

**2024-2025**

Première année de Master

Biologie végétale

**ETUDE COMPARATIVE DES SYNTÉNIES POMMIER VS  
POIRIER**  
***synténies des gènes et Éléments transposables***

**JADE LALOUX**

Sous la direction de Jean-Marc Celton

**STAGE RÉALISÉ À L'IRHS (INSTITUT DE RECHERCHE EN HORTICULTURE ET  
SEMENCES)**

**DU 22/04 AU 13/06**

Maître ou maîtresse de stage : Andréa Bouanich

**Jury**

Philippe Simier : président

Jean-Marc Celton : directeur

Didier Peltier : membre

Romain Berruyer : membre



Soutenue publiquement le 01/07/2025

## AVERTISSEMENT

L'université n'entend donner aucune approbation ni improbation aux opinions émises dans les travaux des étudiant·es : ces opinions doivent être considérées comme propres à leurs auteurs.

## ENGAGEMENT DE NON PLAGIAT

*Engagement à signer et à joindre à tous les rapports, dossiers, mémoires ou thèse*

Je, soussigné·e Jade Laloux

déclare être pleinement conscient·e que le plagiat de documents ou d'une partie d'un document publié sur toutes formes de support, numérique ou papier, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée. En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire cette thèse / rapport / mémoire.

signé par l'étudiant·e le 20 / 05 / 2025

A handwritten signature in black ink, consisting of several loops and a long horizontal stroke extending to the right.

## REMERCIEMENTS

Je tiens à remercier particulièrement ma maîtresse de stage Andréa Bouanich qui m'a donnée l'opportunité d'effectuer ce stage, pour le temps qu'elle a pris sur sa thèse pour m'accompagner et me soutenir mais également à sa gentillesse infinie. Je remercie également mon chef d'équipe Jean-Marc Celton pour tous les conseils qu'il m'a donnés, mais surtout pour ses phrases inspirantes. Mon encadrante Christelle Lalanne pour avoir dû corriger tous mes scripts de débutante, ainsi que toute l'équipe VALEMA qui m'a chaleureusement accueillie.

Je remercie grandement Claudine Landès qui n'a pas hésité à venir me donner des conseils dans le peu de temps libre dont elle dispose et pour son éloquence sans nom. Sandra Pelletier pour toutes les formations qu'elle m'a donné mais surtout pour sa patience extraordinaire face à mon manque de connaissance en informatique. Hermann Kamber pour toutes les fois où il a dû venir dans notre bureau pour nous aider.

Merci également aux personnes qui étaient dans mon bureau ou proches, Olena, Marina et Arthur avec qui on a bien ri mais avec qui on s'est aussi plaint de tout et n'importe quoi. Mais surtout de la chaleur étouffante.

## GLOSSAIRE

**\*(Gènes) ohnologues** : Gènes issus de duplications de génome entier (polyploïdisation) survenues au cours de l'évolution. Ils sont conservés dans le génome et partagent un ancêtre commun.

**\*Blocs de synténie** : Ensemble de gènes conservés dans le même ordre entre deux régions de génomes différents, montrant une origine commune.

**\*Eléments transposables (ETs)** : Séquences d'ADN capables de se déplacer dans le génome. Ils peuvent potentiellement modifier l'organisation des gènes et jouer un rôle dans l'évolution.

**\*Packages ou paquets** : En informatique, ensembles de fonctions ou d'outils regroupés dans des bibliothèques logicielles, utilisés pour programmer ou analyser des données.

**\*Versioning** : Gestion des différentes versions d'un fichier ou d'un projet (code, texte, etc.), permettant de suivre les modifications et de revenir à une version précédente si nécessaire.

## LISTE DES ABRÉVIATIONS

**WGD** = Whole genome duplication

***M. domestica*** = *Malus domestica* ( Pommier domestique )

***P. communis*** = *Pyrus communis* ( Poirier commun )

**ETs** = Éléments transposables

## TABLE DES ILLUSTRATIONS

[Figure 1] : Circos du génome de <i>M. domestica</i> .	10
[Figure 2] : schéma représentant les relations entre gènes.	12
[Figure 3] : Schéma chronologique de la collision de l'Inde dans l'Asie	14
[Figure 4] : schéma du déroulement de mon pipeline.	16
[Figure 5] : Extrait de code du "circos.conf".	20
[Figure 6] : Extrait de code du "stats_ET_alignment.py".	22
[Figure 7] : Circos inter-espèce de <i>M. domestica</i> et <i>P. communis</i> .	26
[Figure 8] : Extrait du fichier de sortie du script.	28

## SOMMAIRE

<b>Avertissement</b>	<b>2</b>
<b>Engagement de non plagiat</b>	<b>3</b>
<b>Remerciements</b>	<b>4</b>
<b>Glossaire</b>	<b>5</b>
<b>Liste des abréviations</b>	<b>6</b>
<b>Sommaire</b>	<b>7</b>
<b>Introduction</b>	<b>8</b>
A. Présentation de l'IRHS et de l'équipe VALEMA	8
B. Génome du pommier et du Poirier	8
C. Histoire géologique et évolution des Rosacées	10
D. Objectifs du stage	12
<b>Matériel et Méthode</b>	<b>14</b>
A. Environnement de travail	14
A.1 Disponibilité des scripts et de la méthodologie	14
B. Données génomiques	16
C. Filtrage des blocs de synténies (BdS)	16
D. Analyse de la conservation inter-espèces	16
D.1 BlastP	16
D.2 Correspondances des gènes	17
D.3 Analyse des multiplicons (blocs de synténies)	17
D.4 Visualisation	17
E. Identification des éléments transposables associés aux gènes homologues	18
E.1 Alignement des éléments transposables avec les gènes homologues pommier poirier	18
E.2 Statistiques	18
<b>Résultats</b>	<b>23</b>
A. Analyse des multiplicons (blocs de synténie)	23
B. Organisation globale de la synténie des gènes pommier, poirier	23
C. Alignement des éléments transposables avec les gènes homologues	24
D. Statistiques sur l'alignement des ETs	24
<b>Discussion</b>	<b>26</b>
A. Analyse des multiplicons (BdS)	26
B. Comparaison de la conservation chromosomique	26
C. Statistiques sur l'alignement des ETs	27
D. Discussion littéraire	27



## INTRODUCTION

### A. PRÉSENTATION DE L'IRHS ET DE L'ÉQUIPE VALEMA

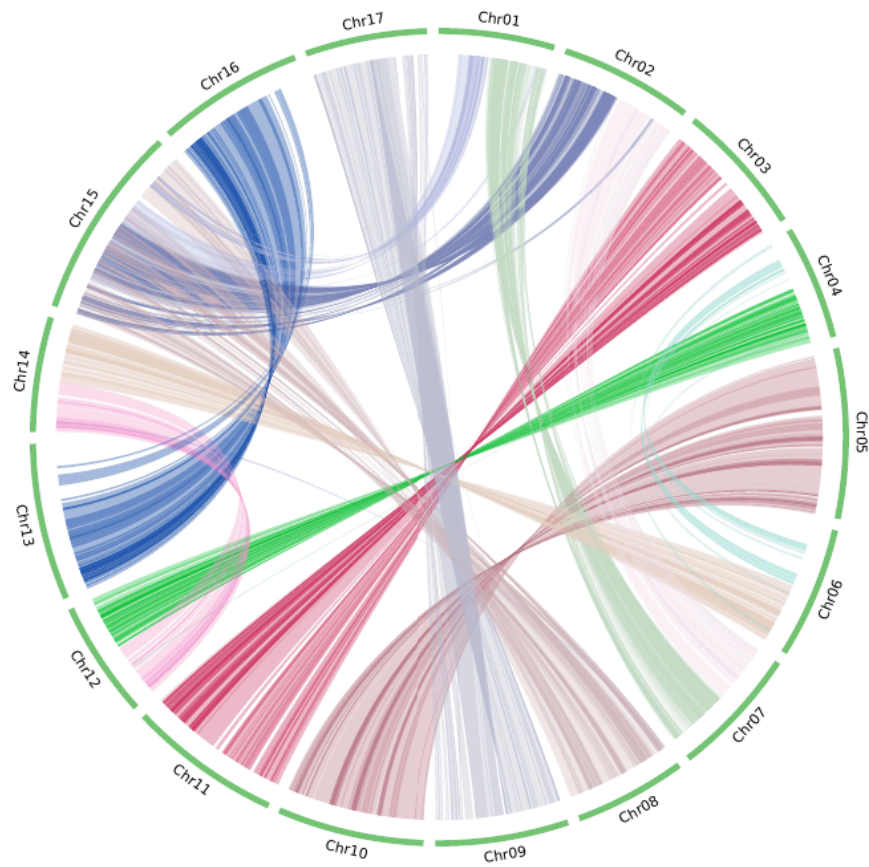
L'IRHS, Institut de Recherche en Horticulture et Semences est une unité mixte de Recherche (UMR 1345) placée sous la tutelle de l'INRAE, de l'institut Agro Rennes-Angers et de l'Université d'Angers. Située à Beaucouzé, cette structure regroupe plus de 250 chercheurs, enseignants-chercheurs, techniciens et ingénieurs. Elle intègre la génétique, la physiologie des plantes, la bioinformatique, la biochimie et la phytopathologie. L'IRHS est un acteur majeur de la recherche en biologie végétale; il œuvre au développement de systèmes de production horticoles durables et performants en réponse aux changements climatiques. Ses travaux visent également à réduire l'utilisation d'intrants, tout en intégrant l'évolution des modes de consommation et les besoins alimentaires croissants liés à l'augmentation de la population.

Au sein de l'IRHS, l'équipe VALEMA (VALorization of Epigenetic Marks in plAnts) dirigée par Jean-Marc Celton depuis sa création en 2019, se spécialise dans l'étude des marques épigénétiques induites par des stress biotiques et abiotiques chez les plantes pérennes. L'objectif final est de développer l'épigénétique comme outil d'amélioration des plantes. L'épigénétique désigne les mécanismes qui modifient l'expression des gènes sans altérer la séquence de l'ADN. Ces mécanismes jouent un rôle clé dans la réponse des plantes aux variations environnementales. La plupart des recherches ont été menées sur des plantes modèles annuelles, laissant un vide de connaissance concernant les plantes pérennes. Pour répondre à ces questions, VALEMA mène plusieurs projets de recherche, dont BIGGER : Etude bioinformatique Intégrative de l'évolution des Génomes et epiGEnomes de Rosaceae, projet dans lequel s'inscrit mon stage.

### B. GÉNOME DU POMMIER ET DU POIRIER

Le Pommier (*Malus domestica*) et le Poirier (*Pyrus communis*) appartiennent à la famille des Rosacées et sont des arbres fruitiers d'importance agronomique majeure dans le monde entier ([Fruit-June-2023.pdf](#), s. d.). Ces deux espèces partagent un ancêtre commun ayant subi plusieurs duplications complètes du génome, appelées WGD (Whole Genome Duplication) ([Lallemant et al., 2023](#)). La duplication du génome entier est un processus biologique récurrent dans l'évolution des espèces qui, face à des changements environnementaux drastiques ([Van De Peer et al., 2021](#)), favorise l'émergence de nouvelles fonctions génétiques, augmentant ainsi les chances de survie des espèces ([De Storme & Mason, 2014](#); [Soltis & Soltis, 2016](#); [Van De Peer et al., 2021](#)). Le génome entier se retrouve doublé voire triplé au sein même de l'espèce. Une WGD peut aboutir à la duplication d'un génome déjà diploïde, comme c'est le cas chez l'ancêtre de *M. domestica* et *P. communis*, on parle alors d'autopolyploïdie ([De Storme & Mason, 2014](#); [Lallemant et al., 2023](#)). Le génome originellement composé de neuf chromosomes ( $2n=18$ ,  $n=9$ ) s'est intégralement dupliqué et réarrangé pour donner dix-sept chromosomes ( $2n=34$ ,  $n=17$ ) ([Daccord et al., 2017](#); [Pont et al., 2019](#); [Velasco et al., 2010](#)).

Chez l'ancêtre commun du pommier et du poirier, la WGD la plus récente est datée à 27 millions d'années ([Lallemant, 2022](#)), elle s'observe encore très bien dans le génome du pommier moderne *M. domestica*. Notamment grâce aux travaux de séquençage *de novo* réalisés sur la lignée haploïde doublée "Golden Delicious" n°13, ou GDDH13 ([Daccord et al., 2017](#)). Ce génotype a été choisi en raison de sa stabilité génétique et de sa faible hétérozygotie. Ce qui facilite grandement l'assemblage du génome et permet d'obtenir une séquence de référence de haute qualité ([Pryszcz & Gabaldón, 2016](#)).



[Figure 1] : Circos du génome de *M. domestica* Les chromosomes sont représentés en vert tout autour, les lignes colorées représentent la synténie des gènes. Un lien coloré représente donc un \*gène ohnologue. Ainsi on distingue nettement la conservation de l'ordre des gènes entre les chromosomes dupliqués.

L'analyse du génome de GDDH13 a révélé une forte synténie entre chromosomes. Autrement dit, la conservation de l'ordre des gènes sur des segments du génome. Ce phénomène peut être observé entre deux espèces différentes, ou bien entre deux régions d'un même génome issu d'une duplication complète (i.e., une WGD). Lorsqu'un ensemble de gènes, un segment de chromosome conservent leur position relative, on les appelle \*blocs de synténie. Ces blocs constituent une trace visible des duplications anciennes du génome, comme chez l'ancêtre du pommier et du poirier [fig 1].

Cette observation est le reflet direct de la WGD [fig 1], (Daccord et al., 2017). Les gènes présents dans les régions synténiques sont appelés \*ohnologues, un sous-type de gènes homologues, provenant spécifiquement d'une WGD (Wolfe, 2000).

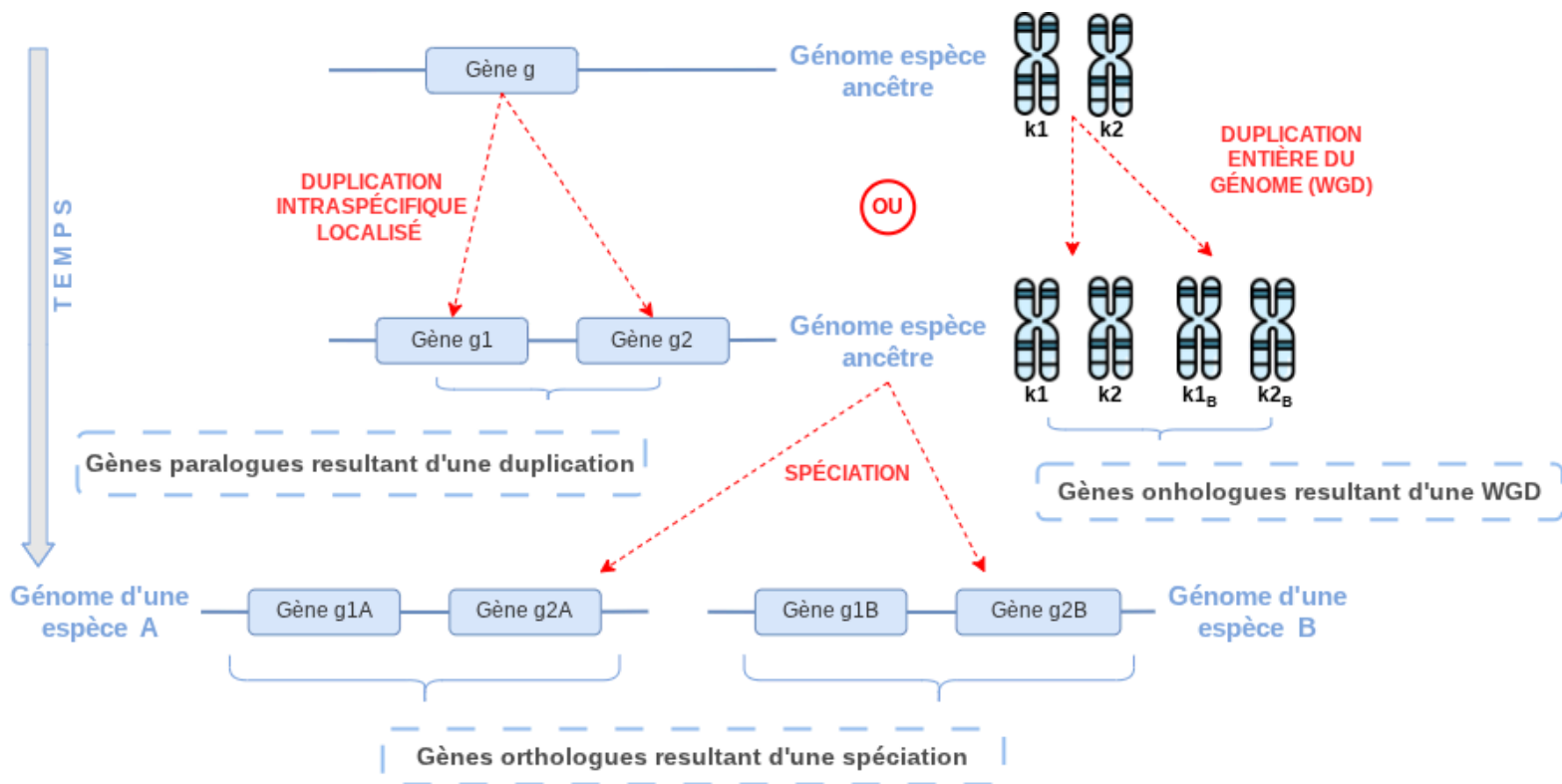
Le terme "\*ohnologue" a été proposé par Kenneth H. Wolfe en 2000, en hommage à Susumu Ohno, qui a introduit l'idée que les duplications complètes du génome jouent un rôle majeur dans l'évolution des vertébrés. Les gènes \*ohnologues se distinguent des autres types de gènes dupliqués (i.e., paralogues) [fig 2] de par leur origine : ils résultent d'une duplication de l'ensemble du génome, plutôt que d'une duplication locale ou segmentaire [fig 2] (Ohno, 1970; Wolfe, 2000).

Des travaux similaires ont déjà été menés sur le poirier, *P. communis*, suggérant une organisation génomique globalement comparable à celle du pommier. Ils mettent notamment en évidence la conservation de \*blocs de synténie et la présence \*d'ohnologues résultant d'une même WGD ancestrale (Li et al., 2019). Toutefois, d'autres recherches sont actuellement en cours, dans le cadre d'une thèse (Andréa Bouanich, données non publiées) avec pour objectif de réanalyser le génome du poirier à l'aide des mêmes outils que ceux utilisés pour *M. domestica*. Cette approche vise à limiter les biais méthodologiques et à s'assurer que les différences observées proviennent bien des espèces elles-mêmes, et non des variations dans les outils utilisés.

## C. HISTOIRE GÉOLOGIQUE ET ÉVOLUTION DES ROSACÉES

L'ancêtre commun du pommier et du poirier situé dans les régions de l'actuel Kazakhstan, a subi il y a environ 27 Ma sa plus récente duplication entière du génome (Lallemant et al., 2023). Cette WGD coïncide temporellement avec des bouleversements géologiques tels que la collision des plaques indienne et eurasienne, amorcée il y a 35 Ma (Vance & Harris, 1999) [fig 3]. Cette convergence tectonique a progressivement engendré l'élévation des chaînes de montagnes, telles que l'Himalaya et, plus au nord, le Tian Shan (Aitchison et al., 2007). L'ancêtre de *Malus* et *Pyrus*, initialement présent dans les vallées, a ainsi été progressivement soumis à une élévation en altitude (Vance & Harris, 1999). Cette élévation progressive a engendré des modifications environnementales importantes : diminution des températures et de la pression atmosphérique, augmentation de l'irradiation UV et apparition de nouvelles contraintes écologiques (Kumar, 1999). Par ailleurs, l'élévation de l'Himalaya a provoqué des variations climatiques sur l'ensemble du continent asiatique, notamment en amplifiant la mousson (Wang et al., 2012). Ces changements abiotiques drastiques ont constitué une pression de sélection non négligeable pour les végétaux locaux, favorisant des réorganisations génétiques et épigénétiques.

Dans ce contexte de stress intense, un "burst" \*d'éléments transposables (ETs) a été détecté dans le génome de *M. domestica* il y a environ 21 Ma (Daccord et al., 2017). Les ETs sont des \*éléments capables de se déplacer et se multiplier dans le génome. Ils sont connus pour jouer un rôle central dans la dynamique évolutive i.e., ils modifient l'architecture génomique, régulent l'expression des gènes et participent à l'émergence de nouvelles fonctions. Leur activation est fréquemment associée à des périodes de stress environnemental ou à des instabilités post-WGD (Ito, 2022).



[Figure 2] : schéma représentant les relations entre gènes. Inspiré d'une figure de l'article de [\(Tekaia, 2016\)](#) et de [\(Thomas Denecker, s. d.\)](#). Les carrés bleus représentent des gènes. *k* = chromosome.

A ce jour, aucun burst comparable n'a été mis en évidence dans le génome de *Pyrus communis*. Si un tel phénomène était mis en évidence, cela pourrait suggérer un rôle important des ETs dans la divergence entre les genres *Malus* et *Pyrus*. Mais possiblement aussi dans d'autres séparations de genre ou spéciations. Cette hypothèse renforcerait l'idée que les \*éléments transposables pourraient être des acteurs de la diversité épi/génétique et phénotypique. Ils pourraient donc être utiles pour la résilience des plantes lors de stress important, notamment en vue du dérèglement climatique actuel.

## D. OBJECTIFS DU STAGE

Ce stage s'inscrit dans le cadre du projet BIGGER (Etude bioinformatique Intégrative de l'évolution des Génomes et epiGENomes de Rosaceae), et vise à étudier l'histoire évolutive des \*éléments transposables chez deux espèces de Rosacées fruitières majeures : *Malus domestica* et *Pyrus communis*.

L'objectif principal de ce stage est de mieux comprendre si les \*éléments transposables ont joué un rôle dans la divergence des genres *Malus* et *Pyrus*. Plus concrètement, il s'agit d'explorer si un burst \*d'éléments transposables, observé chez le pommier, a également eu lieu chez le poirier. Mais également si des différences de nature ou de répartition des \*éléments transposables pourraient expliquer leur divergence phénotypique.

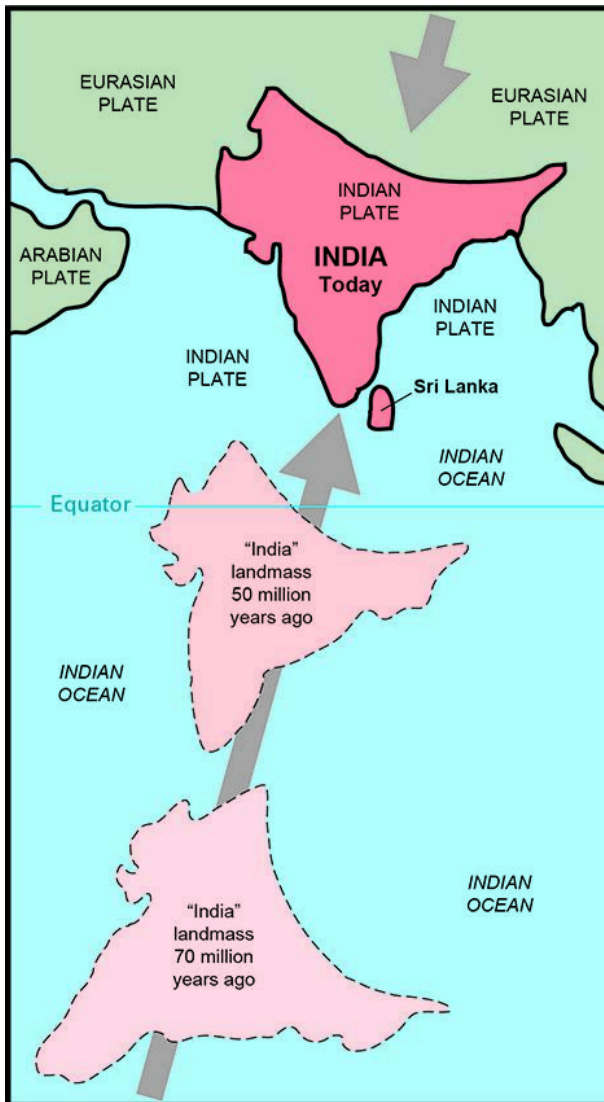
Pour cela, mon travail repose sur deux jeux de données : des annotations en \*éléments transposables des lignées haploïdes doublées de pommier et de poirier, et des \*blocs de synténie de gènes issues de leur dernière duplication complète du génome.

Ma première mission est de comparer l'organisation des gènes dans les \*blocs de synténie partagés entre les deux espèces, afin de vérifier si la synténie génique est bien conservée. Cela permet d'obtenir des régions comparables, qui serviront de base pour la suite des analyses. Puis de créer un visuel (Circos) ([Krzywinski et al., 2009](#)) de l'organisation de ces gènes. La deuxième étape consiste à analyser la répartition des \*éléments transposables au sein de ces mêmes blocs, pour voir si leur présence ou leur organisation diffère entre les deux génomes.

Ces analyses ont été réalisées à l'aide du langage de programmation Python, en suivant les recommandations FAIR (Findable Accessible Interoperable Reusable) :

- Un **environnement conda** reproductible, afin que la pipeline soit réutilisable par autrui.
- Une mise à disposition sur **GitLab** afin de garantir la traçabilité des codes et leur réutilisation au sein de l'équipe.

Ce stage permettra donc de développer des compétences en programmation, en gestion de données génomiques de grandes tailles et en analyse statistique, tout en contribuant à une meilleure compréhension des mécanismes de diversification génétique et épigénétique. Ces connaissances pourraient à terme enrichir les stratégies d'amélioration variétale en contexte de changement climatique.



[Figure 3] : Schéma chronologique de la collision de l'Inde dans l'Asie tiré de ([The Himalayas \[This Dynamic Earth, USGS\], s. d.](#)). Il y a 70 Ma le mouvement de l'Inde a été amorcé, le début de la collision aurait débuté il y a environ 35 Ma.

## MATÉRIEL ET MÉTHODE

### A. ENVIRONNEMENT DE TRAVAIL

Toutes les analyses ont été réalisées dans un environnement Linux, sur le cluster nommé SLURM de l'IRHS. SLURM permet la gestion efficace de tâches informatiques lourdes et en parallèle. Un environnement isolé que j'ai intitulé "applePear", a été mis en place à l'aide de Conda. Conda est un gestionnaire de paquets (\*packages) et d'environnements dédiés à la bioinformatique, afin d'assurer la reproductibilité des analyses. Cela permet de figer les versions des outils utilisés et de garantir que l'ensemble des scripts fonctionnent à l'identique sur différentes machines. Mais aussi dans un futur proche ou lointain, même lorsque les versions des outils auront évolué.

Le code a été documenté, versionné et partagé via un dépôt privé GitLab, outil de suivi de version (\*versionning), garantissant la traçabilité des modifications et favorisant le travail collaboratif. Ce dépôt respecte les principes FAIR (Findable, Accessible, Interoperable, Reusable), pour que les résultats soient reproductibles et que les pipelines soient réutilisables par d'autres membres de l'équipe.

L'ensemble des scripts ont été écrits en Python (version 3.12), à l'aide de bibliothèques standard comme pandas, csv, os etc. Ils ont été développés en suivant plusieurs bonnes pratiques de programmation :

- **Utilisation d'un shebang** (`#!/usr/bin/env python`) en première ligne de chaque script. Cette ligne permet d'indiquer à l'interpréteur quel langage utiliser pour exécuter le code.
- **Ajout d'une fonction principale** (`if __name__ == "__main__":`) garantissant une meilleure lisibilité et modularité du code.
- **Utilisation d'argparse** : cette bibliothèque permet de gérer les arguments passés en ligne de commande. Elle facilite l'exécution flexible des scripts en spécifiant les fichiers d'entrée, de sortie, ou d'autres paramètres directement depuis le terminal, tout en assurant un affichage clair pour l'utilisateur.
- **Gestion des erreurs** avec des blocs `try/except` : ces instructions permettent de capturer et traiter proprement les erreurs potentielles (e.g., des problèmes de lecture de fichiers), améliorant ainsi la robustesse du code.
- **Commentaires et documentation** : chaque script est documenté pour expliciter son fonctionnement, ses entrées et sorties, ce qui facilite sa réalisation et sa compréhension par d'autres utilisateurs.

#### A.1 Disponibilité des scripts et de la méthodologie

Tous les scripts utilisés pour le traitement de données et les analyses sont disponibles sur le dépôt GitHub suivant :

- <https://forge.inrae.fr/evolution-maleae/apple-pear-analysis>

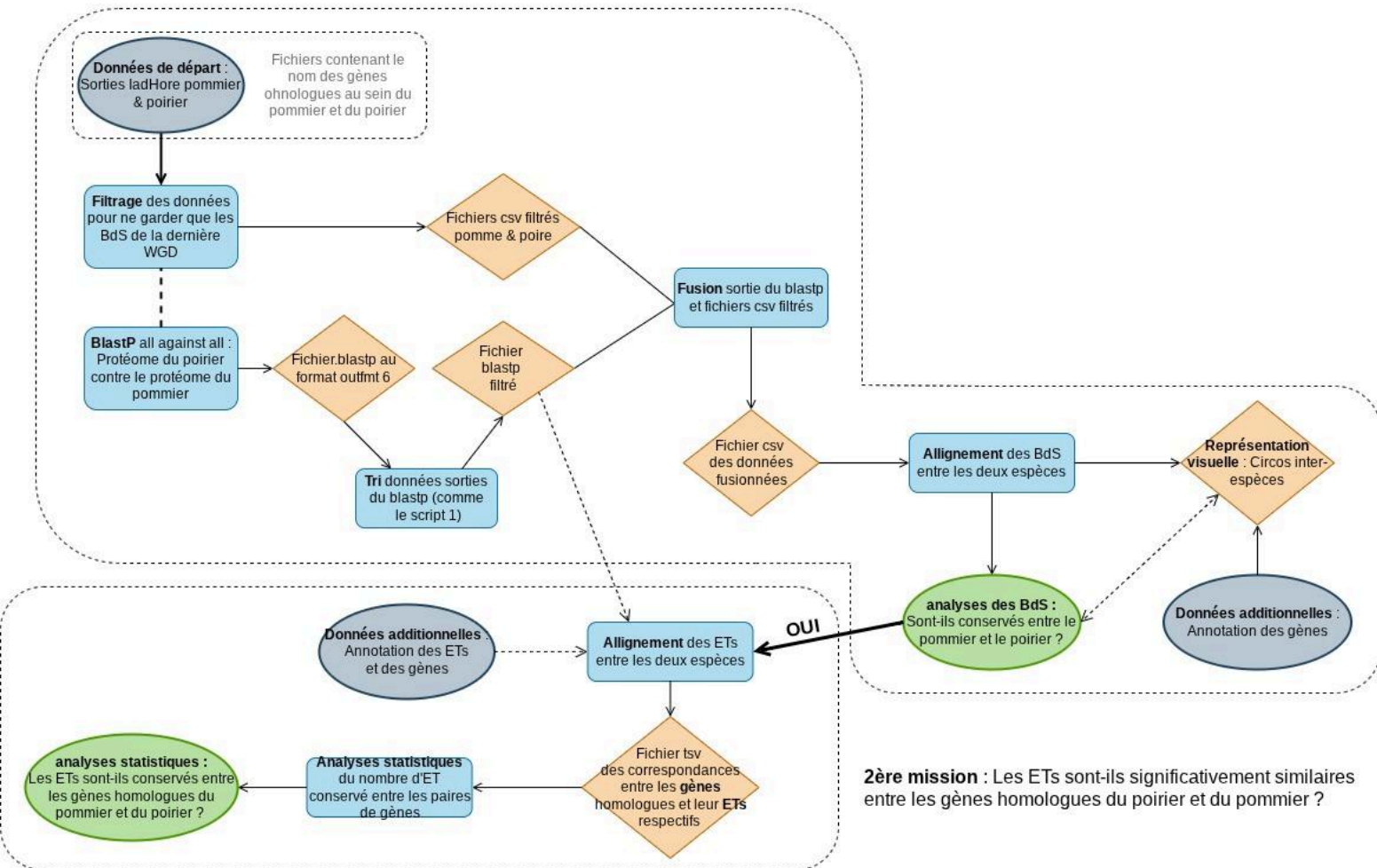
Chaque étape décrite ci-dessous est associée à un script spécifique, indiqué souligné à la fin de chaque section.

Afin de faciliter la compréhension et la reproductibilité de l'ensemble des étapes réalisées, un schéma détaillé du pipeline d'analyse a également été réalisé [fig ].



## Déroulement du pipeline

**1ère mission : Les gènes des deux espèces sont-ils conservés ?**



[Figure 4] : schéma du déroulement de mon pipeline. Les carrés bleus représentent les actions effectuées par les scripts. Les losanges oranges représentent les fichiers de sortie, les résultats des scripts. Les cercles verts représentent les questions de fin de mes missions. Les cercles gris représentent les données additionnelles que je n'ai pas produites mais utilisées.



## B. DONNÉES GÉNOMIQUES

Les données d'entrées sont constituées :

- Des sorties du logiciel i-ADHoRe 3.0, réalisées à partir des \*protéomes déduits des deux doubles haploïdes *Malus domestica* (GDDH13) et *Pyrus communis* (Bartlett v2.0).

i-ADHoRe (Iterative Automatic Detection of Homologous Regions) ([Proost et al., 2011](#)) est un programme qui permet d'identifier des blocs de gènes conservés entre différentes espèces ou au sein de d'une même espèce. Ces blocs de gènes dans le même ordre présents sur au moins deux chromosomes ou régions du génome sont ce qu'on appelle les \*blocs de synténie.

- Des **fichiers d'annotations des gènes** \*ohnologues pour chaque espèce.
- Mais également des **fichiers d'annotation des ETs** pour chaque espèce, celle du pommier est disponible publiquement et celle du poirier est fournie par l'équipe VALEMA.
- D'un **fichier de protéome** pour chaque espèce, utilisé pour les analyses de similarité.

## C. FILTRAGE DES BLOCS DE SYNTÉNIES (BdS)

Afin de ne conserver que les \*blocs de synténie résultant de la dernière WGD commune aux deux genres, les fichiers de sortie i-ADHoRe ont été filtrés selon les paires de chromosomes. Les paires de chromosomes correspondant à la dernière WGD ont été identifiées dans l'article de ([Daccord et al., 2017](#)) : 1-7, 1-15, 2-7, 2-15, 3-11, 4-6, 4-12, 5-10, 6-14, 8-15, 9-17, 12-14, 13-16. a noter qu'il existe des demi paires de chromosomes comme 1-7 et 2-7. Cela signifie qu'une partie du chromosome 7 correspond à une partie du chromosome 1 et que l'autre partie du chromosome 7 correspond au chromosome 2 [fig 1]. Le filtrage de ces paires de chromosomes sert à effacer les paires de chromosomes pouvant provenir de WGD plus anciennes et qui pourraient ne plus être présent dans un des deux génomes, et donc ne garder que du matériel comparable entre les deux espèces. Les deux tableaux de données finaux sont sauvegardés au format .csv.

script : [ohnologous k pairs sorting.py](#)

## D. ANALYSE DE LA CONSERVATION INTER-ESPÈCES

### D.1 BlastP

Un alignement croisé des gènes des blocs synténiques a ensuite été effectué via un BlastP all-against-all, avec le protéome du poirier contre celui du pommier. C'est-à-dire que le protéome du pommier est la séquence de base sur lequel est aligné le protéome du poirier.

Le programme BlastP a été utilisé avec le paramètre outfmt 6 (output format) afin d'obtenir une sortie de fichier tabulaire standard. Il a également été demandé une e-value à 1e-5, afin d'écarter les correspondances faibles tout en conservant les alignements significatifs attendus entre deux espèces proches. Afin de conserver uniquement les correspondances les plus fiables, seul le meilleur alignement pour chaque correspondance (RBH, Reciprocal Best Hits) ont été retenus (max\_target\_seqs 1). Aucun filtrage supplémentaire sur le pourcentage d'identité ou sur la couverture de l'alignement n'a été appliqué.

script : [submit\\_blastdb.sh](#)

Le fichier de sortie .blastp est ensuite trié exactement comme les fichiers de sortie i-ADHoRe afin de ne garder que les paires de chromosomes dérivant de la WGD la plus récente. Cependant, seules les correspondances dans un sens (poirier vers pommier) ont été conservées à cette étape, ce qui a pu conduire dans la suite des analyses à la perte de certaines correspondances réciproques potentielles.



Un exemple, les lignes avec les gènes du chromosome 1 poirier associés aux gènes du chromosome 7 pommier ont été conservés. Cependant les gènes du chromosome 7 poirier associés aux gènes du chromosome 1 pommier n'ont pas été conservés.

script : [blastp\\_pear\\_vs\\_GDDH13\\_results\\_sorting.py](#)

## D.2 Correspondances des gènes

Le fichier **BlastP** filtré ainsi que les fichiers de **sortie i-ADHoRe** également filtrés ont ensuite été **fusionnés** afin d'identifier les correspondances géniques entre les deux espèces dans les régions synténiques.

script : [synthenic\\_blocs\\_comparison\\_apple\\_pear.py](#)

Ce fichier fusionné a lui aussi été filtré, pour **effacer** tous les **duplicatas** entiers, qui fausseraient les résultats. Ce fichier fusionné a également servi de base pour la génération des visualisations circos ([Krzywinski et al., 2009](#)), permettant une représentation graphique des correspondances de gènes entre les deux génomes.

script : [duplicates\\_verification\\_apple\\_pear.py](#)

## D.3 Analyse des multiplicons (blocs de synténies)

Afin de quantifier précisément la conservation des \*blocs de synténie, aussi appelés "multiplicons" entre le pommier et le poirier, un script a été développé. Ce script permet de déterminer, pour chaque \*bloc de synténie identifié dans l'une des deux espèces, à combien de blocs différents il correspond dans l'autre espèce. Cette analyse permet d'évaluer le niveau de fragmentation ou de conservation des blocs entre les deux génomes. Le script prend en entrée le fichier créé dans la partie "D.2 correspondances des gènes". Il contient, pour chaque gène correspondant, l'identifiant du \*bloc de synténie auquel il appartient dans le poirier et le pommier.

Le script calcul combien de multiplicons différents du pommier sont associés à chaque multiplicon du poirier, et inversement. Autrement dit, pour un multiplicon chez le pommier à combien de multiplicons cela correspond chez le poirier, et vis versa.

L'objectif est de détecter les **relations** entre les \*blocs de synténie identifiés :

- **(Relation 1:1)**: Un multiplicon chez le poirier correspond à un unique multiplicon chez le pommier. Cette situation traduit une conservation stricte de la structure génomique entre les deux espèces dans la région concernée ([Xue et al., 2018](#)).
- **(Relation 1:n et n:1)**: Un multiplicon chez le poirier (ou pommier) correspond à plusieurs multiplicons différents chez le pommier (ou poirier). Cette situation révèle des événements de fragmentation, de réarrangement ou de divergence post-duplication entre les deux génomes ([Gutierrez-Gonzalez & Garvin, 2011](#)).

Ce script génère un fichier de sortie textuel contenant :

- Le détail du **nombre de multiplicons** correspondants pour chaque bloc
- Des **statistiques globales** (moyennes, médiane, maximum) sur ces relations de correspondances

script : [multiplicons\\_analysis.py](#)

## D.4 Visualisation

Les \*blocs de synténie communs entre *P. communis* et *M. domestica* ont été visualisés via une représentation de type Circos [fig 1] ([Krzywinski et al., 2009](#)), (version 0.69-9). Les fichiers utilisés comme entrée pour créer le Circos ont été générés à partir des résultats du BLASTP (filtrés) des deux genres ainsi que les fichiers d'annotation des gènes. Les résultats BLASTP servent à définir les synténies entre les deux espèces afin de créer des liens entre les chromosomes des deux genres, représentés par des lignes

```
552 <rule>
553 |   condition = between(AChr13,PChr13)
554 |   color     = E3936F
555 |   thickness = 4
556 |   z         = 15
557 </rule>
558
559 <rule>
560 |   condition = between(PChr13,AChr13)
561 |   color     = E3936F
562 |   thickness = 4
563 |   z         = 15
564 </rule>
565
566 <rule>
567 |   condition = between(AChr14,PChr14)
568 |   color     = 714CE6
569 |   thickness = 4
570 |   z         = 15
571 </rule>
```

[Figure 5] : Extrait de code du "[circos.conf](#)" représentant le choix des couleurs et de l'épaisseur pour les liens entre le chromosome 13 du pommier (AChr13) vers le chromosome 13 du poirier (PChr13). Mais également du chromosome 13 du poirier vers le chromosome 13 du pommier. et enfin du chromosome 14 du pommier vers le chromosome 14 du poirier.

colorées dans le circos. Les fichiers d'annotation de gènes servent à définir les positions de début et de fin des gènes sur les chromosomes respectifs.

Les chromosomes ont été disposés dans l'ordre numérique pour chaque espèce de manière à faciliter l'interprétation globale des synténies interchromosomiques. La même configuration Circos que celui de la thèse de Tanguy Lallemant ([Lallemant, 2022](#)) a été utilisée à l'exception du nombre maximal de liens affichés (`max_links`). Le nombre maximal de liens est normalement de 25000, ce qui ne suffit pas pour la représentation entre les deux genres, composé d'environ 32000 liens. La représentation obtenue permet d'observer les relations globales de colinéarité entre les chromosomes des deux espèces.

L'erreur dans les fichiers précédents, l'absence de certaines correspondances (dans un seul sens) entre les paires de gènes homologues, a été identifiée et corrigée à cette étape. Le stage étant déjà arrivé à sa fin au moment de l'identification de l'erreur, je n'ai pas pu corriger les autres scripts. Le visuel Circos a été refait par ma maîtresse de stage Andréa Bouanich pour que je puisse avoir un résultat correct. Ainsi, la représentation obtenue reflète correctement les relations globales de colinéarité entre les chromosomes des deux espèces.

plusieurs scripts ont dû être fait pour la configuration du circos : [circos](#).

La majeure partie du travail a été effectué dans ce script : [circos.conf](#) [fig 5]

## E. IDENTIFICATION DES ÉLÉMENTS TRANSPOSABLES ASSOCIÉS AUX GÈNES HOMOLOGUES

### E.1 Alignement des éléments transposables avec les gènes homologues pommier poirier

L'objectif de cette analyse est d'identifier et de comparer les \*éléments transposables associés aux gènes homologues entre le pommier et le poirier. L'association d'un ET à un gène homologue est définie par un chevauchement direct avec le gène ou une proximité dans une zone de 5 kb (kilobases) en amont ou en aval. Ce seuil est ajustable via un paramètre du script utilisé. Ce choix repose sur la valeur médiane de l'espace intergénique observée dans ces génomes, estimée à environ 3,6 kb. En sélectionnant un seuil légèrement supérieur (5 kb), il devient possible de capturer non seulement les ETs intragéniques (présents au sein des séquences codantes) mais également ceux localisés dans les régions proches susceptibles d'influencer, de réguler l'expression des gènes. Ce compromis permet donc d'équilibrer entre spécificité et sensibilité de détection des ETs.

Données utilisées :

- **Paires de gènes homologues** : obtenues à partir des résultats filtrés du BLASTP (cf. partie D.1), indiquant pour chaque gène du poirier son gène homologue chez le pommier, afin de pouvoir y associer leur ETs respectifs.
- **Annotations des gènes** : fichiers GFF3 des deux espèces, permettant de récupérer les coordonnées précises des gènes et donc de pouvoir indiquer dans l'ordre les ETs sur chaque gène.
- **Annotation des éléments transposables** :
  - Pour *M. domestica* : annotations disponibles sur REPETdb, base de données de l'outil de l'annotation
  - Pour *P. communis* : annotations fournies par l'équipe VALEMA

Le script python développé permet de croiser les coordonnées des gènes homologues avec celles des ETs correspondants dans chaque espèce. Chaque paire de gènes homologues est alors associée aux ETs trouvés chez le pommier d'un côté et ceux trouvés chez le poirier de l'autre.

```
stats = {
    "total_paires": total_pairs,
    "0_ET": format_count_and_percentage(counts.get(0, 0)),
    "1_ET": format_count_and_percentage(counts.get(1, 0)),
    "2_ET": format_count_and_percentage(counts.get(2, 0)),
    "3_ET": format_count_and_percentage(counts.get(3, 0)),
    "4_ET": format_count_and_percentage(counts.get(4, 0)),
    "5+_ET": format_count_and_percentage(sum(v for k, v in counts.items() if k >= 5)),
    "%_paires_avec_≥1_ET": f"{round((total_pairs - counts.get(0,0)) / total_pairs * 100, 2)}%",
    "moyenne_ETs_communs_par_paire": round(df['nb_ETs_communs'].mean(), 2),
    "médiane_ETs_communs_par_paire": int(df['nb_ETs_communs'].median()),
    "max_ETs_communs_par_paire": int(df['nb_ETs_communs'].max())
}
```

[Figure 6] : Extrait de code du "stats\_ET\_alignment.py" représentant les lignes qui donnent le résultat en [figure 8].

Le fichier final généré contient, pour chaque paire de gènes homologues, les codes des \*éléments transposables associés. Ces codes correspondent à une classification selon l'article de Wicker ([Wicker et al., 2007](#)), standard international pour la catégorisation des ETs en classes, ordres et superfamilles. Cette classification permet une interprétation fonctionnelle, par exemple pour distinguer les rétrotransposons (e.g., code RLC pour LTR/Copia, code RLG pour LTR/Gypsy) des transposons à ADN (code DTA pour DNA/TIR).

Le but est de déterminer si les ETs sont les mêmes entre paires de gènes homologues, et donc s'ils ont été conservés entre les deux espèces au cours du temps.

script : [ETs\\_alignment\\_apple\\_pear.py](#)

## E.2 Statistiques

Pour quantifier l'association entre les \*éléments transposables et les gènes homologues entre le pommier et le poirier, plusieurs analyses statistiques ont été effectuées. Ces statistiques ont été établies à partir du fichier de sortie du script précédent, le fichier des correspondances gènes homologues ETs (cf E.2).

Ce script calcul :

- La répartition du nombre total d'ETs par paire de gènes homologues, en regroupant ces paires selon le nombre d'ETs détectés (0, 1, 2, 3, 4,  $\geq 5$ ).
- La proportion de paires contenant au moins un ET.
- Des statistiques descriptives sur le nombre d'ETs communs par paire de gènes homologues entre le pommier et le poirier (moyenne, médiane et maximum).

[\[fig 6\]](#)

Ces statistiques permettent d'évaluer le niveau de co-occurrence des ETs entre gènes homologues inter-espèces, i.e., à quelle fréquence les mêmes ETs sont présents autour des gènes homologues chez les deux espèces. Ainsi que la variabilité de ces associations. Elles offrent un aperçu global de l'enrichissement des gènes homologues en \*éléments transposables, et permettent de quantifier la conservation éventuelle de ces \*éléments entre le poirier et le pommier.

script : ["stats\\_ET\\_alignment.py"](#) [\[fig 6\]](#)





## RÉSULTATS

### A. ANALYSE DES MULTIPLICONS (BLOCS DE SYNTÉNIE)

L'analyse des correspondances entre les \*blocs de synténie (multiplicons) de *M. domestica* et *P. communis* a permis de quantifier le degré de conservation et de réarrangement entre les deux génomes. Au total, 227 multiplicons ont été identifiés chez le pommier contre 131 multiplicons chez le poirier, dans les régions issues de leur dernière duplication entière du génome commune.

Le calcul des correspondances montre une prédominance des relations 1:n : plusieurs multiplicons chez l'une des espèces correspondent à un seul multiplicon chez l'autre. La moyenne est de 3,53 multiplicons du pommier par multiplicons du poirier, tandis que chaque multiplicon du pommier correspond en moyenne à 2,04 multiplicons du poirier. Ces résultats reflètent une dynamique asymétrique entre les deux genres.

Les médianes, quant à elles, sont égales à 2, et ce, pour les deux espèces. Cela indique que pour la moitié des multiplicons considérés, deux correspondances inter-espèces sont observées. Cette symétrie des médianes traduit une situation relativement équilibrée, malgré des cas extrêmes de fragmentation ou fusion observés aux extrémités des distributions.

Quelques exemples illustrent ces observations :

- Le **multiplicon n°800** du pommier correspond à **1 seul multiplicon** chez le poirier (relation 1:1), traduisant une **conservation** locale **stricte**.
- Les **multiplicons n°856, 861 et 864** du pommier correspondent chacun à **2 multiplicons différents** du poirier (relation 1:n), signe d'une **fragmentation** modérée chez le poirier, ou bien d'une **fusion** chez le pommier.

Le cas le plus extrême observé correspond à 33 multiplicons du pommier associés à un même multiplicon du poirier, traduisant des remaniements plus importants (relation n:1).

Ces résultats indiquent donc une prévalence des relations multiples (1:n ou n:1) entre les deux génomes, suggérant des réarrangements structuraux post-duplication, tels que des fragmentations ou des inversions, fréquents dans l'évolution des espèces pérennes à duplication ancienne ([Gutierrez-Gonzalez & Garvin, 2011](#); [Xue et al., 2018](#)).

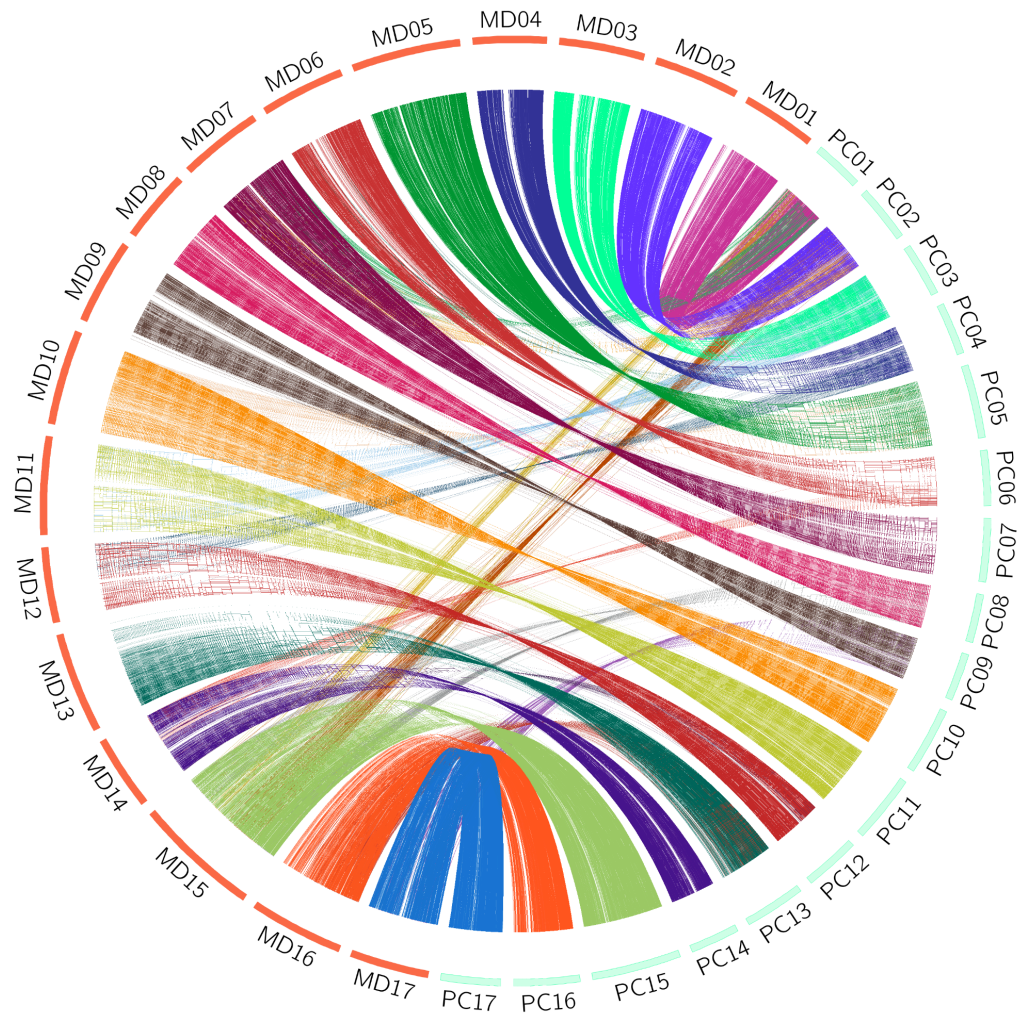
Il est important de noter que l'analyse présentée ici repose uniquement sur une partie des données disponibles. En effet seules les correspondances dans un seul sens (ex : 1→7) ont été considérées, tandis que les correspondances dans le sens inverse (7→1) n'ont pas été intégrées. Cette absence entraîne une sous-estimation potentielle du nombre total de relations entre multiplicons. Toutefois, malgré cette limitation, les tendances observées (prévalence des relations 1:n, valeurs médianes à 2) restent cohérentes avec ce qui est attendu dans un contexte de réarrangements post-WGD.

### B. ORGANISATION GLOBALE DE LA SYNTÉNIE DES GÈNES POMMIER, POIRIER

Le représentation Circos obtenue illustre les synténies interchromosomiques entre *M. domestica* et *P. communis*. Les correspondances homologues entre les gènes des deux espèces sont représentées par des liens colorés. [fig 7]

Le visuel met en évidence une forte correspondance entre les chromosomes homologues des deux espèces. Par exemple, le chromosome 1 du poirier présente de nombreux liens directs avec le chromosome 1 du pommier, et cette correspondance claire se retrouve pour chaque chromosome numéroté. Cette organisation suggère une conservation globale de la structure chromosomique entre *M. domestica* et *P. communis*.

Par ailleurs, les relations intrachromosomiques déjà observées lors de l'analyse spécifique à chaque espèce sont également visibles sur cette représentation inter-espèces.



[Figure 7] : Circos inter-espèce de *M. domestica* et *P. communis*. à l'extérieur les traits rouges représentent les chromosomes du pommier et en bleu/vert les chromosomes du poirier. Les liens représentent les gènes homologues entre le pommier et le poirier.

À titre d'exemple, la correspondance entre les chromosomes 4 et 12 observée chez le pommier se retrouve également entre les deux espèces dans le circos inter-espèce. Ces relations complexes, bien que parfois moins visibles en raison du nombre élevé de liens affichés restent néanmoins perceptibles. [fig 7]

Enfin, l'erreur identifiée dans les fichiers précédents, a été corrigée à cette étape, ce qui garantit que la représentation Circos obtenue reflète fidèlement les relations globales de colinéarité inter-espèces.

### C. ALIGNEMENT DES ÉLÉMENTS TRANSPOSABLES AVEC LES GÈNES HOMOLOGUES

L'alignement des \*éléments transposables avec les gènes homologues a été effectué sur un total de 29610 paires de gènes homologues identifiées entre le poirier et le pommier. Pour chaque paire, les ETs présents sur les gènes homologues respectifs ont été extraits, en considérant les chevauchements directs ainsi qu'une zone de 5 kb en amont et en aval des gènes.

Les \*éléments transposables détectés sont classifiés selon la nomenclature de (Wicker et al., 2007), permettant d'obtenir, pour chaque gène homologue, une liste des ordres ou superfamilles d'ETs associés.

Le fichier généré contient pour chaque paire :

- Le nom du gène poirier
- Les ETs associés trouvés chez le poirier
- Les ETs associés trouvés chez le pommier
- Le nom du gène homologue chez le pommier

Cette structuration permet une lecture directe de la présence ou de l'absence \*d'éléments transposables associés aux gènes homologues dans chaque espèce.

### D. STATISTIQUES SUR L'ALIGNEMENT DES ETs

L'analyse des associations entre \*éléments transposables et gènes homologues entre le poirier et le pommier a porté sur un total de 29610 paires de gènes. Parmi celles-ci 83,73% présentent au moins un \*élément transposable en commun.

La répartition du nombre d'ETs communs par paire montre que les catégories 1 ET (30,7%) et 2 ETs (28,19%) sont les plus fréquentes. Les paires avec 0 ou 3 ETs communs sont représentées dans des proportions comparables (respectivement 16,27% et 16,43%). En revanche, les cas où 4, 5 ETs ou plus sont partagés entre les deux espèces sont beaucoup plus rares (6,53% pour 4 ETs et 1,87% pour 5 ETs).

Sur l'ensemble des paires, le nombre moyen d'ETs communs est de 1,72, avec une médiane de 2. Le maximum observé est de 7 ETs communs pour certaines paires.

Ces résultats mettent en évidence que, dans la majorité des cas, les gènes homologues entre le pommier et le poirier partagent 1 voire 2 ETs commun par paire. [fig 8]

```
1  total_paires: 29610
2  0_ET: 4818 (16.27%)
3  1_ET: 9091 (30.7%)
4  2_ET: 8346 (28.19%)
5  3_ET: 4866 (16.43%)
6  4_ET: 1934 (6.53%)
7  5+_ET: 555 (1.87%)
8  %_paires_avec_≥1_ET: 83.73%
9
10 moyenne_ETs_communs_par_paire: 1.72
11 médiane_ETs_communs_par_paire: 2
12 max_ETs_communs_par_paire: 7
```

*[Figure 8] : Extrait du fichier de sortie du script "stats\_ET\_alignments.py". Les résultats sont explicités dans le corps du texte*

## DISCUSSION

### A. ANALYSE DES MULTIPLICONS (BdS)

Les relations observées entre les multiplicons du pommier et du poirier confirment les dynamiques évolutives attendues pour des espèces issues d'un même événement de duplication génomique ancienne, mais ayant divergé depuis plusieurs millions d'années.

La présence des relations 1:1, bien que minoritaire, reflète la conservation stricte de certaines régions génomiques, signe que certaines parties du génome sont restées relativement stables malgré le temps écoulé. Cette conservation est généralement associée à des régions contenant des gènes essentiels ou à forte contrainte fonctionnelle (Qiao et al., 2019).

La prédominance des relations 1:n, n:1, illustrée par les moyennes élevées et les cas extrêmes (jusqu'à 33 correspondances pour un seul \*bloc de synténie), suggère des événements de réarrangement importants. Ces observations sont cohérentes avec les mécanismes décrits dans d'autres études portant sur des génomes d'espèces végétales ayant subi des duplications anciennes, où les processus de fractionnement post-duplication entraînent la perte ou la dispersion de gènes dans certaines régions (Sankoff et al., 2012; Teng et al., 2022).

Le fait que les médianes soient égales à 2 pour chaque espèce indique que ces réarrangements concernent une part importante mais modérée du génome, tandis que quelques cas extrêmes participent à faire monter la moyenne. Cela reflète une hétérogénéité des trajectoires évolutives des différentes régions de leur génome.

Ces résultats s'inscrivent dans les observations plus générales des études sur les duplications anciennes chez les Rosacées (Li et al., 2019; Xue et al., 2018), où les génomes du pommier et du poirier ont évolué de façon parallèle mais indépendante, accumulant réarrangement et divergences structurelles. Ces modifications structurelles peuvent être liées à des adaptations spécifiques aux contraintes écologiques et aux stratégies de reproduction propres à chaque espèce.

Comme énoncé dans les résultats, une partie des correspondances entre multiplicons n'a pas été prise en compte dans cette analyse. Cette omission limite l'exhaustivité de la quantification des relations. Il est donc probable que certaines relations 1:n ou même des cas potentiels de relations multiples aient échappé à cette analyse. Cependant, les résultats obtenus fournissent déjà des indices clairs sur la structuration des régions synténiques entre les deux espèces, et confirmant l'intérêt d'une exploration complète dans de futures analyses.

### B. COMPARAISON DE LA CONSERVATION CHROMOSOMIQUE

Cette visualisation Circos permet d'observer et de confirmer la forte conservation de la colinéarité entre *P. communis* et *M. domestica*. La correspondance marquée entre les chromosomes homologues est cohérente avec la proximité phylogénétique des deux espèces au sein des Rosaceae (Velasco et al., 2010; Wu et al., 2013).

Le choix de limiter la visualisation aux relations inter-chromosomiques a facilité l'interprétation globale des échanges entre les génomes. Cependant, le grand nombre de liens présents complexifie localement certaines régions, rendant certaines synténies moins visibles. À la suite de ce visuel, il serait intéressant de faire un circos ne représentant que les liens spéciaux (paires de gènes \*ohnologues intraspécifiques) mais entre les deux espèces, sans les correspondances chromosomes 1 pommier avec chromosome 2 pommier. Cela permettrait de visualiser plus concrètement la conservation ou non de la synténie des gènes suite à la dernière WGD.



### C. STATISTIQUES SUR L'ALIGNEMENT DES ETs

Les résultats obtenus montrent qu'une large majorité des gènes homologues entre le pommier et le poirier partagent au moins un \*éléments transposable commun. La distribution du nombre d'ETs communs est globalement centrée autour de 1 à 2 ETs partagés par paire, ce qui pourrait suggérer une certaine conservation de ces \*éléments entre les deux espèces.

Cependant, il est important de nuancer cette interprétation. La méthodologie employée dans cette analyse repose uniquement sur la présence ou l'absence des ETs dans les annotations associées aux gènes homologues, sans prise en compte de leur position précise par rapport aux gènes. Cela limite fortement la portée des conclusions que l'on peut tirer quant à une éventuelle conservation fonctionnelle ou évolutive de ces ETs.

De plus, le nombre total d'ETs associés à chaque gène n'a pas été pris en compte dans cette analyse. Calculer le nombre d'ETs totaux par paire aurait permis de rapporter le nombre d'ETs communs à une base de comparaison plus informative, et d'évaluer par exemple si certaines paires étaient particulièrement enrichies en ETs communs relativement à leur contenu total en \*éléments transposables.

Ces limitations méthodologiques rendent donc difficile toute conclusion solide sur le degré réel de conservation des ETs autour des gènes homologues entre pommier et poirier. Pour affiner ces résultats, des analyses plus poussées seraient nécessaires, incluant une prise en compte des positions relatives des ETs par rapport aux gènes, de leur ordre d'apparition, ainsi que des catégories \*d'éléments transposables concernées.

En résumé, ces statistiques fournissent une première ébauche descriptive du partage d'ETs entre gène homologue inter-espèces, mais ne permettent pas, en l'état, de conclure fermement sur la conservation ou le rôle fonctionnel potentiel de ces \*éléments.

### D. DISCUSSION LITTÉRAIRE

Plusieurs études menées chez d'autres genres végétaux ont exploré la relation entre conservation de synténie et contenu en éléments transposables (ETs), fournissant un cadre pour interpréter nos résultats sur le pommier et le poirier.

Une analyse approfondie de l'amandier *P. dulcis* et le pêcher *P. persica* a montré que , bien que leurs génomes soient fortement synténique et proches (divergés il y a ~5,9 Ma), ils présentent un nombre élevé de variantes de présence/absence attribuables à des mouvements transposables. De plus, les insertions récentes d'ETs sont corrélées à des différences phénotypiques, notamment le caractère doux de l'amande chez certaines variétés (Alioto et al., 2020).

Dans le genre *Brassica*, l'analyse comparative entre *Brassica rapa* et *Brassica oleraceae* a montré que l'expansion différentielle du génome s'explique en grande partie par une accumulation d'éléments transposables dans certaines régions, notamment des LTR et des transposons à ADN (Parkin et al., 2014).

Plus largement, une étude comparative des génomes de *Fragaria vesca*, *Malus domestica* et *Prunus persica* (Velasco et al., 2010). a mis en évidence une conservation globale des blocs génomiques au sein des Rosacées, tout en soulignant la présence de réarrangements spécifiques à chaque genre, probablement liés à des activité transposables post-duplication. Ces résultats soutiennent l'idée que les ETs jouent un rôle dans la plasticité des génomes végétaux, notamment dans les régions synténiques, et renforcent la pertinence de notre approche comparative pour mieux comprendre la dynamique évolutive des génomes *Malus* et *Pyrus*.



## CONCLUSION

Ce stage avait pour objectif d'évaluer la conservation des synténies entre *P. communis* et *M. domestica*, ainsi que d'explorer l'implication possible des éléments transposables dans l'évolution de ces régions conservées. L'approche mise en place a permis d'identifier des gènes homologues organisés en blocs synténiques entre les deux espèces, et d'associer à ces gènes les ETs présents dans leurs régions génomiques.

Les résultats obtenus confirment une conservation globale de l'organisation des gènes entre le pommier et le poirier, ce qui est cohérent avec leur proximité phylogénétique et avec les observations faites dans d'autres espèces de la famille des Rosacées. Toutefois, le lien entre les ETs identifiés et l'histoire évolutive des génomes reste difficile à établir à partir de ces seules données. En particulier, il n'est pas possible à ce stade de déterminer si le burst d'ETs survenu il y a environ 21 Ma est intervenu avant ou après la séparation des genres *Malus* et *Pyrus*.

Ce travail constitue néanmoins une première étape vers une meilleure compréhension des interactions entre synténie, duplication des gènes et éléments transposables chez ces espèces fruitières. Il ouvre des perspectives intéressantes pour la suite du projet, notamment par l'intégration d'analyses plus fines sur les familles d'ETs impliquées, l'élargissement de l'étude à d'autres génomes de Rosacées, et l'exploitation conjointe de données d'expression génique et phylogénétique.



## BIBLIOGRAPHIE

- Bouanich, Andréa. Titre thèse (thèse en cours), Université d'Angers / IRHS, prévue pour 2025. Données non publiées.
- Aitchison, J. C., Ali, J. R., & Davis, A. M. (2007). When and where did India and Asia collide? *Journal of Geophysical Research: Solid Earth*, 112(B5).  
<https://doi.org/10.1029/2006JB004706>
- Alioto, T., Alexiou, K. G., Bardil, A., Barteri, F., Castanera, R., Cruz, F., Dhingra, A., Duval, H., Fernández I Martí, Á., Frias, L., Galán, B., García, J. L., Howad, W., Gómez-Garrido, J., Gut, M., Julca, I., Morata, J., Puigdomènech, P., Ribeca, P., ... Arús, P. (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes : Results from the almond genome sequence. *The Plant Journal: For Cell and Molecular Biology*, 101(2), 455-472.  
<https://doi.org/10.1111/tpj.14538>
- Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E. A., Gouzy, J., Rees, D. J. G., Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., ... Bucher, E. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*, 49(7), 1099-1106.  
<https://doi.org/10.1038/ng.3886>
- De Storme, N., & Mason, A. (2014). Plant speciation through chromosome instability and ploidy change : Cellular mechanisms, molecular factors and evolutionary relevance. *Current Plant Biology*, 1, 10-33. <https://doi.org/10.1016/j.cpb.2014.09.002>
- Fruit-June-2023.pdf*. (s. d.). Consulté 23 mai 2025, à l'adresse  
<https://www.fas.usda.gov/sites/default/files/2025-03/Fruit-June-2023.pdf>
- Gutierrez-Gonzalez, J. J., & Garvin, D. F. (2011). Reference Genome-Directed Resolution of Homologous and Homeologous Relationships within and between Different Oat Linkage Maps. *The Plant Genome*, 4(3).  
<https://doi.org/10.3835/plantgenome2011.01.0004>
- Ito, H. (2022). Environmental stress and transposons in plants. *Genes & Genetic Systems*,

- 97(4), 169-175. <https://doi.org/10.1266/ggs.22-00045>
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S., & Marra, M. (2009). CIRCOS : An information aesthetic for comparative genomics. *Genome research*, 19, 1639-1645. <https://doi.org/10.1101/gr.092759.109>
- Kumar, A. (1999). *Rising Himalaya : Advent and intensification of monsoon*. [https://www.academia.edu/12168522/Rising\\_Himalaya\\_Advent\\_and\\_intensification\\_of\\_monsoon](https://www.academia.edu/12168522/Rising_Himalaya_Advent_and_intensification_of_monsoon)
- Lallemand, T. (2022). *Évolution des gènes dupliqués chez le pommier : Identification et caractérisation de la dominance du sous-génome dans le génome de la pomme* [Phdthesis, Université d'Angers]. <https://theses.hal.science/tel-04081238>
- Lallemand, T., Leduc, M., Desmazières, A., Aubourg, S., Rizzon, C., Landès, C., & Celton, J.-M. (2023). Insights into the Evolution of Ohnologous Sequences and Their Epigenetic Marks Post-WGD in *Malus Domestica*. *Genome Biology and Evolution*, 15(10), evad178. <https://doi.org/10.1093/gbe/evad178>
- Li, Q., Qiao, X., Yin, H., Zhou, Y., Dong, H., Qi, K., Li, L., & Zhang, S. (2019). Unbiased subgenome evolution following a recent whole-genome duplication in pear (*Pyrus bretschneideri* Rehd.). *Horticulture Research*, 6(1), 1-12. <https://doi.org/10.1038/s41438-018-0110-6>
- Ohno, S. (1970). Polyploidy : Duplication of the Entire Genome. In S. Ohno (Éd.), *Evolution by Gene Duplication* (p. 98-106). Springer. [https://doi.org/10.1007/978-3-642-86659-3\\_17](https://doi.org/10.1007/978-3-642-86659-3_17)
- Parkin, I. A., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., Town, C. D., Nixon, J., Krishnakumar, V., Bidwell, S. L., Denoeud, F., Belcram, H., Links, M. G., Just, J., Clarke, C., Bender, T., Huebert, T., Mason, A. S., Pires, J. C., ... Sharpe, A. G. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biology*, 15(6), R77. <https://doi.org/10.1186/gb-2014-15-6-r77>
- Pont, C., Wagner, S., Kremer, A., Orlando, L., Plomion, C., & Salse, J. (2019). Paleogenomics : Reconstruction of plant evolutionary trajectories from modern and

- ancient DNA. *Genome Biology*, 20(1), 29.  
<https://doi.org/10.1186/s13059-019-1627-1>
- Proost, S., Fostier, J., Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., & Vandepoele, K. (2011). i-ADHoRe 3.0—Fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic acids research*, 40, e11.  
<https://doi.org/10.1093/nar/gkr955>
- Pryszcz, L. P., & Gabaldón, T. (2016). Redundans : An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, 44(12), e113.  
<https://doi.org/10.1093/nar/gkw294>
- Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., & Paterson, A. H. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology*, 20(1), 38. <https://doi.org/10.1186/s13059-019-1650-2>
- Sankoff, D., Zheng, C., & Wang, B. (2012). A model for biased fractionation after whole genome duplication. *BMC Genomics*, 13 Suppl 1(Suppl 1), S8.  
<https://doi.org/10.1186/1471-2164-13-S1-S8>
- Soltis, P. S., & Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology*, 30, 159-165.  
<https://doi.org/10.1016/j.pbi.2016.03.015>
- Tekaia, F. (2016). Inferring Orthologs : Open Questions and Perspectives. *Genomics Insights*, 9, GEI.S37925. <https://doi.org/10.4137/GEI.S37925>
- Teng, J., Wang, J., Zhang, L., Wei, C., Shen, S., Xiao, Q., Yue, Y., Hao, Y., Ge, W., & Wang, J. (2022). Paleopolyploidies and Genomic Fractionation in Major Eudicot Clades. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.883140>
- The Himalayas [This Dynamic Earth, USGS]*. (s. d.). Consulté 24 juin 2025, à l'adresse <https://pubs.usgs.gov/gip/dynamic/himalaya.html>
- Thomas Denecker. (s. d.). Consulté 4 juin 2025, à l'adresse <https://thomasdenecker.github.io/thesisWebsite/annexes/orthologue/>
- Van De Peer, Y., Ashman, T.-L., Soltis, P. S., & Soltis, D. E. (2021). Polyploidy : An evolutionary and ecological force in stressful times. *The Plant Cell*, 33(1), 11-26.

<https://doi.org/10.1093/plcell/koaa015>

Vance, D., & Harris, N. (1999). Timing of prograde metamorphism in the Zaskar Himalaya. *Geology*, 27(5), 395-398.

[https://doi.org/10.1130/0091-7613\(1999\)027<0395:TOPMIT>2.3.CO;2](https://doi.org/10.1130/0091-7613(1999)027<0395:TOPMIT>2.3.CO;2)

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavauiolo, M., Coppola, G., Costa, F., Cova, V., Ri, A., & Viola, R. (2010). The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nature genetics*, 390, 833-839. <https://doi.org/10.1038/ng.654>

Wang, L., Schneider, H., Zhang, X.-C., & Xiang, Q.-P. (2012). The rise of the Himalaya enforced the diversification of SE Asian ferns by altering the monsoon regimes. *BMC Plant Biology*, 12(1), 210. <https://doi.org/10.1186/1471-2229-12-210>

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements.

Wolfe, K. (2000). Robustness—It's not where you think it is. *Nature Genetics*, 25(1), 3-4. <https://doi.org/10.1038/75560>

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Khan, M. A., Tao, S., Korban, S. S., Wang, H., Chen, N. J., Nishio, T., Xu, X., Cong, L., Qi, K., Huang, X., Wang, Y., Zhao, X., Wu, J., ... Zhang, S. (2013). The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Research*, 23(2), 396-408. <https://doi.org/10.1101/gr.144311.112>

Xue, H., Wang, S., Yao, J.-L., Deng, C. H., Wang, L., Su, Y., Zhang, H., Zhou, H., Sun, M., Li, X., & Yang, J. (2018). Chromosome level high-density integrated genetic maps improve the *Pyrus bretschneideri* 'DangshanSuli' v1.0 genome. *BMC Genomics*, 19(1), 833. <https://doi.org/10.1186/s12864-018-5224-6>

## RÉSUMÉ

### Etude comparative des synténies pommier vs poirier

Chez les plantes, les éléments transposables (ETs) représentent une part importante des génomes et pourraient jouer un rôle clé dans l'évolution et la diversification des espèces. Le pommier (*Malus domestica*) et le poirier (*Pyrus communis*), deux espèces proches de la famille des Rosacées, présentent des différences phénotypiques importante malgré leur proximité évolutive. Ce stage avait pour objectif de déterminer si les ETs ont pu jouer un rôle dans la différenciation entre ces deux espèces, notamment en relation avec le burst d'ETs supposé il y a 21 millions d'années.

Pour cela, une analyse comparative de la synténie génique entre le pommier et le poirier a été réalisée, couplée à l'étude de la répartition des ETs dans ces régions conservées. L'objectif était d'identifier les ETs présents dans les régions homologues entre les deux espèces, afin d'évaluer leur distribution et leur éventuel lien avec la dynamique des blocs synténiques. Ce travail constitue une première étape vers une meilleure compréhension du rôle potentiel des ETs dans l'évolution différentielle de ces deux espèces fruitières, et ouvre des perspectives pour des analyses plus ciblées sur certaines familles d'ETs ou sur les conséquences fonctionnelles de leurs insertions.

**Mots-clefs :** *Malus domestica*, *Pyrus communis*, synténie, éléments transposables (ETs).

## ABSTRACT

### Comparative study of synteny between apple and pear

In plants, transposable elements (TEs) represent a major part of genomes and may play a key role in the evolution and diversification of species. The apple (*Malus domestica*) and the pear (*Pyrus communis*), two closely related species of the Rosaceae family, exhibit significant phenotypic differences despite their evolutionary proximity. This project aimed to investigate whether TEs could have contributed to the divergence between these two species, particularly in relation to the TE burst hypothesized to have occurred around 21 million years ago.

To this end, a comparative analysis of gene synteny between apple and pear was carried out, combined with an investigation of the distribution of transposable elements (TEs) within these conserved regions. The objective was to identify the TEs present in homologous regions between the two species, in order to assess their distribution and potential link with syntenic block dynamics. This work represents a first step toward a better understanding of the potential role of TEs in the differential evolution of these two fruit species, and opens up perspectives for more targeted analyses on specific TE families or the functional consequences of their insertions.

**Key words :** *Malus domestica*, *Pyrus communis*, synteny, transposable elements (TEs)