

COLLEGE	ECOLOGIE
DOCTORAL	GEOSCIENCES
BRETAGNE	AGRONOMIE ALIMENTATION

# THESE DE DOCTORAT DE

L'INSTITUT AGRO RENNES ANGERS

ECOLE DOCTORALE N° 600

*Ecologie, Géosciences, Agronomie, Alimentation*

Spécialité : Génétique, génomique et bio-informatique

Par

**Marianne LAURENÇON**

**Déterminants génétiques et prédictions de la germination et de la vigueur précoce chez le colza**

Thèse présentée et soutenue à Rennes, le 04 février 2025

Unité de recherche : Institut de Génétique, Environnement et Protection des Plantes (IGEPP),  
INRAE, Institut Agro Rennes-Angers, Université de Rennes, Le Rheu  
Thèse n° : 2025AGRORA03

**Rapporteurs avant soutenance :**

Vincent SEGURA

Chargé de recherche, INRAE,

Sophie BRUNEL-MUGUET

Amélioration génétique et adaptation des plantes méditerranéennes et tropicales

Chargée de recherche, INRAE,

Laboratoire d'écophysiologique végétale, agronomie et nutrition

**Composition du Jury :**

Présidente : Maria MANZANARES-DAULEUX  
Examinateurs : Jean-Christophe AVICE

Professeure, Institut Agro Rennes-Angers, IGEPP

Professeur, Université de Caen,

Laboratoire d'écophysiologique végétale, agronomie et nutrition

Eric MANDROU

Sélectionneur colza, LIDEA

Béatrice TEULAT

Professeure, Institut Agro Rennes-Angers,

Institut de Recherche en Horticulture et Semences

Professeure, Institut Agro Rennes-Angers, IGEPP

Dir. de thèse : Anne LAPERCHE  
Nathalie NESI

Directrice de recherche, INRAE, IGEPP



Déterminants génétiques et  
prédictions de la germination  
et de la vigueur précoce  
chez le colza



# Remerciements

---

Ces trois ans (et quelques mois) au sein de l'IGEPP ont été extrêmement enrichissants, autant sur le plan personnel que professionnel, et ce grâce à l'ensemble des personnes que j'ai pu rencontrer. Je tiens à remercier chaleureusement toutes les personnes qui ont contribué, de près ou de loin, à cette aventure scientifique et humaine.

Dans un premier temps, je souhaiterais remercier les membres de mon jury de thèse, Vincent Segura et Sophie Brunel-Muguet pour avoir accepté d'être les rapporteurs de ce manuscrit, ainsi que Jean-Christophe Avice, Eric Mandrou, Maria Manzanares-Dauleux et Béatrice Teulat pour avoir accepté d'être les examinateurs de mes travaux.

J'aimerais également remercier l'association Promosol ainsi que l'INRAE pour avoir co-financé ma thèse, notamment à travers le projet NAVIG.

Je remercie chaleureusement les membres de mon comité de thèse, Renaud Rincent, Christopher Sauvage et Mathieu Rousseau-Gueutin, pour vos idées, vos conseils avisés et votre écoute. Christopher, un énorme merci d'avoir accepté de suivre la réalisation de mes travaux trois ans de plus, et d'avoir apporté ton enthousiasme et tes conseils éclairés (je ne parle pas que des choix de bières). Tu le sais, mais c'est en grande partie grâce à toi (ou à cause ?) que j'en suis là.

Un grand merci à Marie-Hélène Wagner, Didier Demilly et Sylvie Ducournau pour la génération des données de germination sur les bancs du GEVES, ainsi que pour les échanges, le temps que vous nous avez accordé à Angers avec Cécile et tous vos conseils.

Je tiens également à remercier les organisateurs du groupe R2D2 de m'avoir permis de présenter mes travaux, et aux membres du groupe pour leurs conseils avisés et pour nos échanges durant le séminaire à Angers. Merci également à la section « outils et méthodes en appui à la sélection variétale » de l'ASF et aux organisateurs du séminaire sur la Sélection Phénomique pour m'avoir invitée pour présenter mes travaux. Un grand merci également à Marina Bossard, Martine Leflon, Christophe Jestin et l'équipe organisatrice des Carrefour Colza pour m'avoir donné l'opportunité de présenter ma thèse à plusieurs reprises. Un merci tout particulier à l'unité pédagogique Agronomie et Amélioration des plantes de l'Institut Agro Rennes-Angers, et plus particulièrement à Anne, Maria, Mélanie Jubault et Annie Masson pour m'avoir permis de donner des TD trois années de suite ainsi que d'être rapportrice de stage de M2. Un grand merci également à Jacques David pour m'avoir intégrée dans l'équipe pédagogique de la super aventure qu'est l'atelier de prédiction et de sélection génomique Selgen.



Le plus grand des mercis revient à Anne Laperche et Nathalie Nesi. Merci pour tout, pour m'avoir fait confiance, votre présence sans faille, votre patience, vos conseils, les discussions et votre bonne humeur. Vous avez été particulièrement présentes dans les moments les plus compliqués, je ne vous remercierais jamais assez pour votre soutien et vos conseils. Merci pour votre calme lorsque l'on « naviguait en eaux troubles ». Vous m'avez appris à me canaliser (ce qui n'a pas toujours été simple, désolée, et qui reste encore un peu compliqué) et à me donner confiance en moi. Un énorme merci pour tout ça ! Anne, merci d'avoir partagé ton bureau avec moi. Merci pour nos discussions pas toujours très scientifiques jusqu'à parfois un peu trop tard le soir, pour les samedis « agro » et pour ta bonne humeur perpétuelle. Ces quelques années ont été un vrai plaisir !

Je remercie également l'ensemble des membres de RCA, sans qui ce travail n'aurait pas pu aller aussi loin. Un grand merci à Elise Alix, Solenn Guichard et à Bernard Moulin, la team « expés », pour votre aide sur les différentes expés au tunnel et pour votre bonne humeur à toute épreuve ! Elise et Bernard (ou devrais-je dire Monsieur Teflon ?), on aura fait un sacré paquet de goutteurs, d'installation de matériel au tunnel, de semis et de prises de photos ensemble. Solenn, merci pour ces après-midi à cramer sous le tunnel pour acquérir les NIRs. Et un énorme merci pour votre aide sans faille lors des chantiers de récolte. J'ai énormément appris à vos côtés. Un grand merci à Stéphane Jumel, monsieur Photos, pour tout le temps que tu as consacré à mettre au point la perche et pour ton aide sur les photos et sur leur analyse. Merci également à Aurélien Carillo pour ton aide sur la prise de photos lors des dernières expérimentations et pour nos sessions « pesage », ainsi qu'à Liliya Shveda pour nos journées pesées-broyages hautes en couleur. Merci également à Sophie Rolland pour m'avoir formée à l'acquisition des spectres NIRs, pour tous tes tests et pour nos discussions « spectrales ». J'aimerais également remercier Cécile Baron pour son aide sur le génotypage et sur les expérimentations préliminaires de germination. Merci pour ta bonne humeur et pour tous nos moments « papotages ». Un grand merci également à toutes les personnes qui sont venues nous aider tous les quatre sur les chantiers de récolte : Cécile Baron, Solenne Berardocco, Aurélien Carillo, Vanessa Clouet, Anne-Marie Denis, Clémence Germain, Florian Guilbaud, Patrick Leconte, Kevin Levardois et Sophie Rolland. Merci pour votre aide indispensable, et pour les petits moments conviviaux toujours accompagnés de bons gâteaux ! Mention spéciale à Florian et Aurélien, pour leur énergie (un peu trop !) débordante. Ces récoltes n'auront pas été de tout repos ! Merci également à Younès Dellero pour nos discussions autour du NIRs et de la physio. De manière générale, merci à vous tous pour m'avoir très vite intégrée dans l'équipe, et pour nos nombreuses discussions au détour d'un couloir, durant les pauses ou entre deux découpes de plantes.

Un merci tout particulier pour Julie Legrix, qui a bien voulu faire son stage de M2 avec nous. Ça a été un vrai plaisir de travailler avec toi et d'apprendre ensemble. Merci de m'avoir fait confiance, j'espère



avoir été à la hauteur de tes espérances et je te souhaite le meilleur pour ta thèse et pour la suite de ton avenir professionnel.

Marion Gauthier et Céline Richard-Molard, merci de m'avoir donné envie d'aimer l'écophysiologie. Vous avez été d'une patience sans faille et d'une grande pédagogie. Marion, merci pour ces TDs partagés, pour m'avoir fait confiance et pour m'avoir donné confiance en moi. Merci pour toutes nos discussions, pour ton aide et ta bienveillance. Céline, merci pour tous nos échanges et ton aide sur le fonctionnement CN. Merci pour toutes nos discussions et pour tes conseils. Je garderai un souvenir inoubliable de nos balades dans Sydney.

Merci à l'équipe serre, et plus particulièrement à Laurent Charlon, Jean-Paul Constantin et Frédéric Le Tertre pour toute l'aide apportée durant nos expérimentations en tunnel, pour l'irrigation contrôlée d'une main de maître et pour votre aide sur l'identification des causes de la mort de nos si jolies plantules ! Merci également à Frédéric Brunet pour sa patience. J'en profite pour m'excuser auprès de toi, j'essaierais d'éteindre plus régulièrement mon ordinateur à l'avenir !

Plus généralement, j'aimerais remercier toutes les personnes de l'IGEPP avec qui j'ai pu passer de bons moments. Un grand merci aux équipes « d'à côté », DEBI, MVI et le pôle Ressources, pour l'ambiance conviviale dans le bâtiment. Je me souviendrai longtemps du sacré périple qu'a été le changement de cartouche d'encre, immortalisé à jamais par Nathalie. Concernant les moments papotages, scientifiques (ou pas) mais toujours remplis de bonne humeur et de bons conseils, j'aimerais remercier chaleureusement Françoise Le Caherec. Tu n'étais pas là souvent, mais quel plaisir quand c'était le cas. J'aimerais également remercier la « team du soir » : Anne, Mathieu Tiret, Bernard Rolland, Anne-Marie Chèvre et Remy Perronne pour votre présence et toutes nos discussions. Mathieu, merci pour tes conseils, pour toujours avoir été là lorsque je faisais trop de nourriture et pour ne m'avoir jamais virée de ton bureau, alors qu'il y aurait eu des moments où ça aurait été « mérité ». J'en profite pour remercier également Anouk Zancarini et Marc Galland, pour nos sorties, vos super cocktails et repas et j'en passe. Merci pour toute votre aide et votre soutien. Vous êtes vraiment des personnes en or ! J'aimerais également remercier Cyril Falentin et Mathieu Rousseau-Gueutin. Cyril, merci pour tous ces fous rire et pour ta bienveillance permanente. Je te souhaite le meilleur à Montpellier. Mathieu, merci pour tous tes conseils, nos discussions musique et autres. Par contre, je ne te remercie pas pour avoir rempli ma voiture de ballons et de chips plastiques. Dans les personnes que je ne remercie pas, il y a également toi Florian, pour le nombre insensé de crises cardiaques que j'ai failli faire par ta faute ! Malgré tout, merci pour toutes nos discussions et toutes ces soirées avec Amélie. Hâte d'aller au Hellfest avec vous ! Merci également à Laura Chauvin pour sa bonne humeur, sa douceur et sa gentillesse. Ta recette de gnocchis est à présent l'une de mes recettes préférées. Un merci spécial à



Sarah Ben Sadoun pour ses conseils, nos discussions, et surtout pour tous les restos faits ensemble ! Merci pour tous les super moments passés à Montpellier, Rennes et surtout Vienne. Je tiens également à remercier toutes les personnes avec qui j'ai partagé mon bureau, ainsi que la team midi/pause-café, qui a évolué au fil des années. Merci à Mathilde Petitpas et Marie Devers pour toutes nos discussions, pour votre bonne humeur et pour m'avoir donné le virus de la collection de plantes. Désolée d'avoir lâché en cours de route... J'espère vous revoir très vite, à la maison cette fois-ci, maintenant que vous avez migré dans le sud-ouest ! J'aimerais également remercier Franz Boideau, Youssef Bouargalne et Loeiz Maillet. Merci pour tous ces moments passés ensemble, pour vos conseils ainsi que pour votre joie de vivre. Merci à Julie Ferreira De Carvalho et à Maud Facon pour nos discussions au début de ma thèse. Merci également à la team « doc RCA » : Mathieu Aubert, Pierre-Nicolas Boulc'h, Youcef Haddad et Khadidiatou Demba, merci pour tous ces moments de partage autour des galères de la thèse, notamment des manipes. Un merci tout particulier à Khadidiatou, qui m'a rejoint sur le sujet de la vigueur à la fin de ma thèse. Merci pour nos discussions, pour tous les moments de réflexion sur les traits, les manipes, ou des sujets plus philosophiques. Ça a été un vrai plaisir de réaliser ma dernière année avec toi. Merci également à Rosa Kemmerling, pour tous nos moments autour d'un verre ou d'un bon petit plat après nos journées toutes pourries, ainsi que pour les balades du samedi après-midi. Grâce à toi, je connais presque le centre-ville de Rennes par cœur ! Je tiens également à remercier Valérie Schurdi-Levraud, qui m'a transmis la passion de l'amélioration des plantes dès la L2 et m'a toujours poussée à me dépasser.

Enfin, je tiens à remercier tout particulièrement ma famille, en particulier mes parents, mes sœurs, et mes merveilleux neveux et nièces pour leur amour et pour leur soutien indéfectible. Emilie, Lise, merci pour vos conseils, votre soutien permanent malgré la distance et pour tous ces moments passés ensemble. J'ai pu mesurer au cours de ces années la chance que j'avais d'être votre sœur.



# Sommaire

---

Contexte général.....	1
<b>Enjeux et contraintes de la production de colza.....</b>	1
<b>Contexte de la thèse.....</b>	7
Chapitre I. Synthèse bibliographique.....	9
<b>I.1. Processus physiologiques de l'implantation du colza et vigueur précoce.....</b>	9
I.1.1. Germination de la graine.....	9
I.1.2. Emergence de la nouvelle plantule.....	10
I.1.3. De la levée au stade 4-5 feuilles, la phase délicate du développement précoce.....	10
I.1.4. Leviers pour garantir une implantation réussie de la culture.....	11
<b>I.2. Identification de traits cibles pour la germination et la vigueur précoce.....</b>	15
I.2.1. Décomposition de processus intégratifs en traits fonctionnels.....	15
I.2.2. Les traits fonctionnels liés à la germination ou à la levée.....	16
I.2.3. Les traits morphologiques et d'architecture liés à la vigueur précoce.....	17
I.2.4. Les traits physiologiques candidats pour expliquer la vigueur précoce.....	17
<b>I.3. Contrôle génétique des traits impliqués dans la germination et la vigueur précoce chez le colza.....</b>	20
I.3.1. Déterminants génétiques de la germination.....	20
I.3.2. Déterminants génétiques de la vigueur précoce.....	21
I.3.3. Y-a-t-il encore la place pour des analyses génétiques de la germination et la vigueur précoce chez le colza ? .....	22
<b>I.4. Histoire du colza : impacts sur l'organisation génomique et la diversité génétique disponible.....</b>	23
I.4.1. L'organisation du génome du colza résulte d'une histoire évolutive longue.....	23
I.4.2. Ressources génomiques.....	24
I.4.3. La sélection et l'amélioration génétique du colza ont contribué à réduire la diversité génétique déjà faible chez cette espèce.....	25
I.4.4. Ressources disponibles pour élargir la diversité génétique chez le colza.....	26
I.4.5. Etendue du déséquilibre de liaison dans les populations de colza.....	27
I.4.6. Ressources génétiques utilisées pour la thèse.....	27
<b>I.5. Ressources et méthodes pour la sélection chez le colza.....</b>	28



I.5.1. Mobilisation des ressources génétiques en pre-breeding.....	28
I.5.2. Principaux facteurs impactant le gain génétique en sélection.....	29
I.5.3. La sélection génomique pour la sélection de traits au déterminisme génétique complexe.....	30
I.5.4. Sélection phénotypique : prédictions à l'aide de données endophénotypiques.....	33
I.5.5. Sélection des traits de germination, développement et vigueur : GS ou PS ? .....	35
<b>I.6. Objectifs, hypothèses et stratégie déployée au cours de la thèse.....</b>	<b>36</b>
Chapitre II. Déterminants génétiques de la germination et prédictions.....	39
<b>II.1. Préambule.....</b>	<b>39</b>
<b>II.2. Article.....</b>	<b>40</b>
Chapitre III. Architecture génétique et prédictions de la vigueur précoce chez le colza.....	57
<b>III.1. Préambule.....</b>	<b>57</b>
Questions traitées dans le chapitre III.....	57
Acquisition des données.....	58
<b>III.2. Favorable alleles for early vigor related traits show specific occurrence regarding rapeseed germplasms.....</b>	<b>60</b>
Background.....	61
Material and Methods.....	63
Results.....	67
Discussion.....	71
Supplemental information.....	76
<b>III.3. Prédiction phénotypique de la vigueur précoce.....</b>	<b>77</b>
Matériel et Méthodes.....	78
Résultats.....	83
Eléments de discussion.....	87
Chapitre IV. Discussion générale et conclusion.....	91
<b>IV.1. Traits impliqués dans la germination et la vigueur précoce.....</b>	<b>91</b>
IV.1.1. Les traits de germination et de vigueur précoce étudiés sont faiblement corrélés.....	91
IV.1.2. Corrélation entre le poids de mille grains et les traits de vigueur précoce.....	92
IV.1.3. Quels traits de germination et de vigueur précoce en conditions de contraintes ? .....	93



<b>IV.2. Architecture génétique de la germination et de la vigueur précoce.....</b>	95
<b>IV.3. Valorisation de la diversité génétique pour améliorer la germination et la vigueur précoce.....</b>	96
IV.3.1. Améliorer la germination et la vigueur précoce par la remobilisation de la diversité génétique.....	96
IV.3.2. La prédiction phénotypique et la prédiction génomique pourraient permettre de remobiliser de la diversité génétique dans les populations de pre-breeding.....	97
<b>IV.4. Prédiction des traits d'implantation et intérêt de la prédiction phénotypique pour ces traits.....</b>	99
IV.4.1. Les modèles de prédiction génomique et phénotypique sont pertinents pour les traits d'implantation, mais restent à optimiser.....	99
IV.4.2. La combinaison d'un ensemble d'endophénotypes pourrait permettre de mieux capturer le GxE et les mécanismes impliqués dans la vigueur.....	100
<b>Conclusion.....</b>	103
<b>Références bibliographiques.....</b>	104



# Liste des principales abréviations

---

## A

AUC : Aire sous la courbe

## B

BLUE : *Best Linear Unbiased Estimates*,

meilleur estimateur linéaire sans biais

BLUP : *Best Linear Unbiased Predictions*,

meilleur prédicteur linéaire sans biais

## C

CMS : stérilité mâle cytoplasmique

CNV : *Copy Number Variation*, variabilité du nombre de copies d'un gène

## D

DAS : *Days After Sowing*, jours après semis

DJ : Degrés-Jours

## E

ES : *radicule Elongation Speed*, vitesse

d'elongation de la radicule.

## F

FG : *First Germination time*, temps pour obtenir la première germination du lot de graines.

## G

GxE : Interaction Génotype-Environnement

GBLUP : *Genomic BLUP*, BLUP génomique

GDD : *Growing degree days*. Cumul des degrés-jours en °C.

GEBV : *Genomic-Estimated Breeding Values*, valeurs génétiques estimées par des données génomiques

GP : *Genomic prediction*, Prédiction génomique

GR (GP dans le chapitre 2) : *Germination rate (ou germination percentage)*, Taux de germination à 96h

GS : *Genomic selection*, Sélection génomique

GR36 (GP36 dans le chapitre 2) : *Germination rate at 36h (ou germination percentage at 36h)*, Taux de germination à 36h

GWAS : *Genome Wide Association Study*, analyse d'association le long du génome

## H

H : matrice d'apparentement spectrale

HD : Haploïdes doublés

## K

K : matrice d'apparentement génomique (ou *kinship*)

## L

LASNQ : *Leaf Area / Shoot Nitrogen Quantity ratio*, rapport entre la surface foliaire et la quantité d'azote dans les parties aériennes

LAR : *Leaf Area Ratio*, ratio entre la surface foliaire et la biomasse totale

LWR : *Leaf Weight Ratio*, ratio entre la biomasse aérienne et la biomasse totale

## M

MGT : *Mean Germination Time*, temps moyen de germination

## N

NIRS : *Near Infrared Spectra*, spectre dans l'infrarouge proche

NUE : *Nitrogen Use Efficiency*, efficience de l'azote

NUpE : *Nitrogen Uptake Efficiency*, efficience d'absorption de l'azote

NUtE : *Nitrogen Utilization Efficiency*, efficience d'utilisation de l'azote

## P

PAV : *Presence Absence Variation*, variation présence/absence

PEPV : *Phenomic Estimated Phenotypic Values*, valeurs phénotypiques estimées à partir de données phénotypiques



PA : *Predictive Ability*, capacité prédictive  
PP : *Phenomic Prediction*, Prédiction phénomérique  
PS : *Phenomic Selection*, Sélection phénomérique

## Q

QTL : *Quantitative Trait Loci*, Locus de caractères quantitatifs

## R

SR : *Ratio Shoot / Root*, ratio entre la partie aérienne et la partie racinaire  
rrBLUP : *ridge regression BLUP*, BLUP par régression d'arête

## S

SAM : Sélection Assistée par Marqueurs  
SCA : *Specific Carbon Assimilation*, efficience d'assimilation du carbone  
SLA : *Specific Leaf Area*, ratio entre la surface foliaire et la biomasse aérienne  
SNP : *Single Nucleotide Polymorphisms*, polymorphisme nucléotidique

SNU : *Specific Nitrogen Uptake*, efficience d'absorption spécifique de l'azote par les racines  
SOSR : *Spring Oilseed Rape*, colza oléagineux de printemps  
SRL : *Specific Root Length*, longueur spécifique de la racine

## T

T20-T50-T80 : *Time to reach 20-50-80% of germination*, temps pour obtenir 20-50-80% de germination  
TSW : *Thousand Seed Weight*, poids de mille grains

## V

VI : *Volume Imbibition*, volume de la graine lors de l'imbibition

## W

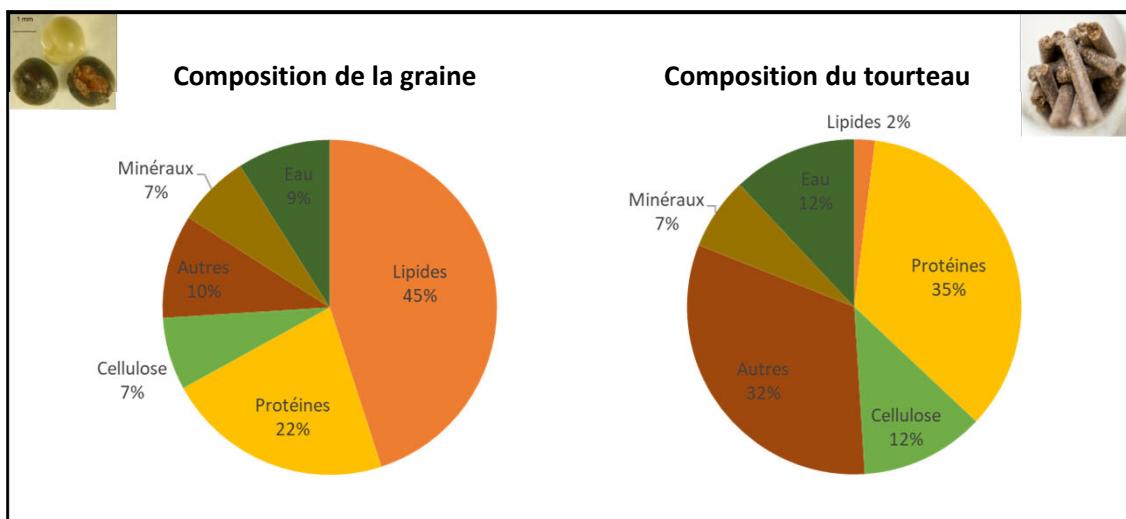
WOSR : *Winter Oilseed Rape*, colza oléagineux d'hiver



---

## **CONTEXTE GENERAL**

---



**Figure 1. Composition de la graine et du tourteau de colza.**

Le compartiment 'Autres' comprend les fibres de façon générale (lignines et hémicelluloses en particulier).

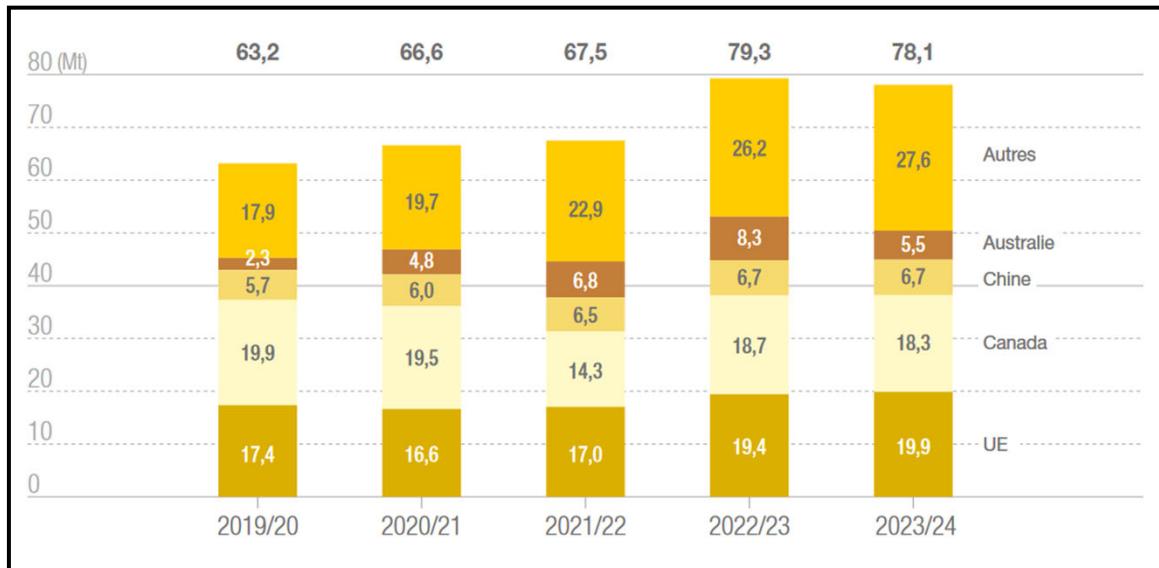
## Enjeux et contraintes de la production de colza

---

### Place du colza dans la production mondiale de graines oléo-protéagineuses

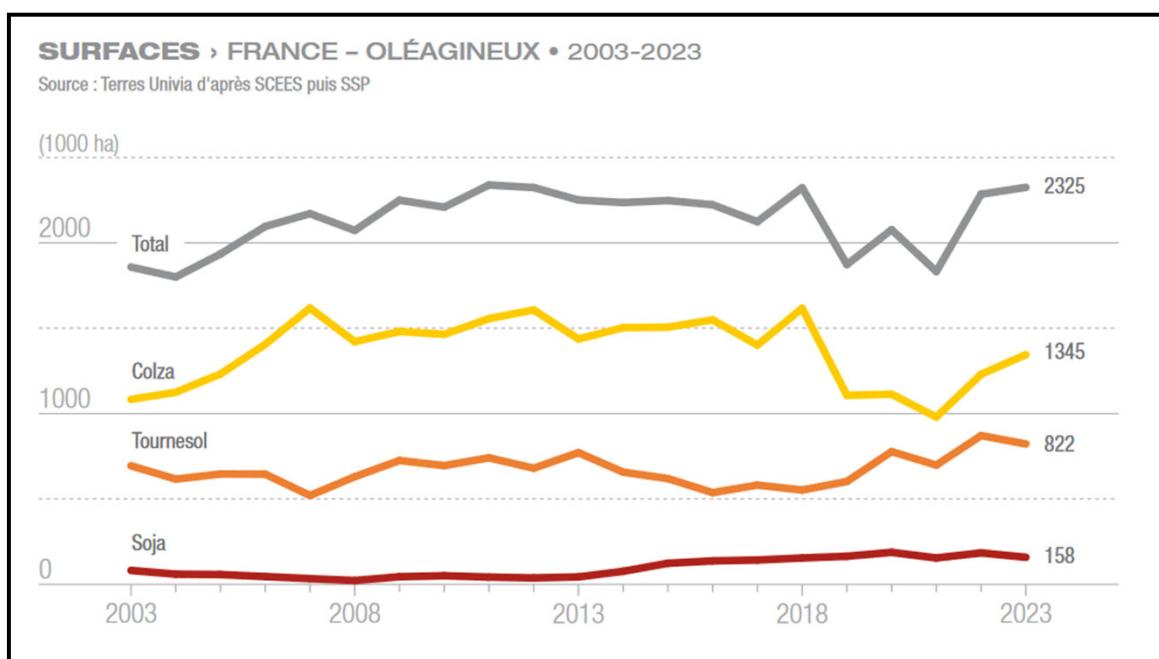
La production mondiale de graines oléo-protéagineuses affiche une croissance soutenue, en raison de l'essor de la démographie mondiale et de modifications dans les régimes alimentaires, qui se caractérisent dans de nombreux pays par une hausse de la consommation d'huiles et de protéines végétales. Par ailleurs, le développement des utilisations industrielles (biocarburants, chimie verte, cosmétiques) contribue également à cette dynamique. Avec près de 630 millions de tonnes (Mt) de graines produites en 2023 (Chatellier, 2022), les oléo-protéagineux génèrent une richesse importante qui profite directement aux producteurs (400€ / t de graines), et bien plus encore après transformation des graines pour les débouchés alimentaires, énergétiques ou chimiques. La multiplicité des cultures oléo-protéagineuses offre une certaine complémentarité en termes d'approvisionnement (distribution géographique) et de qualité des produits pour différents usages (composition en acides gras des huiles et en acides aminés des protéines). Cependant, les échanges de matières sur les marchés internationaux sont impactés et fortement complexifiés par les niveaux de production (fluctuation des récoltes annuelles) et le contexte géopolitique. Ces dernières années ont été fastes en raison de la guerre en Ukraine qui a provoqué de très fortes hausses de prix, notamment pour le colza et le tournesol en Europe. Cette dimension économique provoque un regain d'intérêt des producteurs pour les oléo-protéagineux.

Avec près de 80 Mt de graines produites en 2023-2024, le colza est la 2<sup>ème</sup> culture de graines oléo-protéagineuses sur la scène mondiale, loin derrière le soja (396 Mt, données AGRESTE 2023) et devançant l'arachide et le tournesol (50 Mt chacun). La graine de colza est composée d'eau, de lipides, de protéines et des résidus dont des glucides, des fibres, des minéraux et des cendres (**Figure 1**, Nesi et al. 2008). Les efforts de sélection au cours des cinq dernières décennies ont principalement porté sur la teneur en lipides qui avoisine désormais 45-50% (w/w), faisant de la graine de colza un produit de choix pour la filière des huiles végétales. Par ailleurs, les variétés double zéro ('00') sont les standards cultivés depuis les années 1980 à travers le monde. Elles résultent de programmes de sélection qui ont éliminé l'acide érucique (C22:1) et fortement réduit les glucosinolates (GSL) dans les graines (Busch et al. 1994). Ces variétés produisent une huile de haute qualité, riche en acide gras insaturés (C18:2 et C18:3), précurseurs des oméga-3 et oméga-6, et ainsi adaptée à des usages à la fois alimentaires (huile de table, industrie agroalimentaire) et industriels (biocarburants, chimie verte). Les tourteaux, coproduits issus de l'extraction d'huile, sont principalement utilisés dans les rations alimentaires des ruminants. Les programmes récents sur la qualité de la production de colza visent à

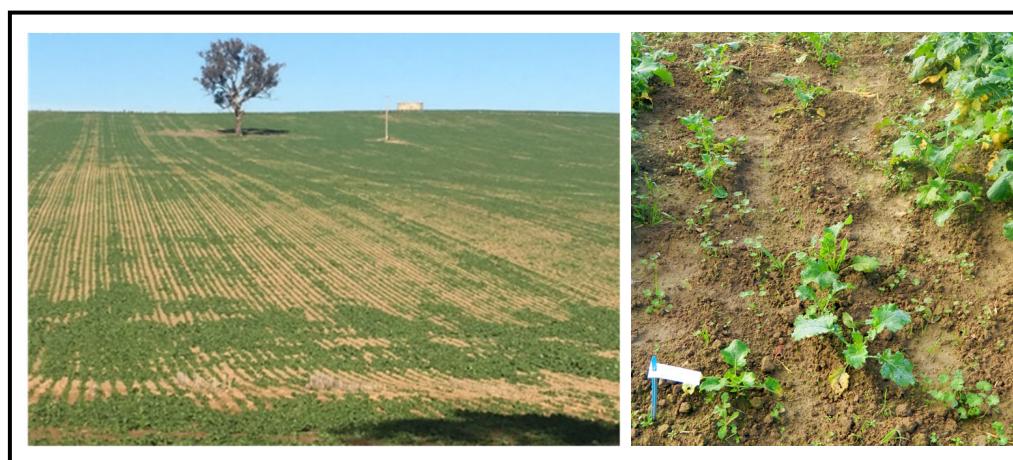


**Figure 2. Production de graines de colza (en Mt) dans le monde entre 2019 et 2024 (données : Terres Univia 2023).**

‘UE’ comprend le Royaume-Uni; ‘Autres’ comprend l’Ukraine et la Russie.



**Figure 3. Surfaces cultivées (en Mha) en France entre 2003 et 2023 pour le colza, le tournesol et le soja (données : Terres Univia 2023).**



**Figure 4. Parcelles de colza où l'implantation est clairsemée.**  
(Photos: Nelson et al. 2022; N. Nesi INRAE).

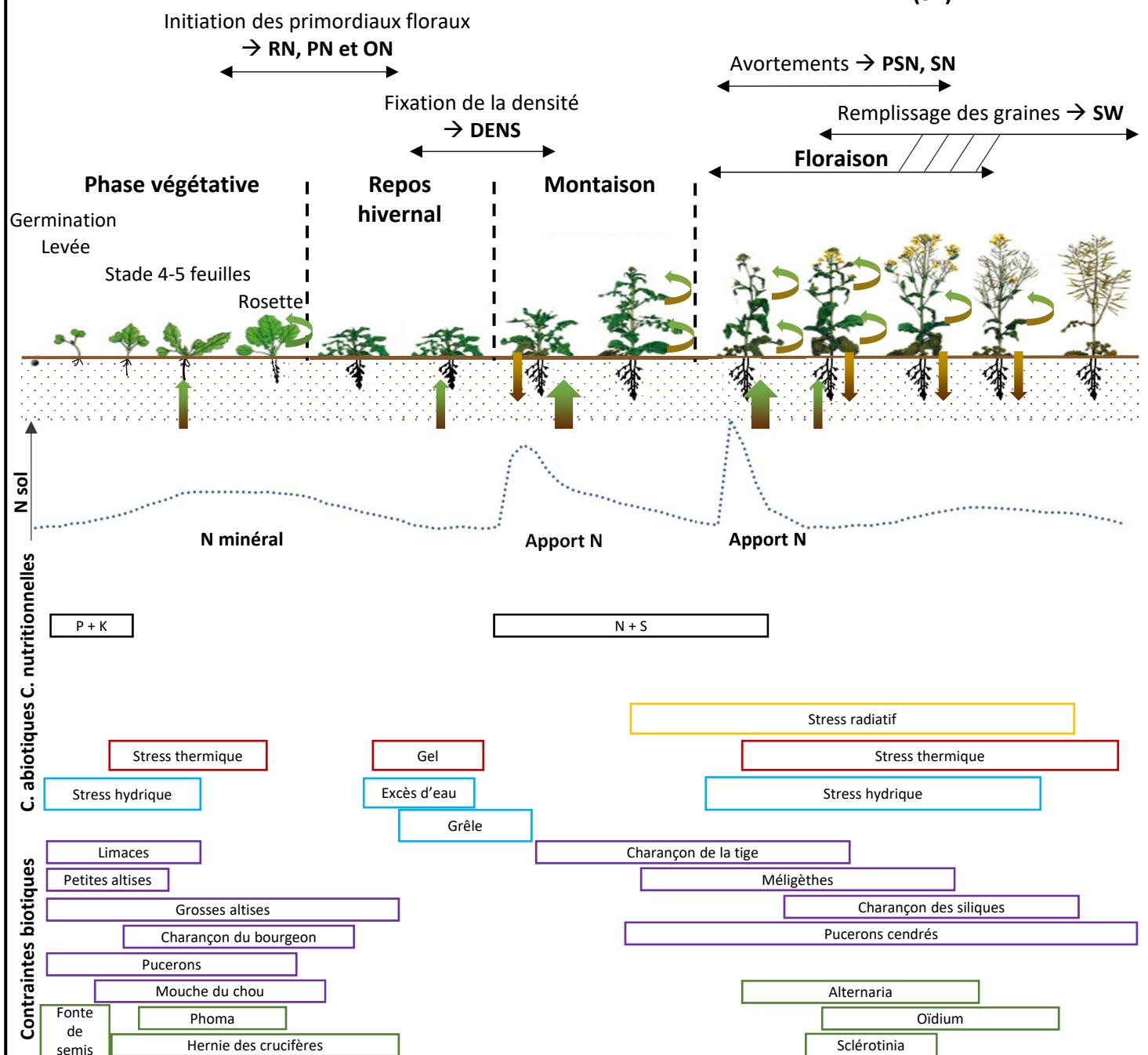
améliorer la qualité du tourteau afin de positionner ce produit comme alternative aux tourteaux de soja importés. Les cibles sont l'augmentation de la teneur en protéines, l'équilibre en acides aminés ainsi que la réduction des fibres et des composés antinutritionnels pour améliorer la digestibilité *in vivo*. Ces programmes devraient permettre une meilleure valorisation des tourteaux de colza, notamment pour l'alimentation des monogastriques (volailles, porcs), favorisant ainsi l'autonomie protéique des zones concernées. En plus de produire de l'huile et de contribuer à l'autonomie protéique des élevages, la culture de colza remplit positivement différents services écosystémiques dans des systèmes de culture agroécologiques : diversification des rotations, amélioration de la structure et de l'état organique des sols grâce à sa racine pivot, couverture végétale durant l'hiver (pour le colza d'hiver), limitation des adventices et interruption des cycles de développement des agents pathogènes des céréales dans les rotations (Sausse et al. 2012). C'est également une espèce mellifère facilitant la biodiversité des insectes dans les parcelles. Le nectar produit est disponible en sortie d'hivernage des insectes, ce qui favorise leur activité printanière.

Les principales zones de production de colza sont l'Union Européenne (UE), le Canada et la Chine (productions respectives de 19.9 Mt, 18.3 Mt, 6.7 Mt de graines en 2023-2024 ; **Figure 2**). Selon les climats, différents types de colza sont cultivés. Ainsi, les variétés de colza d'hiver prédominent dans les régions à hiver doux (Europe de l'Ouest). Elles sont semées à la fin de l'été et nécessitent une période de vernalisation pour fleurir. Leur cycle de développement est long (~ 10 mois). À l'inverse, les variétés de printemps sont semées à la fin de l'hiver et prédominent dans les régions septentrionales (Europe de l'Est, Canada, Asie) ainsi qu'en Australie. Leur cycle de développement est plus court (~ 6 mois) et ne nécessite pas de période de vernalisation. Leur rendement est très inférieur aux types hiver. Enfin, les variétés adaptées à la culture en Asie requièrent une période de vernalisation qui est beaucoup plus courte que les variétés d'hiver (4 semaines vs. 8-10 semaines), elles sont qualifiées de variétés semi-hiver. Le rendement moyen de la culture de colza dépend du type variétal (lignée, hybride F1, association variétale), des conditions environnementales, des pratiques agronomiques (travail du sol, intrants) et des systèmes de culture, et peut varier de 15-20 q/ha au Canada ou en Europe de l'Est (colza de printemps cultivé en système extensif à bas intrants) à 35 q/ha en Europe de l'Ouest (colza d'hiver cultivé en système plus intensif).

En France, le colza est une culture majeure des assolements à l'échelle du territoire national (1.3 Mha en 2023 ; **Figure 3**). La production atteint 4.3 Mt de graines, plaçant ainsi la France en tête au niveau européen. Cependant, les surfaces ont connu une diminution sensible depuis 2018 (**Figure 3**) qui s'explique en partie par des problèmes d'implantation et de végétation à l'automne (**Figure 4**). En effet, les conditions météorologiques souvent défavorables en fin d'été (sécheresses suivies de fortes pluies) provoquent une émergence et un développement difficiles, rendant les plantes d'autant plus sensibles

## Elaboration du potentiel de rendement

## Réalisation du rendement (SY)



**Figure 5. Cycle cultural du colza d'hiver et relations entre croissance de la plante, dynamique de l'azote et mise en place des composantes du rendement (d'après Bouchet et al. 2016 et documents Terres Inovia).**

Le graphique supérieur représente le développement du peuplement végétal au cours du cycle de culture. Les flux d'azote sont représentés par les différents flèches symbolisant l'absorption d'azote ainsi que les processus de remobilisation de l'azote des feuilles sénescentes vers les organes en croissance et les pertes d'azote liées à la chute des feuilles. Les phases principales du cycle sont indiquées avec la mise en place des composantes du rendement (DENS, plant density; ON, ovule number; PN, pod number; PSN, pod seed number; RN, ramification number; SN, seed number; SW, seed weight; SY, seed yield). La courbe en pointillés représente la dynamique de l'azote dans le sol au cours du cycle. Un itinéraire avec un fractionnement de la fertilisation azotée minérale en deux apports (à la reprise de végétation et 2-3 semaines après) est représenté ici. A l'automne, l'azote organique du sol est minéralisé et participe ainsi au stock d'azote minéral disponible pour la plante. Enfin, les différentes contraintes sont positionnées sous le cycle de culture: maladies (vert), ravageurs (violet), nutriments (noir; P-phosphore, K-potassium, N-azote, S-soufre), hydriques (bleu), radiatives (jaune) et thermiques (rouge).

aux dégâts d'insectes (ex. grosse altise) ou de la période hivernale (froid/gel) dont l'intensité peut être plus ou moins extrême. Dans certaines zones de la planète, l'implantation des cultures de colza ne dépasse pas 50-60% des graines semées, comme c'est souvent le cas en Australie (Nelson et al. 2022).

## Cycle de développement du colza d'hiver : phases clés et facteurs limitants

Sous nos latitudes, le colza a la particularité d'être semé à une période souvent sèche et d'avoir un cycle très long (~ 2400°CJ, degrés jour calculés en base 0°C), ce qui l'expose à de nombreux stress biotiques, abiotiques ou contraintes nutritionnelles (**Figures 5, 6**). Mais contrairement à d'autres espèces cultivées, dès lors que les conditions de production ne sont pas trop défavorables, il bénéficie d'une croissance dynamique à l'automne et d'une forte capacité de compensation au printemps. Les principales phases du cycle développemental du colza sont succinctement décrites ci-dessous et illustrées à la **Figure 5**.

De la germination à la levée. La durée de cette phase est variable et dépend principalement de l'humidité du sol et du contact graine-terre. Il est considéré que 7-10 mm de pluie sont nécessaires pour la germination. En conditions optimales, cette phase de germination-levée dure ~100°CJ (base 0°C), durée qui augmente en conditions sèches. Les graines de colza étant de petite taille (2-3 mm), elles sont semées dans un l'horizon très superficiel du sol (< 3 mm) où l'eau est très peu retenue. D'autres facteurs tels que la présence d'oiseaux ou de petits mammifères se nourrissant des graines en surface ou encore la formation de croûte de battance suite à des alternances sécheresse-pluies sont également très préjudiciables aux étapes de germination et de levée. En parallèle, un ensemble de maladies peuvent affecter le peuplement initial voire l'anéantir, telles que la fonte des semis (causée par *Pythium ultimum* ou *Rhizoctonia solani*) ou encore la hernie des crucifères (causée par *Plasmodiophora brassicae*) (Lamichhane et al. 2018; Nelson et al. 2022).

De la levée au stade 4-5 feuilles. C'est la phase critique pour le colza car, à cette période, la croissance est lente et la production de biomasse aérienne faible, la plante développant surtout son système racinaire. Le colza est alors sensible aux attaques de bioagresseurs (limaces, escargots, petites et grosses altises, taupins, chenilles ...). Cette phase dure environ ~400°CJ (base 0°C). Plus la levée est précoce, plus la plante bénéficie des températures élevées de la fin d'été, et plus la durée de cette phase sensible sera courte. Des événements de sécheresse peuvent survenir durant cette phase précoce, ce qui impacte également la croissance et le développement des jeunes plantes.

La phase de croissance végétative rapide. Passé le stade 4-5 feuilles, le colza entre en phase de croissance rapide et est beaucoup moins sensible aux bioagresseurs. L'enjeu consiste donc à favoriser une croissance continue jusqu'à l'entrée en repos végétatif (hiver), et donc éviter les facteurs



**Figure 6. Exemples de dégâts souvent observés sur des cultures de colza.**

Le colza est la cible de multiples bioagresseurs qui, combinés à des stress abiotiques (excès/manque d'eau, fortes/faibles températures, gel, ombrage...) et des contraintes nutritionnelles, vont fortement impacter l'implantation de la culture en début de cycle et avoir des répercussions sur la vigueur de la culture, sa sensibilité par rapport aux adventices, son efficience d'absorption et d'utilisation des nutriments, la rendant ainsi moins apte à exprimer son potentiel de rendement, et par conséquent à sécuriser la marge économique. Les stress peuvent affecter différentes phases du cycle (implantation précoce, croissance automnale, reprise de végétation en sortie d'hiver, floraison, remplissage de la graine).  
(Photos: Terres Inovia).

abiotiques et biotiques perturbateurs (ex. mauvais enracinement, concurrence des adventices, carence en azote ou en phosphore). A l'entrée de l'hiver, le colza aura constitué environ 20 % de sa biomasse aérienne et 70 % de sa biomasse racinaire considérée par rapport à la fin du cycle. Une croissance dynamique et continue à l'automne atténue les dégâts causés par les insectes adultes : les morsures de feuilles, et donc la proportion de surface verte détruite, sont négligeables sur de grosses feuilles. Par ailleurs, des pieds vigoureux, avec de gros pétioles et un gros collet, gênent les larves qui mineront plus difficilement les pétioles et qui auront plus de chemin à parcourir pour atteindre l'aisselle des feuilles et le cœur des plantes.

Le repos végétatif (pendant l'hiver). Cette phase est plus ou moins longue et marquée, principalement selon les conditions de température. Les risques majeurs pour la culture sont alors le gel, la grêle et/ou l'excès d'eau dans les parcelles.

La reprise de végétation en sortie d'hiver (montaison). Elle constitue une autre étape critique : l'augmentation des températures favorise l'activité des bioagresseurs (larves d'altises, charançons de la tige notamment). Le colza doit donc bénéficier d'une reprise rapide, favorisée par un bon enracinement et un statut de nutrition minérale adapté. Une reprise et une croissance dynamique en sortie d'hiver permettent, comme en automne, d'atténuer les dégâts causés au printemps par les larves d'insectes d'automne : l'objectif est que la croissance de la tige s'amorce avant que les larves n'atteignent le cœur des plantes.

La floraison. Cette phase, qui débute à ~1400°CJ après semis, va permettre de déterminer le nombre de siliques et de graines. La qualité de floraison dépend des conditions climatiques, et peut être impactée par les dégâts de larves d'insectes, d'automne notamment. Par exemple les charançons de la tige et les méligrèthes percent les boutons floraux, réduisant ainsi le nombre de siliques.

Le remplissage et la maturation des graines dure ~1000°CJ et correspond à la réalisation du potentiel de rendement (nombre de graines, poids de la graine). Les paramètres de rendement peuvent être altérés par la présence de ravageurs (charançons des siliques, cécidomyies, pucerons cendrés) ou d'agents pathogènes responsables de maladies sur la silique et/ou la graine (alternaria, oïdium, sclerotinia). Ces attaques conduisent à des pertes de rendement plus ou moins lourdes (avortement de graines).

## Implantation de la culture de colza et notion de vigueurs

La qualité d'implantation d'une culture est le résultat d'une interaction complexe entre le génotype, l'environnement pédo-climatique et les pratiques agronomiques, avant, pendant et après le semis (Whittington 1973 ; Hodgkin et Hegarty 1978). On l'a vu précédemment, une bonne implantation du



colza avant l'hiver assure une culture robuste, moins sensible aux ravageurs et aux adventices, économique en intrants à cette période et apte à exprimer son potentiel de rendement et à sécuriser la marge économique (Finch-Savage 1995). A l'inverse, une mauvaise implantation représente un coût d'ensemencement important pour les producteurs, en particulier de colza hybride, et par ailleurs, impacte significativement la biomasse produite, le potentiel de rendement et la compétitivité vis-à-vis des adventices (Sausse et al. 2016). La qualité de l'implantation de la culture de colza a donc des répercussions déterminantes sur tout le cycle, d'où l'intérêt d'y porter la plus grande attention. En effet, l'optimisation des premières phases de développement pourrait permettre de maintenir une densité optimale de la culture, d'améliorer la tolérance aux stress et ainsi de limiter certaines pertes de rendement en graines. Une fois l'implantation réussie, la culture a une bonne capacité de compensation, ce qui lui permet de limiter l'impact des stress qui ont lieu après cette phase. L'amélioration de l'implantation nécessite d'une part l'adaptation des pratiques mises en œuvre pendant l'interculture et lors des années précédentes sur la parcelle, et d'autre part la fourniture aux agriculteurs de variétés vigoureuses aux différentes étapes clé du cycle et plus particulièrement dans les phases précoce et ce, quels que soient les événements météorologiques pouvant être parfois extrêmes.

La vigueur caractérise la performance de la plante, c'est-à-dire sa capacité à croître, se développer et se reproduire, dans une large gamme de conditions environnementales. Ce trait fait partie des traits importants à sélectionner, en particulier pour le développement de variétés adaptées à des systèmes agroécologiques (Gouleau et al. 2021). Cependant, c'est un trait intégratif car composé d'une multitude de traits et donc difficile à définir. Dans le meilleur des cas, son évaluation est basée sur l'estimation de la biomasse aérienne produite ou de la surface foliaire développée ainsi que de leur dynamique de mise en place. Pour des contraintes de moyens, la vigueur est fréquemment phénotypée à l'aide d'une échelle de notation de 1 à 9. Cependant, on peut aussi prendre en compte d'autres critères plus difficiles à appréhender (ex. port de la plante, surface foliaire) ou à accéder (ex. architecture racinaire). Depuis quelques années, des méthodes de phénotypage basée sur l'estimation du taux de couverture à l'aide de prises de photos RGB par drone ont été développées par Terres Inovia. Cependant ces différentes méthodes d'estimation ne permettent pas complètement d'identifier des traits qui sont reliés aux processus biologiques impliqués dans l'élaboration de la vigueur. Pour identifier de tels traits, il est important d'une part de séparer les différentes phases du cycle où se mettent en place les composantes de la vigueur, et d'autre part d'identifier les principaux processus impliqués pour chacune de ces phases.

Pour clarifier la sémantique utilisée dans le manuscrit, il sera question de vigueur germinative à propos de la phase de germination-levée, de vigueur précoce pour la phase de croissance qui suit (jusqu'à 4-



## *Contexte*

---

5 feuilles), de vigueur automnale pour toute la phase de croissance rapide et de vigueur à la reprise pour la phase de sortie d'hiver.



## Contexte de la thèse

---

Au cours de ma thèse, j'ai cherché à identifier et caractériser les déterminants génétiques de la germination et de la vigueur précoce chez le colza. De plus, j'ai exploré le potentiel de la prédiction phénomique pour valoriser la sélection de ces traits, notamment dans des étapes de pre-breeding.

J'ai réalisé ma thèse au sein de l'équipe 'Réponses sous Contraintes Abiotiques' (RCA) de l'Institut de Génétique, Environnement et Protection des Plantes (IGEPP). Ce travail a été co-dirigé par Nathalie Nesi, directrice de recherche INRAe, et Anne Laperche, maître de conférences à l'Institut Agro Rennes-Angers. Les recherches de l'équipe RCA visent principalement à décrire et comprendre l'impact des stress environnementaux sur la croissance du colza, l'élaboration de son rendement et la qualité des graines. La recherche des traits phénotypiques favorisant l'acclimatation de la culture dans un contexte limitant et l'exploration de la variabilité génétique associée permettront d'identifier de nouveaux leviers pour l'amélioration variétale.

Ma thèse a été cofinancée par le département Biologie et Amélioration des Plantes d'INRAe et le projet NAVIG (2021-2024, PROMOSOL - association pour la promotion de la sélection des plantes oléagineuses). Le projet NAVIG a pour ambition d'identifier les déterminants génétiques et physiologiques qui contrôlent les phases initiales de l'implantation d'un couvert de colza et de valider que l'expression de ces déterminants précoces confère un avantage significatif à la culture sous différentes pratiques culturales (variation de la date de semis, association avec une légumineuse gélive, zéro phyto). Outre le financement d'une demi allocation doctorale, NAVIG a permis de soutenir le fonctionnement de ma thèse (expérimentations, recrutement de personnel, missions). Dans ce cadre, j'ai participé d'une part aux réunions de suivi du projet et d'autre part aux Carrefours Colza organisés annuellement avec l'ensemble de la communauté Promosol (représentants des sélectionneurs français de Colza, de Terres Inovia et de Terres Univia). J'ai ainsi exposé mes résultats lors des Carrefours Colza 2023 et 2024.

Ce manuscrit comprend :

- Une synthèse bibliographique (chapitre I) qui vise à présenter les processus majeurs pour l'implantation des cultures, les potentiels traits impliqués dans la germination et la vigueur précoce, les ressources génétiques et génomiques disponibles chez le colza et les méthodes de sélection pour cette espèce. Le chapitre I s'achève par la présentation des questions de recherche traitées dans la thèse ainsi que de la stratégie déployée pour y répondre.
- Un article (chapitre II) publié dans *Theoretical and Applied Genetics* en 2024 qui décrit l'identification de régions génomiques impliquées dans la germination ainsi que la capacité à



prédire ce trait à partir de régresseurs génomiques et spectraux. Ces résultats ont également fait l'objet d'une présentation orale à l'International Rapeseed Conference (IRC) en Septembre 2023 à Sydney, Australie.

- Un troisième chapitre, subdivisé en deux parties majeures :
  - o La partie III.2 est dédiée à l'identification de régions génomiques associées à la vigueur précoce en considérant un ensemble de traits fonctionnels. L'architecture génétique de la vigueur précoce est investiguée et l'impact de la sélection sur la diversité génétique disponible pour ces traits est discuté par rapport à des utilisations en pre-breeding. Cette partie est rédigée sous forme d'un article à soumettre prochainement à la revue 'BMC Genomics'.
  - o La partie III.3 présente des résultats sur les capacités des modèles de prédictions génomiques et phénotypiques pour les traits de vigueur précoce identifiés précédemment.

Les résultats présentés au chapitre III ont été en partie présentés sous forme de posters à l'IRC 2023 et à la 7<sup>ème</sup> International Conference of Quantitative Genetics en Juillet 2024 à Vienne, Autriche.

- Une discussion générale (chapitre IV) qui resitue les principaux résultats de ma thèse au regard de la bibliographie et aborde les liens entre germination et vigueur précoce. Des perspectives de ce travail à court et plus long terme sont proposées, notamment en lien avec l'origine de la diversité génétique et son utilisation dans le cadre de programmes de sélection.
- Une annexe qui présente mes activités scientifiques (article, communications dans des colloques, formations, activités d'enseignement).

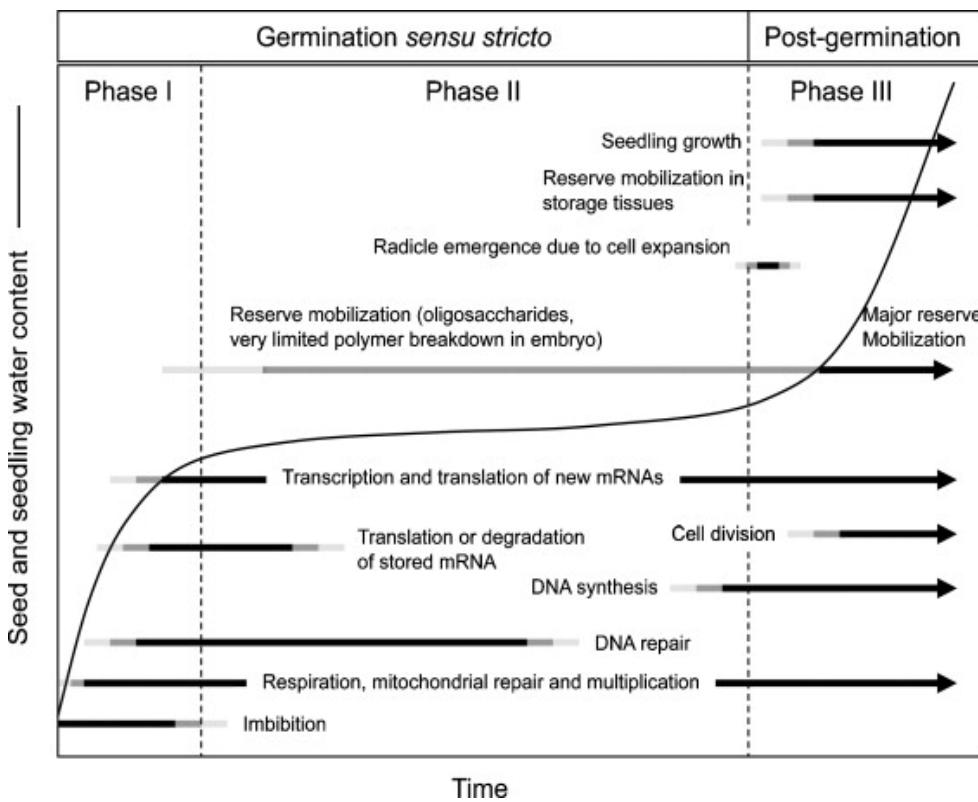


---

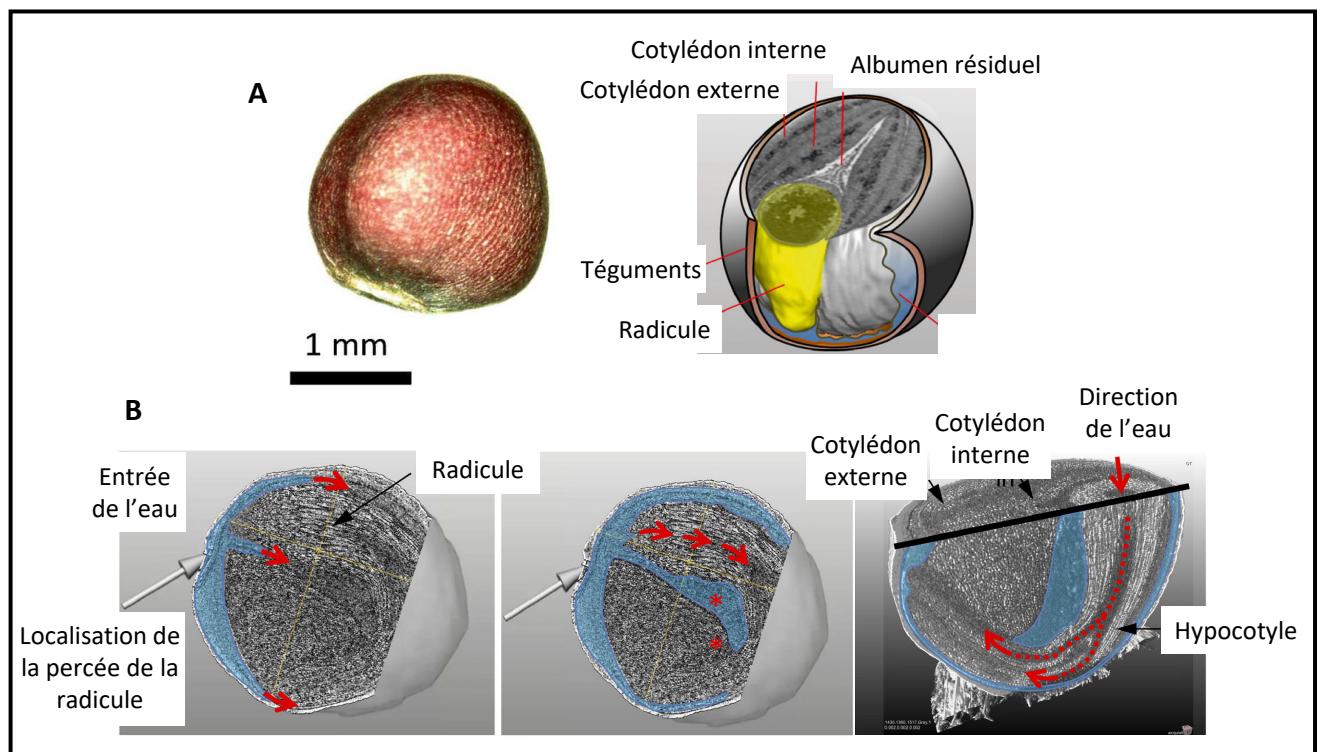
## **CHAPITRE I**

### **SYNTHESE BIBLIOGRAPHIQUE**

---



**Figure I.1.** Principaux processus physiologiques, moléculaires et métaboliques associés à la germination de la graine (d'après Nonogaki et al. 2010).



**Figure I.2.** Organisation de la graine de colza (A) et entrée de l'eau dans la graine lors de l'imbibition (B) (d'après Munz et al. 2017).

Les zones bleues représentent les tissus hydratés. Les flèches rouges indiquent l'entrée d'eau et son cheminement dans les différents tissus de la graine.

## I.1. Processus physiologiques de l'implantation du colza et vigueur précoce

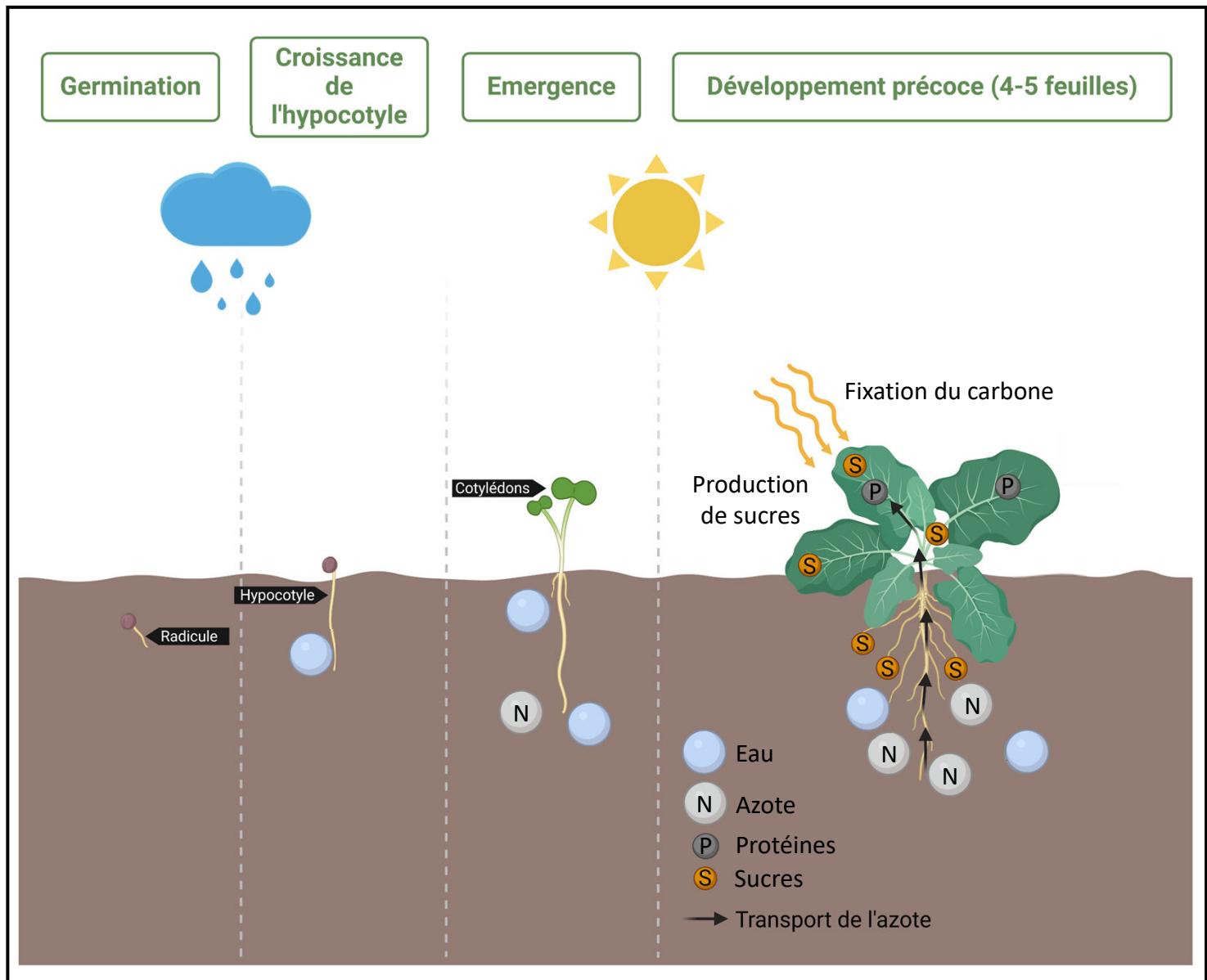
---

Dans cette section nous décrivons les processus physiologiques engagés dans les principaux stades de développement précoce d'un couvert, de la germination au stade 4-5 feuilles, considéré comme un stade critique chez le colza (*cf. Contexte général*), puis nous porterons une attention particulière aux leviers qui offrent des solutions pour garantir une bonne implantation de la culture.

### I.1.1. Germination de la graine

Les graines orthodoxes sont caractérisées par une importante déshydratation au cours de leur maturation, leur teneur en eau finale variant entre 4 et 15% de la matière sèche. En imposant un arrêt quasi-total des réactions métaboliques, cet état sec est essentiel à la survie de la graine pendant la période défavorable à l'établissement de la génération suivante. La germination *sensu stricto* débute par une prise d'eau pendant la phase d'imbibition et s'achève avec la percée des téguments par la radicule (Bewley 1997). La cinétique d'accumulation de l'eau dans les graines permet de diviser la germination en deux phases après le début de l'imbibition (**Figure I.1**). La première se caractérise par une prise d'eau rapide et massive qui permet de réinitier un grand nombre de processus métaboliques telles que la respiration, la réparation de l'ADN, la transcription, la traduction et la dégradation d'ARN messager. La seconde phase se caractérise par une stabilisation de la quantité d'eau dans la graine, du fait de contraintes mécaniques imposées par les téguments (**Figure I.1**). La mobilisation des réserves est enclenchée et soutient une phase de multiplications et d'expansion cellulaires intense qui conduiront *in fine* à la percée des téguments par la radicule en croissance. Après cette phase de germination *sensu stricto*, la quantité d'eau dans les graines augmente à nouveau et la mobilisation des réserves énergétiques supportent le développement de la plantule (Nonogaki et al. 2010). En plus de la présence d'eau, d'autres stimuli tels que la température, la lumière, et la photopériode participent à promouvoir la germination. L'effet de ces stimuli se traduit par une cascade de signalisation qui met en jeu l'action d'hormones, avec un rôle majeur des gibbérellines, de l'éthylène et de l'acide abscissique (Groot and Karssen 1987; Sun and Gubler 2004; Rajjou et al. 2012; Corbineau et al. 2014).

Chez le colza, la prise d'eau au moment de l'imbibition est bien décrite (Munz et al. 2017 ; **Figure I.2**) L'eau pénètre dans la graine via un faisceau vasculaire localisé au niveau du hile (cicatrice du point d'attache de la graine sur la plante mère) et réhydrate dans un premier temps la radicule puis les cotylédons. La réactivation du métabolisme suit spatialement la réhydratation des tissus (Munz et al.



**Figure I.3. Développement précoce chez le colza, de la germination au stade 4-5 feuilles (d'après Chen et al. 2016 et Nelson et al. 2022).**

Lors du stade autotrophe, l'azote absorbé par la plante via les racines permet la synthèse de protéines essentielles pour l'expansion cellulaire et la photosynthèse. Le carbone fixé lors de la photosynthèse est transformé en sucres et est transporté vers les zones de croissance.

2017). La gamme de températures optimales pour assurer la germination du colza est relativement large et comprise entre 12 et 33°C (Dürr et al. 2016; Derakhshan et al. 2018).

### I.1.2. Emergence de la nouvelle plantule

Après la germination et la croissance de la radicule, c'est l'hypocotyle qui se développe. Le colza a une germination qualifiée d'épigée, c'est-à-dire que c'est la croissance de l'hypocotyle qui permet la sortie de terre des cotylédons (**Figure I.3**). Au cours de cette phase hétérotrophe, les réserves contenues dans les cotylédons (huile, protéines, sucres) sont mobilisées pour assurer la croissance et le développement de la jeune plantule (Penfield et al. 2005; Cernac et al. 2006). Ces réserves ont été accumulées sur la plante mère lors de la phase de remplissage-maturation des graines ; leur teneur et leur qualité dépendent donc principalement du génotype de la plante mère mais également des conditions environnementales au moment de la formation et du remplissage des graines, et des interactions entre le génotype de la plante et l'environnement (GxE). Lors de l'émergence, les cotylédons s'ouvrent et verdissent, permettant ainsi à la plantule d'initier la photosynthèse, avec une efficacité relative dans un premier temps. La formation de la première feuille a lieu et représente le passage au stade autotrophe (*cf. I.1.3*). Parallèlement, la jeune racine croît et développe des radicelles afin de capter le maximum de ressources hydriques et nutritives dans le sol. La racine pivot atteint généralement entre 2 et 5 cm lors de l'émergence de la première feuille. L'exploration du sol par la racine permet également le recrutement d'un microbiote adéquat (Jones et al. 2009).

### I.1.3. De la levée au stade 4-5 feuilles, la phase délicate du développement précoce

La phase de développement précoce entre le stade 1 feuille et le stade 4-5 feuilles se caractérise par une production lente de biomasse, due à la mise en place des structures (feuilles). Le développement précoce requiert approximativement 400 degrés jours (°CJ, calculés en base 0°C) (Hebinger 2013) avec un phyllochron moyen de 75°CJ (Triboi-Blondel et al. 1988). Le développement de la plante est régulé par un ensemble d'interactions entre les flux d'azote et de carbone, permettant la croissance des parties aériennes et racinaires (Hellebust and Bidwell 1963).

L'azote est principalement absorbé sous forme de nitrate ( $\text{NO}_3^-$ ) et d'ammonium ( $\text{NH}_4^+$ ) par les racines (Crawford and Forde 2002). La vitesse d'absorption de l'azote peut être modulée par la quantité d'azote présent dans le milieu, ainsi que la quantité de sucres présents dans les racines (Delhon et al. 1996). L'azote absorbé permet majoritairement l'expansion cellulaire, via la synthèse de protéines impliquées dans la paroi cellulaire et le cytosquelette (Radin and Parker 1979; Lawlor et al. 1988). De plus, l'azote est impliqué dans la formation des photosystèmes et d'enzymes telles que la Rubisco afin

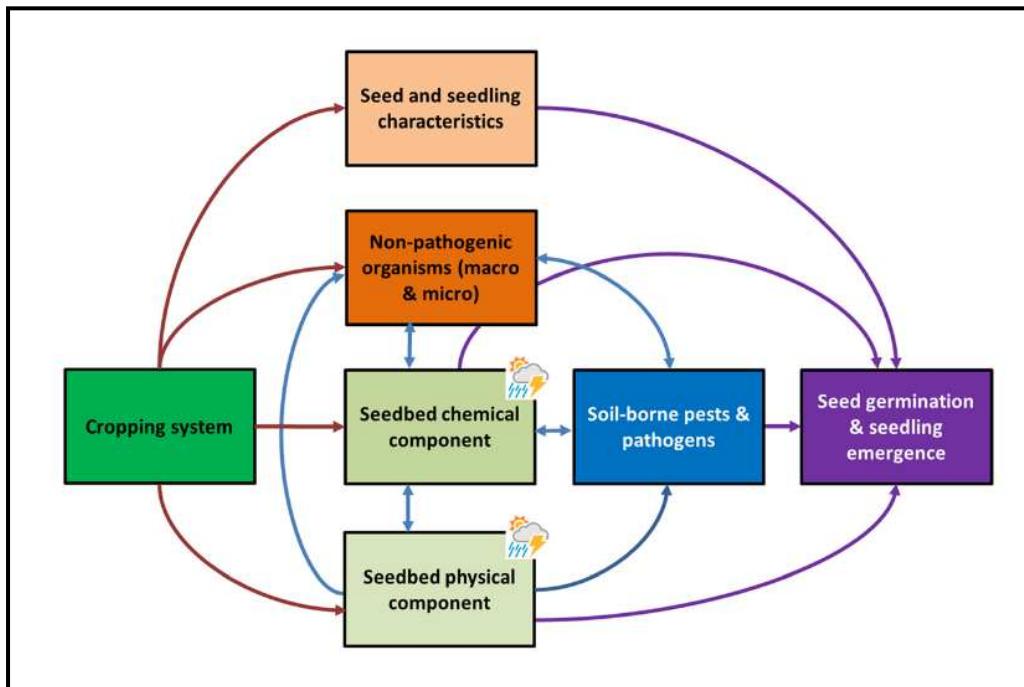


d'absorber de la lumière et de fixer du CO<sub>2</sub>. L'azote contenu dans les feuilles joue donc un rôle crucial dans la photosynthèse, étant impliqué dans la production d'énergie au sein des thylakoïdes et dans la production de sucres (**Figure I.3**). Le carbone nouvellement assimilé est transporté sous forme de saccharose via le phloème vers les zones de croissance (Dale 1982; Schnyder et Nelson 1987), que ce soit dans le compartiment aérien ou racinaire. Dans les feuilles en croissance, un transport de carbone est nécessaire afin de démarrer l'activité photosynthétique. Cette activité augmente progressivement jusqu'à ce que la feuille devienne pleinement autotrophe. Ainsi, le carbone est indispensable pour la croissance des cellules, car il permet de leur fournir les sucres nécessaires à la biosynthèse des parois et à la production d'énergie.

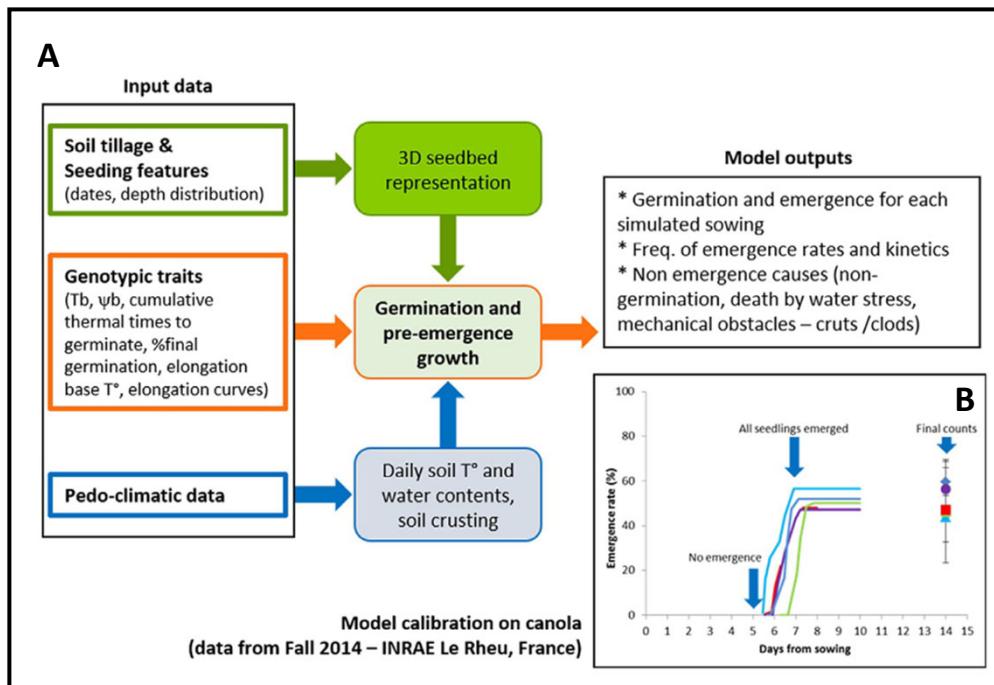
De manière générale, les flux de carbone et d'azote sont interconnectés. Ainsi, la distribution de l'azote dans les feuilles influe sur la capacité photosynthétique de la plante ainsi que sur le développement de la surface aérienne, impactant ainsi l'assimilation de carbone et par conséquent l'élaboration de la biomasse et la croissance (Gastal et Lemaire 2002). L'allocation du carbone et de l'azote dans les compartiments peut dépendre de la disponibilité de ces éléments, permettant à la plante d'équilibrer ses besoins face aux ressources disponibles dans un environnement donné. Ainsi, une carence en azote peut mener à une priorisation de l'allocation de la biomasse vers les racines afin d'augmenter le volume de sol exploré.

#### I.1.4. Leviers pour garantir une implantation réussie de la culture

Une implantation de culture réussie est le résultat d'un taux de germination élevé, d'une germination rapide et homogène, et d'un développement rapide des systèmes aérien et racinaire, et cela dans une large gamme de conditions environnementales. La capacité à germer et à produire une plantule normale dans des conditions suboptimales qualifie la vigueur des semences (Finch-Savage et Bassel 2016). Les paramètres de mise en place du couvert végétal qualifient quant à eux la vigueur précoce, c'est-à-dire la performance de la plante durant la phase précoce de son cycle de développement. Une plantule vigoureuse dès les premières phases du cycle serait plus apte à tolérer des stress ou des contraintes. Ainsi, chez le riz une corrélation positive est observée entre la biomasse à un stade précoce et le rendement en condition de compétition avec des adventices (Namuco et al. 2009). De même, des lignées de blé caractérisées par des feuilles plus larges et des racines plus longues que d'autres montrent une compétitivité supérieure dans un peuplement avec du ray-grass (Hendriks et al. 2024). Chez le colza, il semble que la taille de la graine ainsi que la biomasse et la vitesse de développement des parties aériennes dès l'apparition de la première feuille impactent positivement la tolérance du peuplement face à des attaques d'altises (Elliott et al. 2008). Plusieurs leviers peuvent être mobilisés et combinés pour favoriser l'implantation d'un couvert végétal, en considérant d'une



**Figure I.4. Facteurs biotiques, abiotiques et leurs interactions, affectant la germination et la levée des plantules (d'après Lamichhane et al. 2018).**



**Figure I.5. Principe du modèle SIMPLE pour simuler l'émergence des plantes (d'après Durr et al. 2016).**

(A) Les paramètres d'entrée du modèle concernent les variables agronomiques (travail du sol et pratiques culturales) qui permettent de construire le module « sol » (vert), les variables pédoclimatiques du lieu qui permettent de construire le module « environnement » (bleu; il s'agit en fait du modèle STICS qui est implémenté) et des variables liées à la germination de la plante considérée qui permettent de construire le modèle « germination – émergence » (orange). (B) Le modèle génère des courbes de germination et d'émergence pour l'environnement et la culture ciblé. Il permet également de fournir les causes de non-émergence. Le calibrage du modèle sur colza au Rheu (essais 2014) a été réalisé en ajustant des données acquises au champ par les expérimentateurs (notations de levées, nombre de plantes dans le peuplement final). Le modèle est calibré sur le colza mais ne permet cependant pas une distinction de la variabilité génotypique (Durr et al., 2016).

part la phase de germination-levée et d'autre part la phase de développement précoce. Ces leviers peuvent être de nature agronomique (système de culture, conduite de la culture), génétique (choix variétaux), environnementale (conditions de développement des plantes et de production des graines) ou technologique (conditions de récolte et de stockage des graines).

Sur le plan agronomique, il s'agira de réduire, ou du moins de limiter les facteurs physiques, chimiques et biologiques du sol qui peuvent impacter la réussite des phases de germination-levée et de développement précoce (**Figure I.4**, pour revue voir Lamichhane et al. 2018). A titre d'exemple, nous pouvons citer, une mauvaise structure du sol qui limite l'imbibition des graines ou impose des obstacles mécaniques tels que la formation de croute de battance dans certaines conditions environnementales. De même, la présence d'agents pathogènes dans le sol (bactéries, champignons) cause des levées faibles du fait d'une fonte de semis (Lamichhane et al. 2017, cf. **Contexte général**). Ces facteurs limitants peuvent être contrôlés, voire réduits, via un système de culture et/ou des pratiques culturales adaptés. Ainsi, le choix du précédent cultural permettra de rompre les cycles de développement de certains agents pathogènes et de réduire ainsi la pression en inoculum (Davis and Nunez 1999) mais également de moduler la structure du sol grâce au développement de systèmes racinaires (Chan and Heenan 1996), voire d'enrichir le sol en azote comme c'est le cas avec la culture de légumineuses (St. Luce et al. 2015). Le travail du sol en amont du semis est également un élément majeur pour moduler sa structure et sa composition (Aubertot et al. 1999). Le labour aura des impacts différents de ceux des techniques culturales simplifiées sur la formation de mottes, d'une croute de battance mais également sur la teneur en eau du sol en profondeur (Verhulst et al. 2010). Enfin, la date, la densité et la profondeur de semis sont également des déterminants majeurs pour la réussite de l'implantation. A titre d'exemple, des semis précoces permettent, si les conditions météorologiques s'y prêtent, d'avancer la saison de culture, rendant ainsi les plantes moins vulnérables lors des attaques d'insectes à l'automne (Dejoux et al. 2003). La profondeur de semis est également essentielle : des graines semées profondément, bénéficient d'un meilleur environnement hydrique (teneur en eau plus élevée dans les horizons plus profonds au moment du semis) mais doivent cependant présenter une meilleure croissance de l'hypocotyle pour émerger. La plupart de ces facteurs sont pris en compte dans le modèle SIMPLE qui simule la levée d'une culture (Dürr et al. 2001 ; **Figure I.5**). Adapté au colza, SIMPLE permet de prédire la levée dans un grand nombre d'environnements pédo-climatiques et d'expliquer les causes probables de la non levée (Dürr et al. 2016). Les limites de ce modèle résident dans le fait qu'il ne prend pas en compte la dimension biotique des facteurs limitants l'implantation.

La variabilité génétique apparaît comme une piste prometteuse pour améliorer la vigueur chez le colza (Finch-Savage and Bassel 2016; Nelson et al. 2022). Pour cela, il est dans un premier temps nécessaire



de mieux caractériser phénotypiquement les premières phases de développement pour les conditions environnementales ciblées, identifier les traits d'intérêt, et caractériser la variabilité génotypique disponible. L'étude du déterminisme génétique de ces traits pourra ensuite être réalisée afin d'identifier des régions génomiques d'intérêt pour la sélection variétale. Certains traits d'intérêt ont déjà été identifiés pour caractériser la capacité germinative et la vigueur précoce. Concernant la capacité germinative, les principaux traits sont le taux de germination final, la vitesse de germination et l'homogénéité de germination. Quant à la vigueur précoce, plusieurs traits cibles ont été identifiés tels que la biomasse aérienne et la biomasse racinaire, la surface foliaire, la hauteur des plantes ainsi que la surface foliaire chez les céréales (Regan et al. 1992; López-Castañeda et al. 1995) ou encore chez le riz (Cairns et al. 2009; Shi et al. 2020). Cependant, concernant la vigueur précoce, il est important de bien comprendre le fonctionnement de la plantule pour pouvoir identifier de nouveaux traits d'intérêt, basés sur les processus physiologiques et écophysiologiques mobilisés lors de l'élaboration de la vigueur (acquisition de nutriments, développement (Ryan et al. 2015, Liao et al. 2004; Pang et al. 2013).

Les conditions environnementales au moment du développement et de la maturation des graines sur la plante mère ainsi que les conditions de stockage des lots de semences impactent la viabilité, la capacité germinative et la longévité. Ces facteurs sont souvent globalisés et qualifiés d'effet 'lot de semences'. Concernant l'impact de la période de formation des graines, un effet transgénérationnel peut être observé dans certaines conditions. Ainsi, Hatzig et al. (2018) rapportent que des graines ayant subi des conditions de stress hydrique pendant la maturation produisent une biomasse précoce à 7 jours après semis plus élevées que les graines n'ayant pas subi de stress. L'hypothèse de régulations épigénétiques qui agiraient comme un effet mémoire ('imprinting') est avancée pour expliquer cet effet transgénérationnel (Bilichak and Kovalchuk 2016; Lämke and Bäurle 2017; Gallusci et al. 2017).

La qualité de la semence, estimée non seulement par sa capacité à germer mais également à donner une plantule normale et vigoureuse, dépend également du stade de maturité des graines lors de la récolte. En effet, cette qualité est maximale au stade de maturité physiologique (le stade où la qualité de la graine est maximale, Finch-Savage et Bassel 2016). Elle décroît sensiblement lorsque la graine est récoltée trop tôt. La dégradation mécanique des graines (et plus particulièrement des téguments) par les outils mécaniques de récolte est également source d'une qualité moindre de la semence.

Enfin, s'agissant de l'effet du stockage post-récolte, les graines vieillissent et leur capacité germinative diminue naturellement avec le temps mais cette diminution de la qualité est significativement accélérée par les paramètres environnementaux (taux d'hygrométrie, température, taux d'oxygène).



Des mauvaises conditions de stockage conduisent à des dégradations cellulaires dans la graine qui altèrent *de facto* ses capacités de germination et sa viabilité (Zinsmeister et al. 2020; Buitink and Leprince 2024).



## I.2. Identification de traits cibles pour la germination et la vigueur précoce

---

### I.2.1. Décomposition de processus intégratifs en traits fonctionnels

Les processus tels que la germination et la vigueur précoce permettent de qualifier la performance (fitness) d'un individu ou d'une population *i.e.* sa capacité à croître, se reproduire et survivre. Ils sont dits intégratifs car ils résultent d'un grand nombre de traits fonctionnels sous-jacents et de leurs interactions, ces traits étant eux-mêmes pour la plupart complexes car quantitatifs car leur variation est continue et polygéniques. Une approche pour appréhender ces processus intégratifs et en caractériser la diversité existante consiste en premier lieu à disséquer les traits fonctionnels sous-jacents. La définition de ces traits fonctionnels est peu homogène dans la littérature. Dans cette étude, nous utilisons la définition de trait fonctionnel proposée par Arnold (1983) et adaptée aux plantes par Violle et al. (2007). Selon ces auteurs, les traits fonctionnels sont des traits développementaux, morphologiques ou physiologiques mesurés à l'échelle des individus qui influencent *in fine* leur performance individuelle. Décomposer des processus intégratifs en plusieurs traits fonctionnels est un moyen d'accéder à des données phénotypiques qui permettront de réaliser des analyses génétiques dans une population de diversité et ainsi d'en identifier et caractériser les déterminants génétiques et moléculaires (Cooper et al. 2009, Romero Navarro et al. 2017).

Les traits développementaux (ou phénologiques) sont les traits qui se rapportent à la chronologie du cycle de vie de la plante. Il peut s'agir d'un événement ponctuel ou périodique comme l'apparition d'un organe ou d'un événement inscrit dans une durée déterminée par deux points du cycle comme le phyllochrome (Rickman and Klepper 1995). Les traits morphologiques comprennent ceux qui décrivent la forme de la plante et son architecture tridimensionnelle (Drouet et Pagès 2003; Bucksch et al. 2017). Par exemple, la longueur ou la largeur d'un organe, la couleur d'un tissu. Les traits physiologiques sont des traits reflétant les fonctions biologiques de la plante. Par exemple, la teneur en métabolites d'un organe, la biomasse. Ces traits peuvent être définis comme des traits d'efficience lorsqu'ils caractérisent la capacité de la plante à utiliser des ressources pour son développement. Par exemple, la capacité à assimiler le carbone ou la capacité à utiliser de l'azote pour faire de la biomasse (Gaju et al. 2011; Pang et al. 2013). Les traits d'efficience présenteraient deux avantages majeurs : i) selon certains auteurs, le déterminisme génétique de ces traits serait potentiellement plus simple (Reymond et al. 2004) et ii) ces traits seraient moins impactés par l'effet environnemental et les interactions Génotype × Environnement (Chenu et al. 2018).

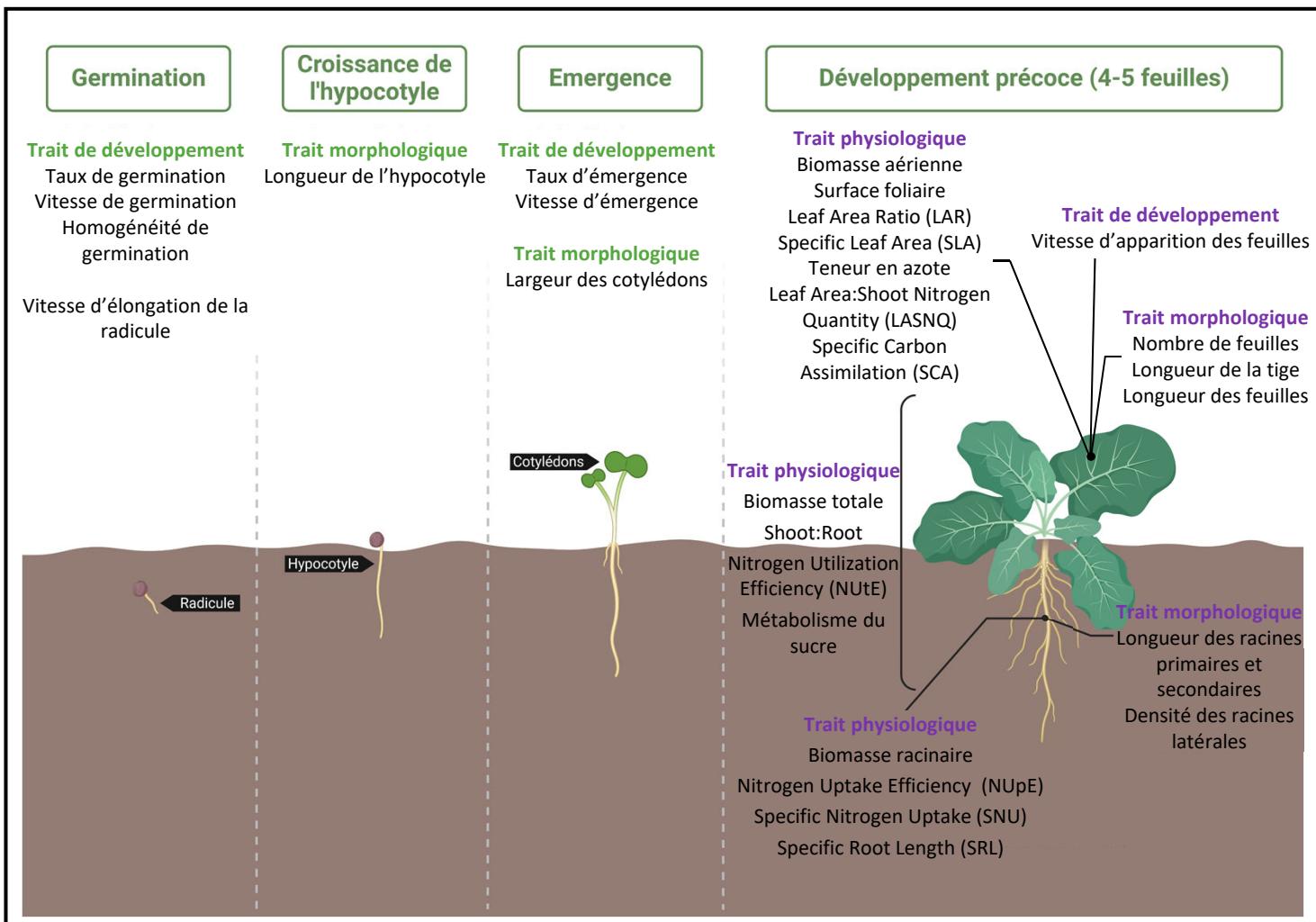


Figure I.6. Traits fonctionnels potentiels à étudier pour caractériser la germination et les stades précoces du développement chez le colza (adapté de Nelson et al. 2022).

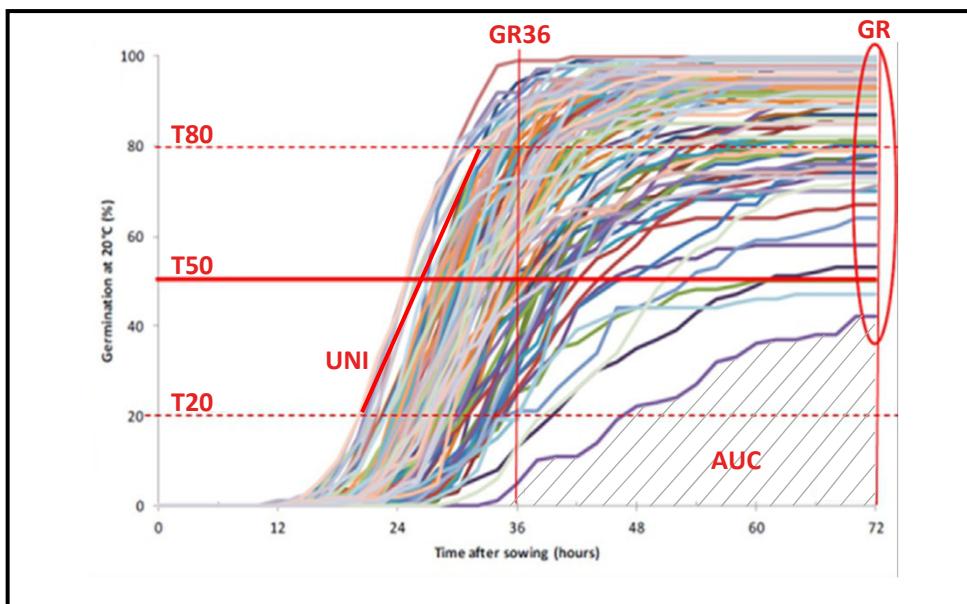
Des modèles de croissance peuvent être utilisés pour identifier des variables impliquées dans les traits à décomposer. Ces modèles, spécifiques ou non d'une espèce, permettent d'identifier et de hiérarchiser les traits fonctionnels impliqués dans la fluctuation du processus ciblé et de caractériser les relations entre ces traits (Di Paola et al. 2016). Parmi ces modèles ont été développés des modèles généraux décrivant l'interaction entre la plante, son environnement et les pratiques culturales (tels que APSIM, Holzworth et al. 2014 ou SIMPLE, Dürr et al. 2001) ainsi que des modèles structure-fonction décrivant le développement de la structure de la plante et son fonctionnement écophysiologique (tels que GreenLab, Yan et al. 2004; Vos et al. 2010). C'est à partir de ces modèles que du phénotypage assisté par modèle peut être réalisé afin de choisir les traits à phénotyper ou afin d'accéder à des traits difficilement mesurables manuellement (Martre et al. 2015).

Les différents traits potentiels à investiguer et leur catégorie sont résumés pour chaque stade de développement par la **Figure I.6**.

### 1.2.2. Les traits fonctionnels liés à la germination ou à la levée

Un ensemble de traits de développement peuvent être déduit à partir de courbes de germination (**Figure I.7**). Parmi ces traits, le taux de germination représente la proportion de graines réussissant à germer dans des conditions données à 36h ou au temps final. Chez le colza, le taux de germination à 36h correspond au moment pour lequel est observé la plus grande variabilité phénotypique (Hatzig et al. 2015; Gu et al. 2016). Les courbes permettent également d'estimer la vitesse de germination après le début de l'imbibition, (T50, Dias et al. 2011) ainsi que l'uniformité de la germination du lot. Enfin, les courbes permettent également d'estimer l'aire sous la courbe (AUC) de germination pour chaque génotype, représentant le taux de germination cumulé à partir du début de l'imbibition. Les courbes de germination peuvent être acquises de manière automatisées sur des bancs de germination, permettant ainsi de cribler de larges populations (Ducournau et al. 2004; Wagner et al. 2011). De plus, la capacité germinative dépend également des conditions de germination. Ainsi, de la diversité génétique a été mis en lumière pour des paramètres du modèle SIMPLE (Dürr et al. 2001), chez *Medicago truncatula*, tels que la température et le potentiel hydrique de base pour la germination (Brunel et al. 2009).

La quantité de malate et d'aspartate, sont identifiées comme des marqueurs métaboliques décrivant la performance de la germination (Boter et al. 2019) et constituent donc des traits physiologiques associés à la vigueur. La balance hormonale entre l'acide abscissique, les gibberellines et l'acide indole acétique à des temps spécifiques explique aussi des différences de comportement entre génotypes. Lors de la germination, la reprise du métabolisme engendre de l'activité enzymatique, de l'activité respiratoire, ainsi que la remobilisation de la réserve lipidique. L'étude du protéome durant la



**Figure I.7. Courbes de germination pour un panel de diversité de colza en conditions non limitantes.**

Un ensemble de paramètres sont obtenus à partir de ces courbes: les temps pour obtenir 20% (T20), 50% (T50), 80% (T80) de germination, le taux de germination à 36h (GR36) et final (GR), ainsi que l'uniformité de germination du lot (UNI) et l'aire sous la courbe de germination (AUC). Adapté de Nelson et al., 2022.

germination a permis de mettre en lumière un ensemble de protéines et d'enzymes synthétisées durant la germination (Galland et al. 2014). Etudier la variabilité de ces processus permettrait d'identifier de nouveaux traits physiologiques permettant de décrire la capacité germinative.

Après la germination, d'autres traits morphologiques et de développement peuvent être étudiés afin de caractériser l'implantation, tels que la vitesse d'élongation de la radicule (Brunel-Muguet et al. 2011; Nelson et al. 2022), le taux et la vitesse d'émergence (Dürr et al. 2016). Nelson et al. (2022) proposaient également d'étudier la longueur de l'hypocotyle. Une radicule plus grande pourrait également permettre à la plantule d'aller chercher plus rapidement des nutriments. Enfin, les auteurs proposaient également d'étudier la largeur des cotylédons (**Figure I.6**). Les cotylédons représentent le premier tissu photosynthétique de la plante. Ainsi, les auteurs suggèrent que de larges cotylédons pourraient permettre à la plantule d'accéder à plus de ressources et donc de se développer plus rapidement.

### 1.2.3. Les traits morphologiques et d'architecture liés à la vigueur précoce

Un ensemble de traits morphologiques peuvent être utilisés pour décrire le développement des parties aérienne et racinaire. Parmi ces traits, le nombre de feuilles chez le riz (Rebolledo et al. 2015) ainsi que la longueur des racines primaires et secondaires (Koscielny and Gulden 2012; Thomas et al. 2016) ont été utilisés pour décrire la vigueur précoce, en termes de capacité à produire de la biomasse et à explorer le sol (**Figure I.6**). Concernant l'architecture racinaire, la densité des racines latérales a été identifiée comme un trait cible pour l'adaptation de la culture de colza en condition de faibles apports en azote (Lecarpentier et al. 2021). Ainsi, la diminution de la densité en racines latérales permettrait d'explorer de manière plus optimale le sol en profondeur (Lecarpentier et al. 2021).

### 1.2.4. Les traits physiologiques candidats pour expliquer la vigueur précoce

Les traits les plus utilisés pour étudier la croissance de la plante et la vigueur précoce sont des traits de biomasse et de surface foliaire. Ces traits ont notamment été utilisés pour caractériser le développement précoce chez *Arabidopsis thaliana* (Bac-Molenaar et al. 2015), *Brassica oleracea* (Bettey et al. 2000), le riz (Rebolledo et al. 2015) et le colza (Nguyen et al. 2018; Knoch et al. 2020). Un grand nombre de métabolites semblent contribuer au développement précoce. Notamment, les métabolites impliqués dans le métabolisme du glucose ont été identifiés comme participant particulièrement au développement, dû à leur rôle dans la formation des parois cellulaires et dans le transport du carbone (Meyer et al. 2007; El-Lithy et al. 2010).

Les traits d'efficience correspondent généralement à des ratios entre la quantité de biomasse ou d'azote absorbé et la quantité de ressources disponibles dans la plante ou dans le milieu. A ce titre, la

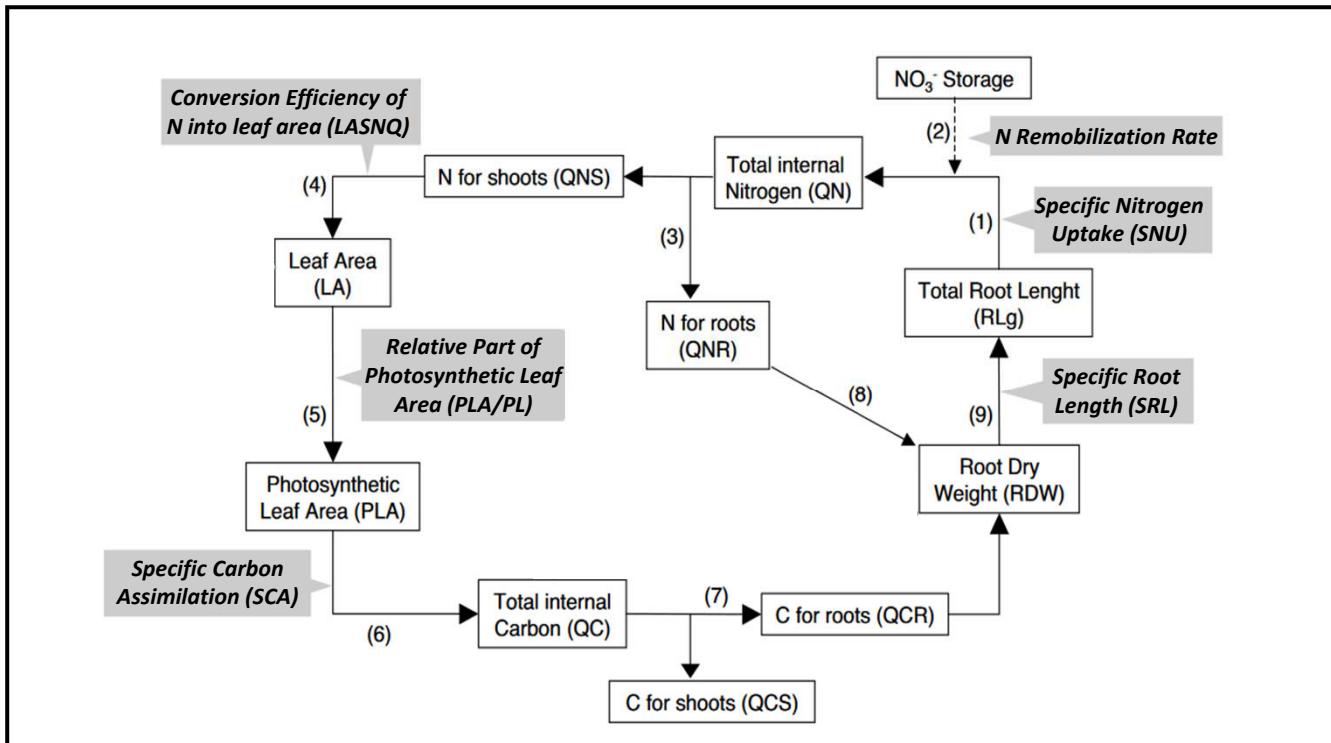


Figure I.8. Schéma conceptuel du modèle ARNICA, représentant les flux de carbone et d'azote.

Les traits d'efficience sont écrits en gras sur fond gris. Les différentes étapes du modèle sont indiquées par des chiffres. Adapté de Richard-Molard et al., 2009.

capacité de la plante à assimiler l'azote disponible dans le milieux, la Nitrogen Use Efficiency (NUE, Hirel et al. 2007), ainsi que ses deux composantes, l'efficience d'absorption de l'azote disponible dans le sol (Nitrogen Uptake Efficiency, NUpE) et l'efficience d'utilisation de l'azote (Nitrogen Utilization Efficiency, NUtE) sont des efficiencies largement utilisées dans les études qui comparent des comportements génotypiques pour des conditions de fertilisation azotée contrastées (Hirel et al. 2007; Vazquez-Carrasquer et al. 2021; Lecarpentier et al. 2021). La capacité photosynthétique et de production de biomasse peuvent quant à elles être évaluées à partir de l'investissement relatif de la plante dans la production de feuilles par rapport aux racines, représentant l'effort potentiel alloué à la photosynthèse par rapport à la croissance racinaire (Hunt 2003). L'investissement dans la production de feuilles est un trait souvent mobilisé pour l'étude des processus liés à la vigueur. Il peut se mesurer à partir de plusieurs ratios : i) le ratio shoot:root (SR), ii) le Leaf Weight Ratio (LWR) mesurant la fraction de la biomasse totale que la plante investit dans les feuilles et iii) le Leaf Area Ratio (LAR) qui estime la quantité de surface foliaire par unité de biomasse totale. De plus, la capacité photosynthétique *per se* peut être i) mesurée ou ii) approximée à partir du ratio entre la surface foliaire et la biomasse produite (Specific Leaf Area, SLA) (Hunt 2003; Violle et al. 2007).

Des efficiencies liées aux flux de carbone et d'azote ont été définies à partir du modèle ARNICA (**Figure 8**, Richard-Molard et al. 2009). Ce modèle, adapté au colza (Vazquez-Carrasquer 2021), représente de façon conjointe les flux de carbone et d'azote dans les compartiments aérien et racinaire. Dans un premier temps, la capacité du génotype à absorber l'azote du milieu ou à le remobiliser est considérée. Deux efficiencies sont mises en avant : l'efficience d'absorption spécifique de l'azote (Specific Nitrogen Uptake, ou SNU ; *cf. #1 sur la Figure I.8*) ainsi que le taux de remobilisation de l'azote à partir des compartiments de stockage (**#2**). L'azote est par la suite réparti entre les compartiments aérien et racinaire (**#3**). Dans les parties aériennes, la capacité du génotype à établir sa surface foliaire en fonction du pool d'azote disponible peut être estimée (Leaf Area:Shoot Nitrogen Quantity, ou LASNQ ; **#4**). Comme seule la surface foliaire photosynthétique permet d'assimiler du carbone, la capacité de la plante à développer prioritairement cette partie peut être estimée (**#5**). Ainsi, cela permet d'estimer la capacité de la plante à assimiler du carbone au cours du temps en fonction de la dynamique de mise en place de la surface foliaire photosynthétique (Specific Carbon Assimilation, ou SCA ; **#6**). Quant aux racines, la part allouée en carbone (**#7**) et en azote (**#8**) à partir des pools générés permet de simuler la biomasse racinaire totale. La capacité du génotype à utiliser ensuite sa biomasse racinaire pour développer la longueur de ses racines (Specific Root Length, ou SRL ; **#9**) est estimée comme étant le ratio entre la longueur racinaire totale et la biomasse racinaire totale. Parmi les efficiencies proposées, les efficiencies d'absorption de l'azote (SNU) et d'assimilation du carbone (SCA) ont été identifiées



comme étant impliquées dans la plasticité des génotypes dans différentes conditions azotées chez *Arabidopsis thaliana* (Richard-Molard et al. 2009).



## I.3. Contrôle génétique des traits impliqués dans la germination et la vigueur précoce chez le colza

---

Comme nous l'avons vu précédemment, la germination et le développement précoce sont des processus multi-trait complexes (*cf. I.1 & I.2*). En effet, ils sont sous contrôle d'une part d'un nombre important de déterminants à effets faibles (nombreuses régions génomiques impliquées) et d'autre part de stimuli environnementaux (lumière, eau, température) qui, la plupart du temps, entrent en interactions avec le génotype (effet GxE). L'ensemble des traits impliqués dans la germination et les phases de levée et de développement précoce sont donc, par nature, des traits quantitatifs (Buckler et al. 2009). De plus, l'architecture génétique d'un trait résulte de l'ensemble des loci associés au trait, ainsi que des effets additifs, de dominance, d'épistasie et de pléiotropie. Décrire l'architecture génétique de traits quantitatifs constitue un véritable enjeu, non seulement pour l'amélioration variétale, mais également pour comprendre les mécanismes sous-jacents à ces processus et leur évolution (Mackay 2001; Holland 2007; Mackay et al. 2009). Déchiffrer l'architecture génétique d'un trait nécessite des outils, ressources et méthodes d'analyse qui seront décrits plus tard (*cf. I.4*). Par ailleurs, chez le colza, la proximité des génomes des *Brassica* mais également des *Brassicaceae*, en particulier *Arabidopsis thaliana*, constitue une ressource de loci/gènes/méabolites impliqués dans un trait en particulier (*cf. I.4.1*). Toutes ces ressources fournissent de l'information quant aux mécanismes impliqués dans des traits d'intérêt, en l'occurrence la germination et le développement précoce. Nous n'en donnons ci-après qu'un aperçu.

### I.3.1. Déterminants génétiques de la germination

Parmi l'ensemble des gènes identifiés pour la germination, des gènes régulateurs et des gènes impliqués dans la synthèse et la réponse à certaines hormones ont été mis en avant : le régulateur maître, WRI1 chez *Arabidopsis thaliana* (Cernac et al. 2006) ou encore le facteur de transcription MADS26 chez le maïs (Ma et al. 2022) connu pour participer à la croissance et au développement de la plante (Martel et al. 2011). Ce facteur de transcription participe à la biosynthèse de l'éthylène, à sa perception et à la réponse à l'éthylène en aval. La surexpression du facteur MADS26 chez *Arabidopsis thaliana* améliore la germination et la croissance racinaire (à 7 jours après germination) (Ma et al. 2022). Les déterminants génétiques de la germination ont été investigués en condition de stress froid et en condition standard chez *Medicago truncatula* (Dias et al. 2011). Cette étude a mis en avant l'implication de gènes liés à la sensibilité à l'ABA (*ABI4*, *ABI5*, *AtLPP2*) dans le contrôle du taux de germination.

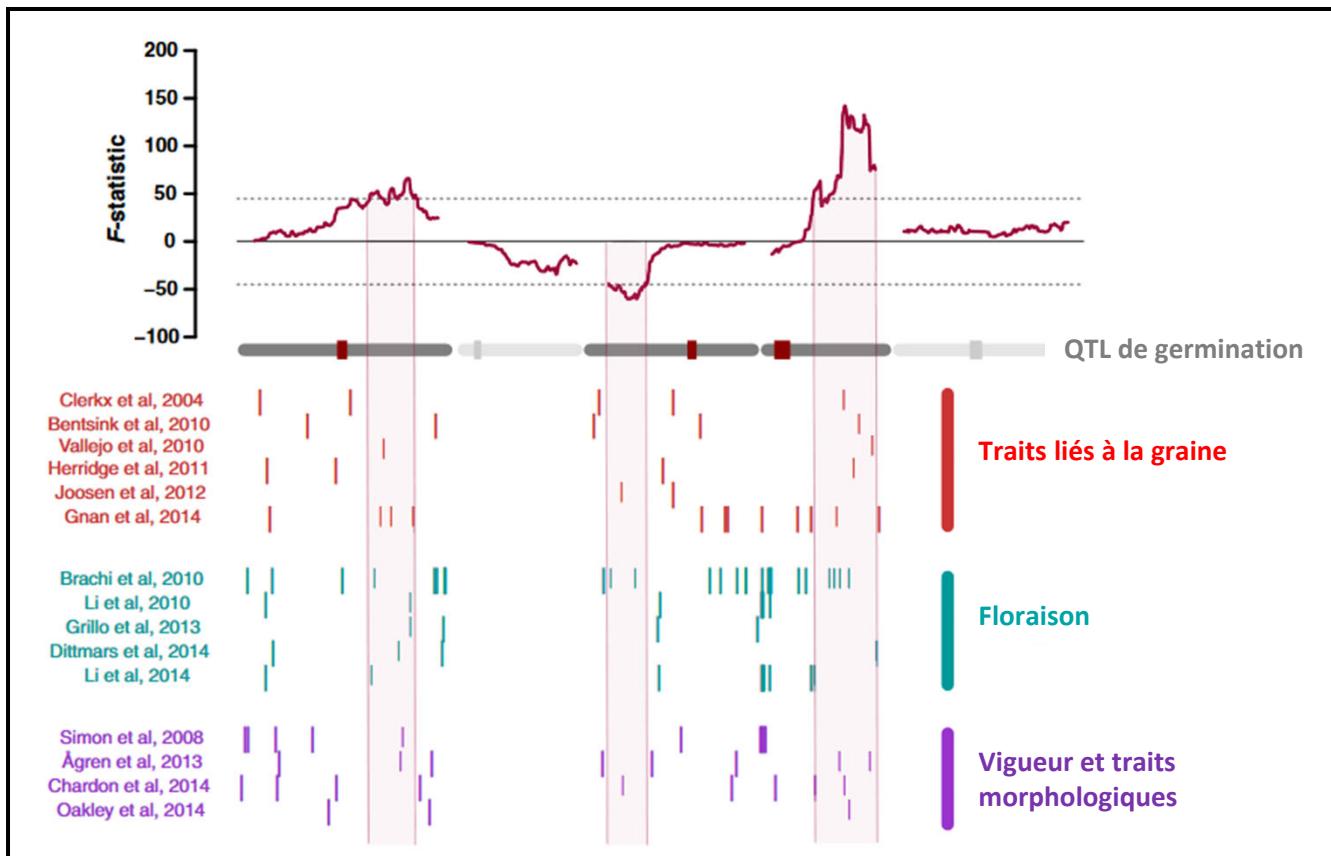


Figure I.9. Co-localisations entre des locis identifiés pour la vitesse de germination et des traits liés à la graine, la phénologie ou la vigueur chez *Arabidopsis thaliana* (d'après Yuan et al. 2016).

Des colocalisations ont aussi été mises en évidence entre de loci impliqués dans la vitesse de germination et d'autres dans la floraison, la qualité de la graine ou la fitness chez *Arabidopsis thaliana* (Yuan et al. 2016 ; **Figure I.9**) et *Brassica rapa* (Basnet et al. 2015). Toutefois, les principaux loci régulant la floraison (*FLC*, *FRI*) ou la dormance (*DOG1*, *DOG6*) ne sont pas impliqués dans ces colocalisations. Ces résultats suggèrent que la vitesse de germination est partiellement dépendante des autres traits d'histoire de vie (Yuan et al. 2016) et qu'il existe un potentiel contrôle génétique commun entre la germination, le développement de la plantule et la floraison.

Chez le colza, quelques études identifient des loci pour des traits liés à la germination (Hatzig et al. 2015; Nguyen et al. 2018; Gad et al. 2021). Elles permettent de mettre en avant en avant trois gènes potentiellement impliqués dans la germination et l'elongation de la radicule : le gène *SCO1* qui code pour un facteur d'elongation essentiel pour la transition entre l'éoplaste (plaste non photosynthétique) et le chloroplaste pendant la germination, le gène *ARR4* qui code pour un régulateur de l'activateur de transcription *ABI5* et le gène *ATE1* code pour des protéines transférases arginyl-t-rna, causant une réduction de la croissance de la radicule chez *Arabidopsis thaliana* (Hatzig et al. 2015).

### I.3.2. Déterminants génétiques de la vigueur précoce

La mise en place de l'architecture aérienne au cours du temps chez *Arabidopsis thaliana* a été investiguée en suivant en cinétique des traits de biomasse, de surface foliaire et de vitesse de développement (Bac-Molenaar et al. 2015; Meyer et al. 2021). Des QTL spécifiques à certaines phases de développement ont été mis en évidence. De même que pour les QTL de germination, les auteurs des deux études identifient des co-localisations entre les QTL détectés et des QTL de floraison. De plus, Bac-Molenaar et al. (2015) identifient un QTL contenant un gène (*DAWDLE*) aux effets pléiotropes, jouant un rôle dans la germination, la croissance et la floraison. En effet, chez *Arabidopsis*, le mutant *dawdle* cause un retard de développement, des malformations des racines, des parties aériennes et des fleurs, ainsi qu'une faible production de graines. Le gène *DAWDLE* serait exprimé dans les méristèmes des racines et des parties aériennes et il serait impliqué de manière précoce dans le développement des organes (Morris et al. 2006).

Chez le riz, le contrôle génétique de la vigueur précoce a été investigué à partir d'une large gamme de traits physiologiques, développementaux et morphologiques. Des loci à effets faibles ont été identifiés pour la longueur de la racine et la biomasse (Cui et al. 2002), à partir de notations visuelles basées sur le nombre de feuilles à 30 jours après semis, ou pour la longueur de la tige (Singh et al. 2017). Un gène master régulateur de la voie de biosynthèse de l'éthylène (*Ethylene insensitive 3*) a été identifié. Il est impliqué dans la différentiation et l'elongation cellulaire, et par conséquent dans un ensemble de

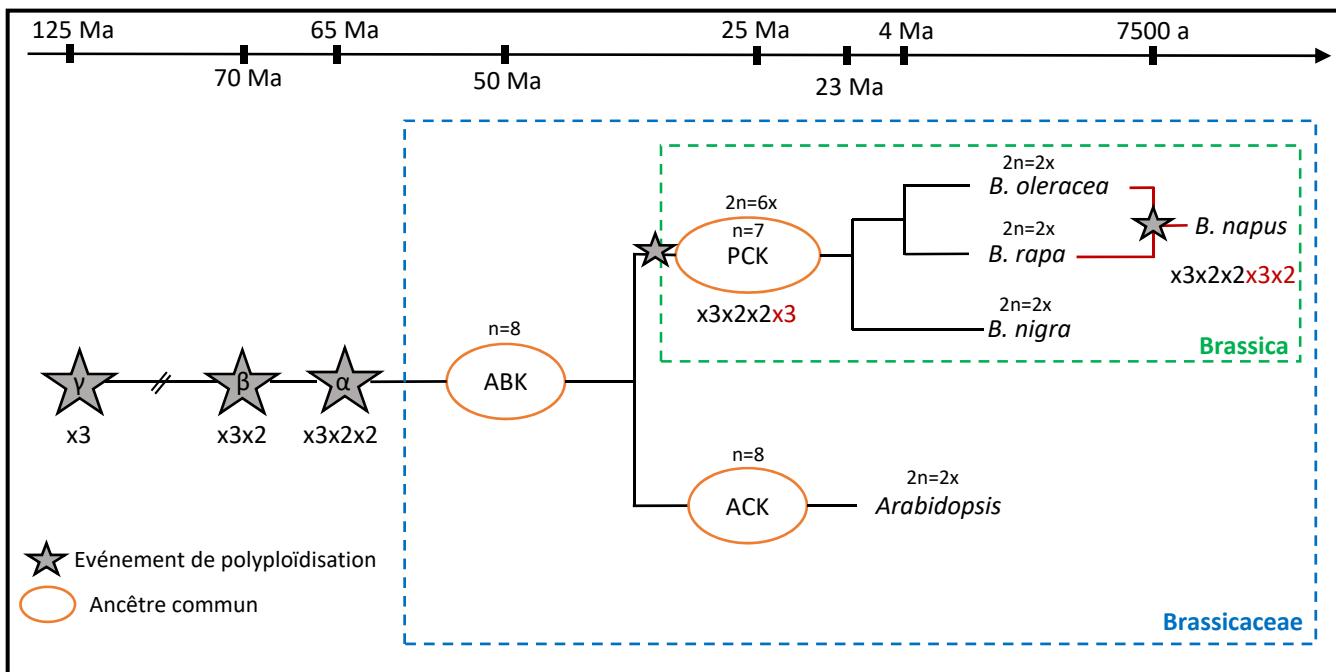


processus liés au développement de la plante (Yang et al. 2015). Enfin, Rebolledo et al. (2015), identifient des régions génomiques impliquées dans des traits fonctionnels morphologiques (nombre de feuilles, nombre de tiges, ...) et dans des traits physiologiques (biomasse, quantité de glucose et de fructose). Pour le colza, des loci ont été identifiés pour des traits physiologiques relatifs à la mise en place de l'architecture aérienne (surface foliaire, volume aérien occupé, biomasse, hauteur) (Knoch et al. 2020, Li et al. 2021) et des traits morphologiques liés au système racinaire (longueur de la racine primaire, longueur racinaire totale, surface racinaire, nombres de racines, biomasse racinaire) (Haelterman et al. 2024).

### I.3.3. Y-a-t-il encore la place pour des analyses génétiques de la germination et la vigueur précoce chez le colza ?

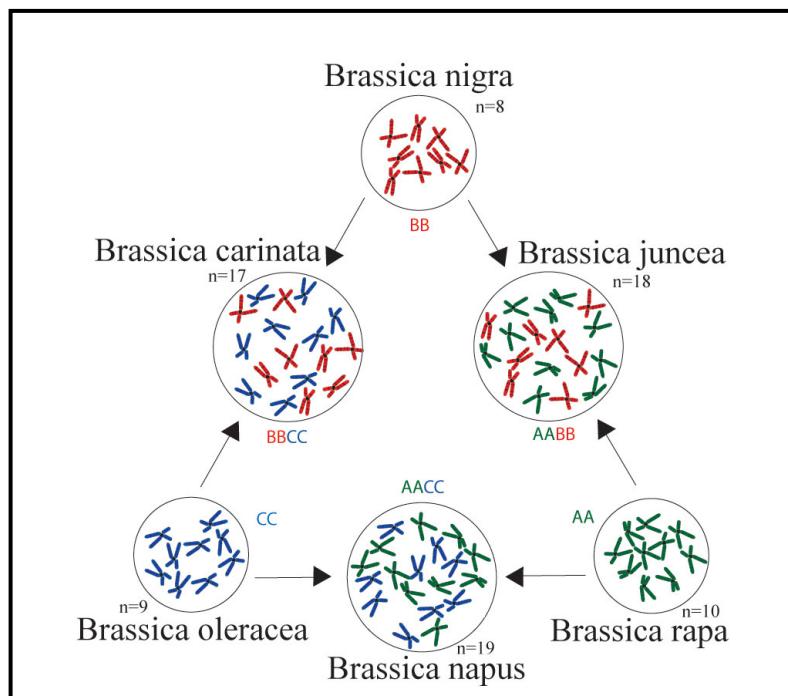
L'analyse de la bibliographie pour plusieurs espèces présentée ci-dessus montre que le déterminisme des traits de germination et de vigueur précoce est polygénique, avec des loci à effets faibles, certains étant par ailleurs pléiotropes. Les traits utilisés sont majoritairement des traits morphologiques, physiologiques (biomasse, surface foliaire, ...) et développementaux (vitesse de germination, ...). Plusieurs études soulignent des colocalisations avec des loci contrôlant la phénologie (précoce, floraison) sans pour autant que les principaux gènes connus pour ces processus (ex. *FLC*) ne soient directement impliqués dans la germination ou le développement précoce. Il est à noter que les études concernent soit la germination soit la vigueur précoce et que peu de publications décrivent des loci impliqués à la fois pour ces deux traits. Enfin, un rôle majeur de la signalisation hormonale (acide abscissique et acides gibbérelliques) a été mis en évidence.

Chez le colza, les études publiées ont pour la plupart été réalisées par analyses de liaison (populations biparentales) ou par analyses d'association en utilisant une faible gamme de diversité génétique (Hatzig et al. 2015; Knoch et al. 2020; Haelterman et al. 2024). Ces analyses concernent la vigueur précoce ou la germination, mais il n'existe a priori pas d'étude mettant en relation les traits de germination et de vigueur précoce ainsi que la comparaison de leur architecture génétique, bien que ces deux traits concourent à la qualité de l'implantation. Ces manques de connaissances restent donc à explorer concernant ces traits et leur corrélation. Pour garantir une étude résolutive, il est important d'utiliser à la fois de dispositifs de phénotypage adaptés pour acquérir des données correspondant aux traits décrits dans les paragraphes précédents, une gamme élargie de diversité génétique, ainsi que de proposer des ponts entre les résultats obtenus aux différentes échelles. Cela peut par exemple être réalisé en caractérisant un grand nombre de génotypes en commun entre les différents dispositifs. Cela pourra permettre d'explorer les corrélations entre traits et potentiellement d'identifier des effets d'interactions ou de pléiotropie entre les locus impliqués.



**Figure I.10. Origine de *Brassica napus*.**

Les étoiles représentent les différents événements de polyploïdisation avec le nombre de multiplications du génome correspondant indiqué en-dessous. Les événements propres aux Brassicaceae sont notés en rouge. Les cercles oranges symbolisent les différents ancêtres communs identifiés (ABK, Ancestral Brassicaceae Karyotype; PCK, Proto-Calepinae Karyotype; ACK, Ancestral Camelinae Karyotype). La double barre entre les duplications  $\gamma$  et  $\beta$  symbolise la séparation de la branche à l'origine des Brassicaceae avec les autres espèces du règne végétal. Schéma adapté de Jiao et al. (2012), Jenczewski et al. (2013), Murat et al. (2015). Figure reprise de Corlouer (2020).



**Figure I.11. Hybridations interspécifiques au sein du genre *Brassica* (U 1935).**

## I.4. Histoire du colza : impacts sur l'organisation génomique et la diversité génétique disponible

### I.4.1. L'organisation du génome du colza résulte d'une histoire évolutive longue

Comme la plupart des espèces végétales, le colza (*Brassica napus* L. subsp. *napus*) résulte d'événements de duplications et de polyploidisations du génome suivis d'étapes de spéciation qui ont jalonné l'évolution (**Figure I.10**). La première triplication ( $\gamma$ ) du génome de l'ancêtre commun, supposée avoir séparé les monocotylédones des dicotylédones, a été suivie de deux duplications ( $\beta$  et  $\alpha$ ) spécifiques des Brassicaceae (Jiao et al. 2012) qui ont abouti aux genres *Brassica* et *Arabidopsis* (Jiao et al. 2012 ; Chalhoub et al. 2014). Au sein du genre *Brassica*, des événements d'hybridations interspécifiques récents ( $\sim 7500$  ans) ont conduit à la formation d'espèces allotétraploïdes (**Figure I.11** ; U 1935). Ainsi l'espèce *Brassica napus* L. (AACC,  $2n=4x=38$ ) est issue du croisement entre *Brassica rapa* (AA,  $2n=2x=20$ ) et *Brassica oleracea* (CC,  $2n=2x=18$ ). Peu d'événements d'hybridation semblent avoir contribué à l'obtention de *Brassica napus* (Chalhoub et al. 2014), limitant ainsi l'intégration de diversité génétique issues des espèces progénitrices. La culture et la diversification de l'espèce *Brassica napus* L. a conduit à différentes sous-espèces plus connues aujourd'hui sous les noms de colza (*Brassica napus* L. subsp. *napus*), rutabaga (*Brassica napus* L. subsp. *rapifera*) et chou kale (*Brassica napus* L. subsp. *pabularia*) (Allender et King 2010).

Les progrès en matière de génotypage et de séquençage des génomes depuis le début des années 2000 ont mis en lumière une forte conservation des séquences codantes entre *Arabidopsis* et le colza (>80% de gènes orthologues; Parkin et al. 2005) ainsi qu'une conservation de la synthénie (ordre des gènes sur les chromosomes) chez les Brassicaceae consolidant ainsi l'hypothèse de l'existence d'un ancêtre commun aux Brassicaceae (ABK, Ancestral Brassicaceae Karyotype) et également au sein des genres *Arabidopsis* (ACK, Ancestral Camelineae Karyotype) et *Brassica* (PCK, Proto-Calepineae Karyotype) (**Figure I.10**). D'autres études basées sur la comparaison des contenus géniques entre Brassicaceae suggèrent que les sous-génomes des espèces polyploïdes ont subi des fractionnements différentiels conduisant à des pertes plus ou moins massives de gènes (sous-génomes hautement fractionnés vs. sous-génomes peu fractionnés) (Liu et al. 2014). Par ailleurs, les gènes dupliqués conservés ont évolué selon différents schémas comprenant la redondance fonctionnelle (les différentes copies conservent la même fonction), la sous-fonctionnalisation (la fonction d'origine est divisée entre les différentes copies) ou la néo-fonctionnalisation (le gène ancestral conserve sa fonction et les copies acquièrent de nouvelles fonctions) (Innan et Kondrashov 2010). Enfin, la

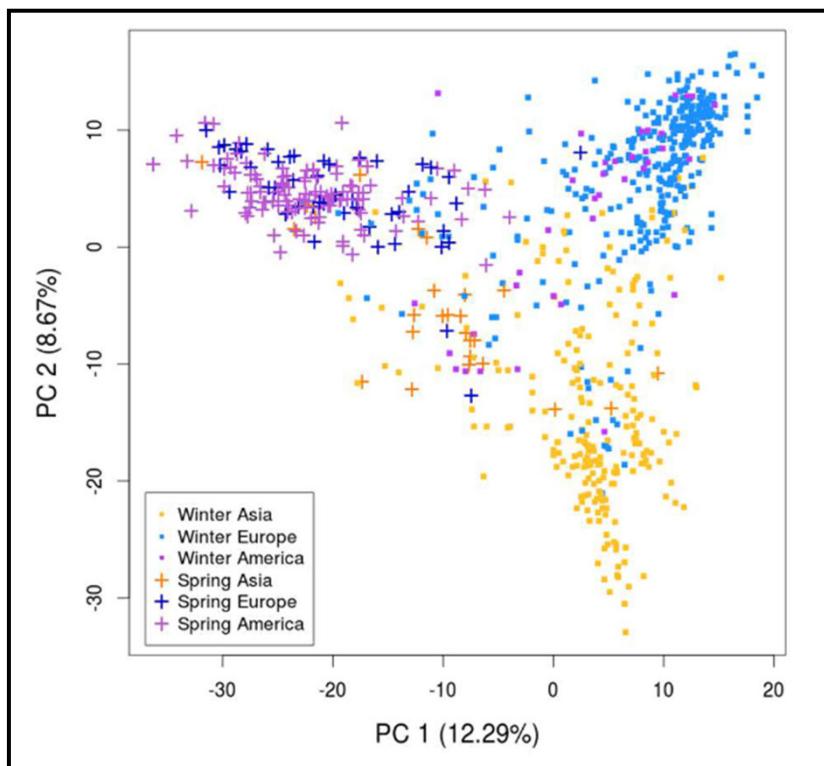


comparaison de plusieurs séquences assemblées de génotypes de colza a révélé l'existence de nombreux événements de translocation chromosomiques entre les génomes homéologues (Chalhoub et al. 2014 ; Song et al. 2020 ; Bayer et al. 2021) pouvant expliquer des variations de type ‘nombre de copies’ (CNV, Copy Number Variation) ou de type ‘présence-absence’ (PAV, Presence Absence Variation). Ces variations structurales expliquent des différences phénotypiques entre accessions de colza dans des processus majeurs comme la précocité ou l’élaboration du rendement (Song et al. 2020). En conclusion, le génome du colza actuel montre un degré de complexité élevé qui résulte d’événements de duplications et de polyploïdisations majeurs auxquels se sont ajoutés des processus de fractionnement génomique et de réarrangements chromosomiques.

#### I.4.2. Ressources génomiques

Des génomes de référence existent pour la plupart des *Brassica* dont les progéniteurs du colza, *Brassica rapa* (Wang et al. 2011) et *Brassica oleracea* (Parkin et al. 2014). Le premier génome de référence de colza, génotype Darmor-*bzh*, a été obtenu en 2014 par un consortium international mené par la France (Chalhoub et al. 2014). Par la suite, des versions améliorées de cet assemblage ont été produites (Bayer et al. 2017, 2021; Rousseau-Gueutin et al. 2020). Par ailleurs, les génomes d'un grand nombre de génotypes de colza ont été assemblés, alignés et annotés dans le but de former des pangénomes (Bayer et al. 2017, 2021 ; Song et al. 2020). L'ensemble de ces données permettent d’élargir les connaissances génomiques et d'accéder à certaines variations inter-génotypiques (ex. CNV et PAV mentionnés précédemment).

Les marqueurs moléculaires les plus utilisés actuellement sont les marqueurs liés au polymorphisme d'un seul nucléotide entre génotypes (SNP, single nucleotide polymorphism). En effet, ces marqueurs, bien que majoritairement bi-alléliques, présentent de nombreux avantages : ils sont abondants et présents tout le long du génome, stables du fait d'un faible taux de mutation et reproductibles. Combinés avec des populations dont l'étendue du déséquilibre de liaison est faible, les SNP sont de puissants outils pour identifier des loci d'intérêt via des analyses génétiques. Les données issues de séquençage nouvelle génération ont permis d'identifier des millions de SNP au sein des ressources génétiques *Brassica* qui ont servi pour le développement de puces de génotypage depuis une dizaine d'années (Edwards 2013; Delourme et al. 2013). La puce commerciale *Brassica* 60K Illumina Infinium™ développée par un consortium international (Clarke et al. 2016) est celle utilisée dans mon travail de thèse. Sur les près de 60 000 marqueurs SNP disposés sur la puce, 33 151 SNP ont été validés dans la population d'étude (fréquence de l'allèle mineur > 5%, moins de 10% de données manquantes, moins de 10% d'hétérozygotie) et ancrés physiquement sur la version améliorée du génome de référence Darmor-*bzh* (Rousseau-Gueutin et al. 2020). Cependant, l'utilisation de puces SNP présente certaines



**Figure I.12. Structure génétique du colza (d'après Gazave et al. 2016).**

L'analyse en composante principale est réalisée sur une population de 782 accessions issues de différentes localisations (Asie, Europe, Amérique du Nord) et classées en type hiver (winter) ou printemps (spring).

limites, dont le fait que la diversité allélique est restreinte aux variants utilisés lors de la conception de la puce (Ganal et al. 2012). Des méthodes de génotypage basées sur du séquençage de régions génomiques ciblées ou du reséquençage de génomes apportent plus de résolution, en particulier elles permettent d'accéder à une diversité allélique plus large, mais sont plus couteuses et chronophages.

### I.4.3. La sélection et l'amélioration génétique du colza ont contribué à réduire la diversité génétique déjà faible chez cette espèce

Aucune population de colza sauvage n'ayant été recensée jusqu'ici (Werner et Snowdon 2018), on parle de ‘sélection’ et non de ‘domestication’ pour cette espèce. Ainsi, la sélection humaine aurait conduit, il y a environ 400 ans, à la sélection d'une part d'individus de type ‘printemps’, caractérisés par un cycle court de développement et sans besoin de vernalisation, et d'autre part d'individus de type ‘hiver’, caractérisés par un cycle de développement plus long et avec un besoin de vernalisation (Lu et al. 2019; *cf. Contexte général*). Dans les années 1930, le colza introduit en Asie y a été sélectionné pour s'adapter aux contraintes environnementales locales, conduisant ainsi aux génotypes de type ‘semi-hiver’ (ou asiatique), dont les besoins en froid sont plus faibles que les types ‘hiver’. Les différences entre les types seraient notamment expliquées par un ensemble de gènes majeurs impliqués dans le besoin de vernalisation et la précocité (Delourme et al. 2013; Chalhoub et al. 2014). La structure génétique chez le colza est expliquée en partie par le type et l'origine géographique (**Figure I.12**; Diers and Osborn 1994; Becker et al. 1995; Bus et al. 2011; Gazave et al. 2016). Cependant, d'autres facteurs rentrent en compte par exemple comme la sélection d'individus pour la production de biomasse foliaire (colzas dits fourragers) par opposition aux individus produisant beaucoup d'huile (colzas dits oléagineux).

La faible diversité présente chez le colza s'explique par son origine évolutive (*cf. I.4.1*) mais elle est encore exacerbée lorsque l'on s'intéresse aux colzas oléagineux. En effet, afin de répondre à des exigences de qualité pour l'alimentation humaine et animale, des programmes d'amélioration de la qualité de la graine ont été menés à partir des années 1945 pour réduire d'une part la teneur en acide érucique et d'autre part la teneur en glucosinolates, aboutissant au déploiement des variétés double zéro (*cf. Contexte général*). Ces programmes ont été conduits sur la base d'introgression par rétrocroisements successifs de caractères favorables à partir de ressources génétiques limitées (un génotype donneur à chaque fois). Ils ont ainsi résulté à deux goulots d'étranglement de la diversité génétique du colza moderne (Becker et al. 1995, Hasan et al. 2006, Bus et al. 2011). A l'inverse, le colza asiatique (ou semi-hiver) semble contenir une diversité allélique plus importante que les autres types et cette diversité n'a pas décliné au cours du temps selon Qian et al. (2006).



#### I.4.4. Ressources disponibles pour élargir la diversité génétique chez le colza

Augmenter la diversité génétique chez le colza est un objectif pour les sélectionneurs, pour ainsi garantir un progrès génétique à long terme, mais aussi pour les programmes de recherche académique comme ceux menés à l'IGEPP. Ainsi, la création, la caractérisation et le maintien de matériel génétique original constitue un enjeu. La diversité génétique présente chez les Brassica est maintenue dans des centres de ressources biologiques comme par exemple le CRB BrACySol d'INRAe. Ces ressources sont disponibles mais plus ou moins faciles à utiliser selon les traits à phénotyper.

La première voie pour augmenter la diversité génétique consiste à utiliser les allèles d'un pool pour enrichir un autre pool. En effet, les différentes sous-espèces et types qui font partie de l'espèce *Brassica napus* (colzas oléagineux, colzas fourragers, rutabaga, chou kale, type hiver, printemps ou asiatique) sont autant de ressources de diversité allélique et s'intercroisent très facilement. Une seconde voie consiste à utiliser la diversité présente chez les espèces apparentées (moutardes, choux, navets ; **Figure I.11**) et sauvages. Des croisements entre espèces diploïdes permettent ainsi de synthétiser *de novo* des colzas. Cependant, ces lignées synthétiques présentent une fertilité réduite et de mauvaises performances agronomiques (Gaeta et al. 2007). Le recours à des lignées semi-synthétiques permet de contourner ce biais : des hybrides allotriploïdes sont obtenus en croisant un colza avec une *Brassica* diploïde. Le passage par la voie allotriploïde AAC augmente le nombre de recombinaisons permettant ainsi le brassage génétique et l'introgression de diversité génétique dans le colza (Leflon et al. 2010; Pelé et al. 2017; Boideau et al. 2021). Un large programme de création de colzas semi-synthétiques utilisant la voie allotriploïde a été mené à l'IGEPP dans le cadre du projet ProBioDiv (Anne-Marie Chèvre, Mathieu Rousseau-Gueutin). Il est également possible d'introgresser des caractères d'intérêt par croisements avec des espèces apparentées en ciblant des loci précis. À titre d'exemple, des gènes de tolérance aux stress abiotiques et de résistance à des bioagresseurs ont été introgressés chez le colza à partir de ressources *Brassica* (pour revue voir Tirnaz et al. 2022). Castillo-Lorenzo et al. (2019) proposent notamment de croiser le colza avec des choux adaptés à des conditions de sécheresse afin d'introgresser les régions génomiques conférant une germination rapide en condition de stress hydrique. L'augmentation de la diversité allélique peut également être obtenue en utilisant des techniques de mutagénèse aléatoire à l'échelle du génome (populations générées par mutagenèse chimique). De même, l'édition de gènes ciblés a prouvé son efficacité pour améliorer certains traits comme l'égrenage des siliques (Zhai et al. 2019).

## Encadré I.1. Principes des analyses d'association et d'identification de signatures de sélection (adapté de Bouchet 2015).

L'analyse QTL (Quantitative Trait Loci) permet d'identifier les régions génomiques responsables des variations d'un trait quantitatif au sein d'une population. Le principe est basé sur la détection d'associations significatives entre des marqueurs génétiquement liés au(x) gène(s) causal(aux) et des variations phénotypiques. Une analyse QTL est définie par :

- la puissance de détection du QTL qui dépend de l'hérédité du trait, de l'effet du QTL et de la représentation des différents allèles dans la population.
- la précision de localisation du QTL, qui dépend de l'étendue du déséquilibre de liaison et de la densité de marqueurs associée.

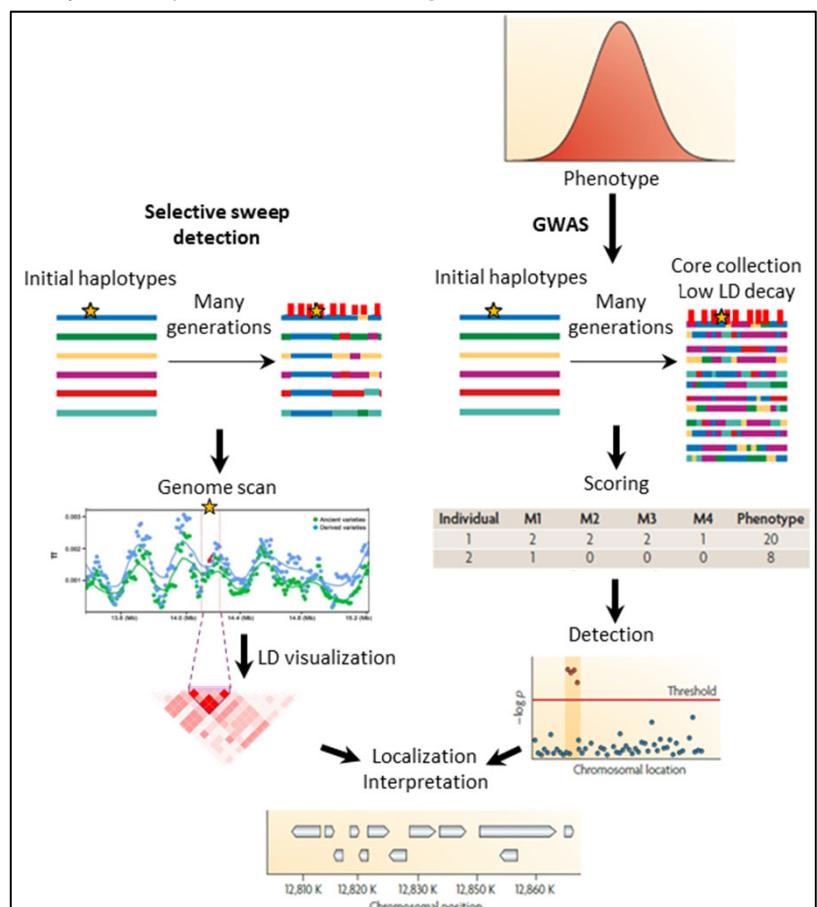
Trois éléments sont nécessaires pour effectuer une analyse QTL:

1. une population présentant une diversité génétique maximisée pour les caractères d'intérêt.
2. des données phénotypiques de haute qualité pour les traits étudiés.
3. des marqueurs moléculaires en densité suffisante sur l'ensemble du génome.

Ainsi, l'analyse QTL peut être étudiée par analyse de liaison ou par analyse d'association (GWAS). L'étude du contrôle génétique de traits associés à des caractères d'adaptation ou ayant subit de la sélection permet d'identifier de nouvelles cibles pour la sélection. Cela peut être réalisé à partir des méthodologies suivantes :

**Les signatures de sélection** représentent l'impact de l'adaptation, de la domestication et de la sélection sur le génome. A l'échelle locale, une signature de sélection peut être identifiée comme étant une augmentation de l'étendu du DL autour de loci impliqués dans des traits d'adaptation ou de sélection. Ainsi, la fréquence de l'allèle sélectionné augmente dans la population, et la variabilité génétique des loci autour de celui sélectionné diminue. Ces signatures peuvent être détectées en cherchant des motifs de DL, une diminution de la diversité génétique le long du génome (ou genome scan) ou une modification de la fréquence allélique par locus.

**L'analyse d'association (GWAS, Genome Wide Association Study)** se base sur l'exploitation des événements de recombinaison qui se sont produits au cours de l'évolution dans les populations naturelles. Les populations d'étude permettent donc de considérer une large diversité génétique, laissant espérer un positionnement relativement précis des QTL. Cependant, pour limiter le nombre de faux-positifs, la structure de la population et l'apparentement entre les génotypes doivent être pris en compte dans les modèles d'analyses (Vilhjálmsson and Nordborg 2013).



### Etapes pour la détection de QTL par analyse d'association (GWAS) et identification de signatures de sélection par genome scan.

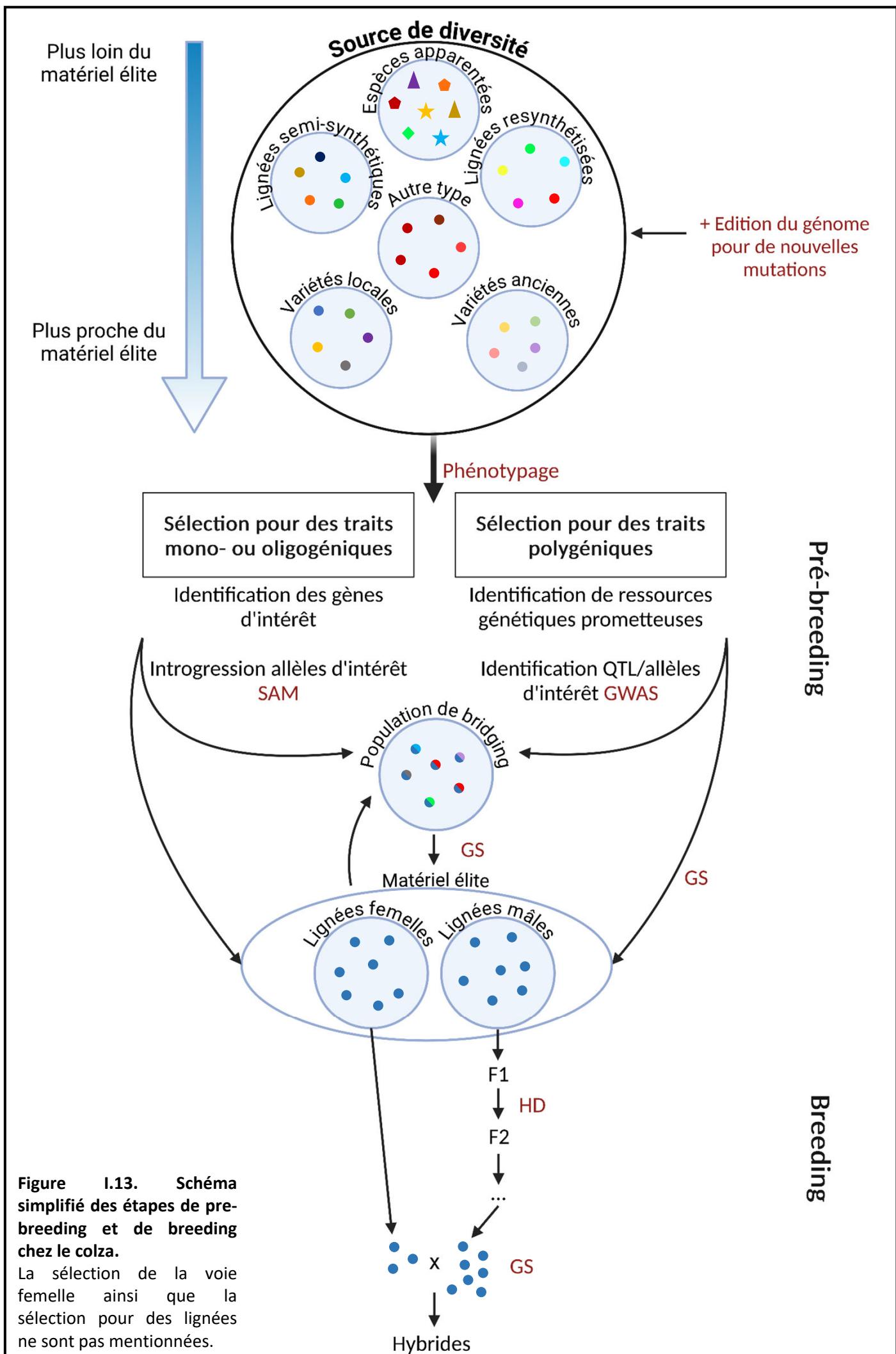
Les marqueurs moléculaires sont représentés par les traits rouges. Les loci d'intérêt sont symbolisés par les étoiles. Adaptée d'après Mackay et al. (2009) et Liu et al. (2023).

#### I.4.5. Etendue du déséquilibre de liaison dans les populations de colza

Le déséquilibre de liaison (DL) est défini comme la différence entre les fréquences gamétiques observées et celles attendues à l'équilibre de Hardy-Weinberg dans une population. Le DL mesure le degré de dépendance statistique entre deux loci. On peut alors définir l'étendue du DL comme la distance maximale pour laquelle deux marqueurs sont considérés en DL (en général avec un seuil de  $r^2=0.2$ ). Les études chez le colza rapportent une étude du DL de l'ordre de 0.2-2.5 Mb et 0.5-1.3 Mb (Hatzig et al. 2015). Ces variations dépendent principalement de la structuration des populations d'étude. De plus, l'étendue du DL varie également en fonction du sous-génome, en étant plus importante sur le sous-génome C que sur le sous-génome A, mais aussi en fonction des chromosomes (Delourme et al. 2013 ; Hatzig et al. 2015), voire même à l'échelle locale notamment en lien avec la présence de signatures de sélection. Enfin, la notion de DL est au centre des principes d'analyses de génétique quantitative et de génétique des populations (**Encadré I.1**). En effet, ces analyses cherchent à identifier et cartographier des régions génomiques par le biais de marqueurs en déséquilibre de liaison avec les loci d'intérêt. Ainsi plus l'étendue du DL est faible, plus la densité de marqueurs devra être élevée pour apporter la puissance nécessaire aux analyses génétiques.

#### I.4.6. Ressources génétiques utilisées pour la thèse

Les traits de germination et de vigueur précoce étant l'objet d'étude de ma thèse, nous avons fait l'hypothèse que nous pouvions phénotyper correctement différents types de colzas ensemble et qu'il n'y aurait pas d'impact majeur de la phénologie à ces stades précoces du développement. Sur cette base, nous avons choisi d'élargir au mieux la diversité génétique du colza oléagineux (*Brassica napus* L.) pour ce travail. Ainsi la population d'étude comprend 233 accessions dont des colzas oléagineux (types hiver, printemps et asiatique) ainsi que des colzas fourragers et des rutabagas. La population présente une étendue du DL à l'échelle du génome de 0.51 Mb. La matrice de génotypage de 33 151 SNP utilisée est donc suffisamment dense pour espérer une bonne résolution des analyses génétiques.



## I.5. Ressources et méthodes pour la sélection chez le colza

---

Le colza est une espèce majoritairement autogame (70% d'autogamie vs. 30% d'allogamie), qui a longtemps été sélectionnée sous forme de lignées pures en France. A partir de 2004, les premières variétés hybrides ont été inscrites. Dans la mesure où ces variétés hybrides présentaient des performances significativement supérieures aux lignées pures, elles ont peu à peu supplanté ces dernières. Ainsi, les hybrides représentaient plus de 50% des inscriptions en 2012, et, aujourd'hui, le marché est composé uniquement de variétés hybrides. La réponse à la sélection est directement conditionnée par la diversité génétique disponible (Falconer et Mackay 1996), il est donc important, en parallèle des programmes de sélection, de réfléchir aux sources de diversité génétique qui vont les alimenter et aux meilleurs moyens de les exploiter.

### I.5.1. Mobilisation des ressources génétiques en pre-breeding

Le pre-breeding correspond aux étapes positionnées en amont de la sélection et qui ont pour objectif d'identifier de nouvelles sources de diversité pour des traits d'intérêt et de mettre cette diversité à disposition des programmes de sélection, notamment en l'introgressant dans des fonds génétiques proches des variétés élites (**Figure I.13**). Bien que les schémas de sélection puissent différer entre les variétés lignées et hybrides, les étapes de pre-breeding sont similaires. L'ensemble des ressources génétiques présentées précédemment (*cf. I.4.4*) sont des entrées valorisables en pre-breeding. Chez le colza, le matériel élite est divisé en deux parties : le matériel femelle et le matériel mâle (**Figure I.13**), et ce pour permettre la création de variétés hybrides maximisant l'hétérosis. Il s'agit donc de réaliser la sélection sur le matériel femelle et le matériel mâle.

Plusieurs stratégies existent pour intégrer les allèles favorables issus des ressources génétiques dans le matériel élite. La première consiste à réaliser des croisements directement entre les ressources génétiques et le matériel élite, puis, laisser les individus de la population s'intercroiser (chez le colza, on peut utiliser le système de stérilité mâle cytoplasmique pour favoriser les intercroisements et donc les recombinaisons), en appliquant à chaque génération une faible intensité de sélection, et ce pendant plusieurs générations. La seconde approche passe par la création d'une population de 'bridging'. Les ressources génétiques à valoriser sont croisées pendant plusieurs générations avec uniquement quelques lignées élites afin d'améliorer leurs performances agronomiques. L'intérêt de ce type de population est de maintenir un maximum d'allèles dans un fond génétique plus proche de celui de la population élite cible que la population initiale de ressources génétiques et présentant de



meilleures performances agronomiques, réduisant ainsi la différence entre les ressources génétiques initiales et le matériel élite.

Afin de réduire la durée de la phase de breeding (environ 10 ans), différentes méthodes existent dont l'objectif est d'accélérer la fixation du matériel génétique. L'haplodiploïdisation permet de fixer les allèles à l'état homozygote à partir de cultures de microspores. Le speed breeding permet d'augmenter le nombre de générations par an grâce à un apport lumineux optimal pour la plante et ainsi de fixer rapidement les allèles par autofécondations successives.

### I.5.2. Principaux facteurs impactant le gain génétique en sélection

Au cours de la sélection, et pour des caractères complexes, on cherche à évaluer et à sélectionner les individus sur la base de leur valeur génotypique. Cette valeur s'estime au travers du phénotype. Cependant, la valeur du phénotype observé est conditionnée par l'effet du génotype, mais aussi par un effet de l'environnement voire un effet de l'interaction génotype × environnement. Au sein d'une population (de sélection) et pour une expérimentation donnée, l'héritabilité informe sur la part de variance d'origine génétique au sein de la variance phénotypique observée et donc sur la capacité d'un dispositif de phénotypage à renseigner sur les valeurs génotypiques des individus impliqués. L'héritabilité  $H^2$  est le rapport de la variance génétique par la variance phénotypique ( $H^2 = VG/VP$ ).

En sélection, on utilise souvent l'héritabilité au sens strict  $h^2$  ( $h^2 = VA/VP$ ) qui limite la prise en compte dans la variance génétique au numérateur aux effets génétiques additifs (VA) et ne prend pas en compte les effets génétiques liés à la dominance (effet d'interaction entre allèles à un même locus) ou à l'épiplatie (effets d'interaction entre plusieurs loci). L'héritabilité au sens strict, informe alors sur la capacité du dispositif à renseigner sur les valeurs génétiques additives des individus ou breeding values.

Durant ce processus de sélection, le gain génétique, ou réponse à la sélection, correspond à la différence de valeur génétique additive entre la population initiale et la population obtenue (théoriquement obtenue après croisement en panmixie des individus sélectionnés au sein de la population initiale). L'équation du sélectionneur (Lush 1943) exprime le gain génétique annuel en fonction de l'intensité de sélection ( $i$  : % de la population choisi pour contribuer à la génération suivante), de la variance génétique additive disponible pour le caractère considéré au sein de la population ( $\sigma_A$ ), de la précision de sélection et qui correspond à la corrélation entre les valeurs phénotypiques et les valeurs génétiques additives ( $r = h$ , racine carrée de l'héritabilité), et enfin de l'intervalle de temps entre deux générations successives ( $T$ ):

$$\Delta G = \frac{ir}{T} \sigma_A,$$

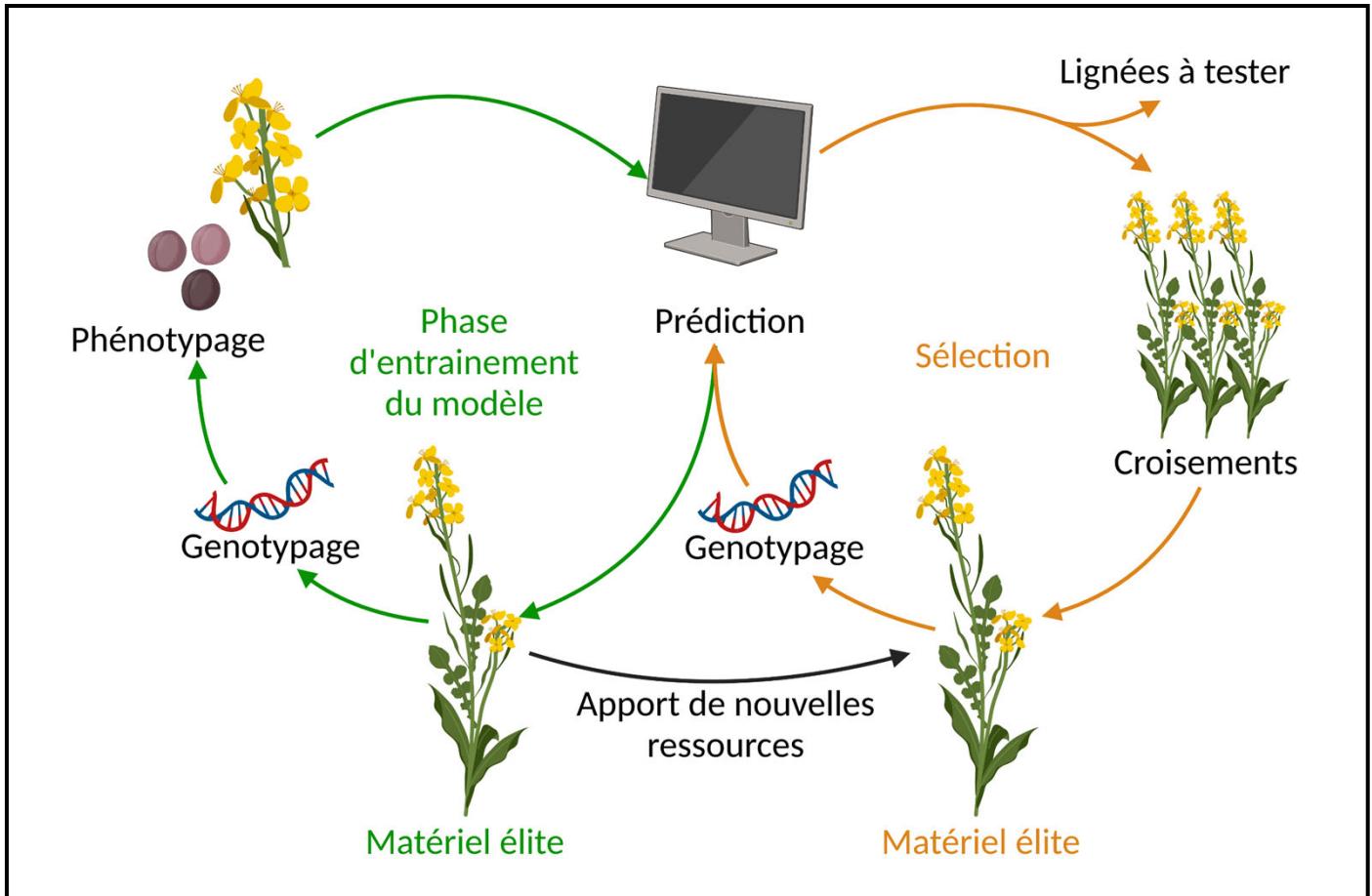


Pour maintenir un potentiel de progrès génétique il est donc nécessaire de maximiser la diversité génétique au sein de la population ( $\sigma_a$ ), ceci soulignant l'intérêt de maintenir un enrichissement du germplasme via les étapes de pre-breeding ; d'étudier le déterminisme génétique et l'héritabilité des traits d'intérêt ( $h^2$ ) ainsi que de pouvoir accélérer la durée des cycles, soit en se basant sur des méthodes d'accélération des cycles, comme mentionné précédemment, ou soit en se basant sur des méthodes qui visent à réduire le temps nécessaire à l'estimation de la valeur génétique des individus via des méthodes de phénotypage précoce ou des méthodes de prédiction de cette valeur sur la base de données moléculaires (sélection génomique) ou d'autres endophénotypes (sélection phénomique). Deux points de vigilance sont à souligner : une intensité de sélection trop forte s'accompagne d'une réduction de la variance génétique. Une forte pression de sélection permet donc un gain génétique important à court terme, mais handicape le potentiel de progrès génétique à moyen/long terme. Ensuite, il est rare que les programmes de sélection ne considèrent qu'un seul caractère à la fois. Quand il s'agit de sélectionner concomitamment pour un ensemble de traits, il faut prendre en compte la corrélation génétique entre les différents traits (Sánchez et al. 2008) et limiter le fait que la sélection pour un trait se fasse au détriment de l'autre.

### I.5.3. La sélection génomique pour la sélection de traits au déterminisme génétique complexe

Les marqueurs moléculaires ont été largement utilisés dans le cadre d'analyses génétiques, que ce soit par des approches de liaison ou d'association (**Encadré I.1**). Chez le colza, leur utilisation a permis de mettre en évidence un grand nombre de QTL pour les principaux traits agronomiques (développement, qualité de la graine, rendement, résistances aux stress biotiques et abiotiques) (pour revue voir Delourme et al. 2018). Certains QTL présentant des effets forts ont pu être valorisés dans des programmes de sélection assistée par marqueurs, par exemple concernant la couleur des graines, la restauration de la fertilité mâle, ou la qualité de l'huile (pour revue voir Hayward et al. 2012). Dans ce cas, ce sont les quelques marqueurs flanquant le QTL qui permettent de le suivre au cours des générations, et de vérifier que les individus conservés possèdent bien les allèles favorables au QTL ciblé. Cependant, la sélection assistée par marqueurs trouve ses limites lorsqu'il s'agit de suivre en descendance un grand nombre de locus à effets faibles simultanément. En se basant sur la valorisation des données moléculaires à l'échelle pangénomique, la sélection génomique (GS, Genomic Selection) permet de contourner cette limite pour la sélection de traits quantitatifs à l'architecture génétique polygénique (Meuwissen et al. 2001).

Les modèles de GS utilisent l'ensemble des marqueurs disponibles couvrant le génome pour prédire la valeur génétique (GEBV, genomic-estimated breeding values) des individus pour le trait d'intérêt.



**Figure I.14. Utilisation de la sélection génomique dans un contexte de sélection (d'après Heffner et al. 2009).**

La première phase (en vert) consiste à entraîner le modèle de sélection génomique sur une population phénotypée et génotypée. Le modèle est ensuite testé pour sa capacité prédictive puis utilisé sur la population test qui génotypée seulement. Lorsque le modèle est suffisamment robuste, il est utilisé en sélection (en orange) et les valeurs génomiques prédites permettent de choisir les individus à intégrer dans les étapes de sélection ou à utiliser pour les croisements.

#### Encadré I.2. Principaux modèles utilisés en sélection génomique.

##### Modèle rrBLUP

$$Y = \mu + X\beta + \epsilon$$

Avec  $Y$  le vecteur des phénotypes de la population,  $\mu$  la moyenne globale de la population,  $X$  la matrice de génotypage à  $N$  lignes ( $N$  génotypes) et  $P$  colonnes ( $P$ , le nombre de marqueurs),  $\beta$  le vecteur des effets additifs des  $P$  marqueurs, où  $\beta \sim N(0, \sigma_B^2 I)$  avec  $B$  l'effet additif de chaque marqueur et  $I$  la matrice d'identité,  $\epsilon$  l'erreur résiduelle où  $\epsilon \sim N(0, \sigma^2 I)$ .

##### Modèle GBLUP

$$Y_i = \mu + G_i + \epsilon_i$$

Avec  $G_i$  l'effet génétique aléatoire suivant une distribution normale  $G \sim N(0, K\sigma_G^2)$ , où  $K$  est la matrice d'apparentement et  $\epsilon \sim N(0, \sigma_e^2 I)$ .

Pour cela, des modèles de régression linéaire sont définis à l'aide d'une population entraînement pour laquelle le génotype et le phénotype de chaque individu sont connus (**Figure I.14, Encadré I.2**). Cette phase d'entraînement permet d'estimer les effets des marqueurs sur le phénotype et de prédire les GEBV des génotypes de la population d'entraînement. Par la suite, une population test est utilisée afin d'évaluer la capacité prédictive du modèle préalablement calibré. Pour cela, seules les données de génotypage de cette population sont intégrées au modèle, et les GEBV sont prédites. La capacité prédictive du modèle, à savoir la corrélation de Pearson entre les valeurs phénotypiques observées et les GEBV est par la suite estimée sur la population test ( $r$ , précision de sélection). Ainsi, sans identifier l'ensemble des marqueurs liés aux traits d'intérêt, mais en considérant l'amélioration du fonds génétique dans son ensemble, la sélection génomique permet de considérer de manière simultanée un ensemble de loci aux effets faibles et, *in fine* d'obtenir un gain génétique sur le trait cible.

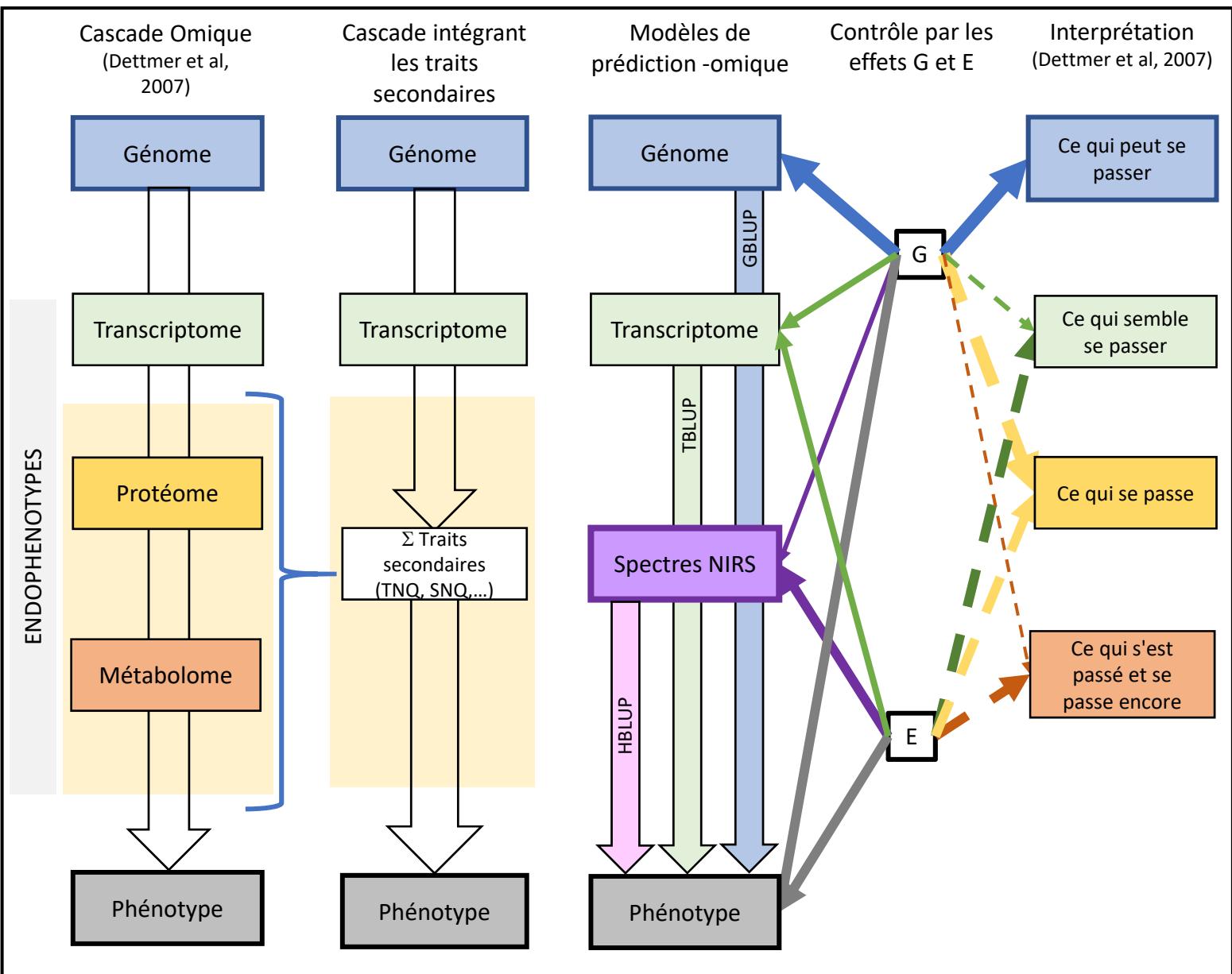
Plusieurs facteurs sont connus pour impacter la précision de la GS. Le premier facteur est l'héritabilité du trait et son architecture génétique. En effet, les capacités prédictives du modèle augmentent avec l'héritabilité du trait (Hayes et al. 2009) et lorsque le trait est caractérisé par un déterminisme polygénique plutôt qu'oligogénique (Daetwyler et al. 2010; Riedelsheimer et al. 2012). L'hypothèse sous-jacente aux modèles de GS est que les effets associés à chacun des marqueurs sont tirés au sein d'une même loi normale centrée sur zéro, hypothèse de laquelle on peut s'écartez dans le cas des traits oligogéniques où seulement quelques loci à effets forts contrôlent le caractère. Le second facteur est la densité de marquage. En effet, pour des densités fortes de marquage, il sera probable que les marqueurs utilisés dans les modèles de GS soient en déséquilibre de liaison avec la plupart des polygènes impliqués dans le contrôle du caractère. La densité de marquage doit donc être raisonnée en fonction de l'étendue du DL au sein de la population d'entraînement (Solberg et al. 2008; Norman et al. 2018) mais également au sein de la population de sélection. Par ailleurs, afin que les modèles soient efficaces, il est important que les deux populations soient le plus apparentées possible, qu'il n'y ai pas de différence de motifs de DL entre les deux populations (Albrecht et al. 2011; Lund et al. 2016; Norman et al. 2018). Enfin, le dernier facteur est le choix du modèle. Afin de contourner les problèmes statistiques causés par le grand nombre de marqueurs (prédicteurs) par rapport au nombre d'individus à prédire ( $n$ ) ( $p>>n$ ) induisant un nombre trop faible de degrés de liberté et la multicolinéarité entre les marqueurs, le choix des modèles de GS s'est tourné vers des modèles utilisant des régressions pénalisées, une sélection des variables ou une réduction des dimensions.

Deux principaux types de modèles sont couramment utilisés en GS : les modèles fréquentistes et les modèles bayésiens. Les modèles fréquentistes les plus utilisés sont le rrBLUP (ridge regression best linear unbiased prediction) et le GBLUP (Genomic best linear unbiased prediction) (**Encadré I.2**). Le rrBLUP est une méthode de régression pénalisée qui permet de tirer les effets des marqueurs dans



une même loi normale et de faire tendre les effets de l'ensemble des marqueurs vers zéro, ce qui permet de modéliser une architecture polygénique avec un très grand nombre de QTL à effets faibles (Meuwissen et al. 2001). Le GBLUP utilise une matrice d'apparentement entre les individus de la population estimée à partir des marqueurs moléculaires (ou matrice de kinship génomique). Cette matrice d'apparentement génomique est de taille  $N \times N$  (où  $N$  est le nombre de génotypes) et permet ainsi une réduction des dimensions initiales du jeu de données ( $n \times p$ ). Les modèles bayésiens (Bayes A, Bayes B, Bayesien Lasso) permettent plus de flexibilité dans l'estimation des effets des marqueurs. Ce type de modèles utilise des lois différentes pour tirer les effets des marqueurs. Ainsi des gènes à effets faibles et des gènes à effets modérés peuvent être considérés dans le même modèle pour mimer la présence de quelques QTL à effet fort. Aux modèles fréquentistes ou bayésiens, s'ajoutent des méthodes d'apprentissage telles que les réseaux neuronaux qui permettent de prendre en compte des interactions faibles entre les marqueurs ainsi que des relations non linéaires (Gianola et al. 2011), ou des méthodes de random forest et de support vector machine (Ogutu et al. 2011). Malgré la multiplicité des modèles utilisés en GS, il semble que les capacités prédictives soient sensiblement équivalentes (Ogutu et al. 2011; Heslot et al. 2012).

L'utilisation de la GS permet de réduire le besoin en phénotypage pour la population cible. Par conséquent, cela diminue le temps entre deux générations ( $T$ ) et augmente le gain génétique tout en réduisant les coûts associés. Un autre intérêt de l'utilisation des GEBV est de sélectionner les individus présentant les meilleures performances pour concentrer l'effort de sélection sur ces derniers. Cependant, plusieurs études ont démontré un lien entre GS et perte de diversité génétique dans les populations élites dû à la sélection d'individus fortement apparentés (Jannink 2010; Lin et al. 2016; Arguello-Blanco et Sneller 2023). Enfin, les modèles de GS couramment utilisés en génétique animale et végétale sont adaptés pour capter des effets génétiques additifs. Cependant, les traits sont également impactés par des effets génétiques non additifs, ainsi que par des interactions  $G \times E$ . Utiliser des informations sur l'endophénotype (*i.e.*, le phénotype intermédiaire) seules ou combinées avec des données génomiques est une voie pour obtenir une capacité prédictive des modèles supérieure à celles des modèles utilisant uniquement des données génomiques.



**Figure I.15. Cascade omique et modèles de prédition potentiels.**

La première colonne représente la cascade omique telle que représentée par Dettmer et al. (2007). Le protéome et le métabolome peuvent être approximés à l'aide d'un ensemble de traits secondaires capturé par les spectres NIRS. Ainsi, trois types de modèles de prédition peuvent être proposés en utilisant des données génomiques (GBLUP), transcriptomique (TBLUP) et spectrales (HBLUP) afin de prédire des phénotypes. Pour chaque donnée omique est représentée son interprétation biologique ainsi que l'influence de l'effet génétique (G) et l'effet environnemental (E) sur les données. L'épaisseur de la flèche indique l'influence de l'effet sur les données. Figure adaptée de Dettmer et al. (2007).

#### I.5.4. Sélection phénomique : prédictions à l'aide de données endophénotypiques

Transcriptome, protéome et métabolome participent au phénotype global et peuvent être définis comme des endophénotypes (ou phénotypes intermédiaires) (Mackay et al. 2009). Capter l'effet de ces endophénotypes peut permettre de mieux comprendre et de prendre en compte une part de la variance du phénotype final qui reste inexpliquée par les effets génétiques additifs seuls. Ces informations permettent en effet de considérer des effets génétiques non additifs et des effets d'interactions G × E sur le phénotype final (**Figure I.15**). Plusieurs études ont été menées pour évaluer l'intérêt d'utiliser les endophénotypes dans des modèles de régression, en supplément ou en remplacement des données génomiques. On parle alors de sélection phénomique (PS, phenomic selection).

Utiliser des données ‘omiques’ pour estimer une matrice d’apparentement et insérer cette matrice dans le modèle GBLUP à la place de la matrice d’apparentement génomique peut améliorer la qualité prédictive de modèles de prédiction (Fu et al. 2012; Riedelsheimer et al. 2012; Xu et al. 2016; Zenke-Philippi et al. 2016). Cependant, acquérir ce type d’endophénotypes est encore coûteux et difficile à mettre en place (Westhues et al. 2017; Schrag et al. 2018; Knoch et al. 2021). Une source de données moins coûteuse et plus facile d'accès consiste dans les données de signatures spectrales acquises à l'échelle d'un organe ou d'un tissu. C'est l'utilisation de données spectrales acquises par spectrométrie dans le proche infrarouge (NIRS) et leur utilisation comme endophénotype pour implémenter la PS au sein des programmes de sélection que nous allons décrire dans la suite de ce manuscrit.

La spectrométrie dans le proche infrarouge permet de mesurer l'absorbance ou la réflectance de la lumière sur un échantillon entre 780 nm et 2500 nm. Les liaisons hydrogènes des constituants absorbent la lumière, permettant ainsi de caractériser l'échantillon par des données quantitatives d'absorbance pour chaque longueur d'onde. Ces valeurs d'absorbance sont proportionnelles à la concentration en molécules (loi de Beer-Lambert). Grâce à cette propriété, des équations de prédiction ont été établies pour prédire, à partir de données acquises par NIRS, la teneur en certains composants majeurs pour la qualité des produits. C'est le cas par exemple pour le colza où des calibrations commerciales existent pour les teneurs en huile, protéines et glucosinolates de la graine. D'autres modèles ont été développés plus récemment pour prédire les teneurs en carbone et en azote dans des tissus variés de plantes de colza (Rolland et al. soumis).

En 2018, Rincent et collaborateurs proposent d'utiliser des données spectrales acquises par NIRS pour estimer une matrice d'apparentement spectrale et prédire les valeurs génétiques des individus pour



différents traits liés au rendement chez le blé d'une part et au développement et à la tolérance à différentes maladies chez le peuplier d'autre part (**Figure I.15**). Ce travail apporte la preuve de concept de l'intérêt d'utiliser les données spectrales pour développer des modèles de PS dont les capacités prédictives se révèlent supérieures à celles des modèles de GS. Depuis 2018, de nombreuses études ont cherché à évaluer la capacité prédictive de modèles de PS basés sur des données spectrales pour la prédiction de traits d'intérêt, en combinaison ou non avec des données génomiques, chez le blé (Krause et al. 2019; Robert et al. 2022a; Winn et al. 2023), le maïs (Cuevas et al. 2019; Lane et al. 2020; Weiß et al. 2022), le triticale (Zhu et al. 2022), le soja (Zhu et al. 2021), le seigle (Galán et al. 2020, 2021), la pomme de terre (Maggiorelli et al. 2024) ou encore la vigne (Brault et al. 2022).

Différents facteurs sont à considérer pour optimiser la PS, en fonction de la population d'étude, des traits à prédire et des données spectrales. Le premier est la capacité des données spectrales à capturer la variabilité du trait cible. Par exemple, la bonne capacité prédictive de la PS pour le rendement chez le blé à partir de données spectrales acquises sur grains peut s'expliquer par le fait que la composition des grains est en partie corrélée au rendement (Dallinger et al. 2023). Le second facteur est l'architecture génétique du trait elle-même. En effet, comme la GS, la PS prédit relativement bien les traits complexes à l'architecture polygénique par comparaison à des traits contrôlés par un faible nombre de gènes à effets forts (Zhu et al. 2022). Le troisième facteur repose sur la capacité des données spectrales à estimer l'apparentement génétique entre les individus. En effet, dès lors que les données spectrales permettent d'estimer précisément l'apparentement entre les individus, la matrice d'apparentement spectrale peut se substituer à la matrice d'apparentement génomique dans le modèle GBLUP tout en conservant une bonne qualité de prédiction. Le quatrième facteur concerne la capacité des données spectrales à rendre compte de variations génétiques. Ainsi, l'héritabilité des longueurs d'ondes varie selon le tissu considéré lors de l'acquisition des spectres (Rincent et al. 2018). L'héritabilité varie également selon l'espèce considérée, malgré le fait que les spectres ont été acquis sur le même type d'organes. Ainsi, pour des spectres acquis sur grains, l'héritabilité des longueurs d'ondes des spectres acquis sur blé était supérieure à celle sur triticale (Rincent et al. 2018; Zhu et al. 2022). Enfin, le cinquième facteur est la structure génétique du panel. Plusieurs études montrent que, contrairement à la prédiction génomique, la capacité prédictive de la prédiction phénotypique semble être moins impactée par la structuration génétique dans les populations d'entraînement et les populations test (Zhu et al. 2021, 2022; Weiß et al. 2022).

La sélection phénotypique est une méthodologie récente qui nécessite plus d'investigations. Certains points sont communs avec les questionnements portant sur la sélection génomique, comme la prise en compte des corrélations entre traits. D'autres questions sont à instruire spécifiquement pour la PS, telles que le choix des longueurs d'ondes, l'utilisation de spectres acquis dans différents



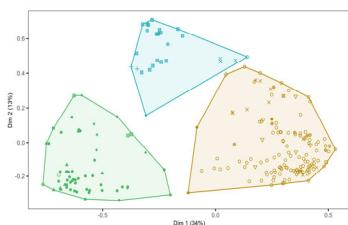
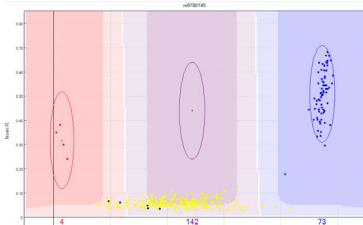
environnements, la prise en compte de l'effet GxE ou encore sur son intégration dans les programmes de sélection.

### I.5.5. Sélection des traits de germination, développement et vigueur : GS ou PS?

Concernant les approches utilisées pour prédire la germination, on ne ressource à ce jour qu'une étude utilisant des modèles de prédiction génomique chez l'orge (Rooney et al. 2022). De plus, pour la vigueur précoce, des modèles de prédiction génomique et phénotypique ont été utilisés pour prédire des traits de biomasse (Fernandes et al. 2018; Galán et al. 2020, 2021) et des notations de vigueur (Brault et al. 2022), chez le sorgho, le seigle et la vigne. Actuellement, les études portant sur la PS et comparant avec la GS concernent majoritairement l'étude de la prédiction du rendement (Rincent et al. 2018; Lane et al. 2020; Robert et al. 2022a; Roscher-Ehrig et al. 2024). Dans ma thèse, je fais l'hypothèse que les spectres NIRS capturent des informations d'apparentement et de la composition des graines qui pourraient être liées à leur capacité germinative et de développement. Ainsi, j'ai choisi d'investiguer et de comparer la capacité prédictive de la GS et de la PS pour ces traits.

## Population de diversité de *B. napus*

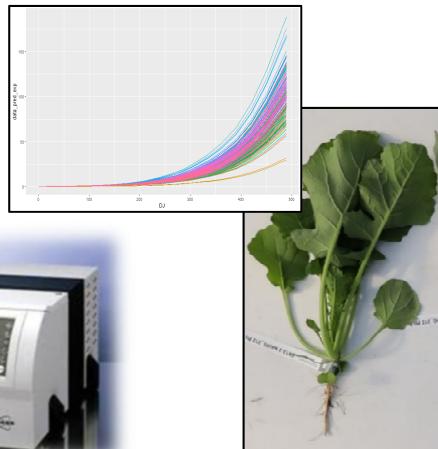
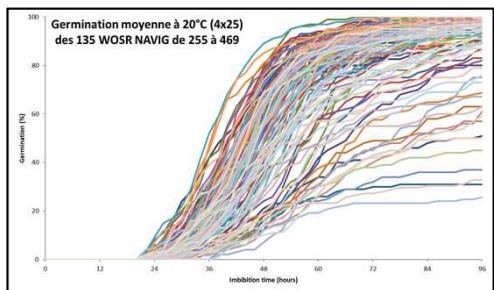
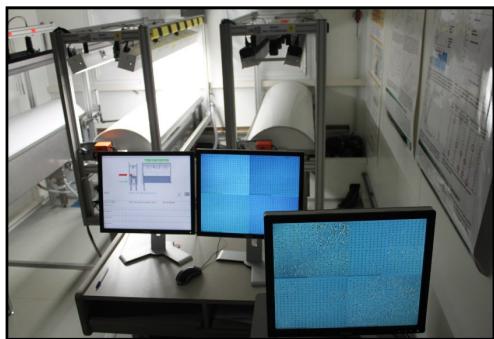
Récolte 2021,  
Le Rheu  
1 lot de graines / accession



**233 accessions**

**33 151 SNP**

**Structure génétique**



**PHENOTIC (GEVES, Angers)**

**Variables GERMINATION**

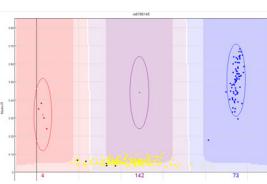
**Graines (1 lot/acc., 3 rep.)**

**Poudre de feuilles (6 plantes/acc., 1 rep.)**

**|GEPP, Le Rheu**

**Variables DEVELOPPEMENT PRECOCE**

**Variables SPECTRALES**



**33 151 SNP**

### LIVRABLES

- Traits et héritabilités
- Loci et gènes sous-jacents
- Allèles favorables
- Ressources génétiques
- Prédictions génomiques et phénotypiques

Figure I.16. Matériel et stratégie globale de la thèse.

## I.6. Objectifs, hypothèses et stratégie déployée au cours de la thèse

---

Réussir l'implantation de la culture de colza est primordial pour garantir le rendement final, notamment dans un contexte de réduction des intrants et de raréfaction des ressources en eau. Cependant, le colza est soumis à un ensemble de stress dès le semis, ce qui va impacter le taux de germination et de levée, puis la tolérance aux attaques et aux contraintes environnementales (*cf. I.1*). Dans ce contexte, la fourniture de matériel vigoureux capable de s'implanter correctement quelles que soient les conditions environnementales est un enjeu majeur pour l'amélioration de la culture et le levier variétal peut en partie y répondre.

L'objectif général de ma thèse a donc été d'identifier les déterminants génétiques impliqués dans la germination et le développement précoce, et d'évaluer la capacité à prédire ces traits pour optimiser le travail de sélection.

Mon projet reposait sur plusieurs hypothèses. La première est que la germination et le développement précoce étant des processus complexes, ils peuvent être plus finement appréhendés par des traits de développement ou physiologiques (*cf. I.2*) à partir desquels il serait possible d'identifier des régions génomiques clés. La seconde hypothèse est que les processus de germination et de développement précoce ont pour le moment été investigués dans des gammes de diversité réduites et qu'il est possible d'étendre la diversité allélique à des ressources plus originales (élargissement des populations d'étude à des ressources issues de l'espèce *Brassica napus*). En effet, ceci pourrait permettre d'identifier des allèles favorables à valoriser en sélection (*cf. I.5*). Enfin la troisième hypothèse est que l'on peut utiliser des données spectrales pour prédire les traits de germination et développement précoce (*cf. I.5*).

Ma thèse s'est organisée autour de deux questions de recherche :

1. Quel est le déterminisme génétique de la germination et de la vigueur précoce ?
2. Ces traits étant potentiellement polygéniques, est-il possible de les prédire ?

Pour cela, j'ai réalisé un phénotypage en dynamique pour la germination et la vigueur précoce sur une large gamme de diversité génétique. A partir des données acquises, des détections de QTL par analyses d'association à l'échelle du génome (GWAS, genome wide association studies) ont été conduites et complétées par des approches de prédiction des traits cibles à partir de données spectrales. Le matériel et la stratégie globale de la thèse sont présentés à la **Figure I.16**.



Le matériel biologique utilisé est constitué d'un panel de diversité de 233 accessions de l'espèce *Brassica napus* L. comprenant : 133 colzas oléagineux d'hiver (WOSR, winter oilseed rape), 85 colzas oléagineux de printemps (SOSR, spring oilseed rape), 13 colzas fourragers d'hiver (WFR, winter fodder rape) et 2 rutabagas. Tous les génotypes ont été multipliés dans les mêmes conditions afin de minimiser l'effet 'lot de graines' (mode et conditions de production et de stockage). La multiplication a été conduite durant la saison 2020-2021 sur une parcelle expérimentale du domaine de la Motte (INRAe). Chaque génotype a été semé sur une micro-parcelle de 10 m<sup>2</sup> en 6 lignes adjacentes. Pour réduire le risque de pollinisation croisée (le colza étant allogame à 30%), les graines ont été récoltées sur les 4 lignes centrales. Nous disposons donc d'un seul lot récolté par génotype qui est conservé dans des conditions de température et d'hygrométrie contrôlées (chambre de stockage des graines de l'IGEPP contrôlée par sondes OCEASOFT®). C'est ce lot de graines qui a été utilisé pour toutes les études de la thèse (quantité de graines suffisante pour la thèse).

Les données alléliques pour ces 233 accessions ont été obtenues par génotypage SNP sur la puce commerciale Brassica 60K Illumina Infinium™ (Clarke et al. 2016; cf. I.4.2) et étaient disponibles dès le début de la thèse.

Le phénotypage pour les variables liées à la germination a été réalisé sur la plateforme à haut-débit PHENOTIC du GEVES, à Angers en 2022 (Ducournau et al. 2004; Wagner et al. 2011; Boureau 2020). Pour chaque génotype, 4 lots de 25 graines ont été suivis pendant 96 h après imbibition. Les images acquises toutes les deux heures permettent, après traitement et analyse, d'accéder à un ensemble de traits : augmentation du volume de la graine au cours de l'imbibition, temps de la 1<sup>ère</sup> germination, temps pour obtenir 20%, 50% ou 80% de germination, taux de germination au cours du temps, vitesse d'élongation de la radicule (Demilly et al. 2014). Il est possible, à partir de ces données, de calculer d'autres variables (homogénéité du lot, aire sous la courbe, ...). Ces données et les résultats qui en ont découlé sont décrits au chapitre II (Laurençon et al. 2024).

Le phénotypage du développement précoce a été réalisé à l'IGEPP (Le Rheu), en tunnel ce qui permettait d'avoir des conditions semi-contrôlées. Le dispositif est celui développé et utilisé par Grégoire Bianchetti (Bianchetti 2021). Les plantes sont conduites dans des caissons de 1m<sup>3</sup> dont l'irrigation est pilotée. Les variables ont été acquises d'une part grâce à un suivi dynamique par prise d'images tous les deux jours et d'autre part lors de la récolte au stade à 4-6 feuilles. Ainsi nous avons pu avoir accès à la dynamique de mise en place des organes aériens (taux de couverture) et à des variables acquises sur les plantes (biomasse racinaire et foliaire, nombre et surface des feuilles, teneurs en C et N) à la fin des expérimentations. Ces données et les résultats correspondants sont présentés font l'objet du chapitre III et d'un article en préparation.



Enfin, des spectres NIRS ont été acquis sur les graines sèches d'une part et sur des poudres issues de feuilles séchées puis broyées issues des expérimentations en tunnel. L'appareil utilisé pour les acquisitions est le spectrophotomètre MPA II FT NIR (Bruker Optic Inc, Allemagne). Pour un type d'échantillon, tous les spectres ont été acquis dans une durée limitée afin de réduire les effets de stockage. Chaque acquisition spectrale résulte de la moyenne de 64 mesures techniques. Pour les graines, trois répétitions ont été réalisées par lot (1 lot par génotype). Pour les poudres de feuilles, les données proviennent de 6 plantes indépendantes récoltées dans un même caisson.



---

## **CHAPITRE II**

**DETERMINANTS GENETIQUES DE LA**

**GERMINATION ET PREDICTIONS**

---



## II.1. Préambule

---

Ce chapitre présente les résultats obtenus pour les deux questions de recherche de la thèse :

1. **Quel est le déterminisme génétique de la germination ?**
2. **Ces traits étant potentiellement polygéniques, est-il possible de les prédire et si oui, comment ?**

Pour cela, nous avons utilisé les données de germination acquises sur 223 génotypes de colzas d'hiver et de printemps. Dans un premier temps, nous avons étudié la relation entre plusieurs traits liés à la germination. Puis nous avons étudié le contrôle génétique de ces traits. Enfin, nous avons comparé différents modèles de prédiction génomique et phénomique afin d'évaluer leurs capacités prédictives pour les traits de germination.

Les résultats majeurs de cette partie sont les suivants :

- L'identification de 17 QTL impliqués dans la germination, illustrant la nature polygénique de ce processus.
- La prédiction génomique et la prédiction phénomique permettent d'obtenir des capacités prédictives modérées à élevées pour la prédiction de traits liés à la germination.
- La prédiction phénomique semble moins impactée par la structure génétique que la prédiction génomique.

Les résultats sont présentés sous la forme d'un article, publié dans le journal *Theoretical and Applied Genetics* en 2024.

Les figures et tableaux supplémentaires de l'article sont disponibles via le lien : [https://static-content.springer.com/esm/art%3A10.1007%2Fs00122-024-04659-0/MediaObjects/122\\_2024\\_4659\\_MOESM1\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1007%2Fs00122-024-04659-0/MediaObjects/122_2024_4659_MOESM1_ESM.xlsx)

Les jeux de données sont téléchargeables via le lien : <https://doi.org/10.57745/JAAO8R>



## II.2. Article

---

Laurençon, M., Legrix, J., Wagner, MH., Demilly, D., Baron, C., Rolland, S., Ducournau, S., Laperche, A., Nesi, N. Genomic and phenomic predictions help capture low-effect alleles promoting seed germination in oilseed rape in addition to QTL analyses. *Theor Appl Genet* **137**, 156 (2024).  
<https://doi.org/10.1007/s00122-024-04659-0>





# Genomic and phenomic predictions help capture low-effect alleles promoting seed germination in oilseed rape in addition to QTL analyses

Marianne Laurençon<sup>1</sup> · Julie Legrix<sup>1</sup> · Marie-Hélène Wagner<sup>2</sup> · Didier Demilly<sup>2</sup> · Cécile Baron<sup>1</sup> · Sophie Rolland<sup>1</sup> ·  
Sylvie Ducournau<sup>2</sup> · Anne Laperche<sup>1</sup> · Nathalie Nesi<sup>1</sup>

Received: 23 February 2024 / Accepted: 25 May 2024

© The Author(s) 2024

## Abstract

**Key message** Phenomic prediction implemented on a large diversity set can efficiently predict seed germination, capture low-effect favorable alleles that are not revealed by GWAS and identify promising genetic resources.

**Abstract** Oilseed rape faces many challenges, especially at the beginning of its developmental cycle. Achieving rapid and uniform seed germination could help to ensure a successful establishment and therefore enabling the crop to compete with weeds and tolerate stresses during the earliest developmental stages. The polygenic nature of seed germination was highlighted in several studies, and more knowledge is needed about low- to moderate-effect underlying loci in order to enhance seed germination effectively by improving the genetic background and incorporating favorable alleles. A total of 17 QTL were detected for seed germination-related traits, for which the favorable alleles often corresponded to the most frequent alleles in the panel. Genomic and phenomic predictions methods provided moderate-to-high predictive abilities, demonstrating the ability to capture small additive and non-additive effects for seed germination. This study also showed that phenomic prediction estimated phenotypic values closer to phenotypic values than GEBV. Finally, as the predictive ability of phenomic prediction was less influenced by the genetic structure of the panel, it is worth using this prediction method to characterize genetic resources, particularly with a view to design prebreeding populations.

## Introduction

Oilseed rape (*Brassica napus* L.) is the leading oilseed crop in Europe (FAO 2023), but it faces numerous constraints, especially in a context of climate change and reduced chemical inputs. One of the keys to overcome these challenges would be to ensure successful plant establishment, enabling the crop to compete with weeds and tolerate stresses during the earliest developmental stages in order to guarantee

a high-yield potential (Nelson et al. 2022). Indeed, oilseed rape establishment is affected by a wide range of biotic and abiotic stresses, such as seed- and soil-borne pathogens causing damping-off, pest attacks reducing young plantlet biomass (slugs and flea beetles), and weed competition as well as environmental limiting factors, especially water (lack or excess) and extreme temperatures that can occur after sowing or during plant emergence (Haj Sghaier et al. 2022). These stresses can dramatically reduce plant density and biomass thus, ultimately affecting yield. Successful plant establishment results from a combination of developmental processes of which germination is the first step (Rajjou et al. 2012). As a matter of fact, rapid and highly efficient germination combined to uniform plant emergence strongly increase the likelihood of plant establishment, which is expected especially under adverse environmental conditions.

Seed germination is a complex trait subject to genetic and environmental controls. The environmental effect is generally referred to the so-called ‘seed-lot’ effect that encompasses the conditions during seed development on the mother plant as well as the post-harvest conditions (Rajjou

---

Communicated by Matthew N Nelson.

---

Co-last authors: Anne Laperche and Nathalie Nesi.

✉ Anne Laperche  
anne.laperche@institut-agro.fr

<sup>1</sup> Institute of Genetics, Environment and Plant Protection (IGEPP), INRAE – Institut Agro Rennes-Angers – Université de Rennes, 35650 Le Rheu, France

<sup>2</sup> Groupe d’Etude et de Contrôle des Variétés Et des Semences (GEVES), 49070 Beaucouzé, France



et al. 2012; Finch-Savage and Bassel 2016). In addition, the cultural practices also lead to variation in germination ability when considering factors such as the sowing date, seedbed preparation, intercropping, and choice of previous crop (Elliott et al. 2007; Lamichhane et al. 2018). Several studies have reported the polygenic nature of seed germination, which is under the control of numerous loci with moderate-effect each (Hatzig et al. 2015; Nguyen et al. 2018; Gad et al. 2021; Luo et al. 2021).

Modeling low additive effects distributed along the genome can be addressed through the recent development of prediction methods. The benchmark article of Meuwissen et al. (2001) conceptualized the prediction of complex traits using genotypic data, called genomic prediction (GP). GP combine phenotypic data and high-density molecular markers obtained on genotypes from a training population, to predict the genomic estimated breeding values (GEBV) of non-phenotyped genotypes from a testing population. Combined with GWAS approach, GP can accelerate genetic gain in breeding (Hickey et al. 2017; Araus et al. 2018). Other emerging methods have focused on using endophenotypes as trait predictors to complement or even to replace molecular data. This enables to improve not only the modeling of additive effects, but also of non-additive and epistatic effects (Mackay et al. 2009; Patti et al. 2012; Ritchie et al. 2015). Multiple studies integrated transcriptomic and metabolomic data, as endophenotypes, into predictive models, either in combination with genomic data or not (Westhues et al. 2017; Schrag et al. 2018; Knoch et al. 2021). These integrations complemented the information provided by molecular markers and have demonstrated the ability to better capture small genetic effects. Rincent and coworkers (2018) proposed a low-cost and high-throughput method as an alternative to the use of -omics data, based on near-infrared spectroscopy (NIRS), called phenomic prediction (PP).

Up to now, most of the studies reporting the use of GP and PP models have focused on traits related to seed yield, seed quality, or plant phenology mainly in cereal crops (Heffner et al. 2011; Albrecht et al. 2011; Rincent et al. 2018; Voss-Fels et al. 2019; Lane et al. 2020; Robert et al. 2022), with additional examples in other crops such as rapeseed (Würschum et al. 2014; Werner et al. 2018; Knoch et al. 2021) or other species such as trees (Resende et al. 2012; Muranty et al. 2015; Isik et al. 2016; Rincent et al. 2018; Brault et al. 2022). Considering seed germination phenotype, a first study reports the interest of GP to decipher the polygenic effect of seed germination capacity in barley (Rooney et al. 2022), with moderate-to-high predictive abilities.

As seed germination has rarely been considered as a selection trait in the past, it is very likely that certain genes and alleles of interest are present in both recent and ancient germplasms. Therefore, we investigate the genetic determinism of seed germination-related traits in rapeseed and the

predictive ability of GP and PP models for these traits in a broad genetic pool including ancient and recent germplasm of winter, semi-winter, and spring type. The combination of different germplasms provided potential access to favorable low-effect polygenes. Our goal was to investigate whether GP and PP were promising methods to improve germination capacity in addition to GWAS. Our results led to (i) identify relationships between germination-related traits, (ii) decipher the genetic control of these traits, and (iii) compare multiple models of GP and PP, integrating one or multiple predictors, to evaluate the predictive abilities for seed germination-related traits.

## Material and methods

### Plant material and genotyping data

The diversity set consisted of 223 genetically diverse *B. napus* inbred lines, including 127 winter oilseed rape (WOSR), 81 spring oilseed rape (SOSR), 13 winter fodder (WFR), and two swedes (Online Resource 1). Seed lots were all produced in the field by open pollination during the 2020/2021 season in Le Rheu, France. Each genotype was sown in a four-row plot. To avoid pollen mixing, the outer rows of each plot were discarded at harvest. All the genotypes were sown and harvested at the same time. Seeds were then stored under the same conditions. After harvest, seed lot was split into sub-lots each dedicated to a specific analysis or experiment (i.e., seed germination monitoring, TSW measurements, and acquisitions of NIRS spectra) as described below. All the accessions were genotyped using the *Brassica* 60 K Infinium® SNP array (Illumina, Inc., San Diego, CA) (Clarke et al. 2016), and the data were visualized using GenomeStudio software (Illumina, Inc., San Diego, CA). A total of 33,151 SNPs were validated using thresholds of 5% for the minor allele frequency (MAF), 10% for the frequency of missing values, and 10% for the heterozygosity level. The missing SNP data were imputed using BEAGLE software following the method described in Browning et al. (2018). SNPs were physically anchored to the *B. napus* Darmor-bzh v10 reference genome (Rousseau-Gueutin et al. 2020).

### Acquisition of germination data

Seed germination dynamics were monitored using the high-throughput phenotyping platform for germination at the National Seed Testing Station in Angers, France (PHE-NOTIC—Angers Seed Phenotyping Facility, Boureau 2020). Seeds were imbibed under controlled conditions at 20 °C, in the dark, during 96 h. The experimental design consisted of blocks of twenty-five seeds per genotype, repeated four times



(Ducournau et al. 2004; Wagner et al. 2011). Imaging and analysis methods are described in Demilly et al. (2014), with pictures taken every 2 h over the experiment.

Germination parameters were monitored for each block of twenty-five seeds as following (Table 1): seed volume increase after 20 h of imbibition (VI, in %), final germination percentage after 96 h (GP, in %), first germination time (FG, in h) corresponding to the time when the first seed germinates, mean germination time (MGT, in h) corresponding to the mean delay to germinate for each seed lot, and radicle elongation speed (ES, in mm/h). Supplementary parameters were deduced from the germination dynamics, such as germination after 36 h of imbibition (GP36, in %), time to reach 20% of germination (T20, in h), time to reach 50% of germination (T50, in h), time to reach 80% of germination (T80, in h), uniformity (UNI) assessed as the difference between T80 and T20, and the area under the curve (AUC) that represents the germination rate as a function of the time from the initiation of the imbibition. AUC was estimated for each genotype using the values for the 100 seeds of the four blocks. Thousand seed weight (TSW, in g) was also measured for each seed lot by weighting seed samples after drying at 105 °C overnight.

### Phenotypic heritability

Broad-sense heritability ( $H^2$ ) was calculated for each phenotypic trait. Genotypic and error variance were extracted from the following linear mixed model:

$$Y_i = \mu + G_i + e_i,$$

where  $Y_i$  is the trait value obtained for the genotype  $i$ ,  $G_i$  corresponds to the random genetic effect for genotype  $i$ , and  $e_i$  is the residual effect. We assumed that  $G_i$  and  $e_i$  were independent, identically distributed and followed a normal distribution.

$H^2$  was then calculated as follows:  $H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_e^2}{n}}$ , where  $\sigma_G^2$  is the genotypic variance,  $\sigma_e^2$  is the error variance, and  $n$  is the number of repetitions.

### NIRS data

NIR spectra were collected on a sub-set of dry seed samples, that originated from the same seed lots as the sub-set used for the germination phenotyping experiments. A MPA FT-NIR spectrophotometer (Bruker Optic Inc., Germany) was used over the range of 4000–12,000 cm<sup>-1</sup> with a 16 cm<sup>-1</sup> optical resolution. NIRS was converted in nm in steps of 1 nm, so the final spectra range from 800 to 2781 nm. Three biological replicates were run per genotype, with each replicate being the average of 64 technical repetitions measured by the spectrophotometer. Due to a lack of seeds, only 210 genotypes were screened (Online Resource 1). NIR spectra were centered and scaled to reduce noise. Then, the first derivative of the Savitzky–Golay filter (Savitzky and Golay 1964) was calculated using the R package signal (Signal Developers 2014) to smooth the curve. The three replicates were then averaged to obtain a mean NIR spectrum per genotype.

**Table 1** Variation of the germination-related traits assessed over a diversity germplasm of 223 rapeseed genotypes

Trait	Abb (unit)	Formula	Min	Max	Mean	SD	$H^2$
Thousand seed weight	TSW (g)		2.550	6.250	4.358	0.623293 /	
Seed volume increase after 20 h of imbibition	VI (%)	Volume <sub>20 h</sub> –Volume <sub>0 h</sub>	0.997	4.440	2.145	0.4961370.40	
Time of the 1st germ	FG (h)	$\frac{\sum_{i=1}^{n_{rep}} t_{1st\ germination}}{n_{rep}}$	22.00	58.00	30.25	4.8051320.51	
Time to 20% of germ	T20 (h)		28.50	64.00	52.26	7.2529350.69	
Time to 50% of germ	T50 (h)		36.00	79.67	51.93	8.8558270.65	
Time to 80% of germ	T80 (h)		43.20	96.00	63.55	11.157510.61	
Mean germ. time	MGT (h)		38.12	75.43	52.26	6.8492420.68	
Uniformity	UNI (h)	T80-T20	11.53	50.00	24.04	7.4721020.43	
Area under the curve	AUC (%·h <sup>-1</sup> )	$\int_{h_2}^{h_1} f_h dh$	1444	5888	4077	972.7179 /	
Germ. percentage at 36 h	GP36 (%)	$\frac{n_{germinated\ seeds\ at\ 36h}}{n_{total}} * 100$	0.00	52.00	15.49	10.352490.69	
Germ. percentage at 96 h	GP (%)	$\frac{n_{germinated\ seeds\ at\ 96h}}{n_{total}} * 100$	46.67	100	49.46	10.919780.46	
Radicle elongation speed	ES (mm·h <sup>-1</sup> )	$\frac{L_{radicle\ at\ 16h}}{16h}$	0.080	0.209	0.149	0.0267880.56	

Abb trait abbreviation Min, Max, Mean, and SD minimum, maximum, mean, and standard deviation for fitted values  $H^2$  broad-sense heritability



## Exploratory analyses and correlations

To explore the relationship between the different seed germination traits, a principal component analysis (PCA) was performed using FactoMineR package (Lê et al. 2008). The number of dimensions was chosen according to the percentage of explained variance. Then, a clustering was carried out, using the partitioning around medoids (PAM) method (Kaufman and Rousseeuw 1990). The number of clusters ( $k$ ) was chosen using silhouette index and gap statistic from cluster R package (Maechler et al. 2022). Due to non-linear relationships, Spearman correlations were calculated between traits.

## Population genetic structure and diversity

Genetic structure of the germplasm was unraveled using the first two components of a principal coordinates analysis (PCoA, Gower 1967) carried out using the SNP data. The PCoA is based on the genetic distances between genotypes estimated by the dissimilarity matrix ( $1-K$ ). The genomic kinship ( $K$ ) was estimated using Astle and Balding (2009) algorithm as following:

$$K = \frac{WW^T}{M}$$

With  $W$ , the matrix scaled on allelic frequencies with dimensions  $N \times M$ ,  $N$  is the number of genotypes and  $M$  is the number of molecular markers and  $W^T$  is the transposed  $W$  matrix.

Nucleotidic diversity ( $\pi$ ) per chromosome was estimated using VCFtools (Danecek et al. 2011).

## Genome-wide association study

For the GWAS only, a kinship matrix was re-estimated for each chromosome tested as described by Rincént et al. (2014). The estimation of the kinship matrix ( $K$ ) is similar to the methodology previously described, except that SNPs located on the chromosome under investigation were discarded. The pairwise linkage disequilibrium (LD) between SNPs ( $r^2$ ) was estimated per chromosome using PLINK software v.1.9 (Purcell et al. 2007). LD decay according to the physical distance between markers for each chromosome is represented in Online Resource 2.

GWAS was performed using FaST-LMM algorithm (Lipper et al. 2011). To reduce false associations (types I and II errors) due to population structure and kinship between genotypes, GWAS was performed with a mixed model that takes these two factors into account (Yu et al. 2006). Due to high average LD calculated between SNP pairs intra- and

inter-chromosomes (mean LD=0.51), the SimpleM approach (Gao et al. 2008, 2010) was used to reduce type I error (false-positive) by estimating the effective number of independent tests ( $M_{\text{eff}}$ ) based on composite LD for the GWAS. Markers were considered as significantly associated with a trait if the  $-\log_{10}(p\text{-value})$  exceeded a 5% threshold (using our data, the 5% threshold was 3.51). In addition, false discovery rate (FDR) adjusted  $q$ -values were calculated to reduce type I errors using the  $q$  value R package (Storey 2002). Markers with a  $q$ -value inferior to 0.2 were retained.

QTL confidence intervals were estimated based on the method proposed by Albert et al. (2016). Briefly, LD calculation was performed between 100,000 randomly chosen pairs of unlinked loci located on different chromosomes. The critical LD threshold was chosen as the 95th percentile of the LD distribution, which equaled 0.16 using our data. Then, local pairwise LD with markers located upstream and downstream the significant marker (on the same chromosome) was calculated in the same way as pairwise LD explained above. Confidence intervals were estimated for each significant marker as the interval between the first (upstream) and last (downstream) markers that presented a LD value higher than the LD threshold when compared to the significant marker.

Effects of each single QTL ( $R^2$  and allelic effect) were estimated at the peak marker that corresponded to the marker with the smallest p-value. Favorable alleles were identified for each QTL as the allele that improved the phenotype, that is to say increasing GP36, TSW, and VI values but decreasing MGT, T20, and T50 values. The number of favorable alleles across all QTL locations for each genotype and trait was calculated, ranging from 0 to the maximum number of QTL identified for the trait. An ANOVA and Tukey test at 5% probability level were performed to compare the effect of accumulating favorable alleles at QTL for each trait.

The genes under the identified QTL were obtained from the *B. napus* Darmor-*bzh* V10 annotation file (Rousseau-Gueutin et al. 2020). The function of candidate genes was obtained from TAIR (<https://www.arabidopsis.org/>).

## Genomic and phenomic predictions

Genomic heritability along NIR-based spectra was estimated for each wavelength from a statistical model considering a random polygenic effect. Genotypic and error variance were extracted from the following linear mixed model:

$$Y_{ij} = \mu + G_i + e_{ij},$$

where  $Y_{ij}$  is the spectrum value obtained for the genotype  $i$ ,  $G_i$  corresponds to the random genetic effect for genotype  $i$ , following a normal distribution  $G \sim N(0, K\sigma_G^2)$  with  $K$  the kinship (see above), and  $e_{ij}$  is the residual effect. We



assumed that  $G_i$  and  $e_i$  were independent, identically distributed and followed a normal distribution.

Genomic heritability was then calculated as follows:  $H_{\text{genomic}}^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \hat{\sigma}_e^2}$ , with  $\hat{\sigma}_G^2$  and  $\hat{\sigma}_e^2$ , the REML estimates of  $\sigma_G^2$  and  $\sigma_e^2$  obtained using the R package sommer (Covarrubias-Pazaran 2016).

A spectral matrix  $H$  was calculated to represent the kinship between genotypes based on the NIR spectra similarity.

$$H = \frac{SS^T}{N_w}$$

With  $S$ , the raw NIR spectra matrix of dimension  $N \times N_w$ . The  $S$  matrix gathered the value of absorbance pretreated as presented above, for each genotype and each wavelength. Values were also centered and scaled.  $N$  represents the number of genotypes, and  $N_w$  represents the number of wavelengths. Mantel test using 999 permutations was realized to compare the spectral relationship matrix  $H$  with the genomic kinship matrix  $K$ .

GBLUP and HBLUP (using the  $K$  or  $H$  matrix, respectively) were used to predict seed germination-related traits which are defined as follows:

$$\text{GBLUP } Y_i = \mu + G_i + \varepsilon_i \quad (1)$$

$$\text{HBLUP } Y_i = \mu + W_i + \varepsilon_i \quad (2)$$

where  $Y_i$  is the mean phenotype value for the genotype  $i$ ,  $\mu$  is the intercept,  $G_i$  or  $W_i$  is the random genetic effect following a normal distribution  $G \sim N(0, K\sigma_G^2)$  with  $K$  the genomic kinship matrix (see above) or  $W \sim N(0, H\sigma_G^2)$  with  $H$  the hyperspectral matrix, and  $\varepsilon$  is the random residual effect following  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ .

Both genomic ( $K$ ) and spectral ( $H$ ) matrices were included simultaneously in a GHBLUP model by integrating two variance–covariance matrices as follows:

$$\text{GHBLUP } Y_i = \mu + G_i + W_i + \varepsilon_i, \quad (3)$$

with  $G \sim N(0, K\sigma_G^2)$ , and  $W \sim N(0, H\sigma_W^2)$ .

GP and PP models were assessed for each trait using cross-validation (CV). Training set was composed of four-fold over five of genotypes chosen randomly, and the testing set was composed of the remaining fold. CV was repeated 100 times. These models were fitted using sommer R package (Covarrubias-Pazaran 2016). PA were estimated for each model by the Pearson correlation between the observed and the predicted values of the validation set.

By homology with the values estimated by genomic prediction, called GEBV, we called the values estimated by phenomic prediction PEPV, for phenomic estimated phenotypic values.

## Results

### Phenotypic distribution and correlations between traits

All germination-related traits displayed moderate-to-high heritabilities (Table 1). The highest heritabilities ( $> 0.60$ ) were observed for GP36, T20, T50, T80, and MGT, and the lowest heritabilities were observed for VI (0.40) and UNI (0.43). Heritabilities could not be estimated for TSW and AUC as only one value per genotype was measured. GP, GP36, T20, T50, T80, AUC, and MGT were highly correlated as shown in Table 2. GP, AUC, and GP36 were

**Table 2** Correlation coefficients for trait pairs associated with seed germination scored on 223 rapeseed genotypes

	AUC	TSW	MGT	T20	T50	T80	UNI	GP	GP36	FG	VI
TSW	0.08										
MGT	<b>-0.97***</b>	-0.06									
T20	<b>-0.94***</b>	-0.04	<b>0.96***</b>								
T50	<b>-0.96***</b>	-0.05	<b>0.96***</b>	<b>0.92***</b>							
T80	<b>-0.91***</b>	-0.12	<b>0.91***</b>	<b>0.84***</b>	<b>0.88***</b>						
UNI	<b>-0.73***</b>	<b>-0.19**</b>	<b>0.67***</b>	<b>0.53***</b>	<b>0.67***</b>	<b>0.88***</b>					
GP	<b>0.89***</b>	0.10	<b>-0.77***</b>	<b>-0.76***</b>	<b>-0.80***</b>	<b>-0.78***</b>	<b>-0.70***</b>				
GP36	<b>0.89***</b>	0.07	<b>-0.93***</b>	<b>-0.95***</b>	<b>-0.88***</b>	<b>-0.81***</b>	<b>-0.50***</b>	<b>0.68***</b>			
FG	<b>-0.61***</b>	-0.08	<b>0.6***</b>	<b>0.63***</b>	<b>0.55***</b>	<b>0.57***</b>	<b>0.40***</b>	<b>-0.55***</b>	<b>-0.62***</b>		
VI	<b>0.27***</b>	<b>0.52***</b>	<b>-0.29***</b>	<b>-0.29***</b>	<b>-0.27***</b>	<b>-0.23***</b>	<b>-0.18**</b>	<b>0.18**</b>	<b>0.28***</b>	-0.11	
ES	<b>0.35***</b>	<b>0.19*</b>	<b>-0.43***</b>	<b>-0.40***</b>	<b>-0.37***</b>	<b>-0.39***</b>	<b>-0.25***</b>	<b>0.20**</b>	<b>0.42***</b>	<b>-0.22**</b>	<b>0.14*</b>

Significant values are shown in boldface

Significances were calculated at levels of: \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$

For details of trait abbreviations refer to Table 1 and “Material and methods”



positively correlated as well as T20, T50, T80, and MGT, which can be partially explained by the non-independence of traits. However, these two groups of variables presented negative correlations one with each other. VI and ES presented significant but moderate correlations with other traits (Table 2). TSW was only significantly correlated to UNI ( $-0.19$ ), ES ( $0.19$ ), and VI ( $0.52$ ). These correlations were strengthened by the correlation circle of the PCA (Fig. 1). Indeed, AUC, GP, GP36, T80, T20, T50, and MGT were highly correlated to the first axis that gathered 60.42% of the variability while TSW and VI were correlated to the second axis that explained 12.43% of the variability. ES, UNI, and FG were less correlated to these two principal components.

### Germplasm structure using genomic, phenotypic, and spectral data revealed different patterns

Genetic structure of the germplasm was assessed using the first two axes of the principal coordinates analysis using SNP (PCoA), explaining 48% of the genetic variance (Fig. 2a). The first axis that encompasses 34% of genetic variance clearly discriminated winter and spring types. The second axis (14% of the genetic variance) separated Asian genotypes on the one side and European and American genotypes on the other side. As a whole, four genetic clusters were identified using the PCoA clustering. Cluster 1 ( $n=114$ , red) was mostly composed of European WOSR and WFR but also included some Asian WOSR. Cluster 2 ( $n=62$ , green) was mostly composed of European and North American

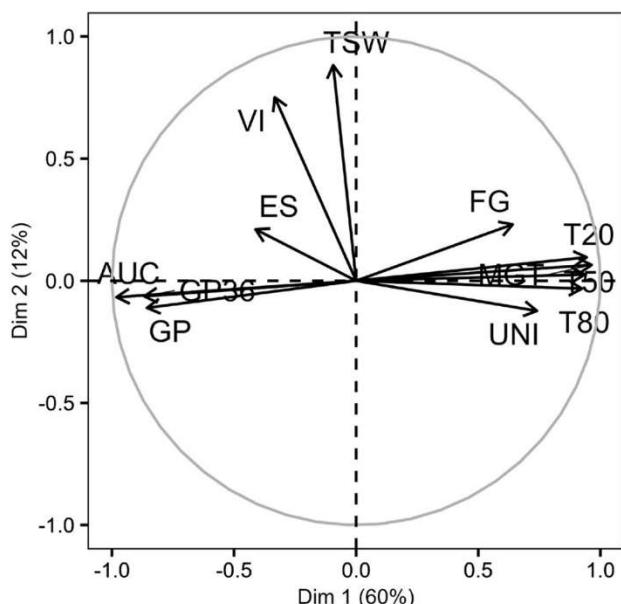
SOSR, but also gathered some Oceanian SOSR. Cluster 3 ( $n=23$ , blue) corresponded to Asian SOSR and WOSR. Finally, cluster 4 ( $n=24$ , purple) was composed of European winter fodder, Asian, and European WOSR. The mean nucleotide diversity  $\pi$  was computed on the whole germplasm and equaled  $7.69e^{-05}$ , indicating an important genetic diversity within the germplasm used in this study.

The relatedness between genotypes was also estimated using NIRS data (Fig. 2b). Heritabilities of the NIR spectra ranged between 0.00 and 0.95 regarding the wavelengths, with 34% and 16% of the wavelengths having a heritability superior or equal to 0.50 and 0.70, respectively (Online Resource 3). This indicates that NIRS is suitable for further genetic analysis. However, no cluster of genotypes was evidenced based on the NIRS data (Fig. 2b). In addition, when the genotypes were labeled according to their genetic cluster, no specific pattern was highlighted. Indeed, the correlation calculated using Mantel test between genomic ( $K$ ) and hyperspectral ( $H$ ) kinship matrices was equal to 0.051.

Based on the germination data, three phenotypic clusters of genotypes representing their seed germination performance were identified by PAM clustering (Fig. 2c). The first phenotypic cluster (from the left, light gray) corresponded to genotypes with a high germination capacity (high GP, GP36, AUC and low UNI, MGT, T80, T20, T50). The second phenotypic cluster (medium gray) consisted of genotypes with an intermediate behavior. Finally, the third cluster (dark gray) was composed of genotypes with poor germination ability (low GP, GP36, AUC and high MGT, T20, T50, T80, UNI). Each cluster gathered SOSR and WOSR types (Fig. 2c). Genomic clusters C1 and C4, which represented WOSR genotypes, were predominantly present in phenotypic clusters 1 (65% and 19%, respectively) and 2 (50 and 5%), i.e., clusters grouping genotypes with medium-to-high germination capacity. In contrast, genotypes from these genomic clusters (C1 and C4) were scarce in phenotypic cluster 3 (25 and 0%, respectively). Genotypes from clusters C2 and C3 were present at low frequency in the first phenotypic cluster (10 and 6%, respectively) and more frequent in the phenotypic clusters 2 (33 and 12%, respectively) and 3 (57 and 18%, respectively). Therefore, a link between the type or origin of rapeseed genotypes and their seed germination behavior was revealed.

### Identification of multiple QTL confirmed the polygenic determinism of seed germination

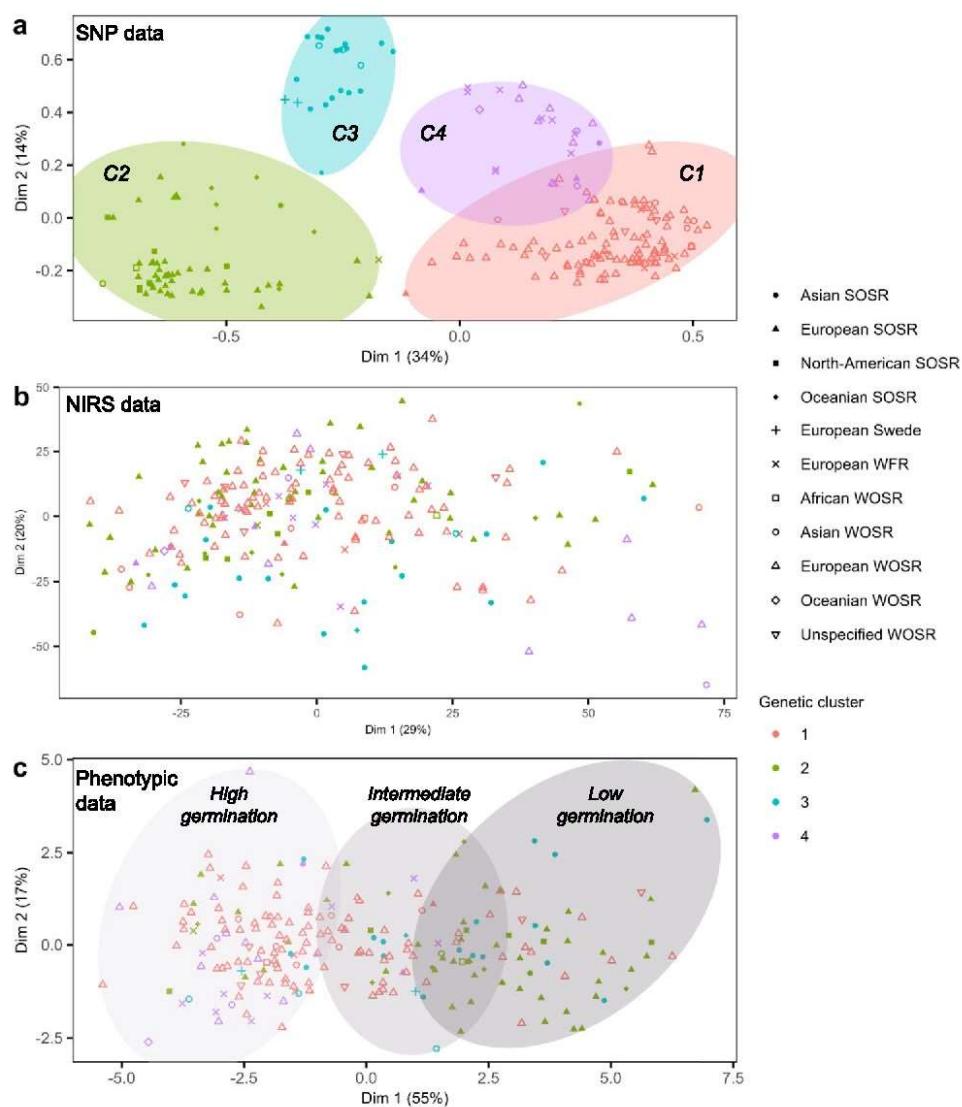
A total of 17 unique regions (QTL) were detected by GWAS for seed germination traits (GP36, MGT, T20, T50, and VI) and thousand seed weight (TSW), with an explained variance ( $R^2$ ) ranging from 6.30 to 8.78% (Table 3). Among these 17 QTL, three were identified for four germination traits (GP36, MGT, T20, and T50) on



**Fig. 1** Plot of principal components for 12 variables associated with seed germination scored on 223 genotypes. Trait description is provided in Table 1



**Fig. 2** Genetic and phenotypic characterization of the panel used. Relatedness of 223 genotypes of rapeseed is highlighted by a principal coordinate analysis (PCoA) using the first two components using a SNP and b NIRS data acquired on seed samples. c PAM clustering on principal components of 12 variables associated with seed germination measured on 223 genotypes. Unfilled symbols correspond to winter types and plain symbols correspond to spring types



A06 (QTL.A06.1), A09 (QTL.A09.1), and C02 (QTL.C02.2). QTL.C01.1 was identified for GP36 and MGT, QTL.A06.2 for MGT and T50, and QTL.C04.2 for MGT and T20. Example of Manhattan and QQ plots for GP36 is available in Online Resource 4. Few QTL detected were trait-specific: three for GP36 (QTL.A07.1, QTL.C02.1, and QTL.C07.1), three for MGT (QTL.A10.1, QTL.C04.1, and QTL.C05.1), two for T50, (QTL.A07.2 and QTL.C07.2), one for TSW (QTL.A10.2), and two for VI (QTL.C03.1 and QTL.C09.1) (Table 3). TSW QTL and VI QTL did not colocate with any QTL detected for the other traits. For most of the QTL, the favorable allele was the major allele in the germplasm (Table 3). Indeed, this was the case for one of the two QTL for VI, 5 QTL out of 9 for MGT, 3 QTL out of 4 for T20, and 5 QTL out of 6 for T50. On the contrary, for GP36, 5 QTL out of 7 had the minor allele as the favorable one. On the same line, the only QTL detected for TSW presented the minor allele

as the favorable one. For all traits, a positive effect of the accumulation of favorable alleles was observed (Online Resource 5).

A search for candidate genes within the confidence intervals of the quantitative trait loci (QTL) identified two genes involved in germination (Online Resource 6): A10p30660.1\_BnaDAR, located under QTL.A10.2, is an ortholog of *A. thaliana* gene involved in seed germination and radicle development, ARGINYLY-t-RNA PROTEIN TRANSFERASE 1 (ATE1); and C03p76020.1\_BnaDAR, located under QTL.C03.1, is an ortholog of another *A. thaliana* gene GDSL-motif lipase 1 (GDSL1).

#### Relative performances of PP and GP at the whole panel level

Then, VI, ES, and TSW were dropped from the further analyses as they were weakly correlated to sensu stricto



**Table 3** QTL controlling germination-related traits detected in a population of 223 rapeseed through GWAS. QTL were ordered regarding their position on the genome. For each QTL, information is given about the peak and flanking markers, the peak SNP position, QTL size, as well as major and minor allele, followed by the minor

allele frequency (MAF). For each trait, the SNP weight indicating the favorable allele version and the phenotypic variance explained by the SNP ( $R^2$ ) are represented. Positions are relative to the reference genome Darmor-*bzh* V10 (Rousseau-Gueutin et al. 2020)

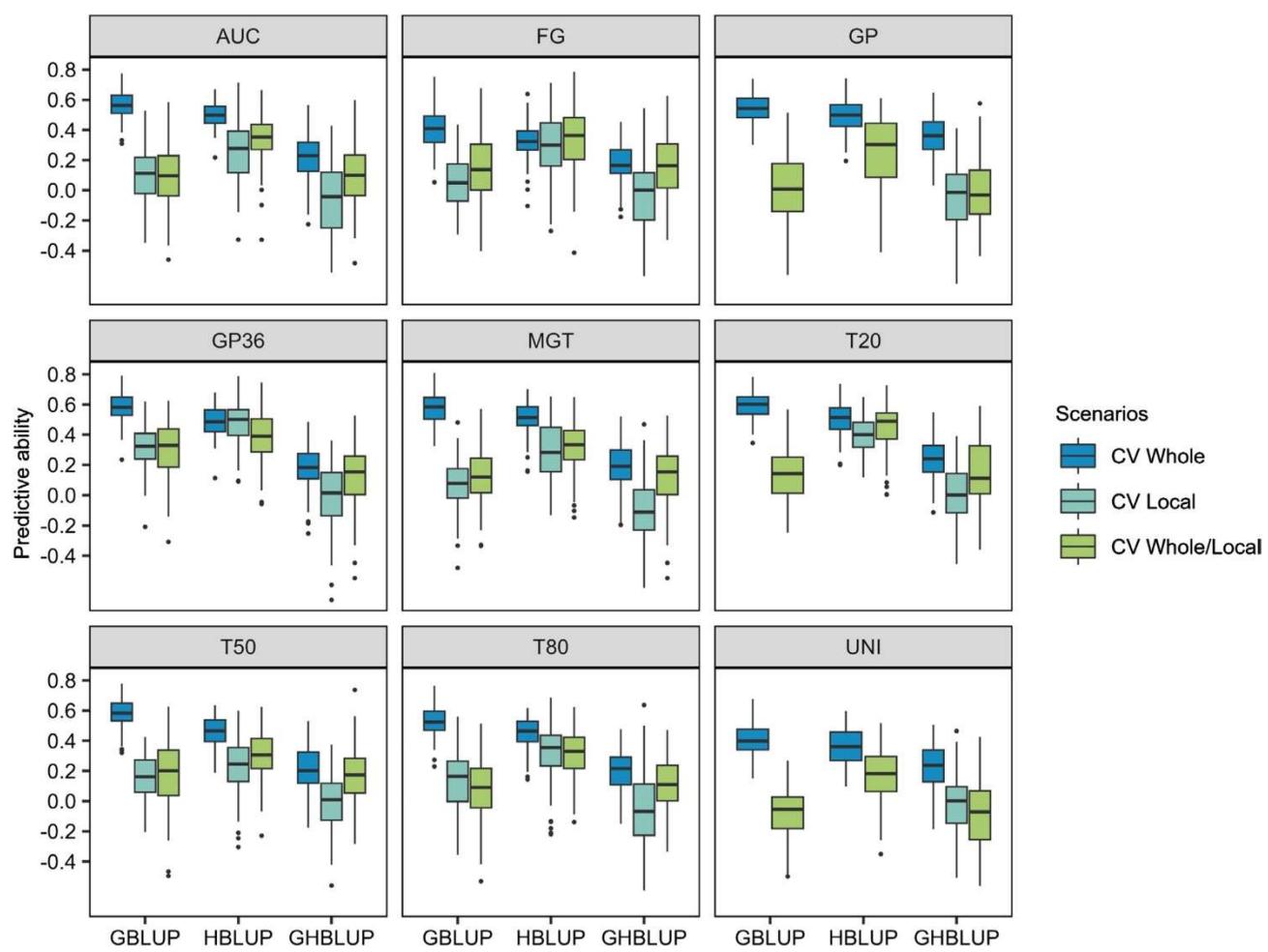
QTL name	Chr	Trait	Peak SNP Peak position (pb)	Flanking marker 1/ Flanking marker 2	QTL size (kb)	Major/ minor allele	MAF	SNP weight	$R^2$
QTL.A06.1	A06	GP36	Bn_A06_p836912 (1,102,221)	Bn_A06_p709547/Bn_A06_p908271	279	T/C	0.15	<b>2.60</b> <b>-1.83</b>	7.60 8.46
		MGT	T20					<b>-1.81</b>	7.64
			T50					<b>-2.38</b>	8.78
QTL.A06.2	A06	MGT	Bn_A06_p14121814 (34,532,655)	Bn_A06_p19574863/Bn_A06_p18040423	4272	T/C	0.20	1.78 2.25	7.99 7.78
QTL.A07.1	A07	GP36	Bn_A07_p15220175 (21,453,055)	Bn_A07_p14390667/Bn_A07_p15885711	1530	A/G	0.25	<b>2.91</b>	8.59
QTL.A07.2	A07	T50	Bn_A07_p16553147 (22,869,578)	Bn_A07_p16157761/Bn_A07_p17026444	909	G/T	0.13	2.19	7.62
QTL.A09.1	A09	GP36	Bn_A09_p2733282 (4,101,875)	Bn_A09_p951202/Bn_A09_p7560188	8824	G/A	0.45	-4.00 2.27 2.43 3.00	8.31 6.48 6.95 8.03
		MGT	T20						
			T50						
QTL.A10.1	A10	MGT	Bn_A10_p9459037 (13,661,179)	Bn_A10_p7835472/Bn_A10_p9645870	1990	C/T	0.27	-1.88	6.60
QTL.A10.2	A10	TSW	Bn_A10_p15838932 (18,901,950)	Bn_A10_p13817274/Bn_A10_p16933405	3226	G/T	0.28	<b>0.23</b>	7.44
QTL.C01.1	C01	GP36	Bn_scaff_15838_5_p446059 (3,600,398)	Bn_scaff_19244_1_p388798/ Bn_scaff_17731_1_p308944	6736	T/G	0.45	<b>3.07</b> <b>-1.91</b>	7.47 6.73
QTL.C02.1	C02	GP36	Bn_scaff_18507_1_p957588 (35,173,323)	Bn_scaff_18507_1_p737065/ Bn_scaff_17067_1_p669413	6469	A/G	0.13	<b>2.66</b>	6.99
QTL.C02.2	C02	GP36	Bn_scaff_16139_1_p720716 (61,523,315)	Bn_scaff_15918_1_p303294/ Bn_scaff_16139_1_p41217	5361	T/C	0.32	-3.13 2.15 2.16 2.47	7.31 8.30 7.75 7.04
		MGT	T20						
			T50						
QTL.C03.1	C03	VI	Bn_scaff_17440_1_p552871 (57,791,809)	Bn_scaff_19740_1_p203934/ Bn_scaff_16394_1_p2229812	5070	T/C	0.39	<b>0.21</b>	6.50
QTL.C04.1	C04	MGT	Bn_scaff_18440_1_p77767 (9,746,347)	Bn_scaff_27469_1_p45054/ Bn_scaff_16576_1_p202308	12,852	T/C	0.24	2.53	6.69
QTL.C04.2	C04	MGT	Bn_scaff_24979_1_p13103 (53,828,390)	Bn_scaff_19208_1_p476848/ Bn_scaff_20079_1_p289504	5714	T/C	0.36	1.88 1.93	7.05 6.73
QTL.C05.1	C05	MGT	Bn_scaff_16414_1_p1774629 (254,723)	Bn_scaff_23107_1_p7608/ Bn_scaff_15712_10_p173981	5460	C/T	0.38	<b>-2.01</b>	6.30
QTL.C07.1	C07	GP36	Bn_A07_p3467400 (15,046,703)	Bn_A07_p3467400/Bn_scaff_16721_1_p1551224	28	C/G	0.07	3.01	8.69
QTL.C07.2	C07	T50	Bn_scaff_16110_1_p1076892 (54,582,081)	Bn_scaff_16110_1_p2439331/Bn_scaff_16110_1_p1076892	1509	A/G	0.07	2.22	7.91
QTL.C09.1	C09	VI	Bn_A09_p694190 (579,848)	Bn_scaff_16486_1_p88236/ Bn_scaff_17297_1_p202207	6571	A/T	0.46	<b>-0.16</b>	6.84

SNP weight in boldface stands for minor alleles that are favorable



seed germination traits (Table 2). Predictions of seed germination traits were run either using genomic (GBLUP) or phenomic (HBLUP) prediction models (Fig. 3). Models that used a combination of the genomic and the spectral kinships were also tested (GHBLUP). In a first approach, we consider a first cross-validation (CV) scenario where the whole panel was used as training and prediction set ( $CV_{whole}$ ). More precisely, the training set included 80% of the whole population chosen randomly, the validation set being the remaining 20% of the whole population. Overall, low-to-medium predictive abilities (PA) were obtained, varying between 0.40 (FG, UNI) and 0.59 (T20) for GBLUP, 0.32 (FG) and 0.50 (T20, MGT) for HBLUP, and between 0.13 (GP36) and 0.34 (GP) for GHBLUP. For all traits, the PA of GBLUP appeared to be slightly higher or similar to the PA of HBLUP (Fig. 3).

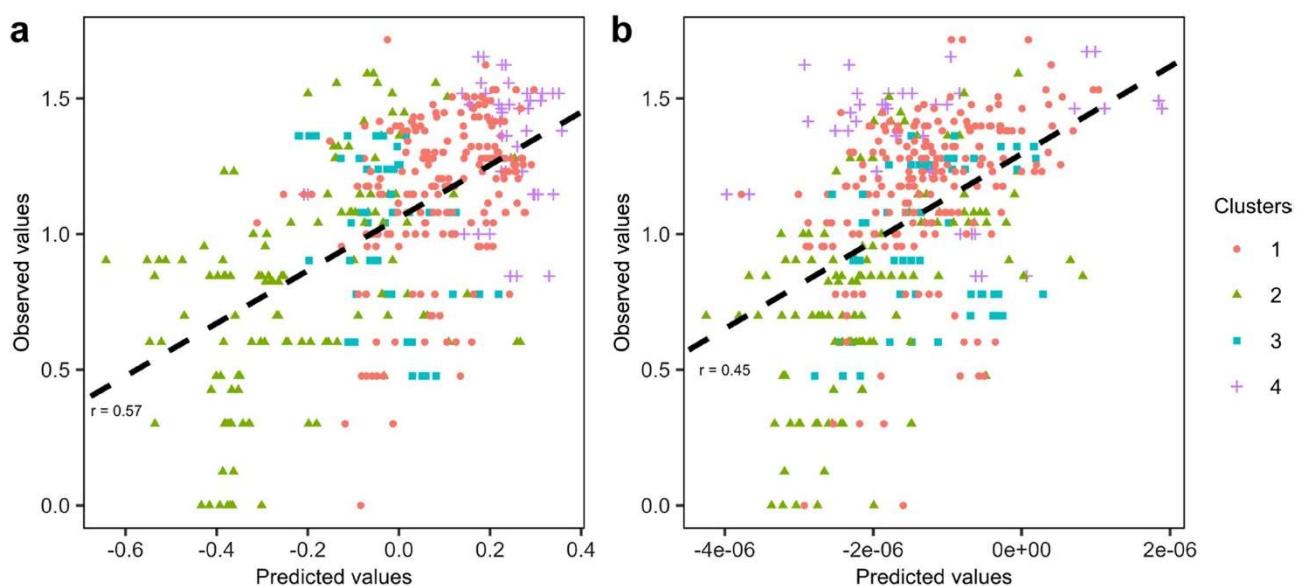
For instance, for GP36, the overall predictive ability of the genomic prediction was moderate (0.59), whereas the PA of HBLUP equaled 0.49. Furthermore, GHBLUP was less performant than GBLUP and HBLUP alone for all traits. Interestingly, when genotypes were labeled according to their genetic cluster affiliation, a separation between the different genetic clusters was revealed when looking at the GBLUP results, each genetic cluster corresponding to a specific stratum (Fig. 4). This result was observed for all traits (Online Resource 7). This highlighted the impact of the genetic structure on the genomic prediction model. Such pattern was less observed for the phenomic prediction model (Fig. 4, Online Resource 7). In conclusion, when the  $CV_{whole}$  scenario was considered, we found that PP was less performant but also less impacted by population structure than GP.



**Fig. 3** Predictive abilities of the nine seed germination-related traits using genomic (GBLUP), phenomic (HBLUP), or combined (GHBLUP) prediction models run for scenarios  $CV_{whole}$ ,  $CV_{local}$ , and  $CV_{whole/local}$ . Scenario  $CV_{whole}$  corresponds to the utilization of the whole germplasm for calibration and estimation of the PA, scenario  $CV_{local}$  corresponds to cluster 1 genotypes used for calibration and

estimation of the PA. Scenario  $CV_{whole/local}$  corresponds to the utilization of the whole germplasm for calibration, and only genotypes of the cluster 1 were used for PA estimation. PA were obtained from a fivefold cross-validation with 100 repetitions. Each boxplot indicates the mean (bold line), the first and third quartiles (boxes), and the first and ninth deciles (whiskers) (color figure online)





**Fig. 4** Predictive ability for GP36 using **a** genomic (GBLUP) and **b** phenomic (HBLUP) prediction models. Dots represent the observed and the predicted trait values across a fivefold cross-validation with 10 repetitions. Colors and dot types represent genomic clusters pre-

viously defined. The mean predictive ability ( $r$ ) calculated across the entire dataset using a fivefold cross-validation with 100 repetitions is represented by a black line

### Relative performances of PP and GP for unstructured population

As genetic structure influenced genomic predictions, particular attention was paid to investigating the most efficient way of predicting a specific genetic cluster. For this purpose, the genotypic cluster 1 was chosen as target population, as it is the only cluster that reached a sufficient size to allow model calibration ( $n=114$ , but only 103 genotypes used due to missing NIRS, Fig. 2a). The cluster 1 gathered most of the winter oilseed rape genotypes. Therefore, two scenarios were compared (Fig. 3). In the first scenario,  $CV_{local}$ , cluster 1 genotypes were used as calibration and validation sets. More precisely, 80% of cluster 1 genotypes were included in the training set, and the remaining 20% composed the validation set. In the second scenario,  $CV_{whole/local}$ , 80% of the whole panel was used in the training set, while the validation set was composed of the genotypes of the cluster 1 that was not present in the calibration set. Each scenario was repeated 100 times. Some PA could not be calculated due to non-convergent model repetitions. A first observation was that PA calculated for these two scenarios were lower than PA obtained for the whole panel (Fig. 3), which can be explained by the fact that the genomic structure of the panel caused a bias in the PA estimation. Furthermore, PA of GBLUP or HBLUP models for scenarios  $CV_{whole/local}$  and  $CV_{local}$  were similar. But we observed that HBLUP model provided higher PA than the other two models, for all traits. In the context of the  $CV_{local}$  scenario, the results

indicated that low-to-medium PA were obtained, with values ranging from 0.06 (FG) to 0.32 (GP36) for GBLUP and from 0.23 (T50) to 0.48 (GP36) for HBLUP. The highest difference in PA between the models was observed for FG (GBLUP PA = 0.09 and HBLUP PA = 0.25), while the lowest was observed for T50 (HBLUP PA = 0.23 and GBLUP PA = 0.15). In addition, GBLUP models outperformed HBLUP models or obtained similar results (Fig. 3). Consequently, in case of unstructured population, PP led to higher PA values than GP.

### Choosing genotypes based on their PEPV rather than on their GEBV provided higher selection differentials

The performance of GBLUP and HBLUP models in selecting the top 10% of genotypes from the whole panel was investigated (Table 4). The mean of the BLUEs values of the whole panel for each trait was compared to the mean BLUEs value of different sets of genotypes. The first set corresponds to the top 10% of the whole panel based on the BLUEs values. The second set corresponds to the top 10% of the whole panel according to their GEBV. And the third set corresponds to the top 10% of the whole panel according to their phenomic estimated phenotypic values (PEPV). For each method, the mean BLUEs value was calculated. The selection differential ( $S$ ) was estimated for each sampling method as the difference between the mean whole population BLUEs and the top 10% mean



**Table 4** Average of the best linear unbiased estimates (BLUEs) for the whole germplasm, and the top 10% individuals selected based on their BLUEs, genomic estimated breeding values (GEBV), and phenomic estimated phenotypic values (PEPV), for each trait, using 210

	Mean whole population	Top 10% BLUEs	S BLUEs	Top 10% GEBV	S GEBV	Top 10% PEPV	S PEPV
GP36	1.07 ± 0.36	1.55 ± 0.07	0.48	1.33 ± 0.21	0.26	1.33 ± 0.22	0.26
T50	51.85 ± 8.87	40.24 ± 1.77	-11.61	48.55 ± 4.83	-3.3	45.33 ± 5.06	-6.52
T20	41.23 ± 7.20	31.68 ± 1.30	-9.55	36.96 ± 4.57	-4.28	35.80 ± 4.51	-5.43
T80	63.48 ± 11.13	48.33 ± 1.97	-15.15	58.89 ± 6.56	-4.59	55.19 ± 5.60	-8.29
AUC	40.84 ± 11.79	60.14 ± 2.43	19.3	53.08 ± 5.50	12.24	53.55 ± 7.71	12.71
FG	30.16 ± 4.75	24.29 ± 0.85	-5.87	27.95 ± 3.27	-2.21	27.26 ± 2.39	-2.90
GP	2.53 ± 1.16	4.40 ± 0.33	1.87	2.81 ± 0.74	0.28	3.17 ± 0.84	0.64
MGT	52.22 ± 6.87	42.43 ± 1.62	-9.79	49.21 ± 4.97	-3.01	47.12 ± 4.03	-5.10
UNI	24.02 ± 7.50	14.49 ± 1.16	-9.53	20.83 ± 4.03	-3.19	20.03 ± 4.20	-3.99

BLUEs. The selection differentials were higher when genotypes were chosen on their PEPV than when they were chosen according their GEBV (average increase in S between S estimated for GEBV and S for PEPV = 29%), indicating that selection decisions based on PEPV would be closer to phenotypic selection than GP-based selection. Concordances between best (and worst, respectively) (Online Resource 8) selected genotypes according to their BLUEs and their GEBV or concordances between best (and, respectively, worst) genotypes selected according to their BLUEs and their PEPV were studied using Jaccard's similarity coefficient. This allowed to compare GP and PP ranking, for each scenario. Average concordances of 14.4% (BLUEs vs PEPV) and 7.2% (BLUEs vs GEBV) were obtained for the best 10% genotypes. Similar concordances were obtained for the 10% worst genotypes, with a concordance rate of 13.6% between BLUEs and PEPV, and 8.2% between BLUEs and GEBV.

## Discussion

In an attempt to get insights into the genetic architecture of seed germination in oilseed rape, we used a combination of genetic methods based on molecular as well as phenotypic and spectral predictors. The main results led to the highlight of 17 genomic regions that control seed germination-related traits in the large genetic diversity used. Moreover, genomic and phenomic prediction methods provided moderate-to-high predictive abilities, demonstrating the capacity to capture small additive and non-additive effects for seed germination. This study also provided the first application of phenomic prediction in oilseed rape and demonstrated the higher ability of phenomic prediction to estimate phenotypic values closer to BLUEs compared to genomic prediction.

rapeseed genotypes. For each average, the standard deviation is associated. The selection differential (S) was calculated for each values, as the difference between the germplasm mean and the mean of the top 10% genotypes

## Alleles promoting seed germination are almost all fixed in oilseed rape

A total of 17 QTL with small effects were detected confirming the polygenic nature of seed germination as illustrated in the previous studies (Bettey et al. 2000; Basnet et al. 2015; Hatzig et al. 2015; Nguyen et al. 2018). Six of these genomic regions controlled several traits (e.g., GP36, MGT, T20, and T50), which was consistent with the high correlations between these traits ( $r < -0.88$  or  $r > 0.96$ ). This suggests pleiotropy or linkage drag. Therefore, it would be interesting to study the underlying genes of these six specific regions. We found a QTL in common with Hatzig et al. (2015) through the identification of a gene orthologous to *A. thaliana* ATE1, also characterized by Holman et al. (2009) as involved in seed germination. In addition, we found C03p76020.1\_BnaDAR gene, highlighted by Ding et al. (2019) as being involved in seed germination in *A. thaliana* and *B. napus*. However, we found no other QTL in common between those identified in our study and those identified in the studies deciphering the genetic control of seed germination *sensus stricto* in *Brassica napus* (Hatzig et al. 2015; Nguyen et al. 2018; Boter et al. 2019; Gad et al. 2021; Luo et al. 2021), *Brassica oleracea* (Bettey et al. 2000), or *Brassica rapa* (Basnet et al. 2015) under optimal or stressed conditions. This lack of correspondence could be explained by the fact that these different studies (except Hatzig et al. (2015)) used restricted genetic diversity populations, limited either to WOSR or SOSR genotypes. Whereas we analyzed a larger genetic diversity.

To estimate the ability to improve seed germination in breeding by marker-assisted selection, we looked at the effect of QTL staking for each trait. The accumulation of several ( $> 3$ ) favorable alleles increased seed germination speed. However, for most of the traits used in this study, the favorable alleles were also the most frequent alleles in our germplasm. Consequently, a low genetic gain is expected. Similarly, Hatzig



et al. (2015) showed an improvement of T50 when stacking multiple favorable haplotypes. They also showed that the favorable haplotypes for this trait were the most frequent in the population, suggesting a strong selection for T50. Interestingly, some minor alleles were identified as favorable, for some traits involved in seed germination capacity, such as GP36. Therefore, improving seed germination capacity could still be achieved by targeting these favorable alleles presenting a low frequency in the population.

#### **SNP-based genomic structure of oilseed rape diversity did not match to the structure observed using spectral data, but influenced seed germination capacity**

No difference between WOSR and SOSR was observed using NIRS data. Furthermore, kinship estimated based on NIRS data ( $H$  matrix) was not correlated to the one based on SNP data ( $K$  matrix). Similar results were obtained for soybean RIL populations (Zhu et al. 2021) or for Dent and Flint maize populations (Weiβ et al. 2022). Brault et al. (2022) also reported a low correlation between NIRS-based  $H$  matrix and SNP-based  $K$  matrix using a diversity panel of grapevine. A single study for triticale revealed a  $H$  matrix well correlated to the genomic (Zhu et al. 2022). The low correlations between  $H$  and  $K$  matrices might result from different histories of domestication and selection or from different ranges of considered genetic diversity.

As shown in the previous studies of oilseed rape genetic diversity (Diers and Osborn 1994; Hasan et al. 2006; Bus et al. 2011), we observed two specific genomic clusters for winter and spring accessions using SNP data. Clustering on seed germination traits revealed a difference in performance between WOSR ( $C_1$  and  $C_4$ ) and SOSR ( $C_2$  and  $C_3$ ) genomic clusters. Nevertheless, within the panel, WOSR are represented in a similar proportion for each breeding period, unlike SOSR which were mainly bred between the 70's and 80's (Online Resource 1). These differences in breeding dates between WOSR and SOSR could partly explain the difference WOSR/SOSR observed for seed germination performance, thus confirming the hypothesis proposed when comparing  $H$  and  $K$  matrices. Indeed, Hatzig et al. (2018) showed that intensive selection conducted between the 70's and 80's to reduce erucic acid and glucosinolates content in seeds had a negative impact on seed germination capacity. Further investigation would be required to distangle the type effect (WOSR/SOSR) from the breeding history effect.

#### **Genomic and phenomic predictions allowed capturing weak polygenic effects**

After seeking to highlight moderate additive effects by GWAS, genomic and phenomic prediction models were

used to consider polygenic background. The combination of spectral and genomic data in the model called GHBLUP performed worse than HBLUP and worse or similar than GBLUP. Similarly, Brault et al. (2022), who identified weak correlations between  $K$  and  $H$  matrices, obtained no gain in PA by combining NIRS and SNP information. However, these results differed from the previous studies using GHBLUP model, which showed superior PA compared to GBLUP or HBLUP models on yield traits (Krause et al. 2019; Galán et al. 2020; Zhu et al. 2021; Robert et al. 2022). In the case of soybean, the rrBLUP model combining SNP and NIRS data showed a higher PA than the genomic rrBLUP, even if  $K$  and  $H$  matrices were not correlated (Zhu et al. 2021). The type of trait studied, its genetic architecture and, in particular, the proportion of phenotypic value explained by non-additive genetic effects, could explain these differences in GHBLUP performance. Therefore, this information could help to choose between genomic, phenomic, and combined predictions models. In particular, the use of NIRS data and PP models could be favored for traits moderately to strongly shaped by non-additive genetic effects. It is also necessary to remain cautious regarding the absorbance values for each wavelength. They can be influenced (i) by the environment in which the seeds were formed and stored, which can have an impact on the PA when predicting in independent environments, and (ii) by the environment in which the spectra were acquired (humidity and temperature), resulting in heterogeneous data over time. The decision to use some or all of the wavelengths, particularly with regard to their heritability, may also have an impact on the accuracy of phenomic predictions models. In addition, PA overestimation can result from the fact that the same samples are used both for the acquisition of NIRS data and for the acquisition of the phenotypic data. Dallinger et al. (2023) also warn against PA overestimation due to the correlation between traits to be predicted and seed compounds as well against unintentional selection of seed compounds in the selected population on the basis of its PEPV. Finally, it is necessary to remain vigilant regarding the traits chosen for prediction.

In this study, both GP and PP methodologies predicted well germination traits in the  $CV_{whole}$  scenario (up to 0.59 for GBLUP on GP36 and 0.50 for HBLUP on T20). However, we showed an impact of the genomic structure on the PA of GP but not with PP models (Fig. 3). This genomic structure impact on genomic prediction models was also reported previously (Riedelsheimer et al. 2013; Schopp et al. 2017; Werner et al. 2020). The absence of impact of genomic structure on PP is in accordance with Zhu et al. (2021, 2022) and Weiβ et al. (2022) studies. Therefore, HBLUP models should be preferred to predict complex traits for genetically diverse germplasm.



## Interest of phenomic prediction in breeding to incorporate genetic diversity in elite pools

Our results confirm that seed germination performance could be increased through genomic and/or phenomic prediction, highlighting the interest in improving the genetic background. We showed that PP provided PEPV closer to BLUEs than GEBV. It would be necessary to confirm these results using independent phenotyping trials. According to the breeder equation (Lush 1943), the impact of PP could be interpreted according to the different terms of the equation: concerning the (i) the generation interval, no impact is expected as seed germination phenotype is acquired at the beginning of the cycle. Considering (ii) the intensity  $i$ , NIRS acquisition is simple, fast, and inexpensive, which would allow to screen a higher number of genotypes. Therefore, more resources could be allocated to construct the training population, which then could lead to a better PA and, therefore, to a better identification of the best promising individuals. This resources reallocation could subsequently also contribute to phenotype these promising individuals more reliably and in different environments. As a whole, PP is a mean to increase the selection intensity. However, a specific attention has to be paid to maintain genetic diversity within the breeding population to assure long-term genetic gain. This can be achieved, (i) by limiting or optimizing the selection intensity, (ii) by considering genetic diversity in the choice of parents to cross, using, for example, optimal contribution selection (Cowling et al. 2017), or (iii) by integrating exotic material into elite genotypes (Simmonds 1993). For this last point, we propose to use PP to efficiently identify genetic diversity to be valorized. Integrating this genetic diversity into elite germplasm would increase the frequency of favorable alleles in the population. Indeed, integrating exotic diversity into elite germplasm is one of the most important challenges in field crops (Cowling 2013), especially in oilseed rape. Oilseed rape has a narrow genetic diversity and could take advantage of this gain of diversity to overcome biotic and abiotic stresses while maintaining or increasing yields. For this purpose, the creation of pre-bridging and bridging populations (i.e., progeny obtain from elite  $\times$  exotic crosses to be integrated into elite germplasm) have been proposed (Gorjanc et al. 2016; Allier et al. 2020; Sanchez et al. 2023). The construction of these (pre-)bridging populations could benefit from phenomic prediction. PP could help identifying material from a broad range of exotic resources for the creation and maintenance of pre-bridging and bridging populations.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00122-024-04659-0>.

**Acknowledgements** The authors would like to thank Renaud Rincent (QGE Le Moulon, INRAE) and Christopher Sauvage (Syngenta) for

many helpful discussions regarding genetic analyses, genomic and phenomic predictions; Mathieu Rousseau-Gueutin and Cyril Falentin (IGEPP, INRAE) for help with the genomic resources of *Brassica napus*. We are very grateful to Vanessa Clouet for the research of candidate genes. The authors would also like to thank the CRB BrACySol for providing genotypes and Elise Alix, Bernard Moulin (IGEPP, INRAE), and the “Domaine de la Motte” Experimental Unit (INRAE) for seed multiplication. A special thanks to Solenn Guichard for NIR spectra acquisition. The GenOuest bioinformatics platform (<https://www.genouest.org/>) is acknowledged for providing informatics resources.

**Author contribution statement** A.L. and N.N. designed the study and obtained funding. M-H.W., D.D., M.L. and C.B. and S.D. carried out the germination assays on the PHENOTIC platform at GEVES (Angers, France). C.B. and S.R. brought expertise with genotyping and NIRS data, respectively. M.L. and J.L. performed all the statistical data, developed the scripts, and ran the genetic analyses. M.L., J.L., A.L., and N.N. interpreted the results. M.L., A.L., and N.N. wrote the manuscript. All authors read and approved the final version.

**Funding** This work was supported by the NAVIG project funded by Promosol association. The PhD grant of M.L. was provided by INRAE and Promosol. The RAPSODY (ANR-11-BTBR-0004) project founded the acquisition of the genomic data.

**Data availability** The datasets generated and analyzed in this study are available using the following link during the review process (<https://entrepot.recherche.data.gouv.fr/privateurl.xhtml?token=4050df7f-5caf-4683-8142-df8b870ef8d5>) and will be freely available with a DOI if the article is accepted.

## Declarations

**Conflict of interest** The authors declare that they have no competing interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abraham S, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36:1627–1639. <https://doi.org/10.1021/ac60214a047>
- Albert E, Segura V, Gricourt J et al (2016) Association mapping reveals the genetic architecture of tomato response to water deficit: focus on major fruit quality traits. J Exp Bot 67:6413–6430. <https://doi.org/10.1093/jxb/erw411>
- Albrecht T, Wimmer V, Auinger H-J et al (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350. <https://doi.org/10.1007/s00122-011-1587-7>



- Allier A, Teyssèdre S, Lehermeier C et al (2020) Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genomics* 21:349. <https://doi.org/10.1186/s12864-020-6756-0>
- Araus JL, Kefauver SC, Zaman-Allah M et al (2018) Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci* 23:451–466. <https://doi.org/10.1016/j.tplants.2018.02.001>
- Astle W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. *Stat Sci* 24:451–471. <https://doi.org/10.1214/09-STS307>
- Basnet RK, Duwal A, Tiwari DN et al (2015) Quantitative trait locus analysis of seed germination and seedling vigor in *Brassica rapa* reveals QTL hotspots and epistatic interactions. *Front Plant Sci* 6:1032. <https://doi.org/10.3389/fpls.2015.01032>
- Bettey M, Finch-Savage WE, King GJ, Lynn JR (2000) Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in *Brassica oleracea*. *New Phytol* 148:277–286. <https://doi.org/10.1046/j.1469-8137.2000.00760.x>
- Boter M, Calleja-Cabrera J, Carrera-Castaño G et al (2019) An Integrative approach to analyze seed germination in *Brassica napus*. *Front Plant Sci* 10:1342. <https://doi.org/10.3389/fpls.2019.01342>
- Boureau T (2020) PHENOTIC Platform
- Brault C, Lazerges J, Doligez A et al (2022) Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods* 18:108. <https://doi.org/10.1186/s13007-022-00940-9>
- Browning BL, Zhou Y, Browning SR (2018) A One-penny imputed genome from next-generation reference panels. *Am J Human Genet* 103:338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Bus A, Körber N, Snowdon RJ, Stich B (2011) Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theor Appl Genet* 123:1413–1423. <https://doi.org/10.1007/s00122-011-1676-7>
- Clarke WE, Higgins EE, Prieske J et al (2016) A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor Appl Genet* 129:1887–1899. <https://doi.org/10.1007/s00122-016-2746-7>
- Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R Package sommer. *PLoS ONE* 11:e0156744. <https://doi.org/10.1371/journal.pone.0156744>
- Cowling WA (2013) Sustainable plant breeding. *Plant Breed* 132:1–9. <https://doi.org/10.1111/pbr.12026>
- Cowling WA, Li L, Siddique KHM et al (2017) Evolving gene banks: improving diverse populations of crop and exotic germplasm with optimal contribution selection. *J Exp Bot* 68:1927–1939. <https://doi.org/10.1093/jxb/erw406>
- Dallinger HG, Löschner F, Bistrich H et al (2023) Predictor bias in genomic and phenomic selection. *Theor Appl Genet* 136:235. <https://doi.org/10.1007/s00122-023-04479-8>
- Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Demilly D, Ducournau S, Wagner M-H (2014) Digital imaging of seed germination. In: Gupta SD, Ibaraki Y (eds) *Plant image analysis: fundamentals and applications*. CRC Press, Boca Raton, pp 147–162
- Diers BW, Osborn TC (1994) Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. *Theoret Appl Genetics* 88:662–668. <https://doi.org/10.1007/BF01253968>
- Ding L-N, Guo X-J, Li M et al (2019) Improving seed germination and oil contents by regulating the GDSL transcriptional level in *Brassica napus*. *Plant Cell Rep* 38:243–253. <https://doi.org/10.1007/s00299-018-2365-7>
- Ducournau S, Feutry A, Plainchault P et al (2004) An image acquisition system for automated monitoring of the germination rate of sunflower seeds. *Comput Electron Agric* 44:189–202. <https://doi.org/10.1016/j.compag.2004.04.005>
- Elliott RH, Mann LW, Johnson EN et al (2007) Vigor tests for evaluating establishment of canola under different growing conditions and tillage practices. *Seed Technol* 29:21–36
- FAO (2023) FAOSTAT Food and agriculture data
- Finch-Savage WE, Bassel GW (2016) Seed vigour and crop establishment: extending performance beyond adaptation. *J Exp Bot* 67:567–591. <https://doi.org/10.1093/jxb/erv490>
- Gad M, Chao H, Li H et al (2021) qtl mapping for seed germination response to drought stress in *Brassica napus*. *Front Plant Sci* 11:629970. <https://doi.org/10.3389/fpls.2020.629970>
- Galán RJ, Bernal-Vasquez A-M, Jebsen C et al (2020) Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. *Theor Appl Genet* 133:3001–3015. <https://doi.org/10.1007/s00122-020-03651-8>
- Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol* 32:361–369. <https://doi.org/10.1002/gepi.20310>
- Gao X, Becker LC, Becker DM et al (2010) Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet Epidemiol* 34:100–105. <https://doi.org/10.1002/gepi.20430>
- Gorjanc G, Jenko J, Hearne SJ, Hickey JM (2016) Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. *BMC Genomics* 17:30. <https://doi.org/10.1186/s12864-015-2345-z>
- Gower JC (1967) Multivariate analysis and multidimensional geometry. *J R Stat Soc Ser D (stat)* 17:13–28. <https://doi.org/10.2307/2987199>
- Haj Sghaier A, Tarnawa Á, Khaeim H et al (2022) The effects of temperature and water on the seed germination and seedling development of rapeseed (*Brassica napus* L.). *Plants* 11:2819. <https://doi.org/10.3390/plants11212819>
- Hasan M, Seyis F, Badani AG et al (2006) Analysis of genetic diversity in the *Brassica napus* L. gene pool using SSR markers. *Genet Resour Crop Evol* 53:793–802. <https://doi.org/10.1007/s10722-004-5541-2>
- Hatzig SV, Frisch M, Breuer F et al (2015) Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. *Front Plant Sci* 6:221. <https://doi.org/10.3389/fpls.2015.00221>
- Hatzig S, Breuer F, Nesi N et al (2018) Hidden effects of seed quality breeding on germination in oilseed rape (*Brassica napus* L.). *Front Plant Sci* 9:419
- Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4:65–75. <https://doi.org/10.3835/plantgenome2010.12.0029>
- Hickey JM, Chiurongwi T, Mackay I, Powell W (2017) Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat Genet* 49:1297–1303. <https://doi.org/10.1038/ng.3920>
- Holman TJ, Jones PD, Russell L et al (2009) The N-end rule pathway promotes seed germination and establishment through removal of ABA sensitivity in *Arabidopsis*. *Proc Natl Acad Sci* 106:4549–4554. <https://doi.org/10.1073/pnas.0810280106>
- Isik F, Bartholomé J, Farjat A et al (2016) Genomic selection in maritime pine. *Plant Sci* 242:108–119. <https://doi.org/10.1016/j.plantsci.2015.08.006>
- Kaufman L, Rousseeuw P (1990) Partitioning around medoids (program PAM). *Find Gr Data* 344:68–125. <https://doi.org/10.1002/9780470316801.ch2>
- Knoch D, Werner CR, Meyer RC et al (2021) Multi-omics-based prediction of hybrid performance in canola. *Theor Appl Genet* 134:1147–1165. <https://doi.org/10.1007/s00122-020-03759-x>



- Krause MR, González-Pérez L, Crossa J et al (2019) Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes Genomes Genetics*. 9:1231–1247. <https://doi.org/10.1534/g3.118.200856>
- Lamichhane JR, Debaeke P, Steinberg C et al (2018) Abiotic and biotic factors affecting crop seed germination and seedling emergence: a conceptual framework. *Plant Soil* 432:1–28. <https://doi.org/10.1007/s11104-018-3780-9>
- Lane HM, Murray SC, Montesinos-López OA et al (2020) Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *Plant Phenom J* 3:e20002. <https://doi.org/10.1002/ppj2.20002>
- Lê S, Josse J, Husson F (2008) FactoMineR: An R package for multivariate analysis. *J Stat Softw* 25:1–18. <https://doi.org/10.18637/jss.v025.i01>
- Lippert C, Listgarten J, Liu Y et al (2011) FaST linear mixed models for genome-wide association studies. *Nat Methods* 8:833–835. <https://doi.org/10.1038/nmeth.1681>
- Luo T, Zhang Y, Zhang C et al (2021) Genome-wide association mapping unravels the genetic control of seed vigor under low-temperature conditions in rapeseed (*Brassica napus* L.). *Plants* 10:426. <https://doi.org/10.3390/plants10030426>
- Lush JL (1943) Animal breeding plans. The Iowa State College Press, Ames
- Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10:565–577. <https://doi.org/10.1038/nrg2612>
- Maechler M, Rousseeuw P, Struyf A, et al (2022) Cluster: cluster analysis basics and extensions
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Muranty H, Troggio M, Sadok IB et al (2015) Accuracy and responses of genomic selection on key traits in apple breeding. *Hortic Res* 2:15060. <https://doi.org/10.1038/hortres.2015.60>
- Nelson MN, Nesi N, Barrero JM et al (2022) Strategies to improve field establishment of canola: a review. *Adv Agron* 175:133–177. <https://doi.org/10.1016/bs.agron.2022.05.001>
- Nguyen TCT, Abrams SR, Friedt W, Snowdon RJ (2018) Quantitative trait locus analysis of seed germination, seedling vigour and seedling-regulated hormones in *Brassica napus*. *Plant Breed* 137:388–401. <https://doi.org/10.1111/pbr.12576>
- Patti GJ, Yanes O, Siuzdak G (2012) Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 13:263–269. <https://doi.org/10.1038/nrm3314>
- Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
- Rajjou L, Duval M, Gallardo K et al (2012) Seed germination and vigor. *Annu Rev Plant Biol* 63:507–533. <https://doi.org/10.1146/annurev-arplant-042811-105550>
- Resende MDV, Resende MFR Jr, Sansaloni CP et al (2012) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol* 194:116–128. <https://doi.org/10.1111/j.1469-8137.2011.04038.x>
- Riedelsheimer C, Endelman JB, Stange M et al (2013) Genomic predictability of interconnected biparental maize populations. *Genetics* 194:493–503. <https://doi.org/10.1534/genetics.113.150227>
- Rincent R, Moreau L, Monod H et al (2014) Recovering power in association mapping panels with variable levels of linkage disequilibrium. *Genetics* 197:375–387. <https://doi.org/10.1534/genetics.113.159731>
- Rincent R, Charpentier J-P, Faivre-Rampant P et al (2018) Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar. *G3 Genes Genom Genet* 8:3961–3972. <https://doi.org/10.1534/g3.118.200760>
- Ritchie MD, Holzinger ER, Li R et al (2015) Methods of integrating data to uncover genotype–phenotype interactions. *Nat Rev Genet* 16:85–97. <https://doi.org/10.1038/nrg3868>
- Robert P, Auzanneau J, Goudemand E et al (2022) Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theor Appl Genet* 135:895–914. <https://doi.org/10.1007/s00122-021-04005-8>
- Rooney TE, Sweeney DW, Sorrells ME (2022) Time series barley germination is predictable and associated with known seed dormancy loci. *Crop Sci* 62:100–119. <https://doi.org/10.1002/csc.20638>
- Rousseau-Gueutin M, Belser C, Da Silva C et al (2020) Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience* 9:giaa137. <https://doi.org/10.1093/gigascience/giaa137>
- Sanchez D, Sadoun SB, Mary-Huard T et al (2023) Improving the use of plant genetic resources to sustain breeding programs' efficiency. *Proc Natl Acad Sci* 120:e2205780119. <https://doi.org/10.1073/pnas.2205780119>
- Schopp P, Müller D, Technow F, Melchinger AE (2017) Accuracy of genomic prediction in synthetic populations depending on the number of parents, relatedness, and ancestral linkage disequilibrium. *Genetics* 205:441–454. <https://doi.org/10.1534/genetics.116.193243>
- Schrag TA, Westhues M, Schipprack W et al (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. *Genetics* 208:1373–1385. <https://doi.org/10.1534/genetics.117.300374>
- Signal Developers (2014) Signal: signal processing. <http://r-forge.r-project.org/projects/signal/>
- Simmonds NW (1993) Introgression and incorporation. Strategies for the use of crop genetic resources. *Biol Rev* 68:539–562. <https://doi.org/10.1111/j.1469-185X.1993.tb01243.x>
- Storey JD (2002) A direct approach to false discovery rates. *J R Stat Soc: Ser B (stat Methodol)* 64:479–498. <https://doi.org/10.1111/1467-9868.00346>
- Voss-Fels KP, Cooper M, Hayes BJ (2019) Accelerating crop genetic gains with genomic selection. *Theor Appl Genet* 132:669–686. <https://doi.org/10.1007/s00122-018-3270-8>
- Wagner M-H, Demilly D, Ducournau S, Durr C (2011) Computer vision for monitoring seed germination from dry state to young seedlings. *Seed Test Int* 142:49–51
- Weiß TM, Zhu X, Leiser WL et al (2022) Unraveling the potential of phenomic selection within and among diverse breeding material of maize (*Zea mays* L.). *Genes Genom Genet* 12:jikab445. <https://doi.org/10.1093/g3journal/jkab445>
- Werner CR, Qian L, Voss-Fels KP et al (2018) Genome-wide regression models considering general and specific combining ability predict hybrid performance in oilseed rape with similar accuracy regardless of trait architecture. *Theor Appl Genet* 131:299–317. <https://doi.org/10.1007/s00122-017-3002-5>
- Werner CR, Gaynor RC, Gorjanc G et al (2020) How Population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. *Front Plant Sci* 11:592977. <https://doi.org/10.3389/fpls.2020.592977>
- Westhues M, Schrag TA, Heuer C et al (2017) Omics-based hybrid prediction in maize. *Theor Appl Genet* 130:1927–1939. <https://doi.org/10.1007/s00122-017-2934-0>



- Würschum T, Abel S, Zhao Y (2014) Potential of genomic selection in rapeseed (*Brassica napus* L.) breeding. Plant Breed 133:45–51. <https://doi.org/10.1111/pbr.12137>
- Yu J, Pressoir G, Briggs WH et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208. <https://doi.org/10.1038/ng1702>
- Zhu X, Leiser WL, Hahn V, Würschum T (2021) Phenomic selection is competitive with genomic selection for breeding of complex traits. Plant Phenom J 4:e20027. <https://doi.org/10.1002/ppj2.20027>
- Zhu X, Maurer HP, Jenz M et al (2022) The performance of phenomic selection depends on the genetic architecture of the target trait. Theor Appl Genet 135:653–665. <https://doi.org/10.1007/s00122-021-03997-7>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



---

## **CHAPITRE III**

# **ARCHITECTURE GENETIQUE ET PREDICTIONS DE LA VIGUEUR PRECOCE CHEZ LE COLZA**

---



## III.1. Préambule

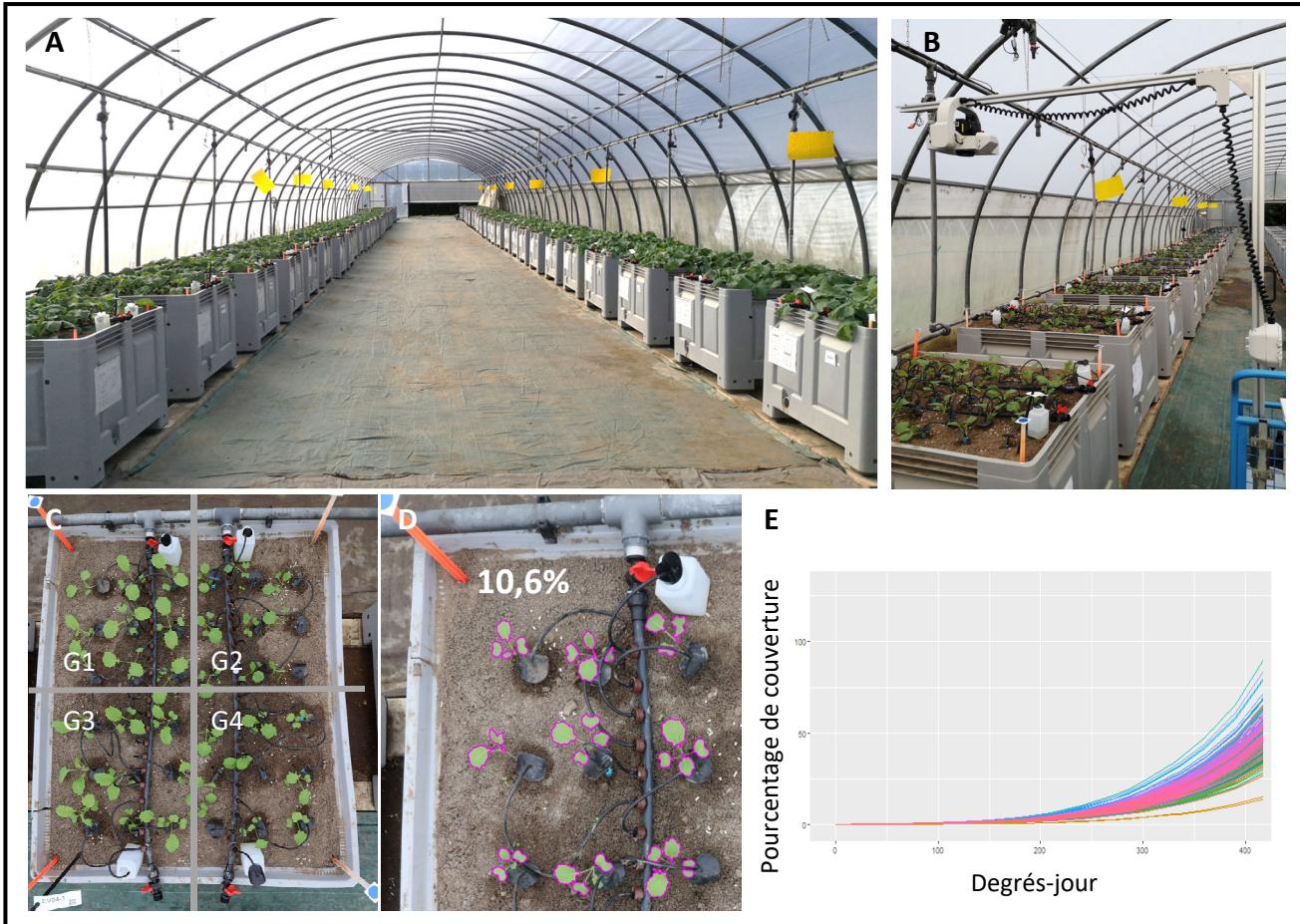
---

### Questions traitées dans le chapitre III

Cette partie comprend d'une part les résultats concernant le déterminisme génétique du développement précoce chez le colza qui sont présentés sous la forme d'un article en préparation dont je suis premier auteur (**III.2**), et d'autre part l'analyse de la capacité des modèles à prédire les traits de vigueur précoce qui pourra être valorisée sous forme de publication ultérieurement (**III.3**).

Le développement précoce est un processus complexe, intégrant un ensemble de traits développementaux, morphologiques et physiologiques (*cf.* **I.2.3 & I.2.4**). Afin de mieux comprendre le contrôle génétique du développement précoce, j'ai investigué les relations entre ces différents traits ainsi que leur organisation à l'échelle du génome du colza par analyse génétique (GWAS) qui ont également permis d'identifier les loci et les allèles favorables à valoriser en sélection (Laurençon et al. article en préparation ; *cf.* **III.2**).

Concernant les modèles de prédiction, l'hypothèse principale est que les spectres NIRS sont de bons prédicteurs des traits de vigueur précoce, en particulier lorsqu'ils sont comparés aux marqueurs moléculaires, car ils permettent de capturer l'effet de traits secondaires corrélés aux traits cibles à prédire et également l'effet de l'apparentement génétique. Une seconde hypothèse est que la prédiction sera d'autant meilleure que les spectres sont acquis sur des organes/tissus proches du phénotype à prédire. Ainsi, j'ai étudié deux facteurs pouvant impacter la qualité des modèles de prédiction phénomique pour les traits de vigueur précoce : la nature des spectres d'une part et les corrections apportées aux données spectrales d'autre part (*cf.* **III.3**). Ces deux facteurs ont été analysés à l'échelle d'un cluster génomique et à l'échelle du panel entier, afin de mettre en évidence d'éventuels biais liés à la structure du panel et/ou aux plans expérimentaux. A propos de la nature des spectres, j'ai comparé les capacités prédictives obtenues avec d'une part des spectres acquis sur graines entières et d'autre part des spectres acquis sur poudre de feuilles. S'agissant des corrections apportées aux données spectrales, l'objectif des différentes corrections testées est de maximiser la prise en compte de l'apparentement entre les individus et de minimiser l'effet environnemental sur la matrice d'apparentement spectral. Dans la littérature, plusieurs méthodes sont utilisées pour corriger les données spectrales. Certaines études utilisent la moyenne des spectres obtenus pour les répétitions d'un génotype, d'autres estiment les valeurs BLUE (best linear unbiased estimate) obtenues pour chaque génotype à l'aide d'un modèle linéaire prenant en compte les différents effets environnementaux (Robert et al. 2022a; Meyenberg et al. 2024), et d'autres utilisent les valeurs BLUP (best linear unbiased prediction) obtenues pour chaque génotype en utilisant une matrice



**Figure III.1.1. Dispositif expérimental utilisé pour phénotypier la vigueur précoce.**

A. Vue générale du tunnel permettant des conditions semi-contrôlées (en particulier, le pilotage de l’irrigation). B. Perche pour la prise de photos. C. Vue à la verticale d’un caisson. Chaque caisson de 1m<sup>3</sup> contient 4 génotypes (G1-G4) et 9 plantes par génotype. D. Analyse d’image pour estimer la surface foliaire (macros développées dans par S. Jumel, IGEPP). Le taux de couverture de cette image est indiqué (ici 10,6%). E. Surface foliaire par plante au cours du temps, pour chaque génotype.



**Figure III.1.2. Plan expérimental des essais semés en Mars, Août et Septembre.**

Pour chaque essai, deux blocs sont concernés. Les rectangles gris correspondent aux caissons. Les rectangles verts correspondent aux caissons contenant les génotypes témoins (2 génotypes par caisson). Les génotypes de printemps (SOSR) et rutabagas ont été répétés 2 fois (correspondant aux blocs 1 et 2) durant l’essai de Mars. Les accessions d’hiver oléagineux (WOSR) et fourrager ont été phénotypés durant l’essai d’Août et l’essai de Septembre. La position des génotypes est aléatoire pour chaque bloc.

d'apparentement génomique pour définir la matrice de variance-covariance de l'effet génétique (Brault et al. 2022; DeSalvio et al. 2024). Ainsi, j'ai étudié l'effet des 3 types de corrections sur les données spectrales (moyenne arithmétique, BLUE, BLUP) sur les capacités prédictives des modèles de prédiction phénomique dans le cas des traits de vigueur précoce.

## Acquisition des données

Le phénotypage de la vigueur précoce a été réalisé à l'IGEPP au Rhei dans un tunnel couvert permettant de travailler en conditions semi-contrôlées (**Figure III.1.1.A**). Les graines imbibées pendant 24 h à 21°C à l'obscurité sur papier filtre sont transférées en terre dans des caissons de 1 m<sup>3</sup> et les plantes sont irriguées individuellement par goutte à goutte pendant l'expérimentation avec une solution nutritive dispensée de manière à ne pas avoir de déficit hydrique (pilotage du potentiel hydrique du sol via des sondes).

Trois expérimentations indépendantes ont été réalisées au cours de l'année 2022. Les 85 génotypes de colza de printemps et les 2 génotypes de rutabagas ont été phénotypés durant le printemps 2022 en une seule expérimentation que l'on appellera par la suite l'expérimentation Mars ('March trial') par référence à la date de semis (15/03/2022). Afin de prendre en compte la différence de rayonnement entre les deux côtés du tunnel (Est vs. Ouest), les 2 répétitions de cette expérimentation 'March trial' ont été disposées de chaque côté du tunnel, les génotypes étant randomisés dans chaque répétition. Les 133 génotypes de colza d'hiver et les 13 fourragers d'hiver ont été phénotypés durant l'automne 2022 en deux expérimentations : la première en Août 2022 ('August trial', semis le 18/08/2022) et la seconde en Septembre 2022 ('September trial', semis le 28/09/2022). Dans ces deux essais, les génotypes étant également disposés de façon randomisée dans les répétitions, à l'exception que ceux qui se trouvaient du côté ouest en Août ont été implantés à l'est en Septembre et vice versa. Le dispositif correspond à un dispositif augmenté composé de 6 blocs (chaque bloc étant représenté par un côté du tunnel lors d'une répétition) (**Figure III.1.2**). Afin de pouvoir comparer les expérimentations entre elles, les deux génotypes Aviso (type WOSR) et Miro (type SOSR) ont été utilisés comme témoins et implantés dans tous les blocs.

Chaque caisson contenait 4 génotypes, avec 9 plantes par génotype (**Figure III.1.1.C**). Dès la levée, un suivi du couvert végétal a été mis en place grâce à une prise de photographie régulière, les lundis, mercredis et vendredis du semis à la récolte. Le dispositif pour la prise de photo a été développé par Stéphane Jumel (IGEPP, équipe DEMECOL) avant mon arrivée au laboratoire et repose sur une perche roulante avec une tête robotisée permettant de fixer un appareil photo à la verticale (**Figure III.1.1.B**). Les images ont ensuite été traitées pour accéder de façon semi-automatique à la dynamique de la



mise en place du couvert végétal (**Figures III.1.1.D & E**). Ainsi, les images ont permis d'acquérir des données de développement des plantes en cinétique.

Les plantes ont été récoltées au stade 5-6 feuilles. Lors de la récolte, les biomasses foliaire et racinaire sèches ainsi que la surface foliaire ont été mesurées pour chaque plante. Les parties aériennes ont été broyées. La prédiction de la teneur en azote a été réalisée pour chaque plante. Pour cela, nous avons utilisé un modèle de régression basé sur les NIRS, qui a été développé dans l'équipe (Rolland et al. soumis).

Les spectres NIRS acquis sur les poudres de feuilles, ainsi que les spectres acquis sur graines (*cf. II.2*) ont été utilisés en tant que régresseurs afin d'évaluer la capacité prédictive des modèles de prédiction phénomique.



## III.2. Favorable alleles for early vigor related traits show specific occurrence regarding rapeseed germplasms

---

Marianne Laurençon<sup>1</sup>, Elise Alix<sup>1</sup>, Solenn Guichard<sup>1</sup>, Stéphane Jumel<sup>1</sup>, Bernard Moulin<sup>1</sup>, Marion Gauthier<sup>1</sup>, Céline Richard-Molard<sup>2</sup>, Nathalie Nesi<sup>1</sup>, Anne Laperche<sup>1\*</sup>

<sup>1</sup> Institute of Genetics, Environment and Plant Protection (IGEPP), INRAE – Institut Agro Rennes-Angers – Université de Rennes, 35650 Le Rheu, France

<sup>2</sup> Functional ecology and ecotoxicology of agroecosystems (Ecosys), INRAE – AgroParisTech – Université de Paris Saclay, 91120 Palaiseau, France

\* Correspondence: [anne.laperche@institut-agro.fr](mailto:anne.laperche@institut-agro.fr)

Manuscript in preparation for submission to BMC Genetics.



## Background

Plant vigor is a key agronomic trait that determines plant performance. It can be defined as the sum of properties that enables plants to transit rapidly through sensitive phases, eventually providing a means to escape from adverse conditions and various stresses throughout the crop cycle (Elliott et al. 2008; Peter et al. 2009; Namuco et al. 2009; Hendriks et al. 2024). Vigor during the early stages of development is particularly critical to ensure correct crop density that helps guarantee yield. In addition, vigor depends on the interplay between plant genotype, environment and crop management practices (Finch-Savage and Bassel 2016). Thus, delivering vigorous plant varieties has been a main goal of plant breeding for years, and is also a lever for the transition of agriculture in the context of reducing chemical inputs.

As vigor is a complex plant trait, it is often approached through visual ratings of plant size and biomass, using scores from 1 (high vigor) to 9 (low vigor) in the field. In addition, biomass traits and leaf area are commonly used to characterize early development in plant species such as *Arabidopsis thaliana* (Bac-Molenaar et al. 2015), rice (Rebolledo et al. 2015) or rapeseed (Nguyen et al. 2018; Knoch et al. 2020). However, plant vigor is the result of a combination of criteria that encompass many more processes. A range of morphological, developmental or physiological traits can be used to unravel plant performance (Violle et al. 2007), several of which have already been identified as influencing early vigor in different crops. For instance, the rate of leaf emergence is reported to be a proxy of vigor in rice (Rebolledo et al. 2012). In addition, the plant abilities to use resources for growth processes are also determinant for vigor; this includes the so-called ‘efficiency traits’. For example, the specific leaf area (SLA), defined as the ratio of leaf area to leaf dry mass, was described as a vigor parameter for rice and wheat since it qualifies the plant ability to compete with weeds (Dingkuhn et al. 1998; Rebetzke et al. 2004). In contrast, the most vigorous genotypes of cotton exhibited a preferential allocation of biomass to the roots (Liu et al. 2019), underlying the interest of traits related to carbon allocation between shoot and root in relation to vigor. Related traits include the shoot:root ratio (SR), the leaf weight ratio (LWR, estimating the allocation of total biomass towards the aerial parts) and the leaf area ratio (LAR, representing the balance between leaf development and biomass) (Hunt 2003). Plant growth also depends on its capacity to absorb and utilize nutrients. Therefore, vigor can also be qualified by the plant capacity to assimilate carbon (specific carbon assimilation, SCA) (Liao et al. 2004; Pang et al. 2013; Liu et al. 2019) or absorb nitrogen (specific nitrogen uptake, SNU) and use it to elaborate leaf area (leaf area:shoot nitrogen quantity ratio, LASNQ) or biomass (nitrogen utilization efficiency, NUTE). LASNQ, SCA and SNU correspond to the efficiencies defined in the ARNICA



conceptual plant functioning model developed on *Arabidopsis thaliana* (Richard-Molard et al. 2009). This model is used to represent carbon and nitrogen fluxes in the aerial and root compartments.

Rapeseed (*Brassica napus* L.) is the second oilseed crop worldwide with a total seed production of 80 Mt. It can be grown as a winter crop with a strong vernalisation requirement in Western Europe, China and USA, whereas in Eastern Europe, Canada and Australia mainly spring types with low vernalisation requirement are grown over mild winter and spring seasons. Episodic drought events over sowing periods highly compromise early crop settlement. In addition, small plants and patchy crops are more vulnerable to pest pressure (insects, weeds), while at the same time pest control options are declining. As a matter of fact, early vigor, more specifically from the emergence to the 4-5 leaf stage (when seedlings exhibit slow growth) has been identified as key traits for rapeseed crop (reviewed in Nelson et al. 2022). Previous studies reported that early vigor is strongly environmentally dependent, but could nevertheless be substantially improved by genomic-based breeding (Finch-Savage and Bassel 2016).

As many plant species, rapeseed resulted from several genome duplication and polyploidization events followed by speciation steps. Within the *Brassica* genus, *Brassica napus* L. (AACC,  $2n=4x=38$ ) originated from recent (~7500 years) and few crosses between *Brassica rapa* (AA,  $2n=2x=20$ ) and *Brassica oleracea* (CC,  $2n=2x=18$ ), thus limiting the integration of genetic diversity from progenitor species (Snowdon et al. 2007; Chalhoub et al. 2014; Lu et al. 2019). The cultivation and diversification of *Brassica napus* L. has led to different subspecies known as rapeseed, rutabaga or kale (Allender and King 2010). Subsequently, selection led to the emergence of three distinct types regarding their phenology (winter, spring and Asian rapeseed types) (Diers and Osborn 1994). In addition, rapeseed was also subjected to intense selection in the 1980s with the objective of reducing seed erucic acid and glucosinolate contents while improving yield and oil quality that significantly reduced the genetic diversity present in the germplasms (Cowling 2007; Fu and Gugel 2010; Bus et al. 2011; Lu et al. 2019). As a consequence, the low genetic diversity observed in rapeseed indicates a limited capacity to adapt to stress and limitations (Rahman 2013). In order to compensate for the lack of diversity within a dedicated germplasm, it is necessary to consider the use of accessions from different types and additional *Brassica napus* subspecies, as well as material from inter-specific crosses for prebreeding (Rahman 2013) that might be reservoirs of favorable alleles for optimizing early vigor.

The genetic control of early development (and therefore early vigor) has been poorly studied in oilseed rape, particularly when the whole plant is considered, and using a wide range of genetic diversity. In the present study, we investigated the genetic determinism of early vigor in rapeseed using a large gene pool, including winter, spring and Asian accessions. Thus, studying the genetic control of a wide range of physiological traits involved in early development is a way of gaining a better understanding of early vigor, and providing access to favorable alleles. The phenotypic variability of



early vigor-related traits was addressed and our results highlighted the genomic regions controlling these traits as well as the genetic diversity available for breeding purposes.

## Material and Methods

### Plant material and growth conditions

The diversity panel consisted of 233 *B. napus* inbred lines that were registered between 1931 and 2010, including 133 winter oilseed rape (WOSR), 85 spring oilseed rape (SOSR), 13 winter fodder (WFR) and 2 swede lines (**Supplemental Table III.2.1**). In particular, 89 accessions were modern double-low lines ('00' quality profile, low in seed erucic acid and glucosinolate contents), 87 were old varieties ('++'), 31 were simple low types ('0+') and only two were deprived of glucosinolates although rich in erucic acid ('+0'). Seed lots were all produced by open-pollination during the 2020/2021 growing season in Le Rheu, France (Laurençon et al. 2024).

Plants were grown under a semi-controlled environment in 1 m<sup>3</sup> containers (Bianchetti et al. 2024). The complete design consisted in an augmented design composed of 6 blocks of 24 containers. Each container consisted in 4 genotypes arranged in 4 squares with 9 plants per genotype. The blocks consisted of 90 genotypes randomly disposed and two checks, Aviso and Miro, that were repeated three times per block to ensure the comparison of the data. The spring genotypes were sown in two experimental blocks in March 2022, with each genotype represented in both blocks ('March trial'). The winter genotypes were sown in August and September 2022, in two blocks for each sowing ('August' and 'September' trials respectively). The containers were filled with a 50/50 mixture of topsoil and sand (Premier Tech GHA, France). To minimize the impact of seed germination ability on early growth capacity, seeds were pre-germinated in Petri dishes on moist filter paper at 21°C during 24 h before being transplanted into the containers. Each single plant was irrigated with a fertilizer solution every two days (Liquoplant FD134 winter, Plantin, France) and the substrate's water potential was monitored daily using a ceramic tensiometer (Tensionic®, SDEC, France) placed adjacent to the checks. The water potential was maintained at a level above -100 mbar to avoid drought stress. Air temperature was recorded at 2 locations for each block at 30-minute intervals throughout the experiment using recorders (Watchdog, SDEC, France). These values were used to estimate the cumulative growing degree days (GDD, base 0°C), thereby enabling the assessment of plant development stages. Daily photosynthetically active radiation (PAR) values were obtained from the weather station located on the experiment site (Le Rheu location) using the INRAe CLIMATIK portal (<https://agroclim.inrae.fr/climatik/ClimatikGwt.html>). Each block was harvested when 50% of all

**Table III.2.1. Traits under study**

Category	Trait	Variable	Formula	Unit
<b>Developmental trait – growth dynamics</b>				
	Area under the curve bw Leaves La and Lb	<b>AUC<sub>La-Lb</sub></b>	$\int_{La}^{Lb} f(x)dx$	
	Speed of dev. bw. leaves La and Lb	<b>Sp<sub>La-Lb</sub></b>	$\frac{x_{Lb} - x_{La}}{GDD_{max} - GDD_{min}}$	
<b>Physiological traits</b>				
<b>Biomass and surface</b>	Aerial Dry Weight	<b>ADW</b>	Measured (weight)	g/plant
	Root Dry Weight	<b>RDW</b>	Measured (weight)	g/plant
	Total Dry Weight	<b>TDW</b>	ADW+RDW	g/plant
	Leaf Area	<b>LA</b>	Measured (planimeter)	cm <sup>2</sup>
	Specific Leaf Area	<b>SLA</b>	LA/ADW	cm <sup>2</sup> .g <sup>-1</sup>
	Leaf Area Ratio	<b>LAR</b>	LA/TDW	cm <sup>2</sup> .g <sup>-1</sup>
	Shoot Nitrogen Content	<b>%SN*</b>	NIRS calibration	g/100 g
	Shoot Nitrogen Quantity	<b>SNQ</b>	(%SN×ADW)/100	g
	Root Nitrogen Content	<b>%RN*</b>	NIRS calibration – check accessions	g/100 g
<b>Efficiencies</b>	Root Nitrogen Quantity	<b>RNQ*</b>	(%RN×RDW)/100	g
	Total Nitrogen Quantity	<b>TNQ</b>	SNQ + RNQ	g
	Shoot:Root Ratio	<b>SR</b>	ADW/RDW	
	Leaf Weight Ratio	<b>LWR</b>	ADW/TDW	
	Leaf Area:Shoot Nitrogen Quantity Ratio	<b>LASNQ</b>	LA/SNQ	cm <sup>2</sup> .g <sup>-1</sup>
	Specific Carbon Assimilation	<b>SCA</b>	TDW/AUC	
	Specific Nitrogen Uptake	<b>SNU</b>	TNQ/RDW	
	Total Nitrogen Utilization Efficiency	<b>NUtE<sub>tot</sub></b>	TDW/TNQ	
	Aerial Nitrogen utilization efficiency	<b>NUtE<sub>aer</sub></b>	ADW/TNQ	

\*, traits that were not considered in the analyses

plants had reached the 4-5 leaf stage. Descriptions of the experiments are given in **Supplemental Table III.2.2.**

### Phenotyping methods

All the 23 traits under study are reported in **Table III.2.1** and the corresponding raw data are provided in **Supplemental Table III.2.2**. The traits consisted of developmental traits acquired in dynamics throughout the experiments and physiological traits acquired at harvest. The first ones qualified growth dynamics while the second ones referred to biomass and surface elaboration as well as efficiencies.

#### *Developmental traits acquired in dynamics*

Pictures of the plants were taken three times a week from the plant emergence to harvest to estimate leaf surface dynamics. In order to limit any bias in the image analysis process, the camera (Canon EOS 250D digital camera with a Canon 10-18mm lens; Canon Inc., Japan) was equipped with an automatic nadir stabilization system (INANIX, France) and automatically controlled using the digiCamControl v. 2.1.1.0 opensource software. Raw pictures were saved as RGB images with a definition of  $6000 \times 4000$  pixel and processed as follows. First, parts of the containers were removed from images by detecting blue markers located at the corners through thresholding from the RGB image. Then, the image was divided into four equal parts corresponding to the four genotypes present in each container. Finally, the plant surface was segmented through thresholding from the RGB image. The leaf coverage percentage was extracted by dividing the mask area by the total image area. A soil surface area per plant was estimated, enabling a leaf area per plant to be calculated over time and reducing the effect of missing plants per block. At the end, an average leaf coverage curve per genotype and per container was obtained throughout time expressed by cumulative GDD. Consolidated data per genotype consisted of the projected leaf area acquired only from the cotyledonary stage to the 4-leaf appearance to avoid confusion due to leaf overlapping from the neighboring plant genotypes. In addition, data were produced for genotypes with at least 6 plants per replicate to ensure reproducibility. All the images and the macro ImageJ are available upon request and will be dowloaded in an open dataverse (under progress).

An exponential was fitted to the data of leaf area development dynamics to estimate the curves per genotype. Areas under the curves (AUC) were assessed using R (DescTools package (Signorell et al. 2023)). Leaf number was counted every day to deduce the phyllochron of each plant (intervening period between the sequential emergence of leaves). Then, AUC values were deduced according to the phenology as follows:  $AUC_{La-Lb}$  was AUC between emergence of leaf La and leaf Lb (variables under



study were **AUC<sub>L1-L4</sub>**, **AUC<sub>L1-L2</sub>**, **AUC<sub>L2-L3</sub>**, **AUC<sub>L3-L4</sub>**). AUC values were used to estimate the speed of development between different leaves. For this, the ratio of the difference in AUC between Lb and La leaves and the difference in GDD between Lb and La leaves was calculated. **Sp<sub>La-Lb</sub>** was the speed of development between leaf La and leaf Lb (variables under study were **Sp<sub>L1-L4</sub>**, **Sp<sub>L1-L2</sub>**, **Sp<sub>L2-L3</sub>**, **Sp<sub>L3-L4</sub>**).

#### *Physiological traits acquired at plant harvest and inferred efficiencies*

For each plant, leaf area (**LA**, cm<sup>2</sup>) was measured using a planimeter (LI-3100C, LI-COR, USA). The aerial and root parts of each plant were dried separately at 70°C during 48 h and weighted afterwards to access aerial and root dry weight (**ADW** and **RDW** respectively, both in g/plant). Three physiological traits were deduced from the values of plant biomass and leaf area, including total dry weight (**TDW**, g/plant) as the sum of ADW and RDW; the specific leaf area (**SLA**, cm<sup>2</sup>.g<sup>-1</sup>) and the leaf area ratio (**LAR**, cm<sup>2</sup>.g<sup>-1</sup>) as ratios of LA to ADW and TDW respectively.

Shoot nitrogen content (**%SN**, g/100 g) was predicted using NIRS-based calibration models (Rolland et al. submitted). Dry leaves were ground to a fine powder and used for collecting NIR spectra (MPA II FT-NIR spectrophotometer, Bruker Optic Inc, Germany) over the range of 800 to 2781 nm with a step of 1 nm. Three biological replicates were run per genotype, with each replicate being the average of 64 technical repetitions measured by the spectrophotometer. Shoot nitrogen quantity (**SNQ**, g) was calculated as the product of %SN by ADW. As root nitrogen content (**%RN**, g/100 g) shows very little variation across a wide range of root samples (Rolland et al. submitted), the %RN value obtained on the roots of control accessions by NIRS calibration was assigned to the entire population. Root nitrogen quantity (**RNQ**, g) was calculated as the product of %RN by RDW and total nitrogen quantity (**TNQ**, g) resulted from the sum of SNQ and RNQ.

In addition, seven additional physiological traits corresponding to efficiencies were calculated, including the shoot:root ratio (**SR**) as the ratio of ADW to RDW; the leaf weight ratio (**LWR**) as the ratio of ADW to TDW; the leaf area:shoot nitrogen quantity ratio (**LASNQ**) as the ratio of LA to SNQ; the specific carbon assimilation (**SCA**) as the ratio of TDW to the **AUC<sub>L1-L4</sub>**; the specific nitrogen uptake (**SNU**) as the ratio of TNQ to RDW; the total and aerial **NUtE<sub>tot</sub>** were estimated as the ratio of TDW (or ADW for aerial **NUtE<sub>aer</sub>**) to TNQ.

#### **Population genetic structure and genetic diversity**

All the accessions were genotyped using the Brassica 60K Infinium® SNP array (Illumina, Inc. San Diego, CA) (Clarke et al. 2016) and the data were visualized using Genome Studio software (Illumina). A total of 33,474 SNP was validated using thresholds of 5% for the minor allele frequency (MAF), 10% for the frequency of missing values and 10% for the heterozygosity level. The missing SNP data were imputed



using BEAGLE software following the method described in Browning et al. (2018). SNP were physically anchored to the *B. napus* Darmor-*bzh* v10 reference genome (Rousseau-Gueutin et al. 2020).

Genetic structure of the germplasm was unraveled using the software ADMIXTURE v1.3 (Alexander et al. 2009) to identify the number of genomic clusters (K number of genomic clusters, varying between 2 and 10). A K-fold cross-validation was used to identify the most likely number of genomic clusters. The first three components of a principal coordinates analysis (PCoA; Gower, 1967) were used to perform a diversity analysis using the SNP data. The PCoA is based on the genetic distances between genotypes estimated by the dissimilarity matrix (1-K). The genomic kinship (K) was estimated using Astle and Balding (2009) algorithm as following :

$$K = \frac{WW^T}{M}$$

With W the matrix scaled on allelic frequencies with dimensions N×M, N is the number of genotypes and M the number of molecular markers and  $W^T$  is the transposed W matrix.

PLINK software v.1.9 (Purcell et al. 2007) was used to estimate the mean  $F_{ST}$  between genetic clusters and linkage disequilibrium (LD) decay according to the physical distance between markers for each chromosome and for each genetic cluster.

### **Estimation of phenotypic values and genomic heritability**

As each genotype was present in 2 blocks only, we used a mixed linear model to estimate phenotypic values. The checks present in each block were declared as fixed to consider the difference between blocks. The following mixed linear model (Federer 1956) was used to calculate best linear unbiased estimators (BLUE) for each genotype.

$$Y_{ij} = \mu + G_{i(j)} + C_k + b_j + \varepsilon_{ij},$$

where  $Y_{ij}$  is the phenotypic value of the  $i^{th}$  genotype in the  $j^{th}$  block,  $\mu$  is the overall mean,  $G_{i(j)}$  is the fixed effect of the  $i^{th}$  genotype in the  $j^{th}$  block,  $C_k$  is the fixed effect of the  $k^{th}$  check genotype,  $b_j$  is the random effect of the  $j^{th}$  block, following a normal distribution  $b \sim \mathcal{N}(0, \sigma_b^2)$ , and  $\varepsilon_{ij}$  the overall residual.

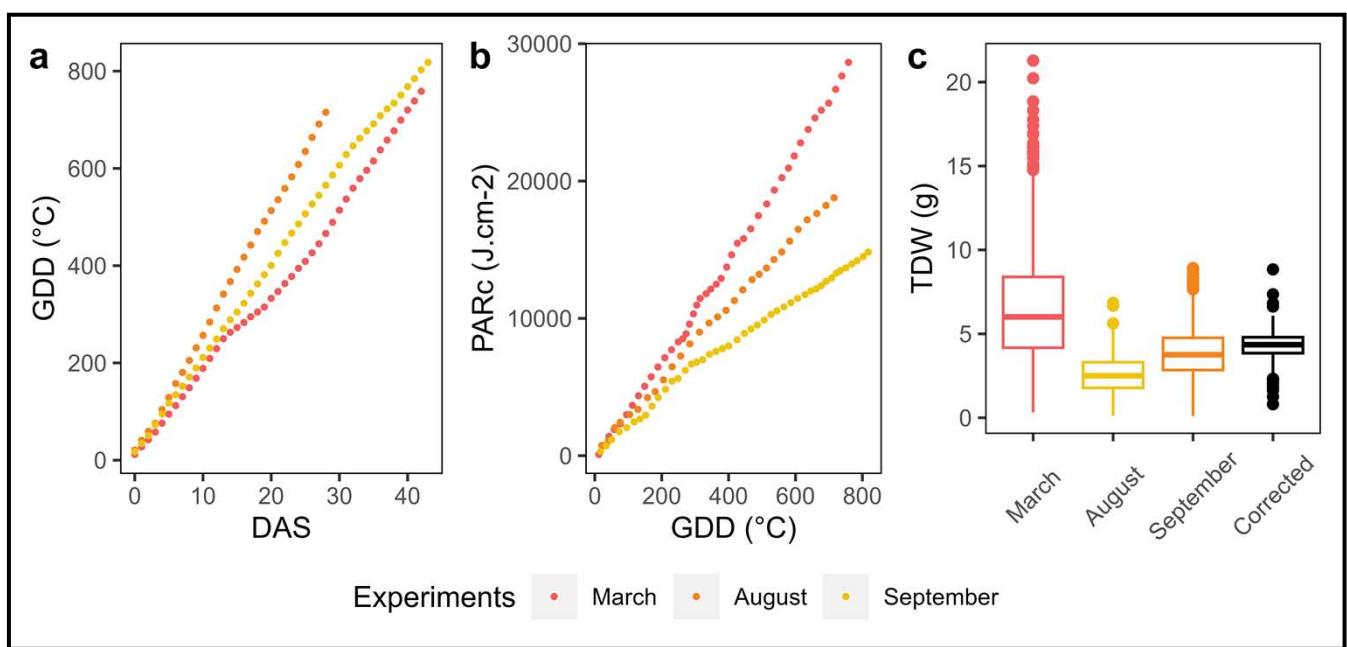
Genomic ( $\sigma_G^2$ ) and error ( $\sigma_e^2$ ) variances were estimated using the following linear mixed model:

$$\hat{Y}_i = \mu + g_i + \varepsilon_i,$$

Where  $\hat{Y}_i$  is the adjusted phenotypic value for the  $i^{th}$  genotype,  $\mu$  is the overall mean,  $g_i$  is the random effect of the  $i^{th}$  genotype following a normal distribution  $g \sim \mathcal{N}(0, K\sigma_G^2)$ , K is the genomic kinship and  $\varepsilon_i$  the overall residual.

From this equation, narrow-sense heritability ( $h^2$ ) was estimated for each phenotypic trait as follows:

$$h^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \hat{\sigma}_e^2},$$



**Figure III.2.1. Comparison of the environmental conditions between the three experiments.**

(a) Thermal time (GDD, growing degree-days) over the experiment (DAS, days after sowing). (b) Cumulative photosynthetic active radiation (PARc) over thermal time. (c) Total dry biomass (TDW) values per each experiment and for the whole population after linear mixed model correction (BLUE values).

where  $\hat{\sigma}_G^2$  and  $\hat{\sigma}_e^2$  are the REML estimations of the genomic and the error variance, calculated using the r package sommer (Covarrubias-Pazaran 2016).

### Trait correlation analyses

Principal component analysis (PCA) of all phenotypic traits was performed using FactoMineR package (Lê et al. 2008) using the BLUE values obtained for each trait and each genotype. The number of dimensions was chosen according to the percentage of explained variance.

For each trait, ANOVA and Tukey tests were performed to compare the phenotypic variability between the genomic clusters defined according the PCoA.

### Genome wide association study

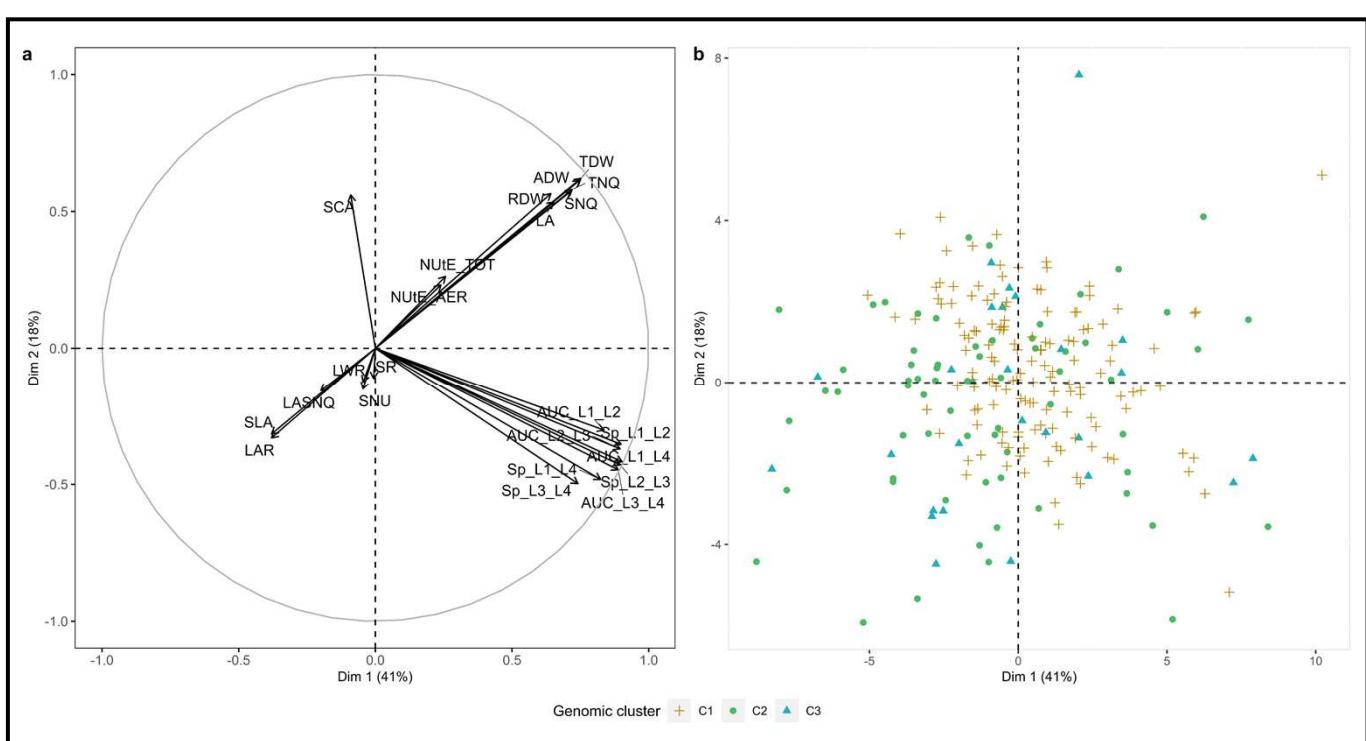
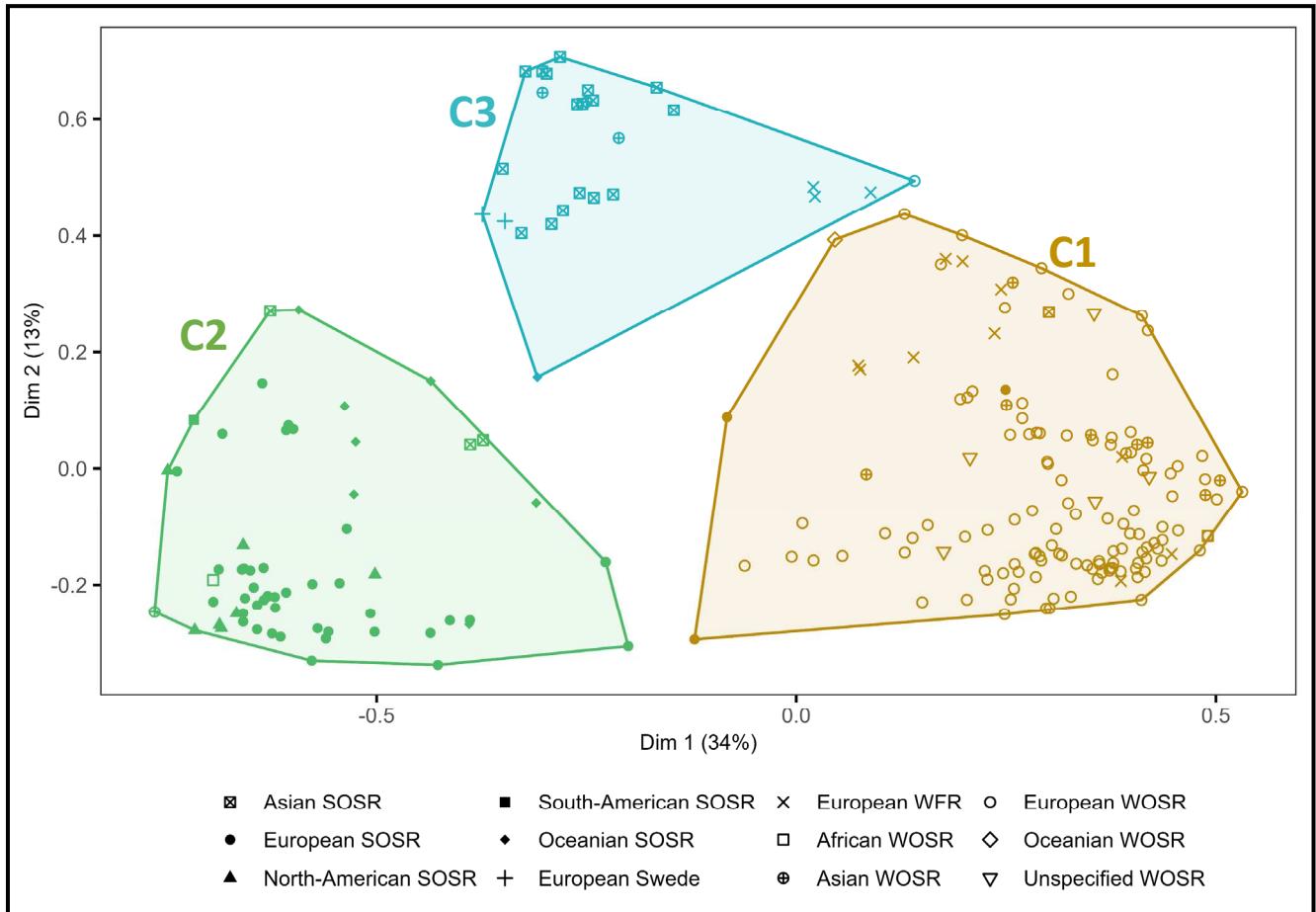
GWAS was performed using FaST-LMM algorithm (Lippert et al. 2011), taking into account population structure and kinship between genotypes to reduce false associations (Yu et al. 2006). A genomic kinship matrix was re-estimated for each chromosome tested as described by Rincent et al. (2014). Markers were considered as significantly associated with a trait if the  $-\log_{10}(p\text{-value})$  exceeded a 5% threshold of the SimpleM approach (Gao et al. 2008, 2010, 5% threshold was equal to 3.49). QTL confidence intervals were estimated as described previously (Laurençon et al. 2024): briefly, confidence intervals were estimated for each significant marker as the interval between the first (upstream) and last (downstream) marker that presented a LD value higher than the LD threshold. LD calculation was performed between 100,000 randomly chosen pairs of unlinked loci located on different chromosomes using PLINK software v.1.9 (Purcell et al. 2007). The critical LD threshold was chosen as the 95<sup>th</sup> percentile of the LD distribution, which equaled 0.174 using our data.

A  $F_{ST}$  genome scan was also carried out. Therefore, Weir and Cockerham's  $F_{ST}$  metric (1984) was estimated per site and along genome using VCFtools (Danecek et al. 2011). The top 5% of  $F_{ST}$  values were considered as significant.

## Results

### Our experimental design was adequate to study early vigor within a large diversity panel of rapeseed

As we assessed early vigor using data from three independent experiments, we used the two check genotypes to compare the datasets with each other on the basis of cumulative GDD and PARc. We showed that cumulative GDD to reach the 4-5 leaf stage slightly differed between the three trials (758



**Figure III.2.2. Genomic relatedness of the 233 rapeseed genotypes.**

The genetic structure of the germplasm was established by using the first two components of a principal coordinates analysis based on the SNP data. Clustering method depicts three different clusters (C1-C3). Accessions are depicted according to their own type.

(a) Relationships between variables (refer to **Table III.2.1** for trait description and calculation). (b) Distribution of genotypes according to their phenotypic values (genotypes are depicted based on their genomic cluster).

GDD in March, 715 GDD in August and 818 GDD in September), but the number of days required to reach this cumulative GDD at harvest was significantly different between the August trial (28 days) and the two other ones (42-43 days) (**Figure III.2.1.a**). In addition, the PARc was also significantly different between the experiments with the highest value in March ( $28638 \text{ J.cm}^{-2}$ ) compared to August ( $18784 \text{ J.cm}^{-2}$ ) and September ( $14826 \text{ J.cm}^{-2}$ ) (**Figure III.2.1.b**). These differences in temperatures and radiations can support differences observed in mean plant total dry biomass, with higher values in March compared to what was obtained in the August and September trials (**Figure III.2.1.c**, **Supplemental Table III.2.2**). Based on these data, we concluded that an environmental effect existed between the three trials that required to be corrected. A linear mixed model was applied on the raw data considering the random effect of blocks as well as the fixed effect of genotypes and checks in each block. The adjusted values obtained for each genotype were therefore BLUE values. After correction, the mean TDW of the whole panel was equal to 4.28 g per plant (**Figure III.2.1.c**).

#### The diversity set was structured in three distinct genomic clusters

Genetic structure of the whole population was assessed on the basis of SNP data using the first two axes of the Principal Coordinates Analysis (PCoA) that explained 34% and 13% of the genetic variance respectively. Three genomic clusters (C1-C3) were identified (**Figure III.2.2**). First axis discriminated winter oilseed types (C1, n=141, incl. 10 winter fodder and three spring oleaginous accessions) from spring oilseed genotypes (C2, n=65, incl. one African WOSR). Second axis primarily distinguished Asian genotypes from other types of rapeseed (C3, n=27, incl. 2 European swede and 3 European winter fodder types). Using the fixation index  $F_{ST}$  metric, the three genomic clusters appeared to be strongly differentiated, with  $F_{ST}$  values of 0.49, 0.46 and 0.47 between C1-C2, C1-C3 and C2-C3 respectively. In addition, setting a threshold on the top 5%  $F_{ST}$ , we identified a total of 226 genomic regions with strong differentiation between clusters C1 and C2, 251 between C1 and C3 and 218 between C2 and C3 (**Supplemental Table III.2.3**).

#### Early vigor variability can be described in terms of growth speed, N and C acquisition, biomass allocation and nitrogen utilization efficiency

When a PCA was performed using values of all the 23 traits selected for our study, we identified four principal dimensions that explained altogether a total of 86% of the phenotypic variability (**Figure III.2.3.a; Supplemental Figure III.2.1**). The first dimension (41% of the explained variance) was positively associated with high values of AUC and Sp traits, illustrating the genotype capacity to exhibit rapid early growth. The second dimension explained 18% of the variance and was positively associated with biomass (TDW, ADW, RDW), LA, nitrogen absorption (SNQ, TNQ) and carbon assimilation (SCA),

**Table III.2.2. Adjusted mean values (BLUE) and heritabilities ( $H^2$ ) for each trait over the whole population and for each genomic cluster (C1-C3)**

Traits	Mean pop.	Mean C1	Mean C2	Mean C3	$H^2$ pop.	$H^2$ C1	$H^2$ C2	$H^2$ C3
<b>Developmental traits – Growth dynamics</b>								
AUC <sub>L1-L4</sub>	8574.66 (1563.82)	8734.28 <sup>a</sup> (1360.46)	8227.97 <sup>a</sup> (1837.72)	8575.68 <sup>a</sup> (1767.72)	0.13	0.24	0.44	0
AUC <sub>L1-L2</sub>	641.77 (128.94)	657.39 <sup>a</sup> (115.26)	608.22 <sup>a</sup> (147.39)	641.01 <sup>a</sup> (138.54)	0.10	0.04	0.49	0.33
AUC <sub>L2-L3</sub>	2629.52 (493.37)	2681.62 <sup>a</sup> (428.75)	2513.84 <sup>a</sup> (582.32)	2635.99 <sup>a</sup> (547.06)	0.13	0.17	0.48	0.06
AUC <sub>L3-L4</sub>	5303.95 (970.13)	5391.35 <sup>a</sup> (834.33)	5114.27 <sup>a</sup> (1161.11)	5304.20 <sup>a</sup> (1095.35)	0.14	0.29	0.41	0
Sp <sub>L1-L4</sub>	0.50 (0.10)	0.51 <sup>a</sup> (0.08)	0.49 <sup>a</sup> (0.12)	0.51 <sup>a</sup> (0.11)	0.13	0.35	0.35	0
Sp <sub>L1-L2</sub>	0.16 (0.03)	0.16 <sup>a</sup> (0.03)	0.15 <sup>a</sup> (0.04)	0.16 <sup>a</sup> (0.03)	0.17	0.19	0.51	0.07
Sp <sub>L2-L3</sub>	0.36 (0.07)	0.37 <sup>a</sup> (0.06)	0.35 <sup>a</sup> (0.08)	0.36 <sup>a</sup> (0.08)	0.18	0.30	0.43	0
Sp <sub>L3-L4</sub>	0.93 (0.19)	0.93 <sup>a</sup> (0.15)	0.92 <sup>a</sup> (0.25)	0.93 <sup>a</sup> (0.20)	0.10	0.38	0.30	0
<b>Physiological traits</b>								
<b>Biomass and surface</b>								
ADW	4.03 (0.90)	4.28 <sup>a</sup> (0.59)	3.55 <sup>b</sup> (1.11)	3.86 <sup>b</sup> (1.16)	0.52	0.50	0.36	0.45
RDW	0.25 (0.08)	0.27 <sup>a</sup> (0.04)	0.23 <sup>b</sup> (0.11)	0.23 <sup>b</sup> (0.12)	0.39	0.004	0.21	0.21
TDW	4.28 (0.96)	4.54 <sup>a</sup> (0.61)	3.78 <sup>b</sup> (1.20)	4.09 <sup>b</sup> (1.25)	0.52	0.45	0.35	0.43
LA	925.66 (169.39)	968.67 <sup>a</sup> (129.39)	932.81 <sup>b</sup> (189.80)	924.55 <sup>ab</sup> (213.56)	0.52	0.12	0.49	0.63
SLA	253.86 (17.66)	252.50 <sup>a</sup> (17.51)	254.30 <sup>a</sup> (18.37)	259.90 <sup>a</sup> (15.90)	0.02	0	0	0.22
LAR	240.22 (16.72)	239.11 <sup>a</sup> (16.55)	240.08 <sup>a</sup> (17.23)	246.30 <sup>a</sup> (15.62)	0.03	0	0	0
SNQ	0.22 (0.04)	0.22 <sup>a</sup> (0.03)	0.20 <sup>b</sup> (0.05)	0.21 <sup>b</sup> (0.05)	0.45	0.19	0.42	0.41
TNQ	0.22 (0.04)	0.23 <sup>a</sup> (0.03)	0.21 <sup>b</sup> (0.05)	0.21 <sup>b</sup> (0.05)	0.46	0.19	0.42	0.40
<b>Efficiencies</b>								
SR	18.76 (3.34)	19.47 <sup>a</sup> (3.55)	17.27 <sup>b</sup> (2.54)	18.67 <sup>ab</sup> (2.71)	0.10	0	0.05	0
LWR	09445 (0.009)	0.9456 <sup>a</sup> (0.007)	0.9416 <sup>b</sup> (0.01)	0.9461 <sup>ab</sup> (0.01)	0.15	0	0.08	0
LASNQ	4594.40 (364.34)	4628.37 <sup>a</sup> (342.47)	4453.39 <sup>b</sup> (363.43)	4753.69 <sup>a</sup> (380.02)	0.20	0.04	0.11	0.49
SCA	4.99x10 <sup>-4</sup> (7.78x10 <sup>-5</sup> )	5.16x10 <sup>-4</sup> <sup>a</sup> (6.36x10 <sup>-5</sup> )	4.70x10 <sup>-4</sup> <sup>b</sup> (9.15x10 <sup>-5</sup> )	8.84x10 <sup>-4</sup> <sup>b</sup> (8.82x10 <sup>-5</sup> )	0.17	0	0.16	0
SNU	1.08 (0.19)	1.11 <sup>a</sup> (0.20)	1.02 <sup>b</sup> (0.13)	1.06 <sup>ab</sup> (0.15)	0.07	0	0	0
NuTe <sub>tot</sub>	18.86 (1.00)	19.08 <sup>a</sup> (0.59)	18.30 <sup>b</sup> (1.34)	19.12 <sup>a</sup> (1.27)	0.15	0.17	0	0
NuTe <sub>aer</sub>	17.80 (0.94)	18.01 <sup>a</sup> (0.58)	17.24 <sup>b</sup> (1.26)	18.09 <sup>a</sup> (1.07)	0.18	0.13	0	0

Standard deviation (SD) are given into parentheses. The superscript letters (a, ab and b) represent the groups of significance.

thus representing the plant capacity to acquire both N and C resources and convert them into biomass. The third dimension (14% of the variance) was positively associated with SR, LWR and SNU, suggesting a preferential biomass allocation to the leaves. The fourth dimension (11% of the variance) was positively associated with LASNQ, SLA, NUT<sub>E<sub>TOT</sub></sub>, NUT<sub>E<sub>AER</sub></sub>, LAR, and negatively associated with SNU, LWR and SR, representing the ability of the plant to utilize the absorbed nitrogen to elaborate leaf area and biomass.

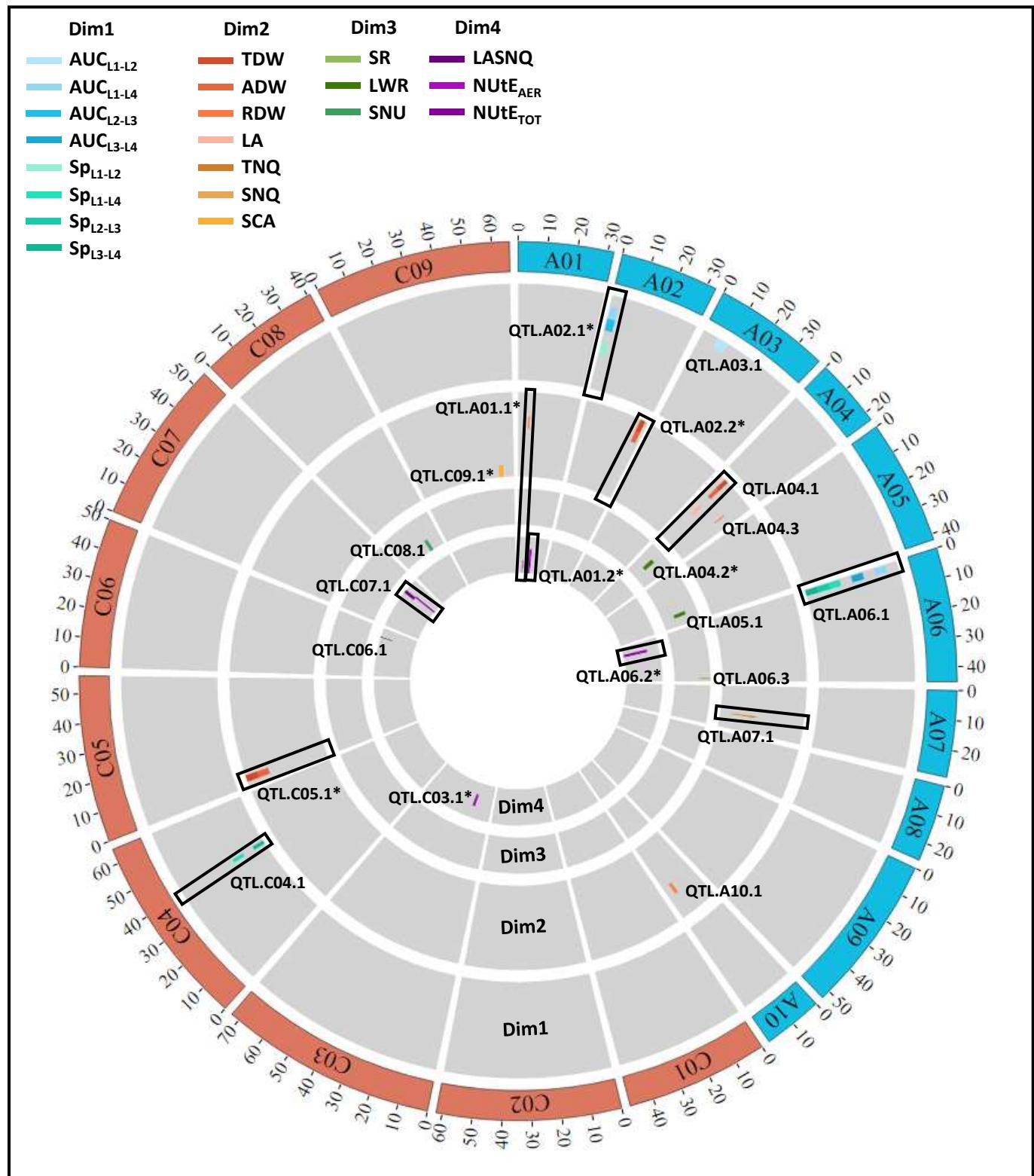
#### Rapeseed genotypes showed different behaviors for early vigor related traits depending on genomic clusters

The repartition of the genotypes within the first two-axes frameworks of PCA showed no preferential distribution according to the genomic cluster whatever the dimensions (**Figure III.2.3.b; Supplemental Figure III.2.2**), suggesting a low impact of the population genetic structure on early vigor traits when the traits were considered altogether. However, putative functional differences between rapeseed types were highlighted when the adjusted mean values (BLUE) for each single trait were compared between the three genomic clusters (**Table III.2.2**). Indeed, higher mean values were always observed in C1 when compared to C2 for the biomass traits (ADW, RDW, TDW), N traits (SNQ, TNQ, SNU, NUT<sub>E<sub>tot</sub></sub>, NUT<sub>E<sub>aer</sub></sub>) and efficiencies (SR, LASNQ, SCA). These results suggested that WOSR genotypes were generally more efficient at assimilating carbon and nitrogen and convert the nutrients into biomass than their spring counterparts. Interestingly, Asian genotypes displayed an intermediate phenotype between C1 and C2 genotypes. In contrast, no difference was observed between the three groups for the developmental traits AUC and Sp as well as for the physiological traits LAR and SLA.

Heritability values were low to moderate at the level of the whole population depending on the targeted trait with values ranging from 0.02-0.20 for developmental traits and efficiencies while reaching 0.39-0.52 for traits related to biomass and surface (**Table III.2.2**). When genomic clusters were considered, our results showed very low to null heritabilities for all the efficiency traits as well as for LAR and SLA. However, a variability exists for other traits between clusters. In addition, heritabilities in the C2 cluster were always higher than for C1 or C3 (except for ADW and TDW).

#### Early vigor is controlled by 21 genomic regions, half of which exhibited strong differentiations between genomic clusters

A total of 38 trait-associated SNP was identified by GWAS using the whole diversity set (**Figure III.2.4; Supplemental Table III.2.4**). For each locus, the explained variance ( $R^2$ ) ranged from 0.65% to 14.69%, indicating low to moderate effects. A first result was the absence of loci identified for SLA and LAR that was consistent with the fact that these two traits displayed the lowest heritabilities (0.02 and 0.03



**Figure III.2.4. Genetic architecture of early vigor traits in rapeseed.**

Each track corresponds to a component of early vigor related traits (PCA dimensions 1, 2, 3 or 4). Colocations of QTL are framed and stars (\*) indicate colocations between QTL and highly differentiated regions ( $F_{ST}$ ).

respectively). When examining the positions of the 38 loci, colocations were shown that reduced the QTL list to 21 genomic regions controlling early vigor (**Figure III.2.4**). Most of the SNP colocations concerned highly correlated traits such those involved in growth dynamics (QTL.A02.1, QTL.A06.1, QTL.C04.1), biomass and surface (QTL.A02.2, QTL.A04.1, QTL.C05.1), N absorption (QTL.A07.1) or N use efficiency (QTL.A01.2, QTL.A06.2), which is far from surprising. In contrast, QTL.A01.1 and QTL.C07.1 supported traits that were not highly correlated (RDW and  $\text{NUtE}_{\text{TOT}}$  for QTL.A01.1 and LASNQ,  $\text{NUtE}_{\text{TOT}}$ ,  $\text{NUtE}_{\text{AER}}$  for QTL.C07.1). At these positions, the favorable allele was the same for the different traits controlled by the locus, indicating possible pleiotropic effects of these regions. In addition, when considering the 21 multi-trait QTL, we showed that four were related to growth dynamics (QTL.A02.1, QTL.A03.1, QTL.A06.1 and QTL.C04.1 for AUC or Sp traits) that were assigned to the first dimension of the PCA (**Supplemental Figure III.2.1**). On the same line, seven regions supported traits related to nutrient acquisition and conversion into biomass as illustrated by the second dimension of the PCA (QTL.A02.2, QTL.A04.1, QTL.A04.3, QTL.A07.1, QTL.A10.1, QTL.C05.1 and QTL.C09.1 for ADW, RDW, TDW, SNQ, TNQ, LA and SCA). Four regions were identified for the third PCA dimension traits related to biomass allocation (QTL.A04.2, QTL.A05.1, QTL.A06.3 and QTL.C08.1 for SR, LWR and SNU). And finally, six QTL were related to N traits and PCA fourth dimension (QTL.A01.1, QTL.A01.2, QTL.A06.2, QTL.C03.1, QTL.C06.1 and QTL.C07.1 for LASNQ,  $\text{NUtE}_{\text{TOT}}$  and  $\text{NUtE}_{\text{AER}}$ ). Altogether, these results highlighted a specific genetic architecture for each component of early vigor (**Figure III.2.4**). To address the question of potential differences in allele distribution between genomic clusters within the 21 QTL, we compared the GWAS results with the genome scan results based on the  $F_{\text{ST}}$  metrics (**Supplemental Tables III.3 and III.2.4**). Our data revealed that 10 of the 21 QTL colocalized with genomic regions with strong differentiation between clusters as follows: two QTL (QTL.A02.1 and QTL.C05.1) corresponded C1-C2  $F_{\text{ST}}$  regions; four QTL (QTL.A01.2; QTL.A04.2; QTL.C03.1 and QTL.C09.1) corresponded to C1-C3  $F_{\text{ST}}$  regions; three QTL (QTL.A01.1; QTL.A02.2 and QTL.A06.2) corresponded to  $F_{\text{ST}}$  regions found between C1-C2 and C1-C3 and one QTL (QTL.A02.2) corresponded to  $F_{\text{ST}}$  regions found between C1-C2 and C2-C3.

**Most of the favorable alleles associated with early vigor traits were found to be specific to a single genomic cluster**

In order to investigate a putative unbalanced distribution of favorable alleles underlying early vigor related traits within the whole population, their frequencies were compared between the genomic clusters for each QTL. Alleles were considered as favorable when they increased the value of the trait. Overall, our results showed that the frequencies of favorable alleles differed significantly between the three groups (**Table III.2.3**). One first striking observation was that for two loci, namely QTL.A06.3 (SR)

**Table III.2.3. Frequencies of favorable alleles underlying early vigor trait regarding genetic clusters**

QTL PCA dimension	Trait/ $F_{ST}$ region	Freq. fav. allele C1 (n=141)	Freq. fav. allele C2 (n=65)	Freq. fav. allele C3 (n=27)
QTL.A02.1 <i>Dim.1</i>	AUC <sub>L1-L4</sub> AUC <sub>L2-L3</sub> Sp <sub>L1-L2</sub> $F_{ST}$ C1-C2 (5 regions)	<b>0.9</b>	0.06	0.19
QTL.A06.2 <i>Dim.4</i>	NUT $E_{aer}$ NUT $E_{tot}$ $F_{ST}$ C1-C2 (2 regions) $F_{ST}$ C1-C3 (1 region)	<b>0.85</b>	0.14	0.33
QTL.A07.1 <i>Dim.2</i>	SNQ TNQ	<b>0.57</b>	0.49	0.37
QTL.C09.1 <i>Dim.2</i>	SCA $F_{ST}$ C1-C3 (2 regions)	<b>0.96</b>	0.28	0.22
QTL.A04.2 <i>Dim.3</i>	LWR $F_{ST}$ C1-C3 (2 regions)	<b>0.81</b>	0.14	0.19
QTL.A05.1 <i>Dim.3</i>	LRW	<b>0.87</b>	0.38	0.4
QTL.A02.2 <i>Dim.2</i>	ADW TDW $F_{ST}$ C1-C2 (1 region) $F_{ST}$ C2-C2 (1 region)	0.05	<b>0.89</b>	0.33
QTL.A03.1 <i>Dim.1</i>	AUC <sub>L1-L2</sub> $F_{ST}$ C1-C2 (1 region) $F_{ST}$ C1-C3 (9 regions)	0.04	<b>0.58</b>	0.11
QTL.A06.1 <i>Dim.1</i>	AUC <sub>L1-L4</sub> AUC <sub>L3-L4</sub> Sp <sub>L1-L4</sub> Sp <sub>L2-L3</sub> Sp <sub>L3-L4</sub>	0.11	<b>0.65</b>	0.37
QTL.C05.1 <i>Dim.2</i>	ADW TDW $F_{ST}$ C1-C3 (1 region)	0.008	<b>0.52</b>	0.11
QTL.A01.1 <i>Dim.4</i>	NUT $E_{tot}$ RDW $F_{ST}$ C1-C2 (1 region) $F_{ST}$ C1-C3 (1 region)	0.11 0.13	0.15 0.45	<b>0.63</b> 0.07
QTL.A04.3 <i>Dim.2</i>	RDW	0.13	0.34	<b>0.52</b>
QTL.C03.1 <i>Dim.4</i>	NUT $E_{aer}$ $F_{ST}$ C1-C3 (2 regions)	0.09	0.28	<b>0.81</b>
QTL.C04.1 <i>Dim.1</i>	Sp <sub>L1-L4</sub> Sp <sub>L3-L4</sub>	0.35	0.15	<b>0.93</b>
QTL.C06.1 <i>Dim.4</i>	LASNQ	0.35	0.09	<b>0.78</b>
QTL.A01.2 <i>Dim.4</i>	NUT $E_{tot}$ NUT $E_{aer}$ $F_{ST}$ C1-C3 (7 regions)	<b>0.96</b> <b>0.96</b>	0.42 0.42	<b>0.93</b> <b>0.93</b>
QTL.A04.1 <i>Dim.2</i>	ADW LA TDW	<b>0.85</b>	0.38	<b>0.96</b>
QTL.C08.1 <i>Dim.3</i>	SNU	0.21	<b>0.97</b>	<b>0.78</b>
QTL.C07.1 <i>Dim.4</i>	LASNQ NUT $E_{aer}$ NUT $E_{tot}$	0.16 0.03	<b>0.57</b> 0.42	<b>0.59</b> 0.48
QTL.A06.3 <i>Dim.3</i>	SR	0.26	0.05	0.41
QTL.A10.1 <i>Dim.2</i>	RDW	0.13	0.38	0.22

and QTL.A10.1 (RDW), the favorable alleles were always the minor alleles in any of the clusters. This suggested that the corresponding traits are far from fixed in our population. Then for 15 out of the 19 remaining QTL, the favorable alleles were the major allele in only one of the three clusters (C1: 6 QTL; C2: 4 QTL and C3: 5 QTL), thus demonstrating that it is possible to use one genomic group to improve the other two. In contrast, the favorable alleles were the major alleles in combination of two clusters for only four QTL (QTL.A01.2; QTL.A04.1; QTL.C07.1; QTL.C08.1). These four loci were shown to control traits related to all the PCA dimensions at the exception of dimension 1 (growth dynamics). Finally, it appeared that genotypes of each single cluster seemed to have accumulated favorable alleles for traits related to all of the early vigor components (PCA dimensions).

## Discussion

To decipher the relationships between early vigor related traits and unravel their genetic architecture, we used a diversity panel gathering 233 rapeseed lines spanning the three types of growing habit (WOSR, SOSR and Asian). Our results revealed that early vigor was qualified by traits supporting growth rate, N and C acquisition, as well as biomass allocation and nitrogen use efficiency. In addition, different behaviors to establish biomass and surfaces were observed between the rapeseed types regarding their growing habits. Then, 21 genomic regions associated with early vigor traits were identified. The majority of the favorable alleles at these regions were specific to a single genomic cluster, thereby demonstrating the value to integrate genetic diversity to improve one gene pool for particular traits related to early vigor.

### **The contrasting early vigor observed between rapeseed types can be attributed to differences in efficiency related to carbon and nitrogen uptake and utilization**

Most studies dealing with early vigor have focused on a few number of traits such as biomass and leaf area (Bac-Molenaar et al. 2015; Rebolledo et al. 2015; Knoch et al. 2020; Rahimi et al. 2023), specific leaf area ratio (Dingkuhn et al. 1998; Rebetzke et al. 2004), glucose quantity (Rebolledo et al. 2015) or leaf development rate (Rebolledo et al. 2012). However, early vigor is a complex process that encompasses a large number of traits. The power of our work relied on the number of traits that were scored at the same time over a large set of genetic diversity. The multivariate analysis highlighted the relationships between early vigor traits, thus helping to unravel the complexity of early vigor. Indeed, we were able to model early vigor as combinations of gradients in leaf area dynamics and biomass production (first and second PCA dimensions), biomass allocation from root parts to aerial parts (third PCA dimension), and a negative relationship between N uptake and utilization (fourth PCA dimension).



Given the low correlations between traits of the four early vigor components, it is likely that favorable alleles for these traits need to be combined to develop high vigorous ideotypes.

No differences in development speed (AUC and Sp), as well as SLA and LAR were observed between the genomic clusters, suggesting that the genetic resources available within each cluster should be used to improve these traits. On the contrary, the C1 genotypes (WOSR) appeared to be the most vigorous accessions in our study regarding a high number of traits (higher plant biomass, organ surface and efficiencies than for C2 or C3 genotypes), indicating that winter genotypes should be favored in breeding schemes for improving for these traits. Consistent results were reported by Haelterman and coworkers (2024) who showed that WOSR accessions had a higher root biomass than their SOSR counterparts. Other studies found a higher root growth rate in SOSR than in WOSR, but the latter produced more secondary roots and longer tap root (Kiran et al. 2019), while root nutrient contents were higher in SOSR than in WOSR (Thomas et al. 2016). The differences observed between rapeseed types could be related to their own requirements during the growing cycle. Indeed, winter types have a long cycle (10-11 months) but must settle 70% of their root biomass and 20% of their above-ground biomass before entering in the vegetative rest phase to overcome the cold period. As a consequence, high uptake efficiency and optimal use of nutrients allows these plants to develop rapidly before winter. In contrast, spring genotypes have a short cycle (5-6 months) with no vegetative pause. In the meantime, the C3 genotypes (mainly Asian) represented an interesting gene pool as they behave in between accessions of C1 and C2 clusters. It is noteworthy that Asian genotypes include more alleles introgressed from *Brassica rapa* than other rapeseed types, supporting a higher genetic diversity than this existing in the other gene pools (Qian et al. 2006). Finally, further investigations of early vigor traits using extended genetic diversity may provide a better understanding of the mechanisms underlying plant establishment and it could also help identifying new targets for breeding.

### New genomic regions controlling early vigor were detected

The polygenic nature of the early vigor related traits in rapeseed has been emphasized by our results, in line with previous studies concerning biomass or leaf area (Knob et al. 2020; Haelterman et al. 2024). A co-location at QTL.C07.1 was observed between traits involved in nitrogen utilization processes (LASNQ and NUTE) that could be supported by a common process involved in nitrogen metabolism (Xu et al. 2012) and therefore explained by pleiotropy. On the contrary, the second trait-based co-location observed at QTL.A01.1 concerned nitrogen utilization (NUTE) and root biomass (RDW). These traits are involved in different processes. Indeed, NUTE does not appear to be associated with root biomass or root morphology traits under nitrogen fertilization conditions at an early growth stage (Wang et al. 2017). As a consequence, we can hypothesize that these traits are controlled by two



genes that are physically linked at the locus QTL.A01.1. In addition, three other loci identified in our study co-located with QTL already published. It concerns QTL.A04.1 involved in total biomass, aerial biomass and leaf area that was shown to co-locate with a QTL for fresh biomass (Knoch et al. 2020). This locus would therefore be critical to increase plant biomass. QTL.A06.3 controlling SR was found to co-locate with a loci associated with lateral root density (Haelterman et al. 2024). However, Wang and coworkers (2017) reported that there is no relation between SR and root density. Finally, QTL.A10.1 involved in root biomass co-located with a QTL involved in total dry biomass and leaf area (Knoch et al. 2020) and a second locus involved in the density of lateral roots (Haelterman et al. 2024). A more detailed investigation of this genomic region would be useful in order to identify potential candidate genes involved in early vigor. Finally, our study revealed a high number of new QTL for early vigor that have never been described elsewhere to our knowledge. The identification of these QTL, combined to the frequency of favorable alleles within each gene pool could provide valuable insight that can be used to enhance early vigor in oilseed rape.

**Efficient allele mining from one gene pool to the other one can help improving early vigor in rapeseed**

Our results pointed out genomic regions controlling early vigor traits whose favorable alleles were carried predominantly by a specific genomic cluster. This highlighted the need to integrate large genetic resources from different types of rapeseed in pre-breeding schemes in order to introduce the maximum of alleles conferring high early vigor into breeding germplasm. In this way, depending on the traits, a specific genomic cluster can be targeted to enrich the gene pool with favorable alleles. For instance, the majority of C1 genotypes possess all the favorable alleles for nitrogen acquisition (SNQ and TNQ), carbon assimilation efficiency (SCA) and biomass allocation into aerial parts (LWR). Therefore, C1 genotypes are recommended for breeding purpose in order to improve these traits. To increase root biomass genotypes from genomic cluster C3 should be favored (RDW trait) while most lines from clusters C1 and C3 carry favorable alleles for nitrogen use efficiency (NUtE). Conversely, some QTL for root related traits showed favorable alleles with minor frequencies in all three populations: RDW (QTL.A01.1 and QTL.A10.1), SR (QTL.A06.3) and NUtE (QTL.C07.1). Specific genotypes within these three gene pools could therefore be used. In order to identify new alleles and genotypes for early vigor improvement, it would be necessary to screen genetic resources that are further genetically distant. These resources will then have to be integrated into the elite material. Several studies have reported the introduction of genetic diversity into WOSR from SOSR and *vice versa* (Butruille et al. 1999; Quijada et al. 2006; Kebede et al. 2010; Gourrion et al. 2020), as well as from Asian into SOSR or WOSR genotypes (Udall et al. 2004; Qian et al. 2007; Qian et al. 2009). For instance,



a significant yield increase was observed in SOSR lines that were derived from SOSR × WOSR crosses (Butruille et al. 1999; Kebede et al. 2010). However, the main issues of this genetic material are their vernalization requirement, precocity or even frost resistance thus imposing additional steps of selection to get rid of undesirable traits (Rahman 2013). In addition, selection for the seed quality must be also conducted when using material from old genetic resources or non-oilseed *Brassica napus* crosses such as swede or kale. To avoid all these drawbacks while targeting favorable alleles for early vigor, genomic selection should be favored in pre-breeding programs (Longin and Reif 2014; Gorjanc et al. 2016). Given that early vigor is polygenic, it is essential to consider the genetic background of genotypes rather than specific QTL. The utilization of genomic selection would facilitate the screening and the identification of valuable individuals that are genetically distant from the elite material (Crossa et al. 2016). To avoid the loss of favorable polygenic variations provided by genetic resources when crossing them by elite material, an intermediate population between diversity panels and elite germplasms is essential. This population, known as the bridging population, provides a means of reducing the disparity between the two populations while incorporating favorable alleles (Gorjanc et al. 2016; Allier et al. 2020).



**Acknowledgements** The authors acknowledge colleagues from INRAe (IGEPP), especially Aurélien Carrillo for help with image capture, Mathieu Rousseau-Gueutin, Cyril Falentin and Cécile Baron for help with the genomic resources of *Brassica napus* and genotyping data, as well as Laurent Charlon, Jean-Paul Constantin and Frédéric Le Tertre for supplying greenhouse services. The authors also thank the CRB BrACySol (INRAe) for providing seed resources and the Experimental Unit of La Motte (INRAe) for seed multiplications in the field. The GenOuest platform (<https://www.genouest.org/>) is acknowledged for providing informatics resources.

**Authors contribution statement** A.L. and N.N. designed the study and obtained fundings. M.L., E.A., S.G. and B.M carried out the tunnel experiments. S.J. developed the image analysis scripts and carried out the image analyses. M.L. performed all the statistical and the genetic analyses. C.R.M. and M.G. provided expertise in ecophysiology. M.L., A.L. and N.N. interpreted the results and wrote the manuscript that was approved by all the authors.

**Funding** This work was supported by the NAVIG project funded by Promosol association. M.L. was the recipient of a three years grant from INRAe and Promosol. The genomic data were acquired in the framework of the RAPSODYN project (ANR-11-BTBR-0004).

**Conflict of interest** The authors declare that they have no competing interest.



## Supplemental information

**Supplemental Table III.2.1.** Passport data of the 233 rapeseed genotypes used in the study.

**Supplemental Table III.2.2.** Overview of the experimental designs and raw datasets.

**Supplemental Table III.2.3.** Genomic regions with high  $F_{ST}$  values between clusters of individuals.

**Supplemental Table III.2.4.** Overview of the 38 loci controlling early vigor traits identified by GWAS in the population of 233 rapeseed lines.

**Supplemental Figure III.2.1.** Eigenvalues and contribution of early vigor traits to PCA dimensions.

**Supplemental Figure III.2.2.** Additional PCA plots of variables (left) and individuals (right).



**Supplemental Table III.2.1. Passport data of the 233 rapeseed genotypes used in the study**

Accession name	Type	Seed quality	Origin	Registration period
Abukuma	SOSR	++	Asia	NA
Argentine	SOSR	++	North America	NA
C.S.S.R	SOSR	++	Europe	NA
Ceska	SOSR	++	Europe	1970-1979
Chang Huzi	SOSR	++	Asia	NA
Chine 28	SOSR	++	Asia	NA
Chine 29	SOSR	++	Asia	NA
Chine 30	SOSR	++	Asia	NA
Chine 32	SOSR	++	Asia	NA
Chine 34	SOSR	++	Asia	NA
Comet	SOSR	++	Europe	1990-1999
Cresus	SOSR	++	Europe	2000-2010
Czyzowski	SOSR	++	Europe	NA
Dairyu	SOSR	++	Asia	NA
Egra	SOSR	++	Europe	1970-1979
Janetzki	SOSR	++	Europe	1931-1969
Kogane	SOSR	++	Asia	NA
Mali	SOSR	++	Europe	NA
Mazowiecki	SOSR	++	Europe	NA
Mlochowski	SOSR	++	Europe	NA
Murasaki	SOSR	++	Asia	1970-1979
Mutsu	SOSR	++	Asia	1970-1979
Nabo	SOSR	++	South America	NA
Norin 9	SOSR	++	Asia	1970-1979
Norin 10	SOSR	++	Asia	1970-1979
Nozowski	SOSR	++	Europe	NA
Regina	SOSR	++	Europe	1931-1969
Sielecki	SOSR	++	Europe	1980-1989
Suehiro	SOSR	++	Asia	NA
Taichung	SOSR	++	Asia	NA
Tanka	SOSR	++	North America	1931-1969
Tie Gieng Quing	SOSR	++	Asia	NA
Tokiwa	SOSR	++	Asia	NA
Yudal	SOSR	++	Asia	1931-1969
Brutor	SOSR	0+	Europe	1970-1979
Crop	SOSR	0+	Europe	1970-1979
Gulliver	SOSR	0+	Europe	1970-1979
Maris Haplona	SOSR	0+	Europe	1970-1979
Nachan	SOSR	0+	Asia	1970-1979
Niklas	SOSR	0+	Europe	1980-1989
Nozeouck	SOSR	0+	Asia	NA
Oro IPK	SOSR	0+	North America	1931-1969
Orpal	SOSR	0+	Europe	1970-1979
Romeo	SOSR	0+	Europe	1970-1979
Wesbell	SOSR	0+	Oceania	1980-1989
Wesreo	SOSR	0+	Oceania	1970-1979
Yeong Dang	SOSR	0+	Asia	NA
Zero erucic acid	SOSR	0+	North America	1970-1979
Bronowski D4	SOSR	+0	Europe	1931-1969
Alto	SOSR	00	Europe	2000-2010
Briol	SOSR	00	Europe	1990-1999
Calisto	SOSR	00	Europe	2000-2010
Caviar	SOSR	00	Europe	1990-1999
Cesar	SOSR	00	Europe	1980-1989
Cyclone	SOSR	00	Europe	1990-1999
Drakkar	SOSR	00	Europe	1980-1989
Dunkeld	SOSR	00	Oceania	1990-1999
Duplo	SOSR	00	Europe	1970-1979
Fidelio	SOSR	00	Europe	1990-1999
Forte	SOSR	00	Europe	1990-1999
Global	SOSR	00	Europe	1980-1989
Golda	SOSR	00	Europe	1980-1989
Hanna	SOSR	00	Europe	1980-1989
Helios	SOSR	00	Europe	1990-1999
Karat	SOSR	00	Europe	1980-1989
Lirawell	SOSR	00	Europe	1980-1989

**Supplemental Table III.2.1. Continued**

Accession name	Type	Seed quality	Origin	Registration period
Maluka	SOSR	00	Oceania	1980-1989
Marnoo	SOSR	00	Oceania	1980-1989
Mary	SOSR	00	Europe	1980-1989
Nora	SOSR	00	Europe	1970-1979
Omega	SOSR	00	Europe	1980-1989
Oscar	SOSR	00	Oceania	1990-1999
RP1	SOSR	00	Europe	NA
RP2	SOSR	00	Europe	1980-1989
Shiralee	SOSR	00	Oceania	1980-1989
Spok	SOSR	00	Europe	1990-1999
Sponsor	SOSR	00	Europe	1990-1999
Spoutnik	SOSR	00	Europe	1990-1999
Stellar	SOSR	00	North America	1980-1989
Tanto	SOSR	00	Europe	1990-1999
Topas	SOSR	00	Europe	1980-1989
Tribute	SOSR	00	North America	1980-1989
Wesbarker	SOSR	00	Oceania	1980-1989
Wesbrook	SOSR	00	Oceania	1980-1989
Westar	SOSR	00	North America	1980-1989
Bangholm	Swede	NA	Europe	1931-1969
Lord Derby	Swede	NA	Europe	NA
Nunsdale	WFR	NA	Europe	NA
Silona	WFR	NA	Europe	1931-1969
Arvor	WFR	++	Europe	1931-1969
Bladkool (Limburgse Bladkool)	WFR	++	Europe	1970-1979
Dwarf Essex	WFR	++	Europe	1931-1969
English Giant	WFR	++	Europe	1970-1979
Fora	WFR	++	Europe	1970-1979
Kentan	WFR	++	Europe	1970-1979
Liragrin	WFR	++	Europe	1970-1979
Palu	WFR	++	Europe	1970-1979
Parapluie	WFR	++	Europe	1931-1969
Samo	WFR	++	Europe	1970-1979
Liratop	WFR	00	Europe	1980-1989
Apex	WOSR	NA	Europe	1990-1999
Apex-93_5 X Ginyou_3 DH line	WOSR	NA	NA	NA
Beluga	WOSR	NA	Europe	2000-2010
Chembere Dzagumhana	WOSR	NA	Africa	NA
Evvin	WOSR	NA	Asia	NA
Exv8-DH_675	WOSR	NA	Europe	NA
Groene Groninger Snijmoes	WOSR	NA	Europe	1970-1979
Kvintett	WOSR	NA	Europe	NA
KW3077	WOSR	NA	Europe	NA
Lion	WOSR	NA	Europe	2000-2010
Lord	WOSR	NA	Europe	NA
Maplus	WOSR	NA	Europe	1990-1999
Missouri	WOSR	NA	Europe	2000-2010
Recital	WOSR	NA	Europe	2000-2010
Robust	WOSR	NA	Europe	NA
Shannon x Winner DH line	WOSR	NA	NA	NA
SLM 0413	WOSR	NA	Europe	NA
V8	WOSR	NA	NA	NA
Verona	WOSR	NA	Europe	2000-2010
Wild accession	WOSR	NA	Africa	NA
Abukuma natane	WOSR	++	Asia	NA
Aburamasari	WOSR	++	Asia	NA
Alsace 1360	WOSR	++	Europe	1931-1969
Blagodatnyi	WOSR	++	Europe	NA
Brauner Schnittkohl	WOSR	++	NA	1931-1969
Debruder Dippes	WOSR	++	Asia	NA
Dippes	WOSR	++	Europe	NA
Dublianski	WOSR	++	Europe	1931-1969
Ezsterhazy N.ST	WOSR	++	Europe	NA
Fertodi Elit	WOSR	++	Europe	NA
Gaspard	WOSR	++	Europe	1980-1989
Goldgelber Zarter Butter	WOSR	++	Europe	1931-1969

**Supplemental Table III.2.1. Continued**

Accession name	Type	Seed quality	Origin	Registration period
Gross Luesewitzer	WOSR	++	Europe	1931-1969
GÜLZOWER ÖLQUELL	WOSR	++	Europe	1931-1969
Hochzucht Repole	WOSR	++	Asia	NA
Hokkai 3-Go	WOSR	++	Asia	NA
K26-96	WOSR	++	NA	NA
Kormovoij Skriverskyi	WOSR	++	Europe	NA
Lembke	WOSR	++	Europe	1931-1969
Ljubljana	WOSR	++	Europe	NA
Major	WOSR	++	Europe	1970-1979
Matador	WOSR	++	Europe	1970-1979
Moldavia	WOSR	++	Asia	1931-1969
Niederarnbacher	WOSR	++	Europe	NA
Novochartoryiskiyi	WOSR	++	Asia	NA
Oleski	WOSR	++	Asia	1931-1969
Panter	WOSR	++	Europe	1931-1969
Pollux	WOSR	++	Europe	NA
Quedlinburger Platzfester	WOSR	++	Europe	1931-1969
Ramses	WOSR	++	Europe	1970-1979
Rangi	WOSR	++	Oceania	1970-1979
Rapol	WOSR	++	Europe	1970-1979
Sarepta	WOSR	++	Europe	1931-1969
Skrzeszowicki	WOSR	++	Europe	1931-1969
Slapska	WOSR	++	Europe	1931-1969
Snitynskyi	WOSR	++	Europe	NA
Sobotkowski	WOSR	++	Europe	1931-1969
Sollux	WOSR	++	Europe	1970-1979
Sonnengold	WOSR	++	Europe	NA
Taiwan	WOSR	++	Asia	NA
Trebicska	WOSR	++	Asia	NA
Victor	WOSR	++	Europe	1990-1999
Warzanowski	WOSR	++	Europe	1931-1969
Aztec	WOSR	0+	Europe	1980-1989
Belinda	WOSR	0+	Europe	1980-1989
Beryl	WOSR	0+	Europe	1980-1989
Brink	WOSR	0+	Europe	1980-1989
Constable	WOSR	0+	Europe	1970-1979
Doral	WOSR	0+	Europe	1980-1989
Emil	WOSR	0+	Europe	1980-1989
JetNeuf	WOSR	0+	Europe	1970-1979
Jupiter	WOSR	0+	Europe	1980-1989
Korina	WOSR	0+	Europe	1980-1989
Krasnodarskyi	WOSR	0+	Asia	NA
Kurander	WOSR	0+	Europe	1980-1989
Ligora	WOSR	0+	Europe	1980-1989
Mikado	WOSR	0+	Europe	1980-1989
Quinta	WOSR	0+	Europe	1970-1979
Sinus	WOSR	0+	Europe	1970-1979
Valdor	WOSR	0+	Europe	2000-2010
Zeruca	WOSR	+0	Europe	1990-1999
Alesi	WOSR	00	Europe	2000-2010
Amber	WOSR	00	Europe	1990-1999
Astrid	WOSR	00	Europe	2000-2010
Bolko	WOSR	00	Europe	1980-1989
Boston	WOSR	00	Europe	1990-1999
Bristol	WOSR	00	Europe	1990-1999
Capricorn	WOSR	00	Europe	1990-1999
Columbus	WOSR	00	Europe	1990-1999
Colvert	WOSR	00	Europe	1990-1999
Contact	WOSR	00	Europe	1990-1999
Coriander	WOSR	00	Europe	NA
Darmor	WOSR	00	Europe	1980-1989
DS009	WOSR	00	Europe	NA
DS036	WOSR	00	Europe	NA
Eurol	WOSR	00	Europe	1990-1999
Expert	WOSR	00	Europe	2000-2010
Express	WOSR	00	Europe	1990-1999

**Supplemental Table III.2.1. Continued**

Accession name	Type	Seed quality	Origin	Registration period
Goeland	WOSR	00	Europe	1990-1999
Hektor	WOSR	00	Europe	1990-1999
Inca	WOSR	00	Europe	1990-1999
Jantar	WOSR	00	Europe	1980-1989
Laser	WOSR	00	Europe	1990-1999
Licorne	WOSR	00	Europe	2000-2010
Lipton	WOSR	00	Europe	NA
Lisabeth	WOSR	00	Europe	1990-1999
Madora	WOSR	00	Europe	1990-1999
Madrigal	WOSR	00	Europe	1990-1999
Mar	WOSR	00	Europe	1990-1999
Maxol	WOSR	00	Europe	1990-1999
Milena	WOSR	00	Europe	2000-2010
Mohican	WOSR	00	Europe	1990-1999
Montego	WOSR	00	Europe	2000-2010
Navajo	WOSR	00	Europe	1990-1999
NK Bravour	WOSR	00	Europe	2000-2010
NK Passion	WOSR	00	Europe	2000-2010
Oase	WOSR	00	Europe	2000-2010
Olimpiade2-DS2	WOSR	00	Europe	NA
Olymp	WOSR	00	Europe	1990-1999
Pacific	WOSR	00	Europe	2000-2010
Pirola	WOSR	00	Europe	NA
Pollen	WOSR	00	Europe	1990-1999
Potomac	WOSR	00	Europe	2000-2010
Remy	WOSR	00	Europe	2000-2010
Roxet	WOSR	00	Europe	2000-2010
Salomont	WOSR	00	Europe	2000-2010
Samourai	WOSR	00	Europe	1980-1989
Savannah	WOSR	00	Europe	2000-2010
Start	WOSR	00	Europe	2000-2010
SW Gospel	WOSR	00	Europe	2000-2010
Tapidor	WOSR	00	Europe	1980-1989
Tenor	WOSR	00	Europe	1990-1999
Twister	WOSR	00	Europe	2000-2010

SOSR, spring oilseed rape

WOSR, winter oilseed rape

WFR, winter fodder rape

++, varieties rich in erucic acid (C22:1) and glucosinolates (GSL)

+0, varieties rich in C22:1 and poor in GSL

0+, varieties poor in C22:1 and rich in GSL

00, varieties poor in C22:1 and in GSL

NA, not available

**Supplemental Table III.2.2. Overview of the experimental designs and raw datasets**

		<b>March trial</b>	<b>August trial</b>	<b>September trial</b>
	<b>Sowing date (month/day/year)</b>	03/15/2022	08/18/2022	09/28/2022
	<b>Mean daily temperature (°C)</b>	17.3	24.6	18.4
	<b>Minimum daily temperature (°C)</b>	6.3	16.3	13.2
	<b>Maximum daily temperature (°C)</b>	28.4	32.9	23.6
	<b>Harvest date (month/day/year)</b>	04/26/2022	09/15/2022	11/10/2022
	<b>Days after sowing</b>	42	28	43
	<b>GDD cumulated at harvest (base 0°C)</b>	758	715	818
	<b>PAR cumulated at harvest</b>	28638	18784	14826
<b>AUC<sub>L1-L4</sub>, Area under the curve bw Leaves L1 and L4</b>	POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	10525.31 (2543.19) 10790.31 (1853.07) 11443.72 (2638.149)	8607.96 (2175.76) 9558.40 (3420.23) 8019.99 (2540.41)	6585.73 (1277.30) 5756.59 (547.52) 6875.99 (687.25)
<b>AUC<sub>L1-L2</sub>, Area under the curve bw Leaves L1 and L2</b>	POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	694.71 (174.43) 744.58 (81.73) 755.47 (151.21)	679.19 (184.31) 752.12 (284.67) 633.45 (229.71)	558.55 (122.89) 511.45 (88.83) 535.30 (31.36)
<b>AUC<sub>L2-L3</sub>, Area under the curve bw Leaves L2 and L3</b>	POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	3105.53 (752.08) 3239.69 (486.06) 3383.25 (742.61)	2676.51 (689.33) 3000.96 (1095.15) 2520.43 (838.80)	2119.26 (422.70) 1882.69 (226.69) 2150.34 (182.17)
<b>AUC<sub>L3-L4</sub>, Area under the curve bw Leaves L3 and L4</b>	POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	6725.08 (1654.40) 6806.04 (1285.76) 7305.00 (1745.51)	5252.26 (1318.89) 5805.32 (2040.97) 4866.11 (1473.53)	3907.92 (753.59) 3362.45 (238.35) 4190.36 (476.37)
<b>Sp<sub>L1-L4</sub>, Speed of dev. bw. leaves L1 and L4</b>	POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	0.68 (0.17) 0.67 (0.14) 0.73 (0.18)	0.48 (0.12) 0.53 (0.18) 0.45 (0.13)	0.35 (0.07) 0.29 (0.01) 0.39 (0.05)
<b>Sp<sub>L1-L2</sub>, Speed of dev. bw. leaves L1 and L2</b>	POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	0.18 (0.05) 0.19 (0.03) 0.20 (0.04)	0.16 (0.04) 0.18 (0.06) 0.15 (0.05)	0.12 (0.02) 0.11 (0.01) 0.13 (0.01)

**Supplemental Table III.2.2. Continued**

		March trial	August trial	September trial
<b>Sp<sub>L2-L3</sub>, Speed of dev. bw. leaves L2 and L3</b>	POPULATION Means (SD)	0.47 (0.12)	0.36 (0.09)	0.26 (0.05)
	MIRO Means (SD)	0.47 (0.09)	0.40 (0.14)	0.23 (0.01)
	AVISO Means (SD)	0.51 (0.12)	0.33 (0.10)	0.29 (0.03)
<b>Sp<sub>L3-L4</sub>, Speed of dev. bw. leaves L3 and L4</b>	POPULATION Means (SD)	1.29 (0.35)	0.87 (0.22)	0.61 (0.13)
	MIRO Means (SD)	1.26 (0.29)	0.96 (0.32)	0.51 (0.02)
	AVISO Means (SD)	1.39 (0.36)	0.80 (0.22)	0.70 (0.10)
<b>ADW, Aerial Dry Weight (g/plant)</b>	POPULATION Means (SD)	5.98 (2.99)	2.42 (1.06)	3.68 (1.42)
	MIRO Means (SD)	6.37 (2.74)	2.57 (1.05)	3.24 (1.18)
	AVISO Means (SD)	6.70 (3.15)	1.99 (0.84)	3.67 (1.37)
<b>RDW, Root Dry Weight (g/plant)</b>	POPULATION Means (SD)	0.45 (0.26)	0.14 (0.07)	0.17 (0.08)
	MIRO Means (SD)	0.57 (0.28)	0.17 (0.07)	0.20 (0.09)
	AVISO Means (SD)	0.44 (0.22)	0.11 (0.05)	0.18 (0.08)
<b>TDW, Total Dry Weight (g/plant) = ADW+RDW</b>	POPULATION Means (SD)	6.43 (3.23)	2.56 (1.12)	3.85 (1.49)
	MIRO Means (SD)	6.94 (3.00)	2.74 (1.12)	3.44 (1.26)
	AVISO Means (SD)	7.14 (3.35)	2.10 (0.88)	3.85 (1.45)
<b>LA, Leaf Area (cm<sup>2</sup>)</b>	POPULATION Means (SD)	1105.47 (447.06)	645.62 (254.41)	1029.88 (338.54)
	MIRO Means (SD)	1167.43 (392.93)	673.64 (216.52)	887.12 (270.91)
	AVISO Means (SD)	1201.54 (377.11)	539.31 (189.24)	1040.47 (342.44)
<b>SLA, Specific Leaf Area (cm<sup>2</sup>.g<sup>-1</sup>) = LA/ADW</b>	POPULATION Means (SD)	197.07 (36.57)	276.36 (47.76)	289.34 (38.09)
	MIRO Means (SD)	192.67 (28.26)	281.24 (88.10)	282.84 (31.30)
	AVISO Means (SD)	196.69 (56.32)	289.26 (83.27)	291.41 (26.72)
<b>LAR, Leaf Area Ratio (cm<sup>2</sup>.g<sup>-1</sup>) = LA/TDW</b>	POPULATION Means (SD)	183.71 (34.76)	261.10 (43.63)	276.98 (36.92)
	MIRO Means (SD)	177.29 (26.85)	263.64 (81.82)	267.02 (29.92)
	AVISO Means (SD)	183.68 (47.24)	237.72 (78.55)	278.01 (26.64)
<b>%SN, Shoot nitrogen content (g/100 g) = (%SN×ADW)/100</b>	POPULATION Means (SD)	4.88 (0.61)	6.25 (0.42)	5.50 (0.42)
	MIRO Means (SD)	4.85 (0.77)	6.03 (0.63)	5.83 (0.42)
	AVISO Means (SD)	4.82 (0.74)	6.47 (0.36)	5.68 (0.35)
<b>SNQ, Shoot Nitrogen Quantity (g) = (%SN×ADW)/100</b>	POPULATION Means (SD)	0.29 (0.14)	0.15 (0.07)	0.20 (0.08)
	MIRO Means (SD)	0.30 (0.11)	0.16 (0.07)	0.19 (0.08)
	AVISO Means (SD)	0.31 (0.13)	0.13 (0.06)	0.21 (0.08)

**Supplemental Table III.2.2. Continued**

				<b>March trial</b>	<b>August trial</b>	<b>September trial</b>
<b>TNQ, Total Nitrogen Quantity (g) =</b> POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	0.30 (15)			0.16 (0.07)		0.21 (0.09)
	0.32 (0.12)			0.16 (0.07)		0.20 (0.08)
	0.33 (0.14)			0.13 (0.06)		0.21 (0.08)
<b>SR, Shoot:Root Ratio = ADW/RDW</b> POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	14.50 (3.94)			18.58 (5.46)		23.39 (5.57)
	11.69 (1.75)			16.02 (3.70)		18.55 (10.05)
	15.75 (2.71)			18.13 (3.22)		20.85 (3.02)
<b>LWR, Leaf Weight Ratio = ADW/TDW</b> POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	0.93 (0.02)			0.95 (0.01)		0.96 (0.01)
	0.92 (0.01)			0.94 (0.01)		0.94 (0.01)
	0.94 (0.02)			0.95 (0.009)		0.95 (0.006)
<b>LASNQ, Leaf Area:Shoot Nitrogen Quantity Ratio = LA/SNQ</b> POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	4088.98 (879.54)			4441.844 (868.989)		5298.01 (907.39)
	4019.18 (601.61)			4798.66 (2087.22)		4888.01 (752.61)
	4161.81 (1590.19)			4481.45 (1272)		5152.69 (580.67)
<b>SCA, Specific Carbon Assimilation = TDW/AUC</b> POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	2.70x10e-4 (1.23x10e-4)			1.34x10e-4 (6.29x10e-5)		2.61x10e-4 (1.02x10e-4)
	2.68x10e-4 (1.20x10e-4)			1.28x10e-4 (6.32x10e-5)		2.47x10e-4 (8.22x10e-5)
	2.88x10e-4 (1.28x10e-4)			1.16x10e-4 (5.51x10e-5)		2.45x10e-4 (9.29x10e-5)
<b>SNU, Specific Nitrogen Uptake = TNQ/RDW</b> POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	0.737 (0.214)			1.19 (0.346)		1.31 (0.303)
	0.601 (0.151)			0.993 (0.247)		1.11 (0.611)
	0.793 (0.18)			1.20 (0.191)		1.22 (0.181)
<b>NUT<sub>tot</sub>, Total Nitrogen Utilization Efficiency = TDW/TNQ</b> POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	21.4 (2.69)			16.6 (1.20)		18.7 (1.48)
	21.8 (3.13)			17.3 (2.03)		17.7 (1.28)
	21.7 (3.18)			16.0 (0.944)		18.1 (1.08)
<b>NUT<sub>Eaer</sub>, Aerial Nitrogen utilization efficiency = ADW/TNQ</b> POPULATION Means (SD) MIRO Means (SD) AVISO Means (SD)	20.0 (2.46)			15.7 (1.10)		17.9 (1.42)
	20.0 (2.78)			16.3 (1.92)		16.8 (1.23)
	20.3 (2.79)			15.1 (0.959)		17.3 (1.04)

Cumulative growing degree days (GDD) and cumulated photosynthetic active radiation (PAR) were calculated from sowing to harvest.

For each trait, the mean is represented as well as the standard deviation (SD) into parentheses.

**Supplemental Table III.2.3. Genomic regions with high  $F_{ST}$  values between clusters of individuals**

Regions that colocalize with a QTL (refer to Supplemental Table III.2.4)

**High  $F_{ST}$  regions between clusters C1 and C3**

$F_{ST}$ region name	Chrom.	Start	End	Length (bp)	Nb. markers
RC1-C3.A01.1	A01	116 121	116 295	174	2
RC1-C3.A01.2	A01	205 917	244 342	38 425	5
RC1-C3.A01.3	A01	292 237	445 272	153 035	8
RC1-C3.A01.4	A01	1 433 103	1 442 065	8 962	3
RC1-C3.A01.5	A01	1 653 302	1 686 115	32 813	2
RC1-C3.A01.6	A01	2 402 978	2 407 419	4 441	3
<b>RC1-C3.A01.7</b>	<b>A01</b>	<b>5 719 114</b>	<b>5 757 919</b>	<b>38 805</b>	<b>4</b>
RC1-C3.A01.8	A01	6 210 234	6 215 140	4 906	2
RC1-C3.A01.9	A01	7 063 265	7 064 344	1 079	4
RC1-C3.A01.10	A01	7 795 389	7 796 010	621	2
RC1-C3.A01.11	A01	8 047 102	8 144 895	97 793	4
RC1-C3.A01.12	A01	8 284 174	8 334 735	50 561	4
RC1-C3.A01.13	A01	9 044 197	9 055 610	11 413	5
<b>RC1-C3.A01.14</b>	<b>A01</b>	<b>11 988 395</b>	<b>12 024 047</b>	<b>35 652</b>	<b>3</b>
<b>RC1-C3.A01.15</b>	<b>A01</b>	<b>12 096 164</b>	<b>12 133 862</b>	<b>37 698</b>	<b>3</b>
<b>RC1-C3.A01.16</b>	<b>A01</b>	<b>12 241 133</b>	<b>12 462 420</b>	<b>221 287</b>	<b>24</b>
<b>RC1-C3.A01.17</b>	<b>A01</b>	<b>12 717 798</b>	<b>12 734 857</b>	<b>17 059</b>	<b>3</b>
RC1-C3.A01.18	A01	12 765 580	12 769 625	4 045	2
<b>RC1-C3.A01.19</b>	<b>A01</b>	<b>12 811 143</b>	<b>12 883 239</b>	<b>72 096</b>	<b>4</b>
<b>RC1-C3.A01.20</b>	<b>A01</b>	<b>13 602 917</b>	<b>13 607 463</b>	<b>4 546</b>	<b>6</b>
RC1-C3.A01.21	A01	13 670 003	13 678 871	8 868	2
<b>RC1-C3.A01.22</b>	<b>A01</b>	<b>14 150 167</b>	<b>14 311 151</b>	<b>160 984</b>	<b>18</b>
RC1-C3.A01.23	A01	19 439 347	19 446 590	7 243	2
RC1-C3.A01.24	A01	19 460 294	19 468 702	8 408	3
RC1-C3.A01.25	A01	19 474 169	19 651 065	176 896	16
RC1-C3.A01.26	A01	19 733 379	19 748 018	14 639	2
RC1-C3.A01.27	A01	19 774 175	19 859 585	85 410	5
RC1-C3.A01.28	A01	19 876 888	19 968 639	91 751	7
RC1-C3.A01.29	A01	20 062 817	20 081 688	18 871	2
RC1-C3.A01.30	A01	20 105 686	20 116 439	10 753	3
RC1-C3.A01.31	A01	20 121 341	20 153 134	31 793	4
RC1-C3.A01.32	A01	20 325 235	20 410 477	85 242	7
RC1-C3.A01.33	A01	21 283 710	21 364 367	80 657	5
RC1-C3.A01.34	A01	21 376 322	21 378 067	1 745	2
RC1-C3.A01.35	A01	23 190 110	23 190 412	302	2
RC1-C3.A01.36	A01	23 460 699	23 461 927	1 228	2
RC1-C3.A01.37	A01	23 831 163	23 875 662	44 499	3
RC1-C3.A01.38	A01	25 345 500	25 345 527	27	2
RC1-C3.A01.39	A01	25 537 786	25 540 580	2 794	2
RC1-C3.A01.40	A01	26 810 610	27 002 061	191 451	6
RC1-C3.A01.41	A01	27 885 419	28 028 386	142 967	4
RC1-C3.A02.1	A02	1 605 346	1 621 791	16 445	2
RC1-C3.A02.2	A02	7 249 759	7 273 068	23 309	2
RC1-C3.A02.3	A02	13 510 021	13 514 430	4 409	2
RC1-C3.A02.4	A02	13 604 454	13 864 755	260 301	2
RC1-C3.A02.5	A02	14 074 280	14 138 270	63 990	3
RC1-C3.A03.1	A03	4 243 278	4 248 947	5 669	4
RC1-C3.A03.2	A03	5 443 631	5 496 941	53 310	2
RC1-C3.A03.3	A03	6 372 802	6 438 702	65 900	2
RC1-C3.A03.4	A03	7 100 429	7 100 750	321	2
<b>RC1-C3.A03.5</b>	<b>A03</b>	<b>7 364 838</b>	<b>7 390 133</b>	<b>25 295</b>	<b>5</b>
RC1-C3.A03.6	A03	7 478 359	7 478 598	239	2
RC1-C3.A03.7	A03	7 532 561	7 540 797	8 236	4
RC1-C3.A03.8	A03	7 543 954	7 550 505	6 551	2
RC1-C3.A03.9	A03	7 854 412	7 886 595	32 183	2
RC1-C3.A03.10	A03	8 171 909	8 208 929	37 020	7
RC1-C3.A03.11	A03	8 269 048	8 294 835	25 787	5
RC1-C3.A03.12	A03	8 540 258	8 556 617	16 359	3
RC1-C3.A03.13	A03	8 587 402	8 598 142	10 740	3
RC1-C3.A03.14	A03	8 643 714	8 660 698	16 984	3
RC1-C3.A03.15	A03	8 675 372	8 686 178	10 806	2
RC1-C3.A03.16	A03	9 149 994	9 158 234	8 240	3
RC1-C3.A03.17	A03	9 331 370	9 331 581	211	2

RC1-C3.A03.18	A03	9 373 654	9 383 070	9 416	2
RC1-C3.A03.19	A03	9 804 163	9 905 843	101 680	5
RC1-C3.A03.20	A03	9 972 745	9 977 710	4 965	2
RC1-C3.A03.21	A03	10 893 630	10 923 355	29 725	2
RC1-C3.A03.22	A03	11 409 233	11 420 483	11 250	2
RC1-C3.A03.23	A03	11 677 269	11 680 826	3 557	4
RC1-C3.A03.24	A03	16 224 287	16 240 081	15 794	2
RC1-C3.A03.25	A03	23 854 377	23 854 739	362	2
RC1-C3.A03.26	A03	25 962 430	25 986 044	23 614	3
RC1-C3.A03.27	A03	26 470 118	26 474 645	4 527	2
RC1-C3.A04.1	A04	326 297	354 705	28 408	4
RC1-C3.A04.2	A04	361 059	416 111	55 052	9
RC1-C3.A04.3	A04	425 292	435 393	10 101	3
RC1-C3.A04.4	A04	888 433	892 636	4 203	2
RC1-C3.A04.5	A04	897 959	902 735	4 776	2
RC1-C3.A04.6	A04	968 992	970 161	1 169	2
RC1-C3.A04.7	A04	1 148 550	1 165 851	17 301	4
RC1-C3.A04.8	A04	1 984 923	1 990 127	5 204	2
RC1-C3.A04.9	A04	2 116 487	2 117 431	944	3
RC1-C3.A04.10	A04	2 768 491	2 770 686	2 195	3
RC1-C3.A04.11	A04	2 777 404	2 802 283	24 879	6
RC1-C3.A04.12	A04	10 126 498	10 168 601	42 103	2
RC1-C3.A04.13	A04	10 310 520	10 331 747	21 227	2
RC1-C3.A04.14	A04	10 336 972	10 342 915	5 943	7
RC1-C3.A04.15	A04	11 488 873	11 488 959	86	2
RC1-C3.A04.16	A04	11 516 673	11 519 835	3 162	3
RC1-C3.A04.17	A04	12 797 448	12 860 499	63 051	4
RC1-C3.A04.18	A04	13 180 069	13 181 367	1 298	3
RC1-C3.A04.19	A04	13 230 464	13 470 718	240 254	38
RC1-C3.A04.20	A04	16 019 998	16 020 115	117	3
RC1-C3.A04.21	A04	20 713 264	20 786 083	72 819	3
RC1-C3.A05.1	A05	999 108	1 026 219	27 111	2
RC1-C3.A05.2	A05	1 115 996	1 138 639	22 643	2
RC1-C3.A05.3	A05	2 552 766	2 576 918	24 152	3
RC1-C3.A05.4	A05	3 422 976	3 429 329	6 353	2
RC1-C3.A05.5	A05	3 501 838	3 517 542	15 704	3
RC1-C3.A05.6	A05	3 733 337	3 733 447	110	2
RC1-C3.A05.7	A05	5 841 244	5 856 160	14 916	4
RC1-C3.A05.8	A05	10 908 072	10 920 163	12 091	2
RC1-C3.A05.9	A05	31 235 412	31 318 401	82 989	6
RC1-C3.A05.10	A05	31 383 666	31 414 354	30 688	3
RC1-C3.A06.1	A06	3 002 227	3 011 574	9 347	2
RC1-C3.A06.2	A06	3 873 816	3 925 559	51 743	7
RC1-C3.A06.3	A06	8 400 121	8 428 814	28 693	2
RC1-C3.A06.4	A06	8 550 111	8 556 741	6 630	3
RC1-C3.A06.5	A06	11 692 292	11 716 022	23 730	5
RC1-C3.A06.6	A06	35 034 612	35 044 704	10 092	2
RC1-C3.A06.7	A06	38 605 305	38 651 388	46 083	4
RC1-C3.A06.8	A06	38 654 625	38 656 565	1 940	3
RC1-C3.A06.9	A06	38 871 296	38 898 756	27 460	2
RC1-C3.A07.1	A07	285 399	390 135	104 736	2
RC1-C3.A07.2	A07	9 188 971	9 201 189	12 218	2
RC1-C3.A07.3	A07	11 513 416	11 513 663	247	2
RC1-C3.A07.4	A07	11 643 259	11 707 639	64 380	2
RC1-C3.A07.5	A07	11 750 310	11 756 629	6 319	2
RC1-C3.A07.6	A07	11 869 161	11 874 787	5 626	2
RC1-C3.A07.7	A07	12 312 656	12 317 200	4 544	3
RC1-C3.A07.8	A07	12 321 380	12 321 534	154	2
RC1-C3.A07.9	A07	12 410 256	12 412 673	2 417	2
RC1-C3.A07.10	A07	12 507 307	12 527 757	20 450	3
RC1-C3.A07.11	A07	12 568 625	12 574 545	5 920	2
RC1-C3.A07.12	A07	12 624 513	12 625 035	522	3
RC1-C3.A07.13	A07	12 795 227	13 073 580	278 353	2
RC1-C3.A07.14	A07	19 425 788	19 430 964	5 176	3
RC1-C3.A07.15	A07	19 479 083	19 486 383	7 300	2
RC1-C3.A08.1	A08	21 282 984	21 321 206	38 222	3
RC1-C3.A09.1	A09	13 034 215	13 186 671	152 456	13
RC1-C3.A09.2	A09	17 939 143	17 967 076	27 933	3
RC1-C3.A09.3	A09	18 000 353	18 027 881	27 528	5
RC1-C3.A09.4	A09	18 147 970	18 184 783	36 813	3

RC1-C3.A09.5	A09	18 313 518	18 423 119	109 601	8
RC1-C3.A09.6	A09	38 473 886	38 474 903	1 017	2
RC1-C3.A09.7	A09	38 475 586	38 510 533	34 947	8
RC1-C3.A09.8	A09	41 243 147	41 371 823	128 676	8
RC1-C3.A09.9	A09	43 328 257	43 333 478	5 221	2
RC1-C3.A09.10	A09	45 187 821	45 303 638	115 817	2
RC1-C3.A09.11	A09	46 040 242	46 084 316	44 074	2
RC1-C3.A09.12	A09	47 820 092	47 893 663	73 571	2
RC1-C3.A09.13	A09	50 050 811	50 081 052	30 241	2
RC1-C3.A10.1	A10	517 738	553 274	35 536	2
RC1-C3.A10.2	A10	909 903	927 433	17 530	2
RC1-C3.A10.3	A10	8 646 666	8 646 873	207	2
RC1-C3.A10.4	A10	11 318 428	11 343 362	24 934	3
RC1-C3.A10.5	A10	11 352 115	11 352 386	271	2
RC1-C3.A10.6	A10	11 725 237	11 746 192	20 955	7
RC1-C3.A10.7	A10	11 869 498	11 879 756	10 258	2
RC1-C3.A10.8	A10	12 058 901	12 134 970	76 069	4
RC1-C3.A10.9	A10	12 191 751	12 287 129	95 378	3
RC1-C3.A10.10	A10	12 950 612	12 974 524	23 912	3
RC1-C3.A10.11	A10	13 149 250	13 164 971	15 721	2
RC1-C3.A10.12	A10	13 263 231	13 280 392	17 161	3
RC1-C3.A10.13	A10	13 783 969	13 799 377	15 408	3
RC1-C3.C01.1	C01	1 431 037	1 438 290	7 253	2
RC1-C3.C01.2	C01	6 065 855	6 094 156	28 301	4
RC1-C3.C01.3	C01	7 140 157	7 149 324	9 167	4
RC1-C3.C01.4	C01	7 150 457	7 162 254	11 797	7
RC1-C3.C01.5	C01	7 167 275	7 169 855	2 580	5
RC1-C3.C01.6	C01	7 185 015	7 217 176	32 161	2
RC1-C3.C01.7	C01	7 217 904	7 360 712	142 808	9
RC1-C3.C01.8	C01	9 190 030	9 224 414	34 384	2
RC1-C3.C01.9	C01	30 508 577	30 521 630	13 053	2
RC1-C3.C01.10	C01	30 564 800	30 566 177	1 377	3
RC1-C3.C01.11	C01	30 597 524	30 611 745	14 221	9
RC1-C3.C02.1	C02	1 872 707	1 899 477	26 770	2
RC1-C3.C02.2	C02	5 206 023	5 211 840	5 817	2
RC1-C3.C02.3	C02	8 233 917	8 236 112	2 195	2
RC1-C3.C02.4	C02	10 985 046	11 806 029	820 983	97
RC1-C3.C02.5	C02	12 092 550	12 119 652	27 102	3
RC1-C3.C02.6	C02	12 151 031	12 201 460	50 429	7
RC1-C3.C02.7	C02	12 247 258	12 277 907	30 649	8
RC1-C3.C02.8	C02	13 139 898	13 530 584	390 686	2
RC1-C3.C02.9	C02	13 608 721	13 631 738	23 017	3
RC1-C3.C02.10	C02	15 223 938	15 224 226	288	2
RC1-C3.C02.11	C02	16 761 205	16 767 619	6 414	2
RC1-C3.C02.12	C02	44 669 097	44 795 106	126 009	2
RC1-C3.C02.13	C02	44 885 165	45 645 018	759 853	42
RC1-C3.C02.14	C02	45 662 441	45 690 048	27 607	2
RC1-C3.C02.15	C02	45 707 052	45 754 415	47 363	7
RC1-C3.C02.16	C02	45 813 714	45 813 869	155	3
RC1-C3.C02.17	C02	45 979 864	46 050 742	70 878	4
RC1-C3.C02.18	C02	48 596 859	48 666 206	69 347	11
RC1-C3.C02.19	C02	48 936 495	49 078 525	142 030	28
RC1-C3.C02.20	C02	50 411 294	50 412 687	1 393	2
RC1-C3.C02.21	C02	51 408 885	51 512 638	103 753	5
RC1-C3.C03.1	C03	8 637 745	8 673 704	35 959	3
RC1-C3.C03.2	C03	9 550 199	9 554 136	3 937	3
RC1-C3.C03.3	C03	9 656 507	9 802 015	145 508	10
RC1-C3.C03.4	C03	16 016 335	16 082 549	66 214	3
RC1-C3.C03.5	C03	19 403 087	19 468 381	65 294	4
RC1-C3.C03.6	C03	19 632 101	19 664 833	32 732	17
RC1-C3.C03.7	C03	40 850 427	40 850 484	57	3
RC1-C3.C03.8	C03	41 213 022	41 527 618	314 596	26
RC1-C3.C04.1	C04	4 954 549	4 963 887	9 338	7
RC1-C3.C04.2	C04	10 135 189	10 147 945	12 756	3
RC1-C3.C04.3	C04	43 795 702	43 927 184	131 482	3
RC1-C3.C04.4	C04	46 910 007	46 925 285	15 278	7
RC1-C3.C04.5	C04	49 959 583	49 977 036	17 453	4
RC1-C3.C04.6	C04	50 162 702	50 666 240	503 538	10
RC1-C3.C04.7	C04	52 142 780	52 172 839	30 059	7
RC1-C3.C04.8	C04	53 962 709	53 968 057	5 348	5

RC1-C3.C04.9	C04	54 396 038	54 454 173	58 135	6
RC1-C3.C04.10	C04	55 633 026	55 696 418	63 392	5
RC1-C3.C04.11	C04	57 405 229	57 528 221	122 992	3
RC1-C3.C04.12	C04	60 742 738	60 744 168	1 430	2
RC1-C3.C04.13	C04	60 755 376	60 755 752	376	3
RC1-C3.C04.14	C04	62 314 313	62 320 856	6 543	3
RC1-C3.C04.15	C04	62 409 193	62 409 880	687	3
RC1-C3.C04.16	C04	62 762 036	62 788 777	26 741	13
RC1-C3.C05.1	C05	14 085 619	14 096 463	10 844	11
RC1-C3.C05.2	C05	14 103 220	14 352 519	249 299	48
RC1-C3.C05.3	C05	15 086 200	15 286 619	200 419	15
RC1-C3.C05.4	C05	49 657 219	49 657 820	601	4
RC1-C3.C06.1	C06	18 635 923	18 691 129	55 206	2
RC1-C3.C06.2	C06	26 533 574	26 563 761	30 187	2
RC1-C3.C06.3	C06	26 761 223	26 761 505	282	4
RC1-C3.C06.4	C06	27 915 103	27 971 582	56 479	4
RC1-C3.C06.5	C06	28 009 500	28 090 392	80 892	11
RC1-C3.C06.6	C06	31 238 084	31 247 321	9 237	2
RC1-C3.C06.7	C06	31 260 879	31 275 213	14 334	3
RC1-C3.C06.8	C06	32 521 655	32 527 033	5 378	3
RC1-C3.C08.1	C08	1 850 361	1 852 539	2 178	2
RC1-C3.C08.2	C08	3 958 900	4 303 315	344 415	3
RC1-C3.C08.3	C08	33 008 990	33 011 585	2 595	3
RC1-C3.C08.4	C08	33 354 240	33 387 167	32 927	3
RC1-C3.C08.5	C08	33 521 827	33 522 128	301	2
RC1-C3.C08.6	C08	33 536 078	33 556 998	20 920	5
RC1-C3.C08.7	C08	33 841 026	33 842 063	1 037	2
RC1-C3.C08.8	C08	34 453 755	34 485 534	31 779	3
RC1-C3.C08.9	C08	34 884 304	34 890 898	6 594	2
RC1-C3.C08.10	C08	34 956 086	34 956 456	370	2
RC1-C3.C08.11	C08	37 612 059	37 677 820	65 761	2
RC1-C3.C09.1	C09	17 821 364	18 006 134	184 770	23
RC1-C3.C09.2	C09	18 014 261	18 185 661	171 400	8
RC1-C3.C09.3	C09	18 441 175	18 448 962	7 787	2
RC1-C3.C09.4	C09	24 448 258	24 558 950	110 692	8
RC1-C3.C09.5	C09	25 803 479	25 944 458	140 979	4
RC1-C3.C09.6	C09	49 891 764	49 893 746	1 982	2
RC1-C3.C09.7	C09	52 497 711	52 811 170	313 459	32
RC1-C3.C09.8	C09	52 969 900	53 039 489	69 589	17
RC1-C3.C09.9	C09	53 040 157	53 085 117	44 960	8
RC1-C3.C09.10	C09	55 033 741	55 037 648	3 907	2
RC1-C3.C09.11	C09	55 532 620	55 580 714	48 094	6
RC1-C3.C09.12	C09	56 935 803	57 072 021	136 218	7
RC1-C3.C09.13	C09	57 818 373	57 818 961	588	2
RC1-C3.C09.14	C09	58 998 195	59 007 215	9 020	2
RC1-C3.C09.15	C09	59 084 856	59 084 915	59	3
RC1-C3.C09.16	C09	59 090 771	59 220 621	129 850	7
RC1-C3.C09.17	C09	59 940 446	59 943 234	2 788	2

**Supplemental Table III.2.3. Genomic regions with high  $F_{ST}$  values between clusters of individuals**

Regions that colocalize with a QTL (refer to Supplemental Table III.2.4)

**High  $F_{ST}$  regions between clusters C1 and C2**

$F_{ST}$ region name	Chrom.	Start	End	Length (bp)	Nb. markers
RC1-C2.A01.1	A01	146 116	171 383	25 267	7
RC1-C2.A01.2	A01	5 263 946	5 330 507	66 561	5
RC1-C2.A01.3	A01	9 446 058	9 446 502	444	2
RC1-C2.A01.4	A01	21 735 710	21 736 893	1 183	2
RC1-C2.A01.5	A01	25 537 786	25 540 580	2 794	2
RC1-C2.A02.1	A02	1 796 615	1 832 139	35 524	2
RC1-C2.A02.2	A02	2 188 094	2 257 726	69 632	10
RC1-C2.A02.3	A02	2 312 274	2 360 362	48 088	8
RC1-C2.A02.4	A02	2 401 747	2 416 704	14 957	4
RC1-C2.A02.5	A02	2 436 949	2 482 200	45 251	4
RC1-C2.A02.6	A02	2 500 473	2 534 877	34 404	3
RC1-C2.A02.7	A02	2 582 124	2 595 223	13 099	2
RC1-C2.A02.8	A02	2 756 488	2 762 607	6 119	2
RC1-C2.A02.9	A02	5 435 861	5 585 365	149 504	17
RC1-C2.A02.10	A02	5 683 113	5 689 144	6 031	2
RC1-C2.A02.11	A02	6 173 427	6 179 125	5 698	6
RC1-C2.A02.12	A02	6 179 871	6 202 102	22 231	4
RC1-C2.A02.13	A02	6 253 822	6 332 189	78 367	2
RC1-C2.A02.14	A02	6 640 629	6 677 093	36 464	10
RC1-C2.A02.15	A02	6 788 065	6 789 235	1 170	2
RC1-C2.A02.16	A02	6 802 324	6 829 284	26 960	3
RC1-C2.A02.17	A02	7 305 421	7 347 120	41 699	4
RC1-C2.A02.18	A02	8 931 011	8 931 639	628	5
RC1-C2.A02.19	A02	9 070 379	9 166 650	96 271	9
RC1-C2.A02.20	A02	9 848 063	9 862 852	14 789	4
RC1-C2.A02.21	A02	9 936 677	9 938 632	1 955	2
RC1-C2.A02.22	A02	10 024 478	10 065 843	41 365	3
RC1-C2.A02.23	A02	10 398 619	10 400 487	1 868	4
RC1-C2.A02.24	A02	10 427 363	10 427 673	310	2
RC1-C2.A02.25	A02	10 484 678	10 489 663	4 985	3
RC1-C2.A02.26	A02	10 660 574	10 682 459	21 885	2
RC1-C2.A02.27	A02	10 685 656	10 702 339	16 683	3
RC1-C2.A02.28	A02	10 709 491	10 715 371	5 880	4
RC1-C2.A02.29	A02	10 796 510	10 800 245	3 735	3
RC1-C2.A02.30	A02	10 826 585	10 842 946	16 361	3
RC1-C2.A02.31	A02	11 155 637	11 156 259	622	3
RC1-C2.A02.32	A02	11 436 008	11 439 802	3 794	3
RC1-C2.A02.33	A02	11 538 079	11 538 241	162	2
RC1-C2.A02.34	A02	27 794 989	27 853 210	58 221	7
RC1-C2.A02.35	A02	29 673 819	29 691 110	17 291	2
RC1-C2.A02.36	A02	30 303 229	30 384 186	80 957	10
RC1-C2.A03.1	A03	6 808 067	6 808 638	571	2
RC1-C2.A03.2	A03	7 909 445	7 913 242	3 797	2
RC1-C2.A03.3	A03	9 886 171	9 905 843	19 672	3
RC1-C2.A03.4	A03	12 223 363	12 243 439	20 076	3
RC1-C2.A03.5	A03	13 404 313	13 406 148	1 835	2
RC1-C2.A03.6	A03	16 751 666	16 751 834	168	2
RC1-C2.A03.7	A03	22 518 972	22 526 882	7 910	2
RC1-C2.A03.8	A03	22 528 388	22 528 698	310	2
RC1-C2.A03.9	A03	22 554 551	22 555 362	811	2
RC1-C2.A03.10	A03	27 822 004	27 833 896	11 892	3
RC1-C2.A04.1	A04	10 511 443	10 512 274	831	2
RC1-C2.A04.2	A04	12 682 103	12 682 362	259	2
RC1-C2.A04.3	A04	12 797 448	12 797 767	319	2
RC1-C2.A04.4	A04	15 911 146	16 020 115	108 969	7
RC1-C2.A05.1	A05	3 422 976	3 429 329	6 353	2
RC1-C2.A05.2	A05	6 108 216	6 111 107	2 891	2
RC1-C2.A06.1	A06	3 448 074	3 450 886	2 812	3
RC1-C2.A06.2	A06	4 504 299	4 564 864	60 565	11
RC1-C2.A06.3	A06	5 678 904	5 681 981	3 077	2
RC1-C2.A06.4	A06	6 366 611	6 405 547	38 936	4
RC1-C2.A06.5	A06	6 419 084	6 420 225	1 141	2
RC1-C2.A06.6	A06	6 578 618	6 659 483	80 865	23

RC1-C2.A06.7	A06	6 936 189	6 948 490	12 301	4
RC1-C2.A06.8	A06	6 990 916	7 024 195	33 279	5
RC1-C2.A06.9	A06	7 863 337	8 042 499	179 162	2
RC1-C2.A06.10	A06	8 550 111	8 556 741	6 630	3
RC1-C2.A06.11	A06	8 778 837	9 077 544	298 707	21
RC1-C2.A06.12	A06	9 354 787	9 423 970	69 183	4
RC1-C2.A06.13	A06	10 255 404	10 277 119	21 715	2
RC1-C2.A06.14	A06	10 346 369	10 376 635	30 266	3
RC1-C2.A06.15	A06	11 699 138	11 708 633	9 495	3
RC1-C2.A07.1	A07	75 970	96 484	20 514	4
RC1-C2.A07.2	A07	117 910	133 065	15 155	2
RC1-C2.A07.3	A07	1 281 464	1 283 185	1 721	4
RC1-C2.A07.4	A07	12 568 625	12 590 973	22 348	3
RC1-C2.A07.5	A07	12 621 971	12 643 758	21 787	15
RC1-C2.A07.6	A07	12 662 165	12 717 959	55 794	11
RC1-C2.A07.7	A07	20 864 868	20 914 706	49 838	5
RC1-C2.A07.8	A07	21 598 892	21 599 869	977	2
RC1-C2.A07.9	A07	24 478 145	24 483 851	5 706	2
RC1-C2.A07.10	A07	24 605 750	24 612 555	6 805	4
RC1-C2.A07.11	A07	26 567 360	26 584 059	16 699	4
RC1-C2.A09.1	A09	11 712 829	11 713 182	353	2
RC1-C2.A09.2	A09	12 276 194	12 333 573	57 379	8
RC1-C2.A09.3	A09	17 939 143	17 967 076	27 933	3
RC1-C2.A09.4	A09	18 000 353	18 027 881	27 528	5
RC1-C2.A09.5	A09	18 068 834	18 660 713	591 879	49
RC1-C2.A09.6	A09	19 261 092	19 277 377	16 285	3
RC1-C2.A09.7	A09	19 351 674	19 379 009	27 335	3
RC1-C2.A09.8	A09	19 540 191	19 925 009	384 818	20
RC1-C2.A09.9	A09	19 984 437	20 067 171	82 734	3
RC1-C2.A09.10	A09	20 108 266	20 418 636	310 370	6
RC1-C2.A09.11	A09	20 599 736	20 904 154	304 418	10
RC1-C2.A09.12	A09	20 920 740	21 034 856	114 116	3
RC1-C2.A09.13	A09	21 053 732	21 171 038	117 306	6
RC1-C2.A09.14	A09	31 106 388	31 579 845	473 457	3
RC1-C2.A09.15	A09	31 845 981	32 008 783	162 802	2
RC1-C2.A09.16	A09	32 084 647	32 110 961	26 314	2
RC1-C2.A09.17	A09	32 348 113	32 776 988	428 875	47
RC1-C2.A09.18	A09	32 796 034	32 943 281	147 247	36
RC1-C2.A09.19	A09	32 965 813	32 968 201	2 388	2
RC1-C2.A09.20	A09	32 992 230	33 003 011	10 781	9
RC1-C2.A09.21	A09	33 004 654	33 018 243	13 589	4
RC1-C2.A09.22	A09	33 018 544	33 020 114	1 570	6
RC1-C2.A09.23	A09	33 021 186	33 396 146	374 960	73
RC1-C2.A09.24	A09	33 402 276	33 404 590	2 314	3
RC1-C2.A09.25	A09	33 536 318	33 658 163	121 845	4
RC1-C2.A09.26	A09	38 473 886	38 474 903	1 017	2
RC1-C2.A09.27	A09	38 475 586	38 510 533	34 947	8
RC1-C2.A09.28	A09	38 600 579	38 641 008	40 429	4
RC1-C2.A09.29	A09	39 014 461	39 145 169	130 708	11
RC1-C2.A09.30	A09	46 685 856	46 685 925	69	2
RC1-C2.A09.31	A09	46 756 320	46 791 548	35 228	5
RC1-C2.A09.32	A09	46 838 236	46 838 294	58	2
RC1-C2.A09.33	A09	47 751 356	47 760 250	8 894	2
RC1-C2.A09.34	A09	48 924 814	48 943 818	19 004	2
RC1-C2.A09.35	A09	53 273 978	53 428 722	154 744	4
RC1-C2.A09.36	A09	53 476 088	53 489 369	13 281	2
RC1-C2.A10.1	A10	107 827	145 449	37 622	3
RC1-C2.A10.2	A10	366 688	368 208	1 520	3
RC1-C2.A10.3	A10	503 502	553 274	49 772	8
RC1-C2.A10.4	A10	946 180	946 195	15	2
RC1-C2.A10.5	A10	1 171 208	1 222 224	51 016	5
RC1-C2.A10.6	A10	1 408 889	1 446 848	37 959	9
RC1-C2.A10.7	A10	1 471 822	1 494 235	22 413	4
RC1-C2.A10.8	A10	4 808 524	4 815 454	6 930	3
RC1-C2.A10.9	A10	4 878 246	4 892 232	13 986	2
RC1-C2.A10.10	A10	4 937 396	4 937 405	9	2
RC1-C2.A10.11	A10	4 938 912	4 939 478	566	4
RC1-C2.A10.12	A10	4 973 649	4 981 328	7 679	2
RC1-C2.A10.13	A10	4 982 072	4 982 177	105	2
RC1-C2.A10.14	A10	5 037 679	5 046 114	8 435	6

RC1-C2.A10.15	A10	5 056 148	5 060 185	4 037	2
RC1-C2.A10.16	A10	5 067 199	5 086 271	19 072	3
RC1-C2.A10.17	A10	5 097 235	5 128 348	31 113	6
RC1-C2.A10.18	A10	7 227 874	7 263 535	35 661	3
RC1-C2.A10.19	A10	8 415 173	8 424 758	9 585	2
RC1-C2.A10.20	A10	8 431 391	8 509 376	77 985	10
RC1-C2.A10.21	A10	8 639 673	8 644 767	5 094	2
RC1-C2.A10.22	A10	8 646 666	8 646 873	207	2
RC1-C2.A10.23	A10	11 318 428	11 352 386	33 958	6
RC1-C2.A10.24	A10	11 594 904	11 604 377	9 473	3
RC1-C2.A10.25	A10	12 950 612	12 971 301	20 689	2
RC1-C2.A10.26	A10	13 395 240	13 409 341	14 101	2
RC1-C2.A10.27	A10	16 877 102	16 881 208	4 106	3
RC1-C2.A10.28	A10	17 710 661	17 716 722	6 061	2
RC1-C2.A10.29	A10	18 000 076	18 062 247	62 171	9
RC1-C2.A10.30	A10	18 113 489	18 113 849	360	2
RC1-C2.A10.31	A10	18 185 727	18 291 833	106 106	8
RC1-C2.A10.32	A10	18 317 070	18 318 279	1 209	2
RC1-C2.A10.33	A10	18 327 816	18 337 894	10 078	2
RC1-C2.A10.34	A10	18 551 526	18 560 050	8 524	3
RC1-C2.A10.35	A10	19 117 467	19 122 731	5 264	3
RC1-C2.A10.36	A10	19 166 226	19 222 747	56 521	2
RC1-C2.A10.37	A10	19 368 573	19 371 442	2 869	2
RC1-C2.C01.1	C01	1 949 032	1 951 048	2 016	2
RC1-C2.C01.2	C01	2 468 038	2 471 948	3 910	4
RC1-C2.C01.3	C01	30 738 063	30 738 401	338	2
RC1-C2.C01.4	C01	30 966 349	30 967 658	1 309	3
RC1-C2.C01.5	C01	31 806 767	31 968 109	161 342	8
RC1-C2.C01.6	C01	32 345 030	32 511 008	165 978	14
RC1-C2.C01.7	C01	32 590 291	32 701 901	111 610	6
RC1-C2.C02.1	C02	6 864 635	6 865 189	554	2
RC1-C2.C02.2	C02	10 985 046	11 954 882	969 836	106
RC1-C2.C02.3	C02	12 249 093	12 249 329	236	2
RC1-C2.C02.4	C02	15 090 176	15 101 508	11 332	2
RC1-C2.C02.5	C02	31 142 416	31 167 548	25 132	3
RC1-C2.C02.6	C02	31 202 263	31 208 620	6 357	4
RC1-C2.C02.7	C02	31 897 992	32 067 756	169 764	4
RC1-C2.C02.8	C02	44 611 792	44 795 106	183 314	4
RC1-C2.C02.9	C02	44 885 165	45 690 048	804 883	46
RC1-C2.C02.10	C02	45 707 052	46 011 253	304 201	17
RC1-C2.C02.11	C02	48 596 859	48 616 616	19 757	2
RC1-C2.C02.12	C02	48 936 495	49 074 264	137 769	27
RC1-C2.C02.13	C02	50 411 294	50 412 687	1 393	2
RC1-C2.C02.14	C02	51 235 428	51 512 638	277 210	13
RC1-C2.C03.1	C03	24 393 084	24 396 880	3 796	2
RC1-C2.C03.2	C03	42 166 731	42 167 689	958	2
RC1-C2.C04.1	C04	3 481 387	3 597 753	116 366	2
RC1-C2.C04.2	C04	3 803 963	3 804 119	156	2
RC1-C2.C04.3	C04	6 053 683	6 074 102	20 419	2
RC1-C2.C04.4	C04	6 146 824	6 148 202	1 378	3
RC1-C2.C04.5	C04	15 907 159	16 084 715	177 556	10
RC1-C2.C04.6	C04	52 142 780	52 159 079	16 299	6
RC1-C2.C05.1	C05	244 401	284 926	40 525	8
RC1-C2.C05.2	C05	2 427 632	2 428 210	578	3
RC1-C2.C05.3	C05	2 448 968	2 450 415	1 447	2
RC1-C2.C05.4	C05	46 441 522	46 441 776	254	2
RC1-C2.C05.5	C05	52 804 512	52 830 735	26 223	4
RC1-C2.C05.6	C05	53 033 107	53 049 272	16 165	2
RC1-C2.C05.7	C05	53 083 920	53 111 088	27 168	3
RC1-C2.C05.8	C05	53 190 696	53 201 624	10 928	3
RC1-C2.C06.1	C06	5 272 320	5 283 338	11 018	2
RC1-C2.C06.2	C06	5 296 613	5 311 581	14 968	5
RC1-C2.C06.3	C06	5 358 479	5 430 648	72 169	19
RC1-C2.C06.4	C06	27 915 103	28 090 392	175 289	16
RC1-C2.C06.5	C06	30 471 040	30 480 127	9 087	8
RC1-C2.C06.6	C06	31 281 167	31 281 245	78	2
RC1-C2.C06.7	C06	31 334 190	31 342 942	8 752	3
RC1-C2.C06.8	C06	33 620 565	33 664 027	43 462	7
RC1-C2.C06.9	C06	33 753 126	33 783 694	30 568	4
RC1-C2.C06.10	C06	35 369 839	35 372 432	2 593	2

RC1-C2.C06.11	C06	37 813 624	38 064 014	250 390	13
RC1-C2.C06.12	C06	43 208 233	43 359 395	151 162	5
RC1-C2.C06.13	C06	43 743 867	43 746 422	2 555	3
RC1-C2.C06.14	C06	44 823 028	44 861 825	38 797	9
RC1-C2.C06.15	C06	47 111 453	47 166 691	55 238	4
RC1-C2.C06.16	C06	47 397 767	47 397 885	118	2
RC1-C2.C06.17	C06	47 408 754	47 414 522	5 768	6
RC1-C2.C06.18	C06	48 083 591	48 206 527	122 936	5
RC1-C2.C08.1	C08	18 292 443	18 293 793	1 350	2
RC1-C2.C08.2	C08	34 145 154	34 156 051	10 897	4
RC1-C2.C08.3	C08	41 575 900	41 580 345	4 445	2
RC1-C2.C09.1	C09	17 778 678	18 298 727	520 049	37
RC1-C2.C09.2	C09	21 484 821	21 680 344	195 523	7
RC1-C2.C09.3	C09	52 497 711	52 811 170	313 459	32
RC1-C2.C09.4	C09	52 969 900	53 022 957	53 057	9
RC1-C2.C09.5	C09	55 033 741	55 037 648	3 907	2
RC1-C2.C09.6	C09	57 071 253	57 072 021	768	2
RC1-C2.C09.7	C09	62 544 378	62 610 869	66 491	5
RC1-C2.C09.8	C09	62 807 775	62 940 380	132 605	2
RC1-C2.C09.9	C09	63 045 186	63 297 767	252 581	18
RC1-C2.C09.10	C09	64 191 522	64 207 793	16 271	3
RC1-C2.C09.11	C09	64 701 910	64 764 580	62 670	3
RC1-C2.C09.12	C09	65 080 013	65 352 420	272 407	10

**Supplemental Table III.2.3. Genomic regions with high  $F_{ST}$  values between clusters of individuals**

Regions that colocalize with a QTL (refer to Supplemental Table III.2.4)

**High  $F_{ST}$  regions between clusters C2 and C3**

$F_{ST}$ region name	Chrom.	Start	End	Length (bp)	Nb. markers
RC2-C3.A01.1	A01	292 237	314 776	22 539	3
RC2-C3.A01.2	A01	658 745	695 649	36 904	3
RC2-C3.A01.3	A01	20 618 161	20 624 574	6 413	8
RC2-C3.A01.4	A01	26 560 192	26 562 766	2 574	3
RC2-C3.A02.1	A02	5 640 344	5 648 058	7 714	2
RC2-C3.A02.2	A02	5 958 214	6 002 434	44 220	3
RC2-C3.A02.3	A02	6 050 634	6 079 277	28 643	2
RC2-C3.A02.4	A02	6 173 427	6 178 906	5 479	4
RC2-C3.A02.5	A02	6 640 629	6 677 093	36 464	10
RC2-C3.A02.6	A02	6 708 038	6 737 138	29 100	2
RC2-C3.A02.7	A02	8 931 619	8 931 639	20	2
RC2-C3.A02.8	A02	9 070 379	9 166 650	96 271	9
RC2-C3.A02.9	A02	9 707 705	9 711 907	4 202	2
RC2-C3.A02.10	A02	10 030 183	10 065 843	35 660	2
RC2-C3.A02.11	A02	10 098 602	10 105 480	6 878	2
RC2-C3.A02.12	A02	10 126 956	10 208 911	81 955	12
RC2-C3.A02.13	A02	10 250 975	10 312 263	61 288	7
RC2-C3.A02.14	A02	10 392 483	10 400 487	8 004	6
RC2-C3.A02.15	A02	10 709 491	10 724 654	15 163	5
RC2-C3.A02.16	A02	10 734 266	10 800 245	65 979	5
RC2-C3.A02.17	A02	10 826 585	10 830 794	4 209	2
RC2-C3.A02.18	A02	10 866 160	10 891 415	25 255	9
RC2-C3.A02.19	A02	10 921 608	10 928 568	6 960	2
RC2-C3.A02.20	A02	10 961 133	10 963 293	2 160	2
RC2-C3.A02.21	A02	10 995 909	11 005 346	9 437	2
RC2-C3.A02.22	A02	11 155 637	11 156 259	622	3
RC2-C3.A02.23	A02	11 436 008	11 439 802	3 794	3
RC2-C3.A02.24	A02	11 579 374	11 597 483	18 109	3
RC2-C3.A02.25	A02	11 632 384	11 635 241	2 857	3
RC2-C3.A02.26	A02	26 919 053	26 943 805	24 752	4
RC2-C3.A02.27	A02	31 160 099	31 266 824	106 725	5
RC2-C3.A02.28	A02	31 386 861	31 387 747	886	2
RC2-C3.A03.1	A03	8 106 918	8 122 347	15 429	2
RC2-C3.A03.2	A03	8 556 395	8 556 617	222	2
RC2-C3.A03.3	A03	10 893 630	10 923 355	29 725	2
RC2-C3.A03.4	A03	12 243 018	12 243 439	421	2
RC2-C3.A03.5	A03	13 136 947	13 145 600	8 653	3
RC2-C3.A03.6	A03	13 444 172	13 448 001	3 829	2
RC2-C3.A03.7	A03	16 751 666	16 751 834	168	2
RC2-C3.A03.8	A03	20 170 995	20 171 841	846	2
RC2-C3.A03.9	A03	20 887 616	20 888 381	765	2
RC2-C3.A03.10	A03	22 518 972	22 526 882	7 910	2
RC2-C3.A03.11	A03	22 528 388	22 528 698	310	2
RC2-C3.A03.12	A03	22 554 551	22 555 362	811	2
RC2-C3.A04.1	A04	3 144 227	3 144 917	690	2
RC2-C3.A04.2	A04	14 742 598	14 776 256	33 658	2
RC2-C3.A04.3	A04	14 954 474	14 960 748	6 274	3
RC2-C3.A04.4	A04	15 179 769	15 210 076	30 307	2
RC2-C3.A05.1	A05	999 108	1 026 219	27 111	2
RC2-C3.A05.2	A05	4 370 891	4 381 828	10 937	3
RC2-C3.A05.3	A05	4 428 105	4 429 313	1 208	2
RC2-C3.A05.4	A05	5 839 690	5 842 170	2 480	4
RC2-C3.A05.5	A05	5 962 539	5 963 329	790	2
RC2-C3.A05.6	A05	9 057 428	9 065 886	8 458	3
RC2-C3.A05.7	A05	10 835 170	10 842 622	7 452	2
RC2-C3.A05.8	A05	10 908 072	10 932 515	24 443	3
RC2-C3.A05.9	A05	30 140 705	30 159 883	19 178	4
RC2-C3.A05.10	A05	30 497 116	30 570 747	73 631	5
RC2-C3.A05.11	A05	35 547 585	35 547 705	120	2
RC2-C3.A05.12	A05	35 697 588	35 717 751	20 163	3
RC2-C3.A06.1	A06	3 836 391	3 838 172	1 781	2
RC2-C3.A06.2	A06	3 914 472	3 925 559	11 087	4
RC2-C3.A06.3	A06	4 670 648	4 681 090	10 442	2

RC2-C3.A06.4	A06	4 712 795	4 717 555	4 760	2
RC2-C3.A07.1	A07	3 191 628	3 193 815	2 187	2
RC2-C3.A07.2	A07	3 205 253	3 205 788	535	3
RC2-C3.A07.3	A07	3 206 298	3 206 658	360	2
RC2-C3.A07.4	A07	11 756 629	11 790 383	33 754	2
RC2-C3.A07.5	A07	14 855 414	14 862 881	7 467	2
RC2-C3.A07.6	A07	22 261 040	22 265 993	4 953	3
RC2-C3.A07.7	A07	22 327 321	22 339 665	12 344	4
RC2-C3.A08.1	A08	1 146 179	1 147 095	916	2
RC2-C3.A08.2	A08	2 677 766	2 759 126	81 360	4
RC2-C3.A08.3	A08	2 762 611	2 774 775	12 164	2
RC2-C3.A08.4	A08	14 051 621	14 091 394	39 773	3
RC2-C3.A08.5	A08	15 516 776	16 397 195	880 419	4
RC2-C3.A08.6	A08	17 568 404	17 576 031	7 627	3
RC2-C3.A08.7	A08	21 237 259	21 244 191	6 932	2
RC2-C3.A08.8	A08	21 399 895	21 407 953	8 058	5
RC2-C3.A08.9	A08	21 487 786	21 489 774	1 988	2
RC2-C3.A08.10	A08	21 555 103	21 556 581	1 478	2
RC2-C3.A08.11	A08	21 727 600	21 744 126	16 526	2
RC2-C3.A08.12	A08	22 004 856	22 015 702	10 846	2
RC2-C3.A09.1	A09	1 981 843	2 083 462	101 619	2
RC2-C3.A09.2	A09	2 545 099	2 555 497	10 398	2
RC2-C3.A09.3	A09	3 613 407	3 614 289	882	2
RC2-C3.A09.4	A09	10 535 751	10 550 133	14 382	7
RC2-C3.A09.5	A09	17 631 010	17 783 301	152 291	3
RC2-C3.A09.6	A09	18 187 721	18 233 732	46 011	5
RC2-C3.A09.7	A09	18 252 156	18 269 931	17 775	6
RC2-C3.A09.8	A09	18 498 215	18 621 232	123 017	15
RC2-C3.A09.9	A09	18 640 169	18 644 227	4 058	3
RC2-C3.A09.10	A09	19 261 092	19 266 918	5 826	2
RC2-C3.A09.11	A09	19 351 674	19 379 009	27 335	3
RC2-C3.A09.12	A09	19 540 191	19 925 009	384 818	20
RC2-C3.A09.13	A09	19 984 437	20 067 171	82 734	3
RC2-C3.A09.14	A09	20 108 266	20 467 794	359 528	7
RC2-C3.A09.15	A09	20 599 736	20 904 154	304 418	10
RC2-C3.A09.16	A09	20 920 740	21 171 038	250 298	10
RC2-C3.A09.17	A09	31 106 388	31 579 845	473 457	3
RC2-C3.A09.18	A09	31 845 981	32 008 783	162 802	2
RC2-C3.A09.19	A09	32 084 647	32 110 961	26 314	2
RC2-C3.A09.20	A09	32 348 113	32 501 514	153 401	19
RC2-C3.A09.21	A09	32 897 126	32 902 753	5 627	4
RC2-C3.A09.22	A09	32 903 427	32 923 005	19 578	7
RC2-C3.A09.23	A09	33 166 199	33 191 939	25 740	3
RC2-C3.A09.24	A09	33 219 434	33 220 930	1 496	5
RC2-C3.A09.25	A09	33 248 979	33 249 858	879	2
RC2-C3.A09.26	A09	33 284 573	33 286 497	1 924	2
RC2-C3.A09.27	A09	33 302 442	33 305 016	2 574	7
RC2-C3.A09.28	A09	33 353 059	33 366 100	13 041	4
RC2-C3.A09.29	A09	33 404 590	33 435 117	30 527	2
RC2-C3.A09.30	A09	34 259 006	34 315 036	56 030	5
RC2-C3.A09.31	A09	48 819 794	48 820 822	1 028	2
RC2-C3.A10.1	A10	3 136 292	3 138 952	2 660	3
RC2-C3.A10.2	A10	16 316 332	16 320 158	3 826	3
RC2-C3.A10.3	A10	18 185 727	18 203 465	17 738	3
RC2-C3.A10.4	A10	18 446 688	18 457 376	10 688	2
RC2-C3.A10.5	A10	18 742 522	18 744 799	2 277	2
RC2-C3.C01.1	C01	606 243	630 609	24 366	2
RC2-C3.C01.2	C01	681 801	698 556	16 755	6
RC2-C3.C01.3	C01	1 949 032	1 951 048	2 016	2
RC2-C3.C01.4	C01	2 208 981	2 230 874	21 893	5
RC2-C3.C01.5	C01	2 266 185	2 268 466	2 281	2
RC2-C3.C01.6	C01	2 337 718	2 471 948	134 230	5
RC2-C3.C01.7	C01	3 328 265	3 348 862	20 597	7
RC2-C3.C01.8	C01	5 489 909	5 494 518	4 609	2
RC2-C3.C01.9	C01	7 334 864	7 360 712	25 848	2
RC2-C3.C01.10	C01	7 432 391	7 439 759	7 368	4
RC2-C3.C01.11	C01	13 399 933	13 412 276	12 343	3
RC2-C3.C01.12	C01	13 497 074	13 654 018	156 944	2
RC2-C3.C01.13	C01	14 318 625	14 553 072	234 447	3
RC2-C3.C01.14	C01	20 449 407	20 458 883	9 476	2

RC2-C3.C01.15	C01	32 510 830	32 511 008	178	2
RC2-C3.C01.16	C01	32 590 291	32 701 901	111 610	6
RC2-C3.C01.17	C01	35 732 368	36 289 939	557 571	25
RC2-C3.C01.18	C01	36 398 229	36 442 365	44 136	4
RC2-C3.C01.19	C01	38 124 781	38 126 804	2 023	2
RC2-C3.C01.20	C01	38 161 950	38 202 529	40 579	8
RC2-C3.C01.21	C01	38 263 325	38 273 969	10 644	16
RC2-C3.C01.22	C01	38 308 879	38 315 097	6 218	3
RC2-C3.C01.23	C01	38 334 734	38 536 835	202 101	16
RC2-C3.C01.24	C01	40 999 195	41 105 487	106 292	8
RC2-C3.C01.25	C01	41 415 441	41 419 202	3 761	2
RC2-C3.C01.26	C01	43 987 355	43 988 494	1 139	3
RC2-C3.C01.27	C01	43 989 349	43 991 590	2 241	4
RC2-C3.C01.28	C01	44 003 009	44 030 639	27 630	2
RC2-C3.C02.1	C02	2 704 589	3 149 796	445 207	6
RC2-C3.C02.2	C02	4 296 110	4 296 864	754	2
RC2-C3.C02.3	C02	4 482 553	4 496 009	13 456	3
RC2-C3.C02.4	C02	53 296 152	53 308 558	12 406	6
RC2-C3.C02.5	C02	53 322 550	53 322 962	412	2
RC2-C3.C02.6	C02	56 922 575	56 931 829	9 254	4
RC2-C3.C02.7	C02	58 533 736	58 537 749	4 013	4
RC2-C3.C02.8	C02	60 891 724	60 916 853	25 129	15
RC2-C3.C03.1	C03	9 283 561	9 304 087	20 526	20
RC2-C3.C03.2	C03	17 053 240	17 057 881	4 641	3
RC2-C3.C03.3	C03	25 810 107	25 812 914	2 807	3
RC2-C3.C03.4	C03	63 857 190	64 335 505	478 315	85
RC2-C3.C03.5	C03	64 344 536	64 756 415	411 879	39
RC2-C3.C03.6	C03	65 114 960	65 126 983	12 023	4
RC2-C3.C03.7	C03	65 805 815	65 806 028	213	2
RC2-C3.C03.8	C03	70 134 519	70 204 586	70 067	7
RC2-C3.C03.9	C03	70 242 969	70 251 441	8 472	7
RC2-C3.C03.10	C03	70 278 360	70 278 885	525	4
RC2-C3.C03.11	C03	70 280 974	70 281 137	163	2
RC2-C3.C03.12	C03	70 302 176	70 402 144	99 968	2
RC2-C3.C03.13	C03	71 127 744	71 142 532	14 788	10
RC2-C3.C03.14	C03	73 563 670	73 564 145	475	3
RC2-C3.C04.1	C04	4 072 885	4 121 136	48 251	2
RC2-C3.C04.2	C04	6 074 102	6 148 202	74 100	5
RC2-C3.C04.3	C04	41 220 261	41 234 422	14 161	10
RC2-C3.C04.4	C04	43 631 996	43 927 184	295 188	15
RC2-C3.C04.5	C04	57 521 668	57 528 221	6 553	2
RC2-C3.C05.1	C05	12 924 291	12 932 523	8 232	2
RC2-C3.C05.2	C05	13 103 176	13 347 673	244 497	47
RC2-C3.C05.3	C05	13 374 224	13 763 119	388 895	66
RC2-C3.C05.4	C05	14 912 358	15 130 531	218 173	31
RC2-C3.C05.5	C05	49 657 219	49 657 820	601	4
RC2-C3.C05.6	C05	53 659 762	53 661 793	2 031	2
RC2-C3.C06.1	C06	6 253 491	6 274 652	21 161	2
RC2-C3.C06.2	C06	26 533 574	26 563 761	30 187	2
RC2-C3.C06.3	C06	26 761 223	26 778 659	17 436	6
RC2-C3.C06.4	C06	27 005 183	27 056 615	51 432	3
RC2-C3.C06.5	C06	27 350 962	27 598 557	247 595	4
RC2-C3.C06.6	C06	37 995 784	37 996 398	614	2
RC2-C3.C06.7	C06	38 035 252	38 064 014	28 762	2
RC2-C3.C07.1	C07	41 221 850	41 251 227	29 377	5
RC2-C3.C07.2	C07	49 035 350	49 278 981	243 631	16
RC2-C3.C07.3	C07	51 175 068	51 179 792	4 724	3
RC2-C3.C07.4	C07	51 536 163	51 536 311	148	2
RC2-C3.C07.5	C07	51 538 520	51 576 393	37 873	16
RC2-C3.C07.6	C07	51 780 235	51 791 173	10 938	4
RC2-C3.C07.7	C07	52 050 990	52 077 105	26 115	2
RC2-C3.C08.1	C08	33 249 618	33 259 488	9 870	3
RC2-C3.C08.2	C08	34 109 415	34 156 051	46 636	22
RC2-C3.C09.1	C09	20 164 626	20 259 165	94 539	2
RC2-C3.C09.2	C09	20 322 413	20 345 716	23 303	10
RC2-C3.C09.3	C09	20 345 974	20 441 114	95 140	18
RC2-C3.C09.4	C09	20 444 329	20 507 949	63 620	20
RC2-C3.C09.5	C09	20 575 098	20 575 103	5	2
RC2-C3.C09.6	C09	21 086 971	21 427 944	340 973	39
RC2-C3.C09.7	C09	21 456 984	21 730 202	273 218	11

RC2-C3.C09.8	C09	21 936 939	23 011 173	1 074 234	44
RC2-C3.C09.9	C09	23 040 259	23 115 335	75 076	7
RC2-C3.C09.10	C09	23 548 229	23 589 882	41 653	3
RC2-C3.C09.11	C09	24 274 900	24 713 638	438 738	11
RC2-C3.C09.12	C09	25 308 501	25 944 458	635 957	6
RC2-C3.C09.13	C09	48 195 146	48 293 315	98 169	16
RC2-C3.C09.14	C09	57 818 373	57 818 961	588	2
RC2-C3.C09.15	C09	58 998 195	59 007 215	9 020	2
RC2-C3.C09.16	C09	59 084 856	59 084 915	59	3
RC2-C3.C09.17	C09	59 090 771	59 094 582	3 811	4
RC2-C3.C09.18	C09	62 544 378	62 610 869	66 491	5
RC2-C3.C09.19	C09	62 807 775	62 940 380	132 605	5
RC2-C3.C09.20	C09	63 045 186	63 297 767	252 581	18
RC2-C3.C09.21	C09	64 191 522	64 207 793	16 271	3
RC2-C3.C09.22	C09	65 080 013	65 080 586	573	2

**Supplemental Table III.2.4. Overview of the 38 loci controlling early vigor traits identified by GWAS in the population of 233 rapeseed lines**

Peak SNP	Chr.	Trait	Position (pb)	Position flanking marker 1	Position flanking marker 2	QTL size (pb)	Major/minor allele	MAF	SNP weight	Favorable allele	R <sup>2</sup>	QTL	Colocalization with F <sub>ST</sub>
Bn_A01_p5318059	A01	NtE <sub>TOT</sub>	5 611 458	5 044 830	6 294 187	1 249 357	G/T	0.18	0.27	T	4.68	QTL.A01.1	RC1-C2 A01.2; RC1-C3 A01.7
Bn_A01_p5697706	A01	RDW	6 018 027	5 305 247	6 648 795	1 343 548	T/C	0.22	0.03	C	6.39	QTL.A01.1	RC1-C3 A01.4; RC1-C3 A01.16; C3 A01.15; RC1-C3 A01.16; RC1-C2 A01 17. RC1-
Bn_A01_p15233095	A01	NtE <sub>TOT</sub>	13 866 158	11 866 490	15 473 155	3 606 665	T/A	0.20	-0.30	T	4.31	QTL.A01.2	RC1-C2 A01.4; RC1-C2 A01.5; C3 A01.15; RC1-C3 A01.16; RC1-C2 A01 17. RC1-
Bn_A01_p15179106	A01	NtE <sub>AER</sub>	13 920 427	11 982 963	15 473 155	3 490 192	A/G	0.20	-0.27	A	4.24	QTL.A01.2	RC1-C2 A02.4; RC1-C2 A02.5; RC1-C2 A02.6
Bn_A02_p3160461	A02	AUC <sub>L1-L4</sub>	2 205 597	267 875	3 398 658	3 398 658	T/G	0.41	-766.81	T	11.56	QTL.A02.1	RC1-C2 A02.4; RC1-C2 A02.5; RC1-C2 A02.6
Bn_A02_p3160461	A02	AUC <sub>L2-L3</sub>	2 205 597	267 875	3 398 658	3 398 658	T/G	0.41	-233.98	T	10.24	QTL.A02.1	RC1-C2 A02.4; RC1-C2 A02.5; RC1-C2 A02.6
Bn_A02_p3160461	A02	Sp <sub>L1-L2</sub>	2 205 597	267 875	3 398 658	3 398 658	T/G	0.41	-0.01	T	2.25	QTL.A02.1	RC1-C2 A02.4; RC1-C2 A02.5; RC1-C2 A02.6
Bn_A02_p24522983	A02	ADW	30 331 518	28 823 104	31 631 508	31 631 508	T/C	0.32	0.33	C	4.80	QTL.A02.2	RC1-C2 A02.36; RC2-C3 A02.27
Bn_A02_p24522983	A02	TDW	30 331 518	28 823 104	31 631 508	31 631 508	T/C	0.32	0.35	C	4.56	QTL.A02.2	RC1-C2 A02.36; RC2-C3 A02.27
Bn_A03_p7919195	A03	AUC <sub>L1-L2</sub>	8 243 389	6 348 789	10 116 857	3 768 068	A/G	0.20	43.11	G	9.03	QTL.A03.1	RC1-C2 A03.3; RC1-C3 A03.5 to RC1-C3 A03.19
Bn_A04_p4834943	A04	ADW	5 308 114	3 344 129	5 734 512	2 390 383	A/G	0.27	-0.24	A	4.03	QTL.A04.1	RC1-C2 A03.3; RC1-C3 A03.5 to RC1-C3 A03.19
Bn_A04_p4834943	A04	LA	5 308 114	3 344 129	5 734 512	2 390 383	A/G	0.27	-46.64	A	8.57	QTL.A04.1	RC1-C2 A03.3; RC1-C3 A03.5 to RC1-C3 A03.19
Bn_A04_p4834943	A04	TDW	5 308 114	3 344 129	5 734 512	2 390 383	A/G	0.27	-0.26	A	4.12	QTL.A04.1	RC1-C2 A03.3; RC1-C3 A03.5 to RC1-C3 A03.19
Bn_A04_p6980459	A04	LWR	9 808 446	8 168 435	11 787 122	3 618 687	G/A	0.46	-0.003	G	0.65	QTL.A04.2	RC1-C3 A04.14; RC1-C3 A04.16
Bn_A04_p13589498	A04	RDW	16 950 391	16 932 185	16 995 761	63 576	G/T	0.23	0.02	T	2.99	QTL.A04.3	RC1-C3 A04.14; RC1-C3 A04.16
Bn_A05_p17822322	A05	LWR	33 251 876	31 644 907	34 793 415	3 148 508	T/C	0.32	-0.003	T	1.38	QTL.A05.1	RC1-C3 A04.14; RC1-C3 A04.16

### **Supplemental Table III.2.4. Continued**

Supplemental Table III.2.4. Continued

Peak SNP	Chr.	Trait	Peak position (pb)	Position flanking marker 1	Position flanking marker 2	QTL size (pb)	Major/minor allele	SNP weight	Favorable allele	R <sup>2</sup>	QTL	Colocalization with F <sub>ST</sub>
Bn_scaff_17972_1_p48	C07	LasNQ	40 056 937	38 295 055	41 911 016	3 615 961	G/A	0.32	99.53	A	8.92	QTL.C07.1
6112												
Bn_scaff_17291_1_p70	C07	NutE <sub>AER</sub>	42 015 695	41 292 508	43 431 882	2 139 374	C/T	0.19	0.26	A	4.93	QTL.C07.1
5967												
Bn_scaff_17291_1_p70	C07	NutE <sub>TOT</sub>	42 015 695	41 292 508	43 431 882	2 139 374	C/T	0.19	0.26	A	4.19	QTL.C07.1
5967												
Bn_scaff_16770_1_p40	C08	SNU	28 089 126	26 690 225	29 914 847	3 224 622	A/C	0.49	0.06	C	3.42	QTL.C08.1
35830												
Bn_scaff_17028_1_p43	C09	SCA	57 072 021	55 532 620	59 007 215	3 474 595	A/C	0.31	-0.00003	A	3.37	QTL.C09.1
2707												RC1-C3.C09.11; RC1-C3.C09.12

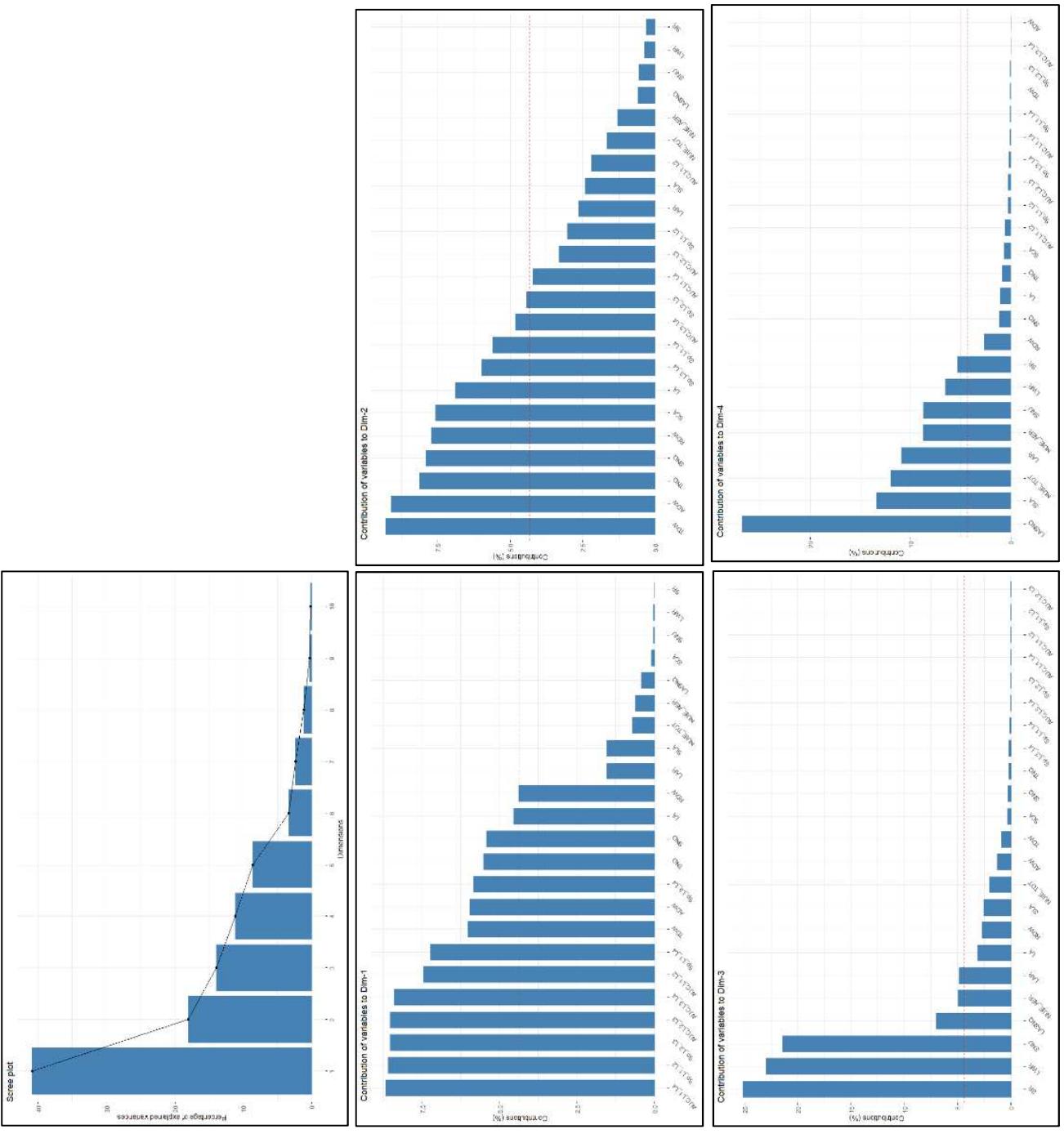
Significant SNP markers were ordered regarding their position on the genome using the Damor-bzh V10 (Rousseau-Gueutin et al. 2020).

For each SNP, information is given about the flanking markers, QTL size as well as major and minor alleles, followed by the minor allele frequency (MAF).

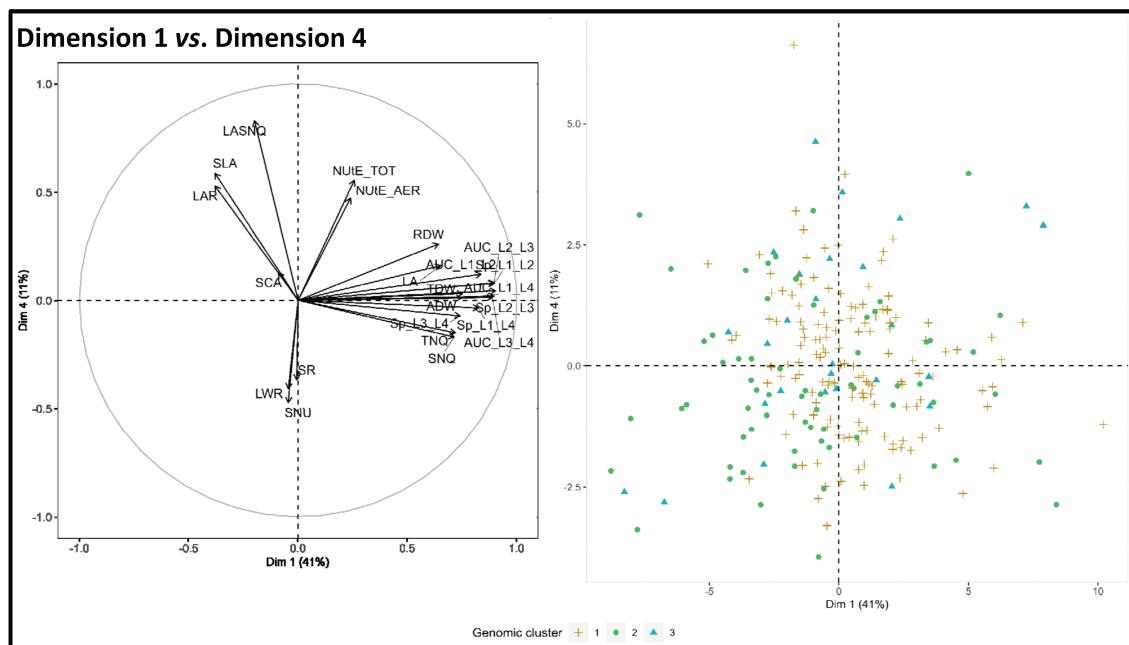
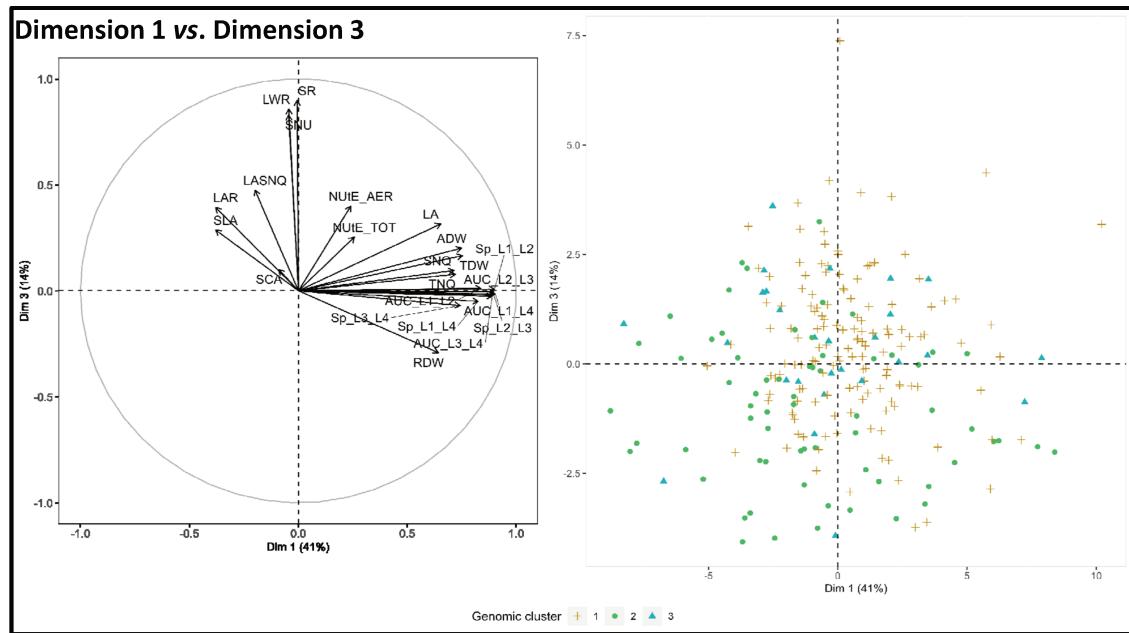
The SNP weight indicating the favorable allele version and the phenotypic variance explained by the SNP (R<sup>2</sup>) are provided.

The 38 trait-associated SNP represent 21 QTL (see last column of the table).

**Supplemental Figure III.2.1. Eigenvalues and contribution of early vigor traits to PCA dimensions**



**Supplemental Figure III.2.2. Additional PCA plots of variables (left) and individuals (right)**



### III.3. Prédiction phénomique de la vigueur précoce

---

Les traits impliqués dans la vigueur précoce sont sous contrôle polygénique et les effets des QTL identifiés sont faibles à modérés (*cf.* III.2). Par conséquent l'utilisation de méthodes de prédiction basées sur l'ensemble du génome pourrait se révéler pertinente afin de sélectionner pour ces traits. Dans cette partie, je me suis intéressée à la capacité de prédiction de traits liés à la vigueur précoce par des données spectrales ou génomiques. Les questions posées dans cette partie sont les suivantes :

- **Quel est l'impact de la nature du spectre sur la capacité prédictive ?** Pour cela, j'ai comparé la capacité prédictive de modèles de PP utilisant des spectres acquis sur différents organes de la plante : i) des spectres acquis sur des graines entières issues d'un même lot de multiplication (*cf.* II) et ii) des spectres acquis sur des poudres de feuilles sèches obtenues lors du prélèvement final de l'expérimentation en tunnel. L'état physiologique des graines, potentiellement capturé par les spectres NIRS, est supposé être faiblement corrélé aux phénotypes de vigueur précoce qui eux se mettent en place après la germination. Concernant les poudres de feuilles, ces dernières ont été obtenues en utilisant les mêmes plantes que celles qui ont servi à acquérir les traits de vigueur précoce. On peut faire l'hypothèse que la composition moléculaire captée par la spectrométrie dans le proche infra-rouge est plus proche des phénotypes à prédire que celle des graines, et que cette proximité permet d'améliorer la capacité prédictive des modèles de PP.
- **Quel est l'impact des corrections réalisées sur les données spectrales sur la matrice d'apparentement et la capacité prédictive des modèles ?** En prédiction phénomique, les données spectrales sont utilisées pour rendre compte de l'apparentement entre les individus de manière plus rapide et moins couteuse que les données de génotypage, cependant les spectres peuvent être soumis à des effets environnementaux lors de leur acquisition, ce qui peut *a priori* impacter la matrice spectrale, et donc la capacité prédictive des modèles. Dans la littérature, plusieurs stratégies sont mises en œuvre pour corriger les données spectrales prises en compte dans les modèles de prédiction (Brault et al. 2022; Meyenberg et al. 2024). Pour étudier cet effet, trois corrections des données spectrales ont été comparées :
  - L'estimation d'une **moyenne des valeurs d'absorbance** pour chaque longueur d'onde et chaque génotype. Aucune correction spécifique n'est apportée pour prendre en compte l'impact des effets environnementaux sur les données spectrales. Concernant les spectres acquis sur graines, le spectre moyen a été obtenu à l'aide de trois valeurs acquises à partir de trois échantillons issus même lot de semences pour chaque génotype. Concernant les spectres acquis sur poudre de feuilles, le spectre moyen a



été obtenu à l'aide de douze à dix-huit valeurs pour chaque génotype correspondant aux spectres individuels acquis sur six à neuf plantes pour chaque génotype et chacune des deux répétitions.

- L'estimation d'une **moyenne ajustée des valeurs d'absorbance** avec correction des effets environnementaux. Cette correction a été appliquée aux spectres obtenus sur poudre de feuilles. Pour cela les moyennes ajustées (BLUE) pour chaque longueur d'onde et chaque génotype ont été obtenues à l'aide d'un modèle mixte modélisant l'impact des effets environnementaux, ici les différentes expérimentations et méthodes d'acquisition des spectres.
- La prédiction de **valeurs génétiques des absorbances** pour chaque longueur d'onde et pour chaque génotype en prenant en compte l'information d'apparentement génomique entre les individus.

Les différentes analyses ont été réalisées soit à l'échelle du panel entier soit à l'échelle du cluster génomique C1. Les données acquises à l'échelle du panel entier sont caractérisées i) par un potentiel effet lié à la structure du panel en trois clusters génomiques, ainsi que ii) par un effet environnemental et un effet d'acquisition des données spectrales sur les feuilles, liés au fait que les génotypes ont été phénotypés selon un dispositif incomplet en « augmented design ». Les données acquises à l'échelle du cluster génomique C1 sont caractérisées par l'absence d'effet de structure de la population, ainsi que par le fait que les spectres « feuilles » ont été acquis en même temps pour tous les génotypes du cluster et selon un dispositif complet et équilibré.

## Matériel et Méthodes

### Matériel végétal et données phénotypiques

Dans cette partie, nous avons utilisé 205 génotypes de colza issus du panel de diversité présenté dans les parties précédentes, ainsi que les moyennes ajustées (BLUE) des phénotypes liés à la vigueur précoce et décrits dans la partie III.2 (**Tableau III.2.1**).

### Données de génotypage et matrice d'apparentement génomique

Les accessions ont été génotypées en utilisant la puce SNP Brassica 60K Infinium® (Clarke et al. 2016). Parmi l'ensemble des SNP, 33 151 SNP ont été validés en appliquant les valeurs seuils des filtres suivants : un seuil de 5% pour la MAF (Minor Allele Frequency), 10% pour les valeurs manquantes et



10% pour le niveau d'hétéozygotie. Les SNP manquants ont été imputés en utilisant le logiciel Beagle (Browning et al. 2018). Les SNP correspondent aux positions du génome de référence Darmor-bzh V10 (Rousseau-Gueutin et al. 2020). A partir de ces marqueurs, une matrice d'apparentement génomique a été estimée à partir de la formule suivante :

$$K = \frac{WW^T}{N_M},$$

Avec W la matrice mise à l'échelle à partir des fréquences alléliques, d'une dimension NxM, N étant le nombre de génotypes et N<sub>M</sub> étant le nombre de marqueurs moléculaires, et W<sup>T</sup> correspond à la matrice W transposée.

### Acquisition des spectres NIRS

Concernant les spectres utilisés, nous avons choisi de réutiliser les spectres NIRS acquis sur graines entières et présentés dans le chapitre II. De plus, nous avons acquis de nouveaux spectres, sur des poudres de feuilles sèches. Pour cela, à la suite de la récolte des expérimentations en tunnel (*cf. III.2*), les échantillons qui avaient été étuvés pour estimer les biomasses aériennes sèches (ADW) ont été broyés individuellement pour chaque plante, puis des spectres NIRS ont été acquis à l'aide du spectromètre NIRS Bruker MPA (Bruker Optic Inc., Allemagne) pour chaque échantillon. En fonction de la quantité de poudre disponible, deux contenants ont été utilisés pour acquérir les spectres. Le premier contenant est un flacon standard, de taille adaptée à la sortie lumineuse du spectromètre. Ce flacon était utilisé lorsqu'une grande quantité de poudre était disponible. A l'inverse, en cas de faibles quantités de poudre disponibles, un deuxième flacon a été placé à l'intérieur du flacon standard. Afin de combler la différence de diamètre entre les deux flacons une pièce métallique sur-mesure a été produite pour combler l'interstice. Pour chaque plante, le spectre obtenu correspond à une moyenne de 64 spectres acquis automatiquement. En moyenne 16 spectres sont disponibles par génotype à l'échelle du dispositif expérimental.

### Corrections des spectres et calcul de l'héritabilité génomique des spectres

Les spectres acquis sur feuilles sèches ont été prétraités de la même façon que les spectres acquis sur graines dans le chapitre II : les spectres ont été centrés et mis à l'échelle, puis un filtre de Savitzky-Golay a été appliqué (Savitzky and Golay 1964). Trois méthodes de correction des spectres ont été testées :

- 1) Correction 1 : L'obtention d'un spectre moyen par génotype, pour chaque tissu, calculé à partir d'une moyenne arithmétique. Les modèles de PP utilisant ces spectres sont nommés «



« HBLUP<sub>leaf1</sub> » lors de l'utilisation de spectres moyens acquis sur poudre de feuilles, ou « HBLUP<sub>seed1</sub> » lors de l'utilisation de spectres moyens acquis sur graines.

- 2) Correction 2 : L'obtention de moyennes ajustées (BLUE, Best Linear Unbiased Estimations) pour chaque longueur d'onde et chaque génotype, afin de corriger les effets environnementaux subis par les plantes et les effets d'acquisition des spectres acquis sur feuilles. Cette correction n'a pas été appliquées sur les spectres « graines » qui ont été obtenus dans les mêmes conditions pour tous les génotypes. Les modèles de prédiction utilisant ces spectres seront nommés « HBLUP<sub>leaf2</sub> ». Le modèle utilisé pour calculer les valeurs BLUE des spectres NIRS acquis sur feuilles est le suivant :

$$(1) \quad A_{ijk} = \mu + bloc_j + Témoin_l + Génotype_{i(j)} + méthode_acquisition_k + \varepsilon_{ijk},$$

Avec  $A_{ijk}$  la valeur d'absorbance pour la longueur d'onde considérée du  $i^{\text{ème}}$  génotype dans le  $j^{\text{ème}}$  bloc obtenue avec la méthode d'acquisition  $k$ ,  $\mu$  la moyenne générale,  $bloc_j$  est l'effet aléatoire du  $j^{\text{ème}}$  bloc suivant une distribution normale  $block \sim \mathcal{N}(0, \sigma_b^2)$ ,  $Témoin_l$  est l'effet fixe du  $j^{\text{ème}}$  génotype témoin,  $Génotype_{i(j)}$  est l'effet fixe du  $i^{\text{ème}}$  génotype du  $j^{\text{ème}}$  bloc,  $méthode\_acquisition_k$  est l'effet aléatoire du type de contenant utilisé pour contenir les poudres lors de l'acquisition des spectres et suivant une distribution normale  $méthode\_acquisition \sim \mathcal{N}(0, \sigma_m^2)$ , et  $\varepsilon_{ijk}$  est la résiduelle globale.

- 3) Correction 3 : L'obtention de valeurs génétiques (BLUP, Best Linear Unbiased Predictions) pour chaque longueur d'onde et chaque génotype. La prédiction des valeurs génétiques permet de s'affranchir des effets environnementaux et de prendre en compte l'apparentement entre les génotypes de la population. De la même manière que précédemment, les modèles utilisant ces spectres corrigés seront nommés « HBLUP<sub>leaf3</sub> » ou « HBLUP<sub>seed3</sub> ». Les formules utilisées pour calculer les valeurs BLUP génomiques des spectres NIRS acquis sur graines (2) et sur feuilles (3) sont les suivantes :

$$(2) \quad \text{Modèle « Graines » : } A_{ij} = \mu + genotype_i + Repetition_j + \varepsilon_{ij},$$

$$(3) \quad \text{Modèle « Feuilles » } A_{ijk} = \mu + bloc_j + Témoin_l + génotype_{i(j)} + méthode_acquisition_k + \varepsilon_{ijk},$$

Avec pour le modèle « Graines »  $A_{ij}$  la valeur d'absorbance obtenue pour le  $i^{\text{ème}}$  génotype et la  $j^{\text{ème}}$  répétition,  $\mu$  la moyenne générale, et  $\varepsilon_{ij}$  est la résiduelle globale.



Et pour le modèle « Feuilles »,  $A_{ijk}$  la valeur d'absorbance pour le  $i^{\text{ème}}$  génotype dans le  $j^{\text{ème}}$  bloc obtenue avec la méthode d'acquisition  $k$ ,  $\mu$  la moyenne générale,  $\text{bloc}_j$  est l'effet aléatoire du  $j^{\text{ème}}$  bloc suivant une distribution normale  $\text{block} \sim \mathcal{N}(0, \sigma_b^2)$ ,  $\text{Témoin}_i$  est l'effet fixe du  $i^{\text{ème}}$  génotype témoin,  $\text{génotype}_{i(j)}$  est l'effet aléatoire du  $i^{\text{ème}}$  génotype observé dans le  $j^{\text{ème}}$  bloc  $\text{génotype} \sim \mathcal{N}(0, K\sigma_G^2)$  avec  $K$  la matrice d'apparentement génomique,  $\text{méthode\_acquisition}_k$  est l'effet aléatoire du type de contenant utilisé pour contenir les poudres lors de l'acquisition des spectres et suivant une distribution normale  $\text{méthode\_acquisition} \sim \mathcal{N}(0, \sigma_m^2)$ , et  $\varepsilon_{ijk}$  est la résiduelle globale.

Afin de calculer l'héritabilité génomique des données spectrales corrigées avec les corrections 1 et 2, un modèle linéaire mixte a été utilisé afin d'estimer les variances génétique ( $\sigma_G^2$ ) et résiduelle ( $\sigma_e^2$ ) :

$$\hat{A}_i = \mu + \text{génotype}_i + \varepsilon_i,$$

Avec  $\hat{A}_i$  la valeur d'absorbance corrigée (moyenne arithmétique ou moyenne ajustée BLUE) pour le génotype  $i$ ,  $\text{génotype}_i$  est l'effet aléatoire du  $i^{\text{ème}}$  génotype ;  $\text{génotype} \sim \mathcal{N}(0, K\sigma_G^2)$  avec  $K$  la matrice d'apparentement génomique,  $\varepsilon_i$  est l'effet de la résiduelle globale. L'héritabilité génomique des spectres pour les corrections 1 et 2, a été calculée à partir de la formule suivante :

$$H^2_{NIRS} = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \hat{\sigma}_e^2}$$

Avec  $\hat{\sigma}_G^2$  et  $\hat{\sigma}_e^2$  les estimations REML des variances génétique et résiduelle obtenues à partir du modèle précédent, en utilisant le package R sommer (Covarrubias-Pazaran 2016).

Les matrices d'apparentement spectrales (matrices H), sont calculées à partir de la formule suivante à partir de chaque type de spectres (« graines » ou « feuilles » combinés aux méthodes de correction présentées précédemment) :

$$H = \frac{SS^T}{N_w},$$

Avec S la matrice des données spectrales pour les différentes longueurs d'ondes et les différents génotypes (moyennes, BLUE ou BLUP) de dimension  $N \times N_w$ , N représente le nombre de génotypes et  $N_w$  représente le nombre de longueurs d'ondes.  $S^T$  correspond à la matrice S transposée.

Afin de comparer les matrices H entre elles et avec la matrice K, des tests de Mantel avec 9999 permutations ont été réalisés. Ces comparaisons entre matrices (spectrales et génomique) ont été réalisées à la fois à l'échelle du panel de diversité entier et à la fois à l'échelle du cluster 1 pour s'affranchir d'un effet lié à la structuration du panel et au protocole d'acquisition des spectres (séries d'acquisition des spectres différentes en fonction du cluster génomique pour les spectres collectés sur feuilles).



## Prédictions génomique et phénotypique

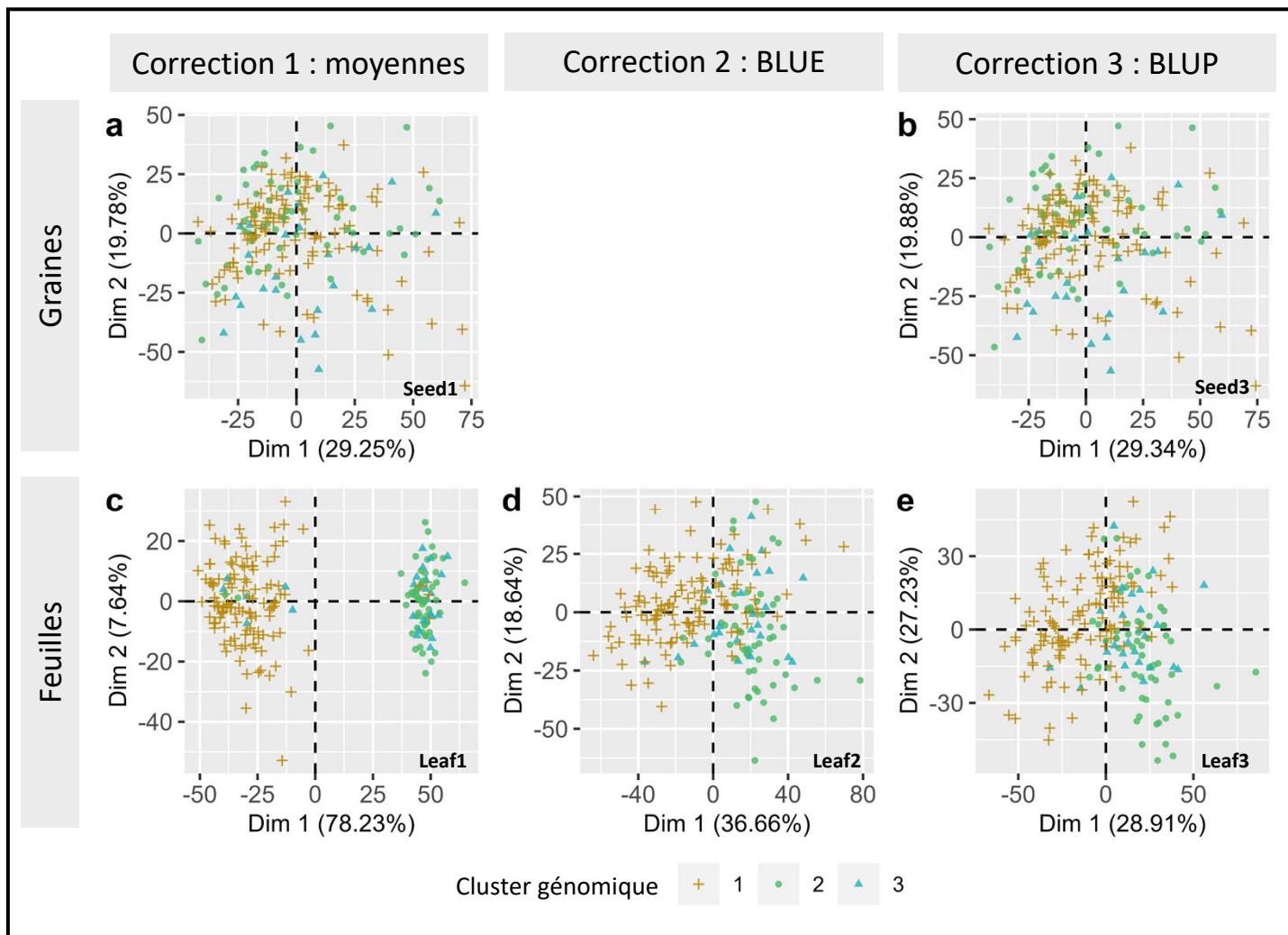
Les prédictions génomiques et phénotypiques ont été par la suite réalisées suivant la même méthode que celle décrite dans le chapitre II (cf. « Genomic and phenomic predictions »). Nous avons utilisé les modèles suivants :

- Le modèle GBLUP, utilisant les données génomiques.
- Le modèle HBLUP, utilisant les données spectrales au lieu des données génomiques. Ce modèle se décline selon les spectres utilisés :
  - o HBLUP<sub>leaf1</sub>, utilisant les données spectrales moyennées acquises sur poudres de feuilles sèches.
  - o HBLUP<sub>seed1</sub>, utilisant les données spectrales moyennées acquises sur graines.
  - o HBLUP<sub>leaf2</sub>, utilisant les moyennes ajustées des données spectrales acquises sur poudres de feuilles sèches (BLUE).
  - o HBLUP<sub>seed3</sub>, utilisant les valeurs génétiques prédictes des spectres acquis sur graines (BLUP).
  - o HBLUP<sub>leaf3</sub>, utilisant les valeurs génétiques prédictes des spectres acquis sur poudres de feuilles sèches (BLUP).

Deux scénarios de prédiction ont été comparés :

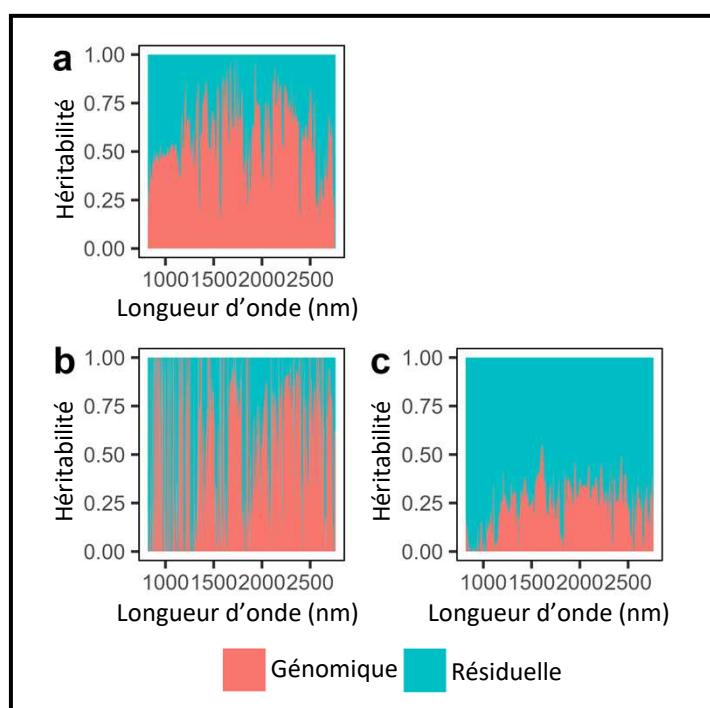
- Scénario<sub>c1</sub> : seuls les génotypes du cluster génomique C1 qui regroupe majoritairement les génotypes de colzas d'hiver ont été considérés (*cf. III.2*). La population de calibration contient 80% des génotypes choisis aléatoirement, et la population de validation les 20% restants.
- Scénario<sub>whole</sub> : l'ensemble du panel a été utilisé en tant que population de calibration et de population de validation. La population de calibration contient 80% des génotypes choisis aléatoirement. La population de validation est composée des 20% restants.

Chaque cross-validation a été répétée 100 fois. Les modèles ont été ajustés en utilisant le package R sommer (Covarrubias-Pazaran 2016). La capacité prédictive des modèles (PA) a été calculée comme la corrélation de Pearson entre les valeurs observées et les valeurs prédictes de la population de validation.



**Figure III.3.1.** Analyses en composantes principales des données spectrales acquises sur graines (a, b) ou sur feuilles sèches (c, d, e).

Trois corrections ont été appliquées sur les données spectrales: moyennes arithmétiques des différents spectres (correction 1 – panels a & c), valeurs BLUE (correction 2 – panel d), valeurs BLUP (correction 3 – panels b & e). Le nom de la combinaison entre la correction et l'organe utilisé est mentionné en bas de chaque figure (ex. seed1 = spectres NIRS acquis sur les graines et valeurs spectrales corrigées par la correction 1).



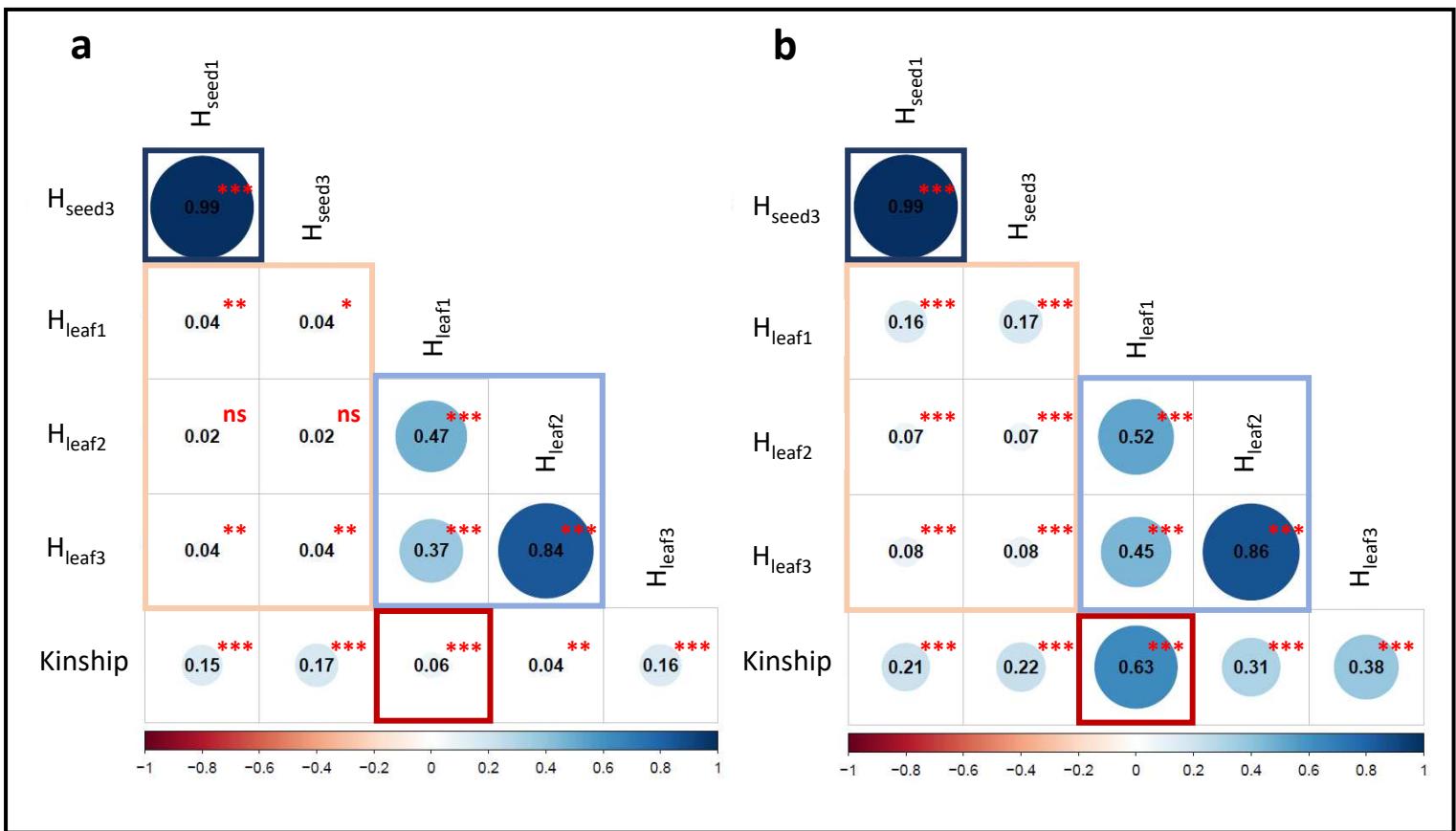
**Figure III.3.2.** Héritabilité génomique en fonction de la longueur d'onde spectrale.  
(a) Données spectrales acquises sur graines et corrigées par les moyennes arithmétiques. Données spectrales acquises sur poudres de feuilles et corrigées par les moyennes arithmétiques (b) ou les moyennes ajustées (c).

## Résultats

**La correction des spectres NIRS pour les effets environnementaux met en avant une faible prise en compte de la structure génomique du panel malgré une héritabilité des données spectrales modérée à forte**

Dans un premier temps, les données spectrales ont été utilisées pour caractériser la diversité génétique présente dans le panel de diversité. La figure III.3.1, présente les ACP réalisées à l'aide des spectres obtenus sur graines (Figure III.3.1 a et b) et sur feuilles (Figure III.3.1 c, d, e) et pour les différentes méthodes de correction. Chaque génotype a été coloré en fonction de son appartenance aux clusters génomiques : le cluster génomique C1 (en marron) correspond majoritairement aux colzas d'hiver, le cluster génomique C2 (en vert) correspond majoritairement aux colzas de printemps et le cluster génomique C3 (en bleu) correspond majoritairement aux colzas asiatiques. Les ACP obtenues à partir des données spectrales « graines » ne permet pas de différencier les trois clusters génomiques. Les ACP obtenues à partir des données spectrales « feuilles » montrent une différenciation entre les clusters génomiques, et ce quelle que soit la correction utilisée (Figure III.3.1 c, d, e). L'effet de la structure est cependant moins marqué quand les spectres sont corrigés à l'aide des méthodes 2 et 3. Cet effet peut être attribué à un phénotype spectral intrinsèquement très différent entre les accessions des différents clusters génomiques, mais aussi au fait que les acquisitions des spectres sur poudres de feuilles ont été réalisées lors d'expérimentations différentes entre les génotypes du cluster C1 (WOSR) et du cluster C2 (SOSR). Dans ce cas, on ne peut pas exclure une confusion entre l'effet génétique lié à l'appartenance à un cluster génomique et l'effet de l'expérimentation. *A contrario*, l'acquisition des données spectrales acquises sur graines a été réalisée dans les mêmes conditions pour l'ensemble du panel.

Dans un deuxième temps, l'héritabilité génomique des spectres a été étudiée pour les données corrigées avec les méthodes 1 et 2. Une forte héritabilité des longueurs d'ondes est observée le long du spectre pour les valeurs d'absorbance moyennées (correction 1) pour le compartiment « graines », avec une héritabilité moyenne de 0.59 (Figure III.3.2 a). Concernant les spectres « feuilles », un profil différent est observé pour l'héritabilité génomique en fonction de la méthode de correction appliquée : les héritabilités sont plus faibles quand les données sont obtenues avec la correction 2 (héritabilité moyenne de 0.25) (Figure III.3.2 c, d, e) par rapport à la correction 1 (héritabilité moyenne de 0.61). Cette différence entre les deux méthodes de correction peut s'expliquer en partie par le même phénomène d'effet confondant entre des effets génétiques et environnementaux, qui a pu entraîner, avec la méthode 1, une surestimation de la variance génétique. Cet effet a pu être corrigé en estimant des moyennes ajustées avec la méthode 2.



**Figure III.3.3. Corrélations entre les matrices d'apparentement spectrales et génomique calculées en utilisant uniquement les génotypes du cluster C1 (a) ou sur l'ensemble du panel (b).**

Les corrélations entre les matrices spectrales « graines » sont encadrées en bleu foncé, celles entre les matrices spectrales « graines » et « feuilles » en beige, et celles entre les matrices « feuille » en bleu clair. La corrélation entre la matrice génomique K et la matrice  $H_{leaf1}$  est encadrée en rouge sur les deux figures. Les étoiles informent sur la significativité des corrélations: \*\*\*p-value<0.001, \*\*P-value=0.001-0.01, \*p-value=0.01-0.05, ns p-value>0.05.

**Les matrices d'apparentement spectral issues des spectres « feuilles » et « graines » sont faiblement corrélées à la matrice d'apparentement génomique**

Pour mieux comprendre à quel niveau les données spectrales permettent de rendre compte des proximités génétiques entre les individus, les matrices d'apparentement calculées sur les données spectrales ( $H_{seed1}$  et  $H_{seed3}$ ,  $H_{leaf1}$ ,  $H_{leaf2}$ ,  $H_{leaf3}$ ) ont été comparées entre elles et à la matrice génomique K. Dans un premier temps, les comparaisons ont été réalisées uniquement au sein du cluster C1 pour limiter les biais liés à la confusion des effets expérimentations et cluster génomique, comme soulevé précédemment. Ensuite, les matrices ont été comparées à l'échelle du panel entier.

A l'échelle du cluster C1, les matrices spectrales issues des données acquises sur feuilles ne sont pas corrélées aux matrices issues des données acquises sur graines (**Figure III.3.3 a**) ( $r<0.04$ ). De plus les matrices spectrales sont faiblement corrélées ou non corrélées à la matrice d'apparentement génomique K ( $r=0.15$  entre K et  $H_{seed1}$  et  $r=0.04$  entre K et  $H_{leaf2}$ ). Ainsi, les spectres acquis sur les deux types de tissus capturent des informations différentes, et des informations différentes des informations de génotypage.

A l'échelle du panel de diversité, on retrouve les mêmes tendances (**Figure III.3.3.b**) : une corrélation nulle à faible entre les matrices spectrales obtenues à l'aide des graines ou des feuilles, une corrélation moyenne à forte des matrices issues des différentes méthodes de correction pour le même organe, et une corrélation modérée des matrices spectrales avec la matrice génomique (entre 0.21 et 0.31) à l'exception d'une forte corrélation entre la matrice  $H_{leaf1}$  et la matrice K ( $r=0.63$ ). Cette dernière valeur est très différente de la même corrélation obtenue à l'échelle du cluster C1 ( $r=0.06$ ) et confirme la confusion possible entre effet expérimentation et effet lié à la structure de la population. Par rapport à la méthode de correction 2, la méthode de correction 3 a pour conséquence une augmentation de la corrélation entre la matrice K et les matrices H. La corrélation entre K et  $H_{leaf2}$  vaut  $r=0.31$  et la corrélation entre K et  $H_{leaf3}$  vaut  $r=0.38$ .

**Au sein du cluster C1, les modèles de prédiction phénotypique sont plus performants que le GBLUP, bien que les PA des différents modèles soient faibles.**

Dans un premier temps, les différents modèles de prédiction phénotypique et génomique ont été comparés au sein du cluster génomique C1 pour s'affranchir des potentiels effets confondants entre la structure de la population et le design expérimental d'acquisition des spectres. A l'échelle globale, on peut poser plusieurs constats : i) tout d'abord le fait que le GBLUP ne permet pas de prédire la valeur des génotypes pour aucun des traits étudiés. Ce constat peut être mis en relation avec le fait que les héritabilités calculées pour le cluster C1 sont faibles, voire nulles, pour 7 des 16 traits considérés. ii) ensuite le fait que les PA pour les modèles de prédiction phénotypique restent faibles

**Tableau III.3.1. Capacités prédictives moyennes des modèles en utilisant les individus du cluster C1 (Scénario<sub>C1</sub>) ou de la population entière (Scénario<sub>whole</sub>).**

	Scénario <sub>C1</sub> (n=120)							Scénario <sub>whole</sub> (n=205)						
	H <sup>2</sup>	GBLUP	HBLUP seed1	HBLUP seed3	HBLUP leaf1	HBLUP leaf2	HBLUP leaf3	H <sup>2</sup>	GBLUP	HBLUP seed1	HBLUP seed3	HBLUP leaf2	HBLUP leaf3	
TDW	<b>0.45</b>	-0.07 (0.14)	<b>0.21</b> (0.16)	0.18 (0.18)	0.17 (0.21)	-0.18 (0.13)	0.008 (0.12)	<b>0.52</b>	<b>0.40</b> (0.14)	0.25 (0.13)	0.25 (0.13)	0.34 (0.14)	0.35 (0.13)	
ADW	<b>0.50</b>	-0.08 (0.15)	<b>0.17</b> (0.17)	0.15 (0.15)	<b>0.17</b> (0.20)	-0.22 (0.14)	-0.07 (0.17)	<b>0.52</b>	<b>0.42</b> (0.14)	0.28 (0.13)	0.24 (0.15)	0.35 (0.14)	0.39 (0.12)	
LAR	0	-0.11 (0.16)	0.07 (0.16)	<b>0.10</b> (0.14)	-0.11 (0.14)	-0.27 (0.12)	-0.22 (0.10)	0.03	-0.05 (0.10)	0.06 (0.13)	0.07 (0.13)	0.20 (0.14)	<b>0.22</b> (0.13)	
RDW	0.004	-0.03 (0.19)	0.20 (0.16)	0.19 (0.15)	<b>0.29</b> (0.20)	0.16 (0.13)	0.15 (0.17)	<b>0.39</b>	0.26 (0.16)	-0.13 (0.10)	-0.14 (0.10)	0.25 (0.14)	<b>0.27</b> (0.14)	
SR	0	-0.24 (0.13)	-0.32 (0.15)	-0.30 (0.13)	<b>0.22</b> (0.17)	0.11 (0.15)	0.04 (0.15)	<b>0.10</b>	<b>0.28</b> (0.13)	0.18 (0.16)	0.19 (0.14)	0.24 (0.14)	0.23 (0.14)	
SNU	0	-0.21 (0.11)	-0.28 (0.09)	-0.32 (0.07)	<b>0.26</b> (0.18)	0.13 (0.16)	0.07 (0.16)	0.07	<b>0.22</b> (0.13)	0.10 (0.13)	0.12 (0.12)	0.17 (0.14)	0.18 (0.16)	
LWR	0	-0.25 (0.13)	-0.41 (0.005)	NA	<b>0.11</b> (0.22)	-0.05 (0.13)	-0.09 (0.12)	<b>0.15</b>	<b>0.21</b> (0.16)	-0.01 (0.09)	-0.03 (0.12)	0.01 (0.13)	0.05 (0.12)	
SNQ	<b>0.19</b>	-0.09 (0.13)	0.18 (0.16)	0.17 (0.17)	<b>0.20</b> (0.25)	-0.22 (0.13)	-0.06 (0.13)	<b>0.45</b>	<b>0.27</b> (0.14)	0.008 (0.10)	0.006 (0.09)	0.14 (0.15)	0.20 (0.12)	
NUtE <sub>tot</sub>	<b>0.17</b>	0.009 (0.16)	-0.39 (0.15)	-0.38 (0.09)	-0.11 (0.12)	<b>0.11</b> (0.20)	0.02 (0.16)	<b>0.15</b>	0.30 (0.17)	0.19 (0.13)	0.21 (0.13)	<b>0.51</b> (0.14)	0.46 (0.13)	
NUtE <sub>aer</sub>	<b>0.13</b>	-0.05 (0.15)	NA	-0.42 (0.07)	-0.06 (0.19)	<b>0.10</b> (0.18)	-0.06 (0.14)	<b>0.18</b>	0.33 (0.16)	0.21 (0.14)	0.23 (0.14)	<b>0.51</b> (0.13)	0.45 (0.14)	
SCA	0	-0.16 (0.09)	-0.38 (0.03)	-0.37 (0.09)	0.14 (0.15)	<b>0.24</b> (0.13)	0.15 (0.14)	<b>0.17</b>	0.21 (0.13)	0.11 (0.13)	0.14 (0.14)	<b>0.35</b> (0.13)	0.30 (0.12)	
LA	<b>0.12</b>	-0.15 (0.12)	-0.02 (0.12)	-0.07 (0.11)	-0.04 (0.23)	-0.19 (0.14)	0.03 (0.15)	<b>0.52</b>	<b>0.41</b> (0.11)	0.27 (0.13)	0.26 (0.12)	0.32 (0.13)	0.33 (0.14)	
Sp L1-L4	<b>0.35</b>	0.09 (0.17)	-0.33 (0.08)	-0.31 (0.11)	-0.31 (0.15)	-0.28 (0.13)	-0.03 (0.11)	<b>0.13</b>	0.11 (0.13)	-0.02 (0.08)	-0.03 (0.09)	0.16 (0.14)	<b>0.19</b> (0.16)	
AUC L1-L4	<b>0.24</b>	0.03 (0.18)	-0.33 (0.10)	-0.35 (0.13)	-0.25 (0.16)	-0.18 (0.09)	-0.01 (0.12)	<b>0.13</b>	0.15 (0.14)	0.01 (0.10)	0.03 (0.09)	0.25 (0.13)	<b>0.27</b> (0.11)	
LASNQ	0.04	-0.03 (0.13)	0.01 (0.13)	0.009 (0.11)	-0.04 (0.19)	-0.32 (0.09)	NA	<b>0.20</b>	0.26 (0.13)	0.15 (0.13)	0.16 (0.14)	0.23 (0.13)	<b>0.28</b> (0.12)	
SLA	0	-0.15 (0.11)	0.03 (0.17)	0.09 (0.17)	-0.16 (0.17)	-0.30 (0.09)	-0.24 (0.11)	0.02	-0.04 (0.10)	0.06 (0.11)	0.07 (0.12)	0.21 (0.15)	<b>0.24</b> (0.12)	

L'héritabilité de chaque trait (H<sup>2</sup>) est indiquée pour Scénario<sub>C1</sub> et Scénario<sub>whole</sub>. Les valeurs d'héritabilité très faibles à nulles sont grisées. Les données relatives aux modèles développés sur les graines (HBLUP<sub>seed</sub>) et feuilles (HBLUP<sub>leaf</sub>) sont indiquées par un fond jaune ou vert respectivement. Les écarts-types sont représentés entre parenthèses. Les modèles les plus performants pour chaque trait sont écrits en rouge et en gras.

(<0.29) quel que soit le trait ; et enfin iii) le fait que les traits d'efficience, les traits de biomasse et les traits relatifs à l'azote (SNU, SR, SCA, SNQ, RDW, TDW : PA >0.20) sont les mieux prédis alors que les traits de vitesse de développement ( $Sp_{L1-L4}$  et  $AUC_{L1-L4}$ ) et de surface foliaire (LA, LASNQ, SLA) ne sont pas prédis par aucun des modèles (**Tableau III.3.1**).

Pour l'ensemble des 11 traits pouvant être prédis par un modèle de prédition phénotypique, la PA du meilleur modèle varie entre 0.10 et 0.29. Les modèles « graines » (HBLUP<sub>seed1</sub> ou HBLUP<sub>seed3</sub>) sont les meilleurs modèles pour les traits de biomasse ADW, TW et LAR. L'ensemble des autres traits sont mieux prédis par des modèles utilisant les données spectrales acquises sur feuilles.

Pour l'ensemble des traits, le modèle HBLUP<sub>leaf3</sub> présente une PA inférieure aux modèles HBLUP<sub>leaf1</sub> ou HBLUP<sub>leaf2</sub>. Les traits RDW, SR, SNU et LWR et SNQ, présentent des heritabilités très faibles (<0.19) et sont des traits pour lesquels le meilleur modèle de PP est le modèle HBLUP<sub>leaf1</sub> (PA comprises en 0.11 et 0.29). Les traits NUTE<sub>tot</sub> ( $h^2=0.17$ ), NUTE<sub>aer</sub> ( $h^2=0.13$ ) et SCA ( $h^2=0$ ) sont quant à eux mieux prédis avec le modèle HBLUP<sub>leaf2</sub> (PA comprises entre 0.10 et 0.24). L'ensemble de ces résultats suggère que les données spectrales permettent d'intégrer une partie d'information environnementale dans la prédition, expliquant ainsi des PA supérieures à 0 pour des traits aux heritabilités faibles à nulle, ainsi que la relative faible performance des modèles HBLUP<sub>leaf3</sub>.

En conclusion, au sein du cluster C1, le GBLUP ne permet pas de prédire les traits. Ainsi la prédition phénotypique apparaît comme une opportunité, même si les valeurs de PA restent modérées à faible (<0.26). Pour peu que l'environnement dans lequel les spectres ont été acquis soit proche de l'environnement cible pour l'agriculteur, les approches de PP peuvent apporter une information intéressante. Ceci pourrait expliquer le fait que les modèles basés sur des spectres de feuilles, acquis lors de la même expérimentation que celle ayant permis l'acquisition des traits de vigueur précoce, montrent des PA supérieures. Cependant ces résultats sont aussi impactés par la faible diversité génétique explorée au sein du cluster C1, comme en témoignent les heritabilités faibles. Pour lever cette limite, la même démarche a été réalisée pour le panel entier, en prenant la précaution de ne pas considérer la méthode HBLUP<sub>leaf1</sub>, en raison de la forte confusion entre effet de structure et effet environnemental.

**A l'échelle du panel entier, le GBLUP prédit mieux la majorité des traits de vigueur précoce que les modèles de PP, à l'exception des traits de développement et des traits liés au métabolisme de l'azote.**

Dans cette partie, il s'agit d'élargir la gamme de diversité génétique considérée pour comparer les différents modèles de prédition. Il s'agit aussi de tester l'effet de la structure génétique sur la PA. En effet, les effets de structure peuvent être à l'origine d'une surestimation de la PA lors du calibrage des



modèles sur le panel entier, comme nous l'avons vu pour l'utilisation du modèle GBLUP et la prédiction des traits de germination (*cf.* II.2). Cependant, pour ces mêmes traits de germination, l'effet de structure n'était pas présent pour les modèles de prédiction phénotomique. Cette hypothèse nécessite d'être retestée ici car les variables phénotypiques et spectrales sont différentes.

A l'échelle du panel entier, on observe une augmentation de l'héritabilité pour tous les traits, en comparaison des résultats obtenus pour le cluster C1. Bien que LAR, SNU et SLA présentent toujours des héritabilités faibles ( $h^2 < 0.07$ ), les autres traits présentent des héritabilités comprises entre 0.10 et 0.52. Cette augmentation des valeurs d'héritabilités est concomitante avec une augmentation de la valeur de PA des modèles GBLUP. Pour le GBLUP, on retrouve une PA nulle pour les caractères non héritables (LAR et SLA) et des PA comprises entre 0.11 et 0.42 pour les autres traits. Il est à noter que le GBLUP est le meilleur modèle de prédiction pour 7 des 16 traits investigués : TDW, ADW, SR, SNU, LWR, SNQ, et LA (PA supérieure à  $r=0.21$ ).

Pour les différents traits, les PA des modèles de prédiction phénotomique, sont supérieures à l'échelle du panel entier, comparées au cluster C1 (modèle HBLUP<sub>leaf1</sub> exclu). Les PA augmentent très significativement pour NUT<sub>tot</sub> ( $r=0.51$ , HBLUP<sub>leaf2</sub>), NUT<sub>aer</sub> ( $r=0.51$  HBLUP<sub>leaf2</sub>) et SCA ( $r= 0.35$  HBLUP<sub>leaf2</sub>). Les traits de développement (Sp<sub>L1-L4</sub>, AUC<sub>L1-L4</sub>) et LASNQ qui n'étaient pas prédictibles à l'échelle du cluster C1, sont prédictibles à l'échelle du panel entier, et présentent des PA comprises entre 0.19 et 0.41 pour le meilleur des modèles de PP testés.

Les PA des modèles de PP « feuilles » sont toujours supérieures aux PA des modèles de PP « graines », y compris pour TDW, ADW et LAR pour lesquels les modèles « graines » présentaient de meilleures performances que les modèles « feuilles » pour le cluster C1. Les modèles de PP « feuilles » surperforment même le GBLUP pour les traits LAR, RDW, NUT<sub>tot</sub>, NUT<sub>aer</sub>, SCA, Sp<sub>L1-L4</sub>, AUC<sub>L1-L4</sub>, LASNQ et SLA. De plus, la PA des modèles de PP « feuilles » est supérieure à l'héritabilité pour certains traits (LAR, SR, SNU, NUT<sub>tot</sub>, NUT<sub>aer</sub>, SCA, Sp<sub>L1-L4</sub>, AUC<sub>L1-L4</sub>, LASNQ et SLA), laissant supposer que ces modèles prennent en compte d'autres effets que les effets génétiques additifs.

Enfin, lorsque l'on compare les méthodes de correction, pour les modèles de PP « graines », on observe une faible augmentation des PA entre les modèles HBLUP<sub>seed1</sub> et HBLUP<sub>seed3</sub>. Pour les modèles de PP « feuilles », on observe une augmentation plus significative des PA entre les modèles HBLUP<sub>leaf2</sub> et HBLUP<sub>leaf3</sub>. A l'échelle du panel entier, la méthode de correction 3 apparaît donc comme la méthode la plus adaptée pour prédire la vigueur précoce.

En conclusion, à l'échelle du panel entier, l'augmentation de la diversité et du nombre d'individus s'est traduite par une augmentation des héritabilités et une meilleure capacité du GBLUP à prédire les traits



de vigueur précoce que les modèles de prédiction phénotypique, à l'exception de quelques traits (SLA, NUTE, SCA et LAR). Pour tous les traits, les spectres acquis sur feuilles ont une meilleure capacité prédictive que les modèles basés sur des spectres acquis sur des graines.

## Eléments de discussion

### Les spectres NIRS capturent peu d'apparentement génétique.

Les spectres NIRS acquis sur graines et sur feuilles ne permettent pas de mettre en évidence une structure du panel similaire à celle obtenue avec les SNP (**Figure III.3.1**). Cette différence est confirmée par la faible corrélation entre les matrices d'apparentement spectral et la matrice d'apparentement génomique, alors que les longueurs d'ondes obtenues sur graines étaient fortement hérétiques. Des observations similaires ont été faites dans les études de Zhu et al. (2021) ; Weiβ et al. (2022) ; Brault et al. (2022) et Roscher-Ehrig et al. (2024). Cette différence peut être liée au fait que les ressemblances/dissemblances spectrales rendent compte d'autres effets que les effets génétiques additifs, dont des effets environnementaux. Dans la littérature, des pistes ont été suggérées afin d'optimiser l'utilisation des spectres dans des démarches de prédiction phénotypique. Ainsi, il est proposé d'éliminer les longueurs d'ondes redondantes, ou de choisir les longueurs d'ondes à utiliser pour le calcul de la matrice d'apparentement spectral sur la base de leur héritabilité, et ce pour améliorer la prise en compte de la structure génomique avec les données spectrales. Jusqu'à présent, peu d'études ont réalisé un filtre sur les longueurs d'ondes selon leur redondance, mais cette démarche n'a pas permis d'améliorer la capacité prédictive des modèles (Galan et al, 2020 ; DeSalvio et al. 2024). D'autres auteurs soulignent l'intérêt de cumuler des spectres de plusieurs natures (spectres acquis sur grains entiers ou sur farine par exemple) pour améliorer la capacité prédictive des modèles, ce qui sous-entend que cumuler des différentes données spectrales permettrait de mieux estimer les apparentements génétiques entre les génotypes au sein de la matrice H (Robert et al, 2022a). Dans notre étude, il pourrait être intéressant de tester l'effet du cumul des données spectrales graines et feuilles pour prédire les traits de vigueur précoce.

### Les spectres sont impactés par des effets environnementaux liés à leurs conditions d'acquisition

Nous avons observé une différence de structure spectrale entre les WOSR et les SOSR. De plus, une corrélation élevée a été obtenue entre la matrice d'apparentement génomique K et la matrice  $H_{leaf1}$  lors de l'utilisation de l'ensemble du panel, alors que la corrélation entre ces matrices est faible (0.06) pour le cluster C1. Ces différences sont expliquées par des effets liés à l'acquisition des spectres et ayant induit une structuration artificielle WOSR/SOSR. Ce phénomène a biaisé les capacités prédictives des modèles à l'échelle du panel entier, menant à une inflation potentielle des PA avec la méthode de



correction 1, comme mentionné par Meyenberg et al. (2024). Ainsi, les auteurs recommandent d'estimer des moyennes ajustées (BLUE) des valeurs phénotypiques et des spectres NIRS afin de corriger pour les effets expérimentaux, ce qui correspond à la méthode de correction 2 investiguée ici.

Il est possible aussi de jouer sur le pré-traitement des spectres pour réduire les bruits et améliorer la capacité prédictive des modèles de prédiction phénotypique. Dans la majorité des études de prédiction phénotypique, les spectres sont centrés et mis à l'échelle, puis un filtre de Savitzky-Golay est appliqué. Cependant, Meyenberg et al (2024) montrent que les paramètres du filtre Savitzky-Golay peuvent être optimisés individuellement pour chaque trait considéré (ordre polynomial, ordre de la dérivée, taille de la fenêtre soit le nombre de longueurs d'ondes considérées) et ainsi permettre une augmentation de la PA des modèles jusqu'à 1500%. Combiner et optimiser les méthodes de prétraitement et de correction des spectres est important, bien que cela nécessite un ajustement spécifique pour chaque modèle, chaque série de spectres et chaque trait à prédire.

Nous avons observé que les modèles de PP permettaient de prédire des traits dont l'héritabilité était très faible à nulle. Ainsi, bien que ce soit l'information comprise dans les spectres NIRS qui permet de prédire ces traits, ce n'est pas l'information génétique qu'elle contient qui contribue à la bonne capacité prédictive du modèle. L'hypothèse sous-jacente est que la nature des spectres, et donc les relations moléculaires qu'ils représentent, consiste en elle-même en un phénotype secondaire proche du phénotype à prédire. C'est un phénomène déjà rapporté dans la bibliographie, notamment pour des traits liés à la composition de la graine chez le blé (Dallinger et al, 2023). Dans notre étude, la quantité d'azote dans les parties aériennes a été prédite à l'aide d'une équation de calibration NIRS (Rolland et al, soumis) qui utilise les mêmes spectres que ceux utilisés pour prédire par PP les traits de vigueur (eux-mêmes calculés pour certains à partir des quantités d'azote). Ainsi, la capacité prédictive des modèles serait surestimée. Dans les programmes de sélection, un nombre important de traits sont obtenus par des modèles de régression PLS utilisant les données spectrales NIRS comme prédicteurs (teneur en huile, teneur en protéines du tourteau, teneur en glucosinolates des graines, teneur en azote des différents organes, ...). Cette bonne capacité des spectres NIRS à prédire ces phénotypes rend compte d'une proximité forte entre l'endophénotype spectral et le phénotype final. On retrouve alors un modèle de PP proche d'une régression phénotype/phénotype où les prédicteurs et le trait à prédire sont estimés sur des organes proches, plutôt qu'un modèle de prédiction « genomic-like » cherchant à prédire les valeurs génétiques des individus sur la base d'une information d'apparentement « spectral » (Robert et al, 2022b).



### **Recommandations pour l'utilisation de la PP pour prédire des traits de vigueur précoce**

Différents avantages ont été mis en avant pour promouvoir l'utilisation de la prédiction phénomique comme alternative ou en complément à la sélection génomique :

- Le moindre coût de l'acquisition des données spectrales, en comparaison aux couts de génotypage
- La capacité des spectres obtenus dans le proche infra-rouge à capturer de l'effet génétique additif et non-additif.

Je vais donner un éclairage sur ces différents points à partir des résultats présentés ci-dessus.

#### La PP alternative ou complément à la GP :

Il a été montré précédemment que les matrices spectrales et génomiques ne rendent pas compte de la même information et sont peu corrélées. La PP n'apparaît donc pas comme une méthodologie équivalente la prédiction génomique. Nous avons aussi mis en évidence que les modèles de prédiction phénomique (modèles de prédiction de la germination présentés dans le chapitre 2) sont peu impactés par la structure génétique de la population. Cette observation est cohérente avec des études menées sur plusieurs espèces (Zhu et al. 2021, 2022 ; Weiβ et al. 2022). Le plus faible impact de la structure de la population sur la PA confère un avantage à l'utilisation des modèles de PP pour le criblage des ressources génétiques, en comparaison à l'utilisation de la prédiction génomique, et permet d'obtenir des modèles adaptés à des populations cibles plus diversifiées. Cependant, nous avons aussi mis en évidence la forte dépendance des données spectrales aux conditions environnementales dans lesquelles les spectres sont acquis, ce qui suggère que la caractérisation spectrale des candidats à la prédiction doit idéalement se réaliser dans les mêmes conditions que la caractérisation spectrale de la population de calibration.

#### Combinaison des données spectrales et génomiques pour la prédiction

Les modèles de PP et de GP apportent des informations complémentaires. Des propositions existent pour valoriser à la fois les données spectrales et les données moléculaires au sein des modèles de prédiction. La première utilisation conjointe des données spectrales et des données moléculaires consiste « simplement » à utiliser la matrice d'apparentement génomique K pour prédire les valeurs spectrales pour chaque longueur d'onde et chaque génotype, comme cela a été le cas avec la méthode de correction 3. Cette méthode de correction revient à utiliser des données spectrales prédites à l'aide d'un modèle de BLUP génomique. Ainsi, pour chaque longueur d'onde, on conserve uniquement la part de la variance d'origine génétique. Nous avons mis en évidence que les PA des modèles de prédiction utilisant cette correction pour les données feuilles n'étaient que très légèrement meilleurs



que les modèles n'utilisant pas les données moléculaires ( $\Delta PA \text{ HBLUP}_{\text{leaf3}} - \text{HBLUP}_{\text{leaf2}} = 0.01$  en moyenne pour tous les traits). Quelques études ont utilisé des BLUP de spectres dans des modèles de prédiction phénotypique (Brault et al. 2022 ; DeSalvio et al. 2024). Brault et al. (2022) observent une augmentation de la PA lors de l'utilisation de cette méthode de correction par rapport à des spectres moyens (équivalent de la méthode de correction 1) pour une population de vigne demi-frères. Ces résultats questionnent l'intérêt de corriger les données spectrales par l'apparentement génomique dans le cadre d'une application en sélection. En effet, les performances des modèles utilisant les moyennes ajustées étaient proches de celles utilisant les valeurs génétiques prédictives, il est important de considérer le cout supplémentaire lié au génotypage de l'ensemble des génotypes de la population de la calibration et de la population de sélection.

Une autre façon de combiner les informations spectrales et génomiques consiste à combiner les deux informations, c'est-à-dire les deux matrices H et K dans le modèle de prédiction (modèles GHBLUP). La faible corrélation entre la matrice d'apparentement génomique et les matrices d'apparentement spectrale suggère que ces matrices portent des informations différentes qui pourraient s'avérer complémentaires. Néanmoins, les gains de PA en combinant ces deux informations sont contrastés (Robert et al. 2022a ; DeSalvio et al. 2024, résultats du chapitre 2 sur la prédiction des traits de germination).



---

## **CHAPITRE IV**

### **DISCUSSION GENERALE ET CONCLUSION**

---



## IV.1. Traits impliqués dans la germination et la vigueur précoce

---

### IV.1.1. Les traits de germination et de vigueur précoce étudiés sont faiblement corrélés

Dans cette thèse, 8 traits de germination sensu stricto et 23 traits de développement et de physiologie ont été utilisés pour décrire la germination et la vigueur précoce respectivement (cf. II et III.2). Des corrélations globalement faibles ont été observées entre les traits de germination et les traits de vigueur précoce, variant entre -0.34 et 0.36 (Tableau IV.1). Les traits de germination ne sont pas ou peu corrélés avec les traits relatifs à l'efficience d'utilisation d'azote (dimension 4) ni avec les traits de vitesse de développement de la surface aérienne (dimension 1). Néanmoins, les traits de germination sensu stricto sont majoritairement corrélés aux traits relatifs à l'acquisition et au développement de la biomasse (dimension 2) et aux traits relatifs à l'allocation de la biomasse (dimension 3). Au sein de chaque catégorie de traits de germination (taux de germination, vitesse de germination, autres traits), les relations significatives avec les traits de vigueur précoce sont soit positives (taux de germination, autres traits) ou soit négatives (vitesse de germination). Ainsi, il ne semble pas y avoir de compromis entre les traits de germination et de vigueur précoce. Afin de mieux caractériser l'implantation, ces corrélations seraient à confirmer en incluant des traits décrivant des stades de développement intermédiaires. Par exemple, plusieurs études se sont intéressées à la biomasse aérienne et racinaire à 7 et 14 jours après le semis, soit équivalent à un stade de croissance de l'hypocotyle voire d'émergence de la plantule (Basnet et al. 2015; Nguyen et al. 2018). Des dispositifs comme ELCONCAM de la plateforme PHENOTIC (Boureau 2020) permettent de caractériser ces stades de croissance via le phénotypage de la longueur et de la vitesse de développement de la radicule et de l'hypocotyle jusqu'à 7 jours après le semis. D'autres traits de germination et de vigueur précoce pourraient également se révéler pertinents à phénotyper. Parmi ces traits, la capacité à réactiver le métabolisme de la graine (Galland et al. 2014), l'efficience d'absorption de l'azote (NUpE, principal contributeur à la NUE durant la phase végétative, Vazquez-Carrasquer et al. 2021), ainsi que des traits d'architecture racinaire (Lecarpentier et al. 2021) et d'accès aux ressources pourraient être envisagés.

**Tableau IV.1. Corrélations entre les traits liés à la germination et les traits liés à la vigueur précoce dans notre étude.**

Germination		Traits relatifs au taux de germination		Traits relatifs à la vitesse de germination					Autres traits		
		GR	AUC	UNI	FG	T20	T80	MGT	VI	ES	TSW
Vigueur précoce											
Dim.1	<b>AUC<sub>L1-L4</sub></b>	ns	ns	-0.24***	ns	ns	-0.17*	ns	0.23***	0.16*	0.35***
	<b>Sp<sub>L1-L4</sub></b>	ns	ns	-0.22**	ns	ns	-0.15*	ns	0.19**	0.14*	0.27***
Dim.2	<b>LA</b>	0.36***	0.34***	-0.24***	-0.17*	-0.34***	-0.29***	-0.31***	0.22**	0.19**	0.17*
	<b>RDW</b>	0.16*	0.17*	-0.19**	ns	ns	-0.2**	-0.14*	0.16*	ns	0.26***
	<b>TDW</b>	0.34***	0.34***	-0.27***	ns	-0.31***	-0.31***	-0.3***	0.23***	0.17*	0.25***
	<b>TNQ</b>	0.3***	0.29***	-0.22**	ns	-0.28***	-0.27***	-0.25***	0.19**	ns	0.2**
	<b>SCA</b>	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Dim.3	<b>LWR</b>	0.21**	0.2**	ns	-0.16*	-0.22**	-0.14*	-0.2**	ns	ns	ns
	<b>SR</b>	0.25***	0.26***	-0.14*	-0.22**	-0.27***	-0.21**	-0.25***	ns	ns	ns
	<b>SNU</b>	0.22***	0.24***	-0.13*	-0.22**	-0.25***	-0.19**	-0.24***	ns	ns	ns
Dim.4	<b>LAR</b>	ns	ns	ns	ns	ns	ns	ns	ns	ns	-0.2**
	<b>SLA</b>	ns	ns	ns	ns	ns	ns	ns	ns	ns	-0.18**
	<b>LASNQ</b>	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
	<b>NUT<sub>E<sub>tot</sub></sub></b>	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.18**
	<b>NUT<sub>E<sub>aer</sub></sub></b>	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.17*

-0.40 -0.20 0 ns 0 0.20 0.40

Les traits de vigueur précoce sont présentés selon les dimensions définies dans le chapitre III.2 (dimensions 1 à 4). Une sélection des traits de vigueur précoce est réalisée de façon à ce que les traits très corrélés entre eux ne soient pas tous présentés ici (ex. AUC, Sp). Les traits de germination sont présentés selon leur signification (taux, vitesse de germination, autres traits). De la même façon que pour les traits de vigueur précoce, les traits très corrélés entre eux ne sont pas tous présentés (ex. GR36, T50).

\*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, ns = non significatif.

#### IV.1.2. Corrélation entre le poids de mille grains et les traits de vigueur précoce

Tandis que le poids de mille grains (TSW), le volume de la graine lors de l'imbibition (VI) et la vitesse d'élongation de la radicule (ES) jusqu'à 16h après germination n'étaient pas ou très faiblement corrélés aux traits de germination *sensu stricto* (Laurençon et al. 2024), ils se révèlent corrélés aux traits de développement de surface foliaire et d'acquisition de ressources et de biomasse (**Tableau IV.1**). De plus, TSW est le seul trait corrélé aux efficiencies d'absorption et d'utilisation de l'azote. A notre connaissance, aucune relation entre le volume d'imbibition ou l'élongation précoce de la radicule et des traits de développement précoce ont été mis en évidence, ni entre TSW et des traits d'efficiency liés à l'azote. De plus, ES n'est pas corrélé significativement à la biomasse racinaire au stade 4-5 feuilles. Ce manque de corrélation suggère que les mécanismes impliqués dans la croissance très précoce de la radicule seraient différents de ceux impliqués dans la croissance précoce de la racine. Suivre le développement de la radicule en utilisant un ensemble de traits fonctionnels dès la percée du tégument jusqu'au stade 4-5 feuilles permettrait d'obtenir un aperçu plus fin du développement précoce de la radicule et des potentiels processus associés. Concernant TSW, plusieurs études ont identifié une corrélation positive avec la biomasse précoce chez le colza (Elliott et al. 2008; Harker et al. 2015) ainsi qu'avec la biomasse de l'hypocotyle chez *Medicago truncatula* (Dias et al. 2011). Continuer à caractériser les stades précoce du développement de manière fine en y intégrant des points cinétiques intermédiaires, comme le stade 7-14 jours après semis permettrait de mettre en évidence des relations entre les traits de germination et de vigueur précoce sur des phases différentes ou à l'inverse, de mettre en évidence des phases indépendantes, qui ne sont pas révélées lorsque l'on se limite à la comparaison de la germination avec le stade 4-5 feuilles. Par ailleurs, comparer ces traits dans différentes conditions environnementales permettrait d'identifier des traits de germination qui pourraient être utilisés comme proxies des traits de vigueur précoce. Ceci permettrait de mettre en lumière des processus communs entre les différentes conditions environnementales et les différents stades.



#### IV.1.3. Quels traits de germination et de vigueur précoce en conditions de contraintes ?

La vigueur est définie comme étant la capacité de la plante à croître et à se développer dans un ensemble de conditions environnementales. Ainsi, il est essentiel d'étudier ces traits et leurs corrélations aussi en condition de stress, uniques ou combinées, afin de mettre en lumière un ensemble de traits cibles pour la sélection, ce set de traits pouvant être potentiellement différent en fonction des conditions environnementales.

Plusieurs études mettent en évidence des modifications de corrélations entre les traits de vigueur en condition non limitantes *versus* en condition limitantes (Wang et al. 2017 ; Zhong et al. 2019). Dans ces études, des traits non corrélés en condition non limitantes, l'étaient en condition de stress. Par exemple, une augmentation de la corrélation entre la biomasse du pivot et les traits d'architecture racinaire a été mise en évidence en conditions de carence azotée en comparaison de conditions non limitantes chez le colza (Wang et al. 2017). Ceci suggère une modification de l'architecture racinaire en condition de carence azotée. En effet, alors que la longueur de la racine pivot semblait être indépendante des autres traits racinaires (longueur totale de la racine, biomasse, volume, ...), en condition de carence azotée la racine pivot semble être plus impliquée dans la biomasse racinaire et le volume exploré par la racine. Zhong et al (2019) ont étudié l'impact du stress hydrique sur la capacité photosynthétiques chez le riz. Ainsi, les auteurs mettent en évidence la corrélation négative entre la tolérance au stress hydrique et l'allocation de l'azote dans les parties photosynthétiques. Ainsi, il y aurait un compromis dans l'allocation de l'azote, impliquant un équilibre entre la croissance et la tolérance au stress hydrique. La dégradation de molécules dont la Rubisco et l'utilisation de protéines solubles permettent l'obtention d'osmolytes, éléments clés pour la tolérance au stress hydrique (Parry et al. 2002 ; Zivcak et al. 2016). Considérer la réponse de la vigueur au stress hydrique est nécessaire car le stress hydrique est le stress abiotique majeur subi lors de l'implantation du colza. Ainsi, dans un premier temps, investiguer la germination et la vigueur précoce en condition de stress hydrique serait particulièrement intéressant afin d'étudier la variabilité des traits de vigueur que nous avons définis. Cependant, étudier la vigueur en réponse à une limitation en eau nécessitera aussi de considérer de nouveaux traits, notamment en lien avec l'efficience d'utilisation de l'eau et l'accès aux ressources hydriques. Ainsi, chez le riz, les traits utilisés pour décrire la vigueur précoce (surface foliaire, biomasse aérienne, nombre de talles) sont anti-correlés aux traits liés à la tolérance au stress hydrique (WUE ou water use efficiency, efficience d'utilisation de l'eau) (Rebolledo et al. 2013). Les auteurs proposent d'étudier plus en détail la relation entre des traits de développement racinaire, d'efficience d'utilisation de l'eau et de vigueur précoce afin d'identifier des compromis potentiels entre certains processus, et à l'inverse des traits conférant une bonne vigueur lors de l'implantation en condition de



stress hydrique. Cela est également valable pour la compétition avec les adventices, où la vigueur confère une capacité à la compétition élevée (Namuco et al. 2009; Asaduzzaman et al. 2014; Hendriks et al. 2024). En effet, la compétition est en partie liée à la capacité à utiliser les ressources nutritives et/ou lumineuses le plus rapidement et le plus efficacement. Ainsi, en cas de fortes densités de plantes dans la parcelle, la compétition pour les ressources va entraîner localement pour la plante des conditions de stress, hydrique et/ou nutritif. Il est donc essentiel de caractériser finement la vigueur à l'implantation ainsi que son contrôle génétique et pour comprendre les mécanismes sous-jacents et apporter des éléments permettant de proposer des stratégies pour l'obtention de cultures vigoureuses pour différentes conditions de culture.

**Tableau IV.2. Co-localisation des QTL de germination et de vigueur précoce identifiés au cours de la thèse et avec la littérature.**

Pour chaque QTL est mentionné le trait, le chromosome et la position physique des marqueurs flanquants, positionnés sur le génome de référence Darmor-Bzh V10 (Rousseau-Gueutin et al. 2020).

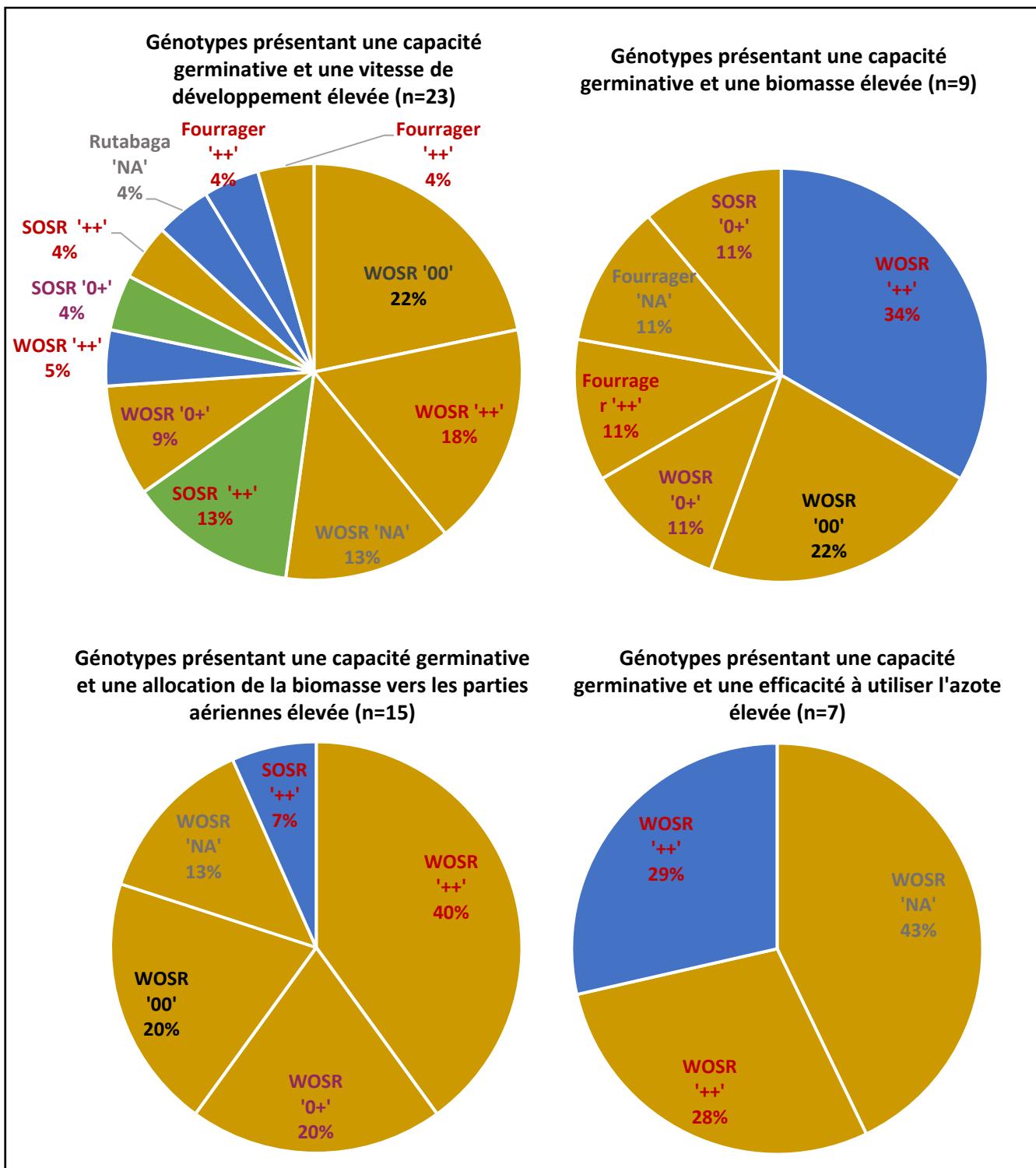
Co-localisations	Littérature	Trait	Chrom.	Position du marqueur flanquant 1	Position du marqueur flanquant 2
		<b>GP36</b>			
	Chapitre II	<b>MGT</b>			
		<b>T20</b>	A06	915357	1194477
		<b>T50</b>			
			<b>AUC<sub>L1-L4</sub></b>		
Entre nos études n°1			<b>AUC<sub>L3-L4</sub></b>		
		Chapitre III.2	<b>Sp<sub>L1-L4</sub></b>	A06	339 984
			<b>Sp<sub>L2-L3</sub></b>		
			<b>Sp<sub>L3-L4</sub></b>		
	Chapitre II	<b>MGT</b>	C05	9233	5469744
Entre nos études n°2	Chapitre III.2	<b>ADW</b>			
		<b>TDW</b>	C05	2 427 632	5 993 034
Avec la littérature n°1	Hatzig et al. 2015	<b>GR72</b>	A03	7 082 702	7 098 892
	III.2	<b>AUC<sub>L1-L2</sub></b>	A03	6 348 789	10 116 857
	Hatzig et al. 2015	<b>GR72</b>	A04	4 102 377	4 144 135
		<b>GR72</b>	A04	4 260 229	4 302 660
Avec la littérature n°2		<b>ADW</b>	A04	3 344 129	5 734 512
	III.2	<b>LA</b>	A04	3 344 129	5 734 512
		<b>TDW</b>	A04	3 344 129	5 734 512

## IV.2. Architecture génétique de la germination et de la vigueur précoce

---

Nous avons identifié un ensemble de QTL à effets faibles à modérés pour les traits impliqués dans la germination et la vigueur précoce, illustrant le fait qu'ils soient sous contrôle polygénique. Ceci est cohérent avec les études antérieures chez colza pour la germination (Hatzig et al. 2015; Luo et al. 2017; Nguyen et al. 2018) et pour la vigueur précoce (Knoch et al. 2020; Li et al. 2021; Haelterman et al. 2024).

En combinant les résultats obtenus dans les chapitres II et III.2 ainsi qu'avec la littérature, nous avons pu mettre en évidence quatre co-localisations entre QTL de germination et QTL de vigueur précoce (**Tableau IV.2**). La première co-localisation est positionnée sur le chromosome A06, entre un QTL lié aux traits de germination GP36, MGT, T20 et T50 et un QTL lié aux traits développement foliaire (AUC et Sp). Bien que les traits soient des traits de vitesse dans les deux cas, nous n'avons pas identifié de corrélation entre eux à l'échelle phénotypique (**Tableau IV.1**). La seconde co-localisation est positionnée sur le chromosome C05, où un QTL impliqué dans MGT co-localise avec un QTL de vigueur précoce impliqué dans le contrôle de la biomasse (ADW et TDW). Ces traits sont modérément corrélés ( $r=-0.3$ ). De plus, nous avons également identifié deux co-localisations entre des QTL mis en évidence par Hatzig et al. (2015) et les QTL que nous avons identifiés pour la vigueur précoce. Ces deux co-localisations correspondent à des QTL impliqués dans le taux de germination final (après 72h d'imbibition) et qui co-localisent avec un QTL de  $AUC_{L1-L2}$  sur le chromosome A03 (QTL.A03.1) et avec un QTL de biomasse (ADW, TDW et LA) sur le chromosome A04 (QTL.A04.1). L'ensemble de ces co-localisations permettent d'émettre l'hypothèse de mécanismes sous-jacents communs entre les différents traits, ou de l'existence de gènes causaux communs ou très proches physiquement. (Solovieff et al. 2013; Mackay and Anholt 2024). Le déséquilibre de liaison (DL) dans notre population étant étendu, des QTL de grande taille (en moyenne 3Mb) ont été identifié. Le grand nombre de gènes sous les QTL ne permet pas d'identifier précisément des gènes candidats pour ces traits. Etudier du matériel présentant un DL moins étendu, et qui présentent de la variabilité pour les traits d'implantation pourraient permettre de caractériser plus finement l'architecture génétique de ces traits. De plus, la caractérisation de l'effet d'un ou de plusieurs loci sur les traits de biomasse précoce et de germination pourrait être obtenue à partir de populations quasi-isogéniques et contrastées pour ces traits.



**Figure IV.1.** Répartition en pourcentage des génotypes les plus performants pour la germination et des traits de vigueur précoce, selon leur appartenance aux clusters génomiques, aux types de colza et la qualité de la graine.

Le cluster génétique est représenté par la couleur du fond (marron = C1, vert = C2, bleu = C3). La couleur de la police indique la qualité de la graine des génotypes. Ainsi, les génotypes '00', '++', '0+' et 'NA' sont représentés en noir, rouge, violet et gris respectivement.

## IV.3. Valorisation de la diversité génétique pour améliorer la germination et la vigueur précoce

---

### IV.3.1. Améliorer la germination et la vigueur précoce par la remobilisation de la diversité génétique

Des différences de performances selon les clusters avaient été observées pour les traits de germination (Laurençon et al. 2024), où les génotypes du cluster C1 (soit les génotypes WOSR majoritairement) semblaient être les plus performants. Cependant, l'absence d'effet de structure sur les traits de vigueur précoce suggère qu'aucun cluster génomique ne semble être plus performant que les autres. Ainsi, cela souligne l'intérêt d'aller chercher des génotypes d'intérêt pour apporter des allèles conférant une amélioration de la germination et de la vigueur précoce dans l'ensemble de la diversité génétique disponible. Des génotypes performants à la fois pour la germination et pour des composantes spécifiques de la vigueur précoce ont été identifiés (**Figure IV.1, Tableau IV.3**). Les génotypes ont été identifié à partir des ACP réalisées sur les traits de germination ou de vigueur précoce. Ceux présentant une capacité germinative élevée avaient préalablement été identifiés par clustering (*cf. II*). Pour les traits de vigueur précoce, les génotypes les plus discriminés par les traits (position du génotype >2 pour une dimension) ont été considéré comme étant les plus extrêmes, et ce pour chaque dimension. La majorité des génotypes identifiés présentait une capacité germinative élevée et des valeurs extrêmes pour une seule dimension de la vigueur précoce (**Figure IV.1**). Parmi ces génotypes, la majorité était des génotypes du cluster génomique C1 et/ou des génotypes ‘++’. Ainsi, le choix de géniteurs d'intérêt pour le pre-breeding dépendra majoritairement de profils de « vigueur » ciblés, en particulier concernant les traits de vigueur précoce. De manière intéressante, quelques génotypes (8 seulement) présentaient une bonne capacité germinative et des valeurs élevées pour plusieurs dimensions de la vigueur précoce (**Tableau IV.3**). Parmi ces 8 génotypes, la moitié était des colzas d'hiver fourragers. Ces génotypes seraient ainsi intéressants pour améliorer à la fois la germination et un ensemble de traits de vigueur précoce. Cela pourrait être expliqué par le fait que ce type de colza a été sélectionné pour un usage en tant que fourrage, et donc en maximisant la production de biomasse aérienne et de surface foliaire pendant la phase végétative. En effet, les colzas fourragers sont généralement semés entre mars et août et ensuite sont soit directement pâturés par les troupeaux pendant l'automne et l'hiver ou récoltés à la sortie de l'hiver. De plus, la majorité des génotypes les plus performants sont des colzas plus anciens qui présentent un profil de qualité de la graine ‘++’. Ainsi, la relation entre la qualité de la graine (et donc le potentiel impact de la sélection moderne) et la vigueur serait à investiguer plus en profondeur. Aucun génotype n'a été identifié

**Tableau IV.3. Description des génotypes combinant une capacité germinative élevée et des valeurs élevées pour plusieurs dimensions de vigueur précoce.**

Les génotypes sont catégorisés selon les types de performances. Pour chaque génotype est indiqué le cluster génétique auquel il appartient, le type de colza et la qualité de la graine. Les chiffres entre parenthèses (1,2,3) représentent les dimensions de l'ACP (**Figure III.2.3**) auxquelles sont associés les traits de vigueur précoce considérés.

Nom	Cluster génomique	Type	Qualité
Génotypes présentant une capacité germinative, une vitesse de développement, une capacité à acquérir du N et du C élevée, ainsi qu'une allocation préférentielle de la biomasse dans les parties aériennes (1,2,3)			
JANETZKI	1	SOSR	++
Génotypes présentant une capacité germinative, une vitesse de développement, une allocation préférentielle de la biomasse dans les parties aériennes et une efficacité élevée à utiliser l'azote (1,3,4)			
Parapluie	1	Fourrager	++
Génotypes présentant une capacité germinative, une Vitesse de développement et une capacité à acquérir du N et du C élevée (1,2)			
Samo	1	Fourrager	++
Silona	1	Fourrager	NA
Génotypes présentant une capacité germinative et une vitesse de développement élevée, ainsi qu'une allocation préférentielle de la biomasse dans les parties aériennes (1,3)			
Victor	1	WOSR	++
Génotypes présentant une capacité germinative et une vitesse de développement élevée ainsi qu'une efficacité élevée à utiliser l'azote (1,4)			
CHINE 29	3	SOSR	++
Génotypes présentant une capacité germinative et une capacité à acquérir du N et du C élevée, ainsi qu'une allocation préférentielle de la biomasse dans les parties aériennes (2,3)			
Liragrun	1	Fourrager	++
Génotypes présentant une capacité germinative élevée, une allocation préférentielle de la biomasse dans les parties aériennes et une efficacité élevée à utiliser l'azote (3,4)			
Dippes	1	WOSR	++

comme présentant une capacité germinative élevée et une efficience à acquérir et à utiliser l'azote et le carbone (dimensions 2 et 4). Investiguer plus de diversité génétique, notamment en utilisant des colzas fourragers et des légumes de l'espèce *Brassica napus* (rutabaga, chou kale), ainsi que du matériel semi-synthétique (obtenu via le croisement entre de génotypes d'espèces apparentées du colza sauvages ou cultivés et des génotypes de colza), pourrait pallier à ce manque de diversité. L'étude des relations génétiques entre ces traits permettrait de mettre en lumière de potentiels effets pléiotropes négatifs expliquant le manque de génotypes performants combinant ces traits au sein des colzas oléagineux. Ou à l'inverse, en cas de non corrélation ou de corrélation positive, d'utiliser des ressources performantes en pré-breeding afin d'aboutir à des génotypes combinant les phénotypes favorables.

#### **IV.3.2. La prédiction phénotomique et la prédiction génomique pourraient permettre de remobiliser de la diversité génétique dans les populations de pre-breeding**

Dans cette étude, nous avons identifié un ensemble de génotypes de type WOSR ou SOSR à valoriser en sélection. Remobiliser ce matériel ainsi que d'autres ressources génétiques (génotypes locaux, haploïdes doublés semi-synthétiques, ...) permettrait d'augmenter la variabilité génotypique et d'incorporer des fonds génétiques conférant des phénotypes favorables dans les pools de sélection (Simmonds 1993; Glaszmann et al. 2010; Longin and Reif 2014; Wang et al. 2017). Les approches de prédiction génomique et phénotomique sont des outils qui permettent d'identifier les ressources à valoriser en pre-breeding (*cf. IV.4.1*). Nous avons mis en évidence pour les traits de germination que la PP semblait être moins impactée par la structure génomique du panel que la GP et que les valeurs phénotypiques prédites par PP sont plus corrélées aux valeurs phénotypiques mesurées que les valeurs génétiques estimées. Autrement dit, la PP permet de prédire des phénotypes proches de ceux mesurés. Cela serait néanmoins à confirmer pour les traits de vigueur précoce. Nous faisons ainsi l'hypothèse que cette méthode pourrait être utilisée afin de remplacer partiellement le phénotypage lors de la création de populations de pré-breeding. La PP a de multiples avantages : cette méthode est i) haut-débit, ii) peu chère, iii) moins influencée par la structure de la population de la GP. Ainsi, cette méthode pourrait être utilisée afin de cibler de larges collections de ressources génétiques pour certains phénotypes d'intérêt, notamment ceux connus pour être sous contrôle polygénique. Jusqu'à présent, la PP a été majoritairement étudiée dans le cadre d'une utilisation dans les schémas de sélection, c'est-à-dire pour être implantée au sein du germplasme élite (Rincent et al. 2018; Robert et al. 2022a; Brault et al. 2022; Roscher-Ehrig et al. 2024). Notre étude montre son intérêt potentiel



pour le criblage de populations de diversité pour la germination et la vigueur précoce. Néanmoins, d'autres études nécessiteraient d'être menées afin d'investiguer plus en détail son intérêt pour le développement de populations de pre-breeding. De plus, il serait pertinent d'évaluer la capacité prédictive des modèles de PP pour prédire des traits d'intérêt, notamment dans un contexte d'agroécologie (efficiences impliquées dans la tolérance aux stress abiotiques, résistances aux maladies et bioagresseurs, phénologie, aptitude à l'association, architecture racinaire, ...). Et plus particulièrement lorsque ces traits ont été acquis dans des conditions environnementales contraignantes. Les corrélations entre les traits d'intérêt et des traits secondaires pourraient être modifiées selon les conditions environnementales, impactant ainsi la capacité prédictive des modèles (*cf. III.3. et IV.1.3.*). Après la création des populations de pré-breeding, la prédiction génomique pourrait être utilisée pour choisir les individus à incorporer au matériel de sélection (Gorjanc et al. 2016).

**Tableau IV.4. Capacités prédictives moyennes des modèles en utilisant les individus de la population entière (Scénario<sub>whole</sub>) pour l'ensemble des traits d'implantation utilisés au cours de la thèse.**

	H <sup>2</sup>	GBLUP	HBLUP seed1	HBLUP seed3	HBLUP leaf2	HBLUP leaf3	R <sup>2</sup> des QTL identifiés (nombre de QTL)
GP36	<b>0.69</b>	<b>0.59</b> (0.1)	0.49 (0.1)				6.99 - 8.69 (7)
T20	<b>0.69</b>	<b>0.59</b> (0.09)	0.5 (0.1)				
MGT	<b>0.68</b>	<b>0.58</b> (0.1)	0.51 (0.11)				6.30 - 8.46 (9)
T50	<b>0.65</b>	<b>0.58</b> (0.09)	0.46 (0.1)				7.04 - 8.78 (6)
T80	<b>0.61</b>	<b>0.53</b> (0.1)	0.45 (0.11)				
TDW	<b>0.52</b>	<b>0.4</b> (0.14)	0.25 (0.13)	0.25 (0.13)	0.34 (0.14)	0.35 (0.13)	4.12 - 4.65 (3)
ADW	<b>0.52</b>	<b>0.42</b> (0.14)	0.28 (0.13)	0.24 (0.15)	0.35 (0.14)	0.39 (0.12)	4.03 - 4.80 (3)
LA	<b>0.52</b>	<b>0.41</b> (0.11)	0.27 (0.13)	0.26 (0.12)	0.32 (0.13)	0.33 (0.14)	8.57 (1)
FG	<b>0.51</b>	<b>0.4</b> (0.13)	0.32 (0.11)				
GP	<b>0.46</b>	<b>0.54</b> (0.09)	0.49 (0.11)				
SNQ	<b>0.45</b>	<b>0.27</b> (0.14)	0.008 (0.1)	0.006 (0.09)	0.14 (0.15)	0.2 (0.12)	2.78 (1)
UNI	<b>0.43</b>	<b>0.4</b> (0.11)	0.36 (0.12)				
RDW	<b>0.39</b>	0.26 (0.16)	-0.13 (0.1)	-0.14 (0.1)	0.25 (0.14)	<b>0.27</b> (0.14)	2.99 - 6.39 (3)
LASNQ	<b>0.2</b>	0.26 (0.13)	0.15 (0.13)	0.16 (0.14)	0.23 (0.13)	<b>0.28</b> (0.12)	8.92 - 11.83 (2)
NUtE <sub>aer</sub>	<b>0.18</b>	0.33 (0.16)	0.21 (0.14)	0.23 (0.14)	<b>0.51</b> (0.13)	0.45 (0.14)	4.03 - 5.81 (4)
SCA	<b>0.17</b>	0.21 (0.13)	0.11 (0.13)	0.14 (0.14)	<b>0.35</b> (0.13)	0.3 (0.12)	3.37 (1)
NUtE <sub>tot</sub>	<b>0.15</b>	0.3 (0.17)	0.19 (0.13)	0.21 (0.13)	<b>0.51</b> (0.14)	0.46 (0.13)	4.19 - 6.33 (4)
LWR	<b>0.15</b>	<b>0.21</b> (0.16)	-0.01 (0.09)	-0.03 (0.12)	0.01 (0.13)	0.05 (0.12)	0.65 - 1.38 (2)
Sp L1-L4	<b>0.13</b>	0.11 (0.13)	-0.02 (0.08)	-0.03 (0.09)	0.16 (0.14)	<b>0.19</b> (0.16)	1.77 - 8.67 (2)
AUC L1-L4	<b>0.13</b>	0.15 (0.14)	0.01 (0.1)	0.03 (0.09)	0.25 (0.13)	<b>0.27</b> (0.11)	11.56 - 13.98 (2)
SR	<b>0.1</b>	<b>0.28</b> (0.13)	0.18 (0.16)	0.19 (0.14)	0.24 (0.14)	0.23 (0.14)	5.79 (1)
SNU	<b>0.07</b>	<b>0.22</b> (0.13)	0.1 (0.13)	0.12 (0.12)	0.17 (0.14)	0.18 (0.16)	3.42 (1)
LAR	<b>0.03</b>	-0.05 (0.1)	0.06 (0.13)	0.07 (0.13)	0.2 (0.14)	<b>0.22</b> (0.13)	
SLA	<b>0.02</b>	-0.04 (0.1)	0.06 (0.11)	0.07 (0.12)	0.21 (0.15)	<b>0.24</b> (0.12)	
AUC		<b>0.56</b> (0.09)	0.5 (0.08)				

L'héritabilité de chaque trait (H<sup>2</sup>) est indiquée. Les données relatives aux modèles développés pour les traits de germination et les traits de vigueur précoce sont indiquées par un fond orange ou gris respectivement. Les écarts-types des capacités prédictives des modèles sont représentés entre parenthèses. Les modèles les plus performants pour chaque trait sont écrits en rouge et en gras.

## IV.4. Prédiction des traits d'implantation et intérêt de la prédition phénomique pour ces traits

---

### IV.4.1. Les modèles de prédiction génomique et phénomique sont pertinents pour les traits d'implantation, mais restent à optimiser

Pour les modèles de prédiction génomique et phénomique, des capacités prédictives modérées à élevées ont été obtenues pour les traits de germination. Néanmoins, elles étaient faibles à modérées pour les traits de vigueur précoce (**Tableau IV.4**). Lorsque les deux approches sont comparées, les modèles de prédiction génomique étaient les plus performants pour prédire les traits possédant une heritabilité modérée à élevée (soit entre 0.4 et 0.7). Nous notons néanmoins que les modèles utilisant des spectres avaient des performances proches des modèles de prédiction génomique. A l'inverse, pour les traits présentant une heritabilité faible, les modèles de prédiction phénomique (utilisant des spectres acquis sur des tissus proches du stade de développement phénotypé) étaient les plus performants. Pour les traits présentant une heritabilité modérée à élevée, jusqu'à 9 QTL ont été identifiés. Ainsi, la bonne capacité prédictive des modèles de GP dans son ensemble suggère qu'une multitude d'autres polygènes impliqués dans le contrôle de ces traits n'ont pas été mis en lumière par la GWAS, confirmant ainsi la nature polygénique des traits de germination et de vigueur précoce.

Une amélioration de la capacité prédictive des traits passe donc d'abord par une optimisation des méthodes de phénotypage (outils et plans expérimentaux) pour permettre une augmentation des valeurs d'heritabilité. En effet, le phénotypage de la germination a été réalisé en conditions contrôlées et de manière automatisée, limitant les effets environnementaux et les effets notateurs. A l'inverse, le phénotypage de la vigueur précoce a été réalisé en conditions semi-contrôlées (gestion de l'irrigation et de la fertilisation uniquement) et en plusieurs séries d'expérimentations. Les effets environnementaux et les effets notateurs ont ainsi impacté le phénotypage, menant à une estimation plus faible de l'heritabilité pour ces traits. Ensuite, la capacité prédictive des modèles pourrait aussi être améliorée en optimisant le choix des génotypes de la population d'entraînement en fonction de leur apparentement. Dans cette étude, les génotypes utilisés dans les panels d'entraînement et de test ont été choisis aléatoirement. Or, la composition des panels d'entraînement peut être optimisée à partir de métriques telles que le coefficient de détermination CDmean (Rincent et al. 2012). Ce coefficient permet par exemple de minimiser l'apparentement entre les génotypes dans la population d'entraînement, et de maximiser l'apparentement entre les génotypes de la population d'entraînement et de la population test, contribuant ainsi à améliorer la prédiction des modèles. Enfin,

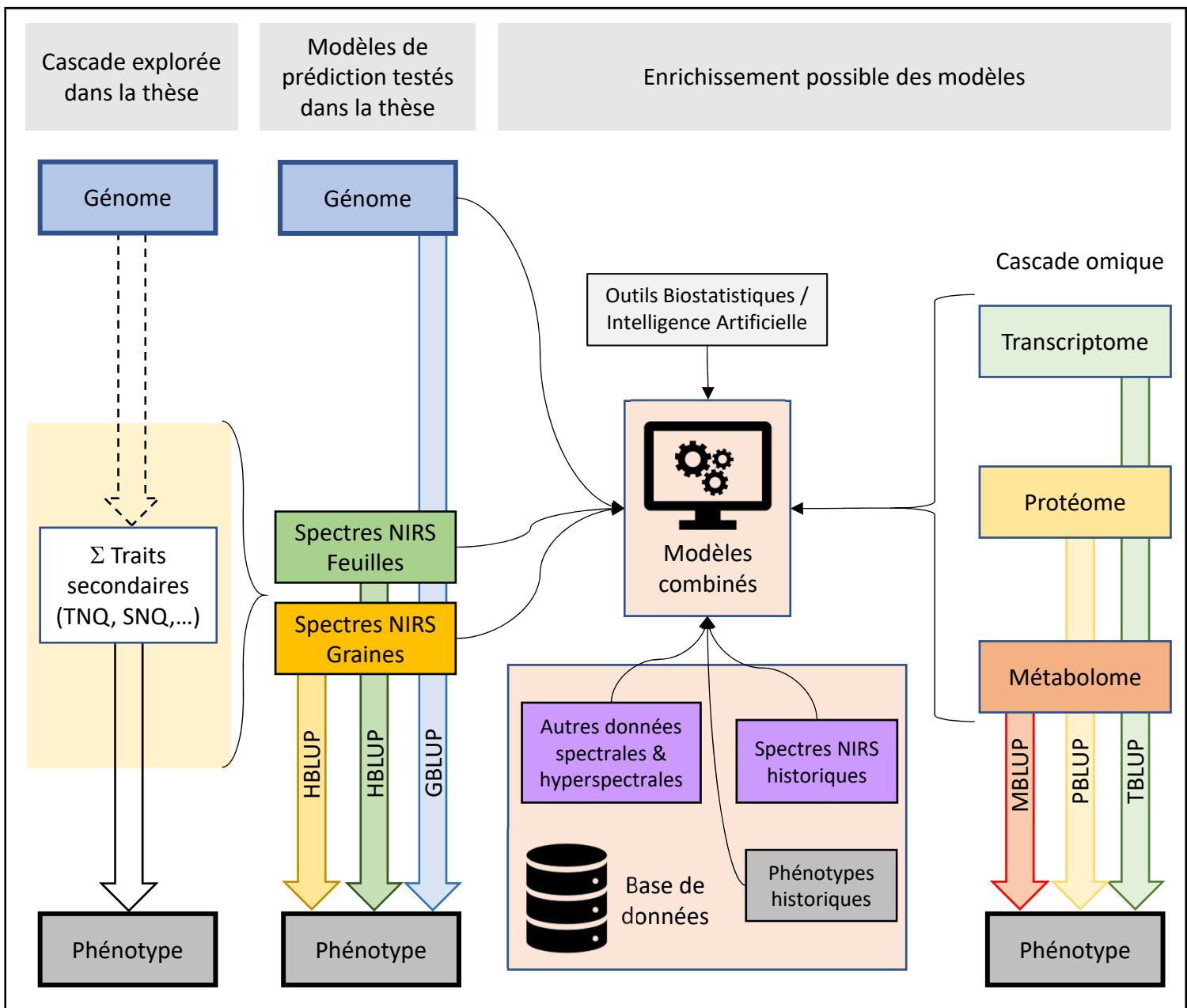


en cas d'utilisation de population structurée, la structure peut être prise en compte dans les modèles, permettant ainsi de réduire les biais de prédition. Cela peut être réalisé en prenant en compte les premières composantes principales des analyses de structure en effet fixe dans les modèles de prédition génomique (Daetwyler et al. 2012; Crossa et al. 2016). Malgré le fait que la prédition phénomique semble moins sensible à la structure que la GP, la considérer dans les modèles pourrait améliorer la PA. Enfin, l'amélioration de la prédition des traits pourrait également être obtenue à travers l'utilisation d'autres types d'endophénotypes.

#### **IV.4.2. La combinaison d'un ensemble d'endophénotypes pourrait permettre de mieux capturer le GxE et les mécanismes impliqués dans la vigueur**

Plusieurs études se sont intéressées à l'utilisation de données de transcriptomique, de protéomique et de métabolomique, seules, combinées entre elles ou combinées avec des données génomiques pour la prédition de traits agronomiques. Un gain de capacité prédictive faible à modéré peut être observé en utilisant ces données, comparé à un modèle de prédition génomique simple (Xu et al. 2016; Westhues et al. 2017; Schrag et al. 2018; Wu et al. 2022; Ali et al. 2024). Pour certains traits comme le rendement en biomasse du maïs fourrage, les données de transcriptomiques seules permettaient de mieux prédire que les données génomiques (Westhues et al. 2017). Néanmoins, dans la plupart des études, le fort effet environnemental capturé par les données omiques peut être à l'origine d'une surestimation de la PA lorsque les données omiques sont acquises dans les mêmes conditions environnementales que la variable à prédire. Dans le cas où les prédicteurs omiques sont acquis dans des conditions différentes des conditions cibles, les modèles de PP ne permettent qu'un faible gain de PA par rapport aux modèles de prédition génomique. Ainsi chez le maïs il a été montré que pour prédire des traits phénotypés au champ, combiner des prédicteurs omiques et génomiques acquis dans des conditions contrôlées ne permettait qu'un faible gain de PA par rapport à un modèle utilisant uniquement des données génomiques (Ali et al. 2024).

La prédition phénomique constitue une perspective intéressante pour valoriser en prédition d'autres endophénotypes que ceux discutés précédemment. En effet, en plus des données omiques, les modèles de PP peuvent être adaptés à l'utilisation de données i) spectrales acquises dans le proche infra-rouge, et ce pour des environnements proches des environnements dans lesquelles sont acquises les variables à prédire (démarche suivie dans le cadre de la thèse), ou dans des environnements différents (Robert et al. 2022b) ; ii) hyperspectrales acquises au champ, en condition de peuplement, à l'aide de capteurs embarqués sur des drones par exemple. Plusieurs études utilisant ces données



**Figure IV.2. Représentation des modèles utilisés durant la thèse et des données potentielles à utiliser pour les enrichir.**

La première colonne représente la cascade omique utilisée durant la thèse. Nous avons considéré que les spectres NIRS permettaient de capturer des traits secondaires. La seconde colonne représente des modèles de prédiction utilisés durant la thèse. Les données génomiques (GBLUP) et spectrales (HBLUP) peuvent être utilisés pour prédire les phénotypes. Enfin, une voie d'enrichissement des modèles est proposée dans la troisième colonne. Ainsi, les données omiques (issues du transcriptome, protéome et du métabolome) pourraient être utilisées seules pour prédire les phénotypes (TBLUP, PBLUP et MBLUP respectivement), ou combinées dans un modèle, avec des données spectrales et génomiques.

dans des modèles de PP notent que les PA obtenues étaient similaires aux PA des modèles de GP (Montesinos-López et al. 2017; Adak et al. 2023; Roth et al. 2023; Maggiorelli et al. 2024). Enfin, la PP ouvre aussi la possibilité de valoriser des données acquises antérieurement, par exemple en réinvestiguant les données de NIRS historiques acquises lors de l'évaluation de certains traits comme la teneur en protéines ou la teneur en huile dans les programmes de sélection. Cependant, de la même manière que pour les données omiques, les spectres sont majoritairement impactés par l'environnement et faiblement par l'apparentement. Ainsi, il est nécessaire de pouvoir corriger les effets environnementaux en prenant en compte les informations contenues dans les métadonnées liées aux données spectrales. Nous avons pu montrer dans la partie **III.3** que pour pouvoir effectuer cette correction, il est nécessaire d'avoir accès i) aux plans expérimentaux dans lesquels les données spectrales de la population de calibration ont été acquises (blocs (in)complets randomisés, design augmenté, ...), ii) d'avoir des génotypes en commun entre chaque expérimentation quand l'acquisition s'est déroulée en plusieurs séries et iii) d'avoir des informations sur la date (conditions environnementales lors de l'acquisition des spectres) et sur la méthode d'acquisition des spectres (appareil, contenant, ...). Disposer de métadonnées constitue un enjeu majeur pour la valorisation et l'utilisation de spectres historiques en prédition phénomique. De plus, dans le cadre où les prédictions seraient réalisées pour différents environnements et utilisant des données phénotypiques et omiques historiques, la modélisation de matrices de covariances génétiques mais aussi non génétiques (environnementales) est également nécessaire pour prendre en compte l'environnement cible pour lequel la prédition est réalisée et ainsi prédire le GxE (Hu et al. 2021; Crossa et al. 2024). Actuellement, la plupart des études privilégiennent des dispositifs dédiés garantissant des conditions homogènes à l'acquisition des données spectrales (Robert et al. 2022a).

L'utilisation et la combinaison de données génomiques et endophénotypiques (que ce soit des données omiques ou spectrales) doit permettre d'avoir une approche plus globale dans la compréhension des traits complexes. En effet, ces données permettent d'avoir accès à l'ensemble de la cascade de réponse, du génome jusqu'au phénotype exprimé (**Figure IV.2**). Ces données peuvent être utilisées à la fois pour la prédition de traits complexes, mais aussi pour contribuer à leur dissection et à l'identification des processus sous-jacents. Pour cela, des modèles de deep learning en sélection génomique ont été développés depuis quelques années (Montesinos-López et al. 2021). Néanmoins, à notre connaissance, aucun modèle de prédition basé sur de l'intelligence artificielle (AI) et cumulant des données omiques et génomiques n'a été développé. Récemment, Crossa et al. (2024) proposaient d'utiliser des données historiques issues des schémas de sélection et d'utiliser de l'AI pour combiner un ensemble de données omiques et des covariables environnementales afin d'identifier des motifs génétiques. Ces motifs, communs aux génotypes apparentés, permettraient de mieux



comprendre les interactions GxE et ainsi de mieux prédire dans des environnements non étudiés. Ces données pourraient également permettre d'identifier des processus impliqués dans la régulation des traits complexes. Par exemple, une base de données cumulant des données génomiques, de transcriptomique et d'annotation de gènes a été développée pour plusieurs espèces de légumineuses (Imbert et al. 2023). Cette base permet de cumuler des données sur plusieurs espèces et de développer des graphes de connaissances, permettant notamment de transférer des connaissances acquises sur une espèce (par exemple *Medicago truncatula*) sur d'autres espèces par orthologie et synthénie. Ce type de données (génomiques, transcriptomiques, protéomiques,...) ont également permis de mettre en évidence des gènes et des réseaux de gènes chez le blé, qui seraient associés par exemple à la fois à la couleur des grains et à la germination sur pieds (Hassani-Pak et al. 2021). Ainsi, combiner de tels larges jeux de données acquis aux différentes échelles omiques pourrait permettre d'aller plus loin dans l'investigation des mécanismes impliqués dans l'implantation et dans la vigueur.



## Conclusion

---

L'ensemble des travaux présentés dans ce manuscrit ont permis d'identifier des traits fonctionnels impliqués dans la germination et la vigueur précoce, leurs relations ainsi que leur déterminisme génétique, et ce en utilisant une large gamme de diversité génétique de colza. L'identification de QTL pour ces traits, ainsi que leur prédiction par des modèles de GP et de PP ouvrent des perspectives sur la compréhension de l'architecture génétique de ces traits et leur utilisation en pre-breeding. Ces données ont été acquises dans des conditions non limitantes pour accéder spécifiquement à ces composantes impliquées dans la réussite de l'implantation de la culture.

Avec le changement climatique, les conditions environnementales vont devenir de plus en plus variables, et ce dès la levée. Ainsi, les traits identifiés dans cette étude pourraient ne plus s'avérer les seuls pertinents à prendre en compte pour assurer un bon développement, en cas de stress combinés intenses (inondations, sécheresses prolongées, fortes pressions biotiques, ...). La plasticité de ces traits en lien avec différentes combinaisons de stress reste à investiguer. De plus, cibler une large gamme de ressources génétiques est une étape essentielle pour identifier des fonds génétiques adaptés aux conditions environnementales. Les travaux réalisés dans la thèse ont notamment permis l'identification de matériel ancien et de matériel fourrager comme étant potentiellement vigoureux pour l'implantation, et donc comme étant des sources porteuses de variants d'intérêt pour ces traits. Aller plus loin, en criblant des ressources génétiques de *Brassica napus* de type légume et des ressources génétiques d'espèces apparentées, constituerait une prochaine étape dans la découverte de variants conférant une vigueur à l'implantation.



---

## **REFERENCES BIBLIOGRAPHIQUES**

---



## Références bibliographiques

---

- Adak A, Murray SC, Anderson SL (2023) Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. *G3-Genes Genomes Genet* 13:jkac294. <https://doi.org/10.1093/g3journal/jkac294>
- Albrecht T, Wimmer V, Auinger H-J, et al (2011) Genome-based prediction of testcross values in maize. *Theor Appl Genet* 123:339–350. <https://doi.org/10.1007/s00122-011-1587-7>
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Ali B, Huguenin-Bizot B, Laurent M, et al (2024) High-dimensional multi-omics measured in controlled conditions are useful for maize platform and field trait predictions. *Theor Appl Genet* 137:175. <https://doi.org/10.1007/s00122-024-04679-w>
- Allender CJ, King GJ (2010) Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biol* 10:54. <https://doi.org/10.1186/1471-2229-10-54>
- Allier A, Teyssèdre S, Lehermeier C, et al (2020) Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genomics* 21:349. <https://doi.org/10.1186/s12864-020-6756-0>
- Arguello-Blanco MN, Sneller CH (2023) The effect of cycles of genomic selection on the wheat (*T. aestivum*) genome. *Theor Appl Genet* 136:70. <https://doi.org/10.1007/s00122-023-04279-0>
- Arnold SJ (1983) Morphology, Performance and Fitness. *Am Zool* 23:347–361. <https://doi.org/10.1093/icb/23.2.347>
- Asaduzzaman M, Luckett D, Cowley R, et al (2014) Canola cultivar performance in weed-infested field plots confirms allelopathy ranking from in vitro testing. *Biocontrol Sci Technol* 24:1394–1411. <https://doi.org/10.1080/09583157.2014.942596>
- Astle W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. *Stat Sci* 24:451–471. <https://doi.org/10.1214/09-STS307>
- Aubertot JN, Dürr C, Kiêu K, Richard G (1999) Characterization of sugar beet seedbed structure. *Soil Sci Soc Am J* 63:1377–1384. <https://doi.org/10.2136/sssaj1999.6351377x>
- Bac-Molenaar JA, Vreugdenhil D, Granier C, Keurentjes JJB (2015) Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. *J Exp Bot* 66:5567–5580. <https://doi.org/10.1093/jxb/erv176>
- Basnet RK, Duwal A, Tiwari DN, et al (2015) Quantitative trait locus analysis of seed germination and seedling vigor in *Brassica rapa* reveals QTL hotspots and epistatic interactions. *Front Plant Sci* 6:1032. <https://doi.org/10.3389/fpls.2015.01032>
- Bayer PE, Hurgobin B, Golicz AA, et al (2017) Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnol J* 15:1602–1610. <https://doi.org/10.1111/pbi.12742>
- Bayer PE, Scheben A, Golicz AA, et al (2021) Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids. *Plant Biotechnol J* 19:2488–2500. <https://doi.org/10.1111/pbi.13674>
- Becker HC, Engqvist GM, Karlsson B (1995) Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. *Theor Appl Genet* 91:62–67. <https://doi.org/10.1007/BF00220859>
- Bettey M, Finch-Savage WE, King GJ, Lynn JR (2000) Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in *Brassica oleracea*. *New Phytol* 148:277–286. <https://doi.org/10.1046/j.1469-8137.2000.00760.x>
- Bewley JD (1997) Seed germination and dormancy. *Plant Cell* 9:1055–1066. <https://doi.org/10.1105/tpc.9.7.1055>
- Bianchetti G (2021) Impacts de stress biotique et abiotique sur l’acquisition de la qualité nutritionnelle et physiologique de la graine de colza (*Brassica napus* L.). Thèse de doctorat, Rennes, Agrocampus Ouest



## Références bibliographiques

---

- Bianchetti G, Clouet V, Legeai F, et al (2024) Identification of transcriptional modules linked to the drought response of *Brassica napus* during seed development and their mitigation by early biotic stress. *Physiol Plant* 176:e14130. <https://doi.org/10.1111/ppl.14130>
- Bilichak A, Kovalchuk I (2016) Transgenerational response to stress in plants and its application for breeding. *J Exp Bot* 67:2081–2092. <https://doi.org/10.1093/jxb/erw066>
- Boideau F, Pelé A, Tanguy C, et al (2021) A modified meiotic recombination in *Brassica napus* largely improves its breeding efficiency. *Biology* 10:771. <https://doi.org/10.3390/biology10080771>
- Boter M, Calleja-Cabrera J, Carrera-Castaño G, Wagner G, Hatzig SV, Snowdon RJ, Legoahec L, Bianchetti G, Bouchereau A, Nesi N, Pernas M and Oñate-Sánchez L (2019) An integrative approach to analyze seed germination in *Brassica napus*. *Front Plant Sci* 10:1342. <https://doi.org/10.3389/fpls.2019.01342>
- Bouchet A-S (2015) Architecture génétique du rendement et de ses composantes chez le colza d'hiver (*Brassica napus* L.) cultivé sous contrainte azotée. Thèse de doctorat, Rennes, Université de Rennes 1
- Bureau T (2020) PHENOTIC Platform. <https://doi.org/10.17180/YKBZ-2V85>
- Brault C, Lazerges J, Doligez A, et al (2022) Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods* 18:108. <https://doi.org/10.1186/s13007-022-00940-9>
- Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet* 103:338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Brunel S, Teulat-Merah B, Wagner M-H, et al (2009) Using a model-based framework for analysing genetic diversity during germination and heterotrophic growth of *Medicago truncatula*. *Ann Bot* 103:1103–1117. <https://doi.org/10.1093/aob/mcp040>
- Brunel-Muguet S, Aubertot J-N, Dürr C (2011) Simulating the impact of genetic diversity of *Medicago truncatula* on germination and emergence using a crop emergence model for ideotype breeding. *Ann Bot* 107:1367–1376. <https://doi.org/10.1093/aob/mcr071>
- Buckler ES, Holland JB, Bradbury PJ, et al (2009) The genetic architecture of maize flowering time. *Science* 325:714–718. <https://doi.org/10.1126/science.1174276>
- Bucksch A, Atta-Boateng A, Azihou AF, et al (2017) Morphological plant modeling: Unleashing geometric and topological potential within the plant sciences. *Front Plant Sci* 8:900. <https://doi.org/10.3389/fpls.2017.00900>
- Buitink J, Leprince O (2024) A seed storage protocol to determine longevity. In: Kawakami, N., Sato, K. (eds) Seed Dormancy. Methods in Molecular Biology, vol 2830. Humana, New York, NY. [https://doi.org/10.1007/978-1-0716-3965-8\\_6](https://doi.org/10.1007/978-1-0716-3965-8_6)
- Butruille D, Guries R, Osborn T (1999) Increasing yield of spring oilseed rape hybrids through introgression of winter germplasm. *Crop Sci* 39:1491–1496. <https://doi.org/10.2135/cropsci1999.3951491x>
- Bus A, Körber N, Snowdon RJ, Stich B (2011) Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theor Appl Genet* 123:1413–1423. <https://doi.org/10.1007/s00122-011-1676-7>
- Busch L, Gunter V, Mentele T, Tachikawa M, Tanaka K (1994). Socializing nature: Technoscience and the transformation of rapeseed into canola. *Crop Sci* 34:607–614. <https://doi.org/10.2135/cropsci1994.0011183X003400030001x>
- Cairns JE, Namuco OS, Torres R, et al (2009) Investigating early vigour in upland rice (*Oryza sativa* L.): Part II. Identification of QTLs controlling early vigour under greenhouse and field conditions. *Field Crop Res* 113:207–217. <https://doi.org/10.1016/j.fcr.2009.05.007>
- Castillo-Lorenzo E, Finch-Savage WE, Seal CE, Pritchard HW (2019) Adaptive significance of functional germination traits in crop wild relatives of *Brassica*. *Agric For Meteorol* 264:343–350. <https://doi.org/10.1016/j.agrformet.2018.10.014>
- Cernac A, Andre C, Hoffmann-Benning S, Benning C (2006) WRI1 is required for seed germination and seedling establishment. *Plant Physiol* 141:745–757. <https://doi.org/10.1104/pp.106.079574>



- Chalhoub B, Denoeud F, Liu S, et al (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345:950–953. <https://doi.org/10.1126/science.1253435>
- Chan KY, Heenan DP (1996) The influence of crop rotation on soil structure and soil physical properties under conventional tillage. *Soil Till Res* 37:113–125. [https://doi.org/10.1016/0167-1987\(96\)01008-2](https://doi.org/10.1016/0167-1987(96)01008-2).
- Chatellier V (2022). Les échanges mondiaux, européens et français d’oléagineux. 16èmes Journées de Recherches en Sciences Sociales, Clermont-Ferrand, 15 et 16 décembre 2022
- Chenu K, Van Oosterom EJ, McLean G, et al (2018) Integrating modelling and phenotyping approaches to identify and screen complex traits: transpiration efficiency in cereals. *J Exp Bot* 69:3181–3194. <https://doi.org/10.1093/jxb/ery059>
- Clarke WE, Higgins EE, Plieske J, et al (2016) A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor Appl Genet* 129:1887–1899. <https://doi.org/10.1007/s00122-016-2746-7>
- Cooper M, van Eeuwijk FA, Hammer GL, et al (2009) Modeling QTL for complex traits: detection and context for plant breeding. *Curr Opin Plant Biol* 12:231–240. <https://doi.org/10.1016/j.pbi.2009.01.006>
- Corbineau F, Xia Q, Bailly C, El-Maarouf-Bouteau H (2014) Ethylene, a key factor in the regulation of seed dormancy. *Front Plant Sci* 5:539. <https://doi.org/10.3389/fpls.2014.00539>
- Corlouer E (2020) Identification des déterminants génétiques et moléculaires contrôlant la stabilité du rendement chez le colza d’hiver. Thèse de doctorat, Rennes, Agrocampus Ouest
- Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One* 11:e0156744. <https://doi.org/10.1371/journal.pone.0156744>
- Cowling WA (2007) Genetic diversity in Australian canola and implications for crop breeding for changing future environments. *Field Crop Res* 104:103–111. <https://doi.org/10.1016/j.fcr.2006.12.014>
- Crossa J, Jarquín D, Franco J, et al (2016) Genomic prediction of gene bank wheat landraces. *G3-Genes Genomes Genet* 6:1819–1834. <https://doi.org/10.1534/g3.116.029637>
- Crossa J, Montesinos-Lopez OA, Costa-Neto G, et al (2024) Machine learning algorithms translate big data into predictive breeding accuracy. *Trends Plant Sci* <https://doi.org/10.1016/j.tplants.2024.09.011>
- Cuevas J, Montesinos-López O, Juliana P, et al (2019) Deep kernel for genomic and near infrared predictions in multi-environment breeding trials. *G3-Genes Genomes Genet* 9:2913–2924. <https://doi.org/10.1534/g3.119.400493>
- Cui K, Peng S, Xing Y, et al (2002) Molecular dissection of seedling-vigor and associated physiological traits in rice. *Theor Appl Genet* 105:745–753. <https://doi.org/10.1007/s00122-002-0908-2>
- Crawford NM, Forde BG (2002) Molecular and developmental biology of inorganic nitrogen nutrition. *Arabidopsis Book* 2002:e0011. <https://doi.org/10.1199/tab.0011>
- Daetwyler HD, Hickey JM, Henshall JM, et al (2010) Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Anim Prod Sci* 50:1004–1010. <https://doi.org/10.1071/AN10096>
- Daetwyler HD, Kemper KE, van der Werf JHJ, Hayes BJ (2012) Components of the accuracy of genomic prediction in a multi-breed sheep population. *J Anim Sci* 90:3375–3384. <https://doi.org/10.2527/jas.2011-4557>
- Dale JE (1982) Some effects of temperature and irradiance on growth of the first four leaves of wheat, *Triticum aestivum*. *Ann Bot* 50:851–858. <https://doi.org/10.1093/oxfordjournals.aob.a086428>
- Dallinger HG, Löschenberger F, Bistrich H, et al (2023) Predictor bias in genomic and phenomic selection. *Theor Appl Genet* 136:235. <https://doi.org/10.1007/s00122-023-04479-8>
- Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Davis RM, Nunez JJ (1999) Influence of crop rotation on the incidence of Pythium- and Rhizoctonia-induced carrot root dieback. *Plant Dis* 83:146–148. <https://doi.org/10.1094/PDIS.1999.83.2.146>



## Références bibliographiques

---

- Dejoux J-F, Meynard J-M, Reau R, et al (2003) Evaluation of environmentally-friendly crop management systems based on very early sowing dates for winter oilseed rape in France. *Agronomy* 23:725–736. <https://doi.org/10.1051/agro:2003050>
- Delhon P, Gojon A, Tillard P, Passama L (1996) Diurnal regulation of NO<sub>3</sub> uptake in soybean plants IV. Dependence on current photosynthesis and sugar availability to the roots. *J Exp Bot* 47:893–900. <https://doi.org/10.1093/jxb/47.7.893>
- Delourme R, Falentin C, Fomeju BF, et al (2013) High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC Genomics* 14:120. <https://doi.org/10.1186/1471-2164-14-120>
- Delourme R, Laperche A, Bouchet A-S, Jubault M, Paillard S, Manzanares-Dauleux MJ, Nesi N (2018) Genes and quantitative trait loci mapping for major agronomic traits in *Brassica napus* L. in The *Brassica napus* Genome Ed: Shengyi Liu, Rod Snowdon and Boulos Chalhoub. [https://doi.org/10.1007/978-3-319-43694-4\\_3](https://doi.org/10.1007/978-3-319-43694-4_3)
- Demilly D, Ducournau S, Wagner M-H (2014) Digital imaging of seed germination. In: *Plant Image Analysis: Fundamentals and Applications*, Gupta, S. D., Ibaraki, Y. CRC Press, Boca Raton, pp 147–162. ISBN : 9780429072345
- Derakhshan A, Bakhshandeh A, Siadat SA, et al (2018) Quantifying the germination response of spring canola (*Brassica napus* L.) to temperature. *Ind Crop Prod* 122:195–201. <https://doi.org/10.1016/j.indcrop.2018.05.075>
- DeSalvio AJ, Adak A, Murray SC, et al (2024) Near-infrared reflectance spectroscopy phenomic prediction can perform similarly to genomic prediction of maize agronomic traits across environments. *Plant Genome* 17:e20454. <https://doi.org/10.1002/tpg2.20454>
- Dias PMB, Brunel-Muguet S, Dürr C, et al (2011) QTL analysis of seed germination and pre-emergence growth at extreme temperatures in *Medicago truncatula*. *Theor Appl Genet* 122:429–444. <https://doi.org/10.1007/s00122-010-1458-7>
- Diers BW, Osborn TC (1994) Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. *Theor Appl Genet* 88:662–668. <https://doi.org/10.1007/BF01253968>
- Dingkuhn M, Jones MP, Johnson DE, Sow A (1998) Growth and yield potential of *Oryza sativa* and *O. glaberrima* upland rice cultivars and their interspecific progenies. *Field Crop Res* 57:57–69. [https://doi.org/10.1016/S0378-4290\(97\)00115-9](https://doi.org/10.1016/S0378-4290(97)00115-9)
- Di Paola A, Valentini R, Santini M (2016) An overview of available crop growth and yield models for studies and assessments in agriculture. *J Sci Food Agric* 96:709–714. <https://doi.org/10.1002/jsfa.7359>
- Drouet J-L, Pagès L (2003) GRAAL: a model of GRowth, Architecture and carbon ALlocation during the vegetative phase of the whole maize plant: Model description and parameterisation. *Ecol Model* 165:147–173. [https://doi.org/10.1016/S0304-3800\(03\)00072-3](https://doi.org/10.1016/S0304-3800(03)00072-3)
- Ducournau S, Feutry A, Plainchault P, et al (2004) An image acquisition system for automated monitoring of the germination rate of sunflower seeds. *Comput Electron Agric* 44:189–202. <https://doi.org/10.1016/j.compag.2004.04.005>
- Dürr C, Aubertot J-N, Richard G, et al (2001) Simple. *Soil Sci Soc Am J* 65:414–423. <https://doi.org/10.2136/sssaj2001.652414x>
- Dürr C, Constantin J, Wagner M-H, et al (2016) Virtual modeling based on deep phenotyping provides complementary data to field experiments to predict plant emergence in oilseed rape genotypes. *Eur J Agron* 79:90–99. <https://doi.org/10.1016/j.eja.2016.06.001>
- Edwards SV (2013) Next-generation QTL mapping: crowdsourcing SNPs, without pedigrees. *Mol Ecol* 22:3885–3887. <https://doi.org/10.1111/mec.12401>
- Elliott RH, Franke C, Rakow GFW (2008) Effects of seed size and seed weight on seedling establishment, vigour and tolerance of Argentine canola (*Brassica napus*) to flea beetles, *Phyllotreta* spp. *Can J Plant Sci* 88:207–217. <https://doi.org/10.4141/CJPS07059>



## Références bibliographiques

---

- El-Lithy ME, Reymond M, Stich B, Koornneef M, Vreugdenhil D (2010) Relation among plant growth, carbohydrates and flowering time in the *Arabidopsis* Landsberg *erecta* × Kondara recombinant inbred line population. *Plant Cell Environ* 33:1369–1382. <https://doi.org/10.1111/j.1365-3040.2010.02155.x>
- Falconer D, Mackay T (1996) Introduction to quantitative genetics. Pearson, Harlow
- Federer WT (1956) Augmented (or hoonuiaku) designs. *Hawaiian Planters' Record* 55:191–208.
- Fernandes SB, Dias KOG, Ferreira DF, Brown PJ (2018) Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theor Appl Genet* 131:747–755. <https://doi.org/10.1007/s00122-017-3033-y>
- Finch-Savage WE (1995) Influence of seed quality on crop establishment, growth and yield. In: Seed quality, basic mechanisms and agricultural implications. Eds Basra AS, Food Products Press, New-York, USA, 361–384.
- Finch-Savage WE, Bassel GW (2016) Seed vigour and crop establishment: extending performance beyond adaptation. *J Exp Bot* 67:567–591. <https://doi.org/10.1093/jxb/erv490>
- Fu J, Falke KC, Thiemann A, et al (2012) Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene expression data. *Theor Appl Genet* 124:825–833. <https://doi.org/10.1007/s00122-011-1747-9>
- Fu Y-B, Gugel RK (2010) Genetic diversity of Canadian elite summer rape (*Brassica napus* L.) cultivars from the pre- to post-canola quality era. *Can J Plant Sci* 90:23–33. <https://doi.org/10.4141/CJPS09073>
- Gad M, Chao H, Li H, et al (2021) QTL mapping for seed germination response to drought stress in *Brassica napus*. *Front Plant Sci* 11:629970. <https://doi.org/10.3389/fpls.2020.629970>
- Gaeta RT, Pires JC, Iniguez-Luy F, et al (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* 19:3403–3417. <https://doi.org/10.1105/tpc.107.054346>
- Gaju O, Allard V, Martre P, et al (2011) Identification of traits to improve the nitrogen-use efficiency of wheat genotypes. *Field Crop Res* 123:139–152. <https://doi.org/10.1016/j.fcr.2011.05.010>
- Galán RJ, Bernal-Vasquez A-M, Jebsen C, et al (2020) Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. *Theor Appl Genet* 133:3001–3015. <https://doi.org/10.1007/s00122-020-03651-8>
- Galán RJ, Bernal-Vasquez A-M, Jebsen C, et al (2021) Early prediction of biomass in hybrid rye based on hyperspectral data surpasses genomic predictability in less-related breeding material. *Theor Appl Genet* 134:1409–1422. <https://doi.org/10.1007/s00122-021-03779-1>
- Galland M, Huguet R, Arc E, et al (2014) Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during *arabidopsis* seed germination. *Mol Cell Proteomics* 13:252–268. <https://doi.org/10.1074/mcp.M113.032227>
- Gallusci P, Dai Z, Génard M, et al (2017) Epigenetics for plant improvement: Current knowledge and modeling avenues. *Trends Plant Sci* 22:610 – 623. <https://doi.org/10.1016/j.tplants.2017.04.009>
- Ganal MW, Polley A, Graner E-M, et al (2012) Large SNP arrays for genotyping in crop plants. *J Biosci* 37:821–828. <https://doi.org/10.1007/s12038-012-9225-3>
- Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol* 32:361–369. <https://doi.org/10.1002/gepi.20310>
- Gao X, Becker LC, Becker DM, et al (2010) Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet Epidemiol* 34:100–105. <https://doi.org/10.1002/gepi.20430>
- Gastal F, Lemaire G (2002) N uptake and distribution in crops: an agronomical and ecophysiological perspective. *J Exp Bot* 53:789–799. <https://doi.org/10.1093/jexbot/53.370.789>
- Gazave E, Tassone EE, Ilut DC, et al (2016) Population genomic analysis reveals differential evolutionary histories and patterns of diversity across subgenomes and subpopulations of *Brassica napus* L. *Front Plant Sci* 7:525. <https://doi.org/10.3389/fpls.2016.00525>



## Références bibliographiques

---

- Gianola D, Okut H, Weigel KA, Rosa GJM (2011) Predicting complex quantitative traits with Bayesian neural networks: A case study with Jersey cows and wheat. *BMC Genet* 12:87. <https://doi.org/10.1186/1471-2156-12-87>
- Glaszmann JC, Kilian B, Upadhyaya HD, et al (2010) Accessing genetic diversity for crop improvement. *Curr Opin Plant Biol* 13:167-173. <https://doi.org/10.1016/j.pbi.2010.01.004>.
- Gorjanc G, Jenko J, Hearne SJ, Hickey JM (2016) Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. *BMC Genomics* 17:30. <https://doi.org/10.1186/s12864-015-2345-z>
- Gouleau A, Gauffreteau A, This P, et al (2021) Saisine du Comité Scientifique CTPS – Quelles variétés pour l'agroécologie ?
- Gourrion A, Simon C, Vallée P, et al (2020) Enlarging the genetic diversity of winter oilseed rape (WOSR) by crossing with spring oilseed rape (SOSR). *OCL* 27:16. <https://doi.org/10.1051/ocl/2020013>
- Gower JC (1967) Multivariate analysis and multidimensional geometry. *J R Stat Soc Ser D-The Statistician* 17:13–28. <https://doi.org/10.2307/2987199>
- Groot SPC, Karssen CM (1987) Gibberellins regulate seed germination in tomato by endosperm weakening: a study with gibberellin-deficient mutants. *Planta* 171:525–531. <https://doi.org/10.1007/BF00392302>
- Gu J, Chao H, Gan L, et al (2016) Proteomic dissection of seed germination and seedling establishment in *Brassica napus*. *Front Plant Sci* 7:1482. <https://doi.org/10.3389/fpls.2016.01482>
- Haelterman L, Louvieaux J, Chiodi C, et al (2024) Genetic control of root morphology in response to nitrogen across rapeseed diversity. *Physiol Plant* 176:e14315. <https://doi.org/10.1111/ppl.14315>
- Harker KN, O'Donovan JT, Smith EG, et al (2015) Seed size and seeding rate effects on canola emergence, development, yield and seed weight. *Can J Plant Sci* 95:1–8. <https://doi.org/10.4141/cjps-2014-222>
- Hasan, M, Seyis, F, Badani, AG et al (2006) Analysis of genetic diversity in the *Brassica napus* L. gene pool using SSR markers. *Genet Resour Crop Evol* 53:793–802. <https://doi.org/10.1007/s10722-004-5541-2>
- Hassani-Pak K, Singh A, Brandizi M, et al (2021) KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. *Plant Biotechnol J* 19:1670–1678. <https://doi.org/10.1111/pbi.13583>
- Hatzig SV, Frisch M, Breuer F, et al (2015) Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. *Front Plant Sci* 6:221. <https://doi.org/10.3389/fpls.2015.00221>
- Hatzig SV, Nuppenau J-N, Snowdon RJ, Schießl SV (2018) Drought stress has transgenerational effects on seeds and seedlings in winter oilseed rape (*Brassica napus* L.). *BMC Plant Biol* 18:297. <https://doi.org/10.1186/s12870-018-1531-y>
- Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Res* 91:47–60. <https://doi.org/10.1017/S0016672308009981>
- Hayward A, Mason AS, Dalton-Morgan J, Zander M, Edwards D, Batley J (2012). SNP discovery and applications in *Brassica napus*. *J Plant Biotechnol* 39:49–61. <https://doi.org/10.5010/JPB.2012.39.1.049>
- Hebinger H (2013) Le colza. In Editions France Agricole. ISBN: 978-2-85557-241-3
- Hellebust JA, Bidwell RGS (1963) Sources of carbon for the synthesis of protein amino acids in attached photosynthesizing wheat leaves. *Botany* 7:985–994. <https://doi.org/10.1139/b63-082>
- Hendriks P-W, Gurusinghe S, Weston PA, et al (2024) Introgression of early shoot vigour in wheat modifies root systems, increases competitiveness and provides options for integrated weed management. *Plant Soil* <https://doi.org/10.1007/s11104-024-06653-3>
- Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. *Crop Sci* 52:146–160. <https://doi.org/10.2135/cropsci2011.06.0297>
- Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J Exp Bot* 58:2369–2387. <https://doi.org/10.1093/jxb/erm097>



## Références bibliographiques

---

- Hodgkin T and Hegarty TW (1978). Genetically determined variation in seed germination and field emergence of *Brassica oleracea*. Ann Appl Biol 88:407-413. <https://doi.org/10.1111/j.1744-7348.1978.tb00732.x>
- Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161. <https://doi.org/10.1016/j.pbi.2007.01.003>
- Holzworth DP, Huth NI, deVoil PG, et al (2014) APSIM – Evolution towards a new generation of agricultural systems simulation. Environ Modell Softw 62:327-350. <https://doi.org/10.1016/j.envsoft.2014.07.009>.
- Hu H, Campbell MT, Yeats TH, et al (2021) Multi-omics prediction of oat agronomic and seed nutritional traits across environments and in distantly related populations. Theor Appl Genet 134:4043–4054. <https://doi.org/10.1007/s00122-021-03946-4>
- Hunt R (2003) Growth and development | Growth Analysis, Individual Plants. In: Thomas B (ed) Encyclopedia of Applied Plant Sciences. Elsevier, Oxford, pp 579–588
- Imbert B, Kreplak J, Flores R-G, et al (2023) Development of a knowledge graph framework to ease and empower translational approaches in plant research: A use-case on grain legumes. Front Artif Intell 6:1191122. <https://doi.org/10.3389/frai.2023.1191122>
- Innan, H., Kondrashov, F., 2010. The evolution of gene duplications: Classifying and distinguishing between models. Nat Rev Genet 11:97–108. <https://doi.org/10.1038/nrg2689>
- Jannink J-L (2010) Dynamics of long-term genomic selection. Genet Sel Evol 42:35. <https://doi.org/10.1186/1297-9686-42-35>
- Jenczewski E, Chèvre AM, Alix K (2013) Chromosomal and gene expression changes in *Brassica* allopolyploids. Polyploid and hybrid genomics, Wiley Online Books. <https://doi.org/doi:10.1002/9781118552872.ch1>
- Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, Wu X, Zhang Y, Wang J, et al (2012) A genome triplication associated with early diversification of the core eudicots. Genome Biol 13:1–14. <https://doi.org/10.1186/gb-2012-13-1-r3>
- Jones DL, Nguyen C, Finlay RD (2009) Carbon flow in the rhizosphere: Carbon trading at the soil–root interface. Plant Soil 321:5–33. <https://doi.org/10.1007/s11104-009-9925-0>
- Kebede B, Thiagarajah M, Zimmerli C, Rahman MH (2010) Improvement of open-pollinated spring rapeseed (*Brassica napus* L.) through Introgression of genetic diversity from winter rapeseed. Crop Sci 50:1236–1243. <https://doi.org/10.2135/cropsci2009.06.0352>
- Kiran A, Wakeel A, Snowdon R, Friedt W (2019) Genetic dissection of root architectural traits by QTL and genome-wide association mapping in rapeseed (*Brassica napus*). Plant Breed 138:184–192. <https://doi.org/10.1111/pbr.12665>
- Knoch D, Abbadi A, Grandke F, et al (2020) Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. Plant Biotechnol J 18:68–82. <https://doi.org/10.1111/pbi.13171>
- Knoch D, Werner CR, Meyer RC, et al (2021) Multi-omics-based prediction of hybrid performance in canola. Theor Appl Genet 134:1147–1165. <https://doi.org/10.1007/s00122-020-03759-x>
- Koscielny CB, Gulden RH (2012) Seedling root length in *Brassica napus* L. is indicative of seed yield. Can J Plant Sci 92:1229–1237. <https://doi.org/10.4141/cjps2012-070>
- Krause MR, González-Pérez L, Crossa J, et al (2019) Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. G3-Genes Genomes Genet 9:1231–1247. <https://doi.org/10.1534/g3.118.200856>
- Lamichhane JR, Dürr C, Schwanck AA, et al (2017) Integrated management of damping-off diseases. A review. Agron Sustain Dev 37:10. <https://doi.org/10.1007/s13593-017-0417-y>
- Lamichhane JR, Debaeke P, Steinberg C, et al (2018) Abiotic and biotic factors affecting crop seed germination and seedling emergence: a conceptual framework. Plant Soil 432:1–28. <https://doi.org/10.1007/s11104-018-3780-9>



## Références bibliographiques

---

- Lämke J, Bäurle I (2017) Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biol* 18:124. <https://doi.org/10.1186/s13059-017-1263-6>
- Lane HM, Murray SC, Montesinos-López OA, et al (2020) Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *Plant Phenome J* 3:e20002. <https://doi.org/10.1002/ppj2.20002>
- Laurençon M, Legrix J, Wagner M-H, et al (2024) Genomic and phenomic predictions help capture low-effect alleles promoting seed germination in oilseed rape in addition to QTL analyses. *Theor Appl Genet* 137:156. <https://doi.org/10.1007/s00122-024-04659-0>
- Lawlor DW, Boyle FA, Keys AJ, et al (1988) Nitrate nutrition and temperature effects on wheat: a synthesis of plant growth and nitrogen uptake in relation to metabolic and physiological processes. *J Exp Bot* 39:329–343. <https://doi.org/10.1093/jxb/39.3.329>
- Lê S, Josse J, Husson F (2008) FactoMineR: An R Package for multivariate analysis. *J Stat Softw* 25:1–18. <https://doi.org/10.18637/jss.v025.i01>
- Lecarpentier C, Pagès L, Richard-Molard C (2021) Genotypic diversity and plasticity of root system architecture to nitrogen availability in oilseed rape. *PLoS One* 16:e0250966. <https://doi.org/10.1371/journal.pone.0250966>
- Leflon M, Grandont L, Eber F, et al (2010) Crossovers get a boost in Brassica allotriploid and allotetraploid hybrids. *Plant Cell* 22:2253–2264. <https://doi.org/10.1105/tpc.110.075986>
- Li K, Wang J, Kuang L, et al (2021) Genome-wide association study and transcriptome analysis reveal key genes affecting root growth dynamics in rapeseed. *Biotechnol Biofuels Bioprod* 14:178. <https://doi.org/10.1186/s13068-021-02032-7>
- Liao M, Fillery IRP, Palta JA (2004) Early vigorous growth is a major factor influencing nitrogen uptake in wheat. *Funct Plant Biol* 31:121–129. <https://doi.org/10.1071/fp03060>
- Lin Z, Cogan NOI, Pembleton LW, et al (2016) Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. *Plant Genome* 9:plantgenome2015.06.0046. <https://doi.org/10.3835/plantgenome2015.06.0046>
- Lippert C, Listgarten J, Liu Y, et al (2011) FaST linear mixed models for genome-wide association studies. *Nat Methods* 8:833–835. <https://doi.org/10.1038/nmeth.1681>
- Liu S, Liu Y, Yang X, et al (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun* 5:3930. <https://doi.org/10.1038/ncomms4930>
- Liu S, Remley M, Nichols RL, Fritschi FB (2019) Morphological traits underlying differences in early vigor among four cotton genotypes. *Crop Sci* 59:1165–1181. <https://doi.org/10.2135/cropsci2018.10.0611>
- Liu H, Pan Y, Cui R et al (2023) Integrating genome-wide association studies with selective sweep reveals genetic loci associated with tolerance to low phosphate availability in *Brassica napus*. *Mol Breeding* 43:53. <https://doi.org/10.1007/s11032-023-01399-9>
- Longin CFH, Reif JC (2014) Redesigning the exploitation of wheat genetic resources. *Trends Plant Sci* 19:631–636. <https://doi.org/10.1016/j.tplants.2014.06.012>
- López-Castañeda C, Richards RA, Farquhar GD (1995) Variation in early vigor between wheat and barley. *Crop Sci* 35:472–479. <https://doi.org/10.2135/cropsci1995.0011183X003500020032x>
- Lu K, Wei L, Li X, et al (2019) Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat Commun* 10:1154. <https://doi.org/10.1038/s41467-019-09134-9>
- Lund MS, Berg I van den, Ma P, et al (2016) Review: How to improve genomic predictions in small dairy cattle populations. *Animal* 10:1042–1049. <https://doi.org/10.1017/S1751731115003031>
- Luo X, Xue Z, Ma C, et al (2017) Joint genome-wide association and transcriptome sequencing reveals a complex polygenic network underlying hypocotyl elongation in rapeseed (*Brassica napus* L.). *Sci Rep* 7:41561. <https://doi.org/10.1038/srep41561>
- Lush JL (1943) Animal Breeding Plans. The Iowa State College Press, Ames, IA



## Références bibliographiques

---

- Ma L, Wang C, Hu Y, et al (2022) GWAS and transcriptome analysis reveal MADS26 involved in seed germination ability in maize. *Theor Appl Genet* 135:1717–1730. <https://doi.org/10.1007/s00122-022-04065-4>
- Mackay TFC (2001) The genetic architecture of quantitative traits. *Annu Rev Genet* 35:303–339. <https://doi.org/10.1146/annurev.genet.35.102401.090633>
- Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10:565–577. <https://doi.org/10.1038/nrg2612>
- Mackay TFC, Anholt RRH (2024) Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat Rev Genet* 1–19. <https://doi.org/10.1038/s41576-024-00711-3>
- Maggiorelli A, Baig N, Prigge V, et al (2024) Using drone-retrieved multispectral data for phenomic selection in potato breeding. *Theor Appl Genet* 137:70. <https://doi.org/10.1007/s00122-024-04567-3>
- Martel C, Vrebalov J, Tafelmeyer P, Giovannoni JJ (2011) The tomato MADS-Box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner. *Plant Physiol* 157:1568–1579. <https://doi.org/10.1104/pp.111.181107>
- Martre P, Quilot-Turion B, Luquet D, et al (2015) Chapter 14 - Model-assisted phenotyping and ideotype design. In: Sadras VO, Calderini DF (eds) *Crop Physiology* (Second Edition). Academic Press, San Diego, pp 349–373
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Meyenberg C, Braun V, Longin CFH, Thorwarth P (2024) Feature engineering and parameter tuning: improving phenomic prediction ability in multi-environmental durum wheat breeding trials. *Theor Appl Genet* 137:188. <https://doi.org/10.1007/s00122-024-04695-w>
- Meyer RC, Steinfath M, Lisec J, et al (2007) The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 104:4759–4764. <https://doi.org/10.1073/pnas.0609709104>
- Meyer RC, Weigelt-Fischer K, Knoch D, et al (2021) Temporal dynamics of QTL effects on vegetative growth in *Arabidopsis thaliana*. *J Exp Bot* 72:476–490. <https://doi.org/10.1093/jxb/eraa490>
- Montesinos-López OA, Montesinos-López A, Crossa J, et al (2017) Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* 13:4. <https://doi.org/10.1186/s13007-016-0154-2>
- Montesinos-López OA, Montesinos-López A, Pérez-Rodríguez P, et al (2021) A review of deep learning applications for genomic selection. *BMC Genomics* 22:19. <https://doi.org/10.1186/s12864-020-07319-x>
- Morris ER, Chevalier D, Walker JC (2006) DAWDLE, a forkhead-associated domain gene, regulates multiple aspects of plant development. *Plant Physiol* 141:932–941. <https://doi.org/10.1104/pp.106.076893>
- Munz E, Rolletschek H, Oeltze-Jafra S, et al (2017) A functional imaging study of germinating oilseed rape seed. *New Phytol* 216:1181–1190. <https://doi.org/10.1111/nph.14736>
- Murat, F., Louis, A., Maumus, F., Armero, A., Cooke, R., Quesneville, H., Crolius, H.R., Salse, J., 2015. Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genome Biol* 1–17. <https://doi.org/10.1186/s13059-015-0814-y>
- Namuco OS, Cairns JE, Johnson DE (2009) Investigating early vigour in upland rice (*Oryza sativa* L.): Part I. Seedling growth and grain yield in competition with weeds. *Field Crop Res* 113:197–206. <https://doi.org/10.1016/j.fcr.2009.05.008>
- Nelson MN, Nesi N, Barrero JM, et al (2022) Strategies to improve field establishment of canola: A review. *Adv Agron* 175:133–177. <https://doi.org/10.1016/bs.agron.2022.05.001>
- Nesi N, Delourme R, Brégeon M, et al (2008) Genetic and molecular approaches to improve nutritional value of *Brassica napus* L. seed. *C R Biol* 331:763–771. <https://doi.org/10.1016/j.crvi.2008.07.018>
- Nguyen TCT, Abrams SR, Friedt W, Snowdon RJ (2018) Quantitative trait locus analysis of seed germination, seedling vigour and seedling-regulated hormones in *Brassica napus*. *Plant Breed* 137:388–401. <https://doi.org/10.1111/pbr.12576>



## Références bibliographiques

---

- Nonogaki H, Bassel GW, Bewley JD (2010) Germination—Still a mystery. *Plant Sci* 179:574–581. <https://doi.org/10.1016/j.plantsci.2010.02.010>
- Norman A, Taylor J, Edwards J, Kuchel H (2018) Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. *G3-Genes Genomes Genet* 8:2889–2899. <https://doi.org/10.1534/g3.118.200311>
- Ogutu JO, Piepho H-P, Schulz-Streeck T (2011) A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc* 5:S11. <https://doi.org/10.1186/1753-6561-5-S3-S11>
- Pang J, Palta JA, Rebetzke GJ, Milroy SP (2013) Wheat genotypes with high early vigour accumulate more nitrogen and have higher photosynthetic nitrogen use efficiency during early growth. *Funct Plant Biol* 41:215–222. <https://doi.org/10.1071/FP13143>
- Parkin, IAP, Gulden, SM, Sharpe, AG, Lukens, L, Trick, M, Osborn, TC, Lydiate, DJ, (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171: 765–781. <https://doi.org/10.1534/genetics.105.042093>
- Parkin IA, Koh C, Tang H, et al (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol* 15:R77. <https://doi.org/10.1186/gb-2014-15-6-r77>
- Parry MAJ, Andralojc JP, Khan S, et al (2002) Rubisco activity: Effects of drought stress. *Ann Bot* 89:833–839. <https://doi.org/10.1093/aob/mcf103>
- Pelé A, Falque M, Trotoux G, et al (2017) Amplifying recombination genome-wide and reshaping crossover landscapes in Brassicas. *PLoS Genet* 13:e1006794. <https://doi.org/10.1371/journal.pgen.1006794>
- Penfield S, Graham S, Graham IA (2005) Storage reserve mobilization in germinating oilseeds: *Arabidopsis* as a model system. *Biochem Soc Trans* 33:380–383. <https://doi.org/10.1042/BST0330380>
- Peter R, Eschholz TW, Stamp P, Liedgens M (2009) Swiss Flint maize landraces-A rich pool of variability for early vigour in cool environments. *Field Crop Res*. 110:157–166. <https://doi.org/10.1016/j.fcr.2008.07.015>
- Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
- Qian W, Meng J, Li M, et al (2006) Introgression of genomic components from Chinese *Brassica rapa* contributes to widening the genetic diversity in rapeseed (*B. napus* L.), with emphasis on the evolution of Chinese rapeseed. *Theor Appl Genet* 113:49–54. <https://doi.org/10.1007/s00122-006-0269-3>
- Qian W, Sass O, Meng J, et al (2007) Heterotic patterns in rapeseed (*Brassica napus* L.): I. Crosses between spring and Chinese semi-winter lines. *Theor Appl Genet* 115:27–34. <https://doi.org/10.1007/s00122-007-0537-x>
- Qian W, Li Q, Noack J, et al (2009) Heterotic patterns in rapeseed (*Brassica napus* L.): II. Crosses between European winter and Chinese semi-winter lines. *Plant Breed* 128:466–470. <https://doi.org/10.1111/j.1439-0523.2008.01597.x>
- Quijada PA, Udall JA, Lambert B, Osborn TC (2006) Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm. *Theor Appl Genet* 113:549–561. <https://doi.org/10.1007/s00122-006-0323-1>
- Radin JW, Parker LL (1979) Water relations of cotton plants under nitrogen deficiency: I. Dependence upon leaf structure. *Plant Physiol* 64:495–498. <https://doi.org/10.1104/pp.64.3.495>
- Rahimi Y, Khahani B, Jamali A, et al (2023) Genome-wide association study to identify genomic loci associated with early vigor in bread wheat under simulated water deficit complemented with quantitative trait loci meta-analysis. *G3-Genes Genomes Genet* 13:jkac320. <https://doi.org/10.1534/g3journal/jkac320>
- Rahman H (2013) Review: Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. *Can J Plant Sci* 93:363–373. <https://doi.org/10.4141/cjps2012-074>
- Rajjou L, Duval M, Gallardo K, et al (2012) Seed Germination and Vigor. *Annu Rev Plant Biol* 63:507–533. <https://doi.org/10.1146/annurev-arplant-042811-105550>
- Rebetzke GJ, Botwright TL, Moore CS, et al (2004) Genotypic variation in specific leaf area for genetic improvement of early vigour in wheat. *Field Crop Res*. 88:179–189. <https://doi.org/10.1016/j.fcr.2004.01.007>



## Références bibliographiques

---

- Rebolledo MC, Dingkuhn M, Péré P, et al (2012) Developmental dynamics and early growth vigour in rice. I. Relationship between development rate (1/phyllochron) and growth. *J Agron Crop Sci* 198:374–384. <https://doi.org/10.1111/j.1439-037X.2012.00528.x>
- Rebolledo MC, Luquet D, Courtois B, et al (2013) Can early vigour occur in combination with drought tolerance and efficient water use in rice genotypes? *Funct Plant Biol* 40:582–594. <https://doi.org/10.1071/FP12312>
- Rebolledo MC, Dingkuhn M, Courtois B, et al (2015) Phenotypic and genetic dissection of component traits for early vigour in rice using plant growth modelling, sugar content analyses and association mapping. *J Exp Bot* 66:5555–5566. <https://doi.org/10.1093/jxb/erv258>
- Regan KL, Siddique KHM, Turner NC, Whan BR (1992) Potential for increasing early vigour and total biomass in spring wheat. II. Characteristics associated with early vigour. *Crop Pasture Sci* 43:541–553. <https://doi.org/10.1071/ar9920541>
- Reymond M, Muller B, Tardieu F (2004) Dealing with the genotypexenvironment interaction via a modelling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters. *J Exp Bot* 55:2461–2472. <https://doi.org/10.1093/jxb/erh200>
- Richard-Molard C, Brun F, Chelle M, Ney B (2009) Modelling N nutrition impact on plant functioning and root architecture in various genotypes of *Arabidopsis thaliana*. *Comp Biochem Physiol A-Mol Integr Physiol* 153:S229. <https://doi.org/10.1016/j.cbpa.2009.04.634>
- Rickman RW, Klepper BL (1995) The phyllochron: Where do we go in the future? *Crop Sci* 35:44–49. <https://doi.org/10.2135/cropsci1995.0011183X003500010008x>
- Riedelsheimer C, Technow F, Melchinger AE (2012) Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. *BMC Genomics* 13:452. <https://doi.org/10.1186/1471-2164-13-452>
- Rincent R, Laloë D, Nicolas S, et al (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192:715–728. <https://doi.org/10.1534/genetics.112.141473>
- Rincent R, Moreau L, Monod H, et al (2014) Recovering power in association mapping panels with variable levels of linkage disequilibrium. *Genetics* 197:375–387. <https://doi.org/10.1534/genetics.113.159731>
- Rincent R, Charpentier J-P, Faivre-Rampant P, et al (2018) Phenomic selection is a low-cost and high throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3-Genes Genomes Genet* 8:3961–3972. <https://doi.org/10.1534/g3.118.200760>
- Robert P, Auzanneau J, Goudemand E, et al (2022a) Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theor Appl Genet* 135:895–914. <https://doi.org/10.1007/s00122-021-04005-8>
- Robert P, Brault C, Rincent R, Segura V (2022b) Phenomic selection: A new and efficient alternative to genomic selection. In: Ahmadi N, Bartholomé J (eds) *Genomic prediction of complex traits: Methods and protocols*. Springer US, New York, NY, pp 397–420
- Rolland S, Leprince F, Guichard S, et al (2024) A comprehensive dataset of near infrared spectroscopy measurements to predict Nitrogen and Carbon contents in a wide range of tissues from *Brassica napus* plants grown under contrasted environments. Soumis à Data in Brief.
- Romero Navarro JA, Willcox M, Burgueño J, et al (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nature Genet* 49:476–480. <https://doi.org/10.1038/ng.3784>
- Rooney TE, Sweeney DW, Sorrells ME (2022) Time series barley germination is predictable and associated with known seed dormancy loci. *Crop Sci* 62:100–119. <https://doi.org/10.1002/csc2.20638>
- Roscher-Ehrig L, Weber SE, Abbadi A, et al (2024) Phenomic selection for hybrid rapeseed breeding. *Plant Phenomics* 6:0215. <https://doi.org/10.34133/plantphenomics.0215>
- Roth L, Fossati D, Krähenbühl P, et al (2023) Image-based phenomic prediction can provide valuable decision support in wheat breeding. *Theor Appl Genet* 136:162. <https://doi.org/10.1007/s00122-023-04395-x>



## Références bibliographiques

---

- Rousseau-Gueutin M, Belser C, Da Silva C, et al (2020) Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience* 9:giaa137. <https://doi.org/10.1093/gigascience/giaa137>
- Ryan PR, Liao M, Delhaize E, et al (2015) Early vigour improves phosphate uptake in wheat. *J Exp Bot* 66:7089–7100. <https://doi.org/10.1093/jxb/erv403>
- Sánchez L, Yanchuk AA, King JN (2008) Gametic models for multitrait selection schemes to study variance of response and drift under adverse genetic correlations. *Tree Genet Genomes* 4:201–212. <https://doi.org/10.1007/s11295-007-0101-5>
- Sausse C, Cerrutti N, Hebinger H (2012) Impact du colza sur l'environnement: évaluation et limites méthodologiques. *OCL* 19:155–163. <https://doi.org/10.1051/ocl.2012.0454>
- Sausse C, Lecomte V, Martin-Monjaret C, et al (2016) Dégâts d'oiseaux dans les cultures d'oléo-protéagineux - Synthèse de l'enquête Terres Inovia
- Savitzky Abraham, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627–1639. <https://doi.org/10.1021/ac60214a047>
- Schnyder H, Nelson CJ (1987) Growth rates and carbohydrate fluxes within the elongation zone of tall fescue leaf blades. *Plant Physiol* 85:548–553. <https://doi.org/10.1104/pp.85.2.548>
- Schrag TA, Westhues M, Schipprack W, et al (2018) Beyond genomic prediction: Combining different types of omics data can improve prediction of hybrid performance in maize. *Genetics* 208:1373–1385. <https://doi.org/10.1534/genetics.117.300374>
- Shi Z, Chang T-G, Chen F, et al (2020) Morphological and physiological factors contributing to early vigor in the elite rice cultivar 9,311. *Sci Rep* 10:14813. <https://doi.org/10.1038/s41598-020-71913-y>
- Signorell A, Aho K, Alfons A, et al (2023) DescTools: tools for descriptive statistics. Available from: <https://cran.r-project.org/web/packages/DescTools/index.html>
- Simmonds NW (1993) Introgression and incorporation. Strategies for the use of crop genetic resources. *Biol Rev* 68:539–562. <https://doi.org/10.1111/j.1469-185X.1993.tb01243.x>
- Singh UM, Yadav S, Dixit S, et al (2017) QTL hotspots for early vigor and related traits under dry direct-seeded system in rice (*Oryza sativa* L.). *Front Plant Sci* 8:286. <https://doi.org/10.3389/fpls.2017.00286>
- Snowdon R, Lühs W, Friedt W (2007) Oilseed Rape. In: Kole C (ed) Oilseeds. Springer, Berlin, Heidelberg, pp 55–114
- Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008) Genomic selection using different marker types and densities. *J Anim Sci* 86:2447–2454. <https://doi.org/10.2527/jas.2007-0010>
- Solovieff N, Cotsapas C, Lee PH, et al (2013) Pleiotropy in complex traits: Challenges and strategies. *Nat Rev Genet* 14:483–495. <https://doi.org/10.1038/nrg3461>
- Song J-M, Guan Z, Hu J, et al (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat Plants* 6:34–45. <https://doi.org/10.1038/s41477-019-0577-7>
- Sun T, Gubler F (2004) Molecular mechanism of gibberellin signaling in plants. *Annu Rev Plant Biol* 55:197–223. <https://doi.org/10.1146/annurev.arplant.55.031903.141753>
- St. Luce M, Grant CA, Zebarth BJ, et al (2015) Legumes can reduce economic optimum nitrogen rates and increase yields in a wheat–canola cropping sequence in western Canada. *Field Crop Res* 179:12–25. <https://doi.org/10.1016/j.fcr.2015.04.003>
- Terres Inovia (2019). Colza - Point technique Réussir son implantation pour obtenir un colza robuste. <https://www.terresinovia.fr/p/guide-technique-reussir-son-implantation-pour-obtenir-un-colza-robuste>
- Terres Inovia (2023) Colza - Point technique Réussir son implantation pour obtenir un colza robuste
- Thomas CL, Alcock TD, Graham NS, et al (2016) Root morphology and seed and leaf ionomic traits in a *Brassica napus* L. diversity panel show wide phenotypic variation and are characteristic of crop habit. *BMC Plant Biol* 16:214. <https://doi.org/10.1186/s12870-016-0902-5>



## Références bibliographiques

---

- Tirnaz S, Zandberg J, Thomas WJW, et al (2022) Application of crop wild relatives in modern breeding: An overview of resources, experimental and computational methodologies. *Front Plant Sci* 13: 1008904. <https://doi.org/10.3389/fpls.2022.1008904>
- Triboi-Blondel A-M, Messaoud J, Rousseau P (1988) Mise en place et fonctionnement des feuilles de colza d'hiver : relations azote-carbone et sénescence. *Agronomie* 8:779–786. <https://doi.org/10.1051/agro:19880904>
- U N (1935) Genome-analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn J Bot* 389–452
- Udall JA, Quijada PA, Polewicz H, et al (2004) Phenotypic effects of introgressing Chinese winter and resynthesized *Brassica napus* L. germplasm into hybrid spring canola. *Crop Sci* 44:1990–1996. <https://doi.org/10.2135/cropsci2004.1990>
- Vazquez-Carrasquer V, Laperche A, Bissuel-Bélague C, et al (2021) Nitrogen uptake efficiency, mediated by fine root growth, early determines temporal and genotypic variations in nitrogen use efficiency of winter oilseed rape. *Front Plant Sci* 12: 641459. <https://doi.org/10.3389/fpls.2021.641459>
- Verhulst N, Govaerts B, Verachtert E, et al (2010) Conservation Agriculture, Improving Soil Quality for Sustainable Production Systems? In: *Advances in soil science: food security and soil quality*. eds R. Lal and B. A. Stewart (Boca Raton, FL: CRC Press), 137–208. <https://doi.org/10.1201/EBK1439800577-7>
- Vilhjálmsson B, Nordborg M (2013) The nature of confounding in genome-wide association studies. *Nat Rev Genet* 14:1–2. <https://doi.org/10.1038/nrg3382>
- Violle C, Navas M-L, Vile D, et al (2007) Let the concept of trait be functional! *Oikos* 116:882–892. <https://doi.org/10.1111/j.0030-1299.2007.15559.x>
- Vos J, Evers JB, Buck-Sorlin GH, et al (2010) Functional-structural plant modelling: a new versatile tool in crop science. *J Exp Bot* 61:2101–2115. <https://doi.org/10.1093/jxb/erp345>
- Wagner M-H, Demilly D, Ducournau S, Durr C (2011) Computer vision for monitoring seed germination from dry state to young seedlings. *Seed Testing International* 49–51
- Wang X, Wang H, Wang J, et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genet* 43:1035–1039. <https://doi.org/10.1038/ng.919>
- Wang J, Dun X, Shi J, et al (2017) Genetic dissection of root morphological traits related to nitrogen use efficiency in *Brassica napus* L. under two contrasting nitrogen conditions. *Front Plant Sci* 8:1709. <https://doi.org/10.3389/fpls.2017.01709>
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370. <https://doi.org/10.2307/2408641>
- Weiβ TM, Zhu X, Leiser WL, et al (2022) Unraveling the potential of phenomic selection within and among diverse breeding material of maize (*Zea mays* L.). *G3-Genes Genomes Genet* 12:jkab445. <https://doi.org/10.1093/g3journal/jkab445>
- Werner, C., Snowdon, R., 2018. Genome-facilitated breeding of oilseed rape, in: Liu, S., Snowdon, R., Chalhoub, B. (Eds.), *The Brassica napus genome*. Springer International Publishing, Cham, pp. 245–269. [https://doi.org/10.1007/978-3-319-43694-4\\_15](https://doi.org/10.1007/978-3-319-43694-4_15)
- Westhues M, Schrag TA, Heuer C, et al (2017) Omics-based hybrid prediction in maize. *Theor Appl Genet* 130:1927–1939. <https://doi.org/10.1007/s00122-017-2934-0>
- Whittington WJ (1973). Genetic regulation of germination. In: *Seed ecology*. Eds Heydecker W, Butterworths, London, UK, 5-30
- Winn ZJ, Amsberry AL, Haley SD, et al (2023) Phenomic versus genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines. *Plant Phenome J* 6:e20084. <https://doi.org/10.1002/ppj2.20084>
- Wu P-Y, Stich B, Weisweiler M, et al (2022) Improvement of prediction ability by integrating multi-omic datasets in barley. *BMC Genomics* 23:200. <https://doi.org/10.1186/s12864-022-08337-7>



## Références bibliographiques

---

- Xu G, Fan X, Miller AJ (2012) Plant nitrogen assimilation and use efficiency. *Ann Rev Plant Biol* 63:153–182. <https://doi.org/10.1146/annurev-arplant-042811-105532>
- Xu S, Xu Y, Gong L, Zhang Q (2016) Metabolomic prediction of yield in hybrid rice. *Plant J* 88:219–227. <https://doi.org/10.1111/tpj.13242>
- Yan H, Kang MZ, De Reffye P, Dingkugn M (2004) A dynamic, architectural plant model simulating resource-dependent growth. *Ann Bot* 93:591–602. <https://doi.org/10.1093/aob/mch078>
- Yang Z-P, Li H-L, Guo D, Peng S-Q (2015) Identification and characterization of the EIN3/EIL gene family in *Hevea brasiliensis*. *Tree Genet Genomes* 11:51. <https://doi.org/10.1007/s11295-015-0877-7>
- Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genet* 38:203–208. <https://doi.org/10.1038/ng1702>
- Yuan W, Flowers JM, Sahraie DJ, et al (2016) Extreme QTL mapping of germination speed in *Arabidopsis thaliana*. *Mol Ecol* 25:4177–4196. <https://doi.org/10.1111/mec.13768>
- Zinsmeister J, Leprince O, Buitink J (2020) Molecular and environmental factors regulating seed longevity. *Biochem J* 477:305–323. <https://doi.org/10.1042/BCJ20190165>
- Zivcak M, Brestic M, Sytar O (2016) Osmotic adjustment and plant adaptation to drought stress. In: Hossain M, Wani S, Bhattacharjee S, Burritt D, Tran LS (eds) *Drought stress tolerance in plants*, Vol 1. Springer, Cham. [https://doi.org/10.1007/978-3-319-28899-4\\_5](https://doi.org/10.1007/978-3-319-28899-4_5)
- Zenke-Philippi C, Thiemann A, Seifert F, et al (2016) Prediction of hybrid performance in maize with a ridge regression model employed to DNA markers and mRNA transcription profiles. *BMC Genomics* 17:262. <https://doi.org/10.1186/s12864-016-2580-y>
- Zhai Y, Cai S, Hu L, et al (2019) CRISPR/Cas9-mediated genome editing reveals differences in the contribution of INDEHISCENT homologues to pod shatter resistance in *Brassica napus* L. *Theor Appl Genet* 132:2111–2123. <https://doi.org/10.1007/s00122-019-03341-0>
- Zhong C, Jian S-F, Huang J, et al (2019) Trade-off of within-leaf nitrogen allocation between photosynthetic nitrogen-use efficiency and water deficit stress acclimation in rice (*Oryza sativa* L.). *Plant Physiol Biochem* 135:41–50. <https://doi.org/10.1016/j.plaphy.2018.11.021>
- Zhu X, Leiser WL, Hahn V, Würschum T (2021) Phenomic selection is competitive with genomic selection for breeding of complex traits. *Plant Phenome J* 4:e20027. <https://doi.org/10.1002/ppj2.20027>
- Zhu X, Maurer HP, Jenz M, et al (2022) The performance of phenomic selection depends on the genetic architecture of the target trait. *Theor Appl Genet* 135:653–665. <https://doi.org/10.1007/s00122-021-03997-7>



## **ANNEXE**

---

### **ACTIVITES SCIENTIFIQUES**

---



## Article avec comité de lecture

**Laurençon M.**, Legrix J., Wagner M-H., Demilly D., Baron C., Rolland S., Ducournau S., Laperche A., Nesi N. Genomic and phenomic predictions help capture low-effect alleles promoting seed germination in oilseed rape in addition to QTL analyses. *Theor. Appl. Genet.* 137, 156 (2024). <https://doi.org/10.1007/s00122-024-04659-0>

## Communication orale avec acte

**Laurençon M.**, Legrix J., Wagner M-H., Demilly D., Baron C., Ducournau S., Laperche A., Nesi N. – Genetic control of germination-related traits in a large panel of rapeseed (*Brassica napus*) and the potential of genomic and phenomic selection to predict fast germinating ability. 16<sup>th</sup> International Rapeseed Congress, 24-27 Sept. 2023, Sydney, Australie

**Laurençon M.**, Legrix J., Wagner M-H., Demilly D., Baron C., Ducournau S., Laperche A., Nesi N. Prédictions génomiques et phénotypiques pour améliorer la germination chez le colza. Séminaire ASF section « outils et méthodes en appui à la création variétale », 18-19 mars 2024, Montpellier, France

## Communications orales sans acte

**Laurençon M.** Déterminants génétiques de la germination et de la vigueur précoce chez le colza. Carrefours Colza 2024, 26-27 Janvier 2024, Bruz, France

**Laurençon M.** Genetic control of germination-related traits in rapeseed and the potential of genomic and phenomic selection to predict fast germinating ability. Séminaire R2D2, 22-24 Novembre 2023, Angers, France

**Laurençon M.** Déterminants génétiques de la germination et de la vigueur précoce chez le colza. Carrefours Colza 2024, 25-26 Janvier 2024, Bruz, France

**Laurençon M.** Prédictions génomiques et phénotypiques pour améliorer la germination chez le colza. Forum de l'IGEPP, 21-22 Mars 2024, Saint-Jacut-de-la-Mer, France

**Laurençon M.** Genomic and phenomic predictions help capture low-effect alleles promoting seed germination in oilseed rape in addition to QTL analyses. Journées de l'ED EGAAL, 21-23 Mai 2024, Paimpol, France

## Communications écrites (posters)

**Laurençon M.**, Rolland S., Alix E., Guichard S., Moulin B., Nesi N., Laperche A. – Impact of the type of spectra on phenomic selection for early vigor prediction in oilseed rape. 18<sup>th</sup> Eucarpia Biometrics in Plant Breeding Conference, 21-23 Sept. 2022, Paris, France

**Laurençon M.**, Alix E., Carrillo A., Guichard S., Jumel S., Moulin B., Richard-Molard C., Nesi N., Laperche A. – Leaf cover dynamics highlighted old spring germplasm as promising resources to improve early vigor in oilseed rape. 16<sup>th</sup> International Rapeseed Congress, 24-27 Sept. 2023, Sydney, Australie

**Laurençon M.**, Elise Alix, Cécile Baron, Solenn Guichard, Stéphane Jumel, Bernard Moulin, Marion Gauthier, Céline Richard-Molard, Nathalie Nesi, Anne Laperche – *Early vigor in rapeseed depends of the germplasm type and the genetic determinants were partly selected during domestication.* 7<sup>th</sup> International Conference of Quantitative Genetics, 22-26 Juillet 2024, Vienne, Autriche



## **Formations**

MOOC Ethique de la recherche – Formation en ligne de 15h, réalisée en 2022.

Formation statistiques avancées – Formation INRAE de 21h, 28-29 novembre et 5 décembre 2022.

MOOC machine learning in Python with scikit-learn – Formation en ligne de 36h, réalisée en 2022.

Cours GWAS, de la formation GAP/GGAP – Institut Agro Rennes, le 16 novembre 2021.

## **Implication dans la vie scientifique**

Organisation de la journée des doctorants 2023 de l'UMR IGEPP – le 23 mars 2023.

Animations récurrentes de journal club - UMR IGEPP équipe RCA, années 2021 à 2024.

Représentant des doctorants de l'IGEPP au conseil d'unité – années 2022 à 2024.

## **Enseignement**

Vacation d'enseignement à Institut Agro Rennes-Angers – Chargé de TD en génétique quantitative (étudiants de niveaux M2 et 3<sup>ème</sup> année ingénieurs) ; 102h soit 25.5h équivalente de formation.

**Titre :** Déterminants génétiques et prédictions de la germination et de la vigueur précoce chez le colza

**Mots clés :** *Brassica napus*, QTL, Germination, Vigueur précoce, Prédictions, NIRS

**Résumé :** L'implantation de la culture de colza est une phase critique qui conditionne le rendement final, et ce en raison des nombreux stress biotiques et abiotiques en début de cycle. Utiliser le levier variétal pour sélectionner des plantes vigoureuses durant les phases précoces du cycle est une solution pour réussir l'implantation de la culture et garantir le potentiel de rendement. Dans ce contexte, ce travail de thèse a cherché à répondre aux questions suivantes : quel est le déterminisme génétique de la germination et de la vigueur précoce ? Ces traits étant potentiellement polygéniques, est-il possible de les prédire ? Pour cela, un panel de diversité de 233 génotypes de colza (*Brassica napus* L.) a été phénotypé pour un ensemble de traits physiologiques et de développement afin de décomposer la germination et la vigueur

précoce. Les analyses génétiques par association ont mis en évidence la nature polygénique de ces traits par l'identification de 17 et 21 QTL respectivement. Le potentiel de prédition des traits a été évalué par des approches de prédictions génomiques (à partir de données moléculaires) ou phénotypiques (à partir de données spectrales). Les deux approches se sont révélées adaptées pour la prédition de ces traits. Néanmoins, la variabilité de performance entre les modèles pour les deux approches suggère l'utilisation de modèles spécifiques en fonction des traits et de la population considérée. L'architecture génétique et les prédictions des traits de germination et de vigueur précoce sont discutées, ainsi que de la valorisation des ressources génétiques dans des schémas de sélection

**Title:** Genetic determinants and predictions of seed germination and early vigor in oilseed rape

**Keywords:** *Brassica napus*, QTL, Seed Germination, Early Vigor, Predictions, NIRS

**Abstract:** Oilseed rape establishment is a critical phase for yield establishment, due to the numerous biotic and abiotic stresses at the beginning of the cycle. Using genetic improvement of varieties to obtain vigorous plants during the early development stages would enable the crop to successfully establish and thus guarantee yield potential. In this context, this thesis aimed to answer the following questions: what is the genetic determinism of seed germination and early vigor? As these traits are potentially polygenic, is it possible to predict them? To this end, a panel of 233 rapeseed genotypes (*Brassica napus* L.) was phenotyped for a set of physiological and developmental traits in order to decompose seed germination and early vigor.

Genome-wide association analyses highlighted the polygenic nature of these traits by identifying 17 and 21 QTL respectively. In addition, the predictive potential of these traits was assessed using genomic prediction approaches (based on molecular data) and phenomic prediction approaches (based on spectral data). Both approaches were found to be suitable for predicting these traits. However, the variability in performance between models of the two approaches suggests the use of specific models regarding traits and the population considered. Genetic architecture and prediction of establishment traits are discussed, as well as the use of genetic resources in pre-breeding.