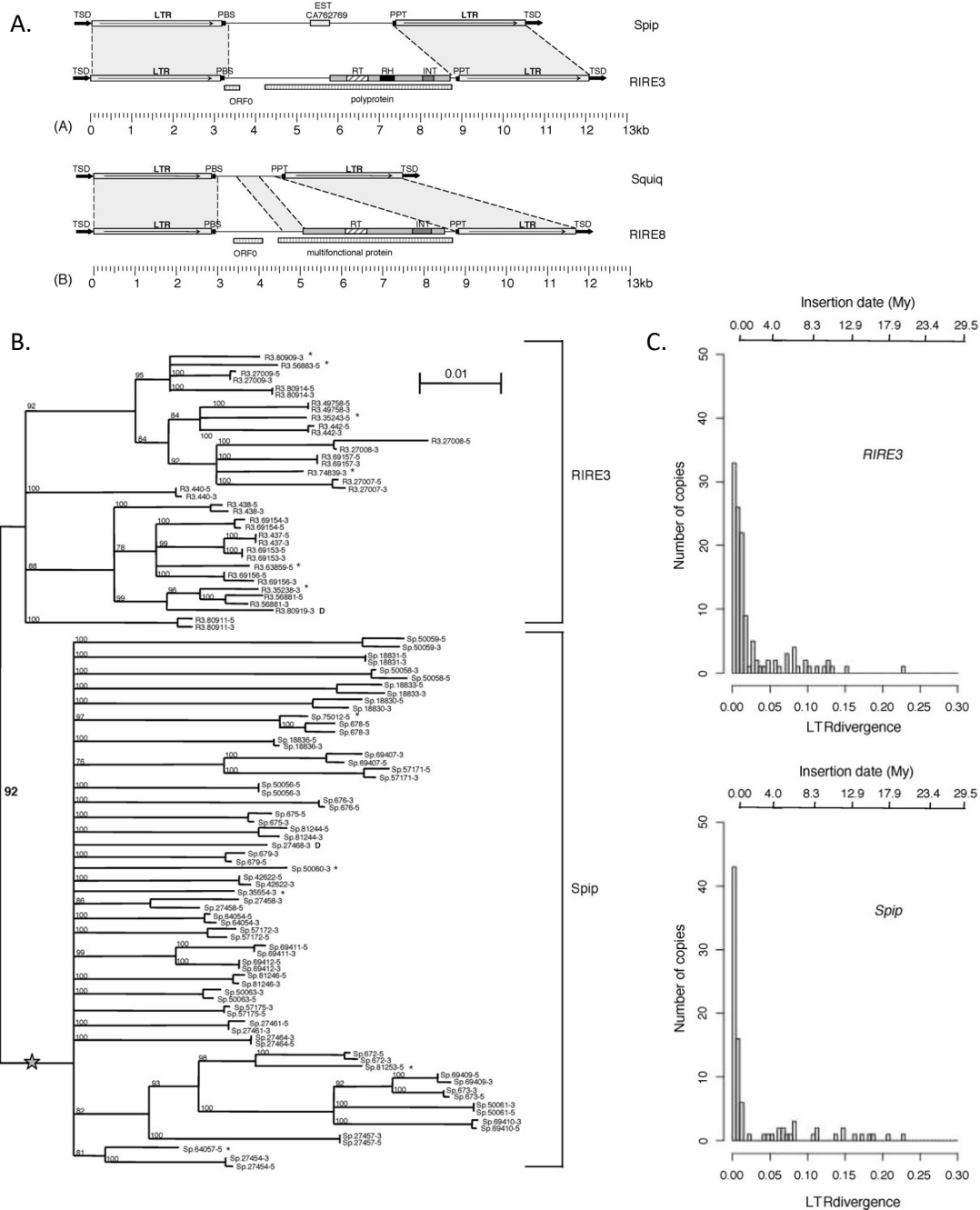
**C. Vitte, O. Panaud, H. Quesneville.** LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss. (2007) BMC Genomics 8:218.

Au cours de cette dernière étude, j'ai mis en évidence et caractérisé plusieurs Large Retrotransposon Derivative (LARD) apparentés à des éléments complets, présentant les mêmes séquences de LTR mais des séquences internes différentes dépourvues de séquences codantes, et que j'avais écartés de l'analyse afin de ne pas biaiser les résultats. Leur découverte permettait néanmoins de mieux comprendre la dynamique des rétrotransposons à LTR non autonomes et leur contribution à la dynamique du génome du riz. J'ai ainsi réalisé une étude approfondie de deux couples de rétrotransposons autonomes/non autonomes : *RIRE3/Spip* et *RIRE8/Squiq* (Figure 3A). Par la combinaison d'approches phénétique et de datation des copies, j'ai pu montrer que *Spip* et *Squiq* sont chacun d'origine monophylétique, et se sont amplifiés au cours des deux derniers millions d'années (Figure 3). Enfin, alors que l'élément *Squiq* s'est amplifié de façon modérée, l'élément *Spip* atteint un nombre de copies similaires à celui de son homologue *RIRE3*, et présente une dynamique d'amplification plus forte que ce dernier au cours des deux derniers millions d'années (Figure 3C). L'ensemble de ce travail a mis en évidence que les éléments LARD ne sont pas rares et participent à la dynamique du génome du riz de façon similaire aux rétrotransposons autonomes.

Ce travail, que j'ai terminé en parallèle de mon séjour post-doctoral, a fait l'objet de la publication :

**C. Vitte, C. Chaparro, H. Quesneville, O. Panaud** (2007). *Spip* and *Squiq*, two novel rice non-autonomous LTR retro-element families related to *RIRE3* and *RIRE8*. Plant Science, 172(1):8-19.

C'est au cours de ce travail que j'ai eu ma première expérience d'encadrement à l'occasion du stage de Marion Vankersschaver (M1, 2003), à qui j'ai appris les méthodes bio-informatiques d'alignements de séquences d'éléments transposables pour étudier l'accumulation de délétions.

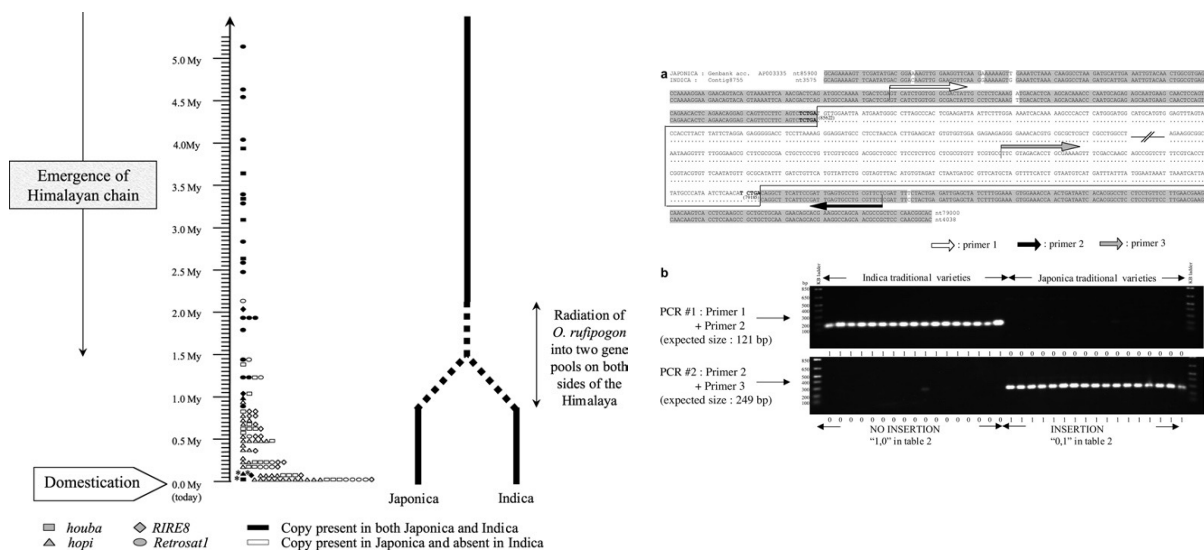


### Figure 3. Présentation et dynamique des deux couples LARD/rétrotransposon à LTR

A. Comparaison des séquences des éléments *Spip* et *RIRE3*, et *Squiq* et *RIRE8*. B. Histogramme des dates d'insertion de *Spip* et *RIRE3*, obtenues par l'analyse de la divergence des LTR des copies complètes. C. Arbre de type 'neighbour-joining' des séquences LTR de *RIRE3* et *Spip*, illustrant la séparation en deux groupes distincts.

Enfin, l'observation d'amplifications très récentes au sein de l'espèce *O. sativa* a ouvert la possibilité d'utiliser les rétrotransposons à LTR comme marqueurs de temps afin d'affiner le modèle de domestication du riz asiatique. En effet, l'origine de domestication de cette espèce était (et est toujours !) très débattue. Les rétrotransposons à LTR n'étant pas capables

d'excision, ils représentaient des marqueurs stables, ne présentant pas réversion du caractère 'présence', et nous avons eu l'idée de les utiliser afin de dater la divergence entre les deux groupes de riz cultivés asiatiques : Indica et Japonica. Grâce à la diffusion publique des génomes de deux variétés de riz, Nipponbare (type Japonica) et 93-11 (type Indica), j'ai analysé la présence/absence de copies 110 de rétrotransposons LTR entre ces deux variétés, afin de distinguer les copies partagées et les copies spécifiques à chaque variété. La datation de l'insertion de ces copies (Figure 4) m'a ensuite permis d'estimer que ces deux variétés ont divergé l'une de l'autre il y a au moins 200 000 ans, donc de façon très antérieure à la domestication du riz (il y a environ 10 000 ans, à la fin du Néolithique). Pour valider ces résultats sur l'ensemble des deux groupes, j'ai ensuite complété cette analyse *in silico* par une étude de biologie moléculaire (« RBIP assay ») des polymorphismes d'insertion au sein d'un large éventail de variétés traditionnelles de riz de types Indica et Japonica. Ces variétés tropicales étant incapables de pousser en France, j'ai réalisé un séjour de 5 mois à l'International Rice Research Institute (IRRI) aux Philippines pour réaliser la culture et l'extraction des ADN. J'ai été accueillie dans l'équipe du professeur Darshan S. Brar, qui a apporté son expertise pour valider le phénotype 'Indica' ou 'Japonica' des variétés que j'ai cultivées. Les données expérimentales ainsi générées m'ont permis une deuxième validation, renforçant la proposition selon laquelle les riz asiatiques de type Indica et Japonica sont issus de deux événements indépendants de domestication.



**Figure 4.** Chronologie des insertions de quatre familles de rétrotransposons LTR et étude comparative de leur présence/absence entre les génotypes de type Indica et Japonica.

A gauche : Age des copies complètes détectées dans le génome de Nipponbare (Japonica) et 93-11 (Indica) sur la base de la divergence entre leurs LTR. Les couleurs indiquent si les copies sont partagées (noir) ou présentes dans Nipponbare mais pas dans 93-11 (blanc). Le trait noir indique le scénario de la radiation des deux génotypes. La partie en pointillé indique la période au cours de laquelle les deux pools génétiques ont divergé, la position du nœud est arbitrairement placée au milieu de cette période. A droite : Exemple de génotypage d'une copie par PCR sur plusieurs génotypes de type Indica et Japonica. a. Localisation des amorces. b. Résultat de la PCR.

Ce travail a fait l'objet de la publication :

**C. Vitte**, T. Ishii, F. Lamy, D.S. Brar and O. Panaud (2004). Genomic paleontology provides evidence for two distinct origins of asian rice (*Oryza sativa*, L.). *Mol. Gen. Genomics*. 272(5):504-11.

Lors de ce séjour dans le groupe 'Wide Hybridization' dirigé par D. Brar, j'ai également formé deux personnes, Jayson Talag (alors technicien), et Jung-Yung Shim (alors doctorant), à la technique de RDA. Ceci a ensuite permis à Jung-Yung Shim d'appliquer la technique sur deux riz distants, *O. minuta* et *O. sativa*, et a permis la publication :

J. Shim, O. Panaud, **C. Vitte**, M. Mendioro, D. S. Brar (2010). RDA derived *Oryza minuta*-specific clones to probe genomic conservation across *Oryza* and introgression into rice (*O. sativa* L.), *Euphytica* 176:269–279

Lors de ce séjour, Jayson, Shung-Yung et moi avons également chercher à développer une approche permettant la validation des résultats de la RDA par une approche d'hybridation sur puce à ADN. Le principe était de valider la présence/absence des régions détectées en RDA par une approche de puce à ADN : sur une puce contenant les ADN issus des clones RDA, les ADN de chacune des espèces, marqués chacun par des fluorochromes rouge ou vert, étaient hybridés. La lecture de la couleur indiquait si le clone était spécifique d'une espèce (couleur rouge ou verte) ou partagée entre les deux espèces (couleur intermédiaire). La mise au point de cette technique a été difficile, et n'a permis d'aboutir que partiellement dans le temps imparti, notamment du fait des problèmes de répétabilité des résultats après les 'dye swap'. Mais l'ensemble de ces mises au point dans un laboratoire nouveau, et avec des technologies nouvelles que je ne connaissais pas, m'a donné le goût de la génomique humide et la capacité de m'adapter, que je conserve encore aujourd'hui.

A la fin de ma thèse, je souhaitais explorer la question de la taille des génomes en lien avec les éléments transposables sous deux nouveaux aspects : (i) analyser la structure et la dynamique de génomes de taille plus grande, (ii) aborder la question à une échelle évolutive plus large que celle du genre.

Pour cela, j'ai choisi de rejoindre le laboratoire du professeur Jeffrey Bennetzen, dont j'avais lu de nombreux articles qui m'avaient passionnée (il avait notamment été le premier à caractériser les structures si complexes d'éléments transposables imbriqués les uns dans les autres, qui sont si caractéristiques du maïs, et avait proposé très tôt le modèle 'increase only' qui a permis de débattre du sujet et de le faire progresser). Beaucoup des travaux de ce laboratoire étaient pionniers en génomique, et j'appréciais chez Jeff sa large culture en biologie des plantes, son côté visionnaire et son inventivité pour développer des techniques qui permettraient de répondre à la question posée de manière 'cost-effective'. Pour replacer cela dans le contexte historique, en 2004, seuls deux génomes de plante avaient été séquencés et assemblés (celui d'*Arabidopsis thaliana*, publié 2000 et celui du riz, publié en 2002), ce qui avait pris plusieurs années à des consortiums internationaux malgré leur taille réduite (125 Mb et 430 Mb, respectivement). La génération de la séquence assemblée pour des génomes de grande taille comme celui du maïs ou du blé était tout juste au stade de réflexion (le génome du maïs - dernier génome de plante à avoir été assemblé par carte physique BAC sous-clonée

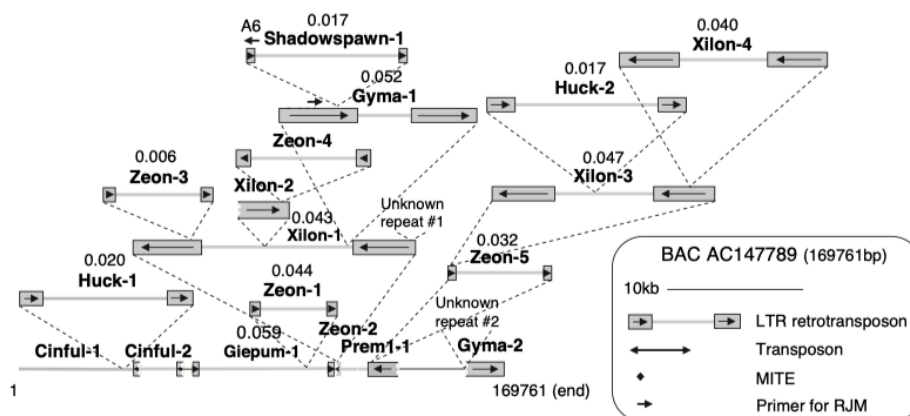
en cosmides - sera publié en 2009 après plusieurs années de travail du consortium américain). Il fallait donc trouver des stratégies permettant d'estimer la structure et la dynamique de tels génomes par des approches incomplètes mais informatives.

## 2.4 Dynamique du génome du maïs

Post-doctorat à l'international au laboratoire de Jeffrey Bennetzen de l'université de Géorgie, Etats-Unis (2005-2006)

*Financement acquis : Bourse Lavoisier du ministère des affaires étrangères*

Mon premier projet, réalisé en collaboration avec Renyi Liu (alors étudiant en thèse), a consisté à caractériser la structure du génome du maïs, par l'analyse de la séquence de 74 BACs sélectionnés de façon aléatoire. Par recherche de gènes candidats ayant des homologues dans d'autres espèces de plantes à fleurs, cette étude a permis d'estimer que le génome nucléaire du maïs contient environ 37 000 gènes fonctionnels (plus un ensemble de 5500 gènes sévèrement tronqués correspondant probablement à des pseudogènes). Nous avons également montré que la distribution des gènes est inégale le long du génome, avec une densité variant de 0 à 7,9 gènes par 100 kb entre les différents BAC. En outre, 30% des BACs ne contenaient aucun gène, et la majorité des BAC à contenu génique ne contenait qu'un seul gène, indiquant qu'une partie importante du génome du maïs est vide de gènes. Grâce à une annotation manuelle en recherchant des séquences proches de séquences d'éléments transposables déjà connus et par une analyse *ab initio* fondée sur la structure des éléments, j'ai pu montrer que 68% de ces régions correspondent à des éléments répétés. L'annotation fine de 6 régions dépourvues de gènes pour lesquelles nous avons pu obtenir un assemblage quasi-complet de la séquence m'a permis de montrer que les régions non géniques contiennent des empilements d'éléments pleine taille (Figure 5), et que les régions géniques et non géniques contiennent des rétrotransposons différents.



**Figure 5. Représentation graphique de la structure du BAC AC147789**

Au-dessus de chaque rétrotransposon à LTR est indiqué le nom de la famille et le numéro de la copie, ainsi que la divergence entre les séquences des LTR. Les emplacements des amorces utilisées pour le marqueur RJM permettant de génotyper l'insertion de la copie *Shadowspawn-1* dans la copie *Gyma-1* sont représentés par des flèches.

D'autre part, l'utilisation de marqueurs PCR uniques marquant la jonction d'éléments transposables empilés (« Repeat Junction Markers », ou RJM) pour génotyper une population

de cartographie m'a permis de montrer que parmi ces 6 régions, seules deux étaient localisées proche de péricentromère ou de renflement hétérochromatinien, suggérant la présence de grandes régions vides de gènes et formées d'empilements d'éléments transposables en dehors de ces régions. L'analyse des dates d'insertion m'a permis de mettre en évidence deux grandes vagues d'activité des rétrotransposons à LTR au cours des deux derniers millions d'années, la vague la plus récente étant plus marquée au niveau des régions géniques et suggérant un renouvellement plus rapide des éléments transposables dans ces régions.

Les résultats ont démontré que le séquençage et l'annotation minutieuse d'un échantillon de BAC sélectionnés au hasard peuvent fournir une description robuste d'un génome végétal complexe comme celui du maïs. Ce travail a été publié dans un article dont je suis première co-auteurice :

R. Liu\*, C. Vitte\*, J. Ma, A.A. Mahama, T. Dhaliwayo, M. Lee, J.L. Bennetzen (2007). A GeneTrek analysis of the maize genome. Proc. Nat. Acad. Sc. U.S.A 104:11844-49.

Cette étude a été le déclenchement de l'intérêt que je porte au génome du maïs, et qui a mené au développement du projet de recherche que j'ai proposé pour mon recrutement au CNRS.

## 2.5 Éléments transposables et taille de génome au sein des Angiospermes

Post-doctorat à l'international au laboratoire de Jeffrey Bennetzen de l'université de Géorgie, Etats-Unis (2005-2006)

*Financement acquis : Bourse Lavoisier du ministère des affaires étrangères*

Mon deuxième projet visait à tester le modèle 'increase/decrease' sur de plus grandes distances évolutives, en comparant le contenu de génomes d'espèces diploïdes ayant divergé à des temps évolutifs variés, ayant des tailles de génomes contrastées et dont les relations phylogénétiques étaient bien décrites : *Lotus japonicus* (470 Mb), *Medicago truncatula* (470 Mb), le maïs (2300 Mb), l'orge (5400 Mb) et *Triticum monococcum* (6100 Mb). L'analyse d'un échantillon de ces génomes (voire méthodologie de la partie 2.4) m'a permis de montrer que toutes ces espèces présentaient une mobilité des rétrotransposons à LTR au cours des 4 derniers millions d'années, mais que seuls les gros génomes contenaient des familles de rétrotransposons LTR ayant plus de 10 000 copies. Ceci indique d'une part que l'amplification différentielle des rétrotransposons à LTR est un facteur important dans la variation de la taille des génomes d'angiospermes, et d'autre part que l'extension des gros génomes a eu lieu par amplification massive de quelques familles, plutôt que par amplification modérée d'un grand nombre de familles distinctes. Enfin, l'analyse de la structure des copies (solo-LTR et copies tronquées) a révélé que les mécanismes de diminution (recombinaison inégale et illégitime) ont lieu dans toutes les espèces, mais avec des efficacités relatives variables. Cependant, l'observation que ces variations ne suivent ni la taille du génome ni la phylogénie indique que l'agressivité relative de ces deux processus change au cours de l'évolution, et n'est pas un facteur majeur des différences de tailles de génome.

L'ensemble de ces résultats m'a permis de proposer un modèle dans lequel les différences de taille de génome observées sont essentiellement liées aux différences relatives d'amplification des rétrotransposons LTR, avec un rôle moins significatif du processus de diminution.

Ces travaux ont fait l'objet de l'article :

**C. Vitte**, J.L. Bennetzen (2006). Analysis of Retrotransposon Structural Diversity Uncovers Properties and Propensities in Angiosperm Genome Evolution. Proc. Nat. Acad. Sc. U.S.A. 103(47):17638-43.

## 2.6 Structure des génomes et dynamique des éléments transposables chez les Asparagales (oignon et asperge)

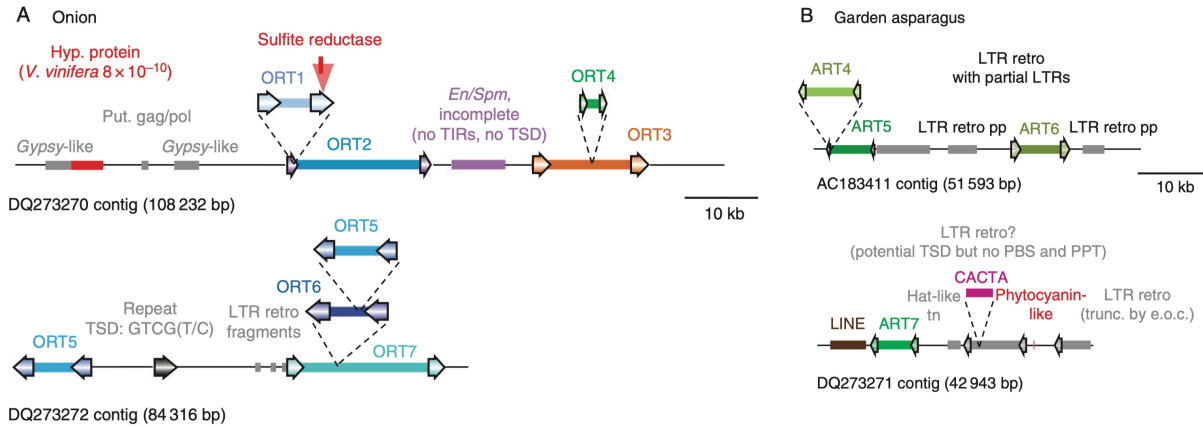
Post-doctorat à l'international au laboratoire de Jeffrey Bennetzen de l'université de Géorgie, Etats-Unis (2005-2006) et poste actuel (2008-2010)

Au sein des monocotylédones, la quasi-totalité des études sur la dynamique des génomes était issue de la famille des Poacées, qui contient les céréales, des plantes d'intérêt agronomique particulièrement important pour lesquelles des budgets conséquents avaient été débloqués. Ce biais posait la question de déterminer si la structure et la dynamique des génomes observée chez les Poacées était aussi vraie dans d'autres taxons. Dans mon troisième projet post-doctoral, j'ai donc étudié les propriétés structurales des génomes et la dynamique des éléments transposables pour deux espèces de l'ordre des Asparagales : l'asperge (*Asparagus officinalis*, 1300 Mb) et l'oignon (*Allium cepa*, 16 400 Mb). Ces deux espèces d'intérêt agronomique ont divergé il y a environ 87 millions d'années (Bremer, 2004), un ordre de grandeur similaire à celui de la radiation des espèces de la famille des Poacées (environ 60 millions d'années). L'analyse fine de la séquence de plusieurs clones BAC m'a permis d'améliorer l'annotation des gènes et de découvrir de nouveaux éléments transposables chez ces deux espèces (Figure 6). Elle a révélé des copies pour la plupart intactes (présentant deux LTR), empilées en structures imbriquées et ayant des dates d'insertion allant jusqu'à 6 millions d'années. L'analyse de la fréquence de lectures 'shotgun' correspondant à chaque élément découvert dans les BAC a révélé une grande variabilité entre les différentes familles (Tableau 1) et m'a permis de mettre en évidence certaines familles représentant plus de 500 Mb du génome de l'oignon.

L'ensemble de cette étude a révélé une série de caractéristiques identiques à celles observées chez les Poacées. L'annotation fine que j'ai réalisée a également fourni des ressources importantes pour la communauté travaillant sur ces espèces, et qui servent encore aujourd'hui.

Ce travail, que j'ai ébauché au laboratoire de Jeff Bennetzen et terminé au laboratoire du Moulon après mon recrutement, a été publié dans l'article suivant, pour lequel je suis autrice correspondante. Il fait référence dans la communauté des Asparagales, et les conclusions sont similaires à celles décrites par l'analyse du génome complet (Finkers *et al.*, 2021).

**C. Vitte**, M. Estep, J. Leebens-Mack, J. Bennetzen (2013). Young, intact and nested retrotransposons are abundant in the onion and asparagus genomes. Annals of Botany. doi: 10.1093/aob/mct155



**Figure 6. Annotation des grands contigs issus des génomes de l'oignon et de l'asperge**

A. Contigs du génome de l'oignon. B. Contigs du génome de l'asperge. Les noms des contigs sont indiqués en dessous de chaque dessin et une échelle de 10 kb est indiquée à droite. Les gènes sont représentés en rouge. Les autres couleurs représentent des éléments transposables entièrement caractérisés. Les barres grises indiquent des régions présentant une homologie avec des éléments transposables dont la structure complète n'a pas pu être caractérisée. Les rétrotransposons à LTR sont représentés par des lignes avec des flèches indiquant la position des LTR. Abréviations : put, putatif ; hyp., hypothétique ; TIR, répétitions terminales inversées; TSD, duplication du site cible ; retro, rétrotransposon ; tn, transposon ; pp, polyprotéine ; trunc, tronqué ; e.o.c., end of contig.

**Tableau 1. Fraction génomique occupée par chacun des rétrotransposons à LTR découverts**

Les rétrotransposons à LTR caractérisés chez l'asperge sont nommés 'ART', ceux caractérisés chez l'oignon sont nommés 'ORT'.

Name	Class	Genomic shotgun sequence occupied		Portion of genome occupied (Mb)	
		Length (bp)	Percentage	From original data set	95 % CI from resampling
ART1	<i>Copia</i>	2802	0.1	2	NE
ART2	<i>Copia</i>	10 931	0.5	6	NE
ART3	<i>Copia</i>	0	0.0	0	NE
ART4	Unknown	109 918	4.7	61	51–70
ART5	Unknown	27 391	1.2	15	12–19
ART6	Unknown	36 565	1.6	20	15–25
ART7	Unknown	67 872	2.9	37	28–50
ART8	Unknown	49 748	2.1	27	23–32
ART9	Unknown	1028	0.0	1	NE
ART10	Unknown	22 434	1.0	12	10–16
ART11	<i>Copia</i>	8488	0.4	5	NE
ORT1	<i>Gypsy</i>	5755	0.2	33	NE
ORT2	<i>Gypsy</i>	1950	0.1	11	NE
ORT3	<i>Copia</i>	111 687	3.9	642	560–730
ORT4	Unknown	23 758	0.8	137	100–185
ORT5	<i>Gypsy</i>	20 798	0.7	120	80–160
ORT6	<i>Gypsy</i>	1378	0.0	8	NE

NE, not estimated.

Note that ORT7 was not analysed due to the possible presence of other elements inserted within it, thus creating the potential for an erroneous prediction of genome coverage.

## 2.7 Lien entre génomique structurale et génomique fonctionnelle

Post-doctorat au laboratoire Dynamique du Génome et Evolution de l'Institut Jacques Monod, Paris (2007)

*Financement acquis : Programme retour post-docs, Fondation pour la Recherche Médicale (FRM)*

A la suite de ce parcours doctoral et post-doctoral essentiellement centré sur la génomique structurale, j'ai souhaité ajouter un aspect plus fonctionnel à mes recherches. J'ai ainsi rejoint un projet visant à comprendre la contribution de la transcription des éléments transposables emboîtés à la maintenance de l'hétérochromatine chez *Drosophila melanogaster*. L'hypothèse était que les emboitements d'éléments transposables pourraient générer des transcrits sens et antisens, et ainsi produire de petits ARN interférants qui cibleraient les séquences des éléments transposables en *cis*, mais également en *trans* par leur appariement avec les ARN sens issus des autres éléments du génome. Ce déclenchement de la voie de l'ARNi induirait en retour la méthylation de ces séquences homologues. Ce projet reposait sur une approche combinée de bio-informatique et de biologie moléculaire. A partir de l'annotation des éléments transposables, j'ai prédit les régions présentant des empilements emboîtés, susceptibles de présenter une transcription sens et antisens. J'ai ensuite interagi avec la compagnie Imaxio (Clermont-Ferrand) pour construire une puce à façon contenant ces régions d'intérêt, et visant à analyser les transcrits et les marques d'histone de ces régions dans des mutants des voies épigénétiques et à différents stades du développement. Enfin, j'ai choisi quelques locus d'intérêt et j'ai analysé la production d'ARN sens et antisens et de petits ARN par des approches de biologie moléculaire. Ce travail a permis de générer des résultats préliminaires qui ont permis de premières réflexions sur le sujet. Après mon départ, le projet a été ré-orienté vers un projet un peu différent auquel je n'ai pas participé. Il a été ma première expérience sur l'analyse des petits ARN, et m'a permis de me familiariser avec les concepts de la régulation de la chromatine dans un organisme non plante.

Riche de cette expérience professionnelle dans laquelle j'avais d'abord analysé les éléments transposables sous l'angle de l'évolution structurale des génomes, puis cherché à faire le lien entre structure du génome, transcription et régulation de la chromatine, j'ai souhaité développer mes recherches autour de l'impact de la variabilité structurale sur la variabilité fonctionnelle des génomes. J'ai choisi de travailler sur le maïs, une espèce sur laquelle j'avais déjà une expertise, et dont le génome était connu pour avoir un grand nombre d'éléments transposables, variables entre individus et présents à proximité des gènes.

J'ai ainsi été recrutée au CNRS en janvier 2008 en tant que Chargée de Recherche 2ème classe, pour développer des recherches fondamentales sur l'impact de la variabilité des éléments transposables sur la variabilité de l'expression des gènes chez le maïs. Avec l'émergence de l'épigénétique, il apparaissait de plus en plus clairement qu'une partie des effets induits par les éléments transposables sur l'expression des gènes était liée à leur régulation épigénétique. La mise en place de cette thématique m'a donc amenée à faire évoluer mes connaissances vers cette discipline. Au cours des années et avec l'évolution des technologies, j'ai également fait évoluer mes compétences de la génomique structurale vers l'épigénomique et la transcriptomique.

En 2008, la séquence assemblée du premier génome de maïs (B73) n'était pas encore disponible (elle arrivera l'année d'après), et les premiers épigénomes d'*Arabidopsis* venaient juste d'être publiés (Cokus *et al.*, 2008; Lister *et al.*, 2008). En préambule pour réaliser ce projet, il fallait donc d'une part que je développe des ressources en génomique du maïs, et d'autre part que je mette au point les technologies d'analyse de l'épigénome sur un génome dix-sept fois plus gros que celui d'*Arabidopsis*.

Ainsi, j'ai choisi de décrire les activités de recherche que j'ai réalisées depuis mon recrutement au CNRS sous la forme de deux chapitres, l'un abordant les travaux que j'ai réalisés en génomique comparée (chapitre 3), et l'autre en épigénomique (chapitre 4).

## 3- Caractérisation de la variabilité structurale intra-spécifique et lien avec l'expression des gènes

### 3.1 Contexte scientifique

Depuis les années 1940, le maïs est connu pour sa grande diversité génétique, en termes de différences de taille du génome, de contenu et de taille des renflements hétérochromatiques, de contenu en ADN répété et de SNP (Brown, 1949; Buckler, Gaut, and McMullen, 2006; Messing and Dooner, 2006). Les premiers travaux de génomique comparée entre plusieurs génotypes de maïs ont révélé la très forte variabilité structurale liée à la présence/absence d'éléments transposables dans les régions intergéniques (Wang and Dooner, 2006). Au-delà de cette grande variabilité intergénique, une variabilité du nombre de gènes au sein de familles de gènes répétés en tandem (Robbins *et al.*, 1991) avait également été observée. L'ensemble de ces résultats a conduit à la proposition de la notion de pangénome (Morgante, De Paoli, and Radovic, 2007), correspondant à l'union de toutes les parties communes et variables des ADN des individus de l'espèce. Ainsi, la partie 'core' du pangénome est partagée par tous les individus de l'espèce, tandis que les parties 'accessoires' sont spécifiques à certains individus seulement. Cette notion de pangénome, étendue de celle purement génique décrite chez les bactéries (Tettelin *et al.*, 2005) va au-delà des gènes, et inclue les éléments transposables.

Si les premières analyses de la variation structurale chez le maïs ont eu lieu au niveau de régions spécifiques par des approches de banques BAC, les approches d'analyse du nombre de copie par hybridation génomique comparée (aCGH) ont révélé des milliers de régions présentes/absentes (« Presence/Absence Variants », ou PAV) et de régions à nombre de copies variables (« Copy Number Variant », ou CNV) entre deux lignées de maïs (Springer *et al.*, 2009; Beló *et al.*, 2010). L'étude de la variation structurale par une analyse globale de la profondeur de lecture dans plus de 100 lignées de maïs a montré que plus de 90% du génome du maïs présente un certain degré de CNV entre les lignées (Chia *et al.*, 2012). L'ensemble de ces études (et d'autres chez d'autres espèces) a révélé qu'un seul génome de référence ne peut pas représenter correctement l'ensemble des séquences fonctionnelles d'une espèce donnée. Afin de mieux décrire la diversité génétique caractéristique de chaque espèce, le séquençage de génomes de plusieurs individus d'une même espèce a ainsi suscité un intérêt croissant (Miga and Wang, 2021).

Si ces séquences 'accessoires' ont d'abord été considérées comme non indispensables à la biologie de l'espèce, la présence de certains gènes dans des individus particuliers pourrait également sous-tendre l'adaptation locale des populations. De même, les éléments transposables étant capables de modifier la régulation transcriptionnelle des gènes proches, certaines insertions d'éléments transposables du génome 'accessoire' pourraient être adaptatives.

Mes travaux se placent dans cette thématique, et se sont articulés autour de plusieurs questions : quelle est la variabilité des éléments transposables dans les régions non géniques, et diffère-t-elle de celle des régions géniques ? Peut-on caractériser des gènes spécifiques des lignées de maïs Européens impliqués dans leur adaptation locale ? Quel est l'impact de la présence/absence de régions génomiques sur la recombinaison ? Quelles sont les fonctions sous-tendues par les gènes 'accessoires' ?

## 3.2 Dynamique intra-spécifique de régions ciblées

Poste actuel (2008-2014)

*Encadrement : Sivasangari Nandy (M2 2012, 50%), Emeline Baptiste (M2, 2012, 100%), Genseric Beydon (M2 2014, 50%).*

Lors des travaux pionniers sur la variabilité structurale entre lignées de maïs, les auteurs ont proposé que celle-ci est sous-tendue par deux forces majeures : l'insertion différentielle des éléments transposables, et la recombinaison (Wang and Dooner, 2006). Ces premières analyses ayant été réalisées dans des régions riches en gènes et donc présentant un fort taux de recombinaison, j'ai voulu tester le rôle de la recombinaison dans la variabilité observée en comparant la structure de régions riches en gènes et de régions non géniques, incluant des régions péricentromériques. Mon hypothèse était que nous devrions observer une différence de structure haplotypique entre ces régions, du fait de la plus forte recombinaison dans les régions géniques. D'autre part, les séquences de maïs qui avaient été analysées jusqu'à présent provenaient de lignées d'origine Américaine ou tropicale. J'ai souhaité que mon travail, en plus de la portée scientifique qu'il devrait avoir, contribue à l'essor de la génomique du maïs Européen. J'ai donc travaillé sur 5 lignées de maïs, que j'ai choisies pour représenter la diversité du maïs et incluant du matériel Européen. Le projet reposait d'une part sur le séquençage et l'analyse structurale de ces régions sur ces 5 lignées, et d'autre part sur l'analyse de la présence/absence des insertions ainsi découvertes par PCR sur un panel de 200 lignées, permettant une validation plus large des observations. En collaboration avec le CNRGV, nous avons construit des banques BAC pour ces lignées, et ces banques ont été criblées avec des sondes provenant des régions d'intérêt, notamment développés grâce aux marqueurs RJM. Les BAC ainsi repérés ont ensuite été séquencés et assemblés. Cela a demandé de gros développements, en particulier pour la détection des clones non géniques, et pour l'assemblage des séquences issues des régions clonées, qui ont d'abord été réalisés avec des lectures paires Illumina, puis 454 et enfin PacBio, qui ont finalement permis leur assemblage complet. Après plusieurs années de travail, nous avons obtenu des résultats préliminaires intéressants qui allaient dans le sens de mon hypothèse. Malheureusement, le temps de développement du projet a été très long, et au moment où nous commençons à obtenir ces résultats, l'approche était devenue obsolète, et les résultats attendus déjà publiés à un niveau plus global (Springer *et al.*, 2009). J'ai donc décidé de réorienter le projet vers des approches permettant une analyse plus globale de la variabilité du génome (voir 3.3).

Ce travail a été partiellement valorisé par l'utilisation des séquences des BAC de la lignée F2 pour valider la qualité de l'assemblage du génome de cette lignée, en particulier dans des régions non géniques comportant essentiellement des empilements d'éléments transposables (voir 3.3). S'il n'a malheureusement pas donné lieu à la production d'une publication *per se*, ce projet a été le point de départ de la génomique du maïs en France. Il a en particulier été clé dans le développement de projets de grande ampleur, notamment Amaizing et SeqOccin, qui nous permettent aujourd'hui de disposer de génomes assemblés pour 29 lignées de maïs d'intérêt pour l'Europe (voir parties 5.1 et 5.2). Il m'a également permis de renforcer mes connaissances sur la diversité des ressources génétiques du maïs et l'histoire de cette espèce, qui ont été clés dans les projets que j'ai développés par la suite. Enfin, il m'a permis d'encadrer ou de co-encadrer trois étudiants en master.

### 3.3 Analyses de régions spécifiques d'un génotype

Poste actuel (2011-2018)

*Financement acquis : ANR Génomique « CNV\_Maize » (2011-2013, partenaire), Projet investissement d'avenir « Amaizing » (2011-2021, coordinatrice du WP3)*

*Encadrement : Kevin Baudry (Post-doctorat, 2020-2021, 20%)*

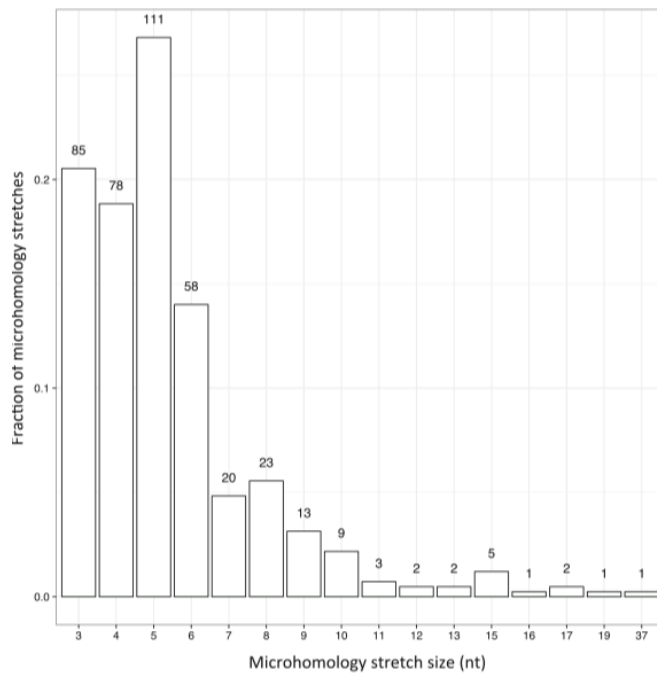
Dans le cadre d'un projet collaboratif, Johann Joets (alors dans l'équipe ABI) et moi avons mis en place un projet qui visait à analyser le contenu de régions présentes dans une lignée Française (F2) et absente de la lignée Américaine de référence (B73). Nous cherchions notamment à estimer si ces régions contenaient des gènes, et si tout ou partie de ces gènes avaient pu jouer un rôle dans l'adaptation du maïs en Europe. A cette période, assembler un génome complet de la taille de celui du maïs avec des lectures pairées de 100 bases était un vrai défi. Afin de répondre à notre question, nous avons généré un ensemble de lectures pairées correspondant à 90 fois la taille du génome, soit 216 Gb de séquence (et plus de 1600 génomes d'*Arabidopsis* !), que nous avons ensuite utilisées pour assembler les parties spécifiques au génome de F2.

Afin de limiter la quantité de génome à assembler *de novo* (étape qui la plus gourmande en temps de calcul), nous avons d'abord essayé d'assembler les régions du génome de F2 absentes du génome de B73 par une approche soustractive : les lectures de la lignée F2 étaient alignées sur le génome assemblé de B73, et les lectures ne s'alignant pas sur celui-ci étaient ensuite assemblées ensemble. Si cette méthodologie était intéressante sur le principe, sa réalisation s'est avérée en-dessous de ce que nous espérions. En effet, le filtre des séquences éléments transposables (présentes dans les régions spécifiques mais aussi ailleurs dans le génome) induisait un assemblage très morcelé des régions spécifiques, ne permettant pas de découvrir ces régions sur toute la longueur des variants. Cette limitation empêchait également la caractérisation des points de cassure, qui nous intéressaient particulièrement pour prédire les mécanismes potentiellement à l'origine de ces variants.

Après plusieurs essais infructueux avec différents paramètres de stringence d'alignement, nous avons changé de stratégie et nous nous sommes lancés dans l'assemblage complet du génome de F2. Nous avons généré un assemblage de 1597 Mb (soit 64,8% du génome) avec une N50 de scaffolds de 13,9 kb et une NG50 de 4,3 kb. S'il n'était pas complet, notre assemblage allait donc bien au-delà des séquences géniques, qui représentent seulement 5% du génome. Les séquences assemblées étaient de bonne qualité, avec une qualité d'assemblage des gènes similaire à celle de B73. Par une approche d'homologie, nous avons identifié que plus 60% de cette séquence correspondait à des éléments transposables, similairement à ce qui est observé pour la séquence du génotype B73 faisant alors référence (Schnable *et al.*, 2009). La comparaison de la séquence assemblée avec celle obtenue à partir de BAC géniques et non géniques (voir partie 3.2) a confirmé la qualité de l'agencement et de l'orientation des contigs dans les scaffolds, et a révélé une absence de chimères, indiquant une bonne qualité d'assemblage même dans les régions très riches en éléments transposables et ayant une structure particulièrement complexe.

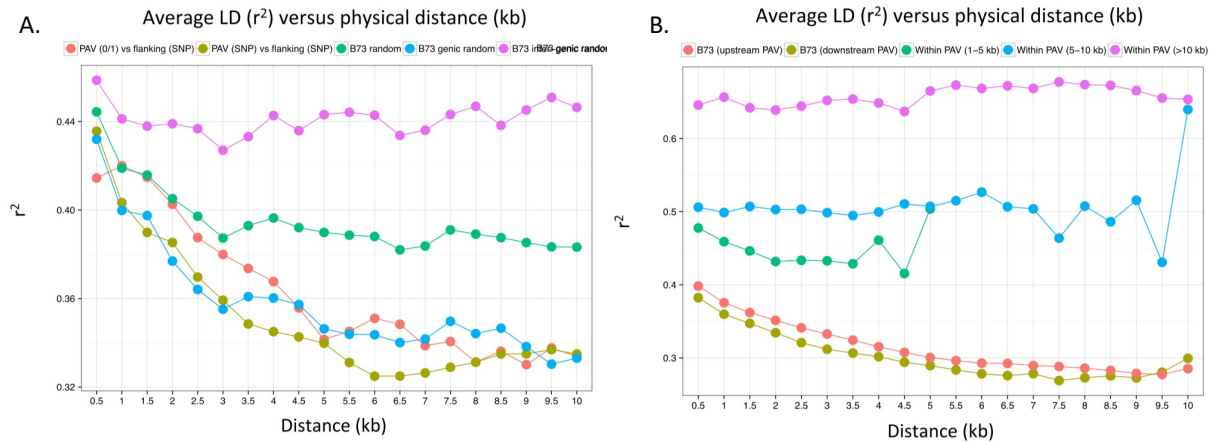
L'alignement des séquences de de B73 sur notre séquence génomique assemblée de F2 nous a permis de mettre en évidence 10 044 régions de plus de 1 kb, spécifiques de F2 et absentes de B73 ('Presence/Absence Variants', ou PAV), représentant un total de 87,4 Mb. Parmi celles-ci, 1028 régions ont pu être ancrées sans ambiguïté au génome B73, et 1960 ont été ancrées de manière ambiguë (plusieurs positions possibles sur la séquence B73). L'utilisation de la même procédure en miroir sur B73 a permis de détecter 691 régions spécifiques de B73.

L'analyse des séquences spécifiques de F2 aux points de cassure a révélé des traces de microhomologie exacte pour 414 régions ancrées (40%), de taille inférieure à 25 bp et montrant une distribution décroissante en fonction de la taille (Figure 7). Cette observation, caractéristique des mécanismes MMEJ (Microhomology-Mediated End Joining) impliqués dans la réparation des cassures double brin (DSBR), suggère qu'une grande partie des variants observés correspondent à des délétions générées par ce mécanisme.



**Figure 7. Distribution de la taille des segments de microhomologie dans les PAV (séquences présentes dans F2 et absentes de B73)**

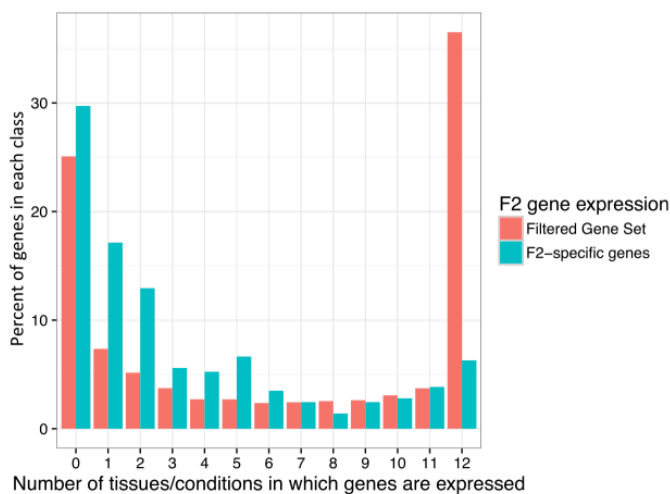
Enfin, l'alignement des lectures Illumina pairée issues d'un ensemble de 23 lignées de maïs tempérés (collaboration Stéphane Nicolas) a révélé que les régions « spécifiques » de la lignée F2 sont également présentes dans d'autres groupes génétiques, avec un nombre de variants communs à F2 augmentant avec la proximité génétique des lignées. Ces deux observations (mécanisme de MMEJ et régions anciennes) suggère qu'une grande partie des régions variables proviennent de délétions. Afin d'analyser l'impact de la présence/absence sur la recombinaison locale, nous avons ensuite analysé les profils de déséquilibre de liaison. L'analyse du lien entre les régions PAV et leurs régions flanquantes ont révélé des profils typiques des régions géniques (Figure 8A). La comparaison du déséquilibre de liaison à l'intérieur des régions PAV et à l'intérieur de leur régions flanquantes indique un déséquilibre plus fort pour les PAV, (Figure 8B) en accord avec l'hypothèse selon laquelle les PAV, du fait de leur absence chez certains génotypes, sont moins souvent impliqués dans des événements de recombinaison que les régions génomiques invariables.



**Figure 8. Analyse du déséquilibre de liaison dans les PAV**

A. Déséquilibre de liaison moyen entre les PAV et leurs régions génomiques flanquantes. Les régions flanquantes sont génotypées à l'aide de SNP, les PAV sont génotypés soit comme 0/1 (rouge), soit à l'aide du SNP intra-PAV le plus proche du point de cassure (kaki). Pour comparaison, le déséquilibre de liaison dans des régions aléatoires issues : du génome complet (vert), de l'espace génique (bleu) et de l'espace intergénique (violet) est également représenté. B. Comparaison du déséquilibre de liaison moyen dans les PAV avec celui de leurs régions flanquantes (en amont, rouge, et en aval, kaki). Les PAV sont séparés en 3 classes de taille, 1-4 kb (vert), 5-10 kb (bleu), 10-40 kb (violet). Le déséquilibre de liaison est estimé par les corrélations au carré de la fréquence des allèles ( $r^2$ ) et tracé en fonction de la distance entre les sites polymorphes (0 à 10 kb).

D'autre part, l'analyse de données mRNA-seq issues de 12 tissus différents de la lignée F2 (produites par Peter Rogowky, RDP, Lyon) nous a permis d'annoter un total de 395 nouveaux gènes dans les régions spécifiques de F2. Leur annotation fonctionnelle a indiqué une sur-représentation en gènes impliqués dans la réponse au stress, les processus biosynthétiques, le développement et le remodelage de la chromatine. Enfin, l'analyse comparée de l'expression de ces gènes PAV et non PAV à travers les tissus a révélé que les gènes communs aux deux lignées sont exprimés dans l'ensemble des tissus, alors que les gènes spécifiques de F2 sont plus souvent tissu-spécifiques (Figure 9).



**Figure 9. Profil d'expression des gènes spécifiques à F2**

'Filtered gene set' : gènes annotés dans le génome de B73 et présents également dans le génome de la lignée F2

Ces résultats ont fait l'objet de la publication suivante, dans laquelle je suis co-première autrice :

Darracq\*, **C. Vitte\***, S. Nicolas, J. Duarte, J.P. Pichon, T. Mary-Huard, C. Chevalier, A. Berard, Le M.C. Paslier, P. Rogowsky, A. Charcosset, J. Joets (2018). Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. *BMC Genomics*, 19 (1) 119

Une méthodologie similaire a été appliquée à deux autres génotypes de maïs tempérés (Ph207 et C103) et représentant, avec F2, trois groupes génétiques distincts de celui de B73. Nous avons ainsi identifié un ensemble de 141 000 régions d'insertion/délétion et de PAV. Ces régions nous ont ensuite servi à construire une puce Axiom® Affymetrix®, représentant ces régions par une combinaison de sondes sélectionnées sur les points de cassure (13%) ou à l'intérieur des régions – sur des sites polymorphes (25%) ou non polymorphes (63%). Cette puce contient un ensemble de 662 772 sondes et cible 105 927 régions, incluant des PAV de taille entre 35 bp et 129kbp. Après contrôles qualité et différents filtres, nous avons ainsi réussi à génotyper 86 648 régions (82%) sur un ensemble de 445 échantillons d'ADN de maïs avec une reproductibilité élevée (> 95%). La matrice de parenté entre les 362 lignées de maïs obtenue à partir des résultats de cette puce s'est montrée fortement corrélée à celle estimée à partir de SNP provenant de la puce Illumina 50 K SNP.

Ce travail a permis la publication suivante, dans laquelle j'ai contribué à la réflexion sur le choix des lignées, au choix des méthodologies de séquençage et de détection des variants, à la génération des régions de la lignée F2, et à la rédaction de l'article :

C. Mabire, J. Duarte, A. Darracq, A. Pirani, H. Rimbart, D. Madur, V. Combes, **C. Vitte**, S. Praud, N. Riviere, J. Joets, J.P. Pichon, S. Nicolas (2019). High throughput genotyping of structural variations in a complex plant genome using an original Affymetrix® Axiom® array. *BMC Genomics*, 1 (20) 848

Enfin, l'ensemble de ces travaux nous ont permis d'acquérir une reconnaissance internationale en génomique du maïs, qui a abouti à notre invitation à écrire deux chapitres d'ouvrage :

J. Joets, C. Vitte, S. Pasternak, D. Ware (2011). Deciphering the multi-faceted aspects of the maize genome: new insights from sequencing projects. In J.-L. Prioul, *Advances in Maize* (Taylor & Francis), the Society for Experimental Biology, London, UK. pp. 111-120.

J. Joets, **C. Vitte**, A. Charcosset (2019). Draft assembly of the F2 European maize genome sequence and its comparison to the B73 genome sequence: a characterization of genotype-specific regions. In *The maize Genome*, editors Jeffrey Bennetzen, Professor. Sherry Flint-Garcia, Professor. Candice Hirsch and Dr. Roberto Tuberosa. Springer Nature, p.3-12

Suite à ces travaux, il m'est apparu indispensable générer une ressource de référence pour le maïs Européen, rassemblant, pour un certain nombre de génotypes d'intérêt, la séquence des génomes ainsi que leur annotation, les variations structurales, les mesures d'expression des gènes dans différents tissus et conditions, et certaines variables environnementales et phénotypiques de la culture. Pour obtenir une ressource de haute qualité et minimiser les biais

expérimentaux et technologiques, notamment entre les génotypes, l'ensemble des données devaient être produites au sein d'un dispositif expérimental unique, avec les mêmes technologies et les mêmes traitements bio-informatique et statistique. Une telle ressource, réutilisable sur le long terme, permettrait d'appréhender de nombreuses questions fondamentales et finalisées, tout en pouvant être incrémentée dans le temps. Ceci a pu être réalisé grâce au financement du programme investissement d'avenir Amaizing (2011-2021).

Nous avons ainsi entrepris la construction d'un pangénome à partir de plusieurs lignées de maïs d'origines contrastées. Le but était de caractériser de nouveaux gènes qui pourraient avoir participé à l'adaptation du maïs aux conditions Européennes, et d'analyser leurs fonctions. Ce projet, réalisé en partenariat avec des entreprises privées de semenciers, avait à la fois un but de recherche fondamentale – mieux comprendre l'étendue de la variabilité génomique du maïs, l'origine moléculaire de cette variabilité, les fonctions sous-tendues par les gènes 'accessoires', et déterminer si ces régions ont contribué à l'évolution fonctionnelle du maïs – et de recherche appliquée – trouver de nouveaux gènes d'intérêt agronomique dans des lignées utiles aux sélectionneurs français.

Je souhaitais également que ces ressources génomiques puissent servir à d'autres projets en cours au Moulon et à l'ensemble de la communauté maïs, et ne soient pas redondants avec des projets similaires mis en place dans la même période. Les discussions avec les différents acteurs nationaux et internationaux de la communauté, que j'ai coordonnées, ont mené au choix de 5 lignées (Tableau 2).

**Tableau 2. Présentation des lignées choisies**

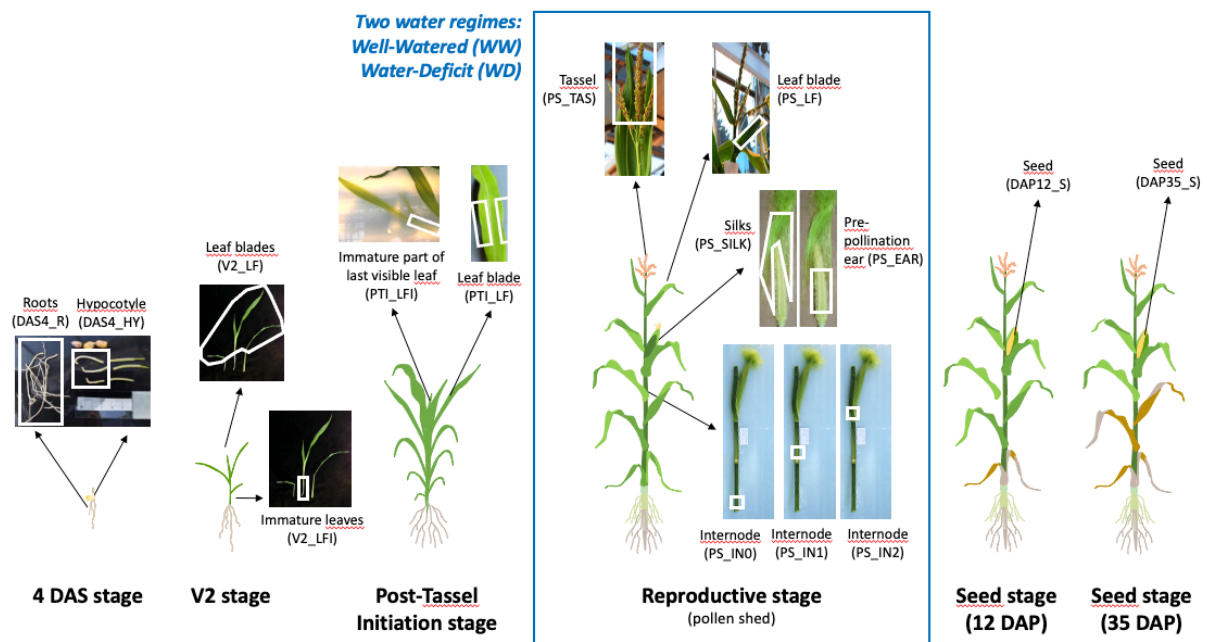
Génotype	Groupe	Sous-groupe	Intérêt biologique
F2	Corné	Denté Européen	Adaptation aux conditions Européennes, précocité de floraison
F4	Corné	Denté du Nord Européen	Adaptation aux conditions Européennes, digestibilité
GF111	Corné	Denté du Nord Américain	Précocité de floraison
MBS847	Denté	Iodent	Sélection divergente pour la précocité de foraison
F252	Denté	Lancaster/Iodent	Sélection divergente pour la précocité de foraison, précocité de floraison
F331	Tropical	Tropical d'altitude	Tolérance au froid
EA1197	Tropical	Tropical Espagnol	Adaptation aux conditions Européennes

Au début du projet, aucune méthodologie permettant d'assembler entièrement la séquence d'un génome de grande taille comme celui du maïs n'était stabilisée, et les coûts des différentes technologies, amplifiés par la taille du génome du maïs, étaient très élevés, sans assurance du résultat. De plus, les technologies de séquençage étaient en plein essor et les compagnies qui les développaient en forte compétitivité, induisant une évolution très rapide des technologies. Dans ce contexte, nous avons choisi de tester différentes stratégies sur la lignée F2, et d'appliquer ensuite la plus prometteuse à l'ensemble des autres lignées.

Nous avons notamment travaillé avec le Génoscope sur l'utilisation de séquences 454 'mate-pair' et 'paired-end', mais la technologie s'est arrêtée avant que nous ayons pu séquencer l'ensemble des banques nécessaires à l'assemblage du génome de F2. En 2015, nous avons entrepris, en collaboration avec le CNRGV de Toulouse, la construction d'une banque BAC profonde (représentant 15 fois le génome) de la lignée F2. Finalement, la compagnie NRGENE a développé une méthodologie permettant d'assembler un génome de maïs à partir de séquences Illumina, en s'appuyant partiellement sur le génome déjà assemblé de B73. Cette méthodologie avait été fructueuse pour l'assemblage de deux génomes de maïs Européens

par nos collègues allemands dans le cadre du programme SYNBREED (<https://webarchiv.it.lis.tum.de/synbreed.tum.de/home/index.html>) et coûtait relativement « peu » cher comparée aux autres. Nous avons donc décidé de l'utiliser pour construire notre pangénome de maïs.

Le but étant de comparer la présence/absence de gènes, il fallait également mettre en place une stratégie pour annoter les gènes, en utilisant des méthodes de prédiction '*ab initio*' et fondées sur la présence de transcrits. J'ai donc mis en place la construction du 'Gene Atlas Amazing', visant à séquencer en triplicat les ARN issus de 15 tissus, dont 7 en deux conditions hydriques, pour l'ensemble des 7 lignées et la lignée de référence B73 (Figure 10).



**Figure 10.** Description des échantillons analysés en transcriptomique pour 7 génotypes de maïs et le génotype de référence B73

Chaque type d'échantillon est représenté par 3 réplicats biologiques (plantes différentes). L'ensemble des échantillons est prélevé sur 8 génotypes différents, représentant un total de plus de 500 échantillons.

J'ai ainsi coordonné une grande aventure collaborative, allant de la culture des plantes à la collecte des échantillons à des stades particuliers et sur des lignées non synchrones en développement, aux extractions d'ARN et leur séquençage. L'ensemble étant réalisé sur trois localisations différentes, par plusieurs expérimentateurs différents venant de laboratoires différents, j'ai particulièrement veillé à limiter ces biais afin de générer un jeu de données de haute qualité. Au cours de ces différentes étapes, j'ai également porté un effort particulier à la traçabilité des lots de graines utilisés et à leur validation génotypique par marqueurs microsatellite, afin de garantir la cohérence entre les lots de graines utilisés pour les assemblages de génome et les analyses de transcriptomique.

Les séquences génomiques ont été assemblées en 2019, et l'ensemble des ARN extraits en 2016. Malheureusement un problème technique sur la plateforme de séquençage des ARN a induit la production de données mRNA-seq ayant des taux de GC inattendus, nécessitant de reséquencer la quasi-totalité des échantillons. L'ensemble des données a finalement été re-

généralisé en 2020, mais à un moment où le programme Amaizing se terminait et ne permettait plus de recrutement de personnel. Du fait de ces complications, ce projet a pris beaucoup de retard, et nous finissons cette année les analyses permettant sa valorisation. Les analyses ont essentiellement été réalisées par Johann Joets, et le manuscrit est en cours d'écriture.

Les données transcriptomiques relatives à B73, qui ont été produites assez tôt, ont été valorisées dans un article analysant les réseaux de gènes impliqués dans la tissu-spécificité, que je présenterai plus en détail dans la partie 4.5 :

M. Fagny, M. L. Kuijjer, M. Stam, J. Joets, O. Turc, J. Roziere, S. Pateyron, A. Venon, C. Vitte§ (2021). Identification of key tissue-specific, biological processes by integrating enhancer information in maize gene regulatory networks. *Front. Genet.*, <https://doi.org/10.3389/fgene.2020.606285>

Ayant fait le choix d'inclure des lignées travaillées par d'autres collègues du Moulon, en particulier celles ayant servi de point de départ à une analyse de sélection divergente de la précocité de floraison (voir Tableau 2), une partie des séquences Illumina générées pour réaliser les assemblages *de novo* des géotypes F252 et MBS847 ont permis l'analyse de l'hétérozygotie au niveau de gènes impliqués dans les variations d'expression de la sélection divergente, et ont participé à la publication :

A. Desbiez-Piat, A. Ressayre, E. Marchadier, A. Noly, C. Remoué, C. Vitte, H. Belcram, A. Bourgeois, N. Galic, M. Le Guilloux, M.I. Tenaillon, C. Dillmann. Pervasive GxE interactions shape adaptive trajectories and the exploration of the phenotypic space in artificial selection experiments (2023). *Genetics*. Oct 12:iyad186. doi: 10.1093/genetics/iyad186. Epub ahead of print. PMID: 37824828.

## 4- Epigénome : des stress abiotiques à la spécificité tissulaire

### 4.1 Contexte scientifique

Si la variation de la séquence d'ADN est connue comme le principal moteur de l'évolution et de l'adaptation des organismes à leur environnement, plusieurs travaux montrent que les variations épigénétiques, notamment de la méthylation de l'ADN, y contribuent également. Bien qu'elles ne provoquent pas de changement de la séquence d'ADN, des modifications de la méthylation de l'ADN peuvent être associées à des modifications héréditaires, au cours du développement ou entre générations, et peuvent modifier la régulation transcriptionnelle des gènes. Elles peuvent ainsi contribuer à la variation phénotypique héréditaire et constituent une source de variation pour la sélection. Ces épimutations sont souvent observées en lien avec l'insertion d'un élément transposable (Baduel and Colot, 2021), et peuvent générer des phénotypes importants.

L'impact de l'environnement sur l'émergence d'épimutations offre de nouvelles pistes de réflexions en termes d'adaptation, en particulier pour les organismes sessiles qui ne peuvent pas échapper à leur environnement et doivent donc constamment s'adapter à ce dernier. De ce fait, de nombreuses recherches se sont penchées sur la réponse épigénétique des plantes aux contraintes abiotiques, et la transmission des épiallèles ainsi induits. Ces travaux ont révélé des modifications de la méthylation de l'ADN pour différentes combinaisons espèces/organes/contraintes environnementales (Baduel and Colot, 2021; Mladenov *et al.*, 2021). Les changements observés correspondent soit à des augmentations (hyperméthylation) ou à des diminutions (hypométhylation) de la méthylation. Ils sont souvent localisés dans ou à proximité de séquences d'éléments transposables (Hashida *et al.*, 2006; Secco *et al.*, 2015; Wibowo *et al.*, 2016). Enfin, la déméthylation induite par l'environnement peut entraîner la transposition de certains éléments transposables (Hashida *et al.*, 2006).

Certains des changements observés sont maintenus au cours de la vie de l'individu après arrêt de l'exposition à la contrainte (Secco *et al.*, 2015), et certains peuvent être transmis à la génération suivante (Ganguly *et al.*, 2017). Mais la transmission de ces épiallèles induits au-delà d'une génération et sans ré-application de la contrainte est rarement observée (Wibowo *et al.*, 2016; Van Dooren *et al.*, 2020) et ces épiallèles peuvent être perdus après quelques générations (Wibowo *et al.*, 2016). Enfin, la transmission semble augmenter avec la répétition de la contrainte sur plusieurs générations (Wibowo *et al.*, 2016).

En conclusion, le méthylome peut varier sous influence des contraintes abiotiques de l'environnement, et peut affecter différentes régions génomiques. La transmission est variable selon les contraintes appliquées, les organes affectés et les séquences modifiées. Ces épiallèles sont souvent associées à des séquences d'éléments transposables, et l'interaction insertion/méthylation au site d'insertion pourrait permettre à la fois une transmission forte (car facilitée) tout en conférant une flexibilité du signal.

Les séquences *cis*-régulatrices de l'expression des gènes ayant un rôle particulier dans le lien avec le phénotype, leur variabilité génétique ou épigénétique pourrait contribuer de façon importante à l'évolution. Elles contiennent des motifs de fixation de facteurs de transcription, et sont caractérisées par une faible méthylation, une forte accessibilité de la chromatine, et

présentent des modifications particulières sur l'histone H3 en lien avec leur rôle régulateur : acétylation sur les lysines de l'histone 3 pour les éléments activateurs, et en marque H3K27me3 pour certains éléments répresseurs (Oka *et al.*, 2017). En particulier, les éléments *cis*-régulateurs (CRE) distaux jouent un rôle important dans le contrôle de l'expression des gènes au cours du développement, en réponse aux signaux hormonaux et environnementaux (Spitz and Furlong, 2012). Par définition, les CRE distaux sont localisés à plus de 2 kb des gènes. Des études récentes ont montré qu'ils sont répandus et que leur nombre est corrélé à la taille du génome, leur distance au gène le plus proche pouvant atteindre 1 Mb chez les espèces à « gros » génome comme le maïs ou l'orge (Lu *et al.*, 2019). Leur effet régulateur est médié par l'interaction de ce complexe ADN/facteur de transcription avec les promoteurs de gènes cibles *via* des boucles de repliement tridimensionnel.

De façon générale, les régions *cis*-régulatrices sont faiblement méthylées (Crisp *et al.*, 2020; Marand *et al.*, 2023), et leur méthylation peut induire une diminution de l'affinité des facteurs de transcription (Hartwig *et al.*, 2023). Ainsi, des modifications de la méthylation de l'ADN au niveau de ces régions pourraient avoir un impact fort sur la régulation des gènes, et jouer un rôle dans la variation phénotypique.

Dans ce contexte, j'ai développé des travaux sur la variabilité du méthylome, en réponse à l'environnement et entre génotypes, puis plus récemment sur la variabilité des éléments *cis*-régulateurs au cours du développement. Dans tous ces projets, j'ai cherché à estimer la contribution des éléments transposables à ces variations.

Ces travaux ont commencé dans un contexte méthodologique où les premiers épigénomes d'*Arabidopsis*, réalisés en bisulfite-seq, venaient juste d'être publiés (Cokus *et al.*, 2008; Lister *et al.*, 2008). L'extension de telles méthodologies n'avaient pas été réalisées pour des espèces à « gros » génomes comme le maïs. Du fait des problèmes de conversion et de leur impact sur les alignements des séquences qui en découlent, le bisulfite-seq apparaissait comme difficilement applicable pour de telles espèces, et a dans un premier temps été mis de côté pour cette espèce au profit du MeDIP-seq, qui était utilisée chez l'homme. Les laboratoires tâtonnaient pour trouver le meilleur moyen d'obtenir des méthylomes pour ces espèces. Aucune de ces techniques n'était travaillée au Moulon. Dans ce contexte, j'ai construit mes différents projets de façon conjointe avec la mise au point de différentes techniques : séquençage ciblé par PCR après traitement de l'ADN au bisulfite de sodium, séquençage du méthylome complet par MeDIP-seq, et séquençage du méthylome complet par bisulfite-seq.

## 4.2 Analyse de la méthylation du gène *VRNA-1* en réponse au froid par une approche ciblée

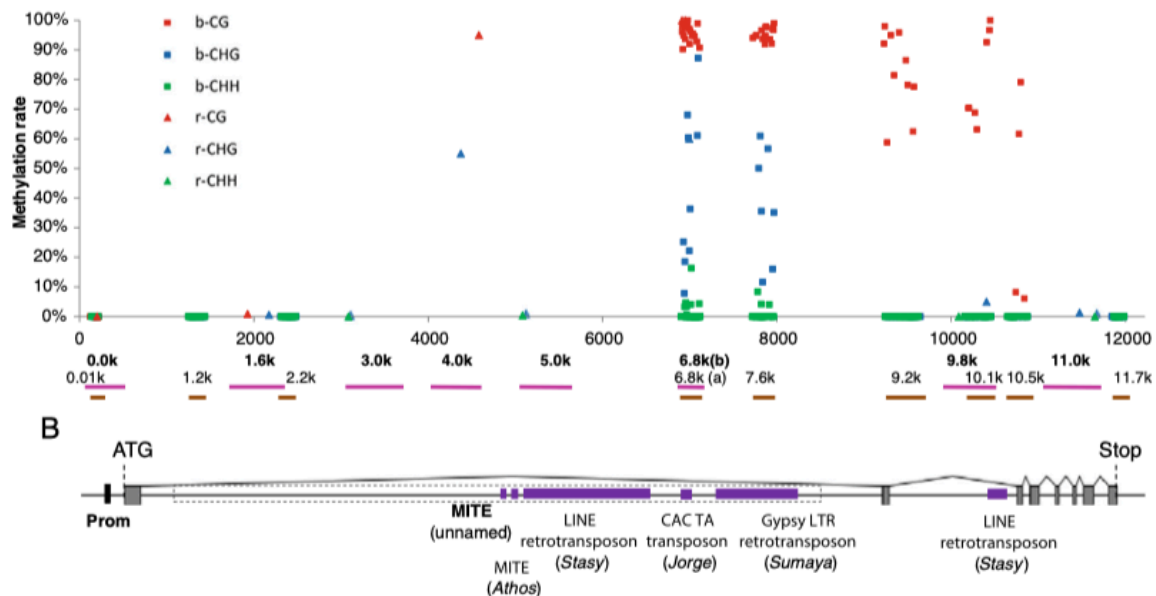
Poste actuel (2009-2013)

*Encadrement : Abdul Rehman-Khan (thèse, 2010-2013, 30%)*

Je souhaitais travailler sur un locus d'intérêt dont la méthylation pourrait être modifiée par l'environnement. L'équipe DEAP s'intéressait à la vernalisation – nécessité de la plante d'être exposée à des températures basses pour passer de la croissance végétative à la floraison – chez le blé. Le gène *VRN1* était connu pour jouer un rôle dans ce processus, et des études antérieures sur le blé et l'orge avaient révélé le rôle fonctionnel de la modification des histones dans l'expression de ce gène (Oliver *et al.*, 2009). J'ai donc développé, en collaboration avec

Jérôme Enjalbert et Isabelle Goldringer de l'équipe DEAP, un projet visant à déterminer si l'expression du gène *VRN-A1* du blé induite par le froid est associée à une modification de la méthylation de l'ADN.

Cette étude a permis de fournir la première analyse de méthylation de l'ADN du gène *VRN-A1*. Nous avons décrit l'existence de méthylation sur les sites CG mais aussi sur les sites non CG (Figure 11). De plus, alors que les sites CG présentent un profil en forme de cloche typique de la méthylation du corps du gène, nous avons montré la présence de méthylation non CG au niveau du premier intron, de taille très élevée (8,5 kb). Nous avons montré d'une part que le froid induit une hyperméthylation spécifique au niveau de ces sites non CG, et d'autre part que cette hyperméthylation est transmise au cours de la mitose et revient à son niveau initial après la reproduction sexuée. Enfin, nous avons montré que cette région hyperméthylée est riche en éléments transposables, suggérant que l'hyperméthylation des éléments transposables dans cet intron pourrait être importante pour le bon fonctionnement de la transcription du gène en condition de froid.



**Figure 11 : Profil de méthylation de l'ADN sur le gène *VRN-A1* de 12 kb de long**

A. Les lignes marron et rose représentent les fragments étudiés à l'aide de techniques fondées sur le bisulfite-PCR et sur les enzymes de restriction, respectivement. Les carrés et les triangles représentent les données du bisulfite et des enzymes de restriction, respectivement. Les couleurs rouge, bleue et verte représentent les contextes CG, CHG et CHH, respectivement. B. Structure du gène *VRN-A1* (fondée sur l'accension Genbank AY747600). Le cadre en pointillé indique la position de la délétion dans l'allèle de Langdon (d'après une comparaison d'alignement avec l'accension Genbank AY747598). Les lignes violettes indiquent les positions des séquences d'éléments transposables. La catégorie et le nom de la famille sont indiqués en-dessous.

Ce travail a été réalisé par Abdul Rehman Khan, alors en thèse dans l'équipe DEAP, et m'a permis ma première expérience de co-encadrement de doctorant. Il a été publié dans l'article suivant, dans lequel Abdul et premier auteur et moi dernière autrice correspondante.

A. R. Khan, J. Enjalbert, A.-C. Marsollier, A. Rousselet, I. Goldringer, C. Vitte<sup>§</sup> (2013). Vernalization treatment induces DNA site-specific hypermethylation at the *VERNALIZATION-A1* (*VRN-A1*) locus in hexaploid winter wheat. *BMC Plant Biology*. 13:209.

### 4.3 Analyse de la variation de méthylation de l'ADN entre lignées de maïs à l'échelle du génome par l'approche MeDIP-seq

Poste actuel (2009-2013)

*Financement acquis : Projet innovant INRAe, projet investissement d'avenir Amaizing (2011-2021)*

*Encadrement : Lilia Moulfi (L3 pro, 2010, 100%), Sara Castelletti (2012-2014, 100%), Iris Grosjean (M1 2014, 50%)*

En parallèle, j'ai mis en place un projet visant à caractériser des régions différentiellement méthylées entre lignées de maïs issues de groupes génétiques différents, et leur relation avec la transcription des gènes adjacents. Pour cela, j'ai mis en place une expérimentation dans les serres de l'IBP (depuis devenu IPS2), où j'ai cultivé 5 génotypes de maïs en 5 réplicats, et pour lesquels j'ai prélevé le limbe de la partie médiane de la feuille de maïs, sur la dernière feuille ligulée, afin de limiter le biais lié aux différences de croissance entre les génotypes. J'ai encadré Lilia Moulfi, étudiante en L3 pro, que j'ai formée à culture des plantes et à la réalisation de PCR quantitative.

Pour la mise en place du MeDIP-seq, ne disposant pas au Moulon du matériel nécessaire pour construire des banques Illumina, j'ai réalisé une collaboration avec la plateforme IMAGIF, qui venait de se construire sur le campus de Gif-sur-Yvette. En parallèle, j'ai mis en place l'étape d'immuno-précipitation au Moulon. Nous avons ainsi développé un protocole permettant l'analyse MeDIP-seq chez le maïs. Sara Castelletti (post-doctorat 2012-2014) a ensuite comparé les profils de méthylation entre lignées de maïs, et analysé le lien entre méthylation et expression des gènes adjacents. Malheureusement, les résultats de MeDIP se sont avérés très complexes à analyser et ont nécessité une profondeur de séquençage beaucoup plus importante qu'anticipée. En effet, du fait de la forte variabilité structurale existant entre lignées de maïs, un effort important de séquençage de l'« input » (correspondant à tout le génome) était nécessaire pour réduire le nombre de faux-positifs et obtenir des pics fiables de méthylation. Le coût devenant presque similaire à celui du bisulfite-seq, pour une détection moins précise et ne donnant pas accès au niveau de méthylation de chaque contexte de cytosine, nous avons, ainsi que d'autres laboratoires travaillant sur cette thématique chez le maïs, abandonné le MeDIP-seq au profit du bisulfite-seq. Les premiers résultats préliminaires nous ont cependant donné des pistes de réflexion sur le lien entre variation génétique et épigénétique entre les lignées, qui sous-tendent mes projets actuels.

### 4.4 Impact de l'environnement sur la méthylation du génome du maïs

Poste actuel (2014-)

*Financement acquis : Projet investissement d'avenir Amaizing (2011-2021)*

*Encadrement : Zeineb Achour (thèse, 2014-2018, 100%), Jonathan Torrel-Salat (CDD IE, 2018-2019, 100%)*

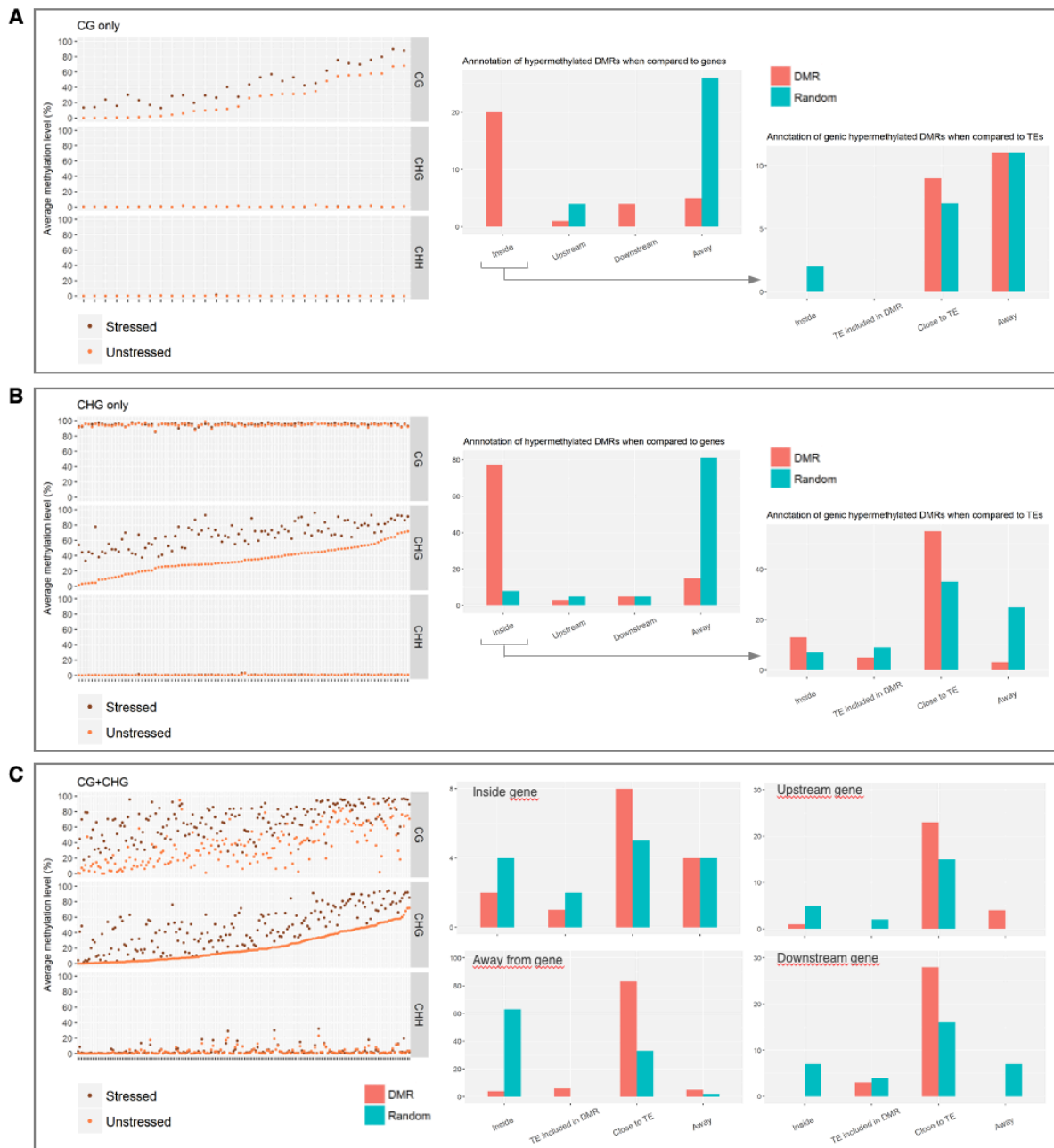
La caractérisation des processus moléculaires qui sous-tendent la réponse des plantes aux contraintes environnementales est un enjeu majeur pour mieux comprendre l'adaptation locale des populations et pour générer des génotypes plus tolérants à ces conditions. Dans un contexte de changements climatiques, le décalage du cycle de vie du maïs par un semis précoce permettrait de maximiser l'accès aux ressources de lumière et d'eau disponibles au printemps. L'adaptation du maïs au froid précoce pourrait ainsi être un moyen de contourner la perte de rendement liée aux risques accrus de sécheresse au moment de la floraison.

Dans ce contexte, j'ai cherché à caractériser l'impact du froid modéré en début de cycle sur la méthylation de l'ADN chez le maïs, et la transmission des modifications induites à la descendance. En collaboration avec Catherine Giauffret (UMR AgroImpact) et l'UE INRAe d'Estrées-Mons, j'ai mis en place un dispositif expérimental en conditions contrôlées dans laquelle un lot de plantes a subi un stress froid prolongé durant les 22 premiers jours de son développement (16h jour à 14°C ; 8h nuit à 10°C), tandis que lot témoin a été cultivé en conditions standard (16h jour à 24°C ; 8h nuit à 18°C). Plusieurs semis décalés des plantes contrôle ont permis d'homogénéiser le stade de développement entre les deux lots.

La génération de données bisulfite-seq (collaboration Jean-Philippe Pichon et Magalie Leveugle de Biogemma) et mRNA-seq issues du limbe foliaire sans nervure en triplicat et leur analyse nous a permis de montrer une augmentation globale de méthylation du génome dans le lot 'froid' par rapport au lot 'contrôle'. Nous avons observé que cette hyperméthylation se produit essentiellement dans les séquences d'éléments transposables, sur les contextes classiquement méthylés dans les différentes régions génomiques : CG et CHG dans les régions distales des gènes, et CHH dans les régions proximales, suggérant que la méthylation des éléments transposables est renforcée en conditions de froid. D'autre part, l'analyse des régions différentiellement méthylées par une approche de scan du génome sur des fenêtres de 200 pb à 1 kb nous a permis de révéler un total de 699 régions différentiellement méthylées (DMR), incluant 377 régions hypométhylées et 322 régions hyperméthylées. La catégorisation des DMR en fonction des contextes a mis en évidence 363 cas modifiés à la fois en CG et CHG localisés essentiellement dans les éléments transposables, 220 cas présentant une modification de CHG seulement, associées à des éléments transposables géniques, et 48 cas présentant une variation uniquement en CG localisées dans les gènes, et parfois associées aux éléments transposables (Figure 12). Nous avons également caractérisé 418 gènes présentant une région différentiellement méthylée à moins de 2kb. Parmi ceux-ci, 28 gènes sont associés à la réponse au stress, dont 6 montrent aussi un changement de niveau de transcription.

Ce travail a été réalisé par Zeineb Achour, doctorante sous ma direction. Les résultats lui ont permis de rédiger un manuscrit, que nous avons soumis à une revue de rang A, et mis à disposition de la communauté sous la forme d'une pré-publication dans Biorxiv :

Z. Achour, J. Joets, M. Leguilloux, H. Sellier, J.-P. Pichon, M. Leveugle, H. Duborjal, J. Caius, V. Brunaud, C. Paysant-Le Roux, T. Mary-Huard, C. Giauffret, C. Vitte (2019). Low temperature triggers genome-wide hypermethylation of transposable elements and centromeres in maize. Biorxiv <https://doi.org/10.1101/573915>.



### Figure 12. Caractéristiques des DMR hyperméthylées

A. Caractéristiques des DMR de type 'CG only'. A gauche : niveaux de méthylation dans les plantes stressées et non stressées. Chaque position de l'axe des x correspond à une DMR, à laquelle correspondent deux points : les valeurs moyennes dans les plantes stressées (marron) et non stressées (orange). Au milieu : position relative des gènes par rapport à des régions aléatoires. Pour la catégorie la plus enrichie (DMRs situées à l'intérieur des gènes), la position par rapport aux éléments transposables est indiquée sur le panneau de droite, par rapport à un ensemble aléatoire de régions situées à l'intérieur des gènes. B. Caractéristiques des DMR de type 'CHG only'. Même légende qu'en A. C. Caractéristiques des DMR de type 'CG+CHG'. Gauche : même légende que les panneaux A et B. Quatre panneaux à droite : Position des DMR par rapport aux éléments transposables et comparaison aux régions aléatoires, après subdivision des DMR par rapport à la distance aux gènes. Les seuils de distance sont de 2 kb pour les gènes et 500 bp pour les éléments transposables.

Nous avons choisi de travailler sur un stress modéré et long pour représenter au plus près les conditions subies par le maïs en champs. D'un point de vue académique, nous souhaitons également tester si le stade de développement au moment de l'application du stress avait un impact sur l'état de méthylation. De plus, comme indiqué plus haut, nous souhaitons estimer le degré de transmission des modifications induites.

Dans une deuxième partie, nous avons donc suivi le niveau de méthylation des 699 DMR détectées précédemment sur différents rangs de feuille, et dans la descendance. Pour cela, nous avons réalisé des autofécondations des plantes stressées, d'une part, et des croisements entre plantes stressées et contrôles d'autres part. L'analyse du niveau de méthylation de ces régions dans les différents échantillons par la technique de SeqCapEpi (collaboration Hervé Duborjal et Magalie Leveugle de Biogemma) a permis à Jonathan Torrel-Salat (CDD ingénieur post-master) de mettre en évidence une grande stabilité des signaux de méthylation entre les différents rangs feuilles à la fois sur les plantes stressées et non stressées, et entre générations.

En parallèle, le retour des 'reviewers' sur le manuscrit de Zeineb a suggéré que les variants observés entre les deux lots pourraient, au moins en partie, ne pas être liés à la différence de traitement des plantes, mais à de la variabilité entre les individus. Cette remarque m'a permis de réinterpréter l'ensemble des résultats sous l'angle de l'interaction entre variabilité épigénétique et environnement. Mais le temps ayant passé entre le moment de la soumission, le retour des 'reviewers' et la reprise des données, la version du génome que nous avons utilisée était devenue obsolète, notamment pour ce qui concerne la qualité de l'assemblage des régions riches en éléments transposables. Je suis donc en train de reprendre les analyses sur la version du génome la plus récente afin de publier ce travail. J'envisage une publication qui agrège les deux parties de l'étude.

#### 4.5 Eléments *cis*-régulateurs impliqués dans la tissu-spécificité chez le maïs : contribution des éléments transposables

Poste actuel (2018-)

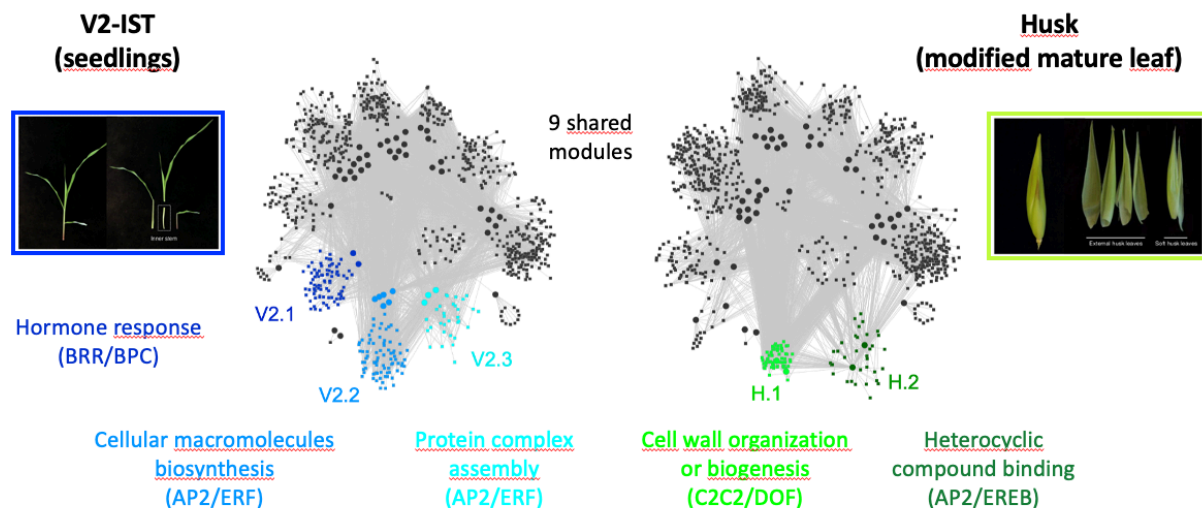
*Financement acquis : Projet investissement d'avenir Amaizing (2011-2021)*

*Encadrement : Maud Fagny (post-doctorat, 2019-2020, 100%)*

Si les données d'épigénomiques, notamment d'ouverture de chromatine, permettent de révéler les CRE distaux dans les génomes, la caractérisation de leur gène cible reste complexe à établir, notamment pour les espèces à gros génome. Chez le maïs, la caractérisation de ces connexions par l'analyse du repliement de la chromatine en 3D a révélé que plus d'un quart des CRE distaux de cette espèce ne ciblent pas leur gène flanquant immédiat et que 34% d'entre eux régulent probablement plusieurs gènes (Lu *et al.*, 2019). Si ces techniques permettent de mettre en évidence les connexions entre CRE et gène-cible, elles sont couteuses et restent limitées si on veut analyser un grand nombre d'individus et de conditions ou tissus. En proposant des approches pour modéliser *in silico* les relations de régulation entre des composants hétérogènes tels que les régions *cis*-régulatrices, les facteurs de transcription et les gènes, la biologie des systèmes offre une alternative puissante et efficace. Ces approches de biologie des systèmes ont contribué à décrypter le rôle des facteurs de transcription se liant au promoteur dans la régulation de leurs gènes cibles, dans différents tissus ou au cours du cycle cellulaire chez les champignons et les animaux (Sonawane *et al.*, 2017; Kuijjer *et al.*,

2019), et à identifier l'impact des maladies sur le câblage des réseaux de régulation tissu-spécifiques chez l'homme (Padi and Quackenbush, 2018).

En combinant les informations de position des régions 'enhancers' caractérisés précédemment pour deux tissus de maïs, la jeune feuille et la spathe (bractée qui entoure l'épi) sur la lignée de référence B73 (Oka *et al.*, 2017), l'annotation des sites de fixation au facteurs de transcription qu'ils présentent, et des données de transcriptomiques générées sur 15 tissus (voir partie 3.3), Maud Fagny (Post-doc 2019-2020) a fait la preuve de concept de ces approches de biologie des systèmes chez le maïs. Nous avons ainsi caractérisé les liens entre CRE et gènes cibles pour la jeune feuille et la spathe, et mis en évidence les réseaux de régulation spécifiques à chacun de ces deux tissus. Nous avons mis en évidence que les réseaux des feuilles au stade jeune (V2-IST) sont caractérisés par la réponse aux hormones et la biogenèse et l'assemblage des macromolécules, qui sont régulés par les familles de facteur de transcription de type *BRR/BPC* et *AP2/ERF*, respectivement (Figure 13). De façon contrastée, les spathe sont caractérisées par la modification de la paroi cellulaire et la réponse aux stress abiotiques, respectivement orchestrées par les familles *C2C2/DOF* et *AP2/EREB*.

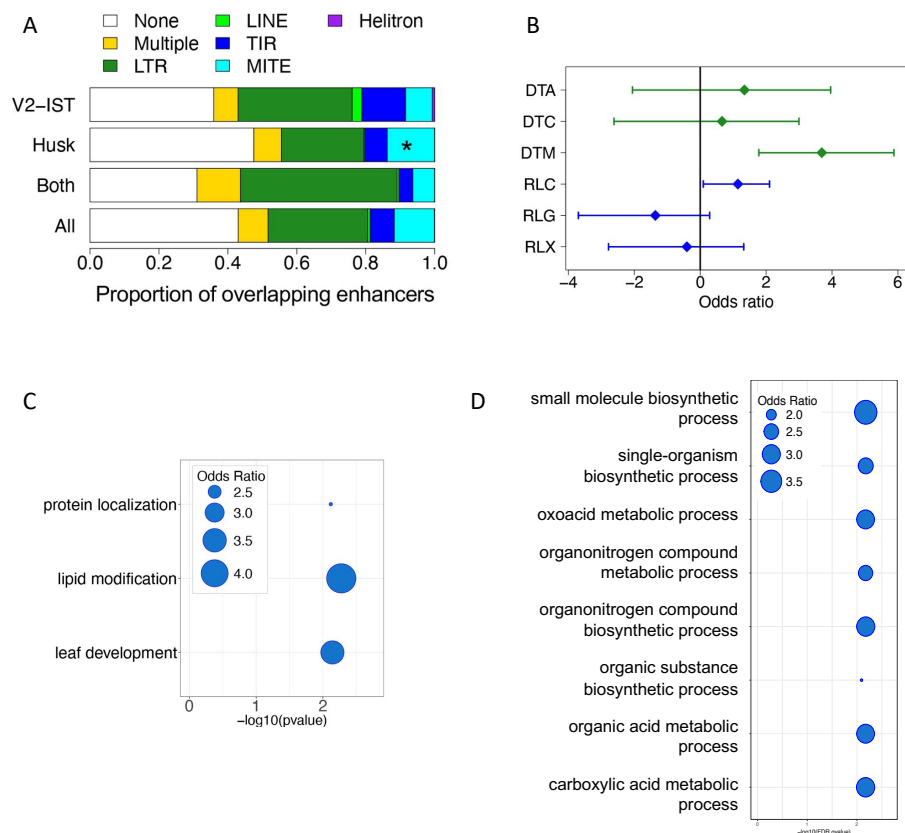


**Figure 13. Structure des modules des réseaux détectés dans les deux tissus**

Gauche : jeune plante. Droite : spathe. Les deux modules spécifiques à la spathe (H.1 et H.2) sont surlignés en vert. Les trois modules spécifiques de la jeune feuille (V2-IST) (V2.1, V2.2 et V2.3) sont surlignés en bleu. Les modules en gris sont partagés entre les réseaux des deux tissus.

L'analyse des éléments transposables dans ces régions régulatrices a révélé que deux catégories différentes sont enrichies dans chacun des tissus : les transposons TIR *Mutator* dans la jeune feuille et les MITE *PIF/Harbinger* dans la spathe, suggérant la contribution de la dynamique de ces éléments à ces réseaux (Figure 14 A et B). Les éléments *Mutator* portaient un TFBS reconnu par la famille *AP2/ERF*. Les MITE *PIF/Harbinger* retrouvés dans les CRE spécifiques des spathe ne portaient pas de TFBS connu, mais l'annotation *de novo* a révélé trois TFBS, dont les plus proches homologues chez *Arabidopsis* sont reconnus par des facteurs de transcription des familles HB, MADS et MYB. De façon intéressante, les gènes ciblés par les 'enhancers' portant le TFBS de type MYB sont impliqués dans des fonctions de type « modification des lipides » et « développement de la feuille », importantes pour le

développement de feuilles adultes comme les spathes (Figure 14 C), alors que les ‘enhancers’ de la jeune feuille dont le TFBS est reconnu par la famille AP2/ERF sont enrichis dans des processus liés à l’azote (Figure 14 D).



**Figure 14. Contribution des éléments transposables dans la régulation de l'expression des gènes en fonction des tissus.**

A. Proportion des différents ordres d'éléments transposables dans les ‘enhancers’ des différents tissus. « None » correspond à l'absence de chevauchement avec une annotation d'élément transposable, et la mention « Multiple » correspond à la présence de chevauchements avec des éléments transposables d'au moins deux ordres différents. L'astérisque indique une valeur de p-value  $p \leq 0.05$ . B. Enrichissement des superfamilles des ordres TIR et LTR parmi les TFBS situés dans les ‘enhancers’ spécifiques de la jeune feuille (V2-IST). C. Fonctions biologiques des gènes régulés par des ‘enhancers’ spécifiques de la spathe chevauchant des MITE et portant des TFBS potentiels dans ces ‘enhancers’. Les bulles représentent le ‘odds ratio’ mesurant un enrichissement dans certaines catégories de la Gene Ontology parmi les gènes potentiellement ciblés par les facteurs de transcription se liant au 15-mère AAGGGATTTTATTT. D. Fonctions biologiques des gènes régulés par des ‘enhancers’ spécifiques de la jeune feuille (V2-IST) chevauchant des séquences reconnues par la famille AP2/ERF dont 70% chevauchent des éléments *Mutator*. DTA : TIR hAT, DTC : TIR CACTA, DTM : TIR Mutator, RLC : LTR Copia, RLG : LTR Gypsy, RLX : LTR inconnu.

Ce travail a été réalisé par Maud Fagny (post-doc 18 mois, 2019-2020), sous ma direction, et a été valorisé sous la forme de l'article :

M. Fagny, M. L. Kuijjer, M. Stam, J. Joets, O. Turc, J. Roziere, S. Pateyron, A.Venon, C.Vitte<sup>§</sup> (2021). Identification of key tissue-specific, biological processes by integrating enhancer information in maize gene regulatory networks. *Front. Genet.*, <https://doi.org/10.3389/fgene.2020.606285>

## 5- Trajectoire et perspectives

Le cheminement de mes travaux passés en France et à l'étranger m'ont permis d'acquérir une expertise à la fois sur la dynamique structurale des génomes, notamment celle liée aux éléments transposables, et sur la régulation chromatinienne de l'information qu'ils contiennent. En toute logique, mes travaux actuels se trouvent à l'interface de ces deux thématiques. Si mon projet initial – caractériser l'impact de la variabilité des éléments transposables sur la variabilité de l'expression des gènes chez le maïs – reste d'actualité, j'ai ouvert mes recherches à tous types de séquences impliquées dans la régulation en *cis* de l'expression des gènes. D'autre part, la possibilité s'offrant désormais d'analyser conjointement la variation structurale et de la méthylation sur plusieurs dizaines d'individus même chez des espèces à grand génome, je souhaite développer des projets étudiant l'impact de ces variations sur l'adaptation et l'évolution. J'ai d'abord abordé cette question à l'échelle de l'adaptation sur plusieurs centaines de générations (parties 5.1 et 5.2), et j'ai depuis cette année ouvert mes recherches à deux autres échelles de temps : l'adaptation rapide d'une espèce sur quelques dizaines de générations (partie 5.3) et l'évolution longue en comparant plusieurs d'espèces ayant divergé il y a plusieurs millions d'années (partie 5.4). Je présente ces projets de façon plus développée dans les parties ci-dessous.

Depuis mon recrutement, j'ai acquis une solide expérience de gestion de projet et d'encadrement de personnel. Ayant toujours fait preuve d'indépendance d'esprit dans mes recherches et ayant acquis très tôt des financements pour les réaliser (Bourse Lavoisier, bourse FRM, projet innovant INRAe), ma première véritable expérience de coordination en tant que chargée de recherches a été celle du WP3 du programme investissement d'avenir Amaizing (2011-2021). J'ai ainsi pu développer mes propres recherches dans une complète autonomie financière et intellectuelle tout en participant à des projets collaboratifs. Cette expérience m'a également donné l'opportunité de collaborer avec plusieurs laboratoires publics et privés, et d'orchestrer des expérimentations sur plusieurs plateformes distantes pour la culture des plantes, la génération des données « omiques », et le traitement bio-informatique des données. Depuis mon recrutement, j'ai encadré une doctorante, un CDD et 2 post-doctorantes, co-encadré un doctorant (50%) et un post-doctorant (20%), et travaillé avec 13 techniciens et assistants ingénieurs issus de plusieurs laboratoires. Mes travaux en tant que chargée de recherche ont conduit à la production de huit articles de recherche publiés dans des journaux à comité de lecture, dont deux en tant qu'auteur senior en 2013 et 2021. J'ai depuis contribué à la rédaction et la mise en place de plusieurs projets collaboratifs en tant que partenaire (SeqOccin 2019-2022, EpiFlorE 2024) ou coordinatrice (easigenomics COMET 2022, ANR SATURN 2025). L'ensemble de ces programmes me permet de développer mes recherches de façon indépendante financièrement.

## 5.1 Contribution de la méthylation de l'ADN à l'adaptation du maïs : rôle des éléments transposables ?

Poste actuel (2019-)

*Financement acquis : Projet investissement d'avenir Amaizing (2011-2021), projet SeqOccin (2019-2022)*

*Encadrement : Arnaud Desbiez-Piat (post-doctorat, 2021, 50%), Gabrielle Minne (M1, 2024, 35%)*

Ce projet a pris ses racines il y a déjà plusieurs années dans le cadre du WP3 du programme Amaizing, à travers ma participation à un projet mené par Maud Tenaillon. A partir d'un panel de 67 lignées de premier cycle de maïs Mexicains, Caribéens, Américains et Européens dont les génomes ont été séquencés avec une profondeur d'environ 15x, nous avons reconstruit l'histoire démographique du maïs, ainsi que les routes d'introduction et les épisodes de sélection qui ont mené à l'émergence des maïs Européens. La combinaison de statistiques de différenciation et de diversité a révélé un ensemble de gènes et de réseaux de gènes montrant de forts signaux de sélection et impliqués notamment dans la réponse à l'environnement (herbivorie, infection pathogène, froid, sécheresse) et dans la qualité du grain, suggérant que les contraintes biotiques et abiotiques ont contribué à l'adaptation du maïs.

Ces résultats ont été publiés sous la forme de l'article :

J. T. Brandenburg, T. Mary-Huard, G. Rigaiil, S. J. Hearne, H. Corti, J. Joets, **C. Vitte**, A. Charcosset, S. Nicolas, M. Tenaillon (2017). Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. *PLoS Genet*, 13 (3) e1006666.

Afin de comprendre si les variations épigénétiques ont pu également contribuer à cette adaptation, j'ai ensuite mis en place un projet visant à estimer la contribution de la méthylation de l'ADN à l'adaptation du maïs suite à son Introduction en Europe. Il consiste à analyser les profils de méthylation de l'ADN de 1166 gènes et leurs régions 1kb flanquantes sur le sous-ensemble de 28 lignées étudiées ci-dessus issues des trois groupes génétiques Européens. Ces 1166 gènes ont été choisis pour contraster un ensemble de 666 gènes potentiellement liés à l'adaptation – 445 gènes présentant des traces de sélection et impliquant les trois groupes européens (Brandenburg *et al.*, 2017), 221 gènes identifiés dans la littérature comme étant potentiellement impliqués dans des adaptations du maïs, dont 42 gènes de floraison – et 500 gènes « contrôle » ne présentant aucune trace de sélection et échantillonnés aléatoirement dans le génome. J'ai ainsi généré, pour ces 1166 régions, des données de méthylation ciblée (SeqCapEpi) sur des prélèvements de feuilles issues de ces 28 lignées en triplicat (collaboration Hervé Duborjal et Magalie Leveugle de Biogemma). Une première analyse par Arnaud Desbiez-Piat (post-doctorant, 4 mois en 2021) a mis en évidence des régions montrant des profils intéressants, notamment dans les introns. Les versions du génome ayant vite progressé, la valorisation de ce travail nécessite une ré-analyse sur la version la plus récente du génome. Cette analyse, réalisée par Gabrielle Minne (étudiante M1 en 2024) a demandé un gros travail de conversion afin de pouvoir analyser la structure intron/exon des gènes, très dépendante de l'annotation et donc assez différente entre les deux versions du génome. Ce travail, réalisé en collaboration avec Maud Tenaillon et Maud Fagny, est encore en cours et nous espérons finaliser une publication dans l'année à venir.

Ayant identifié des différences de méthylation entre groupes génétiques, je souhaite maintenant estimer la contribution des éléments transposables à ces différences. Cependant, l'approche par capture n'est pas adaptée pour ce type d'analyse car elle est très sensible à la répétitivité des séquences, qui sont éliminées lors du « design ». Pour répondre à cette question, il faut avoir accès à un ensemble de séquences génomiques assemblées et ayant des données de méthylation. Ainsi, j'ai coordonné la génération d'assemblages génomiques pleine longueur pour 29 lignées fondatrices des programmes d'amélioration Européens (projet SeqOccin 2019-2022) et de données de méthylome (collaboration Thomas Hartwig et Julia Engelhorn, 2024), avec un effort particulier pour qu'elles soient comparables à celles générées sur un ensemble de 26 lignées fondatrices américaines (Hufford *et al.*, 2021). Ces efforts permettent la comparaison d'un ensemble sans précédent du polymorphisme structural et de méthylation sur un ensemble de 55 lignées de maïs représentant la diversité du maïs et pouvant être analysés de façon homogène. Elles permettront de valider les différences de méthylation observées par capture, et d'analyser le polymorphisme des éléments transposables et leur méthylation dans les régions montrant un profil intéressant.

## 5.2 Contribution des éléments transposables à la variabilité des éléments *cis*-régulateurs du maïs

Poste actuel (2019-)

*Financement acquis* : *Projet investissement d'avenir Amaizing (2011-2021), projet transnational easigenomics COMET (2021), projet ANR JCJC NETWITS (2023-2026), projet SeqOccin (2019-2022)*

*Encadrement* : *Abirami Soundiramourty (M2 2021, 50%), Klaus von Grafenstein (M2 2022, 100%), Thémis Lemarchand (M1 2023, 70%), Matthieu Fressac (M2 2024, 100%)*

Une question qui m'intéresse depuis maintenant plusieurs années est la contribution des éléments transposables aux régions *cis*-régulatrices de l'expression des gènes. Dans l'article Fagny *et al.*, 2021, nous avons mis en évidence une forte contribution des éléments transposables aux enhancers impliqués dans la régulation transcriptionnelle de la jeune feuille et de la spathe. La découverte d'un enrichissement d'éléments transposables différents – TIR *Mutator* dans les enhancers de la jeune feuille et MITE *PIF/Harbinger* dans la spathe – suggère un rôle de la dynamique des éléments transposables dans la mise en place des réseaux qui sous-tendent la spécificité tissulaire. Afin de mieux comprendre la contribution des éléments transposables à la dissémination des éléments *cis*-régulateurs, Abirami Soundiramourty (étudiante M1 en 2021) et Klaus von Grafenstein (étudiant M2 en 2022), ont analysé la dynamique de l'amplification des éléments MITE *PIF/Harbinger*, des TFBS qu'ils portent, et de leur méthylation. Ce travail va être poursuivi cette année afin de produire une publication.

Au-delà de leur implication dans le développement, les éléments transposables pourraient jouer un rôle peut-être encore plus important dans la réponse des plantes à l'environnement (voir introduction). Afin de tester cette hypothèse, j'ai mis en place un projet visant à caractériser des éléments *cis*-régulateurs de la spathe en réponse au déficit hydrique chez le maïs, en collaboration avec Maud Fagny. Nous avons ainsi généré des données d'ATAC-seq et de mRNA-seq pour 6 lignées de maïs en conditions contrôle et condition de déficit hydrique, avec 4 réplicats par condition (collaboration Randy Wisser et Olivier Turc, LEPSE, expertise technique de David Latrasse, IPS2). Une première analyse par Thémis Lemarchand (étudiante M1 en 2023) a permis de mettre en évidence 338 régions différentiellement accessibles entre

les deux conditions pour le génotype B73. L'analyse combinées des données d'ouverture de chromatine, de méthylation, de transcriptomique et d'annotation des gènes et des éléments transposables de ces régions par Matthieu Fressac (étudiant M2 en 2024) a mis en évidence des résultats intéressants sur le lien entre éléments transposables et régions *cis*-régulatrices. A partir de ces résultats, je souhaite développer deux axes. D'une part, identifier des familles d'éléments transposables particulièrement impliquées dans la dissémination des régions *cis*-régulatrices qui répondent au déficit hydrique, et analyser leur dynamique dans différents génotypes afin d'estimer leur contribution à la variabilité de ces régions dans les différents groupes de maïs. D'autre part, analyser la contribution des différents types de mutations (SNP, variants structuraux, méthylation) à la variabilité des séquences *cis*-régulatrices, et estimer si les séquences impliquées dans la réponse à l'environnement ont une dynamique différente de celles impliquées dans la tissu-spécificité. Ces deux axes sont réalisables grâce aux données d'assemblage et de méthylome que nous avons à disposition pour un ensemble de 55 lignées de maïs (voir 5.1). Enfin, je pourrais analyser l'impact de la variabilité génétique de ces régions *cis*-régulatrices sur la transcription, grâce à l'analyse conjointe du typage présence/absence de ces régions (données Illumina 8x, projet SeqOccin), et des niveaux de transcription (données 3' mRNAseq Amaizing) sur un panel de 350 lignées.

### 5.3 Contribution des éléments *cis*-régulateurs à l'évolution phénotypique à court terme

Poste actuel (2024-)

*Financement acquis : Projet INRAE EpiFlorE (partenaire)*

Une question centrale en évolution est la compréhension des processus qui sous-tendent les changements phénotypiques sur des temps évolutifs courts. Pour appréhender cette question, une expérience de sélection divergente de la période de floraison du maïs, a été menée au Moulon pendant plus de 20 ans à partir de deux génotypes de maïs (MBS847 et F252) : chaque année, les plantes les plus précoces des génotypes précoces et les plantes les plus tardives des génotypes tardifs ont été sélectionnées et autofécondées, et les semences produites ont été utilisées pour générer la génération suivante. Cette expérience a permis de générer des familles de plantes à floraisons plus précoces et plus tardives que les lignées d'origine. En seulement 20 générations, cette sélection a permis un décalage de floraison de deux semaines entre les familles précoces et tardives issues d'une même lignée. L'analyse de 18 générations a montré que ce changement résulte de la sélection à la fois d'une hétérozygotie résiduelle et de nouvelles mutations. Sur la base de l'analyse des régions codantes, ces mutations semblent rares. De façon contrastée, les génotypes précoces et tardifs montrent des différences d'abondances moléculaires (Tenaillon *et al.*, 2019). L'ensemble de ces observations suggère que des mutations modifiant l'expression des gènes pourraient avoir évolué au cours de l'expérience. Afin de mieux comprendre les mécanismes moléculaires qui sous-tendent l'évolution rapide des populations, Elodie Marchadier a développé un projet visant à caractériser les modifications génétiques et épigénétiques impliquées dans la réponse à la sélection du maïs pour sa date de floraison. Ces modifications pourraient notamment affecter les régions *cis*-régulatrices, au niveau génétique et/ou épigénétique, et, c'est donc naturellement qu'Elodie et moi avons fait émerger une collaboration. D'autre part, le développement du projet nécessite des ressources et techniques que je contribue à apporter. Le principe du projet résidant dans la comparaison de

familles d'individus contrastés pour la précocité de floraison par rapport au lot de graine initial, il nécessite d'aligner les lectures générées sur un génome de haute qualité de cette lignée, que j'ai justement générée dans le cadre du programme SeqOccin. De plus, l'équipe de Leandro Quadrana a développé une méthode fondée sur les lectures longues nanopore, et qui permet de révéler conjointement l'ouverture de chromatine et la méthylation de l'ADN, sur des lectures de plusieurs kilobases, qui donnent accès à la caractérisation de variants structuraux. Nous avons montré que cette méthode permet d'analyser plus de 90% du génome du maïs, ce qui est un saut qualitatif important par rapport à ce qui est possible de faire avec des approches comme l'ATAC-seq ou le bisulfite-seq sur cette espèce. D'autre part, le typage conjoint des deux marques épigénomiques permet de vérifier la co-occurrence des signaux sur les mêmes noyaux.

Ce travail a permis la publication de l'article :

B. Leduque, A. Edera, **C. Vitte**, L. Quadrana (2024). Simultaneous Profiling of Chromatin Accessibility and DNA Methylation in Complete Plant Genomes Using Long-Read Sequencing. *Nuc. Acid Research* 52(11) : 6285–6297, <https://doi.org/10.1093/nar/gkae306>

L'ensemble des caractéristiques du SAM-seq en fait une technique particulièrement adaptée à la réussite de ce projet, et l'alignement des lectures générées en SAM-seq sur notre génome de haute qualité permettra une analyse fiable des données et la caractérisation fine des variants génétiques (SNP, variants structuraux) et épigénétique (méthylation de l'ADN, ouverture de la chromatine) sur l'ensemble des individus analysés.

## 5.4 Contribution des éléments transposables à l'évolution des espèces

Poste actuel (2024-)

*Financement acquis : ANR SATURN (2025-2029)*

Bien que la contribution des éléments transposables à la régulation en *cis* de certains gènes ait été soigneusement documentée dans certains cas, nous n'avons toujours pas une compréhension complète de la contribution des éléments transposables à la régulation en *cis*. Chez les mammifères, une série d'analyses à grande échelle des profils de régulation de l'expression des gènes a révélé le lien étroit entre éléments transposables et CRE, et suggère l'implication d'une grande variété de séquences d'éléments transposables dans un large éventail de fonctions moléculaires et de processus de régulation (Chuong, Elde, and Feschotte, 2017). Des études comparatives récentes chez les mammifères et les primates ont également révélé que ces éléments contribuent à la fois aux régions régulatrices conservées et aux régions régulatrices spécifiques à l'espèce (Trizzino *et al.*, 2017; Fueyo *et al.*, 2022).

De nombreux CRE de plantes ont été caractérisés en utilisant des combinaisons de données de méthylome, d'accessibilité de la chromatine et marques d'histone. Ces études ont montré une forte contribution des éléments transposables aux CRE (Lu *et al.*, 2019; Noshay *et al.*, 2020; Fagny *et al.*, 2021), et la comparaison d'espèces étroitement apparentées a révélé l'importance des éléments transposables dans les CRE espèce-spécifiques (Lu *et al.*, 2019). Cependant, l'absence d'une analyse comparée entre plusieurs espèces dans un contexte évolutif n'a pas permis d'explorer la contribution des éléments transposables à la régulation des fonctions anciennes. Enfin, bien qu'il ait été démontré que la dynamique des éléments

transposables est étroitement liée aux signaux environnementaux (Ito 2022), l'étendue de leur contribution à la régulation de l'expression des gènes réagissant à différents stress reste encore peu caractérisée à l'échelle globale.

Pour caractériser de façon plus complète la contribution des éléments transposables à l'évolution de l'expression des gènes au cours de l'évolution, deux verrous doivent être levés. Premièrement, nous manquons d'une caractérisation approfondie des éléments transposables, des plus jeunes aux plus anciens. Or, les annotations actuelles des génomes de plantes se sont largement concentrées sur l'identification des copies jeunes et facilement reconnaissables, tandis que les annotations des copies anciennes sont incomplètes ou tout simplement absentes. Deuxièmement, nous manquons d'une caractérisation étendue des CRE sur un ensemble d'espèces et dans plusieurs conditions environnementales. Les approches fondées sur la technologie Illumina classiquement utilisées pour décrire les CRE (bisulfite-seq, ATAC-seq) sont confrontées à des problèmes d'alignement dans les régions répétées, ce qui limite la découverte des CRE à l'échelle du génome entier, et en particulier dans les séquences de TE. Ces deux verrous peuvent être levés grâce aux outils et l'expertise développés à l'URGI, et par la méthodologie SAM-seq (voir partie 5.3).

Pour explorer ces questions, j'ai ainsi mis en place un projet collaboratif avec Leandro Quadrana (IPS2) et Hadi Quesneville et Johann Confais (URGI), reposant sur nos expertises complémentaires, dans lequel nous proposons de caractériser la contribution des TE aux voies moléculaires qui sous-tendent l'évolution de la régulation des gènes chez les plantes à fleurs. Pour étudier l'évolution à court et à long terme, nous utiliserons 6 espèces qui ont divergé les unes des autres entre 20 et 200 Ma : *Arabidopsis thaliana*, le chou (*Brassica oleracea*), la tomate (*Solanum lycopersicum*), *Brachypodium distachyon*, le maïs (*Zea mays*) et le riz (*Oryza sativa*). En utilisant ces espèces comme modèles, nous déploierons une approche interdisciplinaire combinant l'annotation profonde des éléments transposables du génome, la caractérisation épigénomique de pointe des CRE, la génomique et l'épigénomique comparatives, et la validation fonctionnelle, afin d'atteindre trois objectifs : Évaluer la contribution des TE récents et anciens aux CRE impliqués dans la régulation des gènes induits par la différenciation tissulaire ou le stress chez les plantes ; Décrypter la manière dont les éléments transposables récents et anciens contribuent à la régulation génique conservée et spécifique à l'espèce ; Valider les CRE récents et anciens induits par les éléments transposables en tant que régulateurs fonctionnels de l'expression des gènes.

## 6- Faire vivre la science

### 6.1 Former des étudiants

Nous sommes tous différents, et encadrer un étudiant c'est avant tout respecter une personne, accepter ses différences, accepter ses limites pour l'aider à progresser. S'adapter à l'autre pour lui permettre de produire le meilleur de lui-même, lui laisser la place de s'exprimer, d'apporter ses idées. Lui apporter des pistes de progrès et de réussite, lui donner les moyens de faire progresser ses connaissances et ses compétences, mais aussi d'aiguiller son choix professionnel. Il me tient donc particulièrement à cœur à la fois de former les étudiants à la recherche en biologie, et de les accompagner vers leur choix de carrière, en prenant en compte leurs compétences et leurs envies.

Pour former les étudiants à la recherche, il me paraît primordial de leur transmettre le contexte scientifique de l'émergence du projet, les objectifs de leur travail dans ce contexte global, et comment les atteindre. J'espère ainsi leur donner l'envie de la recherche, les motiver à faire un travail qui ait du sens. Il est également important pour moi de transmettre aux étudiants trois valeurs fondamentales : la rigueur (de l'expérimentation et des analyses), l'autocritique (face à ses résultats et à leur interprétation), et l'honnêteté (dans la divulgation des résultats).

Au-delà de ces valeurs, former des étudiants c'est aussi leur transmettre des compétences et un savoir-faire sur une thématique particulière. J'ai ainsi formé des étudiants de biologie aux analyses bio-informatiques et biostatistiques, et des étudiants en bio-informatique/biostatistiques à l'interprétation des résultats pour la biologie. La plupart partaient d'un niveau peu avancé dans ce que je leur ai appris, et j'ai complété, par une approche pratique, les compétences qu'ils avaient acquise au cours de leur parcours académique. Je suis particulièrement fière des deux doctorants que j'ai eu la chance d'encadrer, qui avaient eu chacun à peine une heure de cours en épigénétique et qui n'avaient aucune compétence dans les expérimentations et analyses qu'ils ont menées. Ils ont tous les deux aujourd'hui un poste permanent dans le milieu de la biologie, dans le privé et dans le public, et je suis heureuse d'avoir contribué à leur développement professionnel.

Dans une période où les étudiants s'écartent de plus en plus de la recherche académique, les aider à développer leur projet professionnel de façon objective est particulièrement important. Au cours de ces dernières années, j'ai observé plus de doutes, plus de flou dans leurs réflexions personnelles. Plus de stress aussi, face à l'avenir. Je m'efforce alors de prendre les choses de façon méthodique, en leur demandant de faire un bilan des tâches qu'ils ont aimé et moins aimé faire, de leurs qualités, de leurs compétences, et des manques à combler. Je leur propose ensuite ma vision. C'est impressionnant parfois combien les deux divergent. Confronter ces visions les aide à se positionner, et beaucoup m'ont remerciée d'avoir pris ce temps et de les avoir aidés à y voir plus clair. C'est toujours un grand plaisir pour moi de les aider à évoluer.

Au-delà des étudiants que j'encadre au laboratoire, j'ai participé à 37 comités de thèse pour 20 étudiants. Ces échanges sont importants, car ils permettent aux doctorants d'avoir un autre regard sur leurs recherches, et de discuter ouvertement avec des collègues extérieurs des éventuels problèmes rencontrés. C'est un échange à la fois scientifique et humain, qui demande de la psychologie pour arriver, sur un temps court, à comprendre le contour du

travail, des conditions dans lequel il est réalisé, du moral de l'étudiant, et trouver les moyens de l'aider au mieux. J'ai rencontré des situations complexes, dans lesquelles je pense avoir été utile pour débloquer la situation, apporter un champ des possibles différent de celui envisagé pour que chacun (doctorant et encadrant) y trouve son compte. J'apprécie aussi de pouvoir suivre ces étudiants pendant plusieurs années, c'est une chance de pouvoir les aider au long court à faire les bons choix d'expérimentation et d'analyse, de suivre le développement de leurs travaux et de soutenir le début de leur carrière scientifique.

Depuis le début de mon parcours professionnel, j'ai aussi participé à la formation des étudiants *via* l'enseignement. J'ai d'abord enseigné la microbiologie et la génétique en tant que monitrice pendant mes 3 années de thèse, puis juste après mon recrutement, j'ai participé à une série de cours et de travaux pratiques à la station biologique de Roscoff, et j'ai donné des cours-séminaires aux étudiants du master GCC de l'université Paris 6. Je donne actuellement des cours/conférences en génomique, épigénomique et dynamique des éléments transposables aux étudiants de 3<sup>ème</sup> année d'AgroParisTech depuis 2011, aux mathématiciens du master EvoGEM depuis 2022, et un cours sur les éléments transposables aux étudiants du master GENE en octobre 2024. Interagir avec une classe nécessite de trouver une dynamique pour que chacun puisse participer. Chaque ensemble d'élèves est différent, la dynamique de groupe change entre les masters et d'une année à l'autre, et induit des adaptations du cours, et des digressions que je trouve toujours formatrices, pour eux et pour moi. J'en profite pour discuter un peu d'éthique, leur expliquer le cheminement d'un résultat scientifique, la différence entre un « preprint » et un article publié et l'importance de la revue par les pairs. Ils sont souvent étonnés et très curieux de comprendre les rouages de la recherche. Je profite des inter-cours pour discuter de leurs projets professionnels, j'essaye de leur donner des pistes pour les aider dans leur choix d'orientation. Cette interaction est toujours enrichissante, et me permet aussi de mieux connaître la vision de la génération en cours.

Enfin, sans voir les étudiants en personne, je contribue à leur formation par la rédaction de chapitres d'ouvrages, notamment deux que j'ai co-écrits récemment dans le cadre du GDR « Epigénétique en Ecologie et Evolution » :

C. Vitte, S. Chambeyron and C. Vieira (2024). Epigénétique et Eléments Transposables. In Raconte-moi l'épigénétique, editors C. Grunau and S. Maury. ISTE, p.76-90 (sous presse)

C. Vitte, N. Nègre (2024). L'épigénétique : la même pour toutes les espèces? In Raconte-moi l'épigénétique, editors C. Grunau and S. Maury. ISTE, p.76-90 (sous presse)

## 6.2 Diffuser les connaissances

A quoi sert de générer un résultat s'il n'est pas transmis au reste de la communauté, et à la société ?

Depuis le début de ma carrière, j'ai contribué à la diffusion du savoir par l'écriture de synthèses publiées sous forme de cinq chapitres d'ouvrages et de revues dans des journaux scientifiques à comité de lecture. Ils sont souvent cités, et leurs schémas sont souvent utilisés dans des présentations orales de collègues, en congrès ou en cours. L'invitation par des collègues internationaux à écrire ces chapitres d'ouvrage révèle aussi leur intérêt pour nos travaux, et permet des échanges intéressants sur le sujet.

*Articles de revue publiés dans des journaux à comité de lecture*

M. Mirouze, **C. Vitte** (2014) Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. *J Exp Bot.* 65(10):2801-12.

**C. Vitte**, M.-A. Fustier, K. Alix, M. Tenaillon (2014). The bright side of transposons in crop evolution. *Briefings in Functional Genomics.* 13(4):276-95.

**C. Vitte**, O. Panaud (2005). LTR retrotransposons and plant genome size: emergence of the increase/decrease model. *Cytogenetics and Genome Research.* 110(1-4):91-107. Article invité dans l'édition spéciale « Retrotransposons, Retrotransposition and Genome Evolution ».

*Chapitres de livres (invités)*

J. Joets, **C. Vitte**, A. Charcosset (2019). Draft assembly of the F2 European maize genome sequence and its comparison to the B73 genome sequence: a characterization of genotype-specific regions. In *The maize Genome*, editors Jeffrey Bennetzen, Professor. Sherry Flint-Garcia, Professor. Candice Hirsch and Dr. Roberto Tuberosa. Springer Nature, p.3-12

C.M. Diez, **C. Vitte**, J. Ross-Ibarra, B.S. Gaut, M. Tenaillon (2013). Using next generation sequencing to investigate genome size variation and transposable element content. In M. A. Grandbastien and J. Casacuberta, *Plant transposons (Topics in current genetics)*, Springer-Verlag, Germany.

J. Joets, **C. Vitte**, S. Pasternak, D. Ware (2011). Deciphering the multi-faceted aspects of the maize genome: new insights from sequencing projects. In J.-L. Prioul, *Advances in Maize (Taylor & Francis)*, the Society for Experimental Biology, London, UK. pp. 111-120.

**C. Vitte**, T. Peterson, J. Bennetzen (2011). « Transposable elements and the evolving structure of the maize genome ». In J.-L. Prioul, *Advances in Maize (Taylor & Francis)*, the Society for Experimental Biology, London, UK. pp. 2-36.

P. SanMiguel, **C. Vitte** (2009). « The LTR retrotransposons of maize ». In J. Bennetzen and S. Hake, *The Maize Handbook – Volume II: Domestication, Genetics and Genomics (Springer, ISBN 0387778624, 9780387778624)*, pp.307–327.

Ayant interagit avec les acteurs de la filière maïs dans le cadre du programme Amaizing, j'ai également participé à leur rendre accessible le contenu de notre projet, par l'écriture d'un article pour *Perspectives Agricoles*, le magazine de référence du conseil agricole :

J.P. Cohan, A. Charcosset, F. Tardieu, P. Rogowsky, **C. Vitte**, L. Moreau, J. Joets et C. Giauffret (2020) « Concevoir et évaluer les maïs du futur » pour le journal *Perspectives Agricoles*, septembre 2020, numéro 420

### 6.3 Transmettre le goût de la science et de la recherche

La science est vue comme un enjeu d'importance pour la société. Mais la perception des sciences est très dépendante de la confiance envers les institutions scientifiques, et les personnes qui les représentent. Au cours des dernières années, cette confiance a diminué, en lien notamment avec les aléas des opinions qui circulent sur les réseaux sociaux. En tant que scientifique, il me semble important de participer à la reconquête de cette confiance, notamment en interagissant avec les jeunes générations.

Cette confiance passe d'abord par un rapport humain entre les jeunes et les chercheurs. C'est pourquoi j'ai contribué de façon active à des rencontres avec des classes de terminales, dans le cadre du programme « chercheurs en herbe », puis à faire visiter notre laboratoire à des classes de lycée dans le cadre du programme « des plantes et des hommes », mis en place par le labex SPS. Cette expérience m'a permis à la fois de transmettre des connaissances (et de faire découvrir les éléments transposables !), mais aussi de mettre un visage humain sur mon métier, et d'aider les élèves dans leur choix d'orientation en les faisant réfléchir sur les qualités à avoir pour être chercheur, et sur les matières majeures à maîtriser pour y parvenir. La préparation avec les professeurs m'a également fait réaliser qu'ils manquent de source pour préparer leurs cours. J'ai participé à combler ce manque par une « interview » qui a servi à la rédaction du dossier « Transposons, les gènes qui boostent l'évolution » dans le journal Epsilon en 2022. Cet article m'est également très utile lors de l'accueil des élèves de troisième et de seconde au laboratoire.

Mais la science, comme toute autre façon d'expliquer le monde qui nous entoure, peut s'acquérir chez les enfants dès leur plus jeune âge. Afin de participer à la formation des plus jeunes, j'ai mis en place un atelier autour de la graine, de la diversité et de la domestication, que j'ai proposé à des groupes de CP et CE1 dans le centre de loisir de ma commune. Enfin, cette année je vais me lancer dans l'interaction avec une tranche d'âge intermédiaire, dans le cadre du programme « Déclic collège » mis en place par l'université Paris-Saclay.

Par l'ensemble de ces actions, je sème des graines pour le goût de la biologie. J'espère qu'au moins certaines d'entre elles germeront !

### 6.4 Promouvoir des interactions entre chercheurs

La dynamique de la recherche scientifique dépend aussi de la transmission des résultats acquis entre les acteurs de la communauté scientifique elle-même. Depuis les débuts de ma thèse, j'ai été un membre actif de la communauté des éléments transposables. Je tiens particulièrement au Groupe de Recherches (GDR) « Eléments Transposables », qui permet à notre communauté de se retrouver et d'échanger de façon récurrente. J'ai également apporté une contribution active à la mise en place du Réseau Thématique Pluridisciplinaire « Epigénétique en Ecologie et Evolution » (RTP 3E) et de son évolution en GDR 3E. Cette initiative nous a permis de rapprocher les communautés de biologie moléculaire et d'écologie/évolution intéressées par l'épigénétique. Les communautés se sont finalement organisées autour des organismes (GDR EpiMAR, Epiplant), mais les liens que nous avons créés au sein du comité de pilotage et au-delà, nous aideront à faire des ponts dans le futur. Dans le cadre de ces instances et au-delà, j'ai participé à l'organisation de 4 congrès nationaux et 2 congrès internationaux, 1 symposium international, et j'organise de façon annuelle le workshop « Transposable Elements » au sein du congrès international « Plant and Animal Genomes » à San Diego (Etats-Unis) depuis 2020.

#### Co-organisatrice de congrès nationaux :

Congrès National des Eléments Transposables (CNET), Gif-sur-Yvette, 2022

Colloque « Epigénétique en écologie et évolution », Gif sur Yvette, 2013

Congrès « The central dogma of molecular biology revisited: high-throughput sequencing and new concepts », Institut Curie, Paris, 2013

Colloque « Eléments Transposables, Virus et Domestication des Espèces », Lyon, 2010

#### Co-organisatrice de congrès internationaux :

Workshop « Transposable Elements » au sein du congrès international « Plant and Animal Genomes » à San Diego, Etats-Unis depuis 2020

International Congress on Transposable Elements (ICTE) en 2012 et 2024

Symposium « Epigenetics in ecology and evolution » au sein du Congrès international de la SMBE, Lyon, 2010

L'organisation de ces congrès ma conduite à participer à l'écriture d'un « proceeding » de congrès :

A. Ainouche, M. Betermier, M. Chandler, R. Cordaux, G. Cristofari, J.M. Deragon, P. Lesage, O. Panaud, H. Quesneville, C. Vaury, C. Vieira, **C. Vitte**. (2012) International Congress on Transposable Elements (ICTE) 2012 in Saint Malo and the sea of TE stories. *Mob. DNA*, 3 (1) 17.

et de participer à l'édition d'une collection d'articles pour laquelle j'ai contribué à la rédaction de l'article éditorial :

G. Moissiard, M. Mirouze, C. Carles, **C. Vitte**. Editorial: Plant epigenetics and chromatin dynamics - EPIPLANT 2021-2022. *Frontiers*. <https://doi.org/doi:10.3389/fpls.2023.1260391>

Enfin, j'ai participé comme oratrice à l'école chercheurs « Génomique et diversité des caractères à déterminisme complexe », et, dans le cadre du GDR 3E, j'ai participé à l'organisation d'une école chercheurs sur le thème « Epigénétique en Ecologie et Evolution » dans laquelle j'ai donné des cours interactifs et des travaux pratiques sur l'utilisation de la technique de bisulfite-seq pour l'analyse des méthylomes de plantes.

## 6.5 Produire une science de qualité

Comme indiqué plus haut (partie 6.3), nous avons, en tant que chercheurs, la responsabilité de garantir la qualité de la science, et donc celle des données, des procédures et des résultats que nous produisons. Pour cette raison, je suis très attachée au « reviewing » par les pairs. J'ai contribué à ce processus en expertisant plus de 30 articles dans des journaux à comité de lecture comme *Nature Plant*, *PLoS Genetics*, *The Plant Journal*, *New Phytologist*, *Journal of Experimental Botany*, *Annals of Botany*, *Molecular Biology and Evolution*, *Genome Biology and Evolution*.

Au cours de ces dernières années, j'ai eu l'occasion d'avoir un rôle d'éditeur, dans le cadre d'une collection pour *Frontiers* (voir partie 6.4), puis en tant qu'invitée pour un article du

journal PLoS Genetics. C'est un travail qui demande du temps si on veut le faire bien, mais qui permet d'avoir un impact fort sur la science qui est publiée. C'est pour cette raison que j'ai accepté l'invitation de rejoindre les bureaux éditoriaux des journaux « Mobile DNA » en 2023 et « PLoS Genetics » en 2024.

Au-delà de ce travail de relecture d'article, je participe activement à l'expertise des travaux de thèse. J'ai été examinatrice pour 12 thèses en France et 2 à l'étranger (Brésil, Suisse), et rapportrice pour une thèse à l'étranger (Belgique).

## 6.6 Recruter les chercheurs de demain

Promouvoir la recherche, c'est aussi consolider son futur en recrutant les chercheurs qui l'animeront demain. Au cours de mon parcours, j'ai participé à trois jurys de concours, pour INRAe, et pour diverses universités en France. C'est une expérience un peu frustrante du fait de la limitation du nombre de postes. Mais c'est aussi une grande responsabilité, car nous recrutons sur le long terme et que notre choix a des conséquences sur les recherches qui seront développées, mais aussi sur l'équilibre des laboratoires. Il s'agit pour moi d'arriver à trouver la bonne adéquation entre une personne (sa thématique, ses compétences, mais aussi sa vision de la recherche et de l'enseignement) et un laboratoire d'accueil.

J'ai aussi participé à trois jurys de concours d'école doctorale, deux fois pour l'école doctorale SEVE de Paris-Saclay, et une fois pour l'école doctorale SEVAB de l'université de Toulouse. Les enjeux sont là-encore très forts car le choix du jury oriente fortement la carrière future des étudiants, et implique donc une grande responsabilité. Dans cet exercice, il me semble important de trouver des questions qui permettent aux étudiants de mettre en valeur à la fois leurs connaissances et compétences, leur motivation et leur intérêt pour cet expérience particulière qu'est la thèse.

## 7- Regard sur mon parcours

« Je suis sédimentation » a écrit l'anthropologue Françoise l'Héritier. En effet, je crois que le frottement à l'autre, à ses différences de point de vue, de culture, de connaissances, nous permet de créer de nouveaux sédiments, de faire germer de nouvelles idées, de nouveaux projets. C'est là tout l'intérêt de la diversité et de l'altérité.

Au cours de mon parcours, j'ai fait de nombreuses collaborations, avec des collègues du Moulon, d'autres laboratoires académiques français et étrangers, des partenaires privés, des plateformes, des personnes de tous âges et d'horizons multiples. J'ai eu l'occasion de travailler avec des partenaires de visions très diverses, d'encadrer des étudiants de culture différentes, d'interagir avec des chercheurs et ITA experts en divers domaines comme la biologie moléculaire, la bio-informatique, les biostatistiques.

Partager nos connaissances et compétences, mais aussi nos différences et nos divergences a été pour moi une grande richesse, à la fois scientifique et humaine, qui m'a faite progresser, évoluer, et qui a permis de construire les bases de mes projets actuels. J'en profite ici pour remercier toutes les personnes qui ont participé à l'avancée de mes travaux passés, ou qui se lancent aujourd'hui avec moi dans les projets à venir.

« Pas de patience, pas de science » a écrit Jean-Pierre Jarroux. Je peux effectivement attester que la patience et la persévérance sont des qualités indispensables pour un chercheur en biologie. En reprenant mon parcours, j'ai été frappée de la durée entre le moment de la conception du projet et le moment de sa réalisation. Je me suis remémorée tous les essais/erreurs, toutes les embûches techniques qu'il a fallu surmonter pour arriver à mettre en place les idées que j'ai eues. Le chemin a parfois été long, et le découragement parfois proche. Si je regrette que certains projets n'aient pas pu être mieux valorisés, je suis fière d'avoir tenu le cap, et d'avoir contribué, à mon échelle, à l'avancée des connaissances et des ressources. Dans cette période historique où le sens disparaît de la plupart des métiers, je me sens privilégiée de la liberté d'esprit et du sens que le mien m'apporte. J'espère que les conditions actuelles de la recherche ne dissuaderont pas trop notre jeunesse de s'engager dans la recherche académique, et qu'elles finiront par s'améliorer.

# Références bibliographiques

- Baduel, Pierre, and Vincent Colot. 2021. 'The Epiallelic Potential of Transposable Elements and Its Evolutionary Significance in Plants'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1826): 20200123. <https://doi.org/10.1098/rstb.2020.0123>.
- Beguiristain, Thierry, Marie-Angèle Grandbastien, Pere Puigdomènech, and Josep M. Casacuberta. 2001. 'Three Tnt1 Subfamilies Show Different Stress-Associated Patterns of Expression in Tobacco. Consequences for Retrotransposon Control and Evolution in Plants'. *Plant Physiology* 127 (1): 212–21.
- Beló, André, Mary K. Beatty, David Hondred, Kevin A. Fengler, Bailin Li, and Antoni Rafalski. 2010. 'Allelic Genome Structural Variations in Maize Detected by Array Comparative Genome Hybridization'. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 120 (2): 355–67. <https://doi.org/10.1007/s00122-009-1128-9>.
- Bennetzen, J. L., and E. A. Kellogg. 1997. 'Do Plants Have a One-Way Ticket to Genomic Obesity?' *The Plant Cell* 9 (9): 1509–14. <https://doi.org/10.1105/tpc.9.9.1509>.
- Brandenburg, Jean-Tristan, Tristan Mary-Huard, Guillem Rigau, Sarah J. Hearne, Hélène Corti, Johann Joets, Clémentine Vitte, Alain Charcosset, Stéphane D. Nicolas, and Maud I. Tenaillon. 2017. 'Independent Introductions and Admixtures Have Contributed to Adaptation of European Maize and Its American Counterparts'. *PLoS Genetics* 13 (3): e1006666. <https://doi.org/10.1371/journal.pgen.1006666>.
- Bremer, Kare. 2004. 'The Age of Major Monocot Groups Inferred from 800+ rbcL Sequences'. *Botanical Journal of the Linnean Society*, January. [https://www.academia.edu/24878398/The\\_age\\_of\\_major\\_monocot\\_groups\\_inferred\\_from\\_800\\_rbcL\\_sequences](https://www.academia.edu/24878398/The_age_of_major_monocot_groups_inferred_from_800_rbcL_sequences).
- Brown, W. L. 1949. 'Numbers and Distribution of Chromosome Knobs in United States Maize'. *Genetics* 34 (5): 524–36. <https://doi.org/10.1093/genetics/34.5.524>.
- Buckler, Edward S., Brandon S. Gaut, and Michael D. McMullen. 2006. 'Molecular and Functional Diversity of Maize'. *Current Opinion in Plant Biology* 9 (2): 172–76. <https://doi.org/10.1016/j.pbi.2006.01.013>.
- Chia, Jer-Ming, Chi Song, Peter J. Bradbury, Denise Costich, Natalia de Leon, John Doebley, Robert J. Elshire, et al. 2012. 'Maize HapMap2 Identifies Extant Variation from a Genome in Flux'. *Nature Genetics* 44 (7): 803–7. <https://doi.org/10.1038/ng.2313>.
- Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2017. 'Regulatory Activities of Transposable Elements: From Conflicts to Benefits'. *Nature Reviews. Genetics* 18 (2): 71–86. <https://doi.org/10.1038/nrg.2016.139>.
- Cokus, Shawn J., Suhua Feng, Xiaoyu Zhang, Zugen Chen, Barry Merriman, Christian D. Haudenschild, Sriharsa Pradhan, Stanley F. Nelson, Matteo Pellegrini, and Steven E. Jacobsen. 2008. 'Shotgun Bisulphite Sequencing of the Arabidopsis Genome Reveals DNA Methylation Patterning'. *Nature* 452 (7184): 215–19. <https://doi.org/10.1038/nature06745>.
- Crisp, Peter A., Alexandre P. Marand, Jaclyn M. Noshay, Peng Zhou, Zefu Lu, Robert J. Schmitz, and Nathan M. Springer. 2020. 'Stable Unmethylated DNA Demarcates Expressed Genes and Their Cis-Regulatory Space in Plant Genomes'. *Proceedings of the National Academy of Sciences of the United States of America* 117 (38): 23991–0. <https://doi.org/10.1073/pnas.2010250117>.
- Fagny, Maud, Marieke Lydia Kuijjer, Maike Stam, Johann Joets, Olivier Turc, Julien Rozière, Stéphanie Pateyron, Anthony Venon, and Clémentine Vitte. 2021. 'Identification of Key Tissue-Specific, Biological Processes by Integrating Enhancer Information in Maize Gene Regulatory Networks'. *Frontiers in Genetics* 11 (January):606285. <https://doi.org/10.3389/fgene.2020.606285>.
- Fedoroff, Nina V. 2012. 'Transposable Elements, Epigenetics, and Genome Evolution'. *Science* 338 (6108): 758–67. <https://doi.org/10.1126/science.338.6108.758>.

- Finkers, Richard, Martijn van Kaauwen, Kai Ament, Karin Burger-Meijer, Raymond Egging, Henk Huits, Linda Kodde, *et al.* 2021. 'Insights from the First Genome Assembly of Onion (*Allium Cepa*)'. *G3 Genes/Genomes/Genetics* 11 (9): jkab243. <https://doi.org/10.1093/g3journal/jkab243>.
- Flavell, R. B. 1986. 'Repetitive DNA and Chromosome Evolution in Plants'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 312 (1154): 227–42. <https://doi.org/10.1098/rstb.1986.0004>.
- Fueyo, Raquel, Julius Judd, Cedric Feschotte, and Joanna Wysocka. 2022. 'Roles of Transposable Elements in the Regulation of Mammalian Transcription'. *Nature Reviews. Molecular Cell Biology* 23 (7): 481–97. <https://doi.org/10.1038/s41580-022-00457-y>.
- Ganguly, Deep R., Peter A. Crisp, Steven R. Eichten, and Barry J. Pogson. 2017. 'The Arabidopsis DNA Methylome Is Stable under Transgenerational Drought Stress1[OPEN]'. *Plant Physiology* 175 (4): 1893–1912. <https://doi.org/10.1104/pp.17.00744>.
- Ge, Song, Tao Sang, Bao-Rong Lu, and De-Yuan Hong. 1999. 'Phylogeny of Rice Genomes with Emphasis on Origins of Allotetraploid Species'. *Proceedings of the National Academy of Sciences* 96 (25): 14400–405. <https://doi.org/10.1073/pnas.96.25.14400>.
- Grandbastien, M.-A., C. Audeon, E. Bonnard, J. M. Casacuberta, B. Chalhou, A.-P. P. Costa, Q. H. Le, *et al.* 2005. 'Stress Activation and Genomic Impact of Tnt1 Retrotransposons in Solanaceae'. *Cytogenetic and Genome Research* 110 (1–4): 229–41. <https://doi.org/10.1159/000084957>.
- Hartwig, Thomas, Michael Banf, Gisele Passaia Prietsch, Jia-Ying Zhu, Isabel Mora-Ramírez, Jos H. M. Schippers, Samantha J. Snodgrass, *et al.* 2023. 'Hybrid Allele-Specific ChIP-Seq Analysis Identifies Variation in Brassinosteroid-Responsive Transcription Factor Binding Linked to Traits in Maize'. *Genome Biology* 24 (1): 108. <https://doi.org/10.1186/s13059-023-02909-w>.
- Hashida, Shin-Nosuke, Takako Uchiyama, Cathie Martin, Yuji Kishima, Yoshio Sano, and Tetsuo Mikami. 2006. 'The Temperature-Dependent Change in Methylation of the Antirrhinum Transposon Tam3 Is Controlled by the Activity of Its Transposase'. *The Plant Cell* 18 (1): 104–18. <https://doi.org/10.1105/tpc.105.037655>.
- Hufford, Matthew B., Arun S. Seetharam, Margaret R. Woodhouse, Kapeel M. Chougule, Shujun Ou, Jianing Liu, William A. Ricci, *et al.* 2021. 'De Novo Assembly, Annotation, and Comparative Analysis of 26 Diverse Maize Genomes'. *Science (New York, N.Y.)* 373 (6555): 655–62. <https://doi.org/10.1126/science.abg5289>.
- Ito, Hidetaka. 2022. 'Environmental Stress and Transposons in Plants'. *Genes & Genetic Systems* 97 (4): 169–75. <https://doi.org/10.1266/ggs.22-00045>.
- Kimura, Y., Y. Tosa, S. Shimada, R. Sogo, M. Kusaba, T. Sunaga, S. Betsuyaku, Y. Eto, H. Nakayashiki, and S. Mayama. 2001. 'OARE-1, a Ty1-Copia Retrotransposon in Oat Activated by Abiotic and Biotic Stresses'. *Plant & Cell Physiology* 42 (12): 1345–54. <https://doi.org/10.1093/pcp/pce171>.
- Kuijjer, Marieke Lydia, Matthew George Tung, GuoCheng Yuan, John Quackenbush, and Kimberly Glass. 2019. 'Estimating Sample-Specific Regulatory Networks'. *iScience* 14 (April):226–40. <https://doi.org/10.1016/j.isci.2019.03.021>.
- Lisch, Damon. 2013. 'How Important Are Transposons for Plant Evolution?' *Nature Reviews Genetics* 14 (1): 49–61. <https://doi.org/10.1038/nrg3374>.
- Lister, Ryan, Ronan C. O'Malley, Julian Tonti-Filippini, Brian D. Gregory, Charles C. Berry, A. Harvey Millar, and Joseph R. Ecker. 2008. 'Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis'. *Cell* 133 (3): 523–36. <https://doi.org/10.1016/j.cell.2008.03.029>.
- Liu, Peng, Diego Cuerda-Gil, Saima Shahid, and R. Keith Slotkin. 2022. 'The Epigenetic Control of the Transposable Element Life Cycle in Plant Genomes and Beyond'. *Annual Review of Genetics* 56 (1): 63–87. <https://doi.org/10.1146/annurev-genet-072920-015534>.
- Lu, Zefu, Alexandre P. Marand, William A. Ricci, Christina L. Ethridge, Xiaoyu Zhang, and Robert J. Schmitz. 2019. 'The Prevalence, Evolution and Chromatin Signatures of Plant Regulatory Elements'. *Nature Plants* 5 (12): 1250–59. <https://doi.org/10.1038/s41477-019-0548-z>.

- Marand, Alexandre P., Andrea L. Eveland, Kerstin Kaufmann, and Nathan M. Springer. 2023. 'Cis-Regulatory Elements in Plant Development, Adaptation, and Evolution'. *Annual Review of Plant Biology* 74 (May):111–37. <https://doi.org/10.1146/annurev-arplant-070122-030236>.
- Messing, Joachim, and Hugo K Dooner. 2006. 'Organization and Variability of the Maize Genome'. *Current Opinion in Plant Biology*, Genome studies and molecular genetics: Part 1: Model legumes / edited by Nevin D Young and Randy C Shoemaker; Part 2: Maize genomics / edited by Susan R Wessler. *Plant biotechnology* / edited by John Salmeron and Luis R Herrera-Estrella, 9 (2): 157–63. <https://doi.org/10.1016/j.pbi.2006.01.009>.
- Miga, Karen H., and Ting Wang. 2021. 'The Need for a Human Pangenome Reference Sequence'. *Annual Review of Genomics and Human Genetics* 22 (August):81–102. <https://doi.org/10.1146/annurev-genom-120120-081921>.
- Mladenov, Velimir, Vasileios Fotopoulos, Eirini Kaiserli, Erna Karalija, Stephane Maury, Miroslav Baranek, Naama Segal, *et al.* 2021. 'Deciphering the Epigenetic Alphabet Involved in Transgenerational Stress Memory in Crops'. *International Journal of Molecular Sciences* 22 (13): 7118. <https://doi.org/10.3390/ijms22137118>.
- Morgante, Michele, Emanuele De Paoli, and Slobodanka Radovic. 2007. 'Transposable Elements and the Plant Pan-Genomes'. *Current Opinion in Plant Biology*, Genome Studies and Molecular Genetics / Edited by Stefan Jansson and Edward S Buckler, 10 (2): 149–55. <https://doi.org/10.1016/j.pbi.2007.02.001>.
- Noshay, Jaclyn M, Alexandre P Marand, Sarah N Anderson, Peng Zhou, Maria Katherine Mejia Guerra, Zefu Lu, Christine H O'Connor, *et al.* 2020. 'Assessing the Regulatory Potential of Transposable Elements Using Chromatin Accessibility Profiles of Maize Transposons'. *Genetics* 217 (1): 1–13. <https://doi.org/10.1093/genetics/iyaa003>.
- Oka, Rurika, Johan Zicola, Blaise Weber, Sarah N. Anderson, Charlie Hodgman, Jonathan I. Gent, Jan-Jaap Wesselink, *et al.* 2017. 'Genome-Wide Mapping of Transcriptional Enhancer Candidates Using DNA and Chromatin Features in Maize'. *Genome Biology* 18 (July):137. <https://doi.org/10.1186/s13059-017-1273-4>.
- Oliver, Sandra N., E. Jean Finnegan, Elizabeth S. Dennis, W. James Peacock, and Ben Trevaskis. 2009. 'Vernalization-Induced Flowering in Cereals Is Associated with Changes in Histone Methylation at the VERNALIZATION1 Gene'. *Proceedings of the National Academy of Sciences of the United States of America* 106 (20): 8386–91. <https://doi.org/10.1073/pnas.0903566106>.
- Padi, Megha, and John Quackenbush. 2018. 'Detecting Phenotype-Driven Transitions in Regulatory Network Structure'. *Npj Systems Biology and Applications* 4 (1): 16. <https://doi.org/10.1038/s41540-018-0052-5>.
- Petrov, Dmitri A., Todd A. Sangster, J. Spencer Johnston, Daniel L. Hartl, and Kerry L. Shaw. 2000. 'Evidence for DNA Loss as a Determinant of Genome Size'. *Science* 287 (5455): 1060–62. <https://doi.org/10.1126/science.287.5455.1060>.
- Pulido, Marc, and Josep M. Casacuberta. 2023. 'Transposable Element Evolution in Plant Genome Ecosystems'. *Current Opinion in Plant Biology* 75 (October):102418. <https://doi.org/10.1016/j.pbi.2023.102418>.
- Ricci, William A., Zefu Lu, Lexiang Ji, Alexandre P. Marand, Christina L. Ethridge, Nathalie G. Murphy, Jaclyn M. Noshay, *et al.* 2019. 'Widespread Long-Range Cis-Regulatory Elements in the Maize Genome'. *Nature Plants* 5 (12): 1237–49. <https://doi.org/10.1038/s41477-019-0547-0>.
- Richards, Eric J. 2006. 'Inherited Epigenetic Variation — Revisiting Soft Inheritance'. *Nature Reviews Genetics* 7 (5): 395–401. <https://doi.org/10.1038/nrg1834>.
- Robbins, T. P., E. L. Walker, J. L. Kermicle, M. Alleman, and S. L. Dellaporta. 1991. 'Meiotic Instability of the R-r Complex Arising from Displaced Intragenic Exchange and Intrachromosomal Rearrangement'. *Genetics* 129 (1): 271–83. <https://doi.org/10.1093/genetics/129.1.271>.
- Salazar, Marcela, Enrique González, José A. Casaretto, Josep M. Casacuberta, and Simón Ruiz-Lara. 2007. 'The Promoter of the TLC1.1 Retrotransposon from *Solanum chilense* Is Activated by

- Multiple Stress-Related Signaling Molecules'. *Plant Cell Reports* 26 (10): 1861–68. <https://doi.org/10.1007/s00299-007-0375-y>.
- Schnable, Patrick S., Doreen Ware, Robert S. Fulton, Joshua C. Stein, Fusheng Wei, Shiran Pasternak, Chengzhi Liang, *et al.* 2009. 'The B73 Maize Genome: Complexity, Diversity, and Dynamics'. *Science (New York, N.Y.)* 326 (5956): 1112–15. <https://doi.org/10.1126/science.1178534>.
- Secco, David, Chuang Wang, Huixia Shou, Matthew D Schultz, Serge Chiarenza, Laurent Nussaume, Joseph R Ecker, James Whelan, and Ryan Lister. 2015. 'Stress Induced Gene Expression Drives Transient DNA Methylation Changes at Adjacent Repetitive Elements'. Edited by Detlef Weigel. *eLife* 4 (July): e09343. <https://doi.org/10.7554/eLife.09343>.
- Shirasu, Ken, Alan H. Schulman, Thomas Lahaye, and Paul Schulze-Lefert. 2000. 'A Contiguous 66-Kb Barley DNA Sequence Provides Evidence for Reversible Genome Expansion'. *Genome Research* 10 (7): 908–15.
- Slotkin, R. Keith, and Robert Martienssen. 2007. 'Transposable Elements and the Epigenetic Regulation of the Genome'. *Nature Reviews. Genetics* 8 (4): 272–85. <https://doi.org/10.1038/nrg2072>.
- Sonawane, Abhijeet Rajendra, John Platig, Maud Fagny, Cho-Yi Chen, Joseph Nathaniel Paulson, Camila Miranda Lopes-Ramos, Dawn Lisa DeMeo, John Quackenbush, Kimberly Glass, and Marieke Lydia Kuijjer. 2017. 'Understanding Tissue-Specific Gene Regulation'. *Cell Reports* 21 (4): 1077–88. <https://doi.org/10.1016/j.celrep.2017.10.001>.
- Spitz, François, and Eileen E. M. Furlong. 2012. 'Transcription Factors: From Enhancer Binding to Developmental Control'. *Nature Reviews Genetics* 13 (9): 613–26. <https://doi.org/10.1038/nrg3207>.
- Springer, Nathan M., Kai Ying, Yan Fu, Tieming Ji, Cheng-Ting Yeh, Yi Jia, Wei Wu, *et al.* 2009. 'Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content'. *PLOS Genetics* 5 (11): e1000734. <https://doi.org/10.1371/journal.pgen.1000734>.
- Studer, Anthony, Qiong Zhao, Jeffrey Ross-Ibarra, and John Doebley. 2011. 'Identification of a Functional Transposon Insertion in the Maize Domestication Gene *Tb1*'. *Nature Genetics* 43 (11): 1160–63. <https://doi.org/10.1038/ng.942>.
- Tenaillon, M.I., K. Sedikhi, M. Mollion, Le Guilloux, M., E. Marchadier, A. Ressayre, and C. Dillman. 2019. 'Transcriptomic Response to Divergent Selection for Flowering Time in Maize Reveals Convergence and Key Players of the Underlying Gene Regulatory Network.' *PCI Community in Evolutionary Biology*, 2019.
- Tettelin, Hervé, Vega Masignani, Michael J. Cieslewicz, Claudio Donati, Duccio Medini, Naomi L. Ward, Samuel V. Angiuoli, *et al.* 2005. 'Genome Analysis of Multiple Pathogenic Isolates of *Streptococcus Agalactiae*: Implications for the Microbial "Pan-Genome"'. *Proceedings of the National Academy of Sciences of the United States of America* 102 (39): 13950–55. <https://doi.org/10.1073/pnas.0506758102>.
- Trizzino, Marco, YoSon Park, Marcia Holsbach-Beltrame, Katherine Aracena, Katelyn Mika, Minal Caliskan, George H. Perry, Vincent J. Lynch, and Christopher D. Brown. 2017. 'Transposable Elements Are the Primary Source of Novelty in Primate Gene Regulation'. *Genome Research* 27 (10): 1623–33. <https://doi.org/10.1101/gr.218149.116>.
- Van Dooren, Tom J. M., Amanda Bortolini Silveira, Elodie Gilbault, José M. Jiménez-Gómez, Antoine Martin, Liên Bach, Sébastien Tisné, Leandro Quadrana, Olivier Loudet, and Vincent Colot. 2020. 'Mild Drought in the Vegetative Stage Induces Phenotypic, Gene Expression, and DNA Methylation Plasticity in *Arabidopsis* but No Transgenerational Effects'. *Journal of Experimental Botany* 71 (12): 3588–3602. <https://doi.org/10.1093/jxb/eraa132>.
- Vendrely, R., and C. Vendrely. 1950. 'Sur La Teneur Absolue En Acide Désoxyribonucléique Du Noyau Cellulaire Chez Quelques Espèces d'oiseaux et de Poissons.' *Comptes Rendus de l'Académie Des Sciences*, 230, 788-790.

- Vitte, Clémentine, Margaux-Alison Fustier, Karine Alix, and Maud I. Tenaillon. 2014. 'The Bright Side of Transposons in Crop Evolution'. *Briefings in Functional Genomics* 13 (4): 276–95. <https://doi.org/10.1093/bfgp/elu002>.
- Wallace, Jason G., Peter J. Bradbury, Nengyi Zhang, Yves Gibon, Mark Stitt, and Edward S. Buckler. 2014. 'Association Mapping across Numerous Traits Reveals Patterns of Functional Variation in Maize'. *PLOS Genetics* 10 (12): e1004845. <https://doi.org/10.1371/journal.pgen.1004845>.
- Wang, Qinghua, and Hugo K. Dooner. 2006. 'Remarkable Variation in Maize Genome Structure Inferred from Haplotype Diversity at the Bz Locus'. *Proceedings of the National Academy of Sciences of the United States of America* 103 (47): 17644–49. <https://doi.org/10.1073/pnas.0603080103>.
- Wells, Jonathan N., and Cédric Feschotte. 2020. 'A Field Guide to Eukaryotic Transposable Elements'. *Annual Review of Genetics* 54 (November):539–61. <https://doi.org/10.1146/annurev-genet-040620-022145>.
- Wibowo, Anjar, Claude Becker, Gianpiero Marconi, Julius Durr, Jonathan Price, Jorg Hagmann, Ranjith Papareddy, *et al.* 2016. 'Hyperosmotic Stress Memory in Arabidopsis Is Mediated by Distinct Epigenetically Labile Sites in the Genome and Is Restricted in the Male Germline by DNA Glycosylase Activity'. *eLife* 5 (May):e13546. <https://doi.org/10.7554/eLife.13546>.
- Wicker, Thomas, François Sabot, Aurélie Hua-Van, Jeffrey L. Bennetzen, Pierre Capy, Boulos Chalhouh, Andrew Flavell, *et al.* 2007. 'A Unified Classification System for Eukaryotic Transposable Elements'. *Nature Reviews Genetics* 8 (12): 973–82. <https://doi.org/10.1038/nrg2165>.
- Woodrow, Pasqualina, Giovanni Pontecorvo, Loredana F. Ciarmiello, Amodio Fuggi, and Petronia Carillo. 2011. 'Ttd1a Promoter Is Involved in DNA–Protein Binding by Salt and Light Stresses'. *Molecular Biology Reports* 38 (6): 3787–94. <https://doi.org/10.1007/s11033-010-0494-3>.
- Xu, Gen, Jing Lyu, Qing Li, Han Liu, Dafang Wang, Mei Zhang, Nathan M. Springer, Jeffrey Ross-Ibarra, and Jinliang Yang. 2020. 'Evolutionary and Functional Genomics of DNA Methylation in Maize Domestication and Improvement'. *Nature Communications* 11 (1): 5539. <https://doi.org/10.1038/s41467-020-19333-4>.
- Yang, Qin, Zhi Li, Wenqiang Li, Lixia Ku, Chao Wang, Jianrong Ye, Kun Li, *et al.* 2013. 'CACTA-like Transposable Element in ZmCCT Attenuated Photoperiod Sensitivity and Accelerated the Postdomestication Spread of Maize'. *Proceedings of the National Academy of Sciences of the United States of America* 110 (42): 16969–74. <https://doi.org/10.1073/pnas.1310949110>.
- Zhang, Ming, Yibo Cao, Zhiping Wang, Zhi-Qiang Wang, Junpeng Shi, Xiaoyan Liang, Weibin Song, Qijun Chen, Jinsheng Lai, and Caifu Jiang. 2018. 'A Retrotransposon in an HKT1 Family Sodium Transporter Causes Variation of Leaf Na<sup>+</sup> Exclusion and Salt Tolerance in Maize'. *The New Phytologist* 217 (3): 1161–76. <https://doi.org/10.1111/nph.14882>.

**Titre :** Eléments transposables, dynamique des génomes et des épigénomes : un voyage au cœur de la variabilité structurale et fonctionnelle

**Mots clés :** Eléments transposables, structural variation, évolution, génomes et épigénomes végétaux

**Résumé :** Les progrès de la génomique au cours des 50 dernières années ont permis une description de plus en plus fine du contenu des génomes, de leur organisation et de leur régulation, pour un nombre croissant d'espèces. Ces travaux ont mis en évidence l'ampleur de la fluidité structurale des génomes des plantes, y compris au sein d'une même espèce. Ils ont également révélé la diversité des séquences qui sous-tendent le phénotype, faisant ainsi émerger de nouveaux concepts sur la régulation de l'expression des génomes.

Capables de modifier de façon conjointe la structure et la régulation des génomes, les éléments transposables tiennent une place de premier plan dans leur évolution. D'abord décrits comme une curiosité sans importance, ils sont aujourd'hui considérés comme des contributeurs majeurs à la variabilité phénotypique des individus et à l'évolution des organismes.

Au cours de ma carrière, je me suis intéressée à comprendre comment évolue la structure des génomes, et comment cette évolution sous-tend des variations fonctionnelles, avec un intérêt particulier pour la dynamique des éléments transposables dans ces processus. Après avoir produit des contributions majeures sur la dynamique des éléments transposables et son impact sur l'évolution de la taille des génomes chez les plantes à fleurs, j'ai orienté mes recherches vers des aspects plus fonctionnels, qui me mènent aujourd'hui à étudier la contribution des éléments transposables à l'adaptation et l'évolution des espèces, notamment *via* leur contribution aux séquences *cis*-régulatrices de l'expression des gènes. Mes travaux ont en outre été un levier pour le développement de la génomique et de l'épigénomique du maïs en France, au bénéfice des recherches publiques et privées sur cette espèce.

**Title :** Transposable elements, genome and epigenome dynamics : a journey to the heart of structural and functional variability

**Keywords :** Transposable elements, structural variation, evolution, plants genomes and epigenomes

**Abstract :** Advances in genomics over the last 50 years have led to an increasingly detailed description of genome content, organization and regulation for a growing number of species. This work has highlighted the extent of structural fluidity in plant genomes, even within the same species. It has also revealed the diversity of the sequences that underpin phenotype, leading to the emergence of new concepts in the regulation of genome expression.

Capable of jointly modifying genome structure and regulation, transposable elements play a key role in genome evolution. Initially described as an unimportant curiosity, they are now considered major contributors to the phenotypic variability of individuals and the evolution of organisms.

Throughout my career, I have been interested in understanding how genome structure evolves, and how this evolution underlies functional variation, with a particular interest in the dynamics of transposable elements in these processes. After making major contributions on transposable element dynamics and its impact on the evolution of genome size in flowering plants, I turned my research towards more functional aspects, which today lead me to study the contribution of transposable elements to species adaptation and evolution, notably through their contribution to *cis*-regulatory sequences of gene expression. My work has also been instrumental in the development of maize genomics and epigenomics in France, to the benefit of public and private research on this species.