



**HAL**  
open science

# Identification des déterminants génétiques et moléculaires contrôlant la stabilité du rendement chez le colza d'hiver

Erwan Corlouer

► **To cite this version:**

Erwan Corlouer. Identification des déterminants génétiques et moléculaires contrôlant la stabilité du rendement chez le colza d'hiver. Sciences agricoles. Agrocampus Ouest, 2020. Français. NNT : 2020NSARC149 . tel-03191599

**HAL Id: tel-03191599**

**<https://theses.hal.science/tel-03191599>**

Submitted on 7 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THESE DE DOCTORAT DE

*L'institut national d'enseignement supérieur pour l'agriculture, l'alimentation  
et l'environnement*

Ecole interne AGROCAMPUS OUEST

ECOLE DOCTORALE N° 600

*Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation*

Spécialité : « Génétique, génomique et bio-informatique »

Par

**Erwan CORLOUER**

## **Identification des déterminants génétiques et moléculaires contrôlant la stabilité du rendement chez le colza d'hiver**

Thèse présentée et soutenue à Rennes, le 19 novembre 2020

Unité de recherche : UMR IGEPP, INRAE, Institut Agro, Université de Rennes, le Rheu

Thèse N° : C-149 2020-19

### **Rapporteurs avant soutenance :**

Nicolas Langlade  
Bénédicte Quilot-Turion

Directeur de recherche, INRAE, Laboratoire des interactions plantes microorganismes  
Directrice de recherche, INRAE, Unité Génétique et amélioration des fruits et légumes

### **Composition du Jury :**

Président : Maria Manzanares-Dauleux Professeure, Institut Agro – Agrocampus Ouest

Examineurs : Sophie Brunel-Muguet Chargée de Recherche, UMR Ecophysiologie Végétale, Agronomie  
Christopher Sauvage Genetic Project Leader, Syngenta

Dir. de thèse : Nathalie Nesi  
Co-enc. de thèse : Anne Laperche

Directrice de recherche, UMR IGEPP  
Maître de Conférence, Institut Agro – Agrocampus Ouest



Identification des  
déterminants génétiques et  
moléculaires contrôlant la  
stabilité du rendement chez  
le colza d'hiver



# Remerciements

---

Mon aventure à l'IGEPP de l'INRAe du Rheu commence en réalité en 2016, lorsque j'ai commencé mon stage au cours duquel j'ai commencé à travailler sur les données RAPSODYN pour finalement les valoriser au sein de ce travail de thèse. Toute cette aventure m'a beaucoup apporté sur un bon nombre d'aspects grâce à l'ensemble des personnes que j'ai pu côtoyer de près ou de loin. Ce travail est l'occasion de tous les remercier.

Mais avant de commencer, je tenais à remercier les membres de mon jury de thèse, Bénédicte Quilot-Turion et Nicolas Langlade pour avoir accepté d'être les rapporteurs de ce travail, ainsi que Sophie Brunel-Muguet, Maria Manzanarès-Dauleux et Christopher Sauvage qui ont accepté d'être les examinateurs de mes travaux.

Bien évidemment, j'adresse des remerciements tous particuliers à mes deux encadrantes, sans qui rien de tout cela n'aurait été possible, Nathalie Nesi et Anne Laperche. Merci de m'avoir accordé votre confiance, d'avoir cru en moi quand moi-même je doutais. Merci de m'avoir offert l'opportunité de travailler sur ce sujet de thèse. J'ai énormément appris à vos côtés. Vous avez su également ouvrir mon esprit et ma curiosité à travers d'autres projets. Sans vous, rien de tout cela n'aurait été possible. Encore une fois MERCI.

Je ne saurais oublier chacun des membres de l'équipe « Rendement sous Contraintes Abiotiques » (RCA). Merci pour votre accueil, les diverses discussions, pour votre volontarisme quand je vous forçais à venir faire une présentation en lab meeting, pour tous ces moments conviviaux et puis plus récemment pour ce nouveau format de réunion plénière où le déguisement est de rigueur !

Je remercie également les membres de mon comité de thèse Laurence Moreau, Arnaud Gauffreteau, Marie-Laure Pilet-Nayel pour vos conseils, les discussions et votre regard extérieur sur mon travail.

Je tenais également à remercier l'ensemble des membres du projet RASOPDYN avec lesquels j'ai pu travailler et discuter tout au long de ces années.

Je remercie naturellement le LBDF et l'ensemble de ces occupants passés et présents (Anne-Sophie, Alexandre, Yoann, Sylvain, Benjamin, Julie, Thanina, Grégoire, Franz, Youssef et Anani). Merci pour tous ces bons moments partagés, nos discussions scientifiques et parfois un peu moins, les soirées au bar ou chez Benjamin ! Mention spéciale à Yoann et Sylvain pour notre belle aventure du 4L trophy en pleine thèse mais qui s'est arrêtée après l'achat des 4L, l'association R4LLYES reste le signe de notre folie. Je n'oublie pas non plus la pause-café du bureau BP, merci à Anne-Marie et Mathieu pour



avoir hébergé ces moments de convivialité et à ceux qui la partage avec nous (Gauthier, Julien, Loeiz et tous les autres). Mathieu, je ne t'oublie pas on finira par aller voir le stade Rochelais perdre face au Racing 92 à l'Aréna.

Je remercie Ophélie Guilbert, Lucile Retailleau et Romain Rio pour votre travail respectif lors de vos stages et de m'avoir également appris à encadrer et à partager les connaissances.

Pour finir, je voulais terminer rapidement en remerciant mes proches et en particulier ma conjointe qui m'ont encouragé, soutenu et ont toujours été là quand il fallait !



# Table des matières

---

Liste des abréviations .....	
Contexte général .....	1
Chapitre I : Synthèse bibliographique .....	6
<b>I.1. Génétique et génomique du colza</b> .....	<b>6</b>
I.1.1. Origine du colza et conséquences sur l'organisation du génome .....	6
I.1.2. Ressources génomiques disponibles chez le colza et les autres Brassicaceae .....	7
I.1.3. Les ressources génétiques de <i>Brassica napus</i> .....	10
I.1.4. Etendue du déséquilibre de liaison dans les populations de colza .....	11
<b>I.2. Cycle du colza d'hiver et élaboration du rendement en grains</b> .....	<b>13</b>
I.2.1. Mise en place des composantes du rendement chez le colza .....	13
I.2.2. Les facteurs limitant le rendement chez le colza d'hiver .....	14
<b>I.3. Interactions G×E et QTL×E</b> .....	<b>19</b>
I.3.1. Définition des interactions G×E et utilisation dans un contexte de sélection variétale .....	19
I.3.2. Méthodes descriptives pour l'analyse des interactions G×E .....	20
I.3.3. Prise en compte de l'environnement pour soutenir les interactions G×E .....	22
I.3.4. Analyses QTL et interaction G×E .....	24
<b>I.4. Problématique et objectifs de la thèse</b> .....	<b>28</b>
Chapitre II : Identification de facteurs limitant le rendement chez le colza pour caractériser les interactions G×E .....	30
<b>II.1. Préambule</b> .....	<b>30</b>
<b>II.2. Description du réseau expérimental</b> .....	<b>32</b>
<b>II.3. Envirotypes based on seed yield limiting factors allow to tackle G×E interaction</b> .....	<b>35</b>
<b>II.4. Prise en compte des itinéraires techniques azotés contrastés dans l'envirotypage</b> .....	<b>59</b>
II.4.1. Définitions des indicateurs liés à la fertilisation azotée .....	59
II.4.2. Sélection des indicateurs informatifs et analyse des corrélations .....	61



II.4.3. Envirotypage.....	62
<b>II.5. Conclusions .....</b>	<b>65</b>
<b>Chapitre III : Architecture génétique de la stabilité du rendement chez le colza..</b>	<b>67</b>
<b>III.1. Intérêt de l’approche d’envirotypage pour caractériser les interactions QTL×Essai .....</b>	<b>67</b>
<b>III.2. Un nouvel envirotypage dédié pour les analyses génétiques .....</b>	<b>69</b>
III.2.1. Définitions de nouveaux envirotypes pour le réseau #3.....	69
III.2.2. Présentation des traits étudiés.....	71
<b>III.3. Analyses génétiques .....</b>	<b>72</b>
III.3.1. Plant material and genotyping data .....	72
III.3.2. Traits and phenotypic data analyses .....	72
III.3.3. GWAS analysis .....	75
III.3.4. Comparisons between the three different scales .....	77
III.3.5. Getting deeper into the use of the envirotype QTL .....	78
<b>Chapitre IV : Discussion générale et perspectives .....</b>	<b>92</b>
<b>IV.1. Des indicateurs environnementaux pour expliquer l’interaction G×E.....</b>	<b>92</b>
IV.1.1. Température et radiations à la floraison sont les principaux facteurs limitants du rendement et expliquent une part de l’interaction G×E.....	92
IV.1.2. Intérêt des modèles de culture pour affiner la liste des indicateurs.....	93
IV.1.3. Adapter l’envirotypage à des systèmes de cultures plus complexes.....	94
<b>IV.2. Les déterminants génétiques de l’interaction.....</b>	<b>96</b>
IV.2.1. Peu de QTL détectés pour le rendement .....	96
IV.2.2. QTL de stabilité vs QTL interactifs.....	97
<b>IV.3. Utilisation des envirotypes en sélection .....</b>	<b>98</b>
IV.3.1. Les QTL détectés à l’échelle des envirotypes mettent en évidence des zones d’intérêt pour la sélection assistée par marqueurs .....	98
IV.3.2. Relation entre QTL et facteurs limitants .....	98
IV.3.3. Utilisation des facteurs limitants et des envirotypes en sélection génomique .....	99



Références bibliographiques .....	101
Annexe A : Script PLS.....	
Annexe B : SEEDPROT (side project).....	
Annexe C : Activités Scientifiques .....	



# Liste des abréviations

---

## A

ABK : Ancestral Brassicaceae karyotype  
ACK : Ancestral Camelinae karyotype  
AMMI : Additive main effect and multiplicative interaction  
AN : Aerial nitrogen  
Aut : Autainville

## B

B : Bolting  
BC : Population en Back-Cross  
Blo : Blois

## C

Ch : Châteauroux  
Chl : Châlons-en-Champagne  
Chr : Chartres  
Cup : Cuperly  
Cv : Ormoy  
CW : Climatic winter

## D

DAB : Dry aerial biomass  
DAI : Dry days after input  
DBI : Dry days before input  
DENS : Densité  
Dij : Dijon  
DL : Déséquilibre de liaison  
DS127 : Diversity set of 127 winter oilseed rape

## E

ETP : Evapotranspiration

## F

F : Fall  
FLO : Flowering  
FR : Frost

## G

GDD : Growing degree days  
G×E : Interaction Génotype × Environnement

GWAS : Analyse en association

## H

h<sup>2</sup> : héritabilité  
HD : Haploïde doublé  
HT : High temperature

## I

INN : Indice de nutrition azotée

## K

K : Potassium  
Kc : coefficient de culture

## L

LA : Analyse de liaison  
LF : Less fractionated genome  
LGDD : Length of a period in GDD  
Liv : Liverdy  
Lou : Louville  
LR : Le Rheu  
LSR : Lack of solar radiation  
LT : Low temperature

## M

Md : Mondonville  
MF1 : Medium fractionated genome  
MF2 : Most fractionated genome

## N

N : Nitrogen / Nitrogen Input  
NAE : Nitrogen assimilation efficiency  
NNI : Nitrogen nutrition Index  
NRE : Nitrogen remobilization efficiency  
NUE : Nitrogen use efficiency  
NUpE : Nitrogen uptake efficiency  
NUtE : Nitrogen utilization efficiency

## O

OSR : Oilseed rape



## P

P300 : Period of seed number fixation  
P600 : Period related to reserve allocation to the pod growth  
P1000 : Period related to reserve allocation to the seed growth  
P : Phosphore  
PCK : Proto-Calepineae Karyotype  
PLS : Partial least square  
Pre : Prêmesques

## Q

QPT : Quotient phothermique  
QTL : Quantitative trait loci

## R

R/RF : Rainfalls  
RIL : Lignée recombinante  
RO : Run off

## S

S: Soufre  
Sel : Selommes  
SN : Seed number  
SNP : Single nucleotid polymorphism  
SR : Solar radiation  
SSR : Sum of solar radiation  
SVD : Variables singulères  
SW : Seed weight  
SY : Seed yield

## T

TMAX : Maximal temperature  
TMIN : Minimal temperature  
TMN : Mean temperature  
TPE : Target population of environments

## V

Ver : Verpillères  
Vern : Optimal vernalization condition  
Vil : Villemardy

## W

W<sup>2</sup>g : Ecovalence variétale  
WD : Water deficiency  
WS : Water stress

WSC : Water soil content  
WOSR : Winter oilseed Rape

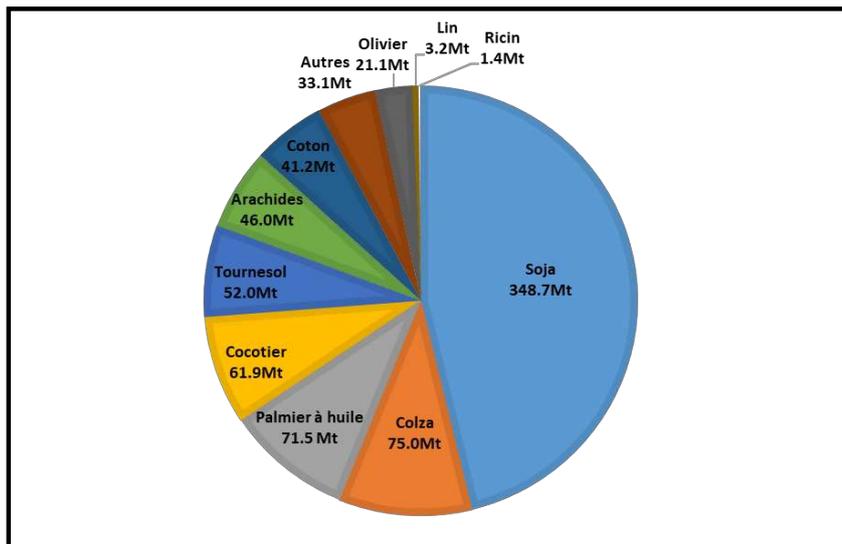
## Y

Yeb : Yéblès



# CONTEXTE GENERAL

---



**Figure 1 : Production mondiale de graines des principales cultures oléo-protéagineuses en 2018 (en millions de tonnes, Mt)**

Le groupe 'Autres' comprend les cultures de sésame, karité, jojoba, carthame, chanvre et moutardes (données FAOSTAT, 2020).

# Enjeux et contraintes de la production de colza

---

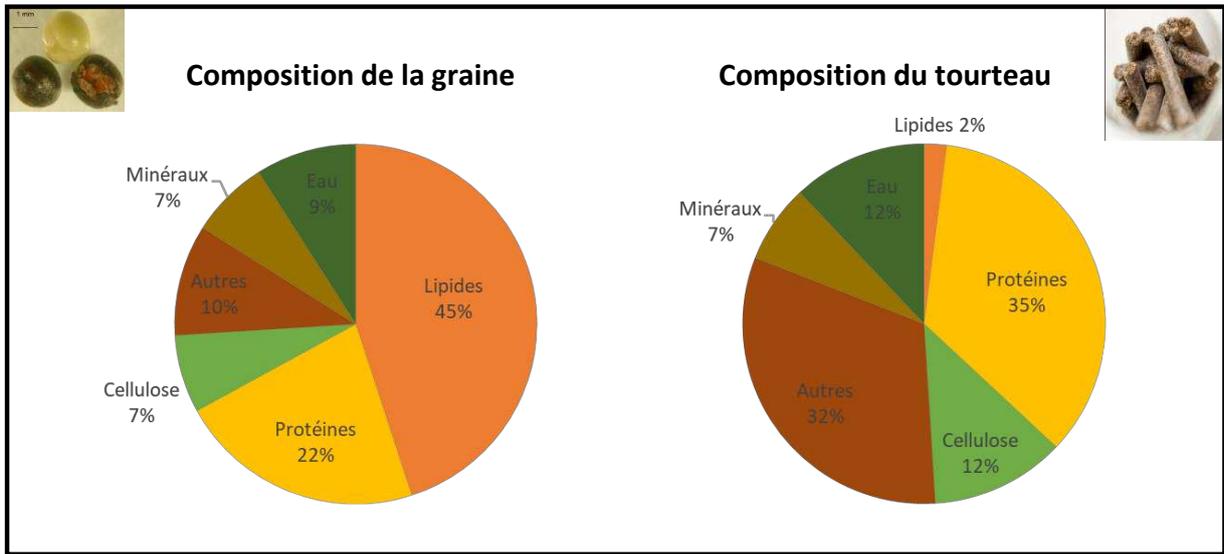
## *Evolution de la demande et de l'offre en huiles et protéines végétales*

En 2030, le monde devra nourrir 8.5 milliards d'habitants (soit +10% par rapport à 2020 ; ONU, 2019). Cette croissance démographique combinée à l'augmentation des niveaux de vie va entraîner une hausse de la consommation alimentaire en huiles et protéines (1<sup>ère</sup> ou 2<sup>nde</sup> transition alimentaire selon les régions). Cette hausse va impacter notablement les besoins en produits d'origine végétale, avec des prévisions de l'ordre de +40% pour les huiles, +43% pour les protéines végétales et +53% pour les tourteaux (BIPE & Sofiproteol, 2014). Concernant l'offre, les ressources en huiles végétales seront importantes dans le futur, dépassant même la demande (+8 Mt en 2030), permettant ainsi de satisfaire les demandes alimentaires, énergétiques et chimiques. En revanche, l'offre en protéines et en tourteaux sera inférieure aux demandes (déficit prévisionnel de 58 Mt de tourteaux en 2030). L'enjeu des cultures oléo-protéagineuses sera donc de pouvoir fournir des produits de qualité en quantité suffisante pour pallier à la demande croissante en huiles et protéines végétales, et cela dans un cadre de production agricole de plus en plus contraint, par la diminution des intrants d'une part et par l'augmentation de l'incidence des événements météorologiques extrêmes liés au changement climatique d'autre part.

En 2018, la production annuelle mondiale de graines oléo-protéagineuses s'élevait à 754.1 Mt pour une surface cultivée totale de 267.4 Mha (Figure 1 ; FAOSTAT, 2020). La multiplicité des cultures oléo-protéagineuses offre une certaine complémentarité en termes d'approvisionnement (répartition géographique) et de qualité des produits pour les différents usages (composition en acides gras des huiles et en acides aminés des protéines). Cependant, les échanges de matières sur les marchés internationaux sont complexifiés et fortement impactés par les niveaux de production (fluctuations des récoltes annuelles) et le contexte géopolitique (pression exercée par la Chine sur la production de soja ; pression de la balance euro/dollar modifiant les équilibres import/export ; prix des huiles ...).

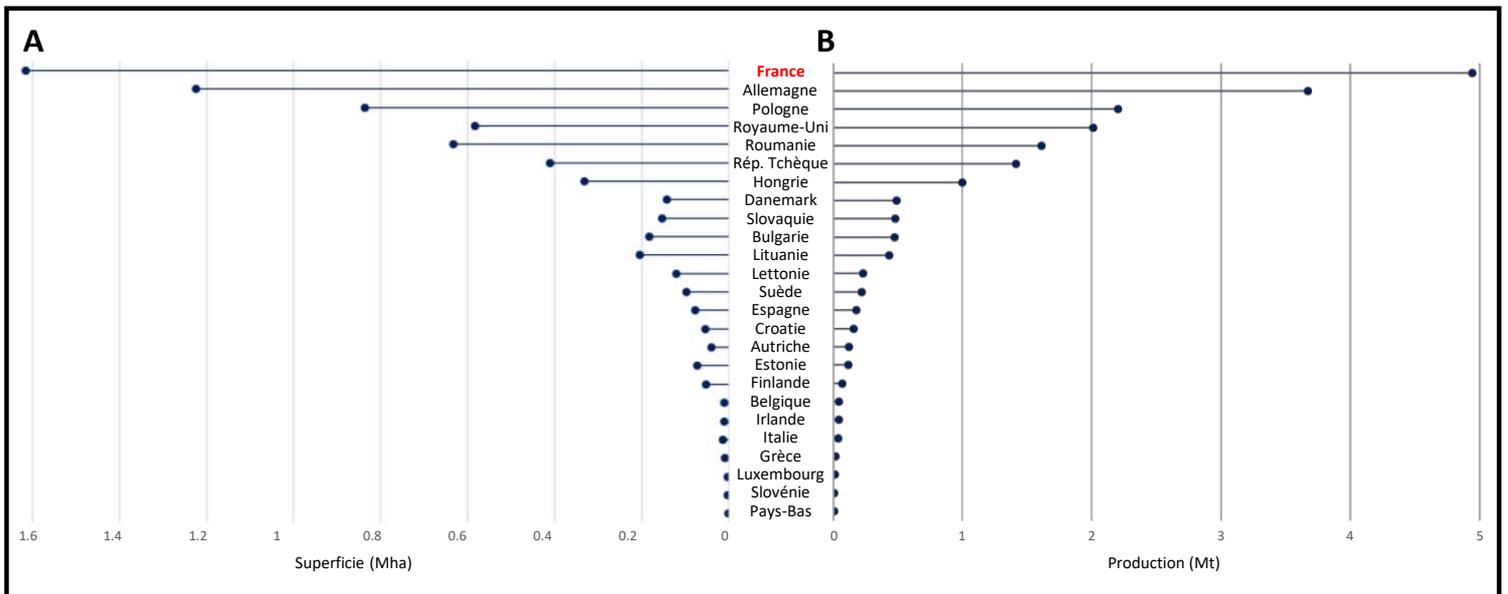
## *Place du colza dans la production d'huile et de protéines végétales*

Avec 75 Mt de graines produites annuellement, le colza occupe la seconde place dans les cultures oléo-protéagineuses, loin derrière le soja et juste devant le palmier à huile (Figure 1). Les constituants de la graine de colza sont l'eau, l'huile, les protéines et les résidus qui comprennent



**Figure 2 : Composition de la graine et du tourteau de colza**

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



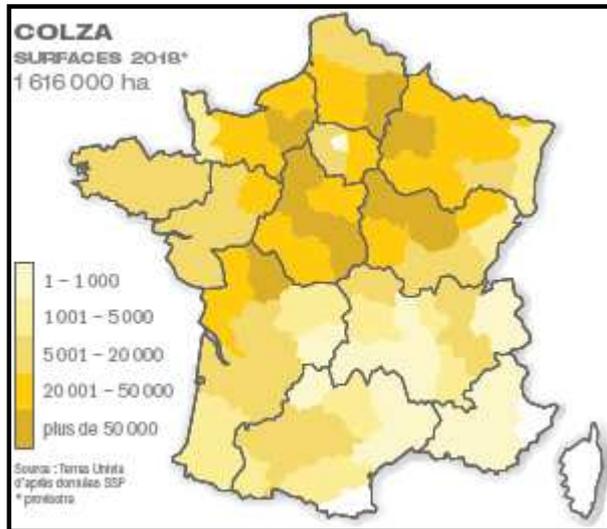
**Figure 3 : Culture du colza dans l'Union Européenne en 2018**

(A) Surfaces cultivées en millions d'hectares (Mha). (B) Production de graines en millions de tonnes (Mt) (données FAOSTAT, 2020).

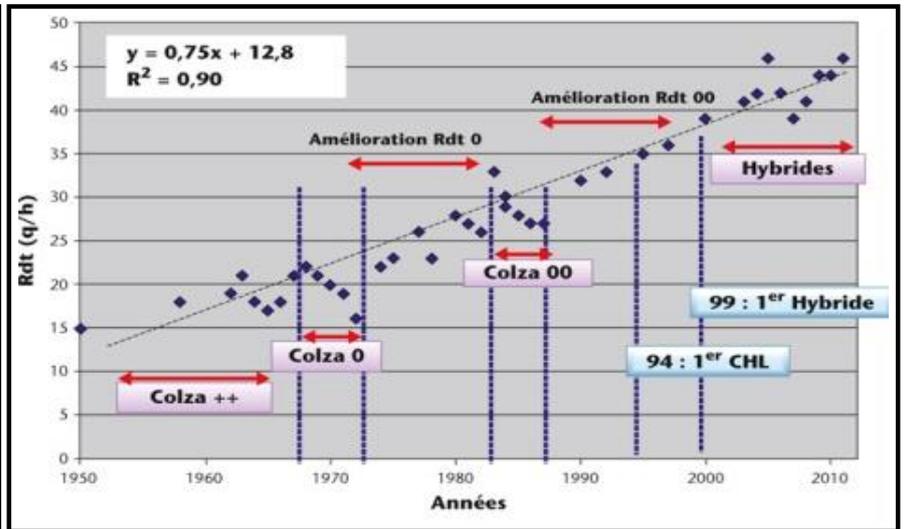
principalement les glucides, les fibres, les minéraux et les cendres (Figure 2 ; Nesi et al., 2008). Par le passé, les efforts de sélection ont principalement porté sur la teneur en lipides qui avoisine désormais 45-50% (w/w), faisant de la graine de colza un produit de choix pour la filière des huiles 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. Très récemment, des programmes de recherche et de sélection ont été initiés en Europe et au Canada visant à améliorer la qualité du tourteau de colza 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.

Les principales zones de production de colza sont l'Union Européenne (UE), le Canada et la Chine (productions respectives de 20 Mt, 20.3 Mt, 13.3 Mt de graines en 2018 ; FAOSTAT, 2020). En fonction des régions, des variétés d'hiver ou de printemps sont implantées. Ainsi, les colzas de type hiver sont cultivés dans les régions à hiver doux (Europe de l'Ouest), ils sont semés à la fin de l'été et requièrent une période de vernalisation pour fleurir. A l'inverse, les cultivars de printemps sont semés à la fin de l'hiver et prédominent dans les régions septentrionales (Europe de l'Est, Canada, Asie) ainsi qu'en Australie. Leur rendement est sensiblement inférieur aux variétés d'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).

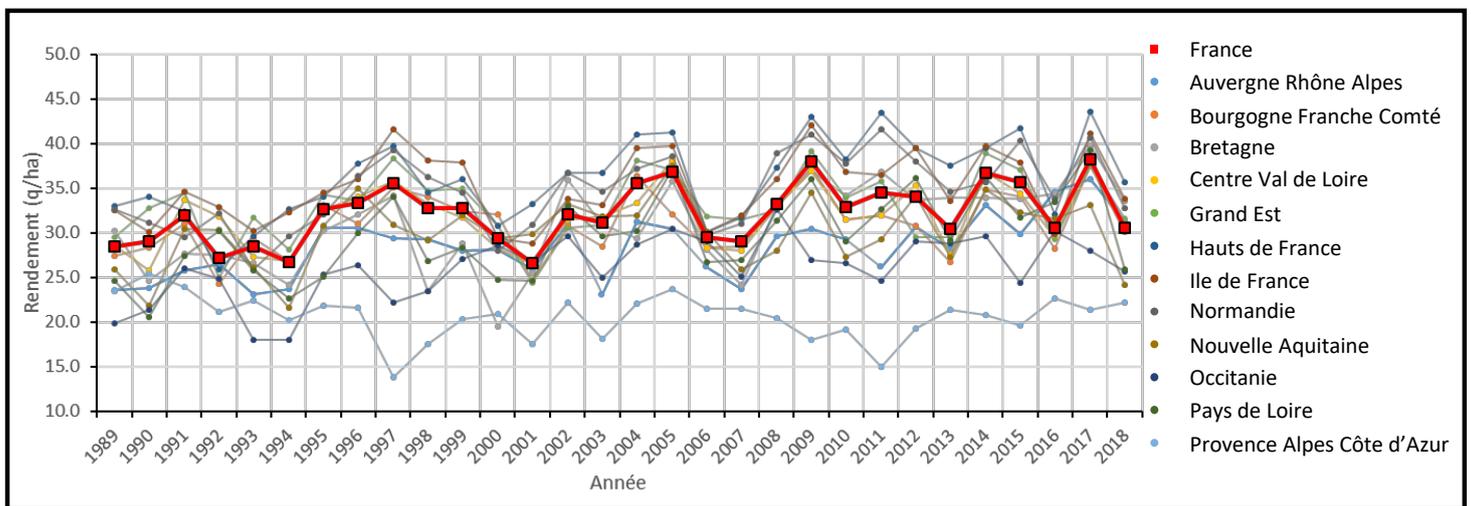
Au sein de l'UE, le colza est la première culture oléagineuse, devant le tournesol (colza/tournesol en 2018 : 20 Mt/10 Mt – 29/24.8 q/ha). Près de 45% de la production européenne est assurée par la France et l'Allemagne (Figure 3) avec un net avantage pour l'Hexagone lors des récoltes 2018 et 2019 (environ +25-30%).



**Figure 4 : Surfaces cultivées de colza par région française en 2018 (d'après Terres Univia, 2018)**



**Figure 5 : Effet du progrès génétique sur le rendement de colza**  
Rendements relevés dans les essais des réseaux d'inscription du Comité Technique Permanent de la Sélection (CTPS) entre 1950 et 2010 (Pinochet et Renard, 2012).



**Figure 6 : Rendements (q/ha) de la culture de colza en France métropolitaine et par région (1989 – 2018)**

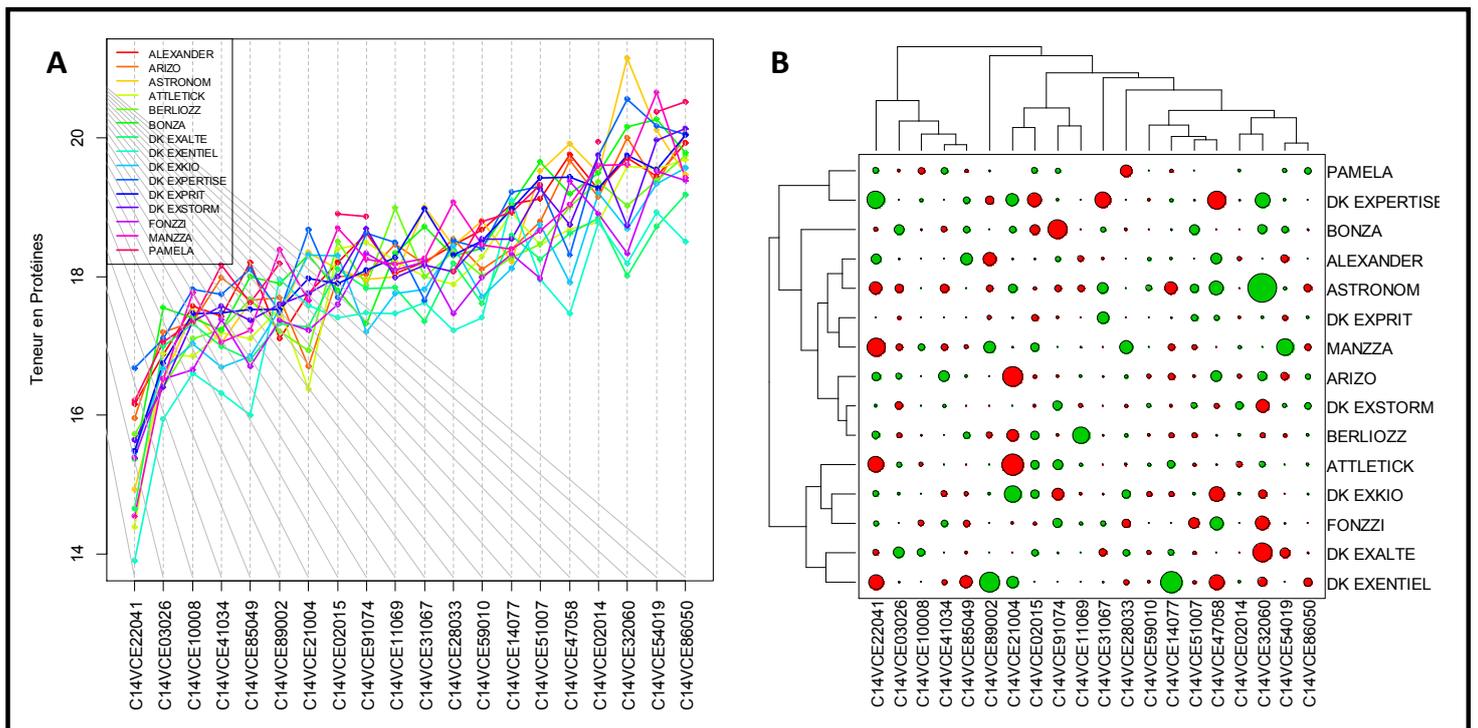
Source: Terres Inovia et Terres Univia d'après les données du Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt, 2020.

## *La culture de colza d'hiver oléagineux en France : entre forces et faiblesses, menaces et opportunités*

Le colza est devenu une culture majeure des assolements en France (~ 1.6 Mha) stimulée depuis 15 ans par le marché du biodiesel et l'effort de création et de renouvellement variétal (développement des hybrides) déployé par les sociétés semencières (Charbonnier et al., 2019). Les principales régions françaises productrices de colza sont la Bourgogne-Franche-Comté, le Centre-Val de Loire, le Grand Est, les Hauts-de-France, l'Île-de-France et la Normandie (Figure 4). Les rendements moyens ont quasiment doublé en 50 ans, soutenus en partie par le progrès génétique (Figure 5 ; FAOSTAT, 2020). Néanmoins, les surfaces plantées en colza en France reculent régulièrement depuis 2015 (-25% en 5 ans ; Terres Inovia, 2020) ce qui s'explique par des baisses sensibles de production (-8 q/ha entre 2017 et 2018 à l'échelle nationale), en partie dues à des problèmes d'implantation et de végétation à l'automne (dégâts d'insectes) et/ou pendant l'hiver (froid/gel). Contrairement au blé qui constitue une valeur refuge, les décisions de semis de colza dépendent de plusieurs facteurs tels que les résultats de l'année précédente, les cours, les difficultés de mise en place en fin d'été et les mesures de la politique agricole commune (PAC). Cependant, en plus de produire de l'huile et de contribuer à l'autonomie protéique des élevages via la production de tourteaux, la culture de colza remplit positivement différents services écosystémiques (ex. diversification des rotations, amélioration de l'état organique et de la couverture des sols, limitation des adventices et des agents pathogènes dans la rotation). C'est également une espèce mellifère facilitant la biodiversité des insectes dans les parcelles. De plus, la culture du colza en association avec des légumineuses gélives est une pratique en fort développement (25% des surfaces de colza biologique sont conduites en association) car elle permet notamment d'optimiser la gestion des insectes ravageurs, la fertilisation azotée et la lutte contre les adventices, tout en limitant l'usage des intrants, ce qui constitue un levier d'amélioration prometteur du système de production, en plus du levier variétal.

### *Variations de rendement*

Des variations de rendement sont observées pour la production de colza en France et celles-ci sont plus ou moins marquées selon les régions et les campagnes de culture (Figure 6). Plusieurs hypothèses peuvent être avancées pour expliquer ces instabilités telles que les impacts du changement climatique et de certaines modifications des pratiques agricoles sur le développement du colza. En effet, le colza d'hiver a la particularité d'être implanté à une période souvent sèche et d'avoir un cycle de culture très long (~ 10-11 mois) l'exposant ainsi à de nombreuses contraintes abiotiques (stress thermique, hydrique, radiatif) et biotiques qui peuvent se combiner. Mais contrairement à d'autres espèces cultivées, dès lors que les conditions de production ne sont pas trop défavorables, le colza bénéficie d'une croissance dynamique à l'automne et d'une forte capacité



**Figure 7 : Teneur en protéines du colza dans le réseau CETIOM 2014**

(A) Les teneurs en protéines sont données pour 15 variétés testées dans 20 environnements. (B) Carte des interactions représentant la matrice des interactions G×E pour la teneur en protéines. Les génotypes sont en ligne et les environnements en colonne. Les interactions positives sont en vert et les négatives sont en rouge, le rayon du cercle est proportionnel à la valeur absolue de l'interaction. (Graphiques fournis par X. Pinochet).

de compensation au printemps (notamment via la mise en place de ramifications), ce qui lui permet bien souvent de maintenir sa productivité. Le changement climatique exerce une pression supplémentaire par la multiplication d'événements climatiques intenses avec, par exemple, une hausse des températures, un manque d'eau ou à l'inverse des précipitations trop importantes (Bell et al., 2004) pouvant exacerber les instabilités de rendement. Enfin, pour un développement optimal, le colza est très gourmand en apports azotés (besoins estimés entre 100 à 250 kg/ha ; Rathke et al., 2006) et soufrés (~75 kg/ha de SO<sub>3</sub> ; Hebinger, 2013), mais également en potassium (30 à 70 kg/ha) et en phosphore (40 à 150 kg/ha ; Terre Inovia, 2020). Or les directives européennes exigent une diminution du recours aux intrants chimiques et notamment azotés (Directive 'Nitrates' 91/676/CEE).

### *Effets de l'environnement et des interactions G×E*

Par ailleurs, au sein de réseaux très larges, il est fréquent de constater qu'une variété donnée est capable de produire différents phénotypes en fonction de son environnement, ce qui témoigne de sa plasticité phénotypique (Figure 7, exemple donné pour la teneur en protéines). La variabilité des traits phénotypiques de plusieurs accessions en fonction des environnements s'explique par la présence d'interactions Génotype × Environnement (G×E) qui seront largement détaillées dans la synthèse bibliographique. Le rendement en grains est fortement dépendant de l'environnement et des interactions G×E.

En conclusion, comme toutes les espèces cultivées, le colza doit s'adapter pour maintenir un service de production, dans des environnements contraints et fluctuants, tout en assurant de nouveaux services pour les agroécosystèmes. Dans ce contexte, un premier objectif vise à mieux identifier et caractériser les stress qui régissent les interactions G×E dans les zones de culture du colza à l'échelle nationale, et un second objectif est de caractériser le déterminisme génétique contrôlant les interactions G×E pour pouvoir, à termes, proposer une offre variétale adaptée.



# Contexte de la thèse

---

Ma thèse a porté sur l'étude des déterminants génétiques impliqués dans les interactions G×E comme un des leviers pour optimiser la stabilité des rendements. Les loci impliqués dans les interactions G×E pourront ensuite être utilisés dans les programmes de sélection pour adapter les variétés en fonction des conditions environnementales.

J'ai réalisé ma thèse au sein de l'équipe RCA (Rendements sous Contraintes Abiotiques) de l'Institut de Génétique, Environnement et Protection des Plantes (IGEPP). Les principaux travaux de recherche de l'équipe RCA consistent à identifier et caractériser les déterminants génétiques et physiologiques du rendement, de l'efficacité d'utilisation des nutriments et de la qualité de la graine chez le colza. Ma thèse a été dirigée par Anne Laperche (MC Institut Agro) et Nathalie Nesi (DR INRAE). Par ailleurs, mes recherches se sont inscrites dans le projet RAPSODYN (2012-2021) financé par l'Agence Nationale de la Recherche (ANR) dans le cadre du programme « Investissements d'Avenir » (projet ANR-11-BTBR-0004). Ce projet a pour objectif d'assurer la compétitivité à long terme du colza dans un contexte de réduction des intrants azotés et de changement climatique.

Ce manuscrit est présenté sous la forme :

- d'une première partie introductive qui vise à présenter la culture du colza, les ressources disponibles pour travailler cette espèce et la problématique des interactions G×E en génétique et amélioration des plantes ;
- d'un chapitre qui correspond à une publication (et des résultats complémentaires) décrivant une méthodologie pour identifier les facteurs limitants qui impactent le rendement et ainsi mieux caractériser les réseaux d'essais ;
- d'un chapitre qui correspond à l'utilisation des envirotypes dans les analyses génétiques, ce chapitre correspond à un article en cours de rédaction ;
- d'une discussion générale qui resitue les principaux résultats de la thèse au regard de la bibliographie et propose des perspectives de ce travail à court et plus long terme ;
- d'annexes, dont les résultats d'un travail sur les protéines de réserve de la graine de colza qui a été mené en « side-project ».

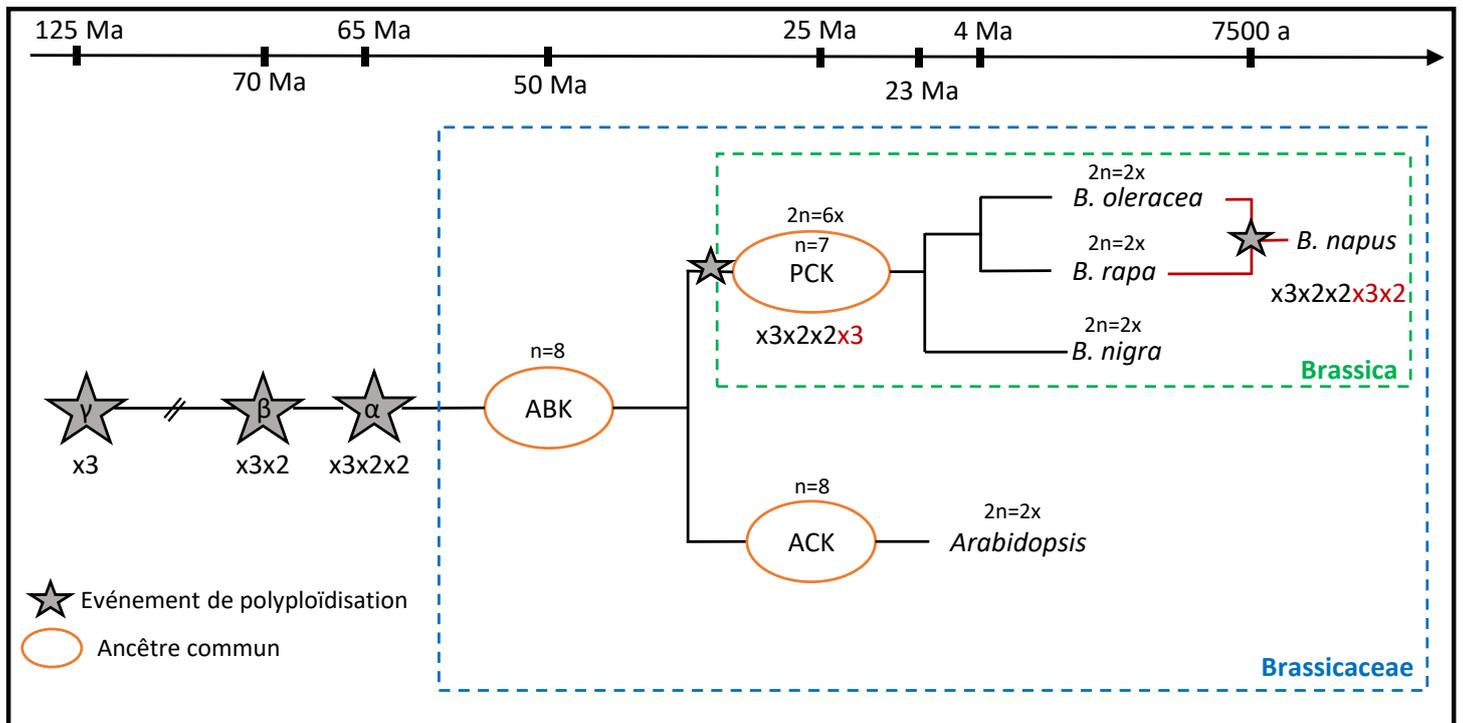


## CHAPITRE I

---

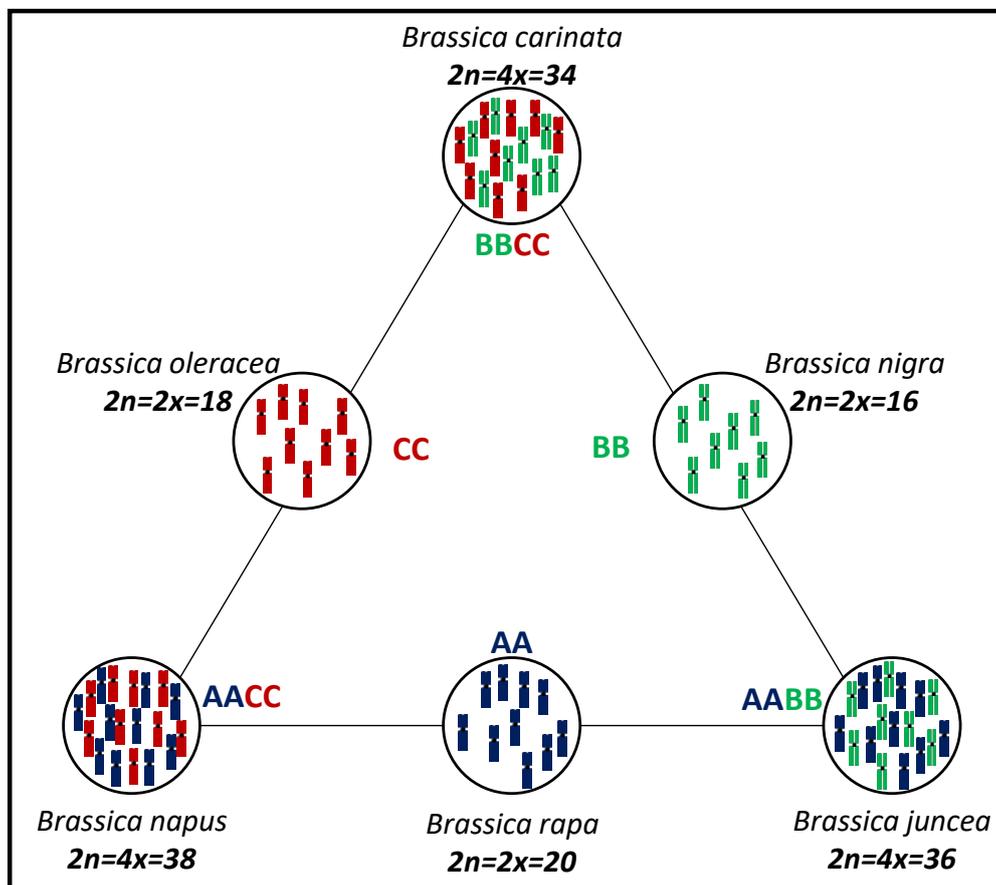
# SYNTHESE BIBLIOGRAPHIQUE

---



**Figure 1.1 : Origine du génome 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-Calepineae 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 1.2 : Relations entre différentes espèces du genre *Brassica* (d'après U, 1935)**

# I.1. Génétique et génomique du colza

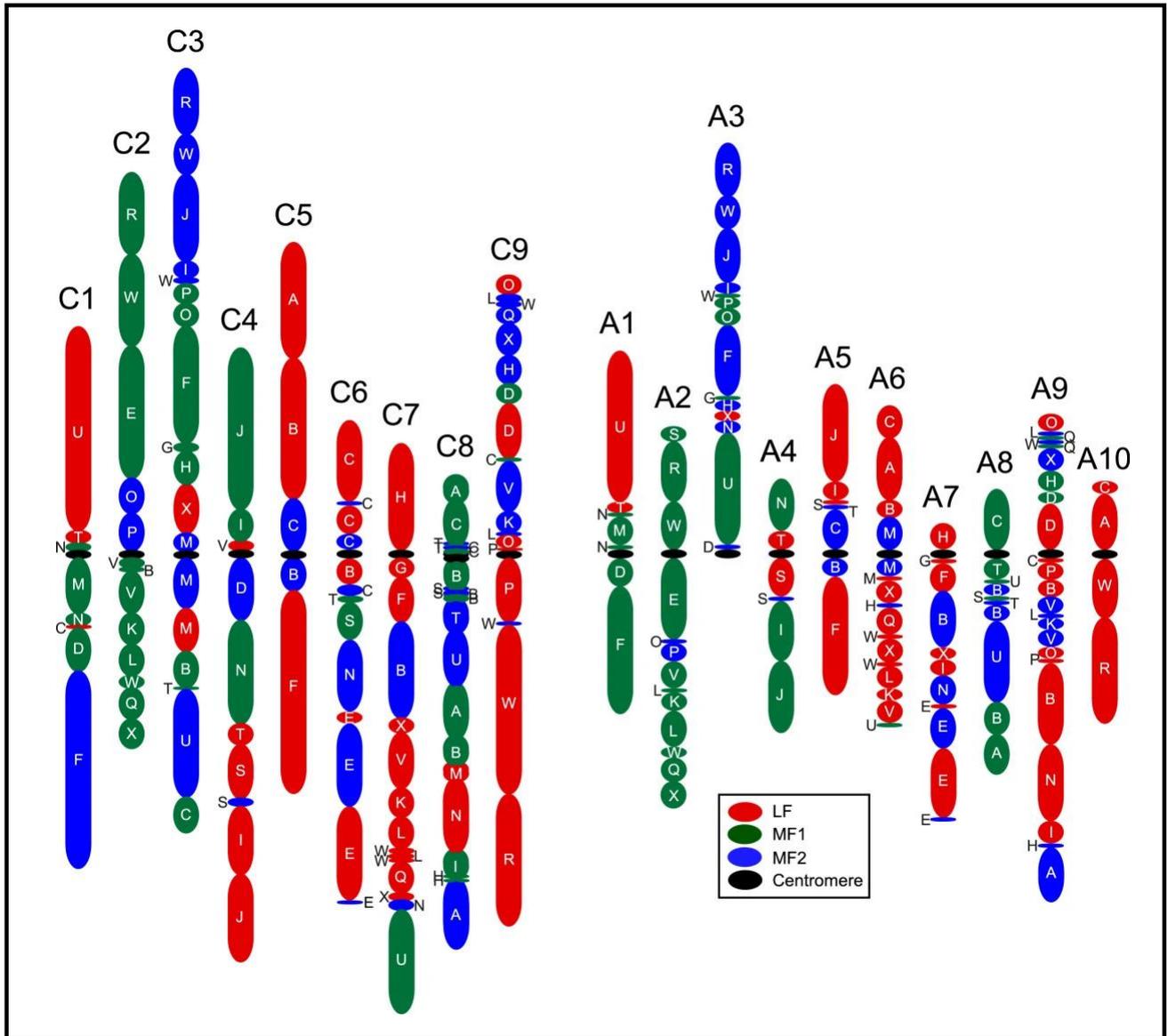
---

Le colza (*Brassica napus* L. subsp. *napus*) est une sous-espèce de l'espèce *Brassica napus* L. qui appartient au genre *Brassica* de la famille des Brassicaceae (anciennement connue sous le nom de Crucifères). Les Brassicaceae comportent environ 4000 espèces différentes, présentes essentiellement dans l'hémisphère nord à l'état sauvage (*Arabidopsis*, capselle par exemple) ou cultivées pour l'alimentation humaine ou animale sous forme de fourrage (colza, chou ou navet fourragers), de graines riches en huile et protéines (colza oléagineux, navette, caméline) ou de légumes (exemple du chou, radis, rutabaga, roquette), ou encore comme plantes ornementales (les giroflées et les alysses) (Friedt et al., 2018). Une des caractéristiques des Brassicaceae est la présence de glucosinolates dans les différents organes de la plante.

## I.1.1. Origine du colza et conséquences sur l'organisation du génome

Comme toutes les espèces végétales, les Brassicaceae ont connu plusieurs événements de duplications de leur génome au cours de l'évolution (Figure I.1). Une première triplication du génome ( $\gamma$ ) est probablement à l'origine de la séparation entre les monocotylédones et les dicotylédones. Cet événement a été suivi de deux duplications ( $\beta$  et  $\alpha$ ) spécifiques aux Brassicaceae (Jiao et al., 2012). Une triplication a ensuite conduit à la séparation des genres *Brassica* et *Arabidopsis* (Jiao et al., 2012 ; Chalhoub et al., 2014). Au sein du complexe d'espèces formé par le genre *Brassica*, des analyses cytologiques ont démontré l'existence d'événements d'hybridations interspécifiques récents (~7 500 ans) qui ont conduit à la formation d'espèces allotétraploïdes (Figure I.2 ; U, 1935). Ainsi l'espèce *Brassica napus* ( $A^nA^nC^nC^n$  ;  $2n=4x=38$ ) est un allotétraploïde résultant du croisement interspécifique entre la navette, *Brassica rapa* ( $A^rA^r$  ;  $2n=2x=20$ ) et le chou, *Brassica oleracea* ( $C^oC^o$  ;  $2n=2x=18$ ). 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).

Depuis les années 1990-2000, l'accumulation d'informations moléculaires (marqueurs moléculaires puis séquences de gènes) a démontré une forte conservation des séquences codantes (> 80% de gènes orthologues entre *Arabidopsis* et le colza ; Parkin et al., 2005) ainsi qu'une certaine conservation de la synténie (contenu et ordre des gènes sur les chromosomes) entre les génomes de Brassicaceae laissant supposer l'existence d'ancêtres communs. Plus récemment l'étude de génomes assemblés de plusieurs Brassicaceae actuelles (notamment *A. thaliana*, *A. lyrata*, *Capsella rubella*, *B.*



**Figure I.3 : Organisation des 24 blocs chromosomiques ancestraux au sein des génomes des espèces diploïdes, *B. oleracea* et *B. rapa* (d'après Parkin et al., 2014)**

Les génomes de *B. oleracea* (C1-C9) et de *B. rapa* (A1-A10) sont représentés respectivement à gauche et à droite sur le schéma. Les 24 blocs chromosomiques identifiés sur le caryotype ABK sont notés par les lettres (A-X). Leur colinéarité a été montrée entre les génomes d'*A. thaliana* et de Brassica. Les couleurs représentent les fractionnements différentiels : LF (Less Fractionated) en rouge, MF1 (Medium Fractionated) en vert et MF2 (Most Fractionated) en bleu. Les centromères sont indiqués par les cercles noirs.

*rapa*, *B. oleracea*, *B. napus*) a permis de proposer un modèle d'évolution des génomes des Brassicaceae. Ainsi, le caryotype de l'ancêtre commun aux Brassicaceae (ABK, Ancestral Brassicaceae Karyotype) comporterait 8 protochromosomes et aurait, après spéciation, évolué pour donner un ancêtre commun hexaploïde (PCK, Proto-Calepineae Karyotype ; n=7) à l'origine des espèces diploïdes du genre *Brassica*, et un ancêtre commun au genre *Arabidopsis* (ACK, Ancestral Camelinaeae Karyotype ; n=8) (Figure I.1 ; Lysak et al., 2006 ; Mandakova et Lysak, 2008 ; Murat et al., 2015). Ce modèle est supporté par l'identification de 24 blocs chromosomiques (notés A-X) au sein du caryotype ABK dont la colinéarité a été décrite entre les génomes d'*A. thaliana* et de *Brassica* (*B. rapa*, *B. oleracea*, *B. napus*) (Parkin et al., 2005 ; Schranz et al., 2006). En particulier, l'identification de 6 copies de chaque bloc chromosomique dans le génome de *B. napus* a conforté l'hypothèse d'un ancêtre hexaploïde des *Brassica* (Tang et al., 2012). Parallèlement, la comparaison des contenus en gènes entre les génomes des espèces diploïdes, *B. rapa* et *B. oleracea*, et celui d'*A. thaliana* suggère que les 3 sous-génomes de PCK ont subi des fractionnements différentiels, avec un sous-génome peu fractionné (LF, Less Fractionated genome) et qui présente donc le plus grand nombre d'orthologues communs avec *A. thaliana* et deux sous-génomes plus fractionnés (MF1, Medium Fractionated genome et MF2, Most Fractionated genome) (Figure I.3 ; Cheng et al., 2012 ; Tang et al., 2012 ; Liu et al., 2014 ; Parkin et al., 2014). Les gènes dupliqués conservés peuvent ensuite avoir 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 caractérisation du génome de *B. napus* a également montré l'existence de nombreux événements de translocation chromosomiques entre les génomes homéologues (Chalhoub et al., 2014) pouvant expliquer des variations de type 'nombre de copies' (CNV, Copy Number Variation) ou de type 'présence-absence' (PAV, Presence Absence Variation) comme cela a été largement observé entre accessions de colza (Schiessl et al., 2014). Ainsi, les événements récurrents de polyploïdisation ainsi que les réarrangements chromosomiques et le fractionnement génomique qui ont suivi, ont largement contribué à la complexité génomique actuelle (Fu et al., 2016).

### I.1.2. Ressources génomiques disponibles chez le colza et les autres Brassicaceae

#### *Génomes de référence assemblés*

La séquence du génome d'*Arabidopsis thaliana* a été relarguée en 2000 (écotype Columbia ; ~ 135 Mb, 27 655 gènes répartis sur 5 chromosomes ; The Arabidopsis Genome Initiative, 2000 ; Cheng et al., 2016). Dès 2005, l'arrivée des technologies de séquençage haut débit (séquençage Illumina),

**Tableau I.1 : Génomes assemblés de colza disponibles en Septembre 2020**

Accession <sup>a</sup>	Type <sup>b</sup>	Taille de l'assemblage en scaffolds (Mb)	Taille des pseudo-chromosomes (Mb)	Nombre de gènes annotés	Références
Darmor- <i>bzh</i> v5	WOSR	850	645	101 040	Chalhoub et al. (2014)
Darmor- <i>bzh</i> v8.1	WOSR	850	798	80 382	Bayer et al. (2017)
Darmor- <i>bzh</i> v10	WOSR	924	867	108 190	Rousseau-Gueutin et al. (2020)
Westar	SOSR	1007	942	97 514	Song et al. (2020)
No2127	SOSR	1011	932	95 385	
Zheyu7	SWOSR	1015	924	96 209	
Gangan	SWOSR	1033	934	96 843	
Shengli	SWOSR	1001	930	94 586	
Tapidor	WOSR	1014	932	96 117	
Quinta	WOSR	1003	933	95 492	
ZS11	SWOSR	1008	961	100 919	
Express 617	WOSR	925	765	89 857	Lee et al. (2020)

(a) Nom des accessions et numéro de version d'assemblage. Le génome Darmor-*bzh* v5 a été utilisé dans la thèse (surligné).

(b) Type de colza: WOSR, colza oléagineux d'hiver; SOSR, colza oléagineux de printemps; SWOSR, colza oléagineux semi-hiver.

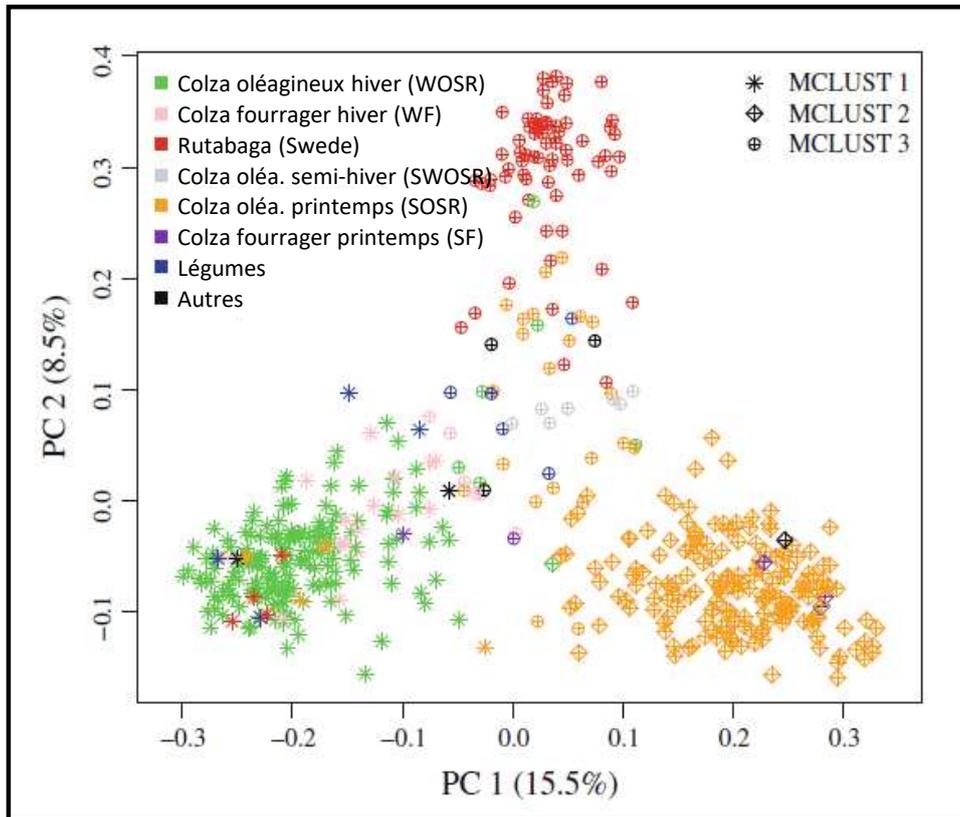
permettant de produire des millions de séquences rapidement et à faible coût, combinée aux développements des méthodes d'analyses bio-informatiques et de stockage des données générées, a permis le séquençage de nombreuses espèces aux génomes plus complexes. Ainsi, depuis une dizaine d'années, des génomes de plusieurs Brassica ont été séquencés, assemblés et mis à disposition dans les bases de données internationales, avec par ordre chronologique le premier génome de *B. rapa* en 2011 (Wang et al., 2011), suivi en 2014 par deux génomes de *B. oleracea* (Parkin et al., 2014) ainsi que le génome de la variété Darmor-*bzh* de colza (Chalhoub et al., 2014). Dans ce travail, réalisé par un consortium international mené par la France, l'analyse des 645 Mb de séquence met en évidence une très forte colinéarité entre les deux sous-génomes ( $A^n$  et  $C^n$ ) et les génomes des deux parents diploïdes ( $A^r$  et  $C^o$  respectivement ; Chalhoub et al., 2014). Des assemblages améliorés ont été relargués depuis (Bayer et al., 2017 ; Rousseau-Gueutin et al., 2020) (Tableau I.1). Cependant la version v5 du génome de Darmor-*bzh* (Chalhoub et al., 2014) fait encore office de référence car elle présente l'avantage d'être associée à un catalogue de gènes consolidés d'une part et à un grand nombre d'analyses génétiques et fonctionnelles d'autre part (identification de QTL et/ou gènes candidats pour de nombreux traits). C'est donc cette version qui a été choisie pour le travail de thèse. En parallèle, d'autres génotypes de colza appartenant à des groupes génétiques différents ont été séquencés, assemblés et annotés (Tableau I.1), permettant ainsi d'élargir les connaissances sur les variations alléliques et structurales (CNV, PAV) entre les ressources génétiques de colza. L'ensemble de ces génomes constituent une première étape vers l'obtention d'un pangénome (Lee et al., 2020 ; Song et al., 2020).

### *Marqueurs moléculaires de type SNP*

De nombreux types de marqueurs moléculaires ont été utilisés chez le colza au cours des 35 dernières années, allant des systèmes basés sur des hybridations aux marqueurs basés sur l'amplification par PCR puis le séquençage en multiplexage (pour revue, voir Obermeier et Friedt, 2015). Les marqueurs liés au polymorphisme d'un seul nucléotide entre individus d'une population (SNP, Single Nucleotide Polymorphism) sont les marqueurs moléculaires les plus utilisés en raison de leur présence en abondance le long du génome (Picoult-Newberg et al., 1999), de leur stabilité (faible taux de mutations), de leur reproductibilité et du fait qu'ils peuvent fournir une haute densité de marqueurs près ou dans un locus d'intérêt (Hayward et al., 2012). La plupart sont bi-alléliques même si des formes tri- ou tétra-alléliques existent. Avec le développement des technologies de séquençage nouvelle génération, la découverte de millions de SNP a été facilitée (Imelfort et al., 2009), ce qui a stimulé la sélection assistée par marqueurs chez le colza (Hayward et al., 2012). Les puces Illumina Infinium™ permettent le génotypage d'un grand nombre de sites polymorphiques (de 3000 à  $>10^6$ ) par échantillon, et plusieurs échantillons peuvent être testés simultanément (Chagné et



al., 2015) ; ces outils sont donc bien adaptés à l'étude de larges populations pour les analyses génétiques. A l'inverse, les puces SNP présentent un certain nombre de désavantages, notamment le fait que la diversité allélique testée est restreinte aux variants utilisés pour la production de l'outil (Ganal et al., 2012), excluant ainsi les allèles rares. Des approches par séquençage sont donc complémentaires (GBS, Genotyping-by-Sequencing) et peuvent être menées à l'échelle du génome complet (on parlera de reséquençage d'accessions) ou de portions de génomes (capture d'exomes ou de régions par exemple). Plusieurs puces SNP haute densité ont été développées chez le colza, initialement avec 6000 – 9000 SNP. La progression rapide de ces outils (Delourme et al., 2013 ; Edwards et al., 2013 ; Dalton-Morgan et al., 2014) a incité la communauté internationale à développer une puce commerciale *Brassica* 60K Illumina Infinium™, contenant un total de 52 157 SNP (Clarke et al., 2016) définis sur la base des programmes de séquençage de transcrits (EST) et des génomes de références disponibles. Cet outil a été utilisé pour un grand nombre d'études (caractérisation des ressources génétiques, analyses génétiques de divers traits : rendement et composantes du rendement, qualité de la graine, germination et vigueur, ramification, floraison, résistances aux maladies ...) et s'est imposée comme une puce de référence (pour revue voir Mason et al., 2017). Une puce réduite contenant 15K SNP a été déduite de la puce 60K permettant un génotypage à plus faible coût pour certaines applications, notamment pour la caractérisation de matériel de pre-breeding (Ganal et al., 2018). En complément des puces SNP, les approches basées sur le reséquençage d'accessions ont permis de caractériser du matériel génétique et de créer de nouvelles données/ressources de génotypage. Le nombre d'accessions reséquencées n'a cessé d'augmenter depuis 5-7 ans, ces informations sont partagées lors des réunions du groupe Multi Brassica Genome Project (MBGP) fédérant la communauté scientifique internationale qui utilise les ressources génétiques et génomiques de Brassica. Très récemment, Wu et collaborateurs (2019) ont publié le reséquençage de 991 accessions. Dans le cadre du projet RAPSODYN, 30 accessions de colza d'hiver oléagineux ont été reséquencées sur la base de génome complet et 206 accessions de colza d'hiver oléagineux sur la base de l'exome (Leveugle et al., 2015). Pour les analyses génétiques menées au cours de la thèse, nous avons mobilisé les données de génotypage issues d'une part du séquençage d'exome cité précédemment et d'autre part de la puce *Brassica* 60K SNP (cf. ChIII). Au total, les données de génotypage de 182 215 SNP répartis de façon homogène sur le génome étaient disponibles sur la population d'étude.



**Figure 1.4 : Organisation des ressources génétiques de *Brassica napus* (d'après Bus et al., 2011)**

Analyse en composantes principales d'une population de 509 accessions de *Brassica napus* appartenant à des germplasmes différents (cf légende). L'analyse montre une répartition en 3 groupes principaux (MCLUST).

### I.1.3. Les ressources génétiques de *Brassica napus*

#### *Diversité génétique*

La diversité génétique présente chez *B. napus* et maintenue dans les centres de ressources biologiques a été étudiée à partir de populations à bases génétiques relativement larges, incluant divers germplasmes dont des formes de colza oléagineux (hiver, printemps ou semi-hiver), des colzas fourragers (hiver ou printemps), des légumes (chou kale Russe, chou frisé sibérien, chou frisé rouge de Russie ...), des rutabagas et d'autres génotypes d'origines géographiques variées (Hasan et al., 2005 ; Bus et al., 2011). Ces études permettent d'identifier trois groupes principaux constitués majoritairement et respectivement des types hiver, des types printemps ou des rutabagas (Figure I.4). Au sein de chaque groupe, les colzas fourragers sont génétiquement distants des colzas oléagineux (Hasan et al., 2005 ; Bus et al., 2011).

Au sein des colzas oléagineux, on trouve principalement des formes de printemps et d'hiver qui se distinguent par leur besoin en vernalisation et leur aire de répartition géographique (Europe de l'Ouest pour le type hiver ; Europe de l'Est, Asie, Canada et Australie pour le type printemps) (Diers et Osborn, 1994 ; Becker et al., 1995, Bus et al., 2011). Le type semi-hiver est caractéristique des colzas d'origine asiatique qui dérivent d'introgressions de *B. rapa*. Il constitue un troisième type de colza oléagineux génétiquement distant des types hiver et printemps (Becker et al., 1995 ; Qian et al., 2006). A ce jour aucune forme sauvage de colza n'a pu être identifiée (Snowdon et al., 2007 ; Werner et Snowdon, 2018). La faible diversité présente chez les colzas oléagineux s'explique par son origine évolutive mais aussi par l'impact de la sélection réalisée à partir des années 1945 sur les traits de qualité de la graine (variétés '0+' à faibles teneurs en C22:1 et variétés '00' à faibles teneurs en C22:1 et en GSL – cf. Contexte général) (Becker et al., 1995). Cet effort de création variétale a été conduit à partir d'un nombre très réduit de géniteurs présentant les traits d'intérêt (Busch et al., 1994 ; Hasan et al., 2005), ce qui explique la faible diversité génétique actuelle présente chez les colzas oléagineux.

Dans la thèse, nous avons limité nos analyses à une population de 173 colzas oléagineux d'hiver (cf. ChII et ChIII). En effet, l'évaluation phénotypique de populations incluant des germplasmes plus diversifiés (ex. colzas de printemps, fourragers, rutabagas...) devient complexe dès lors qu'il s'agit de traits intégrés et complexes tels que le rendement, qui se mettent en place tout au long du cycle et dépendent fortement de la phénologie de la plante.

#### *Voies possibles pour élargir la diversité génétique du colza oléagineux d'hiver*

Plusieurs voies d'augmentation de la diversité génétique peuvent être envisagées. Une première voie d'amélioration consiste à introgresser de la variabilité à partir des colzas oléagineux de type



printemps. Le programme « Printhiver », mené par le GIE Colza, a développé des populations à partir de croisements 'printemps × hiver' et a pu montrer qu'elles étaient génétiquement distantes des types printemps et hiver utilisés comme parents. Des hybrides développés à partir de ces populations montrent des rendements et des qualités d'huile équivalents, voire supérieurs, aux variétés actuelles (Gourrion et al., 2020).

Une seconde voie repose sur l'utilisation de colzas synthétiques obtenus à partir du croisement entre les deux parents diploïdes (*B. rapa* et *B. oleracea*). La descendance obtenue présente généralement une faible fertilité ainsi que des caractères indésirables (Kräling, 1987), mais après plusieurs rétrocroisements, les individus obtenus sont génétiquement distants des variétés classiques et leur fitness est amélioré. Ainsi, l'analyse de 20 accessions de colzas synthétiques a montré que plusieurs événements de délétions, duplications ou d'échanges de chromosomes homéologues sous-tendaient la création de nouvelles ressources génétiques (Samans et al., 2017). Il est également possible d'obtenir des colzas semi-synthétiques par croisement entre un colza et l'un des deux parents diploïdes (Mason et Chèvre, 2016). La voie des hybrides triploïdes augmente le taux de recombinaison et permet d'introduire des fragments génomiques plus petits et sans doute dans des régions peu recombinantes (Leflon et al., 2006 ; Pelé et al., 2017), ouvrant ainsi de nouvelles perspectives pour élargir la diversité grâce à l'introgession rapide de traits d'intérêt à partir des parents diploïdes.

Enfin il est possible d'induire de la variabilité chez le colza en utilisant des techniques de mutagenèse aléatoire ou ciblée. Des populations de variants portant des mutations ponctuelles ont ainsi été générées par mutagenèse chimique à l'éthyl-méthyl-sulfonate chez *B. napus* (Wang et al., 2008 ; Harloff et al., 2012). Ces populations ont permis la découverte de nouveaux variants alléliques par restriction/séquençage dans des gènes d'intérêt comme cela a été fait pour les gènes *FAE1* impliqués dans l'élongation des acides gras (Wang et al., 2008). Récemment, des travaux ont montré la faisabilité d'éditer les différentes copies liées à une fonction chez le colza par l'utilisation du système CRISPR/Cas9 (Braatz et al., 2017 ; Huang et al., 2020 ; Sashidhar et al., 2020).

#### I.1.4. 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  $r^2=0.2$ ). La première étude chez le colza a évalué l'étendue du DL entre 0.5 et 2 cM (Ecke et al., 2010).

## Encadré I.1 : Principes des analyses QTL (repris 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'héritabilité 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.

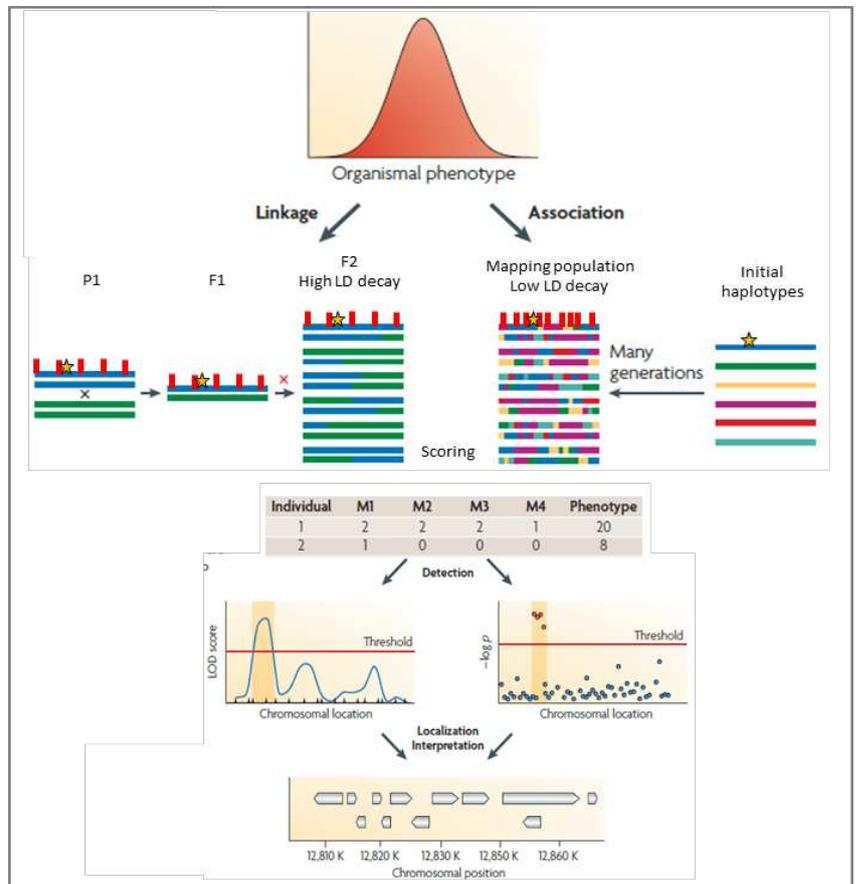
Deux méthodes sont principalement utilisées pour identifier les loci:

**L'analyse de liaison (LA, linkage analysis)** se base sur la ségrégation des allèles dans une descendance entre deux parents (P1, P2), souvent choisis sur la base d'un phénotype contrasté. Les principales populations utilisées sont les haploïdes doublés (HD), les lignées recombinantes (RIL) et les populations de rétrocroisements (BC, Back-Cross). Les fréquences alléliques sont contrôlées permettant de considérer des allèles rares potentiels qui seraient indétectables dans une population naturelle. Cependant, la diversité génétique est réduite par le choix des parents, ce qui limite les résultats à la détection des loci polymorphes entre les deux parents. Par ailleurs, les recombinaisons sont souvent peu nombreuses car issues d'un faible nombre de méioses, conduisant à une mauvaise précision des QTL.

**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 (si tant est que le choix du panel est réfléchi), 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.

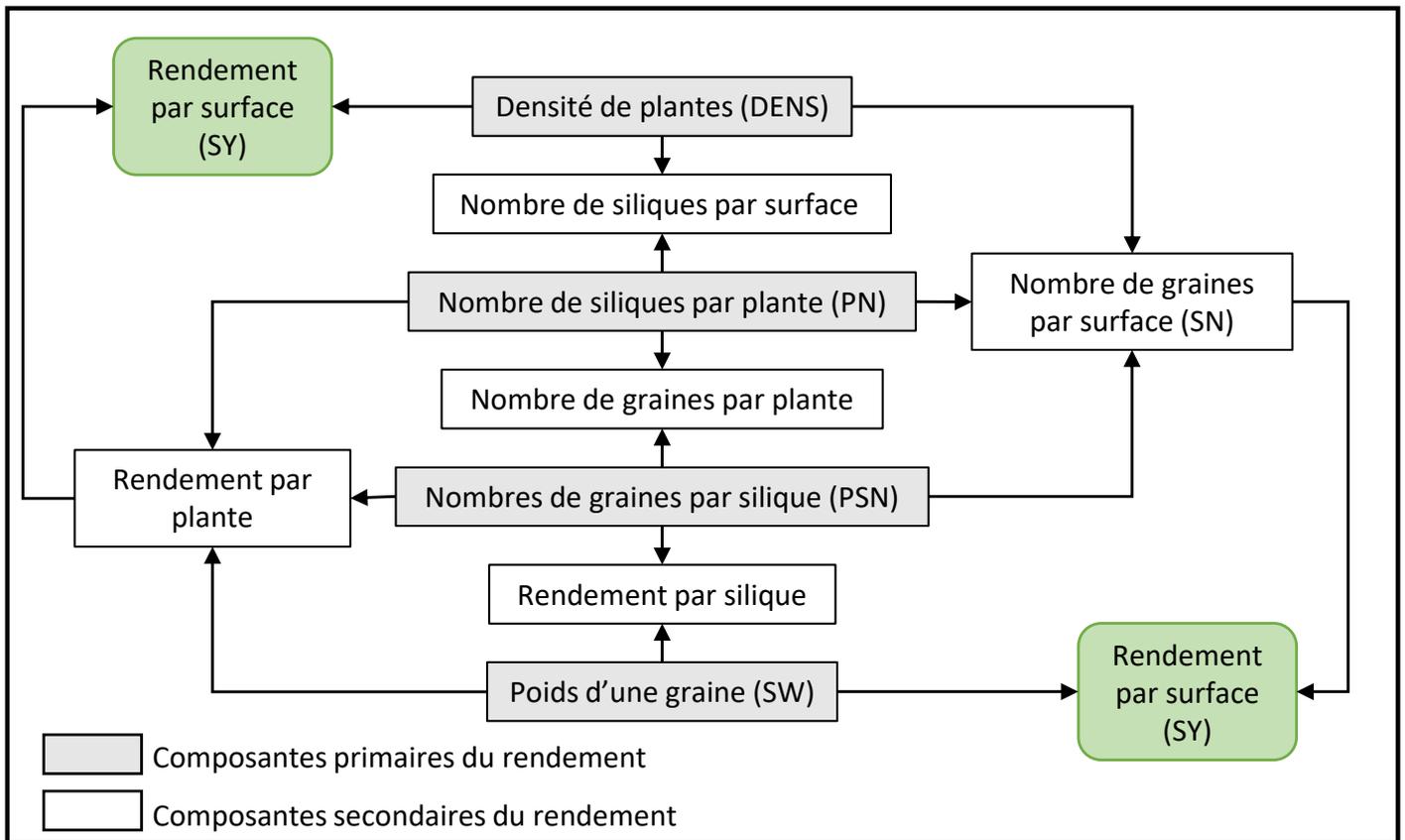
### Etapes pour la détection de QTL par analyse de liaison ou d'association

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)



Avec le séquençage du génome, ces variations sont maintenant exprimées en distance physique et des travaux rapportent une étendue du DL de 0.25-2.5 Mb dans une population de 203 accessions de colzas semi-hiver (Qian et al., 2014) et de 0.48-1.28 Mb dans une population de 248 accessions de colzas d'hiver (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 (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 (et inversement). S'agissant de la population de 173 accessions utilisée dans la thèse, l'étendue du DL à l'échelle du génome est de 174 kb (cf. ChIII, [Supplementary Table III.2](#)), la matrice de génotypage de 182K SNP utilisée (cf. I.1.2) est donc suffisamment dense pour espérer une bonne résolution des analyses génétiques.



**Figure I.5 : Les composantes du rendement en grains chez le colza**

Relations entre les différentes composantes (primaires et secondaires) pour aboutir au rendement grainier par unité de surface. Les abréviations sont données en anglais: DENS, plant density; PN, pod number; PSN, pod seed number; SN, seed number; SW, seed weight; SY, seed yield. Adapté d'après Diepenbrock (2000).

## I.2. Cycle du colza d’hiver et élaboration du rendement en grains

---

### I.2.1. Mise en place des composantes du rendement chez le colza

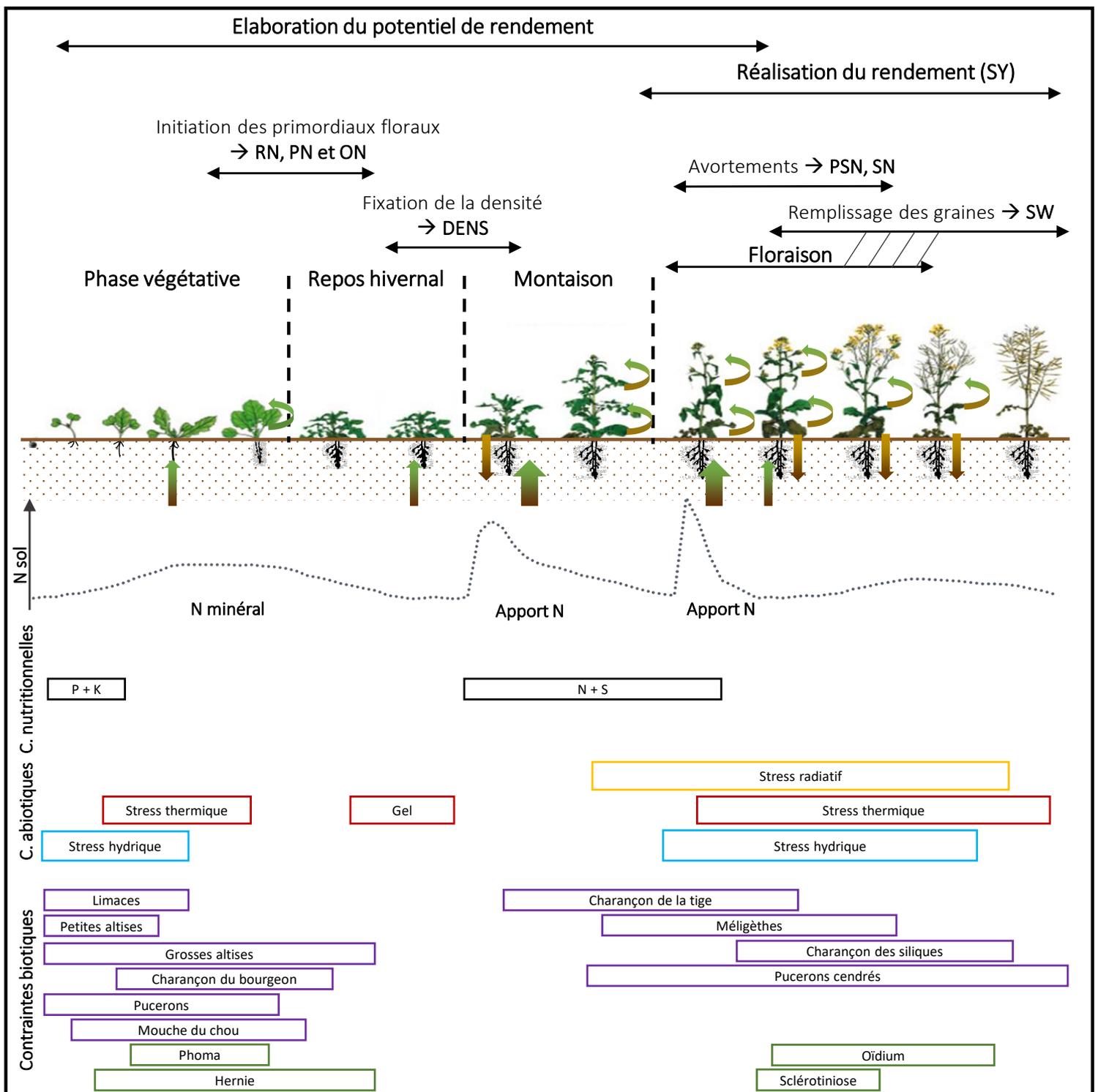
#### *Définition du rendement en graines*

Le rendement en grains (SY, seed yield) est le poids de graines (en tonnes, t) récoltées par unité de surface (en hectares, ha) en considérant un taux d’humidité et d’impuretés à 0%. Le rendement correspond au produit du nombre de graines par unité de surface (SN, seed number) par le poids d’une graine (SW, seed weight). SW peut être estimé via le poids de mille graines (TSW, thousand seed weight) tandis que SN correspond au produit de la densité en plantes (DENS, plant density), du nombre de siliques par plantes (PN, pod number) et du nombre de graines par silique (PSN, pod seed number) (Figure I.5 ; Diepenbrock, 2000). Le rendement est donc un trait complexe qui résulte de plusieurs composantes qui se mettent en place séquentiellement au cours du cycle de culture (Figure I.6).

#### *Le cycle du colza*

La culture du colza d’hiver débute par un semis entre fin août et début septembre et s’achève à la récolte entre fin juin et début juillet. Ce cycle long peut, en première approximation, être découpé selon deux grandes phases, d’une part la phase végétative et d’autre part la phase reproductrice qui correspondent respectivement à l’élaboration et à l’expression du potentiel de rendement (Figure I.6 ; Bouchet et al., 2016b).

La phase végétative commence par la phase de germination-levée qui dure environ 100 °C.J (degrés jours calculés en base 0°C depuis le semis). Le colza rentre ensuite dans une phase de croissance lente jusqu’au stade 4-5 feuilles (~ 400°C.J) marquée par une production de biomasse aérienne faible et un développement racinaire important afin de capter efficacement l’azote minéral du sol pour constituer la biomasse aérienne (Rossato et al., 2001). C’est pendant cette période entre la levée et la mise en place des premières feuilles que la culture est très sensible aux dégâts d’insectes. Puis suit la phase de croissance automnale pendant laquelle la plante a une forte dynamique de mise en place de la biomasse aérienne (croissance et développement de la rosette). Arrivé au repos végétatif (entrée d’hiver), le colza aura ainsi fabriqué 20% de sa biomasse aérienne et 70% de sa biomasse racinaire. Parallèlement, un processus de remobilisation de l’azote qualifiée de séquentielle se met en place dès la fin de l’automne et permet le recyclage de l’azote des feuilles sénescentes vers les



**Figure I.6 : Cycle culturel 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., 2016b et documents Terres Inovia)**

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. Ce stock est ensuite consommé par le peuplement végétal, avec des pics d’absorption principalement à l’automne, à la reprise de végétation en sortie de l’hiver et avant la floraison. 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. La mise en place des composantes du rendement est indiquée sur le graphe, avec les abréviations données en anglais: DENS, plant density; ON, ovule number; PN, pod number; PSN, pod seed number; RN, ramification number; SN, seed number; SW, seed weight. 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).

jeunes feuilles, soutenant ainsi l'initiation des premiers primordiaux floraux. Cette initiation très précoce détermine en partie le nombre de ramifications (RN, Ramification Number), de fleurs et de siliques (PN), et par conséquent le nombre d'ovules (ON) avant la fin de l'hiver (Figure 1.6 ; Bouchet et al., 2016b). Le potentiel de ramification est compris entre 20 et 25 ramifications par plante, mais toutes ne seront pas fertiles à cause de la compétition intra- et inter-plantes (Diepenbrock, 2000). Ensuite, durant l'hiver climatique, le peuplement subit une phase de vernalisation. Au cours de cette période, on note une accélération de la chute des feuilles, ce qui représente une perte d'azote pour le système. Cependant l'azote présent dans les feuilles chutées à l'automne et à l'hiver peut être recyclé et être de nouveau disponible plus tard pour la culture même si des travaux estiment que seul 40% du contenu en azote des feuilles chutées sera recyclé (Dejoux et al., 2000). Parallèlement, les plantes les moins vigoureuses sont sensibles aux dégâts de froid et de gel. C'est donc à la fin de l'hiver que la densité du peuplement (DENS) est fixée. La reprise de végétation a lieu à la sortie de l'hiver où une nouvelle production de biomasse se met en place rapidement via l'émergence et la croissance de nouvelles feuilles puis l'élongation de la tige. Pendant la phase de montaison, la sénescence séquentielle se poursuit et assure la remobilisation de l'azote vers les jeunes organes. Puis la floraison marque l'arrêt de la production de feuilles et le début de la phase reproductive, on observe alors majoritairement un phénomène de sénescence monocarpique qui va sous-tendre la remobilisation de l'azote vers les organes reproducteurs (siliques) (Diepenbrock, 2000 ; Bouchet et al., 2016b). Parallèlement, l'activité photosynthétique est assurée par l'enveloppe des siliques afin de couvrir les besoins en carbone de la plante (Leterme, 1985 ; Müller et Diepenbrock, 2006 ; Jullien et al., 2011). Les itinéraires techniques conventionnels préconisent des apports minéraux à cette période afin de supporter la production de biomasse. A ce stade, les bonnes conditions de début de cycle ont permis d'atteindre un potentiel de rendement élevé, qui va ensuite s'exprimer à partir de la floraison par la mise en place d'un nombre de grains par m<sup>2</sup> (SN) en lien avec le nombre d'avortements d'une part et par le remplissage de ces grains (SW) d'autre part. Enfin, chez le colza, la mise en place séquentielle des ramifications puis des siliques sur les ramifications explique la forte capacité de compensation observée chez cette culture (Diepenbrock, 2000).

### 1.2.2. Les facteurs limitant le rendement chez le colza d'hiver

Les facteurs qui impactent l'élaboration du rendement sont de nature variée et peuvent intervenir à différentes étapes du cycle de culture, seuls mais le plus souvent en combinaisons (Leach et al., 1994). On distingue trois catégories de contraintes qui seront détaillées ci-dessous : les stress biotiques, les contraintes nutritionnelles et les stress abiotiques (Figures 1.6-1.7). En plus de ces contraintes, il faut également considérer les signaux perçus par la plante et qui résultent des



**Figure 1.7 : 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 efficacité 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).

interactions complexes que la plante entretient avec son environnement biotique. Ces interactions de type ‘plante-plante’ et/ou ‘plante-microbiote’ peuvent moduler l’effet des contraintes multiples (en les amplifiant ou à l’inverse en les atténuant).

### *Les stress biotiques*

Le rapport Ecophyto (Potier, 2014) démontre que le colza est, parmi les grandes cultures (hors pomme de terre), l’espèce qui présente les indices de fréquence de traitement (IFT) les plus forts, traduisant un recours élevé à l’utilisation de produits phytosanitaires pour lutter contre les nombreux bioagresseurs auxquels il est sensible. Ainsi, l’IFT moyen du colza oscillait entre 4 (Bretagne) et 7.6 (Nouvelle Aquitaine) en 2017 (Agreste 2019). Au sein des traitements, ce sont les insecticides qui constituent la part majoritaire (50-60%), loin devant les fongicides (12-17%), les herbicides (22-26%) et les molluscicides (6-7%).

En effet, les ravageurs s’attaquent au colza dès la levée. A titre d’exemple, les limaces grises (*Deroceras reticulatum*) ou noires (*Arion hortensis*) dévorent les jeunes plantes, l’une par les racines, l’autre par la tige. La phase automnale est aussi le moment privilégié pour les attaques des insectes ravageurs (Figures 1.6-1.7). En particulier, la petite altise (*Phyllotreta nemorum*) et la grosse altise (*Psylliodes chrysocephala*) se nourrissent des cultures naissantes, visant surtout cotylédons et hypocotyles. Les pertes peuvent aller jusqu’à plusieurs quintaux par hectare. A la reprise de végétation, l’augmentation des températures favorise l’activité des larves d’altises ou de charançons sur bourgeon terminal qui ravagent le cœur des plantes en creusant dans les tiges et pétioles, provoquant des retards de croissance, voire la mort de la plante. Une attaque sévère peut entraîner des pertes jusqu’à 20 q/ha. Juste avant la floraison, des attaques de méligèthes (*Meligethes aeneus*) sont à craindre car ces insectes disposent de fortes connexions avec le colza, notamment dans leur processus de reproduction. Leurs larves installées dans les boutons floraux vont provoquer des avortements de graines et ainsi engendrer des dégâts considérables pour la récolte (Herbinger, 2013). En fin de cycle, des attaques de pucerons et de charançons des siliques peuvent également avoir lieu. A ce jour, il n’existe pas de variétés résistantes aux insectes. Hormis les pesticides, les principaux moyens de lutte visent à utiliser des plantes pièges (stratégie de ‘push pull insect control’) ou à favoriser des stratégies d’échappement de la culture. Pour cela, il est préconisé d’adapter les pratiques culturales (allongement des rotations, avancement de la date de semis) pour, d’une part gérer la présence des ravageurs, et d’autre part obtenir des cultures vigoureuses présentant une meilleure tolérance aux insectes d’automne (Herbinger, 2013).

S’agissant des agents phytopathogènes, le colza est également la cible d’un certain nombre de champignons ou bactéries provoquant des maladies dont l’occurrence et la sévérité dépendent de



conditions environnementales favorables pour le développement des pathogènes (humidité, température, ombrage, résidus de culture ...). En particulier, le phoma causé par l'agent pathogène *Leptosphaeria maculans*, la hernie des crucifères (*Plasmodiophora brassicae*), la sclérotiniose (*Sclerotinia sclerotiorum*) et l'oïdium (*Erysiphe cruciferarum*) sont les maladies les plus fréquentes dans nos conditions pédoclimatiques (Figures 1.6-1.7). Elles entraînent des pertes de rendement notables de l'ordre de 18 q/ha pour le phoma, 4-5 q/ha pour la hernie, 19 q/ha pour la sclérotiniose et 10 q/ha pour l'oïdium. La présence des agents pathogènes peut être contrôlée par les rotations culturales ou une gestion optimale des résidus de culture et des repousses afin de limiter les sources d'inoculum (sclérotiniose, phoma, hernie). Par ailleurs, des variétés de colza portant des résistances spécifiques impliquant des gènes majeurs ont été développées et largement utilisées pour lutter contre le phoma (pour revue voir Delourme et al., 2006) ou la hernie (Rahman et al., 2014; Zhang et al., 2016). Le succès de ces variétés est à mettre en lien avec la relative facilité pour manipuler ce type de résistance en sélection. Cependant, bien que ces résistances spécifiques soient localement et transitoirement très efficaces (McDonald et Linde, 2002), elles sont peu durables, en particulier dans le cas de l'utilisation sur une large portion du territoire de variétés porteuses du même gène de résistance (ex. contournement d'un gène de résistance spécifique au phoma du colza ; Rouxel et al., 2003). Des programmes de recherche complémentaires visent donc à identifier et exploiter les résistances quantitatives, aux effets partiels et basées sur un support génétique polygénique dans la majorité des cas. Elles sont gouvernées par plusieurs gènes dont les effets cumulatifs vont limiter l'agressivité et le développement du pathogène, les rendant ainsi plus difficiles à contourner (Manzanares-Dauleux et al., 2000 ; Delourme et al., 2006 ; Jestin et al., 2011).

Enfin, le colza doit faire face à plusieurs adventices qui sont susceptibles de le concurrencer tant que le couvert n'est pas fermé. Les adventices fréquemment recensées dans les champs de colza sont le vulpin, le coquelicot, le gaillet, le ray-grass, la ravenelle, la capselle ou encore le géranium disséqué (Herbinger, 2013). Le principal moyen de lutte contre les adventices du colza repose sur l'adaptation des itinéraires techniques via la durée des rotations, le travail du sol (faux-semis en amont) ou la date de semis pour assurer une implantation rapide et homogène du colza à l'automne.

### *Les carences nutritionnelles*

Le colza est une culture gourmande en azote (N), soufre (S), potassium (K) et phosphore (P). Les apports de P et K sont réalisés le plus souvent à l'automne après le stade 4-5 feuilles afin d'optimiser la croissance rapide du colza. Des carences en phosphore sont à l'origine de pertes de rendement de l'ordre de 15 q/ha (Herbinger, 2013). Le soufre est également un élément important pour le colza, les pertes de rendement pouvant excéder 20q/ha en cas de carence. Il est majoritairement absorbé au



cours de la montaison, période durant laquelle un apport en S est préconisé (75 kg/ha de SO<sub>3</sub>). Enfin l'azote est le nutriment clé du colza qui conditionne directement le rendement final et doit être apporté à la culture en quantité importante (180 à 250 kg/ha) (Rathke et al., 2006). Malgré une bonne capacité à absorber l'azote minéral, le colza se caractérise par une faible efficacité d'usage de l'azote (NUE, nitrogen use efficiency) (Avice et Etienne, 2014) comparée à d'autres espèces de grande culture. A titre d'exemple, la culture de colza ne produira que 15 kg de graines par unité d'azote contre 35 à 40 kg pour les céréales comme le blé ou l'orge. Sur le plan agronomique, la NUE est définie comme le rendement en grains par unité d'azote disponible pour la plante (Moll et al., 1982). Cette faible NUE peut s'expliquer par une perte non négligeable de l'azote encore contenu dans les feuilles sénescentes et chutées. Ainsi le ratio entre la quantité d'azote dans la plante et la quantité d'azote apporté ne dépasse pas 50-60% (Malagoli et al., 2005). La NUE est donc un processus à améliorer chez le colza et cela revient à optimiser le rendement en fonction de la fertilisation azotée (*i.e.* augmenter le rendement en grains en conservant le niveau de fertilisation ou maintenir le rendement en grains en diminuant les intrants azotés). L'identification, la caractérisation et l'exploitation de la diversité génétique pour l'élaboration du rendement dans des conditions limitantes en azote est un levier pour améliorer la NUE. Ainsi plusieurs travaux rapportent l'identification de QTL de NUE chez le colza (Gül et al., 2003 ; Miro, 2010 ; Bouchet et al., 2014 ; 2016a ; Miersch et al., 2016 ; Wang et al., 2017). Sur le plan fonctionnel, la NUE est le produit de deux composantes : l'efficacité d'absorption de l'azote (NUpE, N uptake efficiency) et l'efficacité d'utilisation de l'azote (NUtE, N utilization efficiency). Cette dernière comprend l'efficacité d'assimilation de l'azote (NAE, Nitrogen Assimilation Efficiency) et l'efficacité de remobilisation de l'azote (NRE, N remobilization efficiency) (pour revue, voir Avice et Etienne, 2014).

### *Les stress abiotiques*

Les stress abiotiques comprennent les stress radiatifs (rayonnement journalier), thermiques (froid, gel, fortes températures) et hydriques (excès ou manque d'eau). Baux et collaborateurs (2015) rapportent que les stress climatiques impactent principalement la floraison et la phase de remplissage de la graine (de l'anthèse à la maturité) chez le colza. En particulier, le coefficient photothermique (QPT), qui se définit comme le quotient entre le rayonnement journalier et la somme des températures moyennes au cours de la floraison, est très corrélé au nombre de graines (SN) et donc au rendement (SY) chez le colza (Bagot et al., 2014). Idéalement, un fort ensoleillement avec des températures fraîches pendant la période de reproduction favorisera un QPT élevé et donc un bon rendement. Ainsi, il a été montré que des faibles radiations, entraînant un QPT faible, constituent un facteur limitant principal du rendement à la floraison et durant le remplissage des graines (Baux et al., 2015). Par ailleurs, le manque de rayonnement récurrent depuis l'implantation



dans certaines régions de production (ex. au nord de la Loire en France) provoque parfois l'avortement des hampes principales. S'agissant de la température, des moyennes faibles à l'automne auront pour conséquence d'allonger la phase de croissance jusqu'au stade 4-5 feuilles et indirectement de rendre le colza plus sensible aux attaques d'insectes ravageurs. Durant l'hiver, des températures inférieures à 5°C sont requises pour une vernalisation correcte (Ferreira et al., 1995 ; Habekotté, 1997 ; Chandler et al., 2005) ; néanmoins le colza est sensible au gel (Bouchet et al., 2016b). Enfin, de fortes températures à la floraison entraînent une baisse sensible des rendements en affectant la fertilité et le remplissage de la graine (Nuttal et al., 1992 ; Morrison, 1993 ; Angadi et al., 2000 ; Young et al., 2004). Les excès d'eau (engorgements de parcelles) impactent le développement des plantes de colza de la levée à la montaison (Zhou et Lin, 1995). Par exemple, cela conduit à une diminution de la taille de la plante et du nombre de ramifications à l'origine d'une baisse de rendement entre 12.5 et 21%. Au niveau physiologique, l'excès d'eau impacte le contenu en chlorophylles et l'activité photosynthétique foliaire, et engendre de nombreuses modifications métaboliques (activités enzymatiques, production de phytohormones) qui perdurent y compris lors de conditions plus favorables (Zhou et Lin, 1995). Par ailleurs, les stress hydriques par manque d'eau ont également un effet drastique sur la croissance et le développement du colza. En effet, ils réduisent la disponibilité de tous les nutriments qui dépendent de l'eau pour leur mobilité dans le sol et bloquent leur absorption par les racines, ce qui peut rapidement entraîner des carences nutritionnelles importantes. La présence d'eau dans le sol est bien sûr essentielle à la germination, mais c'est au cours de la phase reproductrice que le manque d'eau est le plus souvent observé et limitant sous nos climats. Champolivier et Merrien (1996) ont pu montrer qu'un stress hydrique, représentant 37% des besoins en eau, appliqué de la floraison jusqu'à la détermination du nombre de grains provoque une diminution du rendement et de ces composantes de l'ordre de 48%. La sécheresse a également un impact sur la qualité de la graine, en modifiant le rapport huile/protéines (Bouchereau et al., 1996 ; Champolivier et Merrien, 1996 ; Elferjani et Soolanayakanahally, 2018).

Dans la thèse, nous avons utilisé les données issues d'un réseau d'essais en plein champ mis en place dans le projet RAPSODYN et suffisamment étendu sur le territoire français pour représenter la diversité des zones de culture du colza. Dans ce réseau, les essais ont été protégés contre les stress biotiques dans la mesure du possible (gestion des itinéraires et applications de produits phytosanitaires lorsque cela était nécessaire), et supplémentés en éléments minéraux nécessaires pour pallier les carences nutritionnelles en potassium, phosphore et soufre. Les facteurs de variation restants étaient donc les contraintes abiotiques multiples et la contrainte azotée qui a été pilotée à l'échelle de chaque essai avec un protocole commun.

## Encadré I.2 : Modélisation des interactions G×E

Classiquement on explique le phénotype comme le résultat de l'expression du génotype (G) et de l'expression d'éléments externes imputables à l'environnement (E), ce qui se traduit par le modèle linéaire :

$$Y_{ij} = \mu + G_i + E_j + \varepsilon_{ij} \quad (i)$$

où  $\mu$  est la moyenne générale de la population,  $G_i$  est l'effet du i-ème génotype et  $E_j$  correspond à l'effet du j-ème environnement et  $\varepsilon_{ij}$  est la résiduelle du modèle linéaire.

### Modèle linéaire pour l'analyse de variance à 2 facteurs avec interaction

On peut appréhender les interactions comme une déviation au modèle additif (i) comme suit :

$$Y_{ij} = \mu + G_i + E_j + G_i \times E_j + \varepsilon_{ij} \quad (ii)$$

où  $G_i \times E_j$  est le terme interactif entre le génotype  $G_i$  et l'environnement  $E_j$ .

### Ecovalence variétale

Elle est définie comme la contribution du génotype à la somme totale des carrés des écarts de l'interaction :

$$W_g^2 = \sum_{e=1}^E (Y_{ge} - Y_g - Y_e + Y_{..})^2 \quad (iii)$$

avec  $W_g^2$  l'écovalence variétale du génotype  $g$ ,  $Y_{ge}$  la valeur du génotype  $g$  dans l'environnement  $e$ ,  $Y_g$  l'effet moyen du génotype  $g$ ,  $Y_e$  l'effet moyen de l'environnement  $e$  et  $Y_{..}$  représente la moyenne générale de l'ensemble des génotypes et des environnements.

### Modèle AMMI (Additive Main Effects and Multiplicative Interaction)

Le modèle multiplicatif s'écrit de la façon suivante :

$$Y_{ij} = \mu + G_i + E_j + \sum_{k=1}^K \lambda_k \alpha_{ik} \beta_{jk} + \varepsilon_{ij} \quad (iv)$$

avec  $\lambda_k$  la SVD du terme  $k$ ,  $\alpha_{ik}$  le score associé au génotype  $i$  pour le terme  $k$  et  $\beta_{jk}$  le score associé à l'environnement  $j$  pour le terme  $k$ .

### Régression linéaire conjointe

Le terme interactif est partitionné en un premier terme qui décrit l'hétérogénéité du coefficient de régression linéaire ( $b_i$ ) et un second qui représente la déviation :

$$Y_{ij} = \mu + G_i + E_j + (b_i E_j + d_{ij}) + \varepsilon_{ij} \quad (v)$$

où  $b_i$  est le coefficient de régression qui mesure la réponse d'un génotype à des environnements variés et  $d_{ij}$  est la déviation par rapport à la régression du génotype  $i$  dans l'environnement  $j$  (Finlay and Wilkinson, 1963).

### Régression factorielle

Dans le cas où l'on ne considère que les covariables environnementales, le modèle est :

$$Y_{ij} = \mu + G_i + E_j + \sum_{k \in EC} \theta_{ki} z_{kj} + \varepsilon_{ij} \quad (vi)$$

où  $\theta_{ki}$  est le coefficient de régression spécifique au génotype  $G_i$  (sensibilité de la variété au facteur environnemental) et  $z_{kj}$  est la covariable environnementale de l'environnement  $E_j$  (van Eeuwijk et al., 1996 ; Brancourt-Hulmel et al., 2000).

## 1.3. Interactions G×E et QTL×E

---

### 1.3.1. Définition des interactions G×E et utilisation dans un contexte de sélection variétale

Comme pour toutes les cultures, les performances de rendement du colza sont limitées par l'action de contraintes variées et multiples (*cf.* 1.2), en particulier les stress abiotiques dans le cas d'itinéraires conventionnels. Face aux profondes modifications des conditions agricoles (incidence des événements météorologiques extrêmes, réduction des intrants, évolution des systèmes de culture et des itinéraires techniques vers des pratiques agroécologiques), ces contraintes vont probablement s'accroître et surtout se complexifier. La variation des rendements, d'une année à l'autre et/ou d'un lieu à l'autre, a pour origine la sensibilité du matériel végétal à l'effet des contraintes multiples et combinées auxquelles la culture est soumise. La variation des performances génotypiques engendre l'interaction génotype × environnement (G×E), rendant difficile la sélection et la recommandation de génotypes performants. Les interactions G×E représentent, plus précisément, une déviation au modèle additif, où le phénotype est expliqué par un effet principal du génotype et un effet principal de l'environnement (*Encadré 1.2*).

Les interactions G×E sont un enjeu essentiel pour l'amélioration variétale, ceci étant d'autant plus important que l'héritabilité du caractère étudié est moyenne à faible. En présence d'interactions de nature qualitative (*i.e.*, qui modifient le classement des génotypes entre environnements ; Baker, 1988), les sélectionneurs sont à la recherche de méthodes pour exploiter l'interaction G×E dans le processus de sélection. Le recours à l'étude de la stabilité des performances est un moyen qui permet de classer les génotypes selon leurs degrés de stabilité et de performance. Le concept de stabilité revêt différentes acceptations selon les applications (Lin et al., 1986). Becker et Leon (1988) proposent le concept de stabilité statique dans lequel un génotype stable est celui qui présente des performances constantes indépendamment des conditions environnementales. Dans ce cas, le terme statistique d'interaction G×E est significatif puisque la performance observée est différente de celle attendue en sommant les effets additifs du génotype et de l'environnement. À l'inverse, le concept de stabilité dynamique (ou stabilité agronomique) considère un génotype stable lorsque sa réponse sous différents environnements ne dévie pas de la réponse prédite en sommant les effets additifs (Fasahat et al., 2015), le terme interactif G×E est alors non significatif.

Par ailleurs, la sélection pour l'adaptation spécifique (ou locale) est un autre moyen d'exploiter l'interaction G×E (Annicchiarico et al., 2005). En effet, s'il peut paraître séduisant de minimiser



l'interaction G×E en sélectionnant des variétés avec une bonne performance dans la majorité des environnements, cette stratégie est très risquée et ne s'est révélée que très rarement efficace (Ceccarelli, 1989). A l'inverse, l'adaptation spécifique est largement plébiscitée pour sécuriser la production alimentaire locale. De plus, elle n'est pas forcément dirigée uniquement pour répondre à des environnements favorables *versus* défavorables, mais adresse également l'exploitation de potentialités variées entre différents environnements relativement homogènes (ex. exploitation de différents optima de floraison ; Wallace et al., 1993). Cela implique d'exploiter les effets d'interaction G×Lieu positifs qui sont répétables dans le temps tout en minimisant les autres effets d'interaction G×E via la sélection pour une meilleure stabilité du trait étudié (Annicchiarico, 2002). Enfin, il apparaît nécessaire de prendre en compte la réponse des variétés aux stress environnementaux locaux pour mener à bien la sélection spécifique.

L'information issue de l'expérimentation variétale est potentiellement riche. De très nombreux outils biométriques d'analyse de la stabilité et de modélisation des interactions G×E sont décrits dans la littérature (Lin et al., 1986 ; Becker et Leon, 1988 ; Malosetti et al., 2013 ; van Eeuwijk, 2016). Ils sont tous fondés sur une décomposition statistique des performances des génotypes dans différents environnements, du type modèle d'analyse de variance à deux facteurs avec interaction (**Encadré I.2**). Certains modèles sont présentés ci-dessous en commençant par les méthodes descriptives (*cf.* I.3.2), puis les méthodes qui intègrent des informations basées sur la prise en compte, voire la caractérisation environnementale (*cf.* I.3.3). Enfin, la dernière section est dédiée à la détection des QTL et plus précisément à la modélisation des interactions QTL×E (*cf.* I.3.4).

### I.3.2. Méthodes descriptives pour l'analyse des interactions G×E

#### *Les indicateurs de stabilité : exemple de l'écovalence variétale*

Un certain nombre de paramètres de stabilité sont rapportés dans différentes études pour identifier les génotypes performants et stables. Ils reposent sur des méthodes paramétriques ou non paramétriques et permettent de quantifier l'interaction G×E à partir de la qualité d'ajustement des données par rapport au modèle additif. Les études montrent que ces différents paramètres de stabilité ne fournissent pas la même information (Moghaddam et al., 2011 ; Marjanovic-Jeromela et al., 2008) et qu'il est donc prudent de comparer plusieurs de ces indices pour identifier les génotypes stables et performants (Vaezi et al., 2019).

Sur la base du concept de stabilité dynamique (*cf.* I.3.1), le modèle proposé par Wricke (1962) est la méthode la plus simple et la plus couramment utilisée pour évaluer la stabilité d'un génotype. En



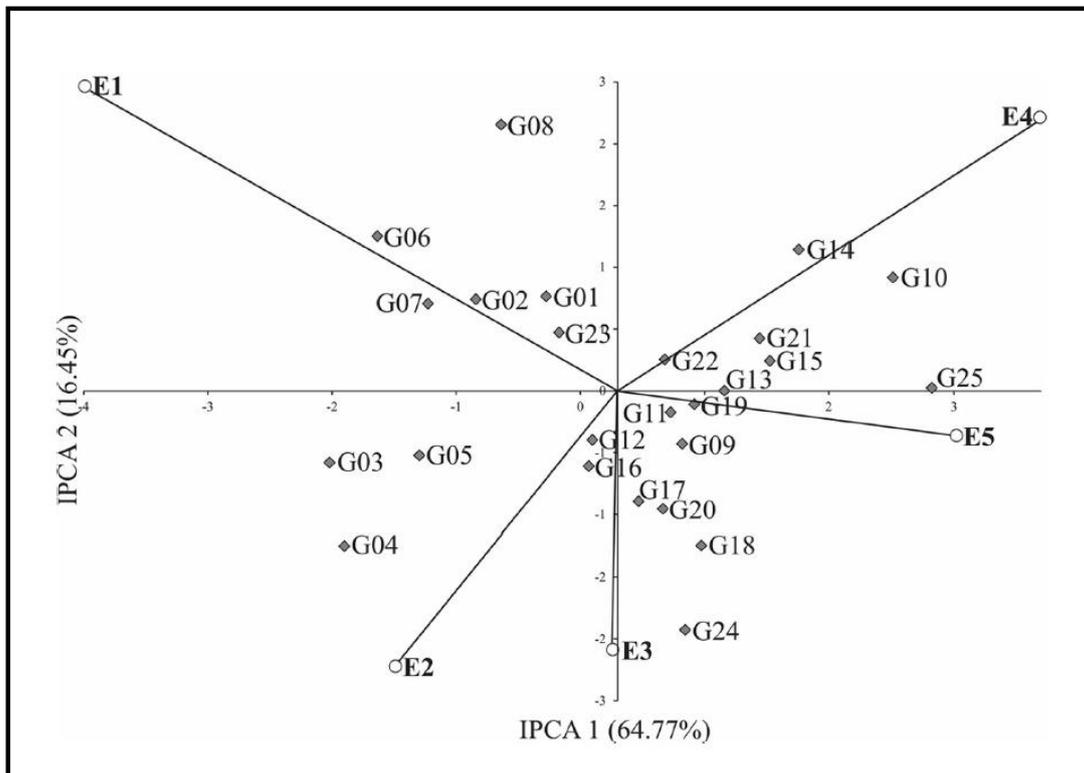
effet, ce modèle définit l'**écovalence variétale** comme la contribution du génotype à la somme totale des carrés des écarts de l'interaction (Encadré I.2). Un génotype est stable lorsqu'il présente une faible écovalence (Brancourt-Hulmel et al., 1997). L'écovalence permet donc d'évaluer la stabilité d'un génotype donné dans une population et de classer les génotypes les uns par rapport aux autres en fonction de leur interactivité/stabilité (Barbottin, 2004). De manière analogue, on peut définir et calculer l'écovalence environnementale pour chaque environnement (Parisot-Baril, 1992).

### *Structuration de l'interaction G×E : clustering, AMMI, GGE*

Des méthodes reposant sur la classification et le regroupement d'environnements ou de génotypes sont proposées dans le but de minimiser l'interaction G×E intra-groupe et à l'inverse de la maximiser entre les groupes (van Eeuwijk et al., 2016).

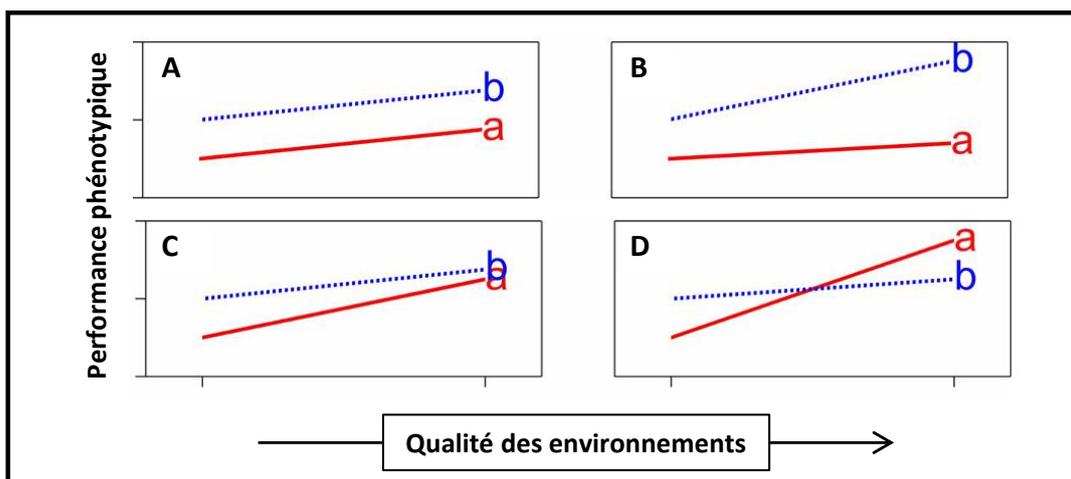
Cependant, une limite des approches de **clustering** réside dans le fait qu'elles imposent un choix pour fixer le nombre de groupes, choix qui doit être guidé pour structurer au mieux l'interaction. A titre d'exemple, les groupes d'environnements peuvent être définis en fonction de l'année, du lieu géographique, des itinéraires techniques, ou encore en fonction de pressions biotiques et/ou abiotiques. On parle alors de TPE (Target Population of Environments) dont la constitution va fortement influencer le regroupement de génotypes (TPG, Target Population of Genotypes) (van Eeuwijk et al., 2016). Dans le cas d'un stress hydrique, le TPE est réalisé en fixant environnements irrigués *versus* non irrigués, les génotypes choisis pour le TPG seront en lien avec la réponse au stress hydrique. Cependant, l'application d'un stress en conditions expérimentales a ses limites qui peuvent biaiser le TPE. Aussi, les TPE doivent être préférentiellement définis en fonction des performances à améliorer ou des stress auxquels les génotypes doivent s'adapter plutôt qu'en fonction d'une répartition basée sur des variations géographiques, temporelles ou de conduite culturale.

En complément des approches décrites précédemment, des approches multidimensionnelles permettent de prendre en compte simultanément de nombreux facteurs pouvant intervenir dans la réponse phénotypique. Parmi ces méthodes, les **modèles AMMI** (Additive Main effect and Multiplicative Interaction, Encadré I.2 ; Gauch, 1988 ; Gauch, 1992) et **GGE** (Yan and Kang, 2002) sont les plus répandus dans la littérature. Le modèle AMMI décompose le terme G×E en valeurs singulières environnementales et génotypiques selon plusieurs termes afin d'identifier les meilleurs discriminants génétiques et environnementaux qui expliquent l'interaction (van Eeuwijk et al., 2016). Le modèle AMMI a été utilisé sur le colza (Shafii et al., 1992 ; Nowosad et al., 2016) où il a permis d'identifier des génotypes peu sensibles à l'interaction tout environnement confondu et par environnement. Cependant il est difficile de donner un sens biologique aux différents termes injectés dans les modèles. Le modèle GGE est sensiblement équivalent au modèle AMMI à l'exception que



**Figure I.8 : Exemple de représentation du modèle AMMI par un biplot (d'après Nowosad et al., 2016)**

Biplot de l'interaction  $G \times E$  de 25 lignées de colza oléagineux d'hiver (G01-G25) dans 5 environnements différents (E1-E5) montrant les effets des deux premières composantes du modèle AMMI (IPCA1 et IPCA2). Les points à proximité de l'origine indiquent un génotype faiblement interactif. Les points proches entre eux présentent des génotypes ayant des comportements similaires.



**Figure I.9 : Utilisation des normes de réaction pour illustrer les interactions  $G \times E$  (d'après Malosetti et al., 2013)**

(A) Situation sans interaction  $G \times E$ , la performance peut être prédite par le modèle additif ((i), Encadré I.2); (B-D) Situations avec interaction  $G \times E$  qui conduisent à un changement du classement des génotypes (D) ou non (B, divergent; C, convergent).

c'est la somme  $G + G \times E$  qui est décomposée, ce qui en fait un modèle moins bien adapté pour évaluer des traits tels que le rendement (Yan and Kang, 2002 ; Fasahat et al., 2015 ; van Eeuwijk et al., 2016). Les sorties des modèles AMMI et GGE sont représentées sous la forme d'un biplot (Figure I.8) permettant d'explorer des groupes basés sur l'interaction (van Eeuwijk et al., 2016). Cette représentation en deux dimensions permet d'identifier facilement les génotypes qui participent le plus à l'interaction  $G \times E$  (Brancourt-Hulmel et al., 1997). Ce type de représentation permet également d'identifier des groupes d'environnements et/ou de génotypes (Yan et al., 2000).

### *Conclusion*

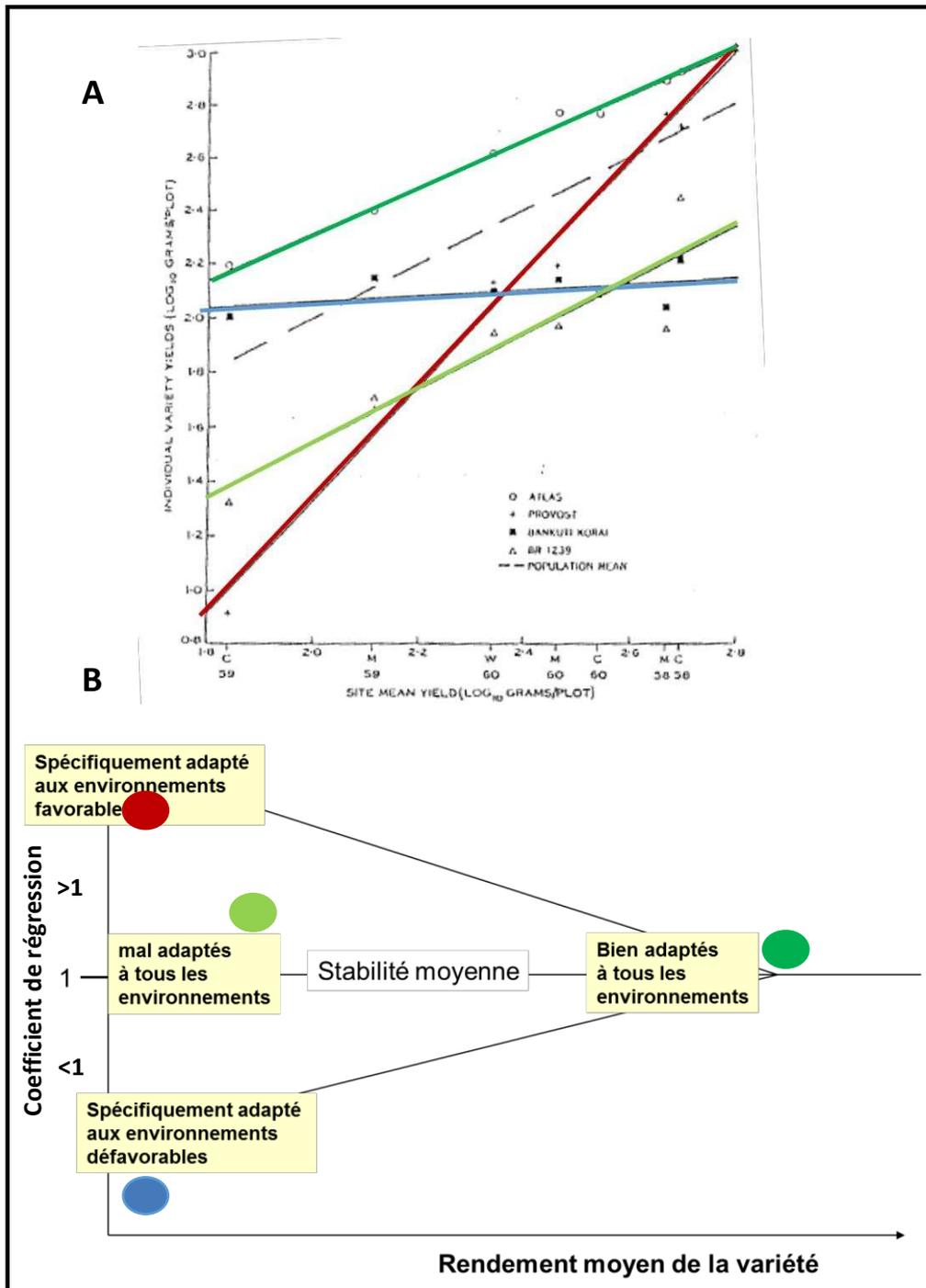
Les méthodes décrites ci-dessus sont très largement plébiscitées pour leurs performances pour décomposer le phénotype entre termes liés à l'environnement, au génotype et à l'interaction  $G \times E$  et offrent des approches descriptives de l'interaction  $G \times E$  en comparant les génotypes et/ou les environnements les uns par rapport aux autres (Brancourt-Hulmel et al., 1997). Cependant, en l'absence de covariables extérieures, il est impossible de donner un sens biologique à ces regroupements. D'ailleurs les biplots ne sont souvent représentés que sur les deux premiers axes car il est difficile de donner un sens aux axes suivants (van Eeuwijk et al., 2016). Une évolution de ces modèles consiste à prendre en compte l'environnement, voire à développer en amont une gamme de descripteurs environnementaux, afin de pouvoir expliquer les interactions  $G \times E$ .

### **1.3.3. Prise en compte de l'environnement pour soutenir les interactions $G \times E$**

#### *Modélisation des normes de réaction par des méthodes de régression*

L'ensemble des différentes réponses d'un seul génotype face à des conditions environnementales variées définit sa sensibilité environnementale ou **norme de réaction** (Falconer, 1990). Autrement dit, la norme de réaction est la relation entre les phénotypes possibles produits par un génotype donné et la gamme de valeurs d'une variable environnementale à laquelle il est exposé (Figure I.9). Dans une population, la variance des normes de réaction est sous-tendue par la présence des interactions  $G \times E$ , c'est-à-dire la différence de réponses des génotypes face aux environnements (Figure I.9).

Pour décomposer les termes d'interaction, c'est-à-dire réduire la part d'incertitude liée à l'interaction  $G \times E$ , différentes méthodes de régression sont décrites dont les résultats dépendent notamment de la qualité d'ajustement du modèle de régression aux données ( $R^2$  du modèle) et de la part de l'interaction qu'il explique.



**Figure I.10 : Illustration de la régression linéaire conjointe**

(A) Représentation des droites de régression linéaire pour des variétés contrastées de blé (d'après Finlay et Wilkinson, 1963). (B) Représentation des coefficients de régression par rapport à la performance des génotypes correspondants. On peut identifier des variétés très adaptées aux milieux favorables (rouge; pente forte et ordonnée à l'origine faible), des variétés adaptées aux milieux défavorables (bleu; pente faible et ordonnée à l'origine correcte), des variétés bien ou mal adaptées à tous les environnements (vert clair ou foncé respectivement; pentes équivalentes mais ordonnées à l'origine différentes).

**La régression linéaire conjointe** est de loin la plus employée, c'est une démarche d'explication empirique de l'interaction G×E par une régression simple où c'est la performance moyenne des variétés dans un milieu qui est prise pour caractériser ce milieu (Finlay et Wilkinson, 1963 ; **Encadré I.2 ; Figure I.10**). La pente de la régression est réajustée par rapport au rendement moyen des variétés dans un milieu. Le modèle statistique considère, la moyenne générale, les effets principaux du génotype et de l'environnement, ainsi que le coefficient de régression génotypique qui mesure la réponse du génotype à la potentialité du milieu qui elle-même est estimée par l'effet principal. On peut comparer les caractéristiques du modèle, comme la pente, par exemple à la valeur de 1 qui correspondrait à dire que le génotype se comporte dans les différents milieux comme la moyenne des génotypes, et qu'il permet d'exploiter la potentialité du milieu.

**La régression factorielle** est une méthode de régression qui décompose le terme d'interaction en une somme de covariables environnementales (et/ou de covariables génotypiques) (Denis, 1988 ; van Eeuwijk et al., 1996 ; Brancourt-Hulmel et al., 1997). L'intérêt majeur consiste à se baser sur ces covariables pour émettre des hypothèses biologiques pour expliquer l'interaction G×E (Malosetti et al., 2013). Des études montrent que cela permet d'expliquer entre 50-70% de l'interaction G×E. Cependant le nombre maximal de facteurs environnementaux que l'on peut utiliser dans ce type de modèle est égal au nombre d'environnements moins un (J-1). Dans le cas où le nombre de covariables est supérieure à J-1, il est alors nécessaire d'utiliser d'autres méthodes de régression telles que la méthode de **régression PLS (Partial Least Square)** (Vargas et al., 2001) permettant de considérer un nombre très important de covariables et de prendre en compte les corrélations entre elles afin de ne pas pénaliser le modèle. Par ailleurs, les modèles de régression permettent de prédire le comportement d'un génotype dans un environnement non testé à condition de connaître la valeur des covariables environnementales dans cet environnement. Afin d'évaluer et d'expliquer l'interaction G×E sur le colza, des approches de régressions PLS ont été testé sur les réseaux CTPS en se basant sur une dizaine de covariables telles que les INN à l'hiver ou le quotient photothermique à la floraison. Cette analyse a permis l'identification de facteurs limitants pour les environnements étudiés (Bagot et al., 2014).

### *Identification des covariables environnementales*

Dans les essais multi-environnementaux, il apparait que l'effet de l'environnement est souvent plus important que celui du génotype sur la variance phénotypique. Il est donc nécessaire de caractériser au mieux les facteurs environnementaux qui limitent les performances phénotypiques. Pour cela, plusieurs outils sont décrits. Une approche couramment utilisée consiste en un **diagnostic agronomique** qui vise à confronter l'état de l'environnement et l'état de culture à différents temps



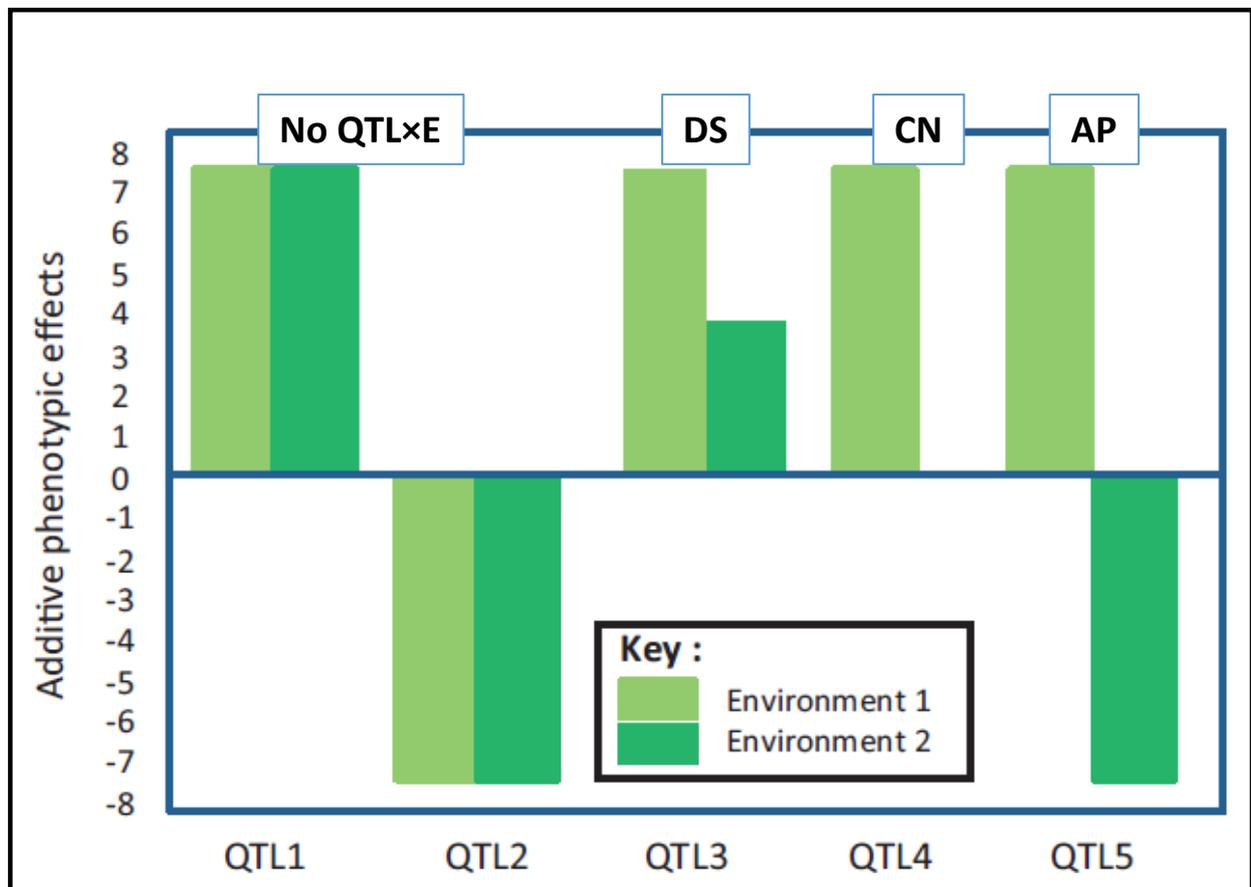
du cycle cultural à l'échelle de la parcelle. Pour faciliter ce diagnostic, il a été proposé d'utiliser des '**génotypes révélateurs**', *i.e.* bien caractérisés, pour lesquels on connaît le fonctionnement sous différentes contraintes et qui couvrent la gamme de précocité expérimentée (Brancourt-Hulmel et al., 2000; Lecomte, 2005). Le fonctionnement de ces génotypes renseigne sur la caractérisation de l'environnement d'expérimentation, et l'avantage est qu'ils peuvent être implantés dans un réseau. Un travail complémentaire consiste également à définir des **indicateurs de stress** en fonction de seuils prédéfinis (ex. un stress thermique défini par une température maximale supérieure à 30°C) (Bouffier et al., 2015). Enfin, l'utilisation des **modèles de culture** est décrite pour accéder à des covariables environnementales pour comprendre les facteurs d'adaptation des génotypes aux milieux (Chapman, 2008 ; Chenu et al., 2011 ; Ly et al., 2018 ; Rincet et al., 2019). Plusieurs modèles de culture sont aujourd'hui disponibles, ils peuvent être généralistes en intégrant des facteurs environnementaux majeurs (STICS - Brisson et al., 1998 ; APSIM - McCown et al., 1996 ; AZODYN - Jeuffroy et al., 2006), et peuvent avoir été adaptés par espèce (APSIM pour le colza oléagineux de printemps - Robertson and Lilley, 2016) ou spécifiquement développé pour une espèce (LINTUL-BRASNAP pour le colza oléagineux d'hiver - Habekotté, 1997 ; - SUNFLO pour le tournesol - Casadebaig et al., 2011). Cependant, il faut noter que les modèles de culture ne s'intéressent pas directement aux interactions G×E. Récemment, certains ont été paramétrés pour une variabilité génotypique très limitée et parfois peu adaptée aux variétés actuelles (ex. AZODYN colza).

#### I.3.4. Analyses QTL et interaction G×E

Afin d'identifier le déterminisme génétique de l'interaction G×E, les réponses phénotypiques à l'environnement doivent être interprétées en termes de QTL sous-jacents et des différences d'effets alléliques entre les environnements. L'interaction G×E est alors sous-tendue par plusieurs QTL ayant des effets spécifiques à un/plusieurs environnement(s) se traduisant comme l'interaction QTL × Environnement (QTL×E ; El Soda et al., 2014).

#### *Génétique quantitative et interaction G×E : l'apport des modèles linéaires mixtes*

L'utilisation du modèle linéaire en déclarant les facteurs en effet fixe est très largement utilisée dans l'étude de l'interaction G×E. Les modèles linéaires sont adaptés lorsque les facteurs sont indépendants et prennent un nombre de niveaux définis. Par exemple un traitement azote à deux niveaux, un limitant et un non limitant. Les modèles mixtes (*i.e.*, comprenant à la fois des facteurs fixes et des facteurs aléatoires) permettent de prendre en compte des facteurs à grand nombre de niveaux et dont les observations correspondent à un sous-ensemble de niveaux pris aléatoirement. Ces niveaux aléatoires suivent une loi normale (Verbeke et Molenberghs, 1997). Ainsi, les modèles



**Figure I.11 : Scénarios possibles d'interaction QTL×E observés concernant l'effet des QTL (d'après El Soda et al., 2014)**

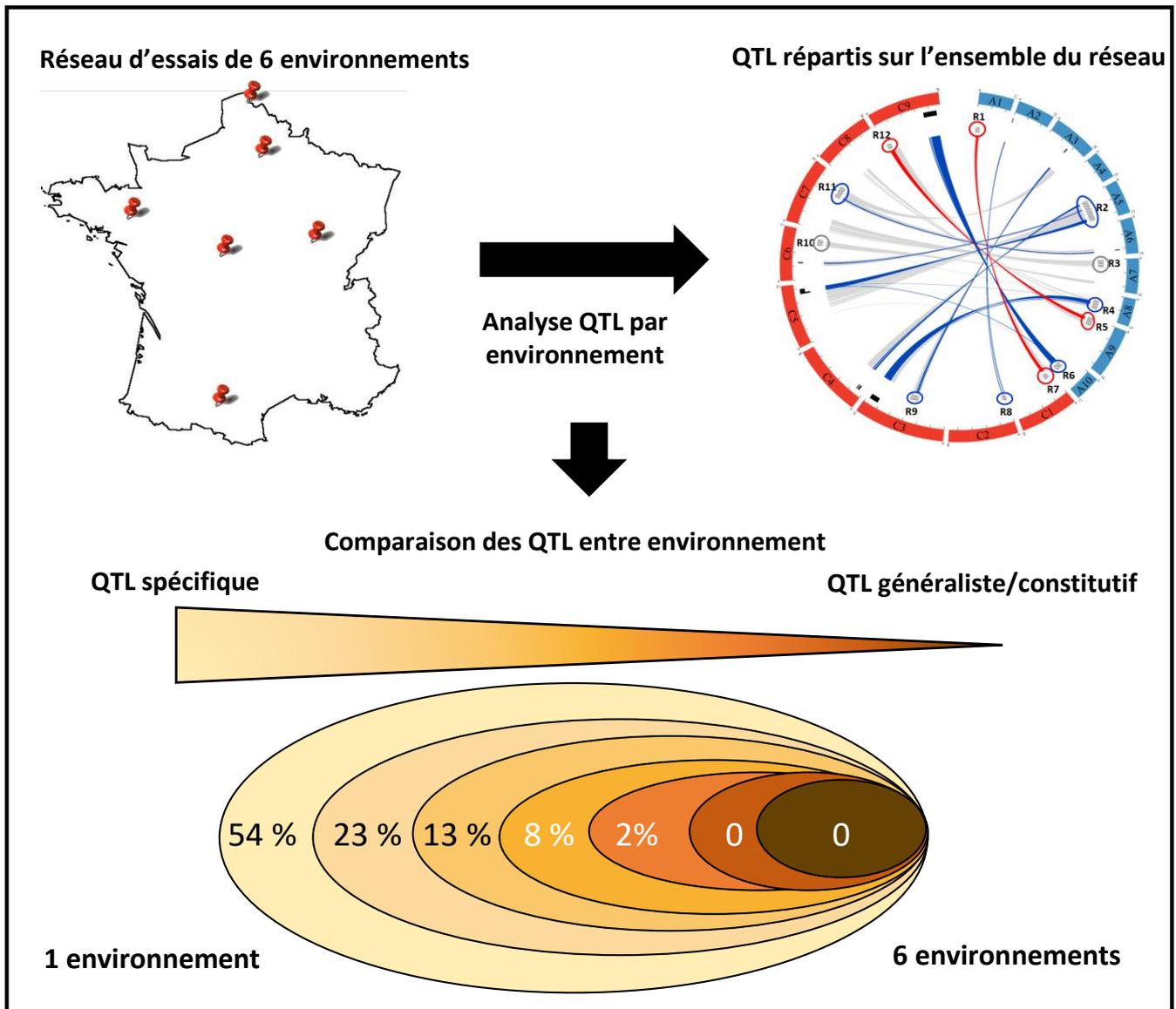
Les effets de 5 QTL identifiés dans deux environnements différents sont représentés. Les QTL1 et QTL2 sont des QTL stables entre environnements avec respectivement un effet positif ou négatif sur le trait étudié, ils ne montrent pas d'interaction QTL×E (No QTL×E). Les QTL3, QTL4 et QTL5 mettent en évidence l'interaction QTL×E en ayant i) un effet plus ou moins fort en fonction de l'environnement (QTL3; DS, sensibilité différentielle), ii) un effet dans un environnement mais pas dans un autre (QTL4; CN, neutralité conditionnelle) ou iii) des effets opposés entre les deux environnements (QTL5; AP, pléiotropie antagoniste).

mixtes facilitent la modélisation de l'hétérogénéité des variances génétiques et des corrélations entre environnements et permettent de prendre en compte le plan expérimental des essais individuels. De plus, la prédiction des traits entre les environnements, l'estimation des variances génétiques et le calcul de l'héritabilité sont simplifiés avec l'utilisation de modèles mixtes (van Eeuwijk et al., 2016). Le choix de définir un facteur comme fixe ou aléatoire dépend de la question à laquelle répondre ou de l'équilibre entre les différents niveaux d'un facteur. On peut alors traiter facilement des réseaux dont les essais ne sont pas indépendants (deux itinéraires techniques différents pour le même environnement), ou dont les populations ne sont pas indépendantes, ou encore lorsqu'il manque des données.

### *Qualification de l'interaction QTL×E*

Le profil d'expression des loci à travers les environnements révèle l'architecture génétique des interactions G×E. Des études ont qualifié la nature de l'interaction QTL×E par la comparaison des effets additifs des QTL pour des traits qui présentent des interactions G×E. Plusieurs scénarios sont ainsi définis (Figure I.11 ; Des Marais et al. 2013 ; El Soda et al., 2014). Si l'effet d'un QTL est avéré dans un environnement et est nul dans l'autre environnement, on parlera de QTL conditionnellement neutre (CN, conditional neutrality). Ces QTL sont relativement simples à utiliser car l'allèle favorable n'aura pas d'effet négatif sur le trait dans un tout autre environnement. Lorsque l'effet d'un QTL est significativement plus fort dans un environnement que dans l'autre, on parle de sensibilité différentielle (DS, differential sensitivity). Lorsque l'effet d'un QTL est opposé dans un environnement par rapport à l'autre, cela suggère que le QTL est sous l'influence d'une pléiotropie antagoniste (AP, antagonistic pleiotropy). Ces QTL doivent être utilisés de manière adéquate dans les différents environnements. Enfin, il peut ne pas y avoir de différence significative entre les effets du QTL dans un environnement et dans l'autre, son effet pouvant être négatif ou positif. Ce QTL est considéré comme un QTL constitutif (ou généraliste), il s'agit d'un QTL stable par rapport à l'environnement.

Toutes ces catégories d'effets de QTL sont à considérer avec recul au regard de la puissance des analyses génétiques pour la détection de QTL. En effet, la diversité génétique considérée dans les populations (nombre de recombinaisons), la densité de marqueurs moléculaires disponible, la précision des données phénotypiques considérées sont autant de paramètres qui vont jouer sur la précision pour la détection et la cartographie des QTL (pour revue voir Des Marais et al., 2013). Par conséquent, de nombreux auteurs interprètent à tort l'absence de QTL dans certains environnements comme relevant de la neutralité conditionnelle alors qu'il s'agit probablement de différences de réponses (sensibilité différentielle). L'utilisation de matériels appropriés (ex.



**Figure I.12 : Mise en évidence de l'interaction QTL×E au sein d'un réseau d'essais**

QTL de rendement issus d'une analyse menée sur un réseau de 6 environnements. La comparaison entre les QTL détectés pour chaque environnement permet d'identifier des QTL spécifiques aux environnements dans lesquels ils sont identifiés. La 1<sup>ère</sup> ellipse représente 54% des QTL, ce sont les loci qui sont spécifiques d'un environnement. La 2<sup>ème</sup> ellipse indique que 23% des QTL sont spécifiques pour une combinaison de 2 environnements. Aucun QTL n'est détecté dans tous les environnements du réseau dans cette étude. Adapté d'après Bouchet (2015).

populations hautement recombinantes, séquences de génomes de référence de qualité, phénotypage adapté) permet de pallier à ces limites. Par ailleurs, force est de constater que les QTL généralistes sont très peu nombreux dans la littérature (Des Marais et al., 2013).

### *Comparaison de QTL obtenus dans différents environnements*

Différentes méthodologies ont été proposées pour investiguer l'interaction G×E. La méthode la plus souvent utilisée consiste à comparer les QTL détectés à l'échelle de chaque environnement. Dans cette approche, la question est de savoir comment et dans quelle mesure l'architecture des effets additifs change en fonction des environnements. En complément, on peut également effectuer une analyse QTL à l'échelle globale d'un réseau (Buckler et al., 2009). Pour cela, la valeur génétique de chaque individu est estimée pour l'intégralité du réseau et utilisée comme variable pour la détection de QTL (El Soda et al., 2014). La détection à l'échelle du réseau se veut plus résolutive que la détection à l'échelle de l'essai, car elle considère un plus grand effectif statistique permettant une meilleure estimation des effets génétiques (Jiang et Zeng, 1995). Ces approches complémentaires ont été menées sur plusieurs espèces dont le colza (Dufey et al., 2012 ; Bouchet et al., 2014 ; 2016a) et permettent de qualifier de stables des QTL identifiés à l'échelle du réseau et/ou dans presque tous les environnements testés et montrant des effets similaires sur le phénotype. A contrario, les QTL sont qualifiés d'interactifs quand ils sont détectés spécifiquement pour certains environnements et/ou ont des effets différents sur le trait. Ainsi sur le colza, Bouchet et al. (2016a) ont pu mettre en évidence 51 QTL stables pour le rendement et ses composantes (détectés à l'aide d'un modèle linéaire ajusté à l'échelle du réseau) ainsi qu'un grand nombre de QTL interactifs détectés spécifiquement dans un environnement (Figure I.12).

Concernant le rendement chez le colza, c'est un caractère complexe et polygénique, chaque déterminant n'expliquant qu'un faible pourcentage de la variance phénotypique (<15%). Dans la littérature, un grand nombre d'études ont décrit la détection de loci contrôlant l'élaboration du rendement en graines et de ses composantes en utilisant des populations variées (HD, RIL, panels d'association) qui couvrent une large diversité génétique (Tableau I.2 ; Delourme et al., 2018). Pour synthétiser cette information, des méta-analyses ont été conduites, révélant plusieurs dizaines de régions génomiques liées au rendement (Shi et al., 2009 ; Zhao et al., 2016 ; Luo et al., 2017 ; Raboanatahiry et al., 2018). Ces études sont complémentaires d'une part parce qu'elles utilisent une diversité génétique variée qui permet de détecter différents loci mais aussi parce qu'elles sont conduites soit en conditions optimales de culture, soit en conditions limitantes en focalisant sur une ou plusieurs contraintes abiotiques et/ou nutritionnelles (Gül et al., 2003 ; Miro, 2010 ; Ding et al., 2012 ; 2014 ; Zhao et al., 2012 ; Shi et al., 2013 ; Bouchet et al., 2014 ; 2016a ; Fletcher et al., 2015 ;

**Tableau I.2 : QTL détectés pour les traits relatifs au rendement en réponse ou non à un stress abiotique chez le colza**

Traits/Traitement	Population(s)	Nombre et type de marqueurs	Nombre d'environnements	Analyses	Références
SY, TSW	4 BC	276 - RFLP	2	LA	Butruille et al. (1999)
SY, TSW	MF216 × P1804 (150 HD) RV128 × P1804 (150 HD)	218 - RFLP 250 - RFLP	4	LA	Quijada et al. (2006)
SY, TSW	MF216 × P1804 (150 HD) RV128 × P1804 (150 HD)	218 - RFLP 205 - RFLP	3	LA	Udall et al. (2006)
SY, TSW	Express 617 × R53 (250 HD) Population de test-cross (250 ind.)	377 - AFLP, SSR	4	LA	Radoev et al. (2008)
TSW	Tapidor × Ningyou7 (202 HD)	786 - AFLP, RFLP, SNP, SSR, STS, SSCP, CAPS	3-10	LA	Shi et al. (2009)
SY, TSW	Express 617 × R53 (250HD et 250 test-cross) Express 617 × V8 (250HD et 250 test-cross)	377 - AFLP, SSR 476 - AFLP, SSR	4	LA	Basunanda et al. (2009)
SY	(150 HD) <i>croisement non décrit</i>	387 - SRAP, SSR	3	LA	Chen et al. (2010)
TSW	SW Hickory × JA177 (238 HD)	327 - SSR	2	LA	Fan et al. (2010)
TSW	HZ396 × Y106 (140 HD)	345 - SSR, AFLP	6	LA	Zhang et al. (2011)
TSW	S1 × S2 (186 RIL) No2127 × S1 (192 RIL)	289 - SSR	2 1	LA	Yang et al. (2012)
TSW	Panel (192 accessions)	1191 - SSR, AFLP	3	GWAS	Cai et al. (2014)
TSW	Zhongshuang11 × No73290 (184 F2, F2:3, F2:4)	529 - SSR, STS, SNP	5	LA	Li et al. (2014)
TSW	8008-9 × 4942C-5 (181 HD)	385 - SSR, AFLP	4	LA	Qi et al. (2014)
SY	Panel (155 accessions)	35791 - SNP	4	GWAS	Luo et al. (2015)
SY	Skipton × Ag-Spectrum (145 HD)	7716 - DArTseq	2	LA	Raman et al. (2016)
SY, SN, TSW	Huashuang × J7005 (254 HD)	Puce 6K SNP	4	LA	Cai et al. (2016)
SY	KenC-8 × N53-2 (348 HD)	400 - SSR, SRAP, STS, IFLP	8	LA	Zhao et al. (2016)
SY	Panels (217 et 188 accessions)	Puce 6K SNP	2	GWAS	Körber et al. (2016)
SN	Zhongshuang11 × 73290 (184 RIL)	Puce 60K SNP	4	LA	Yang et al. (2016)
SY	Panels (175 lignées et 525 hybrides)	Puce 60K SNP	-	LA et GWAS	Zhao et al. (2016)
SN	KenC-8 × N53-2 (348 HD)	403 - SSR, SRAP, STS, IFLP	4	LA	Wang et al. (2016)
SY	Hi-Q × RIL-144 (110 HD)	172 - SSR, AFLP	9	LA	Rahman et al. (2017)
SY	Panels (318 hybrides et 180 lignées parentales)	Puce 60K SNP	3	GWAS, GS	Liu et al. (2017)
SY	Tapidor × Ningyou7 (182 HD)	Puce 60K SNP	19	LA	Luo et al. (2017)
TSW	1167 × HZ396 (167 HD)	Puce 60K SNP	-	LA	Yang et al. (2017)
SY, TSW	Sensibar × Oase (200 HD)	1642 - SNP	1	LA	Fattahi et al. (2018)
SY, TSW	KenC-8 × N53-2 (348 HD)	17978 - SNP 101 - SSR, STS	11	LA	Chao et al. (2019)
TSW	KenC-8 × N53-2 (348 HD)	17978 - SNP 101 - SSR, STS	8	LA	Zhao et al. (2019)
TSW	ZY50 × 7-5 (189 HD) et Panel (505 accessions)	Puce 60K SNP	4	LA et GWAS	Wang et al. (2020)

**Tableau I.2 : (suite)**

Traits/Traitement	Population(s)	Nombre et type de marqueurs	Nombre d'environnements	Analyse QTL	Références
SN, TSW <b>Contrainte N</b>	Mansholt-HR × Samourāi (142 HD)	185 - RFLP	8	LA	Gül et al. (2003)
SY, SN, TSW <b>Contrainte P</b>	B104-2 × Eyou Changjia (124 RL)	840 - SSR, AFLP, SRAP	4	LA	Ding et al. (2012)
SY, SN <b>Contrainte N</b>	Tapidor × Ningyou7 (174/94 HD)	786 - AFLP, RFLP, SNP, SSR, STS, SSCP, CAPS	2	LA	Miro (2010)
SY, SN, TSW <b>Contrainte B</b>	Tapidor × Ningyou7 (181HD) QY10 × Bakow (200 HD)	621 - RFLP, SSR, SNP, ALFP 486 - SSR	2 3	LA	Zhao et al. (2012)
SY, SN, TSW <b>Contrainte P</b>	Tapidor × Ningyou7 (188 HD)	798 - AFLP, RFLP, SNP, SSR, STS, SSCP, CAPS	6	LA	Shi et al. (2013)
SY, SN, TSW <b>Contrainte N</b>	Aviso × Montego (115 HD)	2301 - SNP	6	LA	Bouchet et al. (2014)
SY, SN, TSW <b>Contraintes P ou B</b>	QY10 × Bakow (200 HD) B104-2 × Eyou Changjia (124 RIL)	486 - SSR, AFLP, SRAP 840 - SSR, AFLP, SRAP	6 4	LA	Ding et al. (2014)
SY <b>Contrainte hydrique</b>	Witchita × IMC106RR (225 HD)	1179 - SNP	2	LA	Fletcher et al. (2015)
SY <b>Contrainte N</b>	Alesi-bzh × H30 (242 HD)	471 - SNP, SSR	8	LA	Miersch et al. (2016)
SY, SN, TSW <b>Contrainte N</b>	Panels (69 et 92 accessions) Aviso × Montego (112 HD) Tenor × Express (75 HD)	Puce 60K SNP	14	LA et GWAS	Bouchet et al. (2016a)



Miersch et al., 2016 ; Ding et al., 2014). En cela, elles peuvent apporter des informations sur les déterminants génétiques et moléculaires de la réponse du colza à certains stress environnementaux (Tableau 1.2).

#### *Identifier des QTL sur les indices de stabilité (écovalences, normes de réaction, ...)*

Les paramètres décrits pour analyser la stabilité (cf. 1.3.2 et 1.3.3) sont susceptibles de montrer de la variabilité génétique héritable et donc peuvent être considérés comme des variables quantitatives dans les modèles de détection de QTL (Des Marais et al., 2013 ; Malosetti et al., 2013). Ainsi, Bouchet et al. (2016a) ont effectué des détections de QTL en se basant sur les valeurs d'écovalences variétales (cf. 1.3.2). Une valeur d'interactivité est alors attribuée à chaque génotype. Avec cette méthode, 25 QTL en lien avec l'interaction G×E ont été identifiés suggérant que ces régions confèrent une adaptation spécifique aux environnements. De même, il est possible d'utiliser les paramètres de la régression linéaire conjointe (pente et déviation de la régression) pour des analyses QTL.

#### *Identification de QTL par groupes d'environnements caractérisés par des covariables environnementales*

Les méthodes de détection de QTL précédemment décrites (comparaison de QTL, QTL de stabilité) n'offre pas la possibilité d'émettre des hypothèses sur les fonctions biologiques sous-jacentes à l'adaptation spécifique des génotypes en fonction des contraintes abiotiques subies. Ainsi, une dernière approche décrite ici repose sur l'identification de QTL en considérant des regroupements d'environnements simulant des scénarios de contrainte(s) environnementale(s). Les QTL identifiés sont ensuite appréhendés en fonction de la typologie des groupes d'environnement et des facteurs limitants qui caractérisent ces groupes. Cette approche a été déployée avec succès sur maïs (Millet et al., 2016) et blé (Touzy et al., 2019) en ayant au préalable définis des scénarios climatiques. Dans les deux études, les analyses QTL ont été conduites pour chaque scénario et ont révélé des loci associés à la tolérance aux contraintes environnementales ciblées dans chaque scénario. Cette approche est en mesure d'identifier l'interaction QTL×E tout en permettant de mettre en relation les effets du QTL en fonction de covariables environnementales. Cette relation permet alors d'émettre des hypothèses sur les mécanismes biologiques mis en place et aide à la priorisation de gènes candidats sous-jacents aux QTL. En conclusion, ces méthodes permettent d'approcher l'interaction QTL×E en identifiant les différentes typologies de QTL au regard de covariables environnementales pouvant aider dans la priorisation de cibles pour les programmes de sélection par exemple.



## I.4. Problématique et objectifs de la thèse

---

L'objectif de ma thèse est d'identifier et de caractériser les déterminants génétiques et moléculaires de la stabilité du rendement en grains chez le colza, dans un contexte de réduction des intrants azotés. Nous avons vu précédemment que la notion de stabilité d'un trait pouvait être mise en relation avec l'interaction G×E. Dans ce travail, nous avons choisi d'identifier au préalable les facteurs limitant le rendement dans différents environnements afin de mieux appréhender et expliquer l'interaction G×E et d'en caractériser les déterminants génétiques. Trois questions de recherche ont structuré mon travail :

- **Quels sont les principaux facteurs limitant le rendement chez le colza et comment peuvent-ils aider à disséquer l'effet de l'environnement et de l'interaction G×E ?**
- **Quels sont les déterminants génétiques du rendement et comment se structurent les QTL détectés en termes de stabilité ou d'interactivité ?**
- **Est-il possible d'identifier les déterminants génétiques contrôlant spécifiquement l'interaction G×E et de les caractériser, en fonction de typologies d'environnements ?**

Notre stratégie a consisté à 1) comprendre l'effet environnemental sur le rendement du colza et définir des groupes d'environnements ('envirotypes') sur la base de facteurs pédoclimatiques limitant le rendement et 2) identifier les déterminants génétiques qui contrôlent le rendement en graines à l'échelle d'un large réseau ou des envirotypes.

Le chapitre II est consacré à la définition des envirotypes. Sur la base de données publiées et complétées par le savoir d'experts, nous avons d'abord déterminé des périodes de développement clés au cours du cycle du colza durant lesquelles certains stress pouvaient avoir un impact sur le rendement, puis nous avons associé des indicateurs pédoclimatiques à ces périodes. Nous avons ensuite calculé ces indicateurs pour deux génotypes moyens de colza présents dans un réseau d'essais très large et utilisé une méthode de régression adaptée à la dimension du jeu de données pour identifier les facteurs limitant le rendement dans ce réseau. Ces facteurs ont ensuite permis le regroupement des environnements en fonction de leur impact sur le rendement et sur l'interaction G×E.

Le chapitre III s'intéresse à l'identification des déterminants génétiques de la stabilité du rendement. Des analyses d'association (GWAS) ont été conduites sur les variables liées au rendement (SY, SN, TSW) en considérant une population de 173 génotypes de colza d'hiver oléagineux cultivés dans 22 essais. Des analyses à l'échelle de chaque essai individuel ont permis d'identifier des déterminants



spécifiques d'un essai ou communs à plusieurs essais. Une analyse menée à l'échelle du réseau entier a révélé les QTL liés à la stabilité du rendement. Enfin, l'analyse génétique à partir des envirotypes a permis de cibler des QTL spécifiques de l'interaction ou de la réponse à un facteur limitant. L'ensemble de ces résultats permet de dresser un bilan des déterminants génétiques liés à la stabilité du rendement chez le colza.



## CHAPITRE II

---

Identification de facteurs limitant le rendement chez le colza pour caractériser les interactions G×E

---



## II.1. Préambule

---

Au sein d'un réseau d'essais, le rendement d'une variété résulte de l'expression des effets du génotype, de l'environnement et de l'interaction génotype  $\times$  environnement (Nowosad et al., 2016). Chez le colza d'hiver, Bouchet et al. (2016a) ont montré dans leur étude que l'effet génétique et l'effet environnemental représentaient en moyenne 46% et 43% de la variance respectivement ; les 11% restants s'expliquaient par l'interaction génotype  $\times$  environnement. Cette répartition questionne sur la diversité et la nature des facteurs environnementaux qui impactent le rendement dans un réseau donné et leurs effets sur l'interaction G $\times$ E. Une possibilité pour aborder cette question est d'identifier des typologies d'environnements (appelées envirotypes) en fonction des facteurs ayant impacté le rendement. Dans cette définition, chaque envirotype correspond alors à un groupe d'environnements pour lequel les facteurs environnementaux affectant le rendement présentent un comportement similaire. L'hypothèse sous-jacente est qu'en regroupant ainsi les environnements, on peut réduire, à l'intérieur de ces envirotypes, la part des effets de l'environnement et de l'interaction G $\times$ E. La première étape de cette étude consiste à identifier des covariables environnementales les plus précises possibles et en adéquation avec le cycle cultural pour décrire de manière exhaustive les différents environnements. La seconde étape consiste à expliquer une variable d'intérêt (ici le rendement) à l'aide de ces covariables pour identifier les facteurs qui limitent l'expression de ce trait. La dernière étape consiste enfin à regrouper les environnements sur la base de ces facteurs limitants pour définir les envirotypes.

L'objectif de ce chapitre est de produire une méthode robuste permettant de mieux appréhender l'impact des effets environnementaux et de l'interaction G $\times$ E sur l'élaboration du rendement grainier chez le colza et de l'utiliser pour la caractérisation de réseaux. Pour cela, nous avons développé une approche basée sur l'envirotypage à partir d'un réseau conséquent d'environnements (cf. II.2). Dans un premier temps, nous avons établi une liste d'indicateurs pédoclimatiques en se basant sur des stress abiotiques ou nutritionnels classiquement décrits au cours du cycle du colza ou dans la littérature (Habekotté, 1997 ; Diepenbrock, 2000 ; Lecomte, 2005) ainsi que des validations/retours d'experts. Parmi ceux-ci, les facteurs limitant le rendement ont ensuite été identifiés à l'aide d'une méthode de régression PLS (Partial Least Square) permettant d'avoir un grand nombre de variables explicatives par rapport au nombre d'environnements testés, et déjà décrite dans l'étude des interactions G $\times$ E (Vargas et al., 2001). Nous avons couplé à la régression, une méthode de sélection de variables basée sur la valeur du coefficient de régression de chaque indicateur (Gauchi et Chagnon, 2001) permettant d'identifier individuellement les facteurs limitants. Ces facteurs ont ensuite été utilisés pour la caractérisation des envirotypes. Enfin, l'effet de l'envirotypage sur

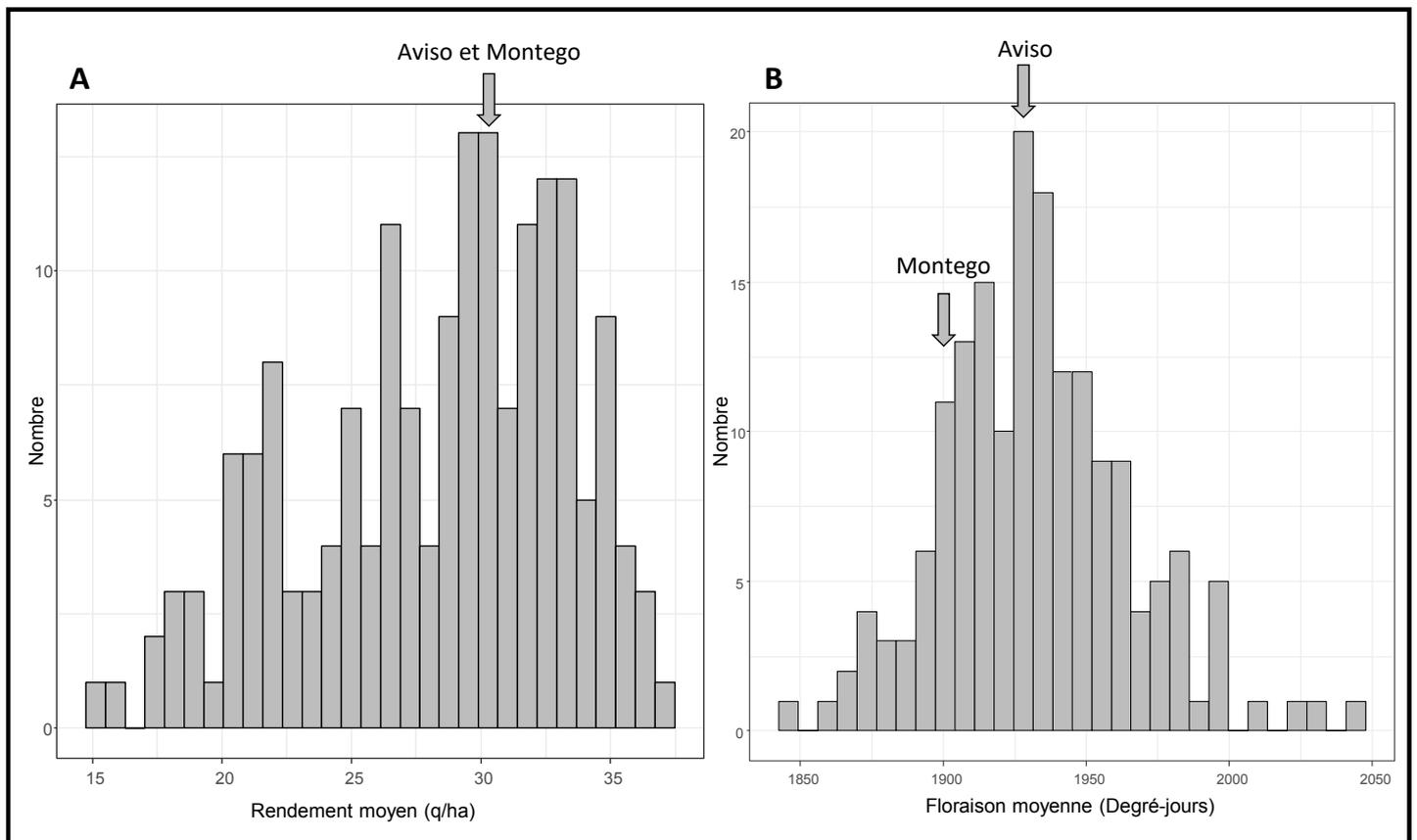


l'interaction G×E a été évalué. Cette méthode a d'abord été testée sur un réseau de 20 environnements (combinaisons 'Lieu x Année') conduits en condition de fertilisation azotée non limitante (cf. II.3) puis étendue à un réseau constitué de 32 environnements pour lesquels la fertilisation azotée était pilotée (régime limitant vs. régime optimal) (cf. II.4). Le réseau élargi est constitué de 47 essais (combinaisons 'Lieu x Année x Condition azotée') dont 28 en condition limitante d'azote et 19 en condition non limitante en azote. Une comparaison de ces approches entre elles et avec la littérature est présentée (cf. II.5).



**Figure II.1 : Localisation des environnements des essais étudiés dans la thèse**

Les environnements d'expérimentation sont localisés sur le territoire français. Chaque environnement est une combinaison 'Lieu x Année'. La nomenclature utilisée pour nommer chaque environnement comprend : des lettres qui font référence à la localité (ex. LR, Le Rheu) et un nombre qui correspond à l'année de récolte (ex. 11, récolte 2011). L'ensemble des essais est donné au **Tab II.1**



**Figure II.2: Positionnement des rendements et des dates de floraison d'Aviso et Montego dans le panel de 173 génotypes**

(A) Rendements en graines d'Aviso et de Montego. (B) Floraison en degré jours après semis d'Aviso et Montego. Les données des deux génotypes contrôles sont indiquées par rapport aux données de 173 génotypes de colza d'hiver (population utilisée pour le réseau #3 GWAS, cf. **Tab II.1**).

## II.2. Description du réseau expérimental

---

### *Réseau*

Mon travail de thèse a été supporté par l'acquisition de données de rendement dans un réseau expérimental très large mis en place dans le cadre des projets GENERGY et RAPSODYN. Ce réseau est constitué de 32 environnements, définis comme les combinaisons 'Lieu × Année', permettant d'explorer des conditions pédoclimatiques variées. Ainsi, les lieux sont répartis sur l'ensemble du territoire français (Figure II.1) et 7 saisons culturales sont concernées avec des récoltes en 2011 (2 lieux), 2012 (1), 2013 (2), 2014 (5), 2015 (7), 2016 (3) et 2018 (12). Dans chaque environnement, les expérimentations ont été menées suivant un itinéraire technique commun (protection contre les stress biotiques notamment) et pour lequel les conditions de fertilisation azotée pouvaient varier mais suivant des règles de décision identiques. Ainsi la condition non limitante (N2) correspondait à un apport azoté X calculé selon la méthode du bilan (Parnaudeau et al., 2009) pour un objectif de rendement final de 30-35 q/ha tandis que la condition limitante (N1) correspondait à la dose X – 80-100 U (kg/ha) d'azote selon les conditions environnementales (NB : N1 peut être égale à zéro dans certains cas). Ainsi, on définit un essai comme la combinaison 'Lieu × Année × Fertilisation azotée'. Au total, le réseau complet comprend 47 essais dont 28 essais 'N1' et 19 essais 'N2' (Tableau II.1).

### *Matériel végétal expérimenté et choix de deux génotypes contrôles*

Le matériel phénotypé dans les différents essais a également varié pour répondre aux différents objectifs des projets GENERGY et RAPSODYN (Tableau II.1). Cependant, les génotypes Aviso et Montego sont communs à l'ensemble des expérimentations. Ces génotypes récents se situent dans la moyenne des génotypes issus du panel d'association pour leur floraison ainsi que leur rendement (Figure II.2). Ils sont néanmoins contrastés pour leur précocité de floraison (différence de 40°C.J) et en hauteur (différence de 13 cm), Montego étant à la fois le génotype le plus précoce et le plus petit. La présence de ces deux génotypes dans l'ensemble du réseau et de leur comportement représentatif du panel d'association permet de les considérer comme « témoins » pour définir les facteurs limitants.

### *Démarche*

La première étape de cette étude consistait à développer la méthode d'envirotypage. Nous avons d'abord choisi de nous affranchir des contraintes nutritionnelles sur un réseau réduit d'environnements, afin de déterminer les principaux facteurs pédoclimatiques ayant limité le

**Tableau II.1 : Réseau expérimental complet et définition des sous-réseaux pour les différentes analyses**

Identifiant <sup>(a)</sup>	Lieu <sup>(b)</sup>	Année <sup>(c)</sup>	Population <sup>(d)</sup>	Azote <sup>(e)</sup>	Réseau #1 <sup>(f)</sup>	Réseau #2 <sup>(g)</sup>	Réseau #3 GWAS <sup>(h)</sup>
LR11_N1	Le Rheu	2010-2011	HD	N1	Non	Oui	Non
LR11_N2				N2	Oui	Oui	Non
Md11_N1	Mondonville	2010-2011	HD	N1	Non	Oui	Non
Md11_N2				N2	Oui	Oui	Non
LR12_N1	Le Rheu	2011-2012	HD	N1	Non	Oui	Non
LR12_N2				N2	Oui	Oui	Non
LR13_N1	Le Rheu	2012-2013	HD	N1	Non	Oui	Non
LR13_N2				N2	Oui	Oui	Non
Dij13_N2	Dijon	2012-2013	HD	N2	Oui	Non	Non
Ch14_N1	Châteauroux	2013-2014	Panel173	N1	Non	Oui	Oui
Ch14_N2				N2	Oui	Oui	Oui
Md14_N2	Mondonville	2013-2014	Panel173	N2	Oui	Non	Non
Pre14_N1	Prémesques	2013-2014	Panel173	N1	Non	Oui	Oui
Pre14_N2				N2	Oui	Oui	Oui
Sel14_N1	Selommès	2013-2014	Panel173	N1	Non	Oui	Oui
Sel14_N2				N2	Oui	Oui	Oui
Ver14_N1	Verpillères	2013-2014	Panel173	N1	Non	Oui	Oui
Ver14_N2				N2	Oui	Oui	Oui
Dij15_N1	Dijon	2014-2015	Panel173	N1	Non	Oui	Oui
Dij15_N2				N2	Oui	Oui	Oui
LR15_N1	Le Rheu	2014-2015	Panel173	N1	Non	Oui	Oui
LR15_N2				N2	Oui	Oui	Oui
Md15_N1	Mondonville	2014-2015	Panel173	N1	Non	Oui	Oui
Md15_N2				N2	Oui	Oui	Oui
Pre15_N1	Prémesques	2014-2015	Panel173	N1	Non	Oui	Oui
Pre15_N2				N2	Oui	Oui	Oui
Sel15_N1	Selommès	2014-2015	Panel173	N1	Non	Oui	Oui
Sel15_N2				N2	Oui	Oui	Oui
Ver15_N1	Verpillères	2014-2015	Panel173	N1	Non	Oui	Oui
Ver15_N2				N2	Oui	Oui	Oui
Yeb15_N1	Yébles	2014-2015	Panel173	N1	Non	Oui	Oui
Yeb15_N2				N2	Oui	Oui	Oui
Liv16_N2	Liverdy	2015-2016	F3	N2	Oui	Non	Non
LR16_N1	Le Rheu	2015-2016	Panel87	N1	Non	Oui	Non
LR16_N2				N2	Oui	Oui	Non
Pre16_N2	Prémesques	2015-2016	F3	N2	Oui	Non	Non
Aut18_N1	Autainville	2017-2018	Elites	N1	Non	Oui	Non
Blo18_N1	Blois	2017-2018	Elites	N1	Non	Oui	Non
Blo18_N2				N2	Non	Oui	Non
Ch218_N1	Chartres	2017-2018	Elites	N1	Non	Oui	Non
Chl18_N1	Châlons-en-Champagne	2017-2018	Elites	N1	Non	Oui	Non
Chr18_N1	Chartres	2017-2018	Elites	N1	Non	Oui	Non
Cup18_N1	Cuperly	2017-2018	Elites	N1	Non	Oui	Non
Cv18_N1	Ormoy	2017-2018	Elites	N1	Non	Oui	Non
Liv18_N1	Liverdy	2017-2018	Elites	N1	Non	Oui	Non
Liv18_N2				N2	Non	Oui	Non
Lou18_N1	Louville	2017-2018	Elites	N1	Non	Oui	Non
LR18_N1	Le Rheu	2017-2018	Panel10	N1	Non	Oui	Non
LR18_N2				N2	Non	Oui	Non
Pre18_N1	Prémesques	2017-2018	Elites	N1	Non	Oui	Non
Vil18_N1	Villemardy	2017-2018	Elites	N1	Non	Oui	Non

(a) Identifiant de l'essai (Lieu × Année de récolte × Fertilisation azotée). (b) Localisation de l'essai (cf. Fig II.1). (c) Campagne de culture. (d) Matériel végétal expérimenté : population d'haploïdes doublés (HD), sets de diversité plus ou moins larges (Panel173, Panel87, Panel10), populations recombinantes (F3), matériel élite des partenaires privés (Elites). (e) Fertilisation azotée pratiquée : N2 non limitante, N1 limitante. (f) Réseau utilisé par Corlouer et al. (2019) (cf. II.3). (g) Réseau étendu (cf. II.4). (h) Réseau utilisé pour l'évaluation de l'interaction G×E et les analyses génétiques (cf. Chapitre III).

rendement. Pour cela nous avons utilisé le réseau #1 (cf. II.3, Tableau II.1) comprenant 20 essais non limitants (N2) organisés entre la campagne 2010-2011 et la campagne 2015-2016. Les essais de ce réseau ont un rendement moyen de 34 q/ha et un indice de nutrition azotée (INN) moyen de 1.18 confirmant l'absence de stress azoté. Cependant le rendement moyen des essais varie entre 24 et 47 q/ha suggérant la présence de contraintes additionnelles au cours du cycle cultural.

Dans un deuxième temps, le réseau d'essais a été étendu à des essais N1 (réseau #2), limités en azote, et à une saison culturale supplémentaire (2017-2018) (cf. II.4 ; Tableau II.1). La prise en compte des contraintes nutritionnelles dans le réseau #2 permet ainsi d'étudier l'effet de différents stress pouvant intervenir de manière concomitante. En effet, il a été montré que la combinaison de stress peut entraîner des effets différents par rapport à l'application de stress de manière séquentielle (Mittler, 2006). De plus, bien que l'azote ait un impact positif sur le rendement (Rathke et al., 2006), certains auteurs suggèrent qu'une surfertilisation azotée pourrait induire une réduction du rendement (Gammelvind et al., 1996 ; Sieling and Christen, 1997). Afin d'identifier si les essais N1 et N2 du réseau #2 sont respectivement et réellement des essais limitants et non limitants, nous avons représenté les indices de nutrition azotée (INN ; Encart II.1) calculés deux semaines environ après fertilisation en fonction des rendements moyens d'Aviso et Montego (Figure II.3). Pour 6 essais du réseau #2 (Sel14\_N1, Sel14\_N2, Sel15\_N1, Sel15\_N2, Pre15\_N1 et Pre15\_N2), les valeurs d'INN obtenues expérimentalement n'étant pas disponibles, des valeurs imputées à l'aide du package *missMDA* (Josse et Husson, 2016) ont été utilisées. On définit deux catégories d'essais sur la base des rendements moyens, les essais à rendement élevé (SY+ ;  $SY > 30$  q/ha) et ceux à rendement plus faible (SY- ;  $SY < 30$  q/ha). On définit également les essais non stressés en azote (INN+) et stressés (INN-) en fonction de la valeur seuil de 0.9 de l'INN. On obtient ainsi 4 groupes d'essais (Figure II.3) : i) 14 essais sont INN+/SY- dont 7 N1 et 7 N2 ; ii) 17 essais sont INN+/SY+ dont 6 N1 et 11 N2 ; iii) 11 essais sont INN-/SY- tous classés N1 et iv) 5 essais sont à INN-/SY+ dont 4 N1 et 1 N2. Cette répartition des essais dans les 4 catégories montre des contradictions entre le rendement obtenu et le statut azoté du couvert estimé via l'INN. En effet, certains essais normalement non stressés ont un rendement inférieur à leur objectif (INN+/SY-) tandis que certains essais supposés stressés ont un rendement supérieur au rendement attendu (INN-/SY+). Ces deux catégories (INN+/SY- et INN-/SY+) illustrent la complexité de la gestion d'un stress azoté en condition de plein champs et suggèrent que d'autres facteurs abiotiques interagissent avec la nutrition azotée pour impacter le rendement. Au vu de ces résultats, il a été nécessaire de mettre au point des indicateurs azotés en complément des indicateurs pédoclimatiques afin d'identifier les facteurs qui limitent le rendement.

Enfin, les essais de 2014 à 2015 des réseaux #1 et #2 comprenaient un panel d'association de 173 lignées de colza d'hiver inscrites entre 1950 et 2010 et retraçant l'histoire récente de la sélection du

### Encadré II.1 : Indice de nutrition azotée (INN)

L'indice de nutrition azotée renseigne sur l'état azoté du couvert. La mesure de la biomasse aérienne et de la quantité d'azote dans les parties aériennes (tiges + feuilles) permettent de calculer l'INN qui a été défini par Lemaire et al. (1989) comme suit :

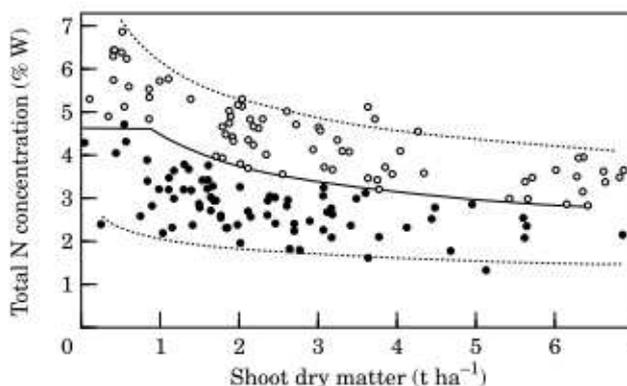
$$INN = N_t / N_c$$

où  $N_t$  est la concentration d'azote mesurée dans les parties aériennes de la plante et  $N_c$  est la concentration d'azote critique calculée pour une biomasse donnée et correspond à une courbe de dilution.

Pour le colza, Colnenne et al. (1998) ont défini la courbe de dilution par la relation suivante :

$$N_c = 4.48 W^{-0.25}$$

où  $W$  est la biomasse des parties aériennes exprimée en t/ha. Les données obtenues sont ensuite projetées sur la courbe de dilution :



Pour la suite on fixe le seuil de 0.9 pour distinguer les essais supposés limitants en azote ( $INN < 0.9$ ) des essais non limitants en azote ( $INN > 0.9$ ).

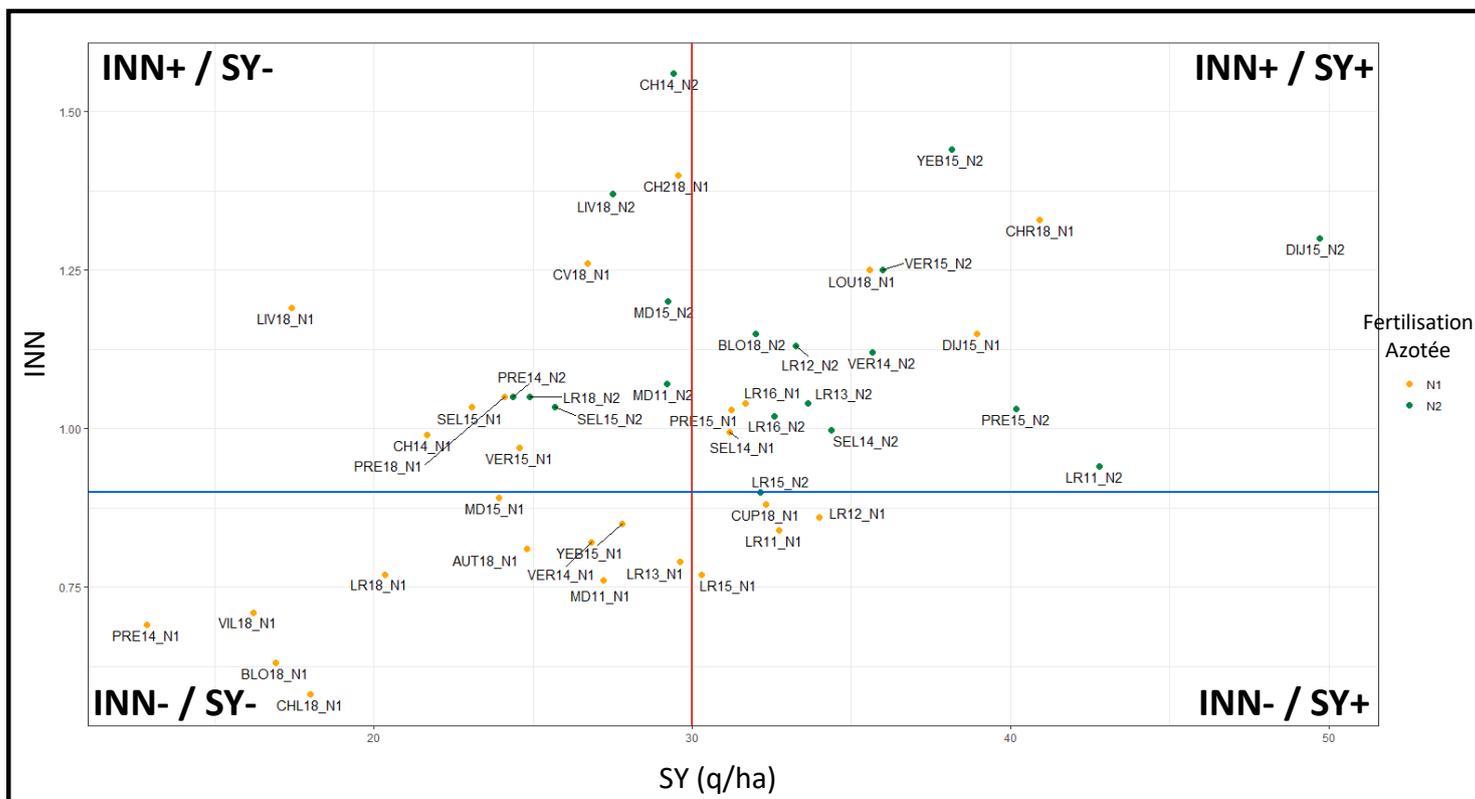


Figure II.3 : Rendements moyens (SY) des essais du réseau #2 en fonction de l'INN

Les rendements sont obtenus à partir des géotypes Aviso et Montego. La ligne verticale rouge fixée à  $SY = 30$  q/ha distingue les essais à fort rendement ( $SY+ > 30$  q/ha) des essais à plus faible rendement ( $SY- < 30$  q/ha). La ligne horizontale bleue fixée à  $INN = 0.9$  distingue les essais non stressés en azote ( $INN+$ ) des essais considérés comme limitants en azote ( $INN-$ ). Les essais ayant reçu un apport azoté non limitant ( $N2$ ) sont indiqués en vert (●) et les essais ayant reçu un apport azoté limitant ( $N1$ ) sont notés en jaune (●).

colza. Ces 22 essais, dont 11 essais N1 et 11 essais N2 constituent le réseau #3 GWAS (Tableau II.1). A l'aide de cette diversité génétique plus large, les essais de ce réseau permettent de quantifier l'impact de l'envirotypage à la fois sur l'effet environnement et sur l'interaction G×E en analysant la décomposition de la variance grâce un modèle linéaire (cf. II.3).



## II.3. Envirotypes Based on Seed Yield Limiting Factors Allow to Tackle G×E Interactions

---

**Agronomy 2019, 9, 798; doi: 10.3390/agronomy9120798**

Erwan Corlouer<sup>1</sup>, Arnaud Gauffreteau<sup>2</sup>, Anne-Sophie Bouchet<sup>1</sup>, Christine Bissuel-Bélaygue<sup>1</sup>, Nathalie Nesi<sup>1</sup>, Anne Laperche<sup>1</sup>

<sup>1</sup>IGEPP, INRA, AGROCAMPUS OUEST, Université de Rennes 1, 35650 Le Rheu, France

<sup>2</sup>UMR Agronomie, INRAE, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France





Article

# Envirotypes Based on Seed Yield Limiting Factors Allow to Tackle $G \times E$ Interactions

Erwan Corlouer <sup>1</sup>, Arnaud Gauffreteau <sup>2</sup>, Anne-Sophie Bouchet <sup>1</sup>, Christine Bissuel-Bélaygue <sup>1</sup>, Nathalie Nesi <sup>1,\*</sup> and Anne Laperche <sup>1</sup>

<sup>1</sup> IGEPP, INRA, AGROCAMPUS OUEST, Université de Rennes 1, 35650 Le Rheu, France; erwan.corlouer@inra.fr (E.C.); asophie.bouchet@gmail.com (A.-S.B.); christine.bissuel@agrocampus-ouest.fr (C.B.-B.); Anne.Laperche@agrocampus-ouest.fr (A.L.)

<sup>2</sup> UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France; arnaud.gauffreteau@inra.fr

\* Correspondence: nathalie.nesi@inra.fr; Tel.: +33-2-23-48-57-07

Received: 14 October 2019; Accepted: 21 November 2019; Published: 23 November 2019



**Abstract:** One challenge in plant breeding is to ensure optimized production under fluctuating environments while reducing the environmental impacts of agriculture. Thus, new rapeseed varieties should be adapted to a wide range of pedoclimatic conditions and constraints. Addressing this issue requires identifying the critical factors limiting production and the genotype by environment ( $G \times E$ ) interaction. Our goal was to characterize the effects of environment and  $G \times E$  interaction on the seed yield of rapeseed grown over a large field network. First, we defined a pedoclimatic indicator set with the ability to highlight the potential limiting factors along the crop cycle by analyzing the yield of two genotypes grown under 20 environments. Out of the 84 pedoclimatic indicators, 10 were identified as limiting after a partial least squares regression analysis. The environments were then clustered into five envirotypes, each characterized by few major limiting factors: low winter temperatures and heat stress during seed filling (1); low solar radiation during seed filling (3); vernalization conditions during winter (4) and high temperatures at flowering (5). A larger genetic diversity was evaluated in a subset of 11 environments to analyze the impact of envirotyping on genotype ranking. Their results were discussed in light of field network management and plant breeding purposes.

**Keywords:** *Brassica napus* L.; seed yield; genotype  $\times$  environment interaction; partial least square regression

## 1. Introduction

Faced with the challenges of adapting agriculture to climate change, as well as more sustainable cultural practices, a major goal is to maintain seed production (quantity and quality) under a wide range of growth conditions, with sometimes highly adverse and unexpected constraints. In this context, plant breeders face a dilemma between designing either new genotypes adapted to diverse pedoclimatic conditions or highly specifically adapted genotypes. This consideration highlights two main issues, first is the definition of field networks for genotype trialing under a large range of limiting factors, and second is the understanding of the  $G \times E$  interactions.

Genotypes are often trialed within field networks for breeding, cultivar registration, or agricultural recommendations. Since the concepts of ecovalence [1] or joint linear regression [2] in the 1960s, several methods have been developed to characterize the genotype reactivity to a given environment using phenotypic observation [3]. These methods are easy to handle and allow quantifying the  $G \times E$  interactions as well as classifying genotypes as reactive or non-reactive. However, they do not provide



any easy way to unravel the interactions due to climate, management practices, or genotype features, nor to access to the biological processes involved.

To get deeper insights into the  $G \times E$  determinants, statistical methods, such as the factorial regression [4] or partial least squares (PLS) regression [5], were developed in the 1980s. They rely on the identification of the main environmental covariates that contribute to the  $G \times E$  effects. A prerequisite of these methods is the definition of *a priori* environmental covariates standing for the potential seed yield limiting factors. According to Van Ittersum et al. [6], the limiting factors can be classified according to their impact on seed yield. The first category corresponds to the “climatic” factors that allow predicting a potential seed yield for a given genotype in a given environment. These factors are related to  $CO_2$  concentration, solar radiation, and soil water holding capacity or temperature. The second category of factors reflects water and nutrient availability that limits the expression of the potential seed yield. This set of environmental factors must be adapted to the considered crop and its agronomic context.

Rapeseed (*Brassica napus* L.) is a major worldwide oil crop with an annual production of around 70 Mt [7]. For winter oilseed rape (WOSR), Bouchet et al. [8] reported that  $G \times E$  interactions could explain up to 10% of the seed yield variation in a field network covering the main crop areas in France. Therefore, there is scope to decipher the main factors that affect seed yield in WOSR in order to improve both the management practices and breeding of cultivars with better resilience towards biotic and/or abiotic constraints. Up until now, most of the studies dedicated to  $G \times E$  in rapeseed were focused on spring oilseed rape (SOSR) accessions grown under a Mediterranean-type climate and showed that water availability and temperature were the two main yield limiting factors under these environmental conditions [9–11]. However, data about other contrasting kinds of climates, such as continental or oceanic types are rather scarce [12].

One main feature of WOSR is its long crop cycle in Western Europe (>10 months from sowing to harvest in France) during which multiple biotic and/or abiotic stresses may occur and impact the seed yield. To define the potential seed yield limiting factors along the crop cycle, it is necessary to divide the whole cycle into different periods, corresponding to homogenous developmental or climatic features, and to list the potential limiting factors over each period. Among all factors defined by Van Ittersum et al. [6], some have already been qualified as limiting for WOSR, such as extreme temperatures at anthesis or vernalization requirement fulfillment during winter [13]. In Northern Europe, solar radiation may also be limiting, and a short photoperiod (less than 9 h) can affect plant development and impact seed yield. Water deficiency, especially at sowing and during seed filling and nitrogen limitation before flowering, has also been described to impact seed yield [14].

The goal of the present work was to cluster environments into different envirotypes in order to describe seed yield variation and explain the  $G \times E$  interaction for WOSR under French pedoclimatic conditions. We based the envirotyping approach on regrouping environments according to their patterns of limiting pedoclimatic factors. We made the hypothesis that inside a single envirotype (cluster of environments presenting the same environmental limiting conditions), the  $G \times E$  interaction was lower than the  $G \times E$  observed at the whole network scale. Therefore, we used the data obtained across a field network of twenty environments (year by location combinations) to develop a four-step strategy: (i) set up a large set of indicators based on pedoclimatic data and phenology of WOSR. (ii) Identify the indicators that most influenced seed yield using PLS regression. (iii) Group the environments into envirotypes based on these limiting factors. (iv) Characterize the impact of the envirotyping on the  $G \times E$  interaction using a large genetic diversity. This pipeline is presented and discussed based on its interest in deciphering the  $G \times E$  interactions to improve plant breeding and field network management.



## 2. Materials and Methods

### 2.1. Field Network and Crop Management Description

Field experiments were conducted in 20 environments (combination of location and year) between 2011 and 2016 in France to cover contrasting pedoclimatic conditions that represented the main areas of rapeseed production and the climate diversity existing in France. The complete field network consisted of 20 environments (Table S1). Each one is defined as the combination of a given location by the year of harvest, as following: Ch14 (Châteauroux in 2014), Dij13 and Dij15 (Dijon in 2013 and 2015 respectively), Liv16 (Liverdy in 2016), LR11, LR12, LR13, LR15, and LR16 (Le Rheu in 2011, 2012, 2013, 2015, and 2016, respectively), Md11, Md14, and Md15 (Mondonville in 2011, 2014, and 2015, respectively), Pre14, Pre15, and Pre16 (Prêmesques in 2014, 2015, and 2016, respectively), Sel14 and Sel15 (Selommes in 2014 and 2015, respectively), Ver14 and Ver15 (Verpillères in 2014 and 2015, respectively) and Yeb15 (Yèbles in 2015). Each individual trial was conducted using classical crop management for WOSR with comprehensive protection against weeds, pests, and pathogens. Optimal Nitrogen (N) fertilization was estimated using the balance sheet method [15,16] for a target yield of 3.5 t ha<sup>-1</sup>. The total amount of required N fertilizer ranged between 40–190 kg N ha<sup>-1</sup> depending on the environment and was provided in one, two, or three applications (Table S1). Plant N status was estimated using the Nitrogen Nutrition Index (NNI) [17] at the stage where flower buds were still enclosed by leaves (BBCH50) [18,19], with a minimum delay of two weeks after the latest N supply. Each trial was designed as a randomized complete block design with two to four repetitions (according to the environment). Individual plot surface ranged from 6.75 m<sup>2</sup> to 14 m<sup>2</sup>.

### 2.2. Plant Material and Seed Yield Assessment

Two probe genotypes, namely Aviso and Montego, were scored for seed yield over the 20 environments of the network. These genotypes were contrasted for earliness (mean difference of 40 growing degree days (GDD) ± 18) at flowering, GDD calculated as Gabrielle et al. [20] and height (mean difference of 13 cm ± 9 cm), with Montego being the earliest and the smallest one. A diversity set of 127 WOSR accessions (hereafter referred to as DS127) released from 1959 to 2007 (Table S2) was scored for seed yield over a subset of 11 environments. Seed yield (SY) was defined as the weight (t) of seeds harvested per ha considering moisture and impurity levels at 0% each. As a first evaluation of the G × E, the ecovalence [1] was calculated for SY for each genotype (Equation (1)). It corresponds to the contribution of each genotype to the G × E. The ecovalence gives information about the stability of a genotype across environments. A high ecovalence means that the genotype is not stable across environments.

$$W_g^2 = \sum_{e=1}^E (Y_{ge} - \bar{Y}_g - \bar{Y}_e + \bar{Y}_{..})^2, \quad (1)$$

where  $W_g^2$  is the ecovalence for the genotype  $g$ ,  $Y_{ge}$  is the SY value for genotype  $g$  in environment  $e$ ,  $\bar{Y}_g$  is the mean SY for the genotype  $g$  across all environments,  $\bar{Y}_e$  is the mean SY for the environment  $e$  across all genotypes, and  $\bar{Y}_{..}$  is the general mean. The DS127 displayed genetic variability for ecovalence, as shown in Figure S1, reflecting a genetic variability of the response to environmental conditions.

### 2.3. Key Periods of the Winter Rapeseed Crop Cycle

The rapeseed crop cycle was divided into seven consecutive and non-overlapping periods based on climatic data and plant phenology (Figure S2). The first period was the fall (F) that covered from the sowing date to the beginning of the climatic winter (CW) defined as the second period. The CW started or ended when 3 consecutive days with daily air temperature were recorded under or above 5 °C, respectively (adapted for Hebing [21]). The third period was the bolting period (B) that spanned from the end of CW to the beginning of flowering of Montego (BBCH60) [18,19]. The fourth period was the flowering (FLO) that lasted three weeks from the Montego BBCH60 stage onwards. The last



three periods were defined on the base of thermal dates calculated from flowering (cumulative thermal time using a base temperature of 0 °C) according to Leterme [22] and Julien et al. [23]. The P300 period lasted 300 GDD after FLO and is related to the seed number fixation. The P600 period started at the end of P300 and lasted 300 GDD. It is related to reserve allocation to the pod growth. Finally, the P1000 period started at the end of P600 and lasted 400 GDD. During P1000, reserves were primarily allocated to seed growth. The vegetative part of the cycle included the F, CW, and B periods, while the FLO, P300, P600, and P1000 defined the reproductive part of the cycle.

#### 2.4. Indicators Used for Environmental Description

##### 2.4.1. Descriptor Definition: Raw Climatic Data and Soil Water Status Evaluation

Specialized climatic data corresponding to mean daily rainfalls ( $R$ ), evapotranspiration of Penmann ( $ETP$ ), global solar radiation ( $SR$ ), mean air temperature ( $T_{mean}$ ), maximum air temperature ( $T_{max_i}$ ), and minimum air temperature ( $T_{min}$ ), that were extracted from the Meteo France database [24].

Water status for each environment was quantified by daily water soil content ( $WSC$ ), calculated from the water balance described in Equation (2). The maximal water soil content was estimated for each location based on soil physical characteristics of two soil layers, 0–30 cm and 30–100 cm, according to Bruand et al. [25]. The soil depth was set up to 100 cm for all locations according to oilseed rape root distribution [26] and French soils mean depth [27].

$$\begin{cases} WSC_i = WSC_{i-1} + (R_i - Kc_i \times ETP_i) - RO_i \\ WSC_i \geq 0, \end{cases} \quad (2)$$

where  $WSC_i$  is the water soil content at day  $i$ ,  $R_i$  is the rainfall of day  $i$ ,  $Kc_i$  is the crop coefficient (with kinetic through adapted from Allen et al. [28]; Figure S2) calculated at day  $i$ ,  $ETP_i$  is the potential evapotranspiration, and  $RO_i$  is the runoff. Runoff at day  $i$  is estimated as the difference between  $[WSC_{i-1} + (R_i - Kc_i \times ETP_i)]$  and  $WSC\_MAX$  when  $[WSC_{i-1} + (R_i - Kc_i \times ETP_i)] > WSC\_MAX$  and is 0 when  $[WSC_{i-1} + (R_i - Kc_i \times ETP_i)] \leq WSC\_MAX$ .

For the initialization of the simulations of water balance, two extreme scenarios were tested in all environments (i)  $WSC = WSC\_MAX$  on 1 January before sowing without any previous crop ( $Kc$  limited to soil evaporation coefficient) and (ii)  $WSC = 0$  at sowing. The two simulations converged rapidly and showed that the  $WSC$  was very close to 0 mm in August before sowing in almost all environments and was at field capacity in winter, as reported by Weymann et al. [29]. Then,  $WSC$  was initialized to 0 mm on August, 1st before each sowing. Based on the previous data, 15 descriptors were identified. Eight descriptors were based on direct calculations of the raw data (mean, range, and sum) and are summarized in Table 1a. They corresponded to the minimal temperature (TMIN), the maximal temperature (TMAX), the mean temperature (TMN) over a period and the length of a period in GDD (LGDD) for periods F, CW, B, and FLO. The water soil content maximal capacity ( $WSC\_MAX$ ) and the mean water soil content ( $WSC\_MN$ ). The sum of the radiations (SSR) and the photothermal quotient (QPT) only for the FLO period calculated as the ratio between SSR and LGDD [30,31]. Seven other descriptors consisted of the quantification of different stress durations by counting the number of days when the considered raw climatic data was under or above a given threshold. These descriptors are summarized in Table 1b and correspond to the number of days with  $T_{max} > 25$  °C (high temperature—HT), the number of days with  $0$  °C  $< T_{min} < 5$  °C (low temperature—LT), the number of days with  $T_{min} < 0$  °C (frost—FR). The number of days when  $WSC < 1/3 WSC\_MAX$  (water stress—WS), the number of days when  $WSC = 0$  mm (water deficiency—WD). The number of days with  $SR < 900$  J cm<sup>-2</sup> (lack of solar radiation—LSR) and the number of days with  $T_{mean} < 5$  °C, and a day length below 9 h (optimal vernalization treatment [13]—VERN\_OPT).



**Table 1.** (a) List of pedoclimatic descriptors: soil data status and descriptors calculated using raw climatic data; (b) List of pedoclimatic descriptors corresponding to stress duration.

(a)			
Descriptor	Description	Category	Unit
LGDD*	Length of a crop period expressed in growing degree-days. Sum of the daily $T_{mean}$ for the period (base temperature 0 °C)	Temperature	°C.d
TMN	Mean of the daily $T_{mean}$ recorded over a given crop period	Temperature	°C
TMIN	Minimal temperature over a given crop period	Temperature	°C
TMAX	Maximal temperature over a given crop period	Temperature	°C
WSC_MAX*	Maximal water soil content for a given environment	Water	mm
WSC_MN	Mean of the water soil content over a given crop period (in proportion of WSC_MAX)	Water	%
SSR	Sum of the daily solar radiation over a given crop period	Solar	J cm <sup>-2</sup>
QPT*	Photothermal quotient of SSR by LGDD (SSR/LGDD)	Solar	MJ m <sup>-2</sup> °C <sup>-1</sup>
(b)			
Descriptor	Description	Category	
HT	Number of days with high temperature ( $T_{max} > 25$ °C)	Temperature	
LT	Number of days with low temperature (0 °C < $T_{min}$ < 5 °C)	Temperature	
FR	Number of freezing days ( $T_{min} < 0$ °C)	Temperature	
WS	Number of days when WSC < 1/3 WSC_MAX	Water	
WD	Number of days when WSC = 0 mm	Water	
LSR	Number of days with a lack of solar radiation ( $SR < 900$ J cm <sup>-2</sup> )	Solar	
VERN_OPT*	Number of days with an optimal vernalization treatment ( $T_{mean} < 5$ °C and day length < 9 h)	Plant	

Each descriptor can be attributed to all crop periods (descriptor × crop period = indicator) excepted for descriptors with a star (\*). The length of a crop period in growing degree days (LGDD) is only calculated for Fall, Climatic Winter, Bolting, and Flowering. The photothermal quotient (QPT) is only calculated for Flowering.  $T_{min}$  and  $T_{max}$  represent daily minimal and maximal temperature, respectively, whereas TMIN and TMAX represent minimal and maximal temperature for a given period.

#### 2.4.2. Set up of Informative Indicators

Combining the different descriptors with the crop periods led to the definition of 84 pedoclimatic indicators falling into four main categories, namely “temperature”, “water”, “solar” and “plant”. The pedoclimatic indicators that did not vary along the crop cycle (11 in this study) or varied in less than 3 environments (5 in this study) were considered as non-informative and eventually removed from our dataset (Figure 1). Among these, 9 were of the temperature category and 7 of the water category. The Pearson correlation coefficients were calculated between each of the 68 remaining pedoclimatic indicators. Environments were clustered based on the pedoclimatic indicator categories using the Ward method [32] to describe the network.

#### 2.5. Statistical Analyses

All statistical analyses were performed with R software version 3.5.1 [33].



### 2.5.1. Partial Least Squares (PLS) Regression

To identify the pedoclimatic indicators that limited seed yield, a univariate partial least squares (PLS) regression [34] was performed on the mean SY value calculated as the mean between Aviso and Montego seed yields across the 20 environments of the network. PLS regression was preferred to linear regression because many indicators were considered and some of them were correlated. The univariate PLS regression model was based on the construction of latent variable ( $T$ ) as linear combinations of the  $X$  variables (the pedoclimatic indicators in this study) so that  $cov(T, Y)$  is maximal and then regressing  $Y$  (the mean seed yield of Aviso and Montego) on those latent variables [35,36]. The number of latent variables to consider was based on the  $Q^2$  indicator [37]; one new component was considered if its  $Q^2$  value was over 0.0975. Finally, the model performance was estimated by its  $Q^2_{cum}$  (ref) value assessed on all its components. Gauchi and Chagnon [38] showed that a selection among the indicators under study could improve the PLS regression model. They proposed methods to perform this selection of variables as the Backward- $Q^2_{cum}$  method (BQ method) [38]. The first step of this method was to perform a PLS regression on all pedoclimatic indicators and then to select the set of indicators that explain SY variability using a backward selection: at each step, the pedoclimatic indicator presenting the smallest regression coefficient in absolute value was discarded and a new PLS regression was performed. The process was repeated  $n-1$  times (with  $n$  the number of pedoclimatic indicators considered). The best PLS regression model corresponded to the model with the highest  $Q^2_{cum}$  value. When two models resulted in the same  $Q^2_{cum}$  value, the one with the smallest number of indicators was chosen. PLS regressions runs were carried out using the `plsreg1` function from the package `plsdepot` [39].

To validate the indicators set identified by the PLS regression, a leave-one-out procedure was used on the 20 environments. Twenty PLS regressions were performed using the same method, each on 19 environments (one environment was removed at each PLS regression). For each of the indicators selected by the PLS regression on the whole network, we calculated a confidence index as the number of times where the pedoclimatic indicator was selected among the 20 “leave-one-out” PLS regressions divided by the total number of environments. If the value of the confidence index was superior to 0.5, the pedoclimatic indicator was considered as consistent.

### 2.5.2. Envirotyping Based on the PLS Regression Results

Using the best PLS regression model, each environment was characterized by its coordinates on the different PLS axes multiplied by the impact of those axes on  $Y$ . Based on these coordinates, Euclidean distances between environments were calculated and used to carry out a clustering using the Ward method [32]. To determine the number of envirotypes (clusters of environments), the inertia gain was observed, and the Krzanowski and Lai index [40] was calculated for a number of clusters between 2 and 15. The index maximal value indicates the optimal number of envirotypes. Finally, each envirotype was described using the function `catdes` of the package `FactoMineR` [41] by the calculation of a test value as defined by Husson et al. [42]. Briefly, the test-value is the normalized variation between the mean value of individuals belonging to a given envirotype and the general mean.

### 2.5.3. Test of the Environment and Genotype by Environment Effects Using Linear Models

Each linear model was run using the function `lm` of  $R$ .

A first linear fixed model (Equation (3)) was fitted on the data of DS127 (11 environments) to test the effects of the environments and the  $G \times E$  interaction across the network:

$$Y_{ijk} = \mu + G_i + E_j + G_i \times E_j + R_{k(j)} + \varepsilon_{ijk}, \quad (3)$$

where  $Y_{ijk}$  is the seed yield (SY) of genotype  $i$  in environment  $j$  for the replicate  $k$ ,  $\mu$  is the population mean,  $G_i$  stands for the effect of genotype  $i$ ,  $E_j$  for the effect of environment  $j$ ,  $R_k$  for the effect of replicate  $k$  nested in the environment  $j$ ,  $G_i \times E_j$  for the effect of interaction between genotype  $i$ , and environment  $j$  and  $\varepsilon_{ijk}$  is the residual.



A second linear model (Equation (4)) was fitted to include an “envirotypage” term and the corresponding interaction effects.

$$Y_{ijkl} = \mu + G_i + C_l + C_{l(j)} + G_i \times C_l + G_i \times C_{l(j)} + C_{l(j \times k)} + \varepsilon_{ijkl}, \quad (4)$$

where  $Y_{ijkl}$  is the SY observed for the genotype  $i$ , in envirotypage  $l$ , in environment  $j$ , and in replicate  $k$ ,  $\mu$  is the population mean,  $G_i$  stands for the effect of genotype  $i$ ,  $C_l$  for the effect of envirotypage  $l$ ,  $E_j$  for the effect of environment  $j$  nested in envirotypage  $l$ .  $R_k$  stands for the effect of replicate  $k$  nested in environment  $j$  also nested in envirotypage  $l$ .  $G_i \times C_l$  the effect of interaction between genotype  $i$  and envirotypage  $l$ ,  $G_i \times C_l(E_j)$  the effect of interaction between genotype  $i$  and environment  $j$  nested in envirotypage  $l$ .  $\varepsilon_{ijkl}$  is the residual.

### 3. Results

#### 3.1. Description of the Field Network and Pedoclimatic Indicators

The mean seed yield of the two probe genotypes Aviso and Montego reached  $3.4 \text{ t ha}^{-1}$  (standard deviation of 0.618) over the whole network, ranging from  $2.4 \text{ t ha}^{-1}$  to  $4.7 \text{ t ha}^{-1}$  depending on the environment. The targeted yield of  $3\text{--}3.5 \text{ t ha}^{-1}$  was reached in most environments (Table S1). In addition, the mean NNI value across the network was 1.18 (standard deviation of 0.25), confirming that the network was not very impacted by a nitrogen stress as the NNI value was higher than 0.9 [17]. No biotic stresses were reported in the network. Taken together, these data suggested that the SY was primarily limited by pedoclimatic conditions.

According to the four classifications, performed for each pedoclimatic categories, environments can be assigned to three climatic categories (Figure 1). The first category gathered the environments Dij13, Dij15, Pre15, Ver15, Sel15, Yeb15, Md11, and Md15, and can be qualified as “continental”. It is characterized by cold winters (high FR\_CW and high LT\_CW) and high temperature during spring and summers (high TMAX\_FLO/P300/P600/P1000 and high HT\_FLO/P300/P600/P1000). No solar deficit was recorded for these environments and five out of eight environments showed water stress during the reproductive phase (Dij15, Pre15, Yeb15, Md11, and Md15). This “continental” category gathered all environments of the year 2015 (except LR15), and all the trials were carried out in location “Dijon” in Eastern France. The second category consisted of the three environments LR12, LR13, and LR16. The environments of this category are characterized by a short duration of the flowering phase when expressed in growing degree-days (low LGDD\_FLO), low temperatures at flowering (high LT\_FLO and low TMN\_FLO), solar deficiency during falls and the reproductive phase (LSR\_F/P300/P600), and no water stress during the whole crop cycle. This category is specific to the LR location in Brittany (Western France), an area known for its oceanic climate with mild and rainy winters and the absence of extreme temperatures. The last group gathered environments Md14, Sel14, Pre14, Ver14, Liv16, Pre16, LR11, Ch14, and LR15. This group can be qualified as “modified oceanic” climate and is characterized by warm falls and winters (high TMAX\_F and low FR\_CW) and cool springs and summers. At the network scale, specific climatic constraints were recorded independently of the climate category: water stress during flowering (WSC\_MN\_FLO) were recorded for Md11, Dij15, Sel14, and LR12, and optimal vernalization conditions were not fulfilled for LR12, LR16, Md14, Md15, Md11, and Pre14. The calculation of the Pearson’s correlation coefficient for the 68 pedoclimatic indicators (Table S3) revealed that 25% of the correlation coefficients were significant at  $\alpha = 0.05$  and corresponded to correlation coefficient absolute values above 0.5 and that 5% of the correlations were highly significant ( $p$ -value  $< 0.001$ ). This 5% of correlations are highlighted in Table S3.

#### 3.2. Identification of the Critical Indicators for Seed Yield

Following the PLS regression analysis, the BQ method was applied, resulting in the selection of a model based on three components and 10 pedoclimatic indicators (Table 2). The confidence

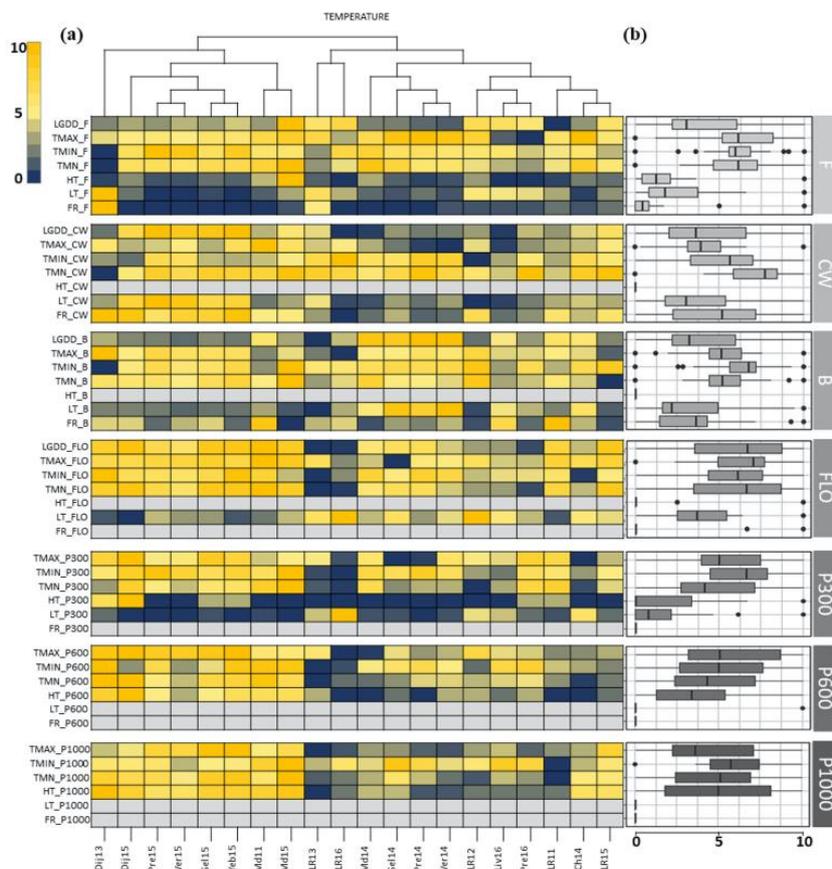


index calculated by PLS across permutations showed that these 10 indicators were identified in at least 60% of the permutations. The observed seed yield and the predicted seed yield were highly correlated ( $R^2 = 0.96$ ,  $RMSE = 1.3$ ). Selected indicators presented highly significant correlations with other indicators excepted for TMAX\_FLO (Table 2).

**Table 2.** The 10 limiting factors identified by partial least squares regression.

Pedoclimatic Indicator	Confidence Index	Correlated Indicators
VERN_OPT	1	LT_CW
TMAX_FLO	0.95	-
HT_P600	0.9	FR_CW (0.72); TMAX_P600 (0.83); TMN_P600 (0.87); HT_P1000 (0.70)
SSR_P600	0.85	LSR_P600 (-0.70)
TMN_CW	0.85	TMIN_B (0.73)
HT_P300	0.8	TMAX_P300 (0.70)
LSR_FLO	0.8	SSR_FLO (-0.70)
WS_P1000	0.75	WS_P600 (0.77); WSC_MN_P600 (-0.68); WSC_MN_P1000 (-0.70)
TMIN_P300	0.65	TMN_FLO (0.70); LT_FLO (-0.68); LGDD_FLO (0.70); TMN_P300 (0.71); LT_P300 (-0.88)
TMN_P1000	0.6	TMN_FLO (0.70); LGDD_FLO (0.70); TMAX_P1000 (0.76); HT_P1000 (0.89); LSR_P1000 (-0.76)

Confidence index corresponds to the percentage of permutations where the indicator was identified by PLS regression. Correlated indicators represent additional indicators that were significantly correlated with the ones identified by the PLS regression. The values given into parentheses represent the Pearson’s correlation coefficient (data given in Table S3).



**Figure 1.** Cont.



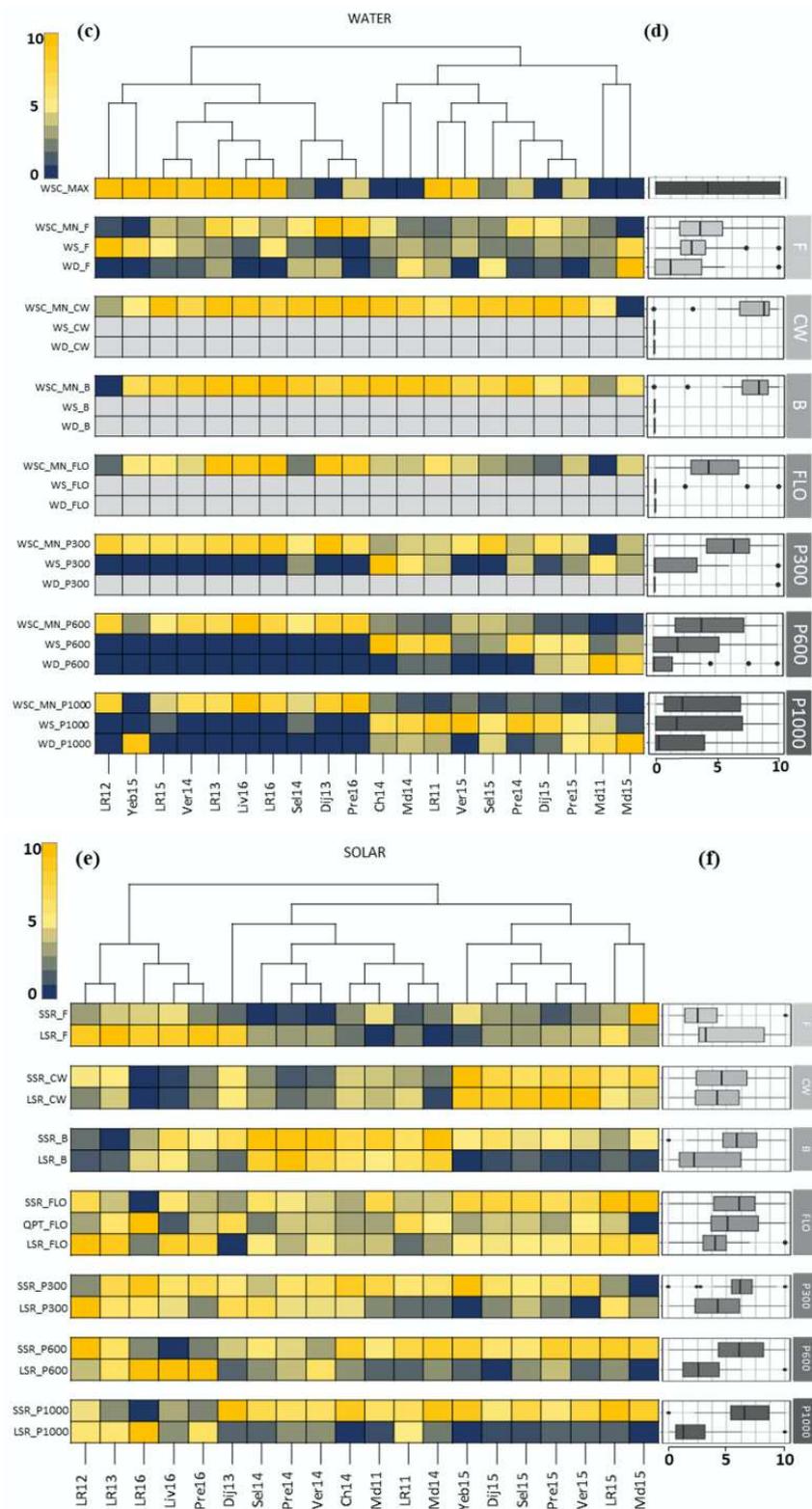
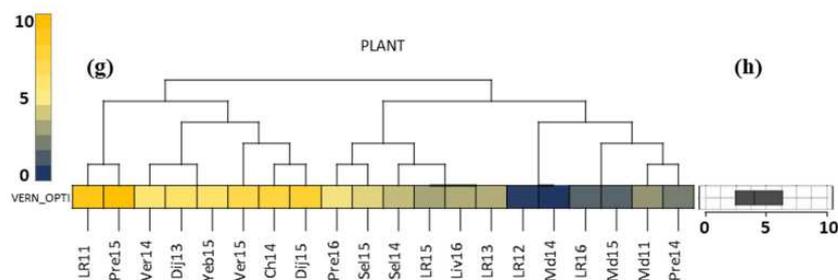


Figure 1. Cont.



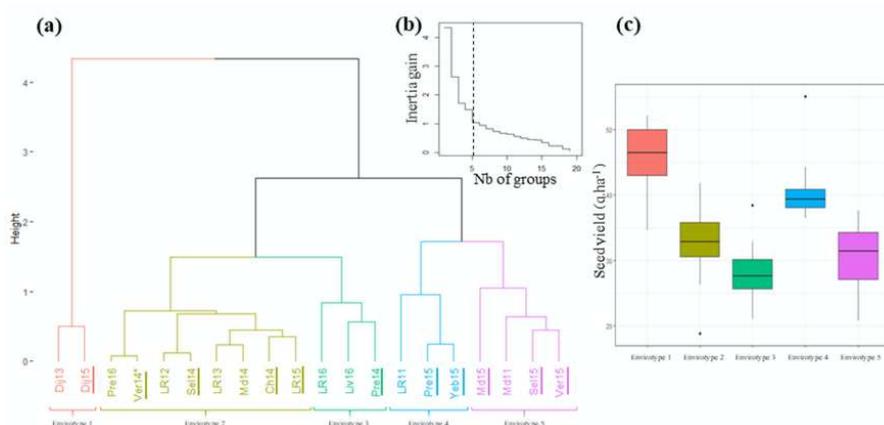


**Figure 1.** Description of the field network by the pedoclimatic indicators and variations of these indicators within the field network. Heatmaps are presented on the left (a,c,e,g) and boxplots on the right (b,d,f,h) for temperature indicators (a,b); water indicators (c,d); solar indicators, (e,f) or plant indicators (g,h). To run comparisons between environments, each indicator was scaled from 0 (dark blue) to 10 (yellow) on the heatmaps. Non-informative indicators are represented in grey and were removed for further studies. Pedoclimatic indicators are grouped by period as following: Fall (F), climatic winter (CW), Bolting (B), Flowering (FLO), Seed number fixation period (P300), Priority allocation to envelope phase (P600), and priority allocation to seeds phase (P1000). Information about environments is given in Table S1.

All categories of descriptors were represented in the 10 indicators revealed by the PLS regression with TMN\_CW, TMAX\_FLO, TMIN\_P300, HT\_P300, HT\_P600, and TMN\_P100 for the temperature category, WS\_P1000 for the water category, LSR\_FLO and SSR\_P600 for the solar category, and finally, VERN\_OPT. Except for VERN\_OPT and TMN\_CW, all selected indicators that explained seed yield variation were post-flowering indicators.

### 3.3. Definition of the Envirotypes

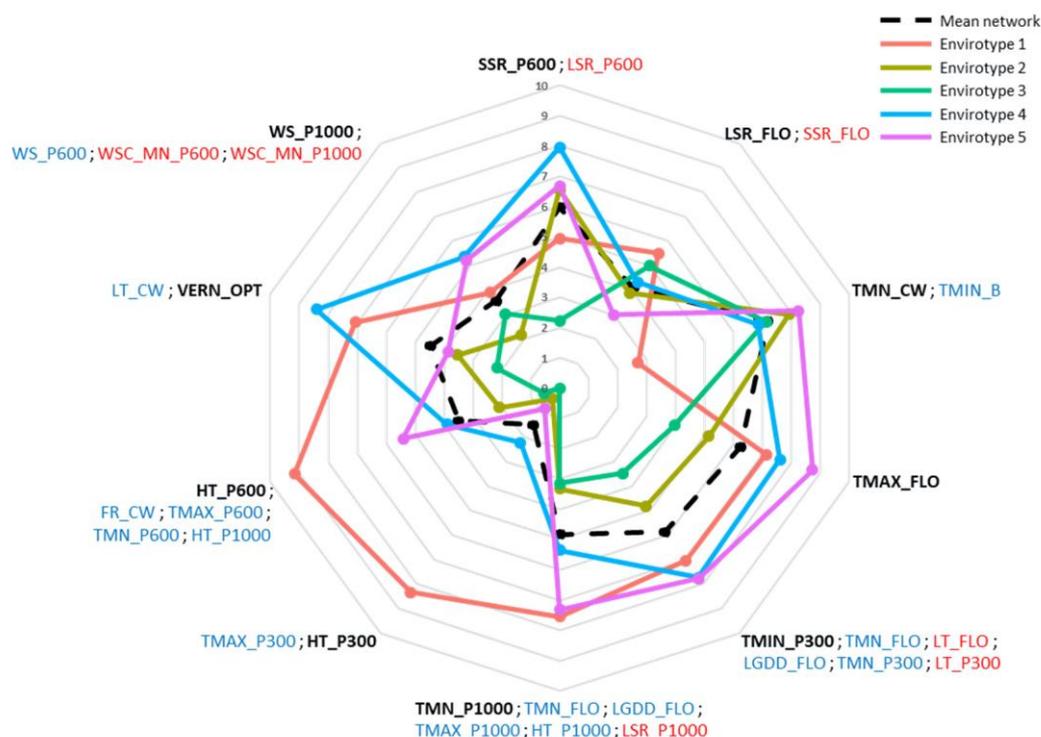
Based on the results of the PLS regression and of clustering of the environments, five envirotypes were identified regarding the inertia gain and the result of the Krzanowski and Lai index [40] (Figure 2a,b). The environments Dij13 and Dij15 constituted the envirotype 1. Pre16, Ver14, LR12, Sel14, LR13, Md14, Ch14, and LR15 constituted the envirotype 2. LR16, Liv16, and Pre14 constituted the envirotype 3. LR11, Pre15, and Yeb15 constituted the envirotype 4 and Md15, Md11, Sel15, and Ver15 constituted the envirotype 5. Noticeably, the different envirotypes did not correspond to a year or location specific classification except for the envirotype 1. The seed yield distribution of Aviso and Montego in each envirotype revealed high-yielding envirotypes (e.g., envirotype 1) and low-yielding envirotypes (e.g., envirotype 3) (Figure 2c).



**Figure 2.** Envirotype definition: (a) dendrogram tree of the 20 environments of the field network based on the PLS regression results. The 11 environments of the DS127 are underlined. (b) Inertia gain for number of groups between 2 to 19. (c) Boxplots showing the distribution of the seed yield ( $q \text{ ha}^{-1}$ ) of Aviso and Montego per envirotype.



The envirotypes were characterized according to their pedoclimatic indicators pattern (Figure 3). Envirotype 1 had a higher mean HT\_P300 and a higher mean HT\_P600 than the global network but a lower mean TMN\_CW. Envirotype 2 was representative of the global network. Envirotype 3 was characterized by a lower mean SSR\_P600. Envirotype 4 was characterized by a more important VERN\_OPT. Last, envirotype 5 had a higher mean TMAX\_FLO than the global network.



**Figure 3.** Characterization of the envirotypes using the 10 indicators selected by the PLS regression. The pedoclimatic indicators are written in black and the correlated indicators in blue (positive correlation) or in red (negative correlation). TMN\_CW (mean temperature during the climatic winter period), TMAX\_FLO (maximal temperature during the flowering period), TMIN\_P300 (minimal temperature during the seed number fixation period), TMN\_P1000 (mean temperature during the seed filling period), HT\_P300 (number of days of high temperature during the seed number fixation period), HT\_P600 (number of days of high temperature during the “allocation to the pod” period), VERN\_OPT (number of days with an optimal vernalization treatment), WS\_P1000 (number of days presenting a water stress during the seed filling period), LSR\_FLO (number of days with a lack of solar radiation during the flowering period). The mean network is represented by a black dash line.

### 3.4. Evaluation and Decomposition of the $G \times E$ Interaction at the Network Level

#### 3.4.1. Effect of the Envirotyping on the $G \times E$ Decomposition

The DS127 was trialed in 11 out of the 20 environments represented in at least one envirotype (Figure 2a). The distribution of seed yield for the DS127 per envirotype displayed the same pattern as shown for Aviso and Montego (Figure S3). The multi-local variance analysis, using the Equation (3) on the DS127 data (Table 3) revealed a high environmental effect with 51.1% of the variation and a high genotype effect and  $G \times E$  interaction reaching 28.9% and 11.7% of the global variation, respectively. When considering the effect of the envirotyping on seed yield, we observed that the envirotype was significant and explained 46.2% of the total variation, and the environmental effect within each envirotype explained 4.9% (Table 4). Clustering the environments into five envirotypes contributed to explaining 90% of the environmental effect observed at the field network scale. The initial  $G \times E$  interaction (Table 3) was split into the  $G \times C$  (42.6%) and  $G \times C \times E$  effect (57.4%) (Table 4).



Finally, within an envirotype, the environmental and  $G \times E$  effects were reduced when compared to the global network.

**Table 3.** Results and corresponding variance partition of the linear fixed model (3) on DS127 data (127 genotypes trialed in 11 environments of the network).

	G	E	G × E	E × R	Residuals
Sum Sq	70821	125131	28620	6069	14190
%Sum Sq	28.9	51.1	11.7	2.5	5.8
Pvalue	***	***	***	***	

Significance of the genotype (G), environment (E), replicate (E × R), and their interaction (G × E) were assessed using model (3). Sum Sq: sum square of the considered effect, %Sum Sq: proportion of the total variation explained for the considered effect, *p* value: results of the F-test. (\*\*\*,  $p < 0.001$ ).

**Table 4.** Results and corresponding variance partition on the linear fixed model (4) on DS127 data.

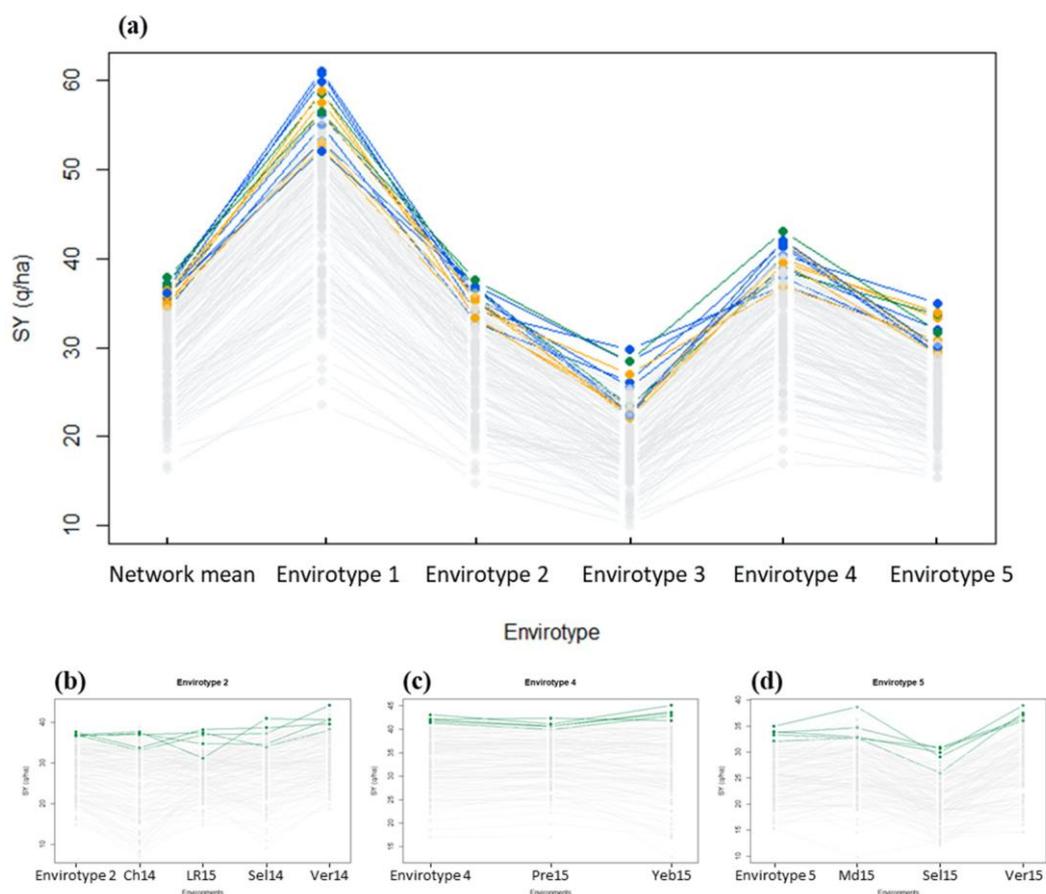
	G	C	C × E	G × C	G × C × E	C × E × R	Residuals
Sum Sq	70821	113195	11936	12195	16425	6069	14190
%Sum Sq	28.9	46.2	4.9	5.0	6.7	2.5	5.8
Pvalue	***	***	***	***	***	***	

Significance of the genotype (G), environment (E), envirotype (C), replicate (C × E × R) and their interactions (C × E, G × C, G × C × E) were assessed using model (4). Sum Sq: sum square of the considered effect, %Sum Sq: proportion of the total variation explained for the considered effect, *p* value: results of the F-test. (\*\*\*,  $p < 0.001$ ).

#### 3.4.2. Ranking of the Genotypes Per Envirotype

The evaluation of the genotype ranking per envirotype also illustrates the impact of the envirotyping on deciphering the  $G \times E$  interaction (Figure 4a). Lists of the top five genotypes of each envirotype were compared. Taken together, these genotypes were always performing in the first quarter list, but the top five genotypes differed between envirotypes. Indeed, only two genotypes (Ecrin and SW Gospel) were identified in the top list of three envirotypes, including envirotype 1, 2, and 5 for Ecrin and envirotype 2, 3, and 4 for SW Gospel. Seven genotypes were common to two envirotypes: Adriana (envirotypes 1 and 5), Capvert (1 and 3), Astrid (1 and 5), Courage (2 and 3), Salomont (2 and 4), Alesi (2 and 4), and Kadore (3 and 4). Finally, five genotypes were envirotype-specific as Remy (1), Navajo (3), Pacific (4) Lewis (5), and Aviso (5). These rankings changed between envirotypes, highlighting qualitative  $G \times E$  at the network scale. The same approach was carried out for envirotypes 2, 4, and 5 that included several environments, and the top list of genotypes per environment of a given envirotype was compared to the global ranking at the envirotype scale (Figure 4b–d). For instance, one genotype (SW Gospel) was identified as one of the best five genotypes of all the constituting environments of envirotype 4. Similar results were observed for genotype Astrid within envirotype 5.





**Figure 4.** Ranking of the genotypes: (a) comparison of the top 5 genotypes in each envirotype across the whole network; network mean represented the mean ranking of the genotypes. Genotypes identified as one of the top 5 genotypes in three envirotypes are represented in green (●); the one identified in 2 envirotypes out of 5 are represented in blue (●) and the one identified in a single envirotype are represented in orange (●). (b,c,d) Ranking of the top 5 genotypes in green (●) for envirotype 2, 4, and 5, respectively, in each environment of the corresponding envirotype.

#### 4. Discussion

The main goal of this study was to characterize the environmental effect and the  $G \times E$  interaction on winter oilseed rapeseed yield variation considering climatic, pedological, and plant phenological features. Using this framework, 68 informative pedoclimatic indicators were defined. A PLS regression coupled with a decision rule (Backward- $Q^2_{cum}$  method) allowed to tag 10 indicators linked to the vernalization and reproductive phase as limiting for seed yield. These 10 indicators were used to cluster the environments of the network into five envirotypes. The envirotyping allowed catching a major part of the environmental effect as well as a smaller part of the  $G \times E$  interaction. These results open new directions in quantitative genetics and genomics selection.

##### 4.1. When Used in A Multi-Constraining Network the PLS Regression Selected Key Indicators That Were Critical for Seed Yield

Under French climatic conditions, a wide range of stresses can limit seed yield, according to the duration of the crop cycle as well as the diversity of climates occurring in France (oceanic, Mediterranean or pseudo-continental climates). To cover most stresses that potentially occurred along the crop cycle, 84 pedoclimatic indicators were defined and split into four categories (thermic, water, solar, and plant) for all developmental stages. This large set was reduced to 68 after consideration of the variability of each single indicator within the network. Other studies have focused on the impact of abiotic



stresses on seed yield of *Brassica napus*. Thermic and water stresses were identified as limiting factors for spring oilseed rape grown under Mediterranean-type climates [9–11], and the time of flowering depends on the photoperiod and temperature [13]. However, these studies did not consider any other abiotic constraints. In this study, an exhaustive environmental screening of potential pedoclimatic indicators affecting seed yield was performed and produced a high number of indicators. Because this number was much higher than the number of environments and because correlations existed between indicators, classical methods to identify limiting factors, such as the factorial regression, cannot be carried out without a preselection of the most important variables. PLS regression permits using all variables, including correlated variables, without selection, and pairing with a decision rule selects indicators to improve the regression model. From the 68 indicators, 10 were identified by the PLS regression as limiting factors and all four categories were represented.

#### 4.2. Critical Factors for Seed Yield in WOSR Were Mostly Related to Heat Stress, Radiation Deficit, and Water Shortage during Vernalization and Reproductive Periods

The 10 pedoclimatic indicators identified for seed yield were related to the vernalization period (TMN\_CW and VERN\_OPT) and the reproductive phase (TMAX\_FLO, LSR\_FLO, TMIN\_P300, HT\_P300, HT\_P600, SSR\_P600, TMN\_P1000, and WS\_P1000). Most of the correlated indicators showed the same processes excepted for the correlation between HT\_P600 and FR\_CW (Table S3). Vernalization controls flowering time and is highly impacted by the photoperiod and the duration at low temperatures (<5 °C for *Brassica napus*) [43]. The non-fulfillment of optimal vernalization conditions impacts seed yield, leading to a delay or an absence of flowering in WOSR [44]. In *Arabidopsis thaliana*, the two genes FRIGIDA (FRI) and FLOWERING LOCUS C (FLC) are known to delay flowering, but their effect can be suppressed by a vernalization treatment. Orthologs have been identified in *Brassica napus* and their impact on flowering time was confirmed [45,46]. With an earlier flowering, the environmental conditions at anthesis could change and affect the seed number elaboration. An earlier flowering also decreases the duration of the vegetative phase resulting in a lower amount of accumulated assimilates to be remobilized to the grain afterward. During the reproductive phase including the flowering period, the seed number fixation period (P300), the period when reserves are allocated to pod growth (P600) and the period when reserves primarily allocated to the pods are used for seed growth (P1000), five thermic indicators were identified as limiting. Heat stress at flowering is known to induce a seed yield reduction in *Brassica napus* by affecting the flower fertility: size and shape of the floral organs [47–49]. Heat stress can also affect the seed number, the number of seeds per pod, and the pod number [47,48], depending on the period affected by the stress. For instance, during the P1000 period, high temperature impact on seed yield was quantified as a loss of 0.4 t ha<sup>-1</sup> for an increase of 3 °C of the mean temperature for canola [50]. Two solar indicators were identified during flowering and P600 periods. Radiation stresses are sometimes confounded with heat stress in the literature, and therefore less described [48]. In our study, these two solar indicators were not correlated with a thermic indicator; we were able to distinguish radiation effects from heat stress and to highlight both effects independently. This distinction is important because radiation stress by itself can lead to seed yield variation, especially during flowering and seed filling, as shown by Baux et al. [31]. Radiation stress during the reproductive phase could affect the pod autotrophy by affecting the pod chlorophyll content and their photosynthetic activity. Finally, one water indicator was identified as limiting during the end of the crop cycle. In France, water shortage is not considered as a major limiting factor for rapeseed, but it can lead to seed yield loss, especially when the stress appears between the BBCH stages 65 and 73, affecting the number of pods per plant [14].

#### 4.3. An Approach to Capture the Components of the Environmental Effect

The five envirotypes identified gathered from two to eight environments each. The environments were not grouped according to their climatic features or the climatic year. Indeed, this clustering was compared to a clustering performed using a PCA carried out using the whole set of pedoclimatic



indicators. The results showed that environments were clustered mainly according to the year of the experiment and secondly according to their location, and no differences were observed for mean seed yield between groups (data not shown). However, the envirotyping based on the PLS regression was reliable as it explained 90% of the environmental variation observed across 11 environments, thus drastically reducing the remaining environmental effect within each envirotype. These results confirmed the important effect of environment on seed yield and the ability of the method to identify *a posteriori* the pedoclimatic components explaining yield instability. This envirotyping allowed us to define a typology of environments. These results can be helpful to design new networks or to improve existing ones. For instance, redundant environments, that are environments attributed to the same envirotype, could be removed from the network, or new environments could be added to increase the representativeness of specific limiting factors. Indeed, the 11 environments where the DS127 was trialed consisted of a first optimization of the network as each envirotype was represented by at least one environment. Such optimized networks could be valuable for registration and post-registration trials to provide agricultural advice in the choice of cultivars adapted to one or several environments.

#### 4.4. An Approach that Provided First Clues to Tackle the $G \times E$ Interaction

After the understanding of the environmental effect, the second goal of this study was to decipher the  $G \times E$  interaction. According to the DS127 data, the  $G \times E$  interaction stood for 11.7% of the variation. The envirotyping permitted to reduce the  $G \times E$  between envirotype to 5% and reducing the  $G \times E$  inside envirotype to only 6.7%. However, as showed by the genotype ranking, the interaction was still present inside envirotypes. When we calculated the ecovalence of each environment as defined by Parisot-Baril [51], which corresponds to the contribution of each environment to the  $G \times E$  interaction, we did not observe that highly interactive (or stable) environments were grouped together. This may result from the fact that period definition, PLS regression, and envirotyping were performed on the two probes genotypes that were not perfectly representative of the whole diversity of DS127. To specifically target the pedoclimatic factors that do affect the  $G \times E$  interaction, it could be worth using the PLS regression approach to explain directly the environmental ecovalence instead of the mean environment seed yield or directly performing a multivariate PLS regression on the interaction matrix estimated for large number of genotypes trialed under numerous environments (more than 20 environments). However, envirotyping directly on  $G \times E$  interaction term remains tricky as  $G \times E$  only accounted for a small part of total variation, leading to potential confounding effects with errors. Such errors could be attributed to the fact that the  $G \times E$  interaction is dependent on both the genotype set and on the environment sets. In such  $G \times E$ -based clustering approach, the environments will be gathered into envirotypes according to their ability to contribute to  $G \times E$  and not to their patterns of seed yield limiting factors, leading to difficulties to interpret the clustering. Moreover, the clustering will be dependent on the given genotype set. It will, therefore, not be possible to use the obtained clustering to optimize a field network. In our study, the envirotyping was based on seed yield variation between environments, leading to an agronomic characterization of the envirotype. We then demonstrated that this envirotyping was also useful to control a part of the  $G \times E$  interaction and could be transferred from few genotypes to a larger range of genetic diversity.

#### 4.5. Get Further into the Genetic Determinant of the $G \times E$ Interaction and Its Interest for Breeding Programs

Handling  $G \times E$  interaction is critical for breeders, as it is a driver of a wide versus specific adaptation of cultivars to environments. Stable cultivars that present a small contribution to the  $G \times E$  are adapted for large areas, whereas cultivars dedicated to specific environments may benefit from high  $G \times E$  interaction. In most cases, the breeding material is evaluated for a variety of locations. Another option consists in breeding genotypes using a single representative environment of the region of interest [52]. However, we showed that trials carried out in the same location but on different years did not present the same limiting factors. Therefore, we propose an alternative strategy where breeding trials are grouped into envirotypes according to their limiting factors and not according to



their geographical proximity. This methodology can help breeders in designing field networks that emphasize limiting factors for their breeding programs, but we must define new indicators related to biotic interaction or cultural conditions. To help in the identification and the assessment of new pedoclimatic indicators, crop models can be used to simulate daily indicators, such as nitrogen nutrition index, soil water content, soil nutrient availability (N, P, K, S...).

## 5. Conclusions

The results of our study could find applications in quantitative genetics (for QTL detection/GWAS) and for breeding (with marker-assisted selection (MAS)) using the envirotypes or by considering the limiting factors as covariables in the models. Indeed, the  $G \times E$  interactions lead to instability of the loci detected from one environment to the other. Understanding this specificity will help breeders considering specific loci for MAS purposes, depending on a given environment (represented here by an envirotype and its pattern of limiting factors). Our results could also be used to improve genomic predictions by calibrating models for each envirotype or using environmental covariates into the calibrations to predict the environment effect and genotypic performances across environments.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/9/12/798/s1>. Figure S1: Description of the DS127 set based on the ecovalence calculated for each genotype. Figure S2: Adaptation of crop coefficient (Kc) dynamics for winter oilseed rape to perform soil water content estimation across the crop cycle. Figure S3: Seed yield distribution of the DS127 set following the envirotypes defined for Aviso and Montego. Table S1: Field network and crop management strategies. Table S2: Description of the DS127 diversity set of winter oilseed rape. Table S3: Matrix of correlation coefficients between the 68 pedoclimatic indicators.

**Author Contributions:** Conceptualization, E.C., A.G., C.B.-B., A.L. and N.N.; methodology, E.C., A.G. and A.L.; software, E.C.; validation, E.C. and A.L.; formal analysis, E.C.; investigation, E.C., C.B.-B., A.L. and N.N.; resources, A.L. and N.N.; data curation, E.C., A.-S.B., A.L. and N.N.; writing—original draft preparation, E.C., A.L. and N.N.; writing—review and editing, A.G., A.-S.B., C.B.-B., A.L. and N.N.; visualization, E.C., A.L. and N.N.; supervision, A.L. and N.N.; project administration, A.L. and N.N.; funding acquisition, N.N.

**Funding:** “This research was supported by two national collaborative projects entitled GENERGY (ANR-07-GPLA-016) funded by the French National Research Agency (ANR) and RAPSODYN (ANR-11-BTBR-0004) funded by the program “Investments for the Future”.

**Acknowledgments:** The authors acknowledge Olivier Godinot (AGROCAMPUS OUEST, France) for providing a useful extraction of the climatic data from Météo France. We thank Sébastien Faure (Innolea, France) and Patrick Bagot (GEVES, France) for helpful discussions about the development of the pedoclimatic indicators, Morgan Hamm (University of Washington, Seattle, USA) for English proofreading of the manuscript. We are grateful to the technical teams for excellent field trial management and data collection, especially Elise Alix, Bernard Moulin, Alina Tollenaere, Tyfanie Bourlet and the Experimental Unit of “La Motte” from INRA (Le Rheu, France); Pierre George, Amandine Larièpe and Sébastien Faure from Innolea (Mondonville, France); Johanna Younous and Thomas Foubert from Euralis (Châteauroux, France); Xavier Pinochet and Guillaume Joly from Terres Inovia (Dijon, France); Clara Simon and Jean-Eric Dheu from Limagrain Europe (Verneuil l’Etang, France); Olivier Maes and Alexandra Heulin from MAS seeds (Rhodon, France); Damien Dugué from RAGT (Prémèsques, France); Julien Moreau from Syngenta (Verpillères, France). We acknowledge the BrACySol Biological Resource Center (INRA Ploudaniel, France) for providing most of the seeds that were used in this study.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. von Wricke, G. Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. *Z. Planznezucht* **1962**, *47*, 92–96.
2. Finlay, K.W.; Wilkinson, G.N. The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.* **1963**, *14*, 742–754. [[CrossRef](#)]
3. Brancourt-Hulmel, M.; Biarnès-Dumoulin, V.; Denis, J. Points de repère dans l’analyse de la stabilité et de l’interaction génotype-milieu en amélioration des plantes. *Agronomie* **1997**, *17*, 219–246. [[CrossRef](#)]
4. Denis, J.B. Two way analysis using covariates. *Stat. A J. Theor. Appl. Stat.* **1988**, *19*, 123–132.
5. Vargas, J.; van Eeuwijk, F.; Sayre, K.D.; Reynolds, P.M.C. Interpreting treatment  $\times$  environment interaction in agronomy trials. *Agron. J.* **2001**, *93*, 949–960. [[CrossRef](#)]



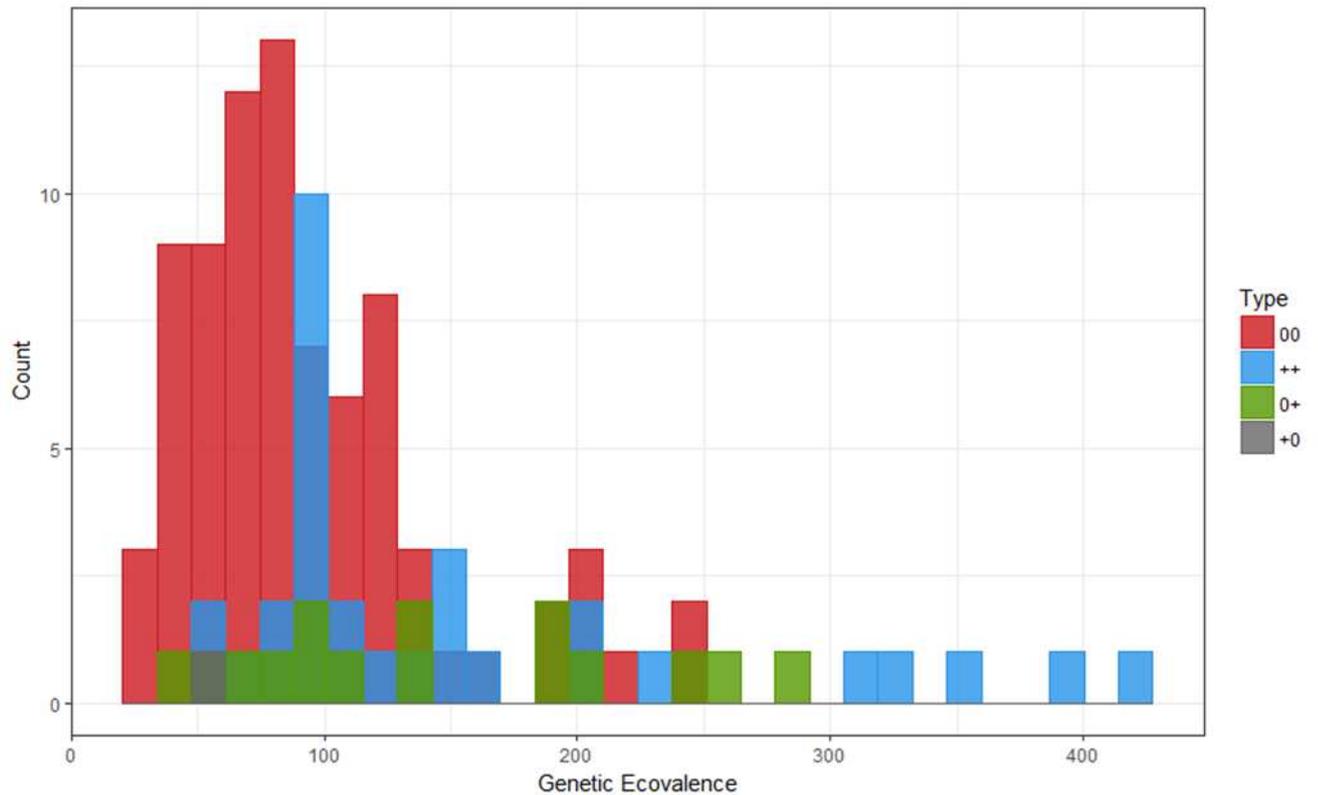
6. Van Ittersum, M.K.; Cassman, K.G.; Grassini, P.; Wolf, J.; Tittonell, P.; Hochman, Z. Yield gap analysis with local to global relevance—A review. *Filed Crop. Res.* **2013**, *143*, 4–17. [[CrossRef](#)]
7. FAOSTAT Production/Yield Quantities of Rapeseed in World. Available online: <http://www.fao.org/faostat/en/#data/QC/visualize> (accessed on 20 April 2019).
8. Bouchet, A.S.; Laperche, A.; Bissuel-Belaygue, C.; Baron, C.; Morice, J.; Rousseau-Gueutin, M.; Dheu, J.E.; George, P.; Pinochet, X.; Foubert, T.; et al. Genetic basis of nitrogen use efficiency and yield stability across environments in winter rapeseed. *BMC Genet.* **2016**, *17*, 131. [[CrossRef](#)]
9. He, D.; Wang, E.; Wang, J.; Lilley, J.M. Genotype × environment × management interactions of canola across China: A simulation study. *Agric. For. Meteorol.* **2017**, *247*, 424–433. [[CrossRef](#)]
10. Moghaddam, M.J.; Pourdad, S.S. Genotype × environment interactions and simultaneous selection for high oil yield and stability in rainfed warm areas rapeseed (*Brassica napus* L.) from Iran. *Euphytica* **2011**, *180*, 321–335. [[CrossRef](#)]
11. Zhang, H.; Berger, J.D.; Milroy, S.P. Genotype × environment interaction studies highlight the role of phenology in specific adaptation of canola (*Brassica napus*) to contrasting Mediterranean climates. *Filed Crop. Res.* **2013**, *144*, 77–88. [[CrossRef](#)]
12. Metzger, M.J.; Bunce, R.G.H.; Jongman, R.H.G.; Mùcher, C.A.; Watkins, J.W. A climatic stratification of the environment of Europe. *Glob. Ecol. Biogeogr.* **2005**, *14*, 549–563. [[CrossRef](#)]
13. Habekotté, B. Evaluation of seed yield determining factors of winter oilseed rape (*Brassica napus* L.) by means of crop growth modelling. *Filed Crop. Res.* **1997**, *54*, 137–151. [[CrossRef](#)]
14. Champolivier, L.; Merrien, A. Effects of water stress applied at different growth stages to *Brassica napus* L. var. *oleifera* on yield, yield components and seed quality. *Eur. J. Agron.* **1996**, *5*, 153–160. [[CrossRef](#)]
15. Parnaudeau, V.; Jeuffroy, M.; Machet, J.; Reau, R.; Bissuel, C. Methods for determining the nitrogen fertiliser requirements of some major arable crops in France. In Proceedings of the International Fertiliser Society, Cambridge, UK, 11 December 2009; pp. 1–26.
16. Rémy, J.; Hébert, J. Le devenir des engrais azotés dans le sol. *Acad. l'Agric. Fr.* **1977**, *63*, 700–714.
17. Colnenne, C.; Meynard, J.; Reau, R.; Justes, E.; Merrien, A. Determination of a Critical Nitrogen Dilution Curve for Winter Oilseed Rape. *Ann. Bot.* **1998**, *81*, 311–317. [[CrossRef](#)]
18. Lancashire, P.D.; Bleiholder, H.; Van Den BOOM, T.; Langelüddeke, P.; Stauss, R.; Weber, E.; Witzemberger, A. A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.* **1991**, *119*, 561–601. [[CrossRef](#)]
19. Weber, E.; Bleiholder, H. Explanations of the BBCH decimal codes for the growth stages of maize, rape, faba beans, sunflowers and peas—with illustrations. *Gesunde Pflanz.* **1990**, *42*, 308–321.
20. Gabrielle, B.; Denoroy, P.; Gosse, G.; Justes, E.; Andersen, M.N. Development and evaluation of a CERES-type model for winter oilseed rape. *Filed Crop. Res.* **1998**, *57*, 95–111. [[CrossRef](#)]
21. Hebing, H. *Le Colza*; Editions France Agricole: Paris, France, 2013; ISBN 978-2-85557-241-3.
22. Leterme, P. *Modelisation De La Croissance Et De La Production Des Siliques Chez Le Colza D'hiver (Brassica napus L.)*; INA Paris-Grignon: Paris, France, 1985.
23. Jullien, A.; Mathieu, A.; Allirand, J.M.; Pinet, A.; De Reffye, P.; Cournède, P.H.; Ney, B. Characterization of the interactions between architecture and sourcesink relationships in winter oilseed rape (*Brassica napus*) using the GreenLab model. *Ann. Bot.* **2011**, *107*, 765–779. [[CrossRef](#)]
24. Météo France Données Quotidiennes Du Modèle De Simulation Des Schémas De Surface. Available online: [https://donneespubliques.meteofrance.fr/?fond=produit&id\\_produit=230&id\\_rubrique=40](https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=230&id_rubrique=40) (accessed on 1 January 2016).
25. Bruand, A.; Duval, O.; Cousin, I. Estimation des propriétés de rétention en eau des sols à partir de la base de données SOLHYDRO: Une première proposition combinant le type d'horizon, sa texture et sa densité apparente. *Étude Gest. Des Sols* **2004**, *11*, 323–334.
26. Fan, J.; McConkey, B.; Wang, H.; Janzen, H. Root distribution by depth for temperate agricultural crops. *Filed Crop. Res.* **2016**, *189*, 68–74. [[CrossRef](#)]
27. Lacoste, M.; Mulder, V.L.; Richer-De-Forges, A.C.; Martin, M.P.; Arrouays, D. Evaluating large-extent spatial modeling approaches: A case study for soil depth for France. *Geoderma Reg.* **2016**, *7*, 137–152. [[CrossRef](#)]
28. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. Crop evapotranspiration—Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. *Irrig. Drain.* **1998**, 1–15.
29. Weymann, W.; Böttcher, U.; Sieling, K.; Kage, H. Effects of weather conditions during different growth phases on yield formation of winter oilseed rape. *Filed Crop. Res.* **2015**, *173*, 41–48. [[CrossRef](#)]



30. Fischer, R.A. Number of kernels in wheat crops and the influence of solar radiation and temperature. *J. Agric. Sci.* **1985**, *105*, 447–461. [[CrossRef](#)]
31. Baux, A.; Wegmüller, J.; Holzkämper, A. Exploring Climatic Impact on Oilseed Rape Yield in Switzerland. *Procedia Environ. Sci.* **2015**, *29*, 123. [[CrossRef](#)]
32. Ward, J.H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [[CrossRef](#)]
33. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
34. Wold, S.; Albano, C.; Dunn, W.J., III; Esbensen, K.; Hellberg, S.; Johansson, E.; Sjöström, H. Pattern Recognition: Finding and Using Regularities in Multivariate Data. In “*Food Research and Data Analysis*”, *Proceedings of the IUFoST Conference, Oslo, Norway, 20–23 September 1982*; Martens, J., Ed.; Applied Science Publications: London, UK, 1983; pp. 147–188.
35. Palermo, G.; Piraino, P.; Zucht, H.D. Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. *Adv. Appl. Bioinform. Chem.* **2009**, *2*, 57–70. [[CrossRef](#)]
36. Tenenhaus, M. *La Régression PLS: Théorie Et Pratique*; Editions Technip: Paris, France, 1998; ISBN 2-7108-0735-1.
37. Schüürmann, G.; Ebert, R.U.; Chen, J.; Wang, B.; Kühne, R. External validation and prediction employing the predictive squared correlation coefficient—Test set activity mean vs. training set activity mean. *J. Chem. Inf. Model.* **2008**, *48*, 2140–2145. [[CrossRef](#)]
38. Gauchi, J.P.; Chagnon, P. Comparison of selection methods of explanatory variables in PLS regression with application to manufacturing process data. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 171–193. [[CrossRef](#)]
39. Sanchez, G. *Plsdepot: Partial Least Squares (PLS) Data Analysis Methods*, R package version 0.1.17; 2012; Available online: <https://CRAN.R-project.org/package=plsdepot> (accessed on 22 November 2019).
40. Krzanowski, W.J.; Lai, Y.T. A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. *Biometrics* **1988**, *44*, 23. [[CrossRef](#)]
41. Lê, S.; Josse, J.; Husson, F. FactoMineR: A Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25*, 1–18. [[CrossRef](#)]
42. Husson, F.; Lê, S.; Pagès, J. *Computer Science and Data Analysis Series Exploratory Multivariate Analysis by Example Using R*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017; ISBN 9781138196346.
43. Chandler, J.; Corbesier, L.; Spielmann, P.; Dettendorfer, J.; Stahl, D.; Apel, K.; Melzer, S. Modulating flowering time and prevention of pod shatter in oilseed rape. *Mol. Breed.* **2005**, *15*, 87–94. [[CrossRef](#)]
44. Ferreira, M.E.; Satagopan, J.; Yandell, B.S.; Williams, P.H.; Osborn, T.C. Mapping loci controlling vernalization requirement and flowering time in Brassica napus. *Theor. Appl. Genet.* **1995**, *90*, 727–732. [[CrossRef](#)] [[PubMed](#)]
45. Sheldon, C.C.; Finnegan, E.J.; Rouse, D.T.; Tadege, M.; Bagnall, D.J.; Helliwell, C.A.; Peacock, W.J.; Dennis, E.S. Control of Flowering By Vernalization. *Curr. Opin. Plant Biol.* **2000**, *3*, 418–422. [[CrossRef](#)]
46. Tadege, M.; Sheldon, C.C.; Helliwell, C.A.; Stoutjesdijk, P.; Dennis, E.S.; Peacock, W.J. Control of flowering time by FLC orthologues in Brassica napus. *Plant J.* **2001**, *28*, 545–553. [[CrossRef](#)]
47. Morrison, M.J. Heat stress during reproduction in summer rape. *Can. J. Bot.* **1993**, *71*, 303–308. [[CrossRef](#)]
48. Angadi, S.V.; Cutforth, H.W.; Miller, P.R.; Mcconkey, B.G.; Entz, M.H.; Brandt, S.A. Response of three Brassica species to high temperature stress during reproductive growth. *Can. J. Plant Sci.* **2000**, *80*, 693–702. [[CrossRef](#)]
49. Young, L.W.; Wilen, R.W.; Bonham-smith, P.C. High temperature stress of Brassica napus during fowering reduces micro- and megagametophyte fertility, induces fruit abortion, and disrupts seed production. *J. Exp. Bot.* **2004**, *55*, 485–495. [[CrossRef](#)]
50. Nuttal, W.F.; Moulin, A.P.; Townley-Smith, L.J. Yield Response of Canola to Nitrogen, Phosphorus, Precipitation, and Temperature. *Agron. J.* **1992**, *84*, 765–768. [[CrossRef](#)]
51. Parisot-Baril, C. *Etude De La Stabilité Du Rendement Chez Le Ble Tendre D’hiver (Triticum Aestivum L. THELL.)*; Université Paris-Sud: Orsay, France, 1992.
52. Yan, W. Analysis and Handling of G × E in a Practical Breeding Program. *Crop Sci.* **2016**, *56*, 2106–2118. [[CrossRef](#)]

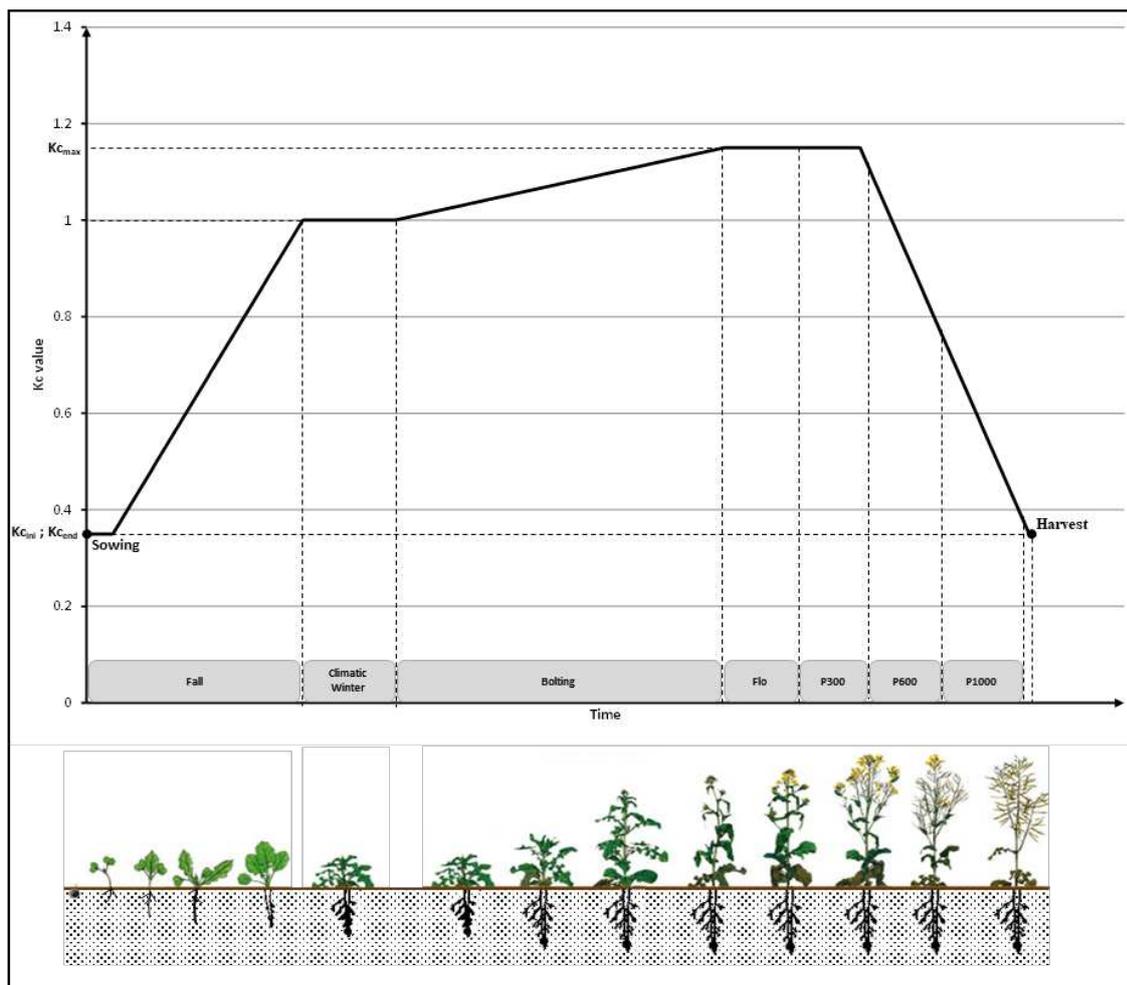




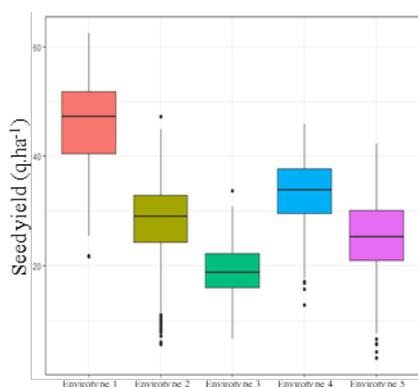


**Figure S1.** Description of the DS127 set based on the ecovalence calculated for each genotype. The ecovalence is defined as:  $W_g^2 = \sum_{e=1}^E (Y_{ge} - Y_{g..} - Y_{.e} + Y_{...})^2$  where g stands for the number of genotype, e stands for the number of environments and Y stands for the seed yield [1]. The ecovalence gives information about the stability of a genotype across environments. A high ecovalence means that the genotype is not stable across environments. The different accession types of the accessions are as following: 00 for low in erucic acid and glucosinolates, ++ for high in erucic acid and glucosinolates, 0+ low in erucic acid and high in glucosinolates, +0 high in erucic acid and low in glucosinolates.





**Figure S2.** Adaptation of crop coefficient ( $K_c$ ) dynamics for winter oilseed rape to perform soil water content estimation across the crop cycle (adapted from Allen et al [xx]).  $K_c$  takes different values during the cycle. The seven periods are represented (Fall, Climatic Winter, Bolting, Flo, P300, P600 and P1000). Four values are fixed for the  $K_c$ . The minimal and maximal values for  $K_c$  ( $K_{c_{ini}}$ ,  $K_{c_{max}}$ ) and the  $K_{c_{end}}$  were chosen according to Allen et al (1998). We adapted the dynamics of  $K_c$  to reflect the climatic winter period: the  $K_c$  was fixed to 0.35 at sowing and during the first 150 GDD, from 150 GDD post sowing up to the beginning of the climatic winter, the  $K_c$  followed a linear dynamic and reached the value of 1 ( $K_{c_{cw}}$ ). During the climatic winter period, the  $K_c$  remained fixed to 1 to mimic the absence of biomass accumulation during the vegetative break. Another linear kinetics was fit to predict the  $K_c$  value from the end of winter up to flowering, where the  $K_c$  reached its maximal value of 1.15 ( $K_{c_{max}}$ ). During the FLO and P300 periods, the  $K_c$  is fixed to 1.15. Finally, from the P300 up to the harvest, the  $K_c$  followed a linear dynamic and decreased to the value of 0.35 ( $K_{c_{end}}$ ).



**Figure S3.** Seed yield distribution of the DS127 set following the envirotypes defined for Aviso and Montego.



**Table S1.** Field network and crop management strategies

Trial information								Cultural practices						Phenotypic value
Environment ID	Crop season	Location	Latitude	Longitude	Maximal Water Soil Content (mm) <sup>1</sup>	Experimental design <sup>2</sup>	Population <sup>3</sup>	Sowing date	Total N fertilization (U = kg.ha <sup>-1</sup> )	First N fertilization (date)	Second N fertilization (date)	Third N fertilization (date)	Harvest date	Seed yield (t.ha <sup>-1</sup> ) <sup>4</sup>
Ch14	2013-2014	Châteauroux	46.914158	1.756584	122.1	2 replicates	DS127	09/03/2013	140 U	70 U (02/17/2014)	70 U (03/15/2014)	none	07/02/2014	2.94 (± 0.27)
Dij13	2012-2013	Dijon	47.230468	5.10036	122.1	3 replicates	Probes	09/12/2012	180 U	60 U (02/14/2013)	70 U (03/14/2013)	50 U (04/19/2013)	08/05/2013	4.70 (± 0.393)
Dij15	2014-2015				122.1	2 replicates	DS127	09/04/2014	140 U	60 U (02/19/2015)	80 U (03/13/2015)	none	07/08/2015	4.43 (± 0.781)
Liv16	2015-2016	Liverdy	48.7048247	2.7564469	197.4	3 replicates	Probes	03/09/2015	140 U	NA	NA	NA	07/17/2016	2.72 (± 0.257)
LR11	2010-2011	Le Rheu	48.13916	-1.799776	197.4	3 replicates	Probes	09/14/2010	130 U	40 U (02/08/2011)	90 U (03/10/2011)	none	06/28/2011	4.28 (± 0.631)
LR12	2011-2012				197.4	4 replicates	Probes	09/14/2011	80 U	40 U (02/21/2012)	40 U (03/14/2012)	none	07/16/2012	3.34 (± 0.406)
LR13	2012-2013				197.4	4 replicates	Probes	09/05/2012	80 U	40 U (02/25/2013)	40 U (03/28/2013)	none	07/15/2013	3.36 (± 0.186)
LR15	2014-2015				197.4	3 replicates	DS127	09/10/2014	40 U	40 U (02/27/2015)	none	none	07/01/2015	3.21 (± 0.336)
LR16	2015-2016				197.4	2 replicates	Probes	09/09/2015	70 U	70 U (02/29/2016)	none	none	07/11/2016	3.27 (± 0.428)
Md11	2010-2011	Mondonville	43.7825	1.341389	122.1	2 replicates	Probes	09/08/2010	120 U	40 U (02/16/2011)	80 U (03/10/2011)	none	06/09/2011	2.92 (± 0.276)
Md14	2013-2014				122.1	2 replicates	Probes	09/19/2013	190 U	50 U (02/25/2014)	70 U (03/14/2014)	70 U (04/02/2014)	06/19/2014	2.70 (± 0.656)
Md15	2014-2015				122.1	2 replicates	DS127	09/02/2014	120 U	120 U (03/10/2015)	none	none	06/27/2015	2.92 (± 0.436)
Pre14	2013-2014	Prémesques	50.66556	2.926667	153.6	2 replicates	DS127	09/05/2013	170 U	50 U (03/07/2014)	120 U (03/31/2014)	none	08/01/2014	2.44 (± 0.292)
Pre15	2014-2015				153.6	2 replicates	DS127	09/10/2014	175 U	75 U (03/27/2015)	100 U (04/15/2015)	none	07/23/2015	4.02 (± 0.315)
Pre16	2015-2016				153.6	3 replicates	Probes	09/10/2015	160 U	60 U (03/11/2016)	100 U (04/01/2016)	none	07/27/2016	3.73 (± 0.443)
Sel14	2013-2014	Selommès	47.44324	1.14943	138.9	2 replicates	DS127	09/04/2013	160 U	50 U (02/27/2014)	50 U (03/07/2014)	60 U (03/15/2014)	07/17/2014	3.44 (± 0.411)
Sel15	2014-2015				138.9	2 replicates	DS127	09/09/2014	160 U	80 U (02/17/2015)	50 U (02/27/2015)	30 U (03/12/2015)	07/06/2015	2.74 (± 0.483)
Ver14	2013-2014	Verpillères	49.663611	2.834444	189.6	2 replicates	DS127	09/04/2013	90 U	40 U (02/27/2014)	50 U (03/13/2014)	none	07/18/2014	3.57 (± 0.231)
Ver15	2014-2015				189.6	2 replicates	DS127	09/02/2014	130 U	50 U (02/20/2015)	80 U (03/13/2015)	none	07/27/2015	3.60 (± 0.11)
Yeb15	2014-2015	Yèbles	48.62417	2.7641667	197.4	2 replicates	DS127	09/03/2014	150 U	90 U (02/12/2015)	60 U (03/09/2015)	none	07/19/2015	3.78 (± 0.147)

NA, not available.

All the dates are given as MM/DD/YYYY.

<sup>1</sup>Maximal Water soil content defined according to the soil texture and without consideration of the density according to Bruand et al [xx]

<sup>2</sup>All the trials were designed as randomized complete block design.

<sup>3</sup>Genotypes used in the present study consisted of a set of two probe genotypes (Aviso and Montego included in DS127) that were phenotyped in the 20 environments as well as 127 winter oilseed rape genotypes (DS127 described in Table S2) that were phenotyped in 11 environments

<sup>4</sup>Seed yield is given as the mean seed yield values from the two probe genotypes.



**Supplementary Table 2.** Description of the 127 WOSR accessions composing DS127

ACCESSION NAME	TYPE		
Alsace 1360	++	Licorne	00
Askari	++	Lioness	00
Blagodatnyi	++	Liporta	00
Cetmeki	++	Lirabon	00
Debruder Dippes	++	Lirajet	00
Fertodi Elit	++	Madora2	00
Gaspard	++	Madora3	00
Gesunder	++	Madrigal	00
Goldgelber Zarter Butter	++	Makila	00
Hochzucht Repole	++	Mar	00
Kormovoi Skrivskyi	++	Milena	00
Lembke B	++	Mohican	00
Lembke R	++	Montego	00
Major	++	Navajo	00
Mansholts	++	NSL 96-25	00
Markus	++	Oase	00
Matador	++	Olymp	00
Moldavia	++	Pacific	00
Nain de Hambourg	++	Payrol	00
Oleski	++	Pollen	00
Olquell Gulzover	++	Potomac	00
Orleans	++	Prince	00
Polux	++	Recital	00
Prerouska	++	Remy	00
QuedLinburger Platzfester	++	Rocket	00
Rangi	++	Roxet	00
Rapol	++	Sahara	00
Regal2	++	Salomont	00
Sarepta	++	Shogun	00
Titus	++	Start	00
Tonus	++	SW Gospel	00
Aztec	0+	Tarox	00
Belinda	0+	Tenor	00
Beryl	0+	Tor	00
Bienvenu	0+	Tradition	00
Brink	0+	Twister	00
Constable	0+	Zorro	00
Doral	0+		
Emil	0+		
Jupiter	0+		
Kurander	0+		
Lesira	0+		
Mikado	0+		
Rafal	0+		
Ramses	0+		
Valdor	0+		
Zeruca	+0		
Adriana	00		
Akamar	00		
Alesi	00		
Aligator	00		
Amber	00		
Ariana	00		
Astrid	00		
Atlantic	00		
Aurum	00		
Aviso	00		
Bellini	00		
Bolko	00		
Boston	00		
Campala	00		
Canberra	00		
Cancan	00		
Capitol	00		
Capricorn	00		
Capvert	00		
Ceres	00		
Chelsi	00		
Cheops	00		
Colvert	00		
Contact	00		
Courage	00		
Darmor	00		
Doublol	00		
Ecrin	00		
Eleonore	00		
Escort	00		
Evri	00		
Expander	00		
Express	00		
Frederic	00		
Goeland	00		
Ibex	00		
Idol	00		
Inca	00		
Jetton	00		
Kadore	00		
Kosto	00		
Lewis	00		
Liberator	00		







## II.4. Prise en compte des itinéraires techniques azotés contrastés dans l'envirotypage

---

Dans cette partie, l'objectif est d'appliquer la méthode développée précédemment (Corlouer et al., 2019) à un réseau élargi pour lequel deux niveaux de fertilisation azotée contrastés sont appliqués (réseau #2, Tableau II.1). Pour traiter ce réseau, nous avons défini de nouveaux indicateurs en lien avec la fertilisation azotée (cf. II.4.1). Ces indicateurs sont relatifs i) au statut azoté de la plante, ii) à l'itinéraire technique, ou iii) aux dynamiques hydriques au moment de l'apport azoté. Une fois ces indicateurs définis, ils ont été ajoutés au set d'indicateurs pédoclimatiques déjà existants (cf. II.3) afin d'identifier les facteurs limitants permettant de définir les envirotypes (cf. II.4.2 et II.4.3).

### II.4.1. Définition des indicateurs liés à la fertilisation azotée

#### *Indicateurs du statut azoté de la plante : NNI, DAB et AN*

L'indice de nutrition azotée (INN ; l'indicateur correspondant sera noté **NNI** par la suite) permet d'évaluer le statut azoté du couvert végétal à l'échelle de l'essai. De par sa définition (cf. Encadré II.1), le calcul de l'INN dépend du prélèvement de plantes effectué au champ, prélèvement qui doit se faire au moins 15 jours après le dernier apport azoté afin de mesurer sans biais la biomasse sèche aérienne produite ainsi que la teneur en azote des parties aériennes. Ces données sont ensuite projetées sur la courbe de dilution établie par Colnenne et al. (1998) pour définir la teneur en azote critique. L'INN est alors défini comme le rapport entre la teneur en azote observée et la teneur en azote critique. L'INN est donc dépendant des conditions dans lesquelles le prélèvement a été effectué et correspond à l'état azoté du couvert à un instant  $t$ . Idéalement, les valeurs d'INN auraient pu être obtenues à un pas de temps journalier à l'aide de modèles de culture tels que le modèle AZODYN-COLZA (Jeuffroy et al., 2006). Ces données permettant alors d'obtenir une cinétique de l'état azoté du couvert à chaque période de développement du colza afin de repérer les carences en azote précoces ou tardives. Cependant AZODYN-COLZA est en cours de développement (Gervois et al., 2019) et les premières simulations obtenues ne permettent pas de l'utiliser de manière fiable pour prédire l'INN, notamment en début de cycle. Nous avons donc choisi de ne pas utiliser ce modèle pour simuler des INN dans le cadre de mon travail de thèse et de nous satisfaire de l'utilisation des valeurs d'INN obtenues par prélèvements/mesures au stade montaison.



Afin de prendre en compte d'éventuelles imprécisions dans le calcul de l'INN engendrées par des biais expérimentaux (ex. prélèvements trop proches de la date de fertilisation), il nous a semblé opportun de considérer des variables liées à la gestion de l'azote par le couvert végétal comme des indicateurs de statut azoté supplémentaires. Nous avons donc défini l'indicateur **DAB** (dry aerial biomass) qui correspond à la matière sèche aérienne mesurée au stade montaison après le dernier apport azoté, et l'indicateur **AN** (aerial nitrogen) qui représente la quantité d'azote aérien mesurée au stade montaison après le dernier apport azoté.

*Indicateurs d'itinéraire technique azoté et de statut hydrique au moment des apports azotés : N, DBI, DAI, RF, RO et WSC\_I*

A partir de l'itinéraire technique appliqué pour la fertilisation azotée sur chaque essai, on peut élaborer un premier indicateur, **N** (Nitrogen input), qui correspond à la quantité totale d'azote (en kg/ha) apportée au champ.

Par ailleurs, les apports azotés au sein du réseau #2 ont pu être fractionnés en 1-3 fois en fonction des pratiques locales. Lors de chaque apport, les conditions hydriques étaient différentes pouvant se traduire par des variations en termes de disponibilité de l'azote apporté. L'azote est principalement absorbé sous forme d'ions  $\text{NO}_3^-$  (labiles) et plus minoritairement sous forme de  $\text{NH}_4^+$ , l'absence d'eau entraîne une baisse en disponibilité de l'azote ainsi qu'une altération du transport de l'azote aux racines (Rathke et al., 2006). De plus, en fonction des conditions environnementales, l'azote apporté peut être perdu notamment par un processus de lessivage des sols (Sieling et al., 1997). Par ailleurs, l'absorption de l'azote par les racines est dépendante de plusieurs facteurs tels que la perte de nitrate par lessivage qui peut représenter 26% de l'apport dans le cas de sols limoneux (MacDonald et al., 1997 ; Rathke et al., 2006) ainsi que la quantité d'eau dans le sol. Afin de prendre en compte ces éléments au moment de la fertilisation des essais, des indicateurs hydriques ont été calculés pour chaque date d'apport afin de quantifier l'aptitude du couvert à absorber l'azote disponible en prenant en compte les précipitations journalières, les jours secs (*ie* sans précipitation), les risques de lessivage ainsi que l'état de la réserve en eau des sols. Cinq indicateurs liés aux dynamiques hydriques au moment de l'apport azoté ont donc été définis :

- **DBI (Dry days Before Input)**: Nombre de jours consécutifs sans précipitation 10 jours avant l'apport azoté
- **DAI (Dry days After Input)**: Nombre de jours sans précipitation 10 jours après l'apport azoté
- **RF (Rainfall)** : Cumul des précipitations 10 jours après l'apport azoté
- **RO (Run Off)** : Nombre de jours de lessivage 10 jours après l'apport azoté



- **WSC\_I (Water Soil Content around Input)**: Etat moyen de la réserve en eau au moment de l'apport azoté (3 jours avant et 10 jours après).

Afin d'obtenir un seul indicateur par essai alors que les dynamiques de fractionnement ont pu être différentes, les indicateurs azotés finaux sont définis de la manière suivante :

$$\text{Indicateur final} = \sum_{i=0}^n \text{Indicateur}_i \times \frac{N_i}{N_{total}} \quad (i)$$

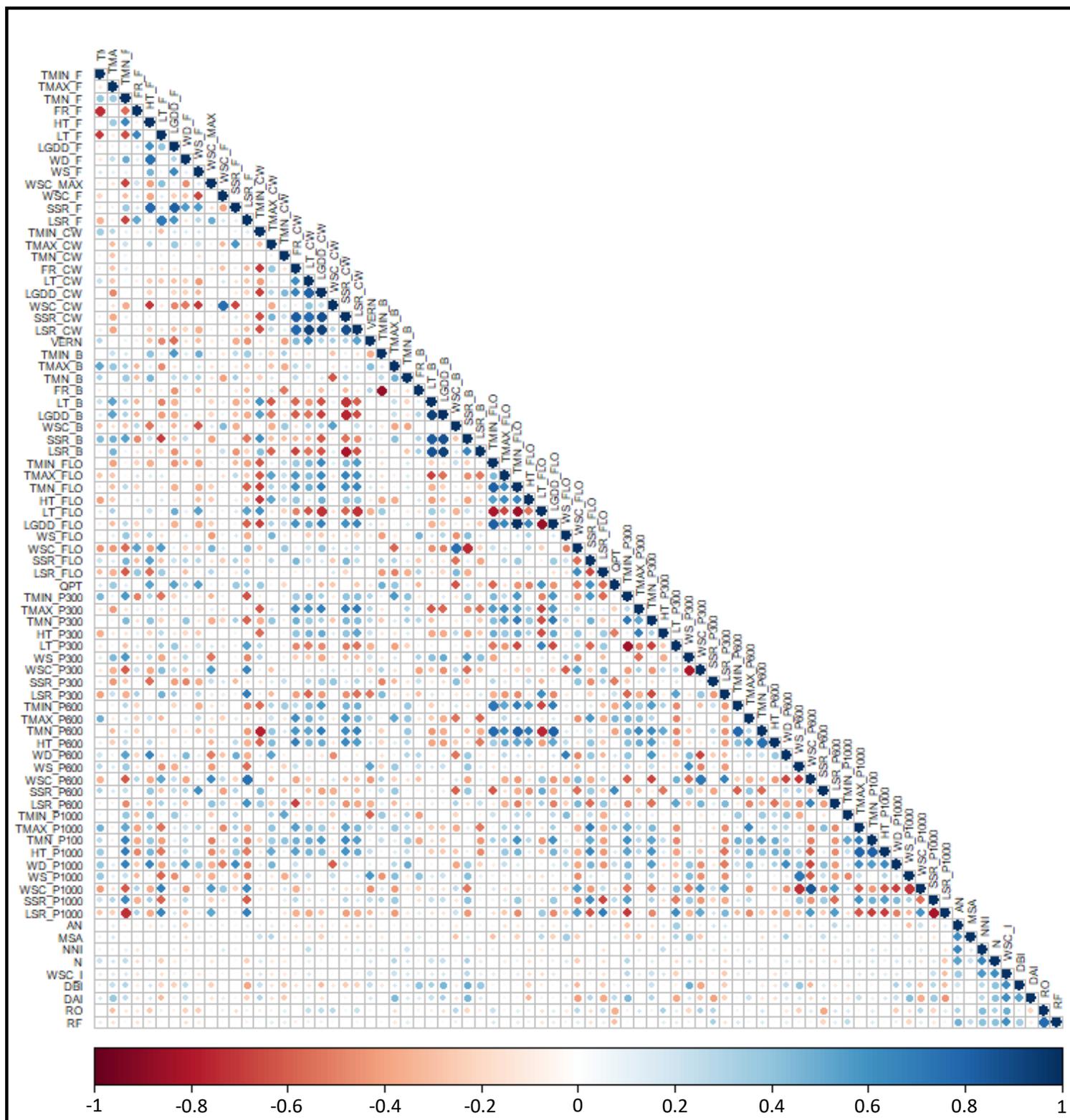
où  $\text{Indicateur}_i$  est l'indicateur calculé pour l'apport azoté  $i$  et  $N_i$  est la quantité d'azote apportée à la culture au moment de l'apport  $i$ . Pour les essais n'ayant pas reçu d'apport, nous avons fixé l'hypothèse suivante : si l'apport N n'a pas été réalisé, l'azote minéral n'est pas disponible pour la plante, l'azote apporté ne peut pas être lessivé, ni absorbé, les indicateurs définis ci-dessus prennent donc la valeur de 0. Ainsi tous les indicateurs azotés calculés en fonction de la date de fertilisation sont réunis en un seul indicateur qui prend en compte i) l'aptitude de la plante à bien valoriser les différents apports d'une part, et ii) l'absence de fertilisation le cas échéant.

En complément, nous aurions pu utiliser les reliquats azotés à la sortie de l'hiver. Ces reliquats renseignent sur la disponibilité en azote dans le sol au moment de la reprise de végétation et sont utilisés pour ajuster la fertilisation à l'aide de la méthode du bilan (Parnaudeau et al., 2009). Ils peuvent être mesurés par prélèvement de sol ou estimés. Les reliquats azotés n'ont pu être utilisés dans cette étude du fait d'une imprécision dans la donnée ou d'informations non remontées par les expérimentateurs.

## II.4.2. Sélection des indicateurs informatifs et analyse des corrélations

### *Identifications des indicateurs non informatifs*

Les indicateurs azotés décrits ci-dessus viennent s'ajouter au set d'indicateurs pédoclimatiques définis lors de la première étude (cf. II.3). Au total, 93 indicateurs pédoclimatiques et azotés sont disponibles pour l'analyse du réseau #2. Cependant, tous les indicateurs ne sont pas informatifs, certains présentent très peu de variation (variations sur 1 ou 2 essais par exemple) et représentent des cas particuliers pouvant entraîner un biais dans l'analyse, tandis que d'autres sont invariants sur le réseau. Ainsi, 14 indicateurs peu ou pas variants au sein du réseau #2 sont écartés des analyses. Ils correspondent à 8 indicateurs thermiques (HT\_CW, HT\_B, FR\_FLO, FR\_P300, FR\_P600, LT\_P600, FR\_P1000 et LT\_P1000) et 6 indicateurs hydriques (WD\_CW, WS\_CW, WD\_B, WS\_B, WD\_FLO, WD\_P300). A l'issue de ce premier filtre, 79 indicateurs ont donc été considérés pour les analyses.



**Figure II.4 : Corrélogramme des 79 indicateurs utilisés pour caractériser le réseau #2**

Chaque corrélation est représentée par un cercle de diamètre et de couleur variable. Les corrélations positives sont indiquées en bleu et les corrélations négatives en rouge. Plus une corrélation est proche de  $|1|$ , plus le diamètre du cercle est important.

### Analyse des corrélations

Les corrélations de Pearson entre les 79 indicateurs retenus sont calculées (Figure II.4). On considère que deux d'indicateurs sont fortement corrélés lorsque la  $p$ -value associée est inférieure à 0.001 et la valeur absolue de la corrélation est supérieure à 0.65. L'analyse de la matrice de corrélations montre un faible nombre d'indicateurs fortement corrélés : ils ne représentent que 4.3% des corrélations. De plus, les nouveaux indicateurs développés en lien avec la fertilisation azotée présentent un niveau de corrélation moyen de 16% avec les autres indicateurs pédoclimatiques (Figure II.4). Ce résultat suggère que les nouveaux indicateurs apportent des informations complémentaires aux indicateurs précédents. D'autre part, certains indicateurs azotés présentent de fortes corrélations entre eux tels que le nombre de jours de lessivage après apport (RO) et le cumul de précipitations après apport (RF) ( $r=0.77$ ). Cette relation illustre le fait que le lessivage est limité voire nul lorsque les précipitations sont faibles. Néanmoins le lessivage dépend aussi de l'état de la réserve utile au moment de l'apport, cette dernière devant être pleine pour que le processus de lessivage se mette en place.

### II.4.3. Envirotypage

#### Identification des facteurs limitants par régression PLS

Les 79 indicateurs pédoclimatiques définis précédemment sont utilisés pour l'identification des facteurs limitant le rendement au sein du réseau #2 selon la méthode décrite par Corlouer et al. (2019).

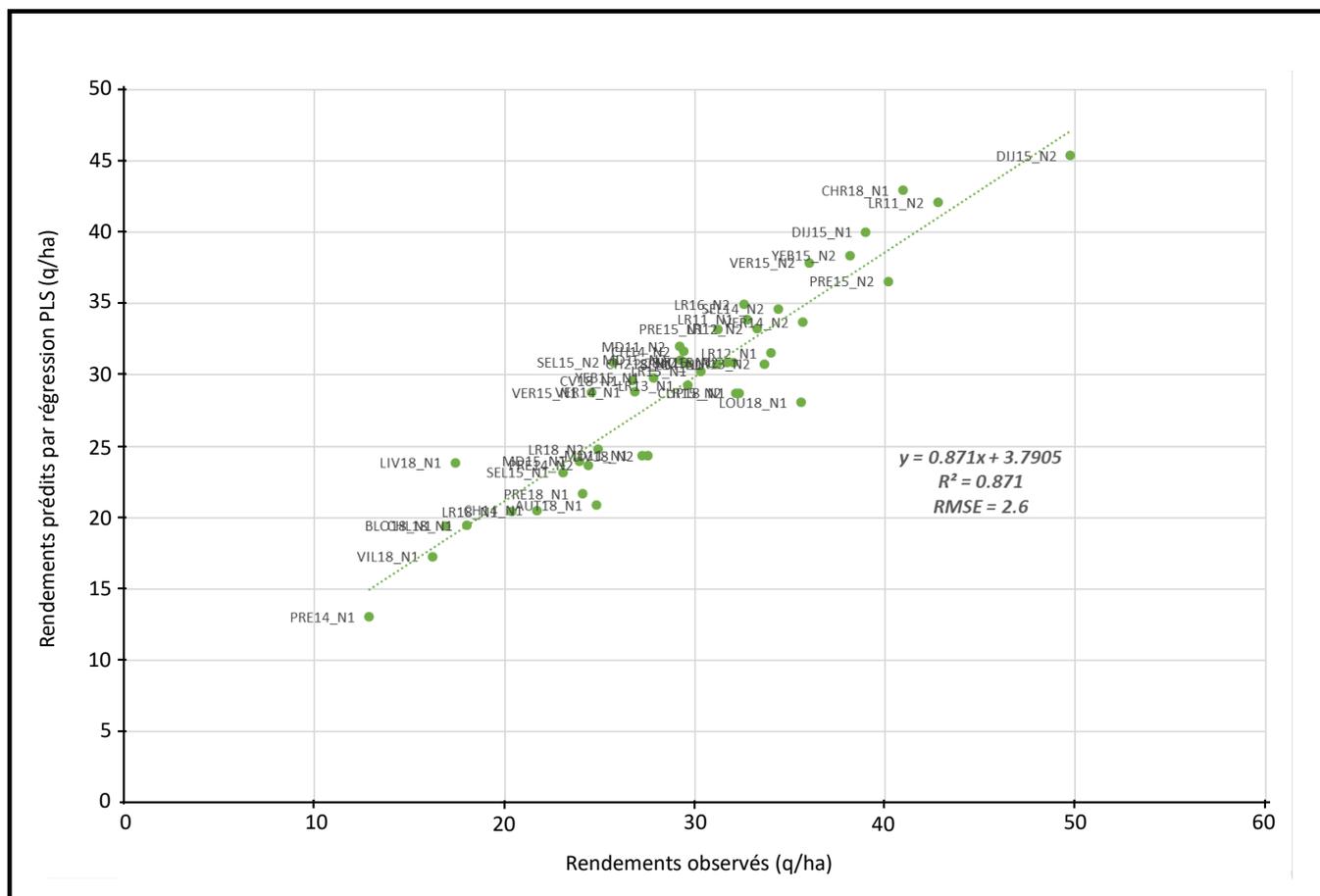
Le meilleur modèle de régression PLS retenu est un modèle à deux composantes. Il met en évidence 13 indicateurs comme impactant le rendement dans le réseau #2 (Tableau II.2) allant de la période hivernale (CW) à la période d'allocation des réserves aux graines (P600). Parmi ces indicateurs, 6 appartiennent à la catégorie thermique, 2 sont des indicateurs hydriques, un indicateur est radiatif, un indicateur est relié à la vernalisation et 3 indicateurs sont en lien avec la fertilisation azotée (Tableau II.2). Les rendements prédits par le modèle sont fortement corrélés aux rendements observés ( $R^2 = 0.87$ ,  $RMSE = 2.6$  ; Figure II.5) mais le  $R^2$  est plus faible que dans l'étude du réseau #1 ( $R^2 = 0.96$ ,  $RMSE = 1.3$  ; Corlouer et al., 2019). Ces différences de  $R^2$  et  $RMSE$  peuvent s'expliquer par le nombre d'essais plus important dans le réseau #2 par rapport au réseau #1.

Parmi les 13 indicateurs identifiés comme limitants, 8 sont spécifiques du réseau #2 (TMAX\_B, HT\_FLO, TMIN\_P600, WD\_P600, WSC\_MAX, NNI, N, DAI), 3 sont communs avec l'étude menée sur le réseau #1 (TMN\_CW, VERN, TMAX\_FLO ; Corlouer et al., 2019), et 2 indicateurs (TMAX\_P300,

**Tableau II.2 : Facteurs limitant le rendement dans le réseau #2 identifiés par régression PLS**

Indicateur <sup>(a)</sup>	Catégorie <sup>(b)</sup>	Indicateurs corrélés <sup>(c)</sup>	Réseau #1 <sup>(d)</sup>
TMN_CW	Thermique	-	Oui
TMAX_B	Thermique	-	
TMAX_FLO	Thermique	SSR_CW (0.67) ; LT_B (-0.69) ; LT_FLO (-0.66)	Oui
HT_FLO	Thermique	TMIN_CW (-0.7) ; TMN_FLO (0.66) ; LGDD_FLO (0.66) ; TMN_P600 (0.7)	
TMAX_P300	Thermique	LGDD_CW (0.68) ; SSR_CW (0.67) ; LSR_CW (0.7) ; LT_FLO (-0.68) ; HT_P300 (0.68)	Corrélé
TMIN_P600	Thermique	TMIN_FLO (0.73) ; TMN_FLO (0.7) ; LT_FLO (-0.67) ; LGDD_FLO (0.7), TMN_P600 (0.78)	
WD_P600	Hydrique	HT_F (0.66) ; WS_FLO (0.7), WSC_P300 (-0.68) ; WSC_P600 (-0.66)	
WSC_MAX	Hydrique	TMN_F (-0.67)	
SSR_FLO	Radiatif	LSR_FLO (-0.67) ; TMAX_P1000 (0.69) ; HT_P1000 (0.65)	Corrélé
VERN	Plante	LT_CW (0.65)	Oui
NNI	Azote	AN (0.65) ; N (0.66)	
N	Azote	NNI (0.66)	
DAI	Azote	-	

(a) Nom de l'indicateur identifié comme facteur limitant dans le réseau #2; (b) Catégorie de l'indicateur; (c) Indicateurs fortement corrélés aux facteurs limitants identifiés; (d) Comparaison avec les facteurs limitants identifiés dans le réseau #1, oui : indicateur identifié dans le réseau #1, non: indicateur spécifique du réseau #2, corrélé : indicateur fortement corrélé à un facteur limitant du réseau #1.



**Figure II.5 : Comparaison des rendements observés dans le réseau #2 et des rendements prédits par le modèle de régression PLS à 13 indicateurs pédo-climatiques**

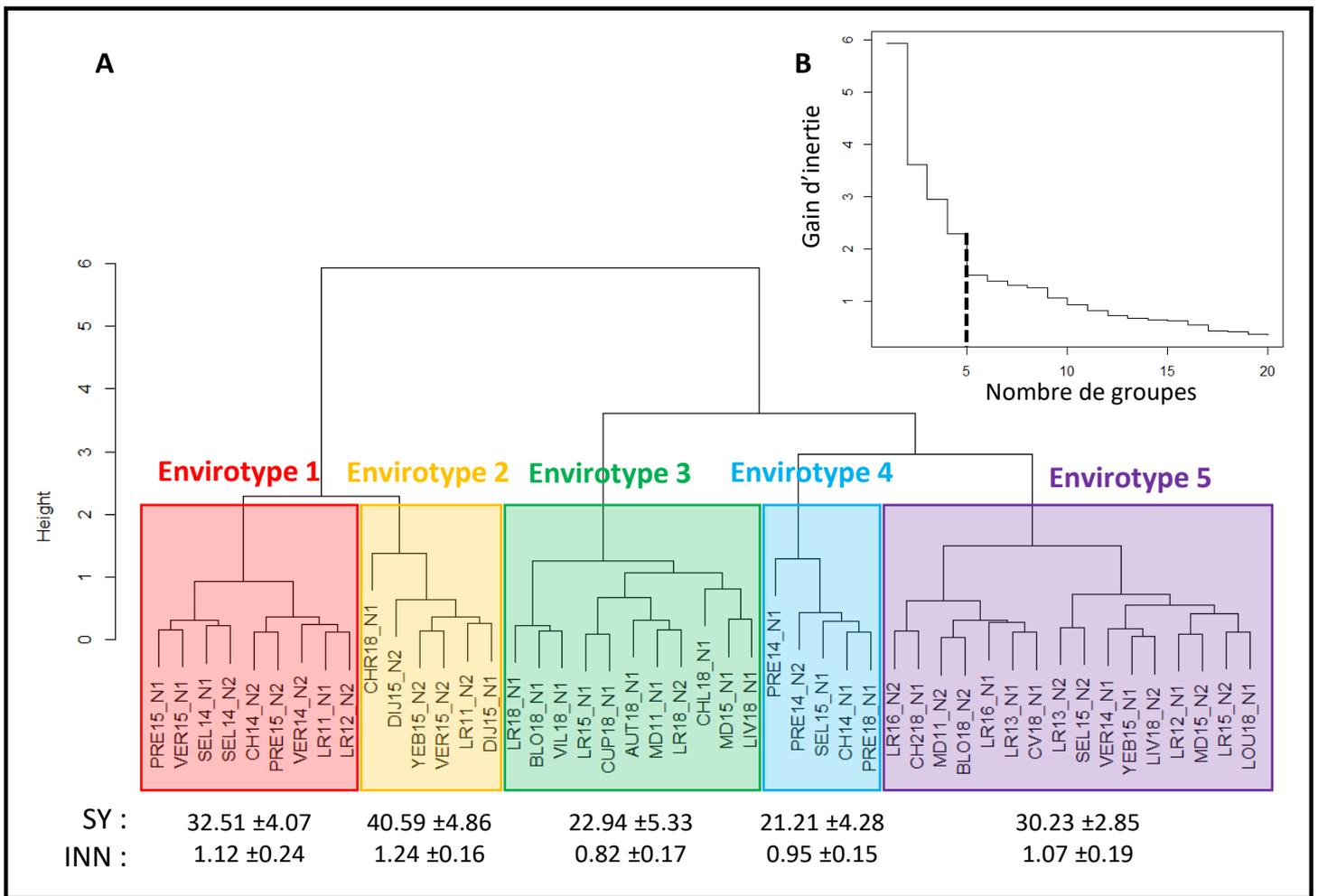
SSR\_FLO) présentent une corrélation significative avec des indicateurs identifiés lors de l'envirotypage du réseau #1 (HT\_P300 et LSR\_FLO respectivement) (Tableau II.2).

Les corrélations des facteurs limitants identifiés et précédemment décrites (Figure II.4) montrent une surreprésentation d'indicateurs en lien avec la période floraison ainsi que des indicateurs de l'hiver climatique. Ces corrélations renforcent l'hypothèse de l'importance de la vernalisation et de la floraison sur l'élaboration du rendement chez le colza d'hiver. La PLS a également mis en évidence l'impact de la température lors de la montaison. Outre ces trois périodes, c'est l'ensemble de la période reproductive du colza qui est principalement ciblée comme période critique pour le rendement par cette étude.

Comme attendu, les indicateurs azotés ont permis d'expliquer les variations de rendement au sein du réseau #2 : parmi les 13 facteurs limitants identifiés, 3 sont des indicateurs azotés (NNI, N et DAI). Les indicateurs NNI et N sont corrélés entre eux (0.66 ; Tableau II.2). Cette valeur, relativement basse, peut s'expliquer par des différences de dynamique hydrique lors des apports azotés entre les environnements, modifiant la disponibilité des nutriments. Cette hypothèse est confirmée par l'identification de l'indicateur DAI qui comptabilise le nombre de jours secs après apport.

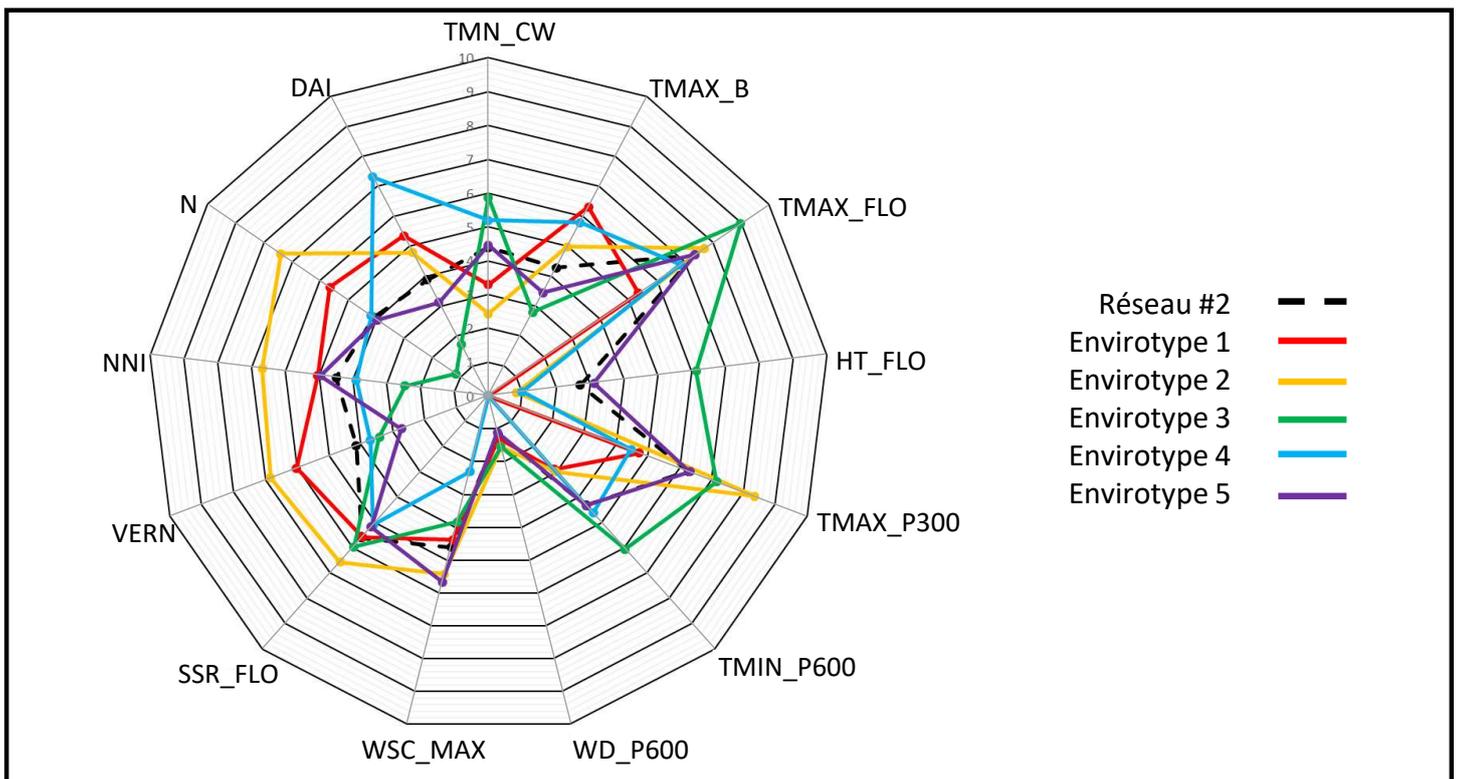
### *Définition et caractérisation des envirotypes*

A partir des résultats du modèle de régression PLS à 13 indicateurs, 5 envirotypes ont été définis par classification hiérarchique ascendante comme décrit par Corlouer et al. (2019) (Figure II.6). Chacun des envirotypes regroupe plusieurs années de culture, plusieurs localisations géographiques ainsi que des essais N1 et N2. Dans un premier temps, chaque envirotype peut être caractérisé par son rendement moyen ainsi que son INN moyen. **L'envirotype 1** présente un rendement moyen de 32.51 q/ha et un INN de 1.12, il est constitué de 9 essais dont 6 essais INN+/SY+, 2 essais INN+/SY- et 1 essai INN-/SY+. **L'envirotype 2** est composé de 6 INN+/SY+ avec un rendement moyen de 40.59 q/ha et un INN moyen de 1.24. Ces deux envirotypes, majoritairement constitué d'essais INN+ et SY+ peuvent être qualifiés d'envirotypes à fort INN et fort SY. **L'envirotype 3**, avec un rendement moyen de 22.94 q/ha et un INN moyen de 0.82 comporte 11 essais dont 7 essais INN-/SY-, 2 essais INN-/SY+ et 2 essais INN+/SY-. **L'envirotype 4**, d'un rendement moyen de 21.21 q/ha et d'un INN moyen de 0.95 est constitué par 5 essais dont 4 INN+/SY- et 1 INN-/SY-. Majoritairement constitués d'essais INN-/SY-, ces deux envirotypes sont définis comme des envirotypes à faible INN et faible rendement. Enfin **l'envirotype 5** est constitué de 16 essais dont 6 INN+/SY-, 5 INN+/SY+, 3 INN-/SY- et 2 INN-/SY+ pour un rendement moyen de 30.23 q/ha et un INN moyen de 1.07 (Figure II.6). Cet envirotype est un envirotype moyen constitué par un grand nombre d'essais et proche de la moyenne du réseau #2 pour l'INN et le rendement. Ces résultats illustrent aussi que l'azote n'est pas le seul facteur à



**Figure II.6: Définition des envirotypes**

(A) Dendrogramme des essais du réseau #2 et définition de 5 envirotypes. Chaque envirotype est caractérisé par sa moyenne de rendement et d'INN ainsi que les écarts-types associés. (B) Graphique des gains d'inertie en fonction du nombre de groupes considéré.



**Figure II.7 : Profil moyen des facteurs limitants par envirotype**

Graphique radar représentant le profil moyen des 13 facteurs limitants du réseau #2 et chaque envirotype. Chaque indicateur est exprimé sur une échelle de 0 à 10.

prendre en compte dans la structuration d'un réseau et qu'il est nécessaire de prendre en considération une combinaison de facteurs abiotiques pour expliquer les différences entre les envirotypes.

Chaque envirotype est caractérisé par chacun des 13 facteurs limitant identifiés (Figure II.7) et comparé par rapport à la moyenne du réseau #2. Les 5 envirotypes présentent des profils de facteurs limitants différents les uns des autres et différents de la moyenne du réseau #2. Pour l'ensemble des envirotypes seuls 3 indicateurs ne présentent pas en moyenne de différences significatives par rapport au réseau #2 moyen, 2 indicateurs hydriques (WSC\_MAX et WD\_P600) et 1 indicateur radiatif (SSR\_FLO).

**L'envirotype 1** se caractérise par une température maximale à montaison (TMAX\_B) et un nombre de jours optimal de vernalisation (VERN) plus élevés ainsi qu'un nombre de jours chauds pendant la floraison (HT\_FLO), une température maximale à P300 et pendant la floraison (TMAX\_FLO et TMAX\_P300) plus faibles. **L'envirotype 2** se caractérise par une température maximale à P300 (TMAX\_P300), une vernalisation (VERN), un INN (NNI) ainsi qu'un apport azoté (N) plus important. Il se caractérise également par température moyenne en hiver (TMN\_CW) plus faible. **L'envirotype 3** se distingue par des températures à l'hiver (TMN\_CW), à floraison (TMAX\_FLO, HT\_FLO) et à P600 (TMIN\_P600) plus élevées que la moyenne du réseau ainsi que par une température à montaison (TMAX\_B), un nombre de jours secs après apport azoté (DAI), un INN (NNI) et un apport azoté (N) plus faibles. Enfin **l'envirotype 4** connaît un plus grand nombre de jours secs après l'apport azoté (DAI) tandis que **l'envirotype 5** montre un nombre de jours de vernalisation (VERN) optimal moins important. Pour finir, l'envirotype 5 présente le profil le plus proche du réseau #2 moyen.

Les envirotypes 1 et 2 (fort rendement et fort INN) ont en commun l'indicateur de vernalisation plus élevé que la moyenne du réseau. Ce lien laisse penser que plus l'indicateur de vernalisation est important, plus le rendement final est élevé. Les envirotypes 1 et 3, avec un profil de rendement et d'INN opposé se distinguent par des indicateurs ayant des comportements opposés par rapport au réseau moyen. Ainsi TMAX\_FLO, HT\_FLO et TMAX\_B sont opposés entre les deux envirotypes. On peut alors émettre que l'hypothèse qu'un meilleur rendement est possible avec une température plus élevée durant la montaison et plus faible pendant la floraison.



## II.5. Conclusion

---

Malgré les différences entre les deux réseaux expérimentaux analysés (taille et conduite azotée), nos résultats montrent l'importance de la température moyenne hivernale (TMN\_CW) et des conditions de vernalisation (VERN). De plus, les données pointent des stress thermiques et radiatifs pendant la période de floraison comme facteurs limitant le rendement chez le colza d'hiver quel que soit le régime de nutrition azotée. L'identification de ces facteurs s'explique par le fait que l'absence ou la mauvaise qualité de la vernalisation peut entraîner une perte de rendement causée par un retard ou une absence de floraison (Ferreira et al., 1995 ; Chandler et al., 2005). De plus un stress thermique pendant la floraison peut impacter la fertilité des fleurs mais aussi le nombre de grains et le nombre de siliques par plante (Morrison, 1993 ; Angadi et al., 2000 ; Young et al., 2004). Enfin les deux études montrent un faible rôle des stress hydriques par rapport aux stress thermiques, majoritaires dans les deux études, bien qu'ils soient plus marqués dans le réseau #2.

Le réseau #2 a permis d'apporter une année de culture supplémentaire (2017-2018) ainsi que des essais présentant une fertilisation azotée différenciée. L'envirotypage met en évidence un stress thermique lors de la montaison (TMAX\_B) mais également un stress hydrique plus précoce et plus sévère que dans le réseau #1 (WD\_P600 dans le réseau #2 contre WS\_P1000 dans le réseau #1). Trois indicateurs azotés ont également été identifiés (NNI, N et DAI) confirmant l'impact d'une limitation en azote sur l'élaboration du rendement chez le colza (Rathke et al., 2006). L'étude de ce réseau #2 a également illustré la difficulté de mettre en place un stress azoté en conditions de plein champ (cf. II.2) et confirme l'impact de stress abiotiques multiples sur le rendement. L'envirotypage permet alors de regrouper les différents essais en fonction des combinaisons de stress réellement subis et constitue une approche sans *a priori* sur l'étude de l'impact de stress multiples sur une culture.

Par ailleurs en plus de mettre en évidence les différents stress et leurs impacts sur l'élaboration du rendement, l'envirotypage a permis de faire un premier pas dans la compréhension de l'interaction G×E. En effet, l'analyse des variances du réseau #1 (cf. II.3) révèle que l'envirotypage permet de prendre en compte une grande part de l'effet environnemental et dans une moindre mesure de réduire la part de variance de l'interaction G×E au sein de chaque envirotype. L'analyse des variances n'a pas été conduite sur le réseau #2 car il présentait un set trop limité de géotypes communs présents dans tous les essais pour tester correctement l'effet géotype et l'effet de l'interaction G×E.

En conclusion, ces résultats ouvrent de nouvelles perspectives pour l'identification des déterminants génétiques de l'interaction G×E en imaginant des détecteurs de QTL au sein de chaque envirotype. Dans le chapitre III, nous proposons d'utiliser les facteurs limitants identifiés sur le réseau #2 afin



d'effectuer un nouvel envirotypage basé uniquement sur les essais du réseau #2 ayant été conduits avec le panel de génétique d'association. Ce sous-réseau est appelé réseau #3 ou réseau GWAS (Tableau II.1). Ces résultats pourront par la suite être exploités pour identifier des déterminants génétiques de la stabilité du rendement chez le colza d'hiver.



## CHAPITRE III

---

# Architecture génétique de la stabilité du rendement chez le colza

---



## III.1. Intérêt de l'approche d'envirotypage pour caractériser les interactions QTL×Essai

---

Différentes méthodologies sont décrites pour investiguer l'interaction QTL×Essai (QTL×T ; Trial) et certaines d'entre elles ont été présentées dans la synthèse bibliographique (cf. I.3.4). Parmi ces méthodes on retrouve classiquement la comparaison des QTL détectés à différentes échelles (par essai vs. à l'échelle globale du réseau). Cette méthode a notamment été mise en œuvre sur le colza pour identifier des loci stables vs. des loci interactifs (Bouchet et al., 2014 ; 2016a). Cependant dans ces approches, la spécificité des QTL est peu mise en relation avec une caractérisation des environnements qui permettrait d'émettre des hypothèses sur les fonctions biologiques de l'adaptation spécifique des génotypes aux contraintes abiotiques subies. Des études sur maïs et blé (Millet et al., 2016 ; Touzy et al., 2019) ont proposé d'utiliser des regroupements d'environnements selon des scénarios climatiques (stress hydrique, stress thermique) et qui permettent d'expliquer une part de l'interaction G×E présente à l'échelle du réseau. Des analyses QTL menées à l'échelle de ces groupes d'environnements ont permis de mettre en évidence une spécificité de certains QTL à des scénarios climatiques.

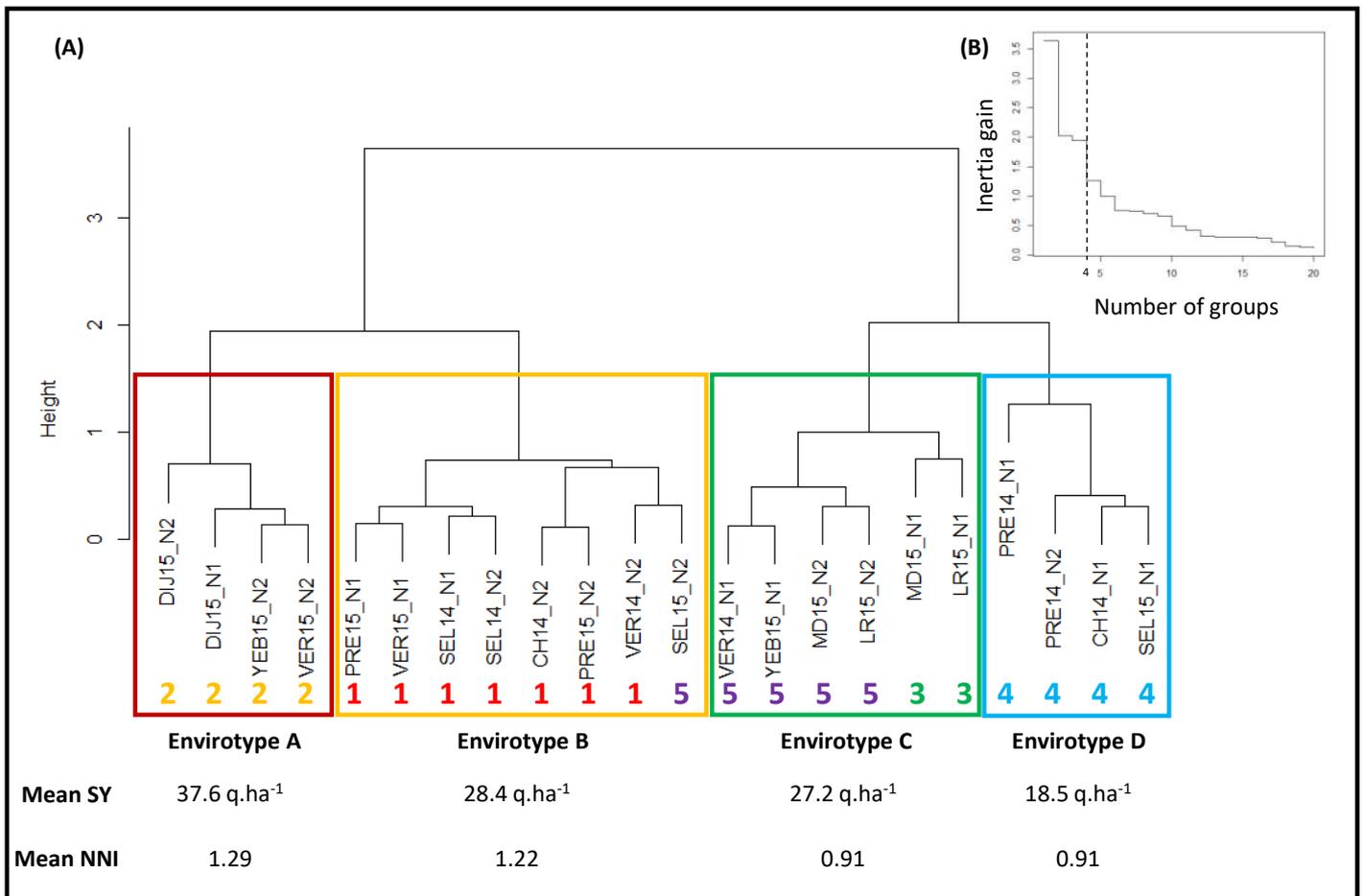
Au cours du chapitre précédent, nous avons mené une caractérisation des essais et identifié les facteurs limitant le rendement. Les environnements ont ensuite été regroupés en envirotypes en fonction de leur profil de facteurs limitants. L'envirotypage a permis d'expliquer une part importante de l'effet de l'environnement ainsi qu'une part de l'interaction G×T à l'échelle du réseau. L'objectif de ce chapitre est de valoriser cette compréhension de l'effet « essai » (T) et des interactions G×T pour étudier le déterminisme génétique de l'interaction G×T, et plus précisément pour identifier des QTL impliqués dans cette réponse différentielle aux variations du milieu. Il s'agira de différencier des QTL stables, *i.e.* détectés dans un grand nombre d'essais, des QTL interactifs, *i.e.* impliqués dans l'interaction G×T et spécifiques de certaines typologies d'essais et donc a priori de certains envirotypes. Pour ces QTL interactifs, on regardera les effets alléliques afin de tenter une qualification des interactions (cf. I.3.4).

### *Stratégie*

Ainsi, nous avons utilisé une démarche d'envirotypage afin d'identifier les déterminants génétiques de l'interaction G×T pour le rendement (SY), le nombre de grains (SN) et le poids de mille grains (TSW). Une détection QTL par GWAS a été réalisée pour trois niveaux d'intégration. Un premier



niveau a pour objectif d'identifier les QTL stables, pour cela il a consisté à utiliser les estimations des effets génotypiques à l'échelle de l'ensemble du réseau. Le second niveau a consisté en des analyses d'association à l'échelle des envirotypes. Ces QTL ont ensuite été mis en relation avec la typologie des envirotypes basée sur des facteurs limitants. Enfin, le troisième niveau a adressé la recherche de QTL à l'échelle de chaque essai individuel du réseau pour identifier les QTL spécifiques de certains essais qui contrôleraient la part de l'interaction G×T résiduelle au sein de chacun des envirotypes. Cette approche devrait aussi permettre d'identifier des QTL communs à plusieurs envirotypes mais dont les effets changent en fonction de l'environnement. Enfin, la comparaison des différentes approches nous permettra de prioriser certains QTL et de discuter de leur intérêt en sélection.



**Figure III.1: New envirotypes dedicated to the network #3**

(A) Dendrogram of the 22 trials from the GWAS network #3 based on the PLS regression results. Mean seed yield (SY) and mean nitrogen nutrition index (NNI) values are indicated for each envirotype. Numbers (1-5) under the trial names indicate the relationships between the envirotypes related to network #2 (cf. Figure II.6) and the current envirotypes. (B) Variation of the inertia gain depending on the number of groups chosen for the hierarchical clustering.

## III.2. Un nouvel envirotypage dédié pour les analyses génétiques

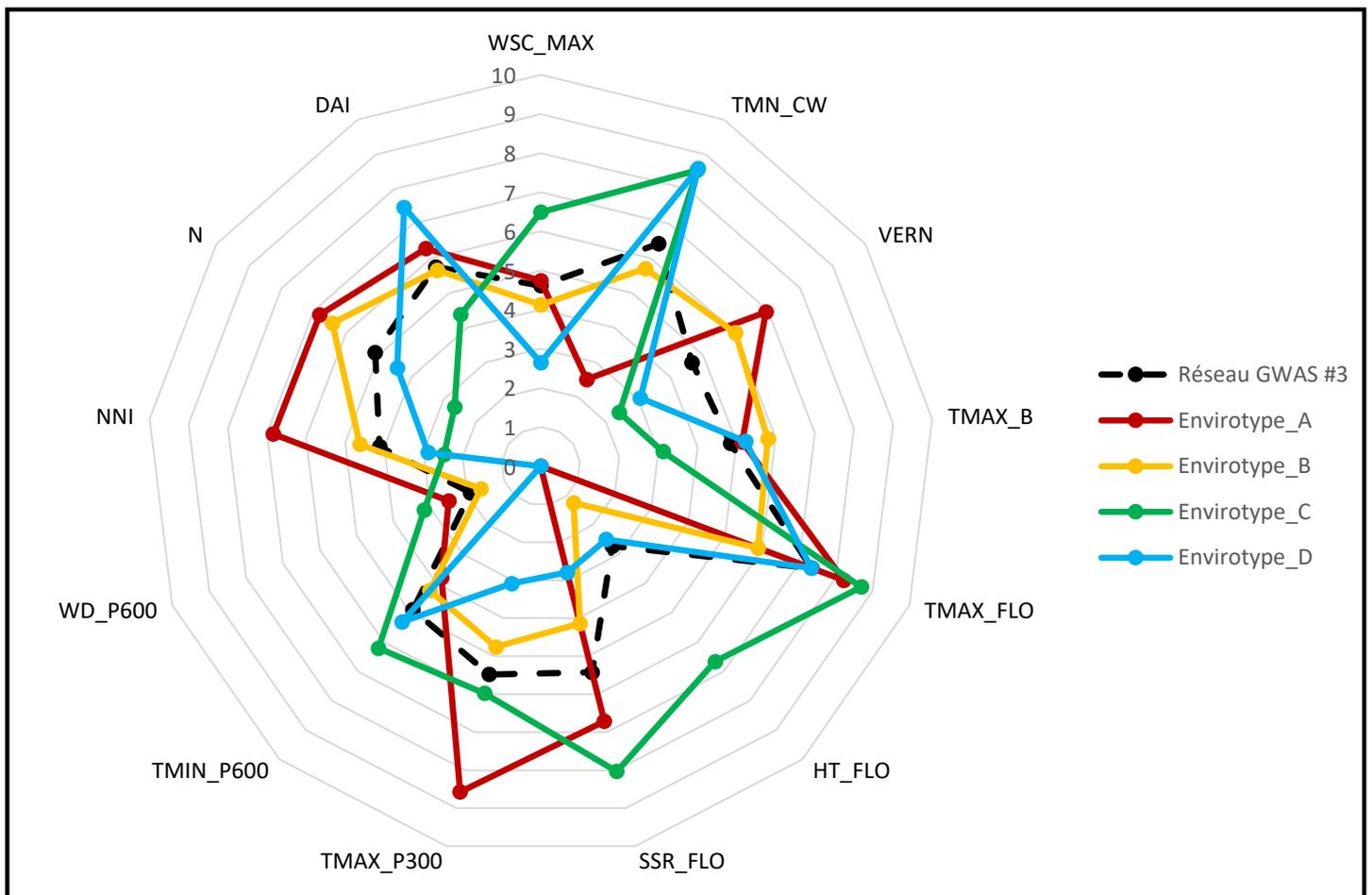
---

Au cours du chapitre II, l'impact de l'envirotypage sur l'explication de l'interaction G×T n'a pu être testé pour le réseau #2 du fait d'un trop faible nombre de génotypes communs sur l'ensemble des essais du réseau. Le réseau GWAS #3 a été défini comme un sous-réseau du réseau #2 au sein duquel une population de 173 génotypes (P173) a été expérimentée. Le réseau #3 (Tableau II.1) est composé de 22 essais répartis de la manière suivante : 11 environnements (soit 11 essais en N1 et 11 essais en N2) sur deux années de culture, 2014 (4 lieux : Châteauroux, Ch14 ; Prêmesques, Pre14 ; Selommes, Sel14 ; Verpillères, Ver14) et 2015 (7 lieux expérimentés : Dijon, Dij15 ; Le Rheu, LR15 ; Mondonville, Md15 ; Prêmesques, Pre15 ; Selommes, Sel15 ; Verpillères, Ver15 ; Yèbles, Yeb15). Ce réseau sera utilisé à la fois pour évaluer l'impact de l'envirotypage sur l'effet de l'essai et sur la décomposition de l'interaction G×T ainsi que pour les analyses génétiques.

Afin de décomposer les effets de l'essai (T) et de l'interaction G×T identifiés à l'échelle du réseau #3, nous choisissons d'utiliser la même méthodologie que celle développée dans le chapitre précédent pour le réseau #1. L'envirotypage des 22 essais du réseau #3 est réalisé à partir des facteurs limitants identifiés préalablement pour le réseau #2 puisqu'il inclut le réseau #3. L'objectif de ce nouvel envirotypage est de redéfinir les groupes d'essais à l'échelle du jeu de données disponible pour quantifier et caractériser les interactions G×T. On attend potentiellement une structuration plus fine du réseau #3, puisque des essais « extrêmes » du réseau #2 sont absents dans le réseau #3. La prise en compte de ces essais extrêmes aurait pu conduire à regrouper au sein d'un même envirotype des essais pourtant « contrastés » à l'échelle du réseau #3, et ainsi limiter la prise en compte des interactions G×T à l'échelle du réseau. Ce nouvel envirotypage permet de ne pas donner trop de poids à certains essais du réseau #3 qui se retrouveraient isolés dans un envirotype défini à l'échelle du réseau #2 et dont la caractérisation en termes de facteurs limitants serait principalement définie par les essais spécifiques au réseau #2 (Figure II.6).

### III.2.1. Définition de nouveaux envirotypes pour le réseau #3

Le nouveau clustering sur la base des 13 facteurs limitants identifiés sur le réseau #2 a permis d'identifier 4 envirotypes au sein du réseau #3 : A, B, C et D (Figure III.1). L'envirotype A est composé des essais Dij15\_N1, Dij15\_N2, Ver15\_N2 et Yeb15\_N2. Il présente un rendement moyen élevé (37,6 q/ha) et un INN moyen de 1,29. Cet envirotype est composé de 3 essais N2 et de l'environnement



**Figure III.2: Envirotype description based on the seed yield limiting factors**

The profile of each limiting factor is represented on a 0-10 scale for each envirotype as well as for the GWAS whole network. Abbreviations of the indicators are as following: TMN\_CW: mean temperature at climatic winter, Vern: vernalization condition, TMAX\_B: maximal temperature at bolting, HT\_FLO: number of days with high temperature (>25°C) at flowering, SSR\_FLO: sum of solar radiation at flowering, TMAX\_P300: maximal temperature during the seed number fixation period, TMIN\_P600: minimal temperature during the pod growth, WD\_P600: water deficit during the pod growth, NNI: Nitrogen Nutrition Index, N: Amount of Nitrogen bring to the trial, DAI: number of dry days after the N-input. WSC\_MAX: maximal water soil capacity.

Dij15 qui est caractérisé par des forts rendements quelle que soit la fertilisation azotée apportée (Dij15\_N1 et Dij15\_N2). Il est défini comme un envirotype globalement non stressé en azote et à fort rendement. A l'opposé, l'envirotype D est caractérisé par rendement moyen très faible (18,5 q/ha) et l'INN moyen de 0,91, ce qui suggère que les essais constitutifs de cet envirotype ont subi des stress plus marqués. Il est constitué de 3 essais N1 (Ch14\_N1, Pre14\_N1, Sel15\_N1) et d'un environnement (Pre14) caractérisé par de faibles rendements quelle que soit la fertilisation apportée (Pre14\_N1 et Pre14\_N2). Enfin les envirotypes B et C, composés respectivement de 8 essais (Ch14\_N2, Pre15\_N1, Pre15\_N2, Sel14\_N1, Sel14\_N2, Sel15\_N2, Ver14\_N2 et Ver15\_N1) et 6 essais (LR15\_N1, LR15\_N2, Md15\_N1, Md15\_N2, Ver14\_N1 et Yeb15\_N1) sont des envirotypes dont les valeurs moyennes de rendements (28.4 q/ha et 27.2 q/ha respectivement) sont similaires entre elles et intermédiaires des envirotypes A et D. Les envirotypes B et C se distinguent néanmoins par leur INN moyen observé (1,22 et 0,99, respectivement). Ces deux envirotypes, au statut azoté contrasté, ne présentent sans doute pas de stress azoté.

Lorsque l'on compare les envirotypes A-D du réseau #3 avec les envirotypes 1-5 du réseau #2, on observe que les deux envirotypes extrêmes A et D correspondent aux envirotypes extrêmes 2 et 4 du réseau #2 respectivement (Figure III.1). Tous les essais de l'envirotype B étaient groupés dans l'envirotype 1 du réseau #2, à l'exception de l'essai SEL15\_N2 (envirotype 5). Enfin l'envirotype C comporte 4 essais de l'envirotype 5 et 2 essais de l'envirotype 3. Ce nouvel envirotypage (réseau #3) permet de conserver une répartition des essais en envirotypes très similaire à l'envirotypage précédent (réseau #2), à l'exception de l'envirotype C qui est l'envirotype le plus remanié.

A l'instar de ce qui a été réalisé dans le chapitre II, les envirotypes A-D du réseau #3 ont été caractérisés en fonction des 13 facteurs limitants et comparés au réseau #3 moyen (Figure III.2). A un risque de 10%, pour l'envirotype A, les indicateurs NNI et TMAX\_P300 sont plus élevés que dans le réseau #3 moyen, tandis que TMN\_CW est inférieur. L'envirotype B présente un TMAX\_FLO plus faible que le réseau #3. L'envirotype C est caractérisé par les indicateurs HT\_FLO, SSR\_FLO et TMN\_CW plus élevés tandis que DAI, NNI, N, TMAX\_B et VERN sont plus faibles. Enfin l'envirotype D se distingue par un DAI plus élevé ainsi que SSR\_FLO et TMAX\_P300 plus faibles que dans le réseau #3. Ces résultats peuvent s'interpréter de la manière suivante : l'envirotype A, non limitant en azote, se caractérise ainsi par des températures post-floraison plus importantes. L'envirotype B est marqué par un déficit thermique à floraison. L'envirotype C est marqué par de fortes températures et un fort rayonnement à floraison ainsi que pendant l'hiver, mais également par de moins bonnes conditions de vernalisation et un stress azoté. Enfin l'envirotype D, à plus faible rendement, est marqué par un rayonnement et des températures plus faibles à floraison et post-floraison. On remarque que les



profils en facteurs limitants des envirotypes extrêmes, A et D, sont similaires aux profils des envirotypes 2 et 4 du réseau #2.

### III.2.2. Présentation des traits étudiés

Le rendement dépend d'un grand nombre de composantes (*cf.* I.2.1) telles que le nombre de grains par m<sup>2</sup> (SN) et le poids de mille grains (TSW). L'envirotypage ainsi que la détermination des facteurs limitants se sont basés uniquement sur les variations de rendement, et non des autres composantes. SY est fortement corrélé à SN ( $r \geq 0.85$  ; valeurs décrites dans la littérature) mais très faiblement corrélé à TSW ( $r \leq 0.07$ ). De plus la corrélation entre SN et TSW est généralement négative et peu significative ( $r \leq -0.32$ ) (Bouchet et al., 2014). Il y a donc un risque pour que l'envirotypage proposé ne soit pas adapté, notamment pour l'analyse du TSW. Il sera donc nécessaire au préalable de toute analyse génétique, de vérifier la pertinence de l'envirotypage pour l'analyse de chacun des traits en évaluant à l'aide d'un modèle linéaire la capacité de l'envirotypage à expliquer une part non négligeable de l'effet principal de l'essai et de l'interaction G×T.



## III.3. Analyses génétiques

---

La suite de ce chapitre est présentée en anglais, car elle fait partie d'un article en cours de préparation. Cependant les paragraphes ne suivent pas la construction classique d'un article afin d'éviter les redondances au sein du manuscrit de thèse. Le matériel et méthodes, les résultats et les éléments de discussion sont présentés ensemble.

### III.3.1. Plant material and genotyping data

A population of 173 WOSR accessions (hereafter referred to as P173) was trialed across the network's GWAS. The accessions originated primarily from Western Europe, and most of them represented commercial varieties released from 1959 to 2010. The P173 (Supplementary Table III.1) comprised 31 accessions of the “++” type (high contents in erucic acid, C22:1, and glucosinolates, GSL), 15 accessions of the “0+” type (low C22:1, high GSL), 1 accession of the “+0” type (high C22:1, low GSL), and 126 accessions of the “00” type (low C22:1, low GSL) of which 46 were elite varieties. Seeds were provided by the BrACySol Biological Resource Center or private companies. The whole population was genotyped using the Brassica 60K Illumina Infinium™ array (Clarke et al., 2016) and an exome sequence capture assay (Leveugle et al., 2015). A total of 182 215 SNP were scored and validated on P173 using a threshold of 2.5% for the minor allele frequency (MAF) and of 10% for the missing values. The missing genotyping data were inferred using Beagle v3 software (Browning and Browning, 2009). All the SNP were physically anchored onto the *Brassica napus* reference genome, Darmor-*bzh* version 5 (Chalhoub et al., 2014).

### III.3.2. Traits and phenotypic data analyses

#### *Correlations between phenotypic traits*

The targeted traits were i) the seed yield (SY in  $\text{q}\cdot\text{ha}^{-1}$ ) determined for each genotype in each trial and adjusted to 0% water content and 0% impurities, ii) the thousand seed weight (TSW in g) and iii) the seed number ( $\text{SN} = (\text{SY}\cdot 100\ 000)/\text{TSW}$ , in  $\text{seeds}\cdot\text{m}^{-2}$ ). The Pearson correlation coefficients between SY, SN and TSW were calculated. As expected, the results showed a high positive correlation between SY and SN ( $r = 0.94$ ). On the opposite, TSW was poorly correlated with SY ( $r = 0.08$ ) and SN ( $r = -0.24$ ). These results are consistent with those reported by Bouchet et al. (2014).



*Evaluation of the envirotyping impact on the Trial effect and the G×T interaction*

**METHODS**

For each trait, three different mixed linear models were fitted using the lme4 package (Bates et al., 2015) and lmerTest package (Kuznetsova et al., 2017) corresponding to the different scales of the study. The trial scale used the model (1), and the envirotype as well as the whole network scales used the model (3). In addition, the model (5) evaluates the impact of the envirotyping on the trial effect and G×T interaction and variance decomposition. These models are presented below.

A first linear model (1) was fitted at the level of each trial:

$$Y_{ij} = \mu + G_i + \underline{R_j} + \underline{\varepsilon_{ij}} \quad (1)$$

where  $Y_{ij}$  is the phenotypic value,  $\mu$  is the population mean,  $G_i$  stands for the effect of genotype  $i$ ,  $R_j$  for the replicate  $j$  and  $\varepsilon_{ij}$  is the residual. The replicate effect and the residual were declared as random. The heritability was calculated as:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_\varepsilon^2 / r} \quad (2)$$

where  $\sigma_G^2$  is the genetic variance,  $\sigma_\varepsilon^2$  the residual variance and  $r$  the number of replicate per genotype.

A second linear model (3) was applied at the envirotype or at the whole network scale:

$$Y_{ijk} = \mu + G_i + T_k + G_i \times T_k + \underline{R_{j(k)}} + \underline{\varepsilon_{ijk}} \quad (3)$$

where  $Y_{ijk}$  is the phenotypic value,  $\mu$  is the population mean,  $G_i$  stands for the effect of genotype  $i$ ,  $T_k$  for the trial  $k$ ,  $R_{j(k)}$  for the replicate  $j$ , nested in the trial  $k$  and  $\varepsilon_{ijk}$  is the residual. The corresponding heritability was defined as follow with all terms declared as random in equation (4):

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{G \times T}^2 / t + \sigma_\varepsilon^2 / r * t} \quad (4)$$

where  $\sigma_G^2$  is the genetic variance,  $\sigma_{G \times T}^2$  the G×T interaction variance,  $\sigma_\varepsilon^2$  the residual variance,  $t$  the number of trials and  $r$  the number of replicate per genotype.

A last linear model (5) was fitted with an envirotype effect to evaluate the impact of the envirotyping on the trial effect and the G×T interaction:

$$Y_{ijkl} = \mu + G_i + En_l + En_{l(k)} + G_i \times En_l + G_i \times En_{l(k)} + \underline{R_{j(l \times k)}} + \underline{\varepsilon_{ijkl}} \quad (5)$$

**Table III.1: Heritabilities estimated for each trait and each trial.** Components of the heritability were obtained using the linear model (1) :  $Y_{ij} = \mu + G_i + R_j + \varepsilon_{ij}$ ; Heritabilities ( $h^2$ ) were calculated according to equation (2) :  $h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{\varepsilon}^2/r}$

Envirotype	Trial	$h^2$		
		SY	SN	TSW
A	Dij15_N1	0.9	-	-
	Dij15_N2	0.91	-	-
	Ver15_N2	0.91	-	-
	Yeb15_N2	0.92	0.93	0.93
	Ch14_N2	0.93	0.93	0.91
	Pre15_N1	0.81	0.85	0.92
	Pre15_N2	0.8	0.85	0.93
	Sel14_N1	0.94	0.93	0.71
	Sel14_N2	0.95	0.93	0.8
	Sel15_N2	0.63	-	-
B	Ver14_N2	0.91	0.9	0.95
	Ver15_N1	0.86	-	-
	LR15_N1	0.85	0.88	0.97
	LR15_N2	0.87	0.9	0.97
	Md15_N1	0.67	0.53	0.76
	Md15_N2	0.74	0.65	0.75
	Ver14_N1	0.77	0.71	0.95
	Yeb15_N1	0.95	0.96	0.96
	Ch14_N1	0.81	0.78	0.91
	Pre14_N1	0.68	0.75	0.93
C	Pre14_N2	0.87	0.87	0.91
	Sel15_N1	0.96	-	-
D				

where  $Y_{ijkl}$  is the phenotypic value,  $\mu$  is the population mean,  $G_i$  stands for the effect of genotype  $i$ ,  $En_l$  for the envirotypic effect  $l$ ,  $En_{l(k)}$  for the trial  $k$  nested in the envirotypic effect  $l$ ,  $R_{j(l \times k)}$  for the replicate  $j$ , nested in the trial  $k$  and envirotypic effect  $l$ ; and  $\varepsilon_{ijkl}$  is the residual

## **RESULTS**

At the **trial scale**, heritabilities ranged from 0.63 to 0.96 for SY, from 0.53 to 0.86 for SN and from 0.71 to 0.97 for TSW (Table III.1).

The linear model (3) was used to evaluate the trial effect and the genotype by trial interaction (G×T) at the **network scale** (Table III.2). The results showed a high trial effect for SY and SN (55.5% and 40.4% of total variance respectively) and a G×T interaction of 5.8% for SY and 8.1% for SN. For TSW, trial effect counted only for 21.3% of the variation and the G×T interaction for 12.5%. All traits were heritable at the network scale ( $h^2=0.98$  for each trait) (Table III.2).

### ***Comparison of the G×T effect at the network scale and within each envirotypic effect***

The genetic effect for SY and SN increased in comparison to the network scale for most of the envirotypes. And consequently, the trial effect (in variance proportion) decreased for most of the envirotypes when compared to the network scale (Table III.2). The same results were observed for TSW except for the trial effect, which was lower than for SY and SN. These changes of the trial and genotypic effects may partly reflect the reduction of the trials number. Heritabilities varied from 0.83 to 0.96 for SY, from 0.81 to 0.96 for SN and from 0.82 to 0.94 for TSW (Table III.2)

For SY, the G×T interaction within each envirotypic effect ranged from 2.3% to 15.9% of the total variance. Only envirotypic effect A presented a lower proportion of variance dedicated to the G×T effect than the proportion observed at the network scale. Envirotypes B and D presented a higher proportion of G×T variance (Table III.2). For SN, the G×T variances ranged from 8.1% to 11.5% and were higher or equal to the variance observed in the network. Finally, for TSW, the G×T variances ranged from 7.4% to 19.3%, envirotypes A and C presented a higher G×T variance than the network and envirotypic effect B and C a lower G×T variance.

### ***Comparison of the partition of the G×T variance between G×En and G×En×T***

The linear model (5) was used to evaluate the envirotyping effect (Table III.3). Except for the TSW, the envirotypic effect explained a main part of the trial effect (70% for SY and 74% for SN). The remaining trial effect nested to the envirotypic effect (Envirotypic effect × Trial effect) represented less than 20% of the total variance at the network scale. Concerning the G×T interaction, it was split between a

**Table III.2: Results of the mixed linear model applied for the whole network and for each envirotyping considering all trials as confounded (linear model (3))**

$Y_{ijk} = \mu + G_i + T_k + G_i \times T_k + \frac{R_{j(k)}}{r} + \varepsilon_{ijk}$  ; with the genotype (G), the trial (T), the genotype by trial interaction (GxT) and the block (R) nested in the trial.

Trait	Group	Number of trials <sup>(a)</sup>	G	T	GxT	R	$\varepsilon$	$h^2$ <sup>(c)</sup>
			% Var <sup>(b)</sup>	% Var	% Var	% Var	% Var	
<b>SY</b>								
	GWAS network	22	24.7	55.5	5.8	3.8	10.1	0.98
	Envirotyping A	4	37.7	48.3	2.3	3.8	7.8	0.96
	Envirotyping B	8	38.5	34	8.7	4.8	14.1	0.95
	Envirotyping C	6	45.7	8.7	5.9	10.8	29	0.94
	Envirotyping D	4	28.9	34.8	15.9	5.3	15	0.83
<b>SN</b>								
	GWAS network	18	33	40.4	8.1	4.9	13.5	0.98
	Envirotyping A	2	46.9	34.1	8.1	1.1	9.9	0.88
	Envirotyping B	7	49	27	9.4	2.1	12.5	0.96
	Envirotyping C	6	43.8	6.9	9.1	13.7	26.5	0.93
	Envirotyping D	3	29.6	32.5	11.5	7.4	19	0.81
<b>TSW</b>								
	GWAS network	18	52	21.3	12.5	1.6	12.5	0.98
	Envirotyping A	2	61.9	2.9	19.3	0	15.8	0.82
	Envirotyping B	7	57.7	15.5	7.4	3	16.4	0.86
	Envirotyping C	6	43.3	32.2	13.6	0.8	9.2	0.94
	Envirotyping D	3	71.2	1.3	10.5	2.3	14.7	0.92

(a) Number of trials considered for the evaluation of the different effects

(b) Percentage of variance explained by each effect

(c) Heritabilities ( $h^2$ ) were calculated according to equation (4) :  $h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{G \times T}^2 / t + \sigma_\varepsilon^2 / r * t}$

**Table III.3: Results of the mixed linear model applied for the whole network to evaluate the impact of the envirotyping using linear model (5)**

$Y_{ijkl} = \mu + G_i + En_l + En_{l(k)} + G_i \times En_l + G_i \times En_{l(k)} + \frac{R_{j(l \times k)}}{r} + \varepsilon_{ijkl}$  ; with the genotype (G), the envirotyping (En), the trial nested in the envirotyping ( $En_{l(k)}$  or  $En \times T$ ), the genotype by envirotyping interaction (GxEn), the genotype by trial by envirotyping interaction ( $G_i \times En_{l(k)}$  or  $G \times En \times T$ ) and the block nested in the trial, nested in the envirotyping (R)

Trait	G	En	EnxT	GxEn	GxEnxT	R	$\varepsilon$
	% Var <sup>(a)</sup>	% Var	% Var	% Var	% Var	% Var	% Var
SY	20	44.4	19.2	1.2	3.9	3.1	8.2
SN	26.8	37.6	13.4	1.1	5.8	4.1	11.1
TSW	51.5	1.7	19.9	1.4	11.4	1.6	12.4

(a) Variance of each effect

genotype × envirotype (G×En) interaction and a remaining G×T inside each envirotype (G×En×T). The G×En explained around 1% of the total variation for all traits and the G×En×T has been reduced for each trait when compared to the G×T at the network scale (3.9% for SY, 5.8% for SN and 11.4% for TSW) but remained important (Table III.2). Finally, the G×En interaction explained 24.6% of the G×T interaction for SY, 16.7% for SN and 11.8% for TSW.

## **DISCUSSION**

The envirotyping allowed reducing drastically the remaining effect of the trial and also impacted the G×T interaction repartition, which is consistent with previous work (Corlouer et al., 2019) where a part of the network #3 and P173 was included. As the envirotypes were also characterized according to their pattern of limiting factors, biological hypotheses can be proposed to explain the phenotypic variations due to the trial effect and G×T interaction.

The use of envirotypes defined according to SY variations appeared to be valuable to study G×T interaction for SN, as the variance repartition showed a similar pattern as observed for SY. However for TSW, the results were different probably because the limiting factors identified using SY variations were not relevant to explain TSW variations and G×T interaction for that trait. To work further on the genetic determinant of the interaction for TSW, a new envirotyping analysis should be performed. For the rest of this study, the genetic analyses for TSW will be carried out only at the network and trial scales in order to identify the stable and interactive QTL of TSW. For SY and SN, genetic analyses will be carried out at the network, the trial and the envirotype scales.

### **III.3.3. GWAS analysis**

#### *GWAS method*

For each trait, each genotype and each trial, the Best Linear Unbiased Estimates (BLUE) were extracted from model (1). For each envirotype and for the whole network, the linear model (3) was used to estimate the BLUE of each genotype. The BLUE defined at each scale (trial, envirotype, network) were used to perform the GWAS analyses.

GWAS analyses were conducted using the FastLMM algorithm (Lippert et al., 2011) using a kinship calculated with the Astle algorithm (Astle and Balding, 2009) for each linkage group as described by Rincent and coworkers (2014). The threshold to set a significant associated SNP corresponded to a Bonferroni threshold of 10% based on a corrected SNP population as proposed by the simpleM method developed by Gao et al. (2008). This method was based on the composite linkage

**Table III.4: List of QTL detected at the different scales**

QTL name	Trait(s)	Chromosome	Start (pb)	End (pb)	Trial(s) <sup>(a)</sup>	Envirotype(s) <sup>(b)</sup>	GWAS Network <sup>(c)</sup>
A03a	SY	A03	2 180 395	2 180 405	Sel14_N2	-	-
A03b	SY	A03	5 607 548	5 607 548	Pre15_N2 / Ver15_N2	B	-
A03c	SY	A03	20 719 243	20 758 120	Pre15_N2	-	-
A04a	SY	A04	7 408 414	7 408 414	Ch14_N2	-	-
A04b	SY	A04	8 280 430	8 280 430	Ch14_N2	-	-
A05a	SY	A05	1 939 257	1 941 241	Ch14_N2	-	-
A05b	SN	A05	3 335 006	3 335 006	Pre14_N2	-	-
A07a	SN	A07	244 837	282 933	Pre14_N1 / LR15_N2	D	-
A07b	SY/SN	A07	754 861	829 734	Pre14_N1 / LR15_N2	-	-
A07c	SN	A07	1 127 600	1 144 941	LR15_N2	-	-
A07d	TSW	A07	17 323 571	17 391 242	Md15_N2 / Ver14_N2	-	Yes
A09	SY/SN	A09	2 517 797	3 051 850	Ch14_N2 / Yeb15_N2 / Pre14_N2 / Ver15_N2 / Ch14_N1 / Ver14_N2	A/D	Yes
A10a	TSW	A10	1 864 577	1 865 963	Ch14_N2 / Pre14_N1	-	-
A10b	TSW	A10	2 067 138	2 115 740	Ch14_N1 / Ch14_N2 / Pre14_N1 / Ver14_N1 / Ver14_N2	-	-
A10c	TSW	A10	2 416 579	2 505 644	Ch14_N2 / Ver14_N1	-	-
A10d	TSW	A10	2 696 446	2 696 612	Ver14_N1	-	-
A10e	TSW	A10	15 639 575	15 639 575	Ch14_N1	-	-
A10f	TSW	A10	16 256 530	16 256 530	Ver14_N1	-	-
A10g	SY	A10	16 721 480	16 721 480	Sel14_N2	-	-
C02a	TSW	C02	<sup>24</sup> 973	34 259	Pre14_N1 / Ver14_N1	-	-
C02b	SY	C02	45 059 324	45 059 324	Yeb15_N2 / Ver14_N2	-	-
C03a	SN	C03	20 039 808	20 039 808	-	C	-
C03b	SY	C03	30 194 954	30 195 362	Ver14_N1	-	-
C04	SY	C04	4 063 677	4 063 677	Dij15_N2 / Pre15_N1	-	-
C08	SY	C08	35 215 389	35 215 692	Ver14_N2	-	-
C09a	SY/SN	C09	816 997	2 784 948	Md15_N2 / Sel14_N1 / Ver14_N2 / Ver15_N2 / Dij15_N1 / Ch14_N2 / Yeb15_N2 / Ch14_N1 / Ver14_N1 / LR15_N2	A/B/C	-
C09b	TSW	C09	4 463 887	4 466 434	Md15_N1	-	-
C09c	SY	C09	38 469 046	38 469 046	Ver14_N1	-	-

(a) Trials where the QTL were detected. Colors are given upon the corresponding envirotype with Envirotype A in red, envirotype B in yellow, envirotype C in green and envirotype D in blue (cf. Figure III.1).

(b) Envirotypes where the QTL were detected.

(c) Indicate whether the QTL was detected (or not) at the network scale.

disequilibrium (CLD) correlation between SNP. The CLD was used to calculate the effective number of independent test (note  $M_{eff}$ ) (Gao et al., 2008). For our study, the  $M_{eff}$  was calculated and fixed to 17 882 SNP. We defined the threshold of highly significant associated SNP as:

$$t = -\log_{10}(\alpha_G/M_{eff}) \quad (6)$$

where  $t$  is the value of the threshold,  $\alpha_G$  is the threshold of the  $p$ -value and was fixed to 0.1 and  $M_{eff}$  is the effective population of SNP. Finally, we defined a QTL as an area of one or multiple associated SNP with a maximal distance between two SNP inferior to the LD decay (calculated for  $r^2=0.2$ ) of each chromosome (Supplementary Table III.2).

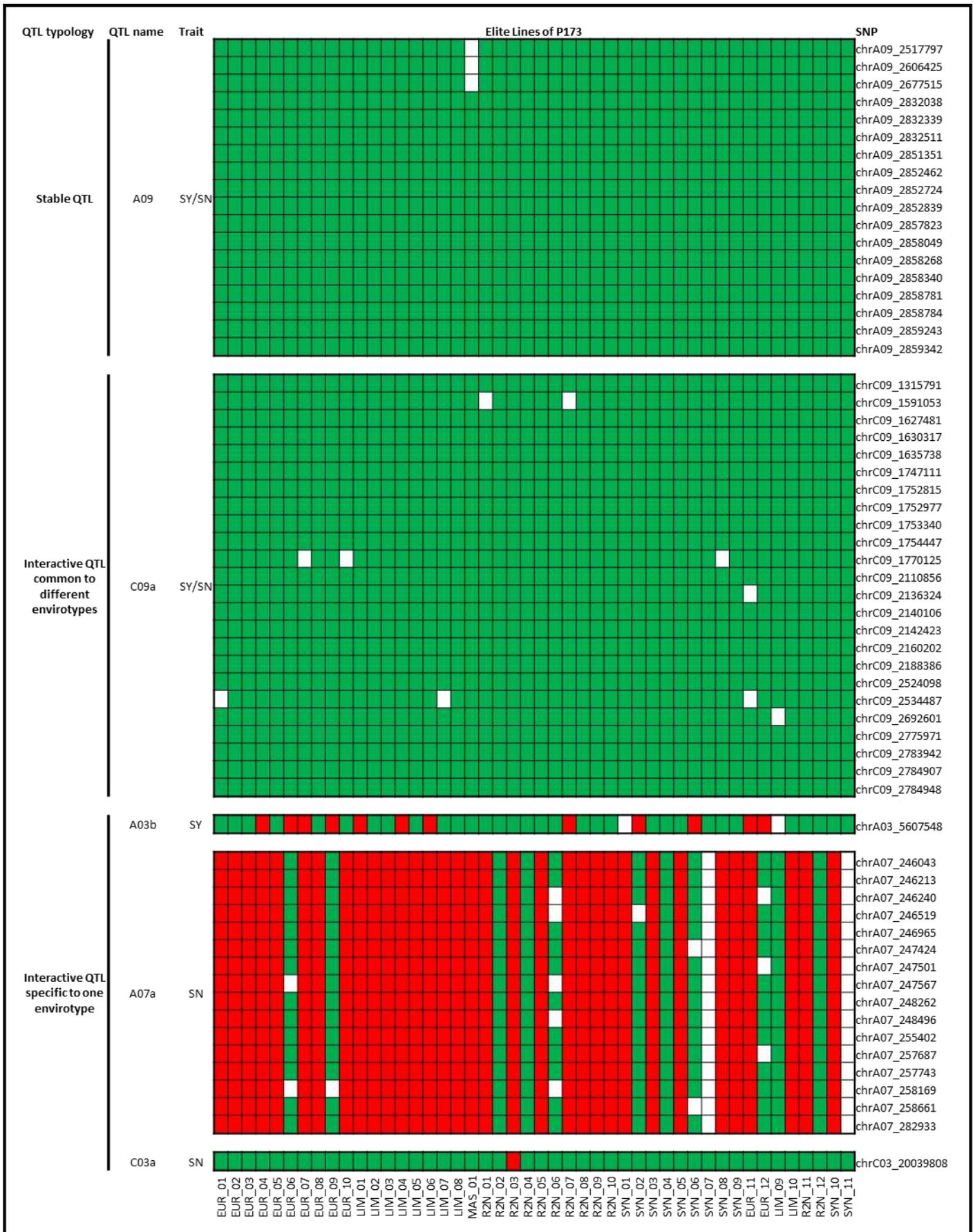
### *GWAS results at the network scale*

At the network scale, only two QTL were identified: one QTL for SY on chromosome A09 between 2 832 038 and 2 859 342 bp that explained 13% of the variation. The second QTL was detected for TSW on chromosome A07 (A07d) between 17 323 571 and 17 326 232 bp and explained 11% of the variation (Table III.4). For both QTL, the favorable allele, which increased the trait value, was the major allele in the population (Supplementary Table III.3.a).

### *GWAS results at the trial scale*

The GWAS analyses conducted for each trial let to identify 27 QTL for all traits across 15 different trials (Table III.4). Among these QTL, 12 were specific to SY (3 on A03, 2 on A04, 1 on A05, 1 on A10, 1 on C02, 1 on C03, 1 on C04, 1 on C08 and 1 on C09). For SN, 3 QTL were defined (1 on A05, 2 on A07). Nine QTL were detected for TSW (1 on A07, 6 on A10, 1 on C02 and 1 on C09). Finally, 3 QTL were identified for both SY and SN (1 on A07, 1 on A09 and 1 on C09). QTL on A09 and C09 for SY/SN presented homeologous relationship. Each SY QTL explained between 10-15% of the variation, for SN QTL explained between 17-23% each and for TSW, QTL explained between 9-17%. All traits confounded, for 35.3% of the SNP identified, the favorable allele was the minor allele, and consequently, for 64.4% of the SNP, the favorable allele was the major one (Supplementary Table III.3.c).

No QTL was identified for 7 trials out of the 22 trials of the network. Among the QTL detected, 59.3% were specific to a single trial, 29.6% were detected in 2 trials, 3.7% in 5 trials, 3.7% in 6 trials and 3.7% in 10 trials.



**Figure III.3: Distribution of the favorable alleles in the elite varieties of the P173 population for QTL detected at the envirotype scale**

Associated SNPs are in lines with SNP names corresponding to the chromosome and position in bp on the *Darmor-bzh* V.5 reference assembly. Genotypes are in column. Favorable alleles are indicated in green and unfavorable ones in red. White cases correspond to missing genotypic values.

### GWAS results at the envirotype scale

At the envirotype scale, across all envirotypes and all traits, 5 QTL were identified (Table III.4). Among these QTL, one was specific to SY (A03b) and was detected only in envirotype B, two were detected for SN : A07a and C03a in envirotypes D and C respectively. Finally, two QTL were detected for both SY and SN on chromosome A09 (envirotypes A and D) and C09 (C09a, envirotypes A, B and C) and corresponded to the QTL already identified at the trial scale. Each SY QTL explained between 11% and 14% of the variation, SN QTL explained between 18% and 21 % of the variation (Supplementary Table III.3.b ; Supplementary Figure III.3.1).

Comparison between QTL detected at the envirotype scale for each trait showed that 3 QTL were specific to a single envirotype, 1 was common to 2 envirotypes (A09, envirotypes A and D) and 1 QTL was common to 3 envirotypes (C09a, envirotypes A, B and C). No QTL were identified across all envirotypes. As each trial belonged to a single envirotype, we can connect the specificity of the SY and SN QTL detected at the trial scale to an envirotype. Among all QTL detected at the trial scale, we showed that 11 QTL were specific to a single envirotype, 5 QTL were identified in trials that belong to two different envirotypes, 1 QTL in trials from 3 envirotypes and 1 QTL was identified in trials belonging to all the envirotypes. However, this last QTL (C09a) was not detected at the network scale.

#### III.3.4. Comparisons between the three different scales

At the network scale, only two QTL were detected (A07d and A09). The QTL A07d detected for TSW was also detected in two trials. The QTL A09 were also detected at the envirotype scale (envirotypes A and D) and in different trials (6 trials for A09 belonging to the envirotypes A, B and D). The QTL A09 was detected at all scales (trial, envirotype and network). The favorable allele at this QTL was present in all elite lines (Figure III.3). These two QTL (A07d and A09) can be qualified as stable loci since as they are detected regardless of the envirotype or trial. According to the definition of “conditional neutrality” QTL (Des Marais et al., 2013), they were detected only in some trials/envirotypes, but always with the same effect, then they can as well be considered as interactive. However, conditional neutrality could also be considered as a special situation of the differential sensitivity. The distinction between these concepts can't be made with our results.

The envirotype scale allowed detecting 4 new QTL in comparison to the ones already detected at the network scale; namely QTL A03b, A07a, C03a and C09a. Except for QTL C03a, the QTL were also identified in the corresponding trials. However, these QTL were also sometimes detected in trials that did not belong to the considered envirotype. This latter result suggests that the envirotyping did



not explain all the G×T interaction as confirmed by the results of the linear model (5). The QTL C03a detected for SN was specifically retrieved at the envirotype scale and was detected only for envirotype C suggesting that envirotyping is useful to highlight interactive QTL. The QTL C09a was detected in 3 envirotypes and the favorable allele was the major allele in the P173 population (MAF=0.25). The  $r^2$  of the QTL were not significantly different between envirotypes, thus leading us to interpret this QTL as interactive with a conditional neutral effect.

Analysis of the gene content underlying this locus (QTL C09a) revealed genes involved in the flowering initiation with 7 genes in relation to *FRIGIDA* (Johanson et al., 2000) and two genes encoding flowering promoters (BnaC09g04240D and BnaC09g04280D). However, no genes related to abiotic stress response was identified with this study (data not shown). In addition, an homeology relationship was identified between this QTL and the QTL A09 already characterized. The typology of these two QTL (A09 and C09a) must be further investigated. One way to confirm their stability/interactivity can be to test the interaction between the QTL and the main limiting factors of each corresponding envirotype as proposed by Malosetti et al. (2004).

At the trial scale, 22 new QTL were detected of which 5 were already detected at the envirotype/network scales. These new QTL were also mostly specific to a single trial. This specificity is a first estimation of the QTL×Trial interaction in a network (Garin et al., 2020).

### III.3.5. Getting deeper into the use of the envirotype QTL

For the three main QTL defined as interactive (A03b, A07a and C03a) that were specific to a single envirotype, the allelic distribution was investigated in the P173 population, with a focus made on the allelic combinations occurring in the 46 elite lines, in order to address whether the favorable alleles had already been fixed or not.

For QTL A03b that was specific of the envirotype B, the favorable allele was fixed in most of the elite lines except for 12 of them (Figure III.3). Interestingly, 6 out of the 12 correspond to the germplasm of a unique breeding company (Euralis). We can hypothesize that the target population of environments chosen by Euralis for their breeding programs are absent or underrepresented in the network #3 GWAS. The gene content analysis within the vicinity of that locus did not reveal any obvious genes in relation to a response to any pedoclimatic factors.

For QTL A07a, the favorable allele was only present in 24.4% of the elite lines (Figure III.3). As a matter of fact, this QTL appears to be a very good candidate for breeding as most of the elite lines could benefit from QTL introgression to improve SN and eventually SY. In addition, this QTL is



explains 19% of the phenotypic variability and is specific of the envirotype D, which is the most stressed envirotype of the network. This envirotype might have been under investigated up to now by breeding programs. The main limiting factors identified for the envirotype D were the lack of radiation and the low temperature at flowering stage.

Finally, the favorable allele at QTL C03a was fixed in the elite lines of the P173 population (Table III.3).

Interactive QTL had particular interest, especially when compared to stable QTL. Here for the stable QTL C09a, all elites lines had already fixed the favorable allele reflecting the fact that breeding programs have already been able to valorize stable seed yield QTL which effects was detected in a large number of trials. In comparison with a classical QTL analyses, our interactive QTL A07a was not detected (Bouchet et al., 2016a).

To conclude, the QTL detected at the envirotype scale could be used to improve seed yield related traits. Envirotypes can be viewed as environmental scenarios that affected seed yield. For instance, the A07a QTL detected for SN is a good candidate for marker assisted selection to improve QY in low N-environments, with low solar radiation and low temperature during flowering. The envirotype D related scenario must be taken into account in the design of the TPE of each breeding company or during registration process.



Supplementary Table III.1: List of the genotypes of the population P173

ACCESSION NAME	TYPE	ACCESSION NAME	TYPE	ACCESSION NAME	TYPE
Alsace 1360	++	Courage	00	LIM_04*	00
Askari	++	Darmor	00	LIM_05*	00
Blagodatnyi	++	Doublol	00	LIM_06*	00
Cetmeki	++	Ecrin	00	LIM_07*	00
Debruder Dippes	++	Eleonore	00	LIM_08*	00
Fertodi Elit	++	Escort	00	LIM_09*	00
Gaspard	++	Evri	00	LIM_10*	00
Gesunder	++	Expander	00	MAS_01*	00
Goldgelber Zarter Butter	++	Express	00	R2N_01*	00
Hochzucht Repole	++	Frederic	00	R2N_02*	00
Kormovoi Skrivskyi	++	Goeland	00	R2N_03*	00
Lembke B	++	Ibex	00	R2N_04*	00
Lembke R	++	Idol	00	R2N_05*	00
Major	++	Inca	00	R2N_06*	00
Mansholts	++	Jetton	00	R2N_07*	00
Markus	++	Kadore	00	R2N_08*	00
Matador	++	Kosto	00	R2N_09*	00
Moldavia	++	Lewis	00	R2N_10*	00
Nain de Hambourg	++	Liberator	00	R2N_11*	00
Oleski	++	Licorne	00	R2N_12*	00
Olquell Gulzover	++	Lioness	00	SYN_01*	00
Orleans	++	Liporta	00	SYN_02*	00
Polux	++	Lirabon	00	SYN_03*	00
Prerouska	++	Lirajet	00	SYN_04*	00
QuedLinburger Platzfester	++	Madora2	00	SYN_05*	00
Rangi	++	Madora3	00	SYN_06*	00
Rapol	++	Madrigal	00	SYN_07*	00
Regal2	++	Makila	00	SYN_08*	00
Sarepta	++	Mar	00	SYN_09*	00
Titus	++	Milena	00	SYN_10*	00
Tonus	++	Mohican	00	SYN_11*	00
Aztec	0+	Montego	00		
Belinda	0+	Navajo	00		
Beryl	0+	NSL 96-25	00		
Bienvenu	0+	Oase	00		
Brink	0+	Olymp	00		
Constable	0+	Pacific	00		
Doral	0+	Payrol	00		
Emil	0+	Pollen	00		
Jupiter	0+	Potomac	00		
Kurander	0+	Prince	00		
Lesira	0+	Recital	00		
Mikado	0+	Remy	00		
Rafal	0+	Rocket	00		
Ramses	0+	Roxet	00		
Valdor	0+	Sahara	00		
Zeruca	+0	Salomont	00		
Adriana	00	Shogun	00		
Akamar	00	Start	00		
Alesi	00	SW Gospel	00		
Alligator	00	Tarox	00		
Amber	00	Tenor	00		
Ariana	00	Tor	00		
Astrid	00	Tradition	00		
Atlantic	00	Twister	00		
Aurum	00	Zorro	00		
Aviso	00	EUR_01*	00		
Bellini	00	EUR_02*	00		
Bolko	00	EUR_03*	00		
Boston	00	EUR_04*	00		
Campala	00	EUR_05*	00		
Canberra	00	EUR_06*	00		
Cancan	00	EUR_07*	00		
Capitol	00	EUR_08*	00		
Capricorn	00	EUR_09*	00		
Capvert	00	EUR_10*	00		
Ceres	00	EUR_11*	00		
Chelsi	00	EUR_12*	00		
Cheops	00	LIM_01*	00		
Colvert	00	LIM_02*	00		
Contact	00	LIM_03*	00		

The accession type (++, +0, 0+ or 00) is indicated with:  
 ++, 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  
 Stars (\*) highlight the elite varieties



**Supplementary Table III.2: Evaluation of linkage disequilibrium (LD) decay at the whole genome, sub-genome and chromosome scale**

	LD (bp) <sup>(a)</sup>
<b>Whole Genome</b>	174 130
<b>Subgenome A</b>	115 159
<b>Subgenome C</b>	320 432
<b>Chr A01</b>	91 345
<b>Chr A02</b>	318 762
<b>Chr A03</b>	97 633
<b>Chr A04</b>	198 122
<b>Chr A05</b>	177 649
<b>Chr A06</b>	115 323
<b>Chr A07</b>	124 353
<b>Chr A08</b>	291 463
<b>Chr A09</b>	239 481
<b>Chr A10</b>	180 366
<b>Chr C01</b>	993 088
<b>Chr C02</b>	239 022
<b>Chr C03</b>	255 935
<b>Chr C04</b>	1 152 120
<b>Chr C05</b>	161 666
<b>Chr C06</b>	482 903
<b>Chr C07</b>	165 837
<b>Chr C08</b>	223 599
<b>Chr C09</b>	496 404

(a) Values of the LD decay in base pairs (bp). The calculation was made for  $r^2=0.2$  using the method described by Sved (1971).



**Supplementary Table III.3.a : QTL detected at the network scale**

QTL Name	Trait	SNP_Name <sup>(a)</sup>	Allelic affect	R <sup>2</sup>	Minor Allele	Major Allele
A07d	TSW	chrA07_17323571	-0.14	0.10	T	<b>C</b>
		chrA07_17323829	-0.15	0.12	A	<b>T</b>
		chrA07_17324698	-0.14	0.11	G	<b>A</b>
		chrA07_17326232	-0.14	0.10	T	<b>G</b>
A09	SY	chrA09_2832038	-1.76	0.14	G	<b>C</b>
		chrA09_2832339	-1.76	0.14	A	<b>G</b>
		chrA09_2832511	-1.76	0.14	G	<b>A</b>
		chrA09_2851351	-1.78	0.13	T	<b>G</b>
		chrA09_2852462	-1.78	0.13	T	<b>C</b>
		chrA09_2852724	-1.78	0.13	T	<b>A</b>
		chrA09_2852839	-1.78	0.13	G	<b>T</b>
		chrA09_2857823	-1.78	0.13	C	<b>T</b>
		chrA09_2858049	-1.78	0.13	T	<b>A</b>
		chrA09_2858268	-1.78	0.13	C	<b>G</b>
		chrA09_2858340	-1.78	0.13	C	<b>A</b>
		chrA09_2858781	-1.78	0.13	C	<b>G</b>
		chrA09_2858784	-1.78	0.13	G	<b>T</b>
		chrA09_2859243	-1.78	0.13	T	<b>A</b>
		chrA09_2859342	-1.78	0.13	C	<b>T</b>

(a) SNP names correspond to the chromosome chr and the position in base pairs (bp)

The favorable alleles are indicated in bold in the last two columns.



**Supplementary Table III.3.b : QTL detected at the envirotpe scale**

QTL Name	Trait	Envirotpe	SNP_Name <sup>(a)</sup>	Allelic affect	R <sup>2</sup>	Minor Allele	Major Allele
A09	SY	EA	chrA09_2517797	-2.14	0.12	A	G
			chrA09_2606425	-2.14	0.12	T	C
			chrA09_2677515	-2.14	0.12	G	A
C09a	SY	EA	chrC09_1315791	-2.60	0.13	A	G
			chrC09_1635738	-2.67	0.12	G	A
			chrC09_1747111	-2.67	0.12	T	C
			chrC09_1752815	-2.67	0.12	A	G
			chrC09_1752977	-2.67	0.12	A	C
			chrC09_1753340	-2.67	0.12	A	C
			chrC09_1754447	-2.67	0.12	T	G
			chrC09_1770125	-2.67	0.12	G	A
			chrC09_2110856	-2.67	0.12	T	C
			chrC09_2136324	-2.67	0.12	G	A
			chrC09_2140106	-2.67	0.12	C	G
			chrC09_2142423	-2.67	0.12	C	A
			chrC09_2160202	-2.67	0.12	T	C
chrC09_2188386	-2.67	0.12	A	G			
A09	SN	EA	chrA09_2517797	-5934.44	0.19	A	G
			chrA09_2677515	-5934.44	0.19	G	A
C09a	SN	EA	chrC09_1591053	-7510.43	0.21	T	G
			chrC09_1627481	-7510.43	0.21	A	G
			chrC09_1630317	-7510.43	0.21	A	C
			chrC09_1635738	-7203.91	0.20	G	A
			chrC09_1747111	-7203.91	0.20	T	C
			chrC09_1752815	-7203.91	0.20	A	G
			chrC09_1752977	-7203.91	0.20	A	C
			chrC09_1753340	-7203.91	0.20	A	C
			chrC09_1754447	-7203.91	0.20	T	G
			chrC09_1770125	-7203.91	0.20	G	A
			chrC09_2110856	-7203.91	0.20	T	C
			chrC09_2136324	-7203.91	0.20	G	A
			chrC09_2140106	-7203.91	0.20	C	G
			chrC09_2142423	-7203.91	0.20	C	A
			chrC09_2160202	-7203.91	0.20	T	C
			chrC09_2188386	-7203.91	0.20	A	G
			chrC09_2524098	-6742.06	0.18	T	G
			chrC09_2534487	-6742.06	0.18	A	G
			chrC09_2692601	-6742.06	0.18	A	G
chrC09_2775971	-6742.06	0.18	G	T			
chrC09_2783942	-6742.06	0.18	C	G			
chrC09_2784907	-6742.06	0.18	T	C			
chrC09_2784948	-6742.06	0.18	C	T			
A03b	SY	EB	chrA03_5607548	1.27	0.11	T	C
C09a	SY	EB	chrC09_1635738	-2.06	0.11	G	A
			chrC09_1747111	-2.05	0.12	T	C
			chrC09_1752815	-2.05	0.12	A	G
			chrC09_1752977	-2.05	0.12	A	C
			chrC09_1753340	-2.05	0.12	A	C
			chrC09_1754447	-2.05	0.12	T	G
			chrC09_1770125	-2.06	0.11	G	A
			chrC09_2110856	-2.06	0.11	T	C
			chrC09_2136324	-2.06	0.11	G	A
			chrC09_2140106	-2.06	0.11	C	G
			chrC09_2142423	-2.06	0.11	C	A
			chrC09_2160202	-2.06	0.11	T	C
			chrC09_2188386	-2.06	0.11	A	G

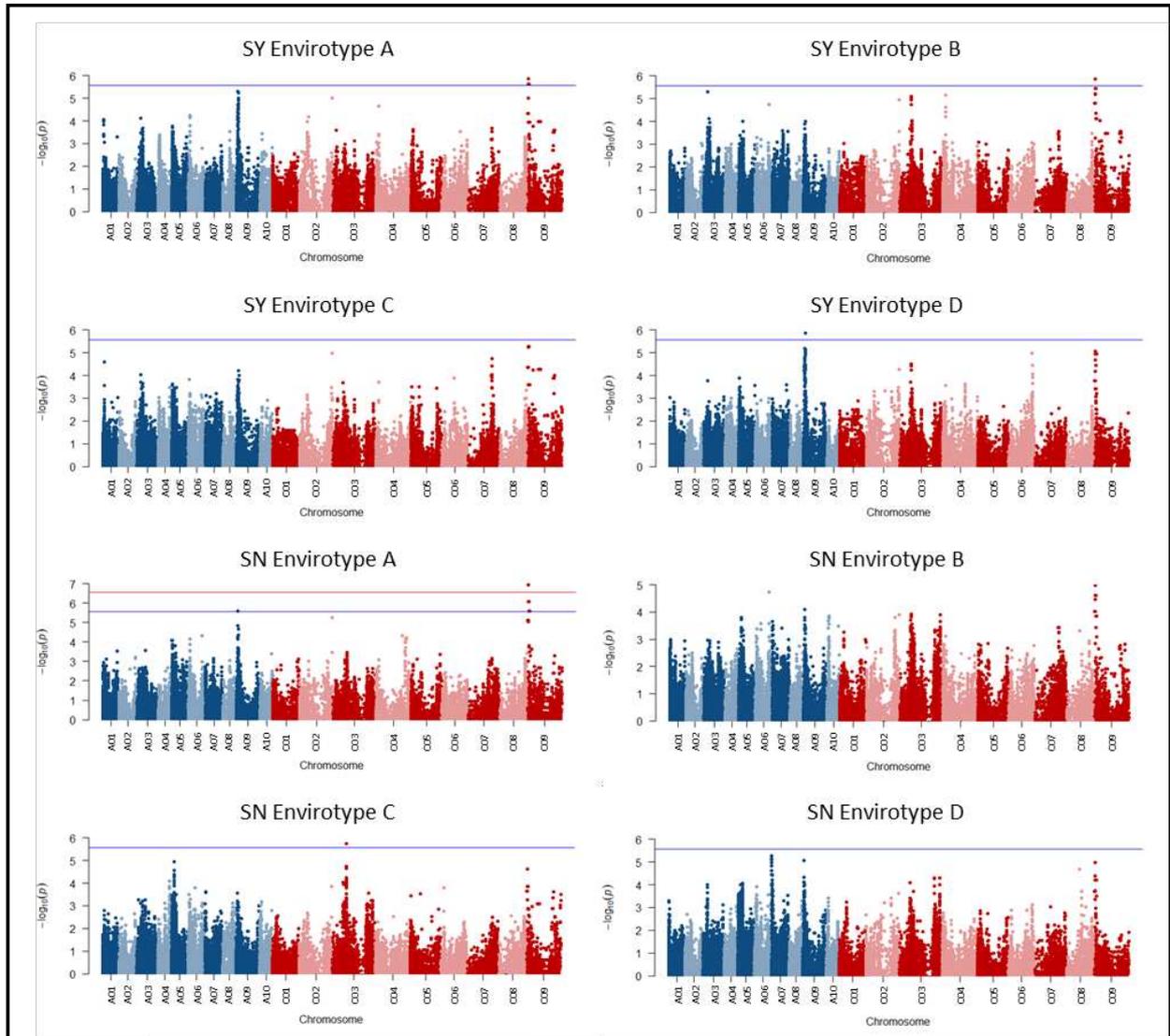


			chrC09_1591053	-1.96	0.11	T	<b>G</b>
			chrC09_1627481	-1.96	0.11	A	<b>G</b>
			chrC09_1630317	-1.96	0.11	A	<b>C</b>
			chrC09_1635738	-1.96	0.11	G	<b>A</b>
			chrC09_1770125	-1.96	0.11	G	<b>A</b>
C09a	SY	EC	chrC09_2110856	-1.96	0.11	T	<b>C</b>
			chrC09_2136324	-1.96	0.11	G	<b>A</b>
			chrC09_2140106	-1.96	0.11	C	<b>G</b>
			chrC09_2142423	-1.96	0.11	C	<b>A</b>
			chrC09_2160202	-1.96	0.11	T	<b>C</b>
			chrC09_2188386	-1.96	0.11	A	<b>G</b>
C03a	SN	EC	chrC03_20039808	-3200.46	0.21	G	<b>C</b>
			chrA09_2851351	-1.57	0.11	T	<b>G</b>
A09	SY	ED	chrA09_2852462	-1.57	0.11	T	<b>C</b>
			chrA09_2852724	-1.57	0.11	T	<b>A</b>
			chrA09_2852839	-1.57	0.11	G	<b>T</b>
			chrA07_246043	2742.37	0.19	<b>T</b>	C
			chrA07_246213	2742.37	0.19	<b>T</b>	C
			chrA07_246240	2742.37	0.19	<b>G</b>	C
			chrA07_246519	2742.37	0.19	<b>C</b>	G
			chrA07_246965	2742.37	0.19	<b>T</b>	A
			chrA07_247424	2742.37	0.19	<b>A</b>	G
			chrA07_247501	2742.37	0.19	<b>T</b>	G
A07a	SN	ED	chrA07_247567	2742.37	0.19	<b>G</b>	C
			chrA07_248262	2742.37	0.19	<b>C</b>	G
			chrA07_248496	2742.37	0.19	<b>G</b>	T
			chrA07_255402	2742.37	0.19	<b>A</b>	G
			chrA07_257687	2742.37	0.19	<b>G</b>	A
			chrA07_257743	2742.37	0.19	<b>T</b>	A
			chrA07_258169	2742.37	0.19	<b>C</b>	T
			chrA07_258661	2742.37	0.19	<b>G</b>	A
			chrA07_282933	2742.37	0.19	<b>A</b>	G

(a) SNP names correspond to the chromosome chr and the position in base pairs (bp)

The favorable alleles are indicated in bold in the last two columns.





**Supplementary Figure III.1: Manhattan plot of the GWAS analyses at the envirotypes scale for yield-related traits**

The blue lines correspond to the GAO threshold of 5%, and the red line corresponds to a classical Bonferroni threshold of 5%. The chromosomes of the A and C subgenomes of *B. napus* are highlighted in blue and red respectively. SY, seed yield; SN, seed number.



**Supplementary Table III.3.c : QTL detected at the trial scale**

QTL Name	Trait	Trial	SNP_Name <sup>(a)</sup>	Allelic affect	R <sup>2</sup>	Minor Allele	Major Allele
C09a	SY	Md15_N2	chrC09_816997	-2.81	0.11	C	T
			chrC09_872059	-2.81	0.11	C	T
			chrC09_872443	-2.81	0.11	A	G
			chrC09_1591053	-2.86	0.11	T	G
			chrC09_1627481	-2.86	0.11	A	G
			chrC09_1630317	-2.86	0.11	A	C
			chrC09_1635738	-3.05	0.12	G	A
			chrC09_1747111	-2.95	0.12	T	C
			chrC09_1752815	-2.95	0.12	A	G
			chrC09_1752977	-2.95	0.12	A	C
			chrC09_1753340	-2.95	0.12	A	C
			chrC09_1754447	-2.95	0.12	T	G
			chrC09_1770125	-3.05	0.12	G	A
			chrC09_2110856	-3.05	0.12	T	C
			chrC09_2136324	-3.05	0.12	G	A
			chrC09_2140106	-3.05	0.12	C	G
chrC09_2142423	-3.05	0.12	C	A			
chrC09_2160202	-3.05	0.12	T	C			
chrC09_2188386	-3.05	0.12	A	G			
A09	SY	Ch14_N1	chrA09_2851351	-2.65	0.11	T	G
			chrA09_2852462	-2.65	0.11	T	C
			chrA09_2852724	-2.65	0.11	T	A
			chrA09_2852839	-2.65	0.11	G	T
			chrA09_2854970	-2.57	0.11	T	C
			chrA09_2951392	-2.59	0.11	A	T
			chrA09_2966543	-2.59	0.11	C	T
			chrA09_2971275	-2.59	0.11	A	G
			chrA09_2973779	-2.59	0.11	T	G
chrA09_2999088	-2.59	0.11	T	C			
C09a	SY	Ch14_N1	chrC09_1635738	-2.86	0.10	G	A
			chrC09_1747111	-2.74	0.10	T	C
			chrC09_1752815	-2.74	0.10	A	G
			chrC09_1752977	-2.74	0.10	A	C
			chrC09_1753340	-2.74	0.10	A	C
			chrC09_1754447	-2.74	0.10	T	G
			chrC09_1770125	-2.86	0.10	G	A
			chrC09_2110856	-2.86	0.10	T	C
			chrC09_2136324	-2.86	0.10	G	A
			chrC09_2140106	-2.86	0.10	C	G
			chrC09_2142423	-2.86	0.10	C	A
chrC09_2160202	-2.86	0.10	T	C			
chrC09_2188386	-2.86	0.10	A	G			
A04a	SY	Ch14_N2	chrA04_7408414	-2.20	0.11	C	T
A04b	SY	Ch14_N2	chrA04_8280430	-2.33	0.11	G	A
A05a	SY	Ch14_N2	chrA05_1939257	-2.21	0.12	C	T
			chrA05_1941241	-2.28	0.12	A	G
A09	SY	Ch14_N2	chrA09_2517797	-2.90	0.11	A	G
			chrA09_2666437	-2.90	0.11	C	G
			chrA09_2666461	-2.90	0.11	A	G
			chrA09_2677515	-2.90	0.11	G	A
			chrA09_2851351	-3.37	0.13	T	G
			chrA09_2852462	-3.37	0.13	T	C
			chrA09_2852724	-3.37	0.13	T	A
			chrA09_2852839	-3.37	0.13	G	T
			chrA09_2854117	-3.19	0.11	C	T
chrA09_2854970	-3.04	0.11	T	C			



			chrA09_2857823	-3.06	0.11	C	T
			chrA09_2858049	-3.06	0.11	T	A
			chrA09_2858268	-3.06	0.11	C	G
			chrA09_2858340	-3.06	0.11	C	A
			chrA09_2858781	-3.06	0.11	C	G
			chrA09_2858784	-3.06	0.11	G	T
			chrA09_2859243	-3.06	0.11	T	A
			chrA09_2859342	-3.06	0.11	C	T
			chrA09_2949785	-3.04	0.12	C	T
			chrA09_2951392	-3.09	0.11	A	T
			chrA09_2966543	-3.09	0.11	C	T
			chrA09_2971275	-3.09	0.11	A	G
			chrA09_2973454	-3.04	0.12	A	G
			chrA09_2973779	-3.09	0.11	T	G
			chrA09_2999088	-3.09	0.11	T	C
			chrA09_3031214	-3.04	0.12	C	A
			chrA09_3034006	-3.04	0.12	G	T
			chrA09_3035316	-3.04	0.12	G	T
			chrA09_3047043	-3.04	0.12	C	G
			chrA09_3048739	-3.04	0.12	T	C
			chrA09_3051850	-3.04	0.12	T	C
			chrC09_1591053	-3.46	0.11	T	G
			chrC09_1627481	-3.46	0.11	A	G
			chrC09_1630317	-3.46	0.11	A	C
			chrC09_1635738	-3.90	0.13	G	A
			chrC09_1747111	-3.94	0.14	T	C
			chrC09_1752815	-3.94	0.14	A	G
			chrC09_1752977	-3.94	0.14	A	C
			chrC09_1753340	-3.94	0.14	A	C
			chrC09_1754447	-3.94	0.14	T	G
			chrC09_1770125	-3.90	0.13	G	A
			chrC09_2110856	-3.90	0.13	T	C
			chrC09_2136324	-3.90	0.13	G	A
C09a	SY	Ch14_N2	chrC09_2140106	-3.90	0.13	C	G
			chrC09_2142423	-3.90	0.13	C	A
			chrC09_2160202	-3.90	0.13	T	C
			chrC09_2188386	-3.90	0.13	A	G
			chrC09_2369409	-3.70	0.14	T	C
			chrC09_2524098	-3.61	0.12	T	G
			chrC09_2534487	-3.61	0.12	A	G
			chrC09_2692601	-3.61	0.12	A	G
			chrC09_2775971	-3.61	0.12	G	T
			chrC09_2783942	-3.61	0.12	C	G
			chrC09_2784907	-3.61	0.12	T	C
			chrC09_2784948	-3.61	0.12	C	T
A07b	SY	LR15_N2	chrA07_754861	1.66	0.13	A	C
			chrA07_755814	1.63	0.13	C	A
			chrC09_1747111	-2.26	0.10	T	C
C09a	SY	LR15_N2	chrC09_1752815	-2.26	0.10	A	G
			chrC09_1752977	-2.26	0.10	A	C
			chrC09_1753340	-2.26	0.10	A	C
			chrC09_1754447	-2.26	0.10	T	G
			chrA09_2517797	-2.60	0.13	A	G
A09	SY	Yeb15_N2	chrA09_2606425	-2.52	0.13	T	C
			chrA09_2677515	-2.60	0.13	G	A
C02b	SY	Yeb15_N2	chrC02_45059324	-3.08	0.13	A	C
			chrC09_1591053	-3.24	0.14	T	G
			chrC09_1627481	-3.24	0.14	A	G
C09a	SY	Yeb15_N2	chrC09_1630317	-3.24	0.14	A	C
			chrC09_1635738	-3.31	0.14	G	A



			chrC09_1747111	-3.31	0.14	T	C
			chrC09_1752815	-3.31	0.14	A	G
			chrC09_1752977	-3.31	0.14	A	C
			chrC09_1753340	-3.31	0.14	A	C
			chrC09_1754447	-3.31	0.14	T	G
			chrC09_1770125	-3.31	0.14	G	A
			chrC09_2110856	-3.31	0.14	T	C
			chrC09_2136324	-3.31	0.14	G	A
			chrC09_2140106	-3.31	0.14	C	G
			chrC09_2142423	-3.31	0.14	C	A
			chrC09_2160202	-3.31	0.14	T	C
			chrC09_2188386	-3.31	0.14	A	G
			chrC09_2524098	-2.89	0.11	T	G
			chrC09_2534487	-2.89	0.11	A	G
			chrC09_2692601	-2.89	0.11	A	G
			chrC09_2775971	-2.89	0.11	G	T
			chrC09_2783942	-2.89	0.11	C	G
			chrC09_2784907	-2.89	0.11	T	C
			chrC09_2784948	-2.89	0.11	C	T
C09a	SY	Sel14_N1	chrC09_816997	-3.00	0.13	C	T
			chrC09_872059	-3.00	0.13	C	T
			chrC09_872443	-3.00	0.13	A	G
			chrC09_1591053	-3.00	0.12	T	G
			chrC09_1627481	-3.00	0.12	A	G
			chrC09_1630317	-3.00	0.12	A	C
			chrC09_1635738	-3.05	0.11	G	A
			chrC09_1747111	-2.97	0.12	T	C
			chrC09_1752815	-2.97	0.12	A	G
			chrC09_1752977	-2.97	0.12	A	C
			chrC09_1753340	-2.97	0.12	A	C
			chrC09_1754447	-2.97	0.12	T	G
			chrC09_1770125	-3.05	0.11	G	A
			chrC09_2110856	-3.05	0.11	T	C
			chrC09_2136324	-3.05	0.11	G	A
			chrC09_2140106	-3.05	0.11	C	G
			chrC09_2142423	-3.05	0.11	C	A
chrC09_2160202	-3.05	0.11	T	C			
chrC09_2188386	-3.05	0.11	A	G			
chrC09_2369409	-2.93	0.12	T	C			
A03a	SY	Sel14_N2	chrA03_2180395	-1.91	0.11	A	G
			chrA03_2180405	-1.91	0.11	A	T
A10g	SY	Sel14_N2	chrA10_16721480	-1.83	0.11	G	C
A09	SY	Pre14_N2	chrA09_2517797	-1.72	0.11	A	G
			chrA09_2677515	-1.72	0.11	G	A
C04	SY	Pre15_N1	chrC04_4063677	-1.62	0.15	G	A
A03b	SY	Pre15_N2	chrA03_5607548	1.49	0.12	T	C
A03c	SY	Pre15_N2	chrA03_20719243	-1.35	0.11	A	G
			chrA03_20740616	-1.35	0.11	A	G
			chrA03_20741855	-1.35	0.11	G	T
			chrA03_20743384	-1.35	0.11	A	G
			chrA03_20744971	-1.35	0.11	A	G
			chrA03_20755573	-1.35	0.11	A	C
			chrA03_20758119	-1.35	0.11	G	T
			chrA03_20758120	-1.35	0.11	G	A
C03b	SY	Ver14_N1	chrC03_30194954	1.36	0.11	T	C
			chrC03_30195010	1.36	0.11	T	C
			chrC03_30195362	1.36	0.11	C	T
C09a	SY	Ver14_N1	chrC09_1635738	-2.07	0.10	G	A
			chrC09_1770125	-2.07	0.10	G	A
			chrC09_2110856	-2.07	0.10	T	C



Déterminants génétiques de la stabilité – III.3. Analyses GWAS (Supp. Data)

			chrC09_2136324	-2.07	0.10	G	A
			chrC09_2140106	-2.07	0.10	C	G
			chrC09_2142423	-2.07	0.10	C	A
			chrC09_2160202	-2.07	0.10	T	C
			chrC09_2188386	-2.07	0.10	A	G
C09c	SY	Ver14_N1	chrC09_38469046	1.63	0.12	T	G
			chrA09_2851351	-2.40	0.12	T	G
			chrA09_2852462	-2.40	0.12	T	C
			chrA09_2852724	-2.40	0.12	T	A
			chrA09_2852839	-2.40	0.12	G	T
A09	SY	Ver14_N2	chrA09_2854970	-2.23	0.11	T	C
			chrA09_2951392	-2.23	0.11	A	T
			chrA09_2966543	-2.23	0.11	C	T
			chrA09_2971275	-2.23	0.11	A	G
			chrA09_2973779	-2.23	0.11	T	G
			chrA09_2999088	-2.23	0.11	T	C
C02b	SY	Ver14_N2	chrC02_45059324	-2.61	0.13	A	C
C08	SY	Ver14_N2	chrC08_35215389	-1.51	0.12	T	G
			chrC08_35215692	-1.51	0.12	A	C
			chrC09_816997	-2.44	0.12	C	T
			chrC09_872059	-2.44	0.12	C	T
			chrC09_872443	-2.44	0.12	A	G
			chrC09_1315791	-2.69	0.14	A	G
			chrC09_1591053	-2.81	0.14	T	G
			chrC09_1627481	-2.81	0.14	A	G
			chrC09_1630317	-2.81	0.14	A	C
			chrC09_1635738	-2.94	0.14	G	A
			chrC09_1747111	-2.88	0.15	T	C
			chrC09_1752815	-2.88	0.15	A	G
			chrC09_1752977	-2.88	0.15	A	C
			chrC09_1753340	-2.88	0.15	A	C
			chrC09_1754447	-2.88	0.15	T	G
C09a	SY	Ver14_N2	chrC09_1770125	-2.94	0.14	G	A
			chrC09_2110856	-2.94	0.14	T	C
			chrC09_2136324	-2.94	0.14	G	A
			chrC09_2140106	-2.94	0.14	C	G
			chrC09_2142423	-2.94	0.14	C	A
			chrC09_2160202	-2.94	0.14	T	C
			chrC09_2188386	-2.94	0.14	A	G
			chrC09_2524098	-2.57	0.11	T	G
			chrC09_2534487	-2.57	0.11	A	G
			chrC09_2692601	-2.57	0.11	A	G
			chrC09_2775971	-2.57	0.11	G	T
			chrC09_2783942	-2.57	0.11	C	G
			chrC09_2784907	-2.57	0.11	T	C
			chrC09_2784948	-2.57	0.11	C	T
A03b	SY	Ver15_N2	chrA03_5607548	1.61	0.11	T	C
A09	SY	Ver15_N2	chrA09_2517797	-2.17	0.11	A	G
			chrA09_2677515	-2.17	0.11	G	A
C09a	SY	Ver15_N2	chrC09_1315791	-2.44	0.11	A	G
			chrC09_1315791	-3.19	0.13	A	G
			chrC09_1635738	-3.20	0.11	G	A
			chrC09_1747111	-3.05	0.11	T	C
			chrC09_1752815	-3.05	0.11	A	G
C09a	SY	Dij15_N1	chrC09_1752977	-3.05	0.11	A	C
			chrC09_1753340	-3.05	0.11	A	C
			chrC09_1754447	-3.05	0.11	T	G
			chrC09_1770125	-3.20	0.11	G	A
			chrC09_2110856	-3.20	0.11	T	C
			chrC09_2136324	-3.20	0.11	G	A



Déterminants génétiques de la stabilité – III.3. Analyses GWAS (Supp. Data)

			chrC09_2140106	-3.20	0.11	C	<b>G</b>
			chrC09_2142423	-3.20	0.11	C	<b>A</b>
			chrC09_2160202	-3.20	0.11	T	<b>C</b>
			chrC09_2188386	-3.20	0.11	A	<b>G</b>
C04	SY	Dij15_N2	chrC04_4063677	-2.15	0.12	G	<b>A</b>
			chrC09_816997	-6229.24	0.19	C	<b>T</b>
C09a	SN	Md15_N2	chrC09_872059	-6229.24	0.19	C	<b>T</b>
			chrC09_872443	-6229.24	0.19	A	<b>G</b>
			chrC09_1315791	-6293.22	0.18	A	<b>G</b>
			chrC09_1635738	-7835.71	0.18	G	<b>A</b>
			chrC09_1747111	-7768.81	0.19	T	<b>C</b>
			chrC09_1752815	-7768.81	0.19	A	<b>G</b>
			chrC09_1752977	-7768.81	0.19	A	<b>C</b>
			chrC09_1753340	-7768.81	0.19	A	<b>C</b>
			chrC09_1754447	-7768.81	0.19	T	<b>G</b>
C09a	SN	Ch14_N2	chrC09_1770125	-7835.71	0.18	G	<b>A</b>
			chrC09_2110856	-7835.71	0.18	T	<b>C</b>
			chrC09_2136324	-7835.71	0.18	G	<b>A</b>
			chrC09_2140106	-7835.71	0.18	C	<b>G</b>
			chrC09_2142423	-7835.71	0.18	C	<b>A</b>
			chrC09_2160202	-7835.71	0.18	T	<b>C</b>
			chrC09_2188386	-7835.71	0.18	A	<b>G</b>
			chrC09_2369409	-7323.95	0.19	T	<b>C</b>
			chrA07_246043	4192.30	0.18	T	C
			chrA07_246213	4192.30	0.18	T	C
			chrA07_246240	4192.30	0.18	<b>G</b>	C
			chrA07_246519	4192.30	0.18	<b>C</b>	G
			chrA07_246965	4192.30	0.18	T	A
			chrA07_247424	4192.30	0.18	<b>A</b>	G
			chrA07_247501	4192.30	0.18	T	G
A07a	SN	LR15_N2	chrA07_247567	4192.30	0.18	<b>G</b>	C
			chrA07_248262	4192.30	0.18	<b>C</b>	G
			chrA07_248496	4192.30	0.18	<b>G</b>	T
			chrA07_255402	4192.30	0.18	<b>A</b>	G
			chrA07_257687	4192.30	0.18	<b>G</b>	A
			chrA07_257743	4192.30	0.18	T	A
			chrA07_258169	4192.30	0.18	<b>C</b>	T
			chrA07_258661	4192.30	0.18	<b>G</b>	A
			chrA07_282933	4192.30	0.18	<b>A</b>	G
A07b	SN	LR15_N2	chrA07_754861	4697.58	0.22	<b>A</b>	C
			chrA07_755814	4685.25	0.23	<b>C</b>	A
			chrA07_1127600	4008.18	0.19	<b>G</b>	C
A07c	SN	LR15_N2	chrA07_1144881	4066.27	0.19	<b>G</b>	C
			chrA07_1144932	4066.27	0.19	T	C
			chrA07_1144941	4066.27	0.19	T	A
A07a	SN	Pre14_N1	chrA07_244837	2421.88	0.19	<b>A</b>	T
A07b	SN	Pre14_N1	chrA07_829693	2351.46	0.19	<b>A</b>	G
			chrA07_829734	2351.46	0.19	T	C
A05b	SN	Pre14_N2	chrA05_3335006	-3566.44	0.20	C	<b>G</b>
			chrC09_1591053	-5952.54	0.17	T	<b>G</b>
C09a	SN	Ver14_N2	chrC09_1627481	-5952.54	0.17	A	<b>G</b>
			chrC09_1630317	-5952.54	0.17	A	<b>C</b>
			chrC09_4463887	0.15	0.11	<b>G</b>	A
C09b	TSW	Md15_N1	chrC09_4466434	0.15	0.11	T	G
			chrA07_17369498	-0.13	0.09	T	<b>C</b>
			chrA07_17369511	-0.13	0.09	T	<b>G</b>
			chrA07_17369525	-0.13	0.09	T	<b>C</b>
A07d	TSW	Md15_N2	chrA07_17370030	-0.13	0.09	G	<b>A</b>
			chrA07_17370364	-0.14	0.11	T	<b>G</b>
			chrA07_17370793	-0.14	0.11	C	<b>G</b>



Déterminants génétiques de la stabilité – III.3. Analyses GWAS (Supp. Data)

			chrA07_17370911	-0.14	0.11	G	<b>A</b>
			chrA07_17371135	-0.14	0.11	G	<b>A</b>
			chrA07_17376493	-0.14	0.10	C	<b>T</b>
			chrA07_17378000	-0.13	0.09	T	<b>C</b>
			chrA07_17390063	-0.13	0.09	A	<b>T</b>
			chrA07_17390189	-0.13	0.09	C	<b>T</b>
			chrA07_17391242	-0.14	0.10	A	<b>C</b>
A10b	TSW	Ch14_N1	chrA10_2074182	0.14	0.09	<b>T</b>	C
			chrA10_2115701	0.14	0.09	<b>C</b>	A
			chrA10_2115740	0.14	0.09	<b>G</b>	A
A10e	TSW	Ch14_N1	chrA10_15639575	-0.15	0.10	C	<b>T</b>
A10a	TSW	Chr14_N2	chrA10_1864577	0.15	0.10	<b>A</b>	C
			chrA10_1865945	0.15	0.10	<b>A</b>	G
			chrA10_1865963	0.15	0.10	<b>T</b>	A
A10b	TSW	Ch14_N2	chrA10_2067138	0.15	0.10	<b>C</b>	A
			chrA10_2067168	0.15	0.10	<b>G</b>	C
			chrA10_2067171	0.15	0.10	<b>T</b>	A
			chrA10_2074025	0.15	0.10	<b>A</b>	T
			chrA10_2074182	0.14	0.09	<b>T</b>	C
			chrA10_2115701	0.16	0.12	<b>C</b>	A
			chrA10_2115740	0.16	0.12	<b>G</b>	A
A10c	TSW	Ch14_N2	chrA10_2504179	0.15	0.09	<b>T</b>	G
			chrA10_2505644	0.15	0.09	<b>A</b>	C
A10a	TSW	Pre14_N1	chrA10_1864577	0.16	0.08	<b>A</b>	C
			chrA10_1865945	0.16	0.08	<b>A</b>	G
			chrA10_1865963	0.16	0.08	<b>T</b>	A
A10b	TSW	Pre14_N1	chrA10_2074025	0.16	0.08	<b>A</b>	T
			chrA10_2115701	0.22	0.14	<b>C</b>	A
			chrA10_2115740	0.22	0.14	<b>G</b>	A
C02b	TSW	Pre14_N1	chrC02_24973	0.18	0.10	<b>C</b>	A
			chrC02_34259	0.18	0.10	<b>A</b>	C
A10b	TSW	Ver14_N1	chrA10_2067138	0.16	0.10	<b>C</b>	A
			chrA10_2067168	0.16	0.10	<b>G</b>	C
			chrA10_2067171	0.16	0.10	<b>T</b>	A
			chrA10_2074182	0.17	0.12	<b>T</b>	C
			chrA10_2115701	0.21	0.17	<b>C</b>	A
			chrA10_2115740	0.21	0.17	<b>G</b>	A
A10c	TSW	Ver14_N1	chrA10_2416579	0.17	0.10	<b>C</b>	A
			chrA10_2444600	0.17	0.10	<b>T</b>	G
A10d	TSW	Ver14_N1	chrA10_2696446	0.18	0.10	<b>G</b>	C
			chrA10_2696512	0.18	0.10	<b>T</b>	C
			chrA10_2696612	0.18	0.10	<b>A</b>	G
A10f	TSW	Ver14_N1	chrA10_16256530	0.16	0.09	<b>C</b>	T
C02a	TSW	Ver14_N1	chrC02_24973	0.16	0.09	<b>C</b>	A
			chrC02_34259	0.16	0.09	<b>A</b>	C
A07d	TSW	Ver14_N2	chrA07_17371135	-0.16	0.09	G	<b>A</b>
			chrA07_17390063	-0.16	0.09	A	<b>T</b>
			chrA07_17390189	-0.16	0.09	C	<b>T</b>
			chrA07_17391242	-0.16	0.09	A	<b>C</b>
A10b	TSW	Ver14_N2	chrA10_2115701	0.16	0.10	<b>C</b>	A
			chrA10_2115740	0.16	0.10	<b>G</b>	A

(a) SNP names correspond to the chromosome chr and the position in base pairs (bp)

The favorable alleles are indicated in bold in the last two columns.



## CHAPITRE IV

---

### Discussion générale et perspectives

---



## IV.1. Des indicateurs environnementaux pour expliquer l'interaction G×E

---

### IV.1.1. Température et radiations à la floraison sont les principaux facteurs limitants du rendement et expliquent une part de l'interaction G×E

La première question soulevée dans la thèse était d'identifier les principaux facteurs limitants et d'investiguer leurs rôles sur les effets de l'environnement et de l'interaction G×E. L'utilisation de la régression PLS, couplée à une règle de sélection de variables, a permis d'identifier les principaux facteurs limitants du rendement sur deux réseaux différents (#1 et #2). La température et les radiations durant les périodes de floraison jusqu'à la maturation de la graine ainsi que les conditions de vernalisation sont les principaux facteurs limitant le rendement. L'impact de la température (Habekotté, 1997 ; Weymann et al., 2015), de la combinaison température et radiation à floraison (Baux et al., 2015) ainsi que le rôle de la vernalisation sur la floraison (Sheldon et al., 2000) chez le colza ont déjà été décrits. Le réseau #2, grâce à une gestion différenciée des apports azotés, a révélé l'impact de l'azote sur le rendement, confirmant des études antérieures (Rathke et al., 2006).

Les envirotypes, construits sur la base des facteurs limitants, expliquent jusqu'à 90% de l'effet de l'environnement et ont permis de regrouper les environnements de manière à ce que l'effet environnement résiduel au sein de chaque groupe soit fortement réduit. De plus, l'envirotypage explique en moyenne 15% de l'interaction G×E présente à l'échelle du réseau global et laisse une interaction G×E résiduelle non expliquée au sein de chaque envirotype de l'ordre de 4% de la variance totale contre plus de 5% sans envirotypage. Des résultats similaires ont été décrits sur le blé où l'utilisation de regroupement d'environnements en fonction de différents scénarios hydriques a permis d'expliquer 20% de l'interaction G×E présente à l'échelle du réseau (Touzy et al., 2019). Deux stratégies coexistent pour regrouper les environnements : des stratégies basées sur le choix des facteurs environnementaux qui expliquent les variations du trait d'intérêt (comme proposé dans la thèse ainsi que dans l'étude de Touzy et al., 2019), et des stratégies basées sur le choix des facteurs qui expliquent les variations de la matrice d'interaction G×E directement. La méthode qui a été retenue dans cette étude permet d'interpréter une part de l'interaction à l'aide des facteurs limitants tout en expliquant la majeure partie des effets de l'environnement, qui restent un des principaux facteurs expliquant les variations de rendement et permettant ainsi de mieux décrire le réseau d'essais. Cependant, nous avons pu voir que l'utilisation des envirotypes sur d'autres traits pouvait être questionnée notamment lorsque le trait d'envirotypage et le trait testé ne sont pas corrélés (ex.



TSW et SY). Cela est dû au fait que ce ne sont pas les mêmes facteurs limitants qui vont expliquer les variations des deux traits.

Dans notre étude, le stress hydrique, qui est décrit comme un stress récurrent et majeur sur la culture de colza (Champolivier et Merrien, 1996), n'est pas apparu comme facteur limitant malgré l'observation de certaines périodes plutôt sèches dans nos réseaux (sur la base des retours d'experts). On peut donc s'interroger sur la puissance de nos indicateurs et/ou des réseaux considérés pour capter des stress plus fugaces, ce qui ouvre la réflexion sur les outils pour améliorer la définition des indicateurs. En complément, on se pose également la question de la transférabilité des indicateurs développés pendant la thèse à d'autres itinéraires techniques et/ou systèmes de culture.

#### IV.1.2. Intérêt des modèles de culture pour affiner la liste des indicateurs

Les indicateurs utilisés dans la thèse sont basés sur des variables acquises expérimentalement et donc, pour la plupart, à un instant donné. A titre d'exemple, l'indicateur NNI est dérivé de la valeur d'INN (*Encadré II.1*) qui repose sur un seul prélèvement de biomasse aérienne réalisé à la montaison. Le prélèvement en lui-même peut générer un biais significatif (en particulier, la date de prélèvement par rapport à la fertilisation azoté ou encore la représentativité des plantes prélevées par rapport au peuplement). Idéalement, les modèles de culture permettraient de palier à ces limites, car ils prédisent ce type de variables de façon dynamique, à un pas de temps journalier. Ils peuvent également être mobilisés pour calculer de nouveaux indicateurs. Par exemple, le modèle AZODYN (Jeuffroy et Recous, 1999), en cours de calibration sur colza, offre la possibilité de simuler des variables telles que la production de biomasse, des INN, des dynamiques hydriques à des pas de temps journaliers (Gervois et al., 2019). Cependant, du fait de certaines limites du modèle à ce jour (ex. mauvaises prédictions des INN à la floraison), nous ne l'avons pas utilisé dans la thèse.

Par ailleurs, notre méthode se base sur le comportement d'un petit nombre de génotypes considérés comme moyens pour définir les périodes ainsi que certains indicateurs « plante ». Cependant, ces génotypes ne représentent pas forcément la gamme de diversité génétique expérimentée au sein du réseau. Par exemple au sein de notre population P173, une différence de 200°C.J sur la date de floraison est observée (*Figure II.B*). Des modèles de culture, tels que SUNFLO sur le tournesol, utilisent un paramétrage génétique spécifique pour simuler les performances (Casadebaig et al., 2011). Ce type de paramétrage pourrait être implémenté sur les modèles de culture en cours de développement sur le colza. A défaut, les périodes et les indicateurs pourraient être définis pour chaque génotype. Il faudrait alors utiliser une régression PLS multivariée (Tenenhaus, 1998) pour



expliquer les variations de rendement propres à chaque génotype. Enfin, une autre piste consiste également à considérer d'autres structures génétiques que les lignées. En effet, les variétés de colza cultivées aujourd'hui sont quasiment exclusivement des hybrides F1, offrant de meilleures performances et une meilleure stabilité (Diepenbrock, 2000 ; Gehringer et al., 2007).

Enfin, les modèles de culture peuvent être utilisés pour établir des scénarios climatiques en se basant sur des réseaux expérimentaux plus larges à l'échelle nationale (ex. réseaux CTPS) ou européenne, voire en simulant des scénarios du futur sur la base de prédictions climatiques. Cette approche a été conduite sur le maïs en se basant sur un réseau de 35 années x 55 lieux pour établir des scénarios climatiques (Millet et al., 2016).

### **IV.1.3. Adapter l'envirotypage à des systèmes de culture plus complexes**

La méthode d'envirotypage développée dans la thèse s'est basée sur du colza en culture pure menée sous itinéraires techniques conventionnels (protection contre les stress biotiques et apports minéraux en P, K, S) et dans un système bas intrants azotés. Cependant d'autres modèles émergent, promouvant des systèmes de culture et des itinéraires culturaux innovants (variations de la date de semis, association avec des légumineuses gélives, zéro phyto) qui permettront de mieux répondre aux enjeux de l'agriculture de demain (réduction des intrants, agroécologie). Dans la suite, je propose de réfléchir sur le développement de nouveaux indicateurs en lien avec les contraintes biotiques d'une part, ou avec la culture du colza en association d'autre part.

Le développement d'indicateurs de stress biotique peut passer par des référencements au champ (adventices, pièges à insectes, apparition et évaluation d'une maladie) mais aussi par la modélisation. Certains auteurs ont développé des modèles sur la base de données climatiques pour prédire des attaques de ravageurs (Olatinwo et Hoogenboom, 2014) ou l'apparition et la sévérité de maladies comme le phoma (Pullens et al., 2019). D'autres modèles, basés sur l'utilisation de paramètres du paysage, sont utilisés pour prédire des attaques de méligèthes sur le colza à des échelles régionales (Rusch et al., 2012).

En France, la culture du colza en association avec les légumineuses gélives est une pratique en fort développement car elle permet notamment d'optimiser la gestion des insectes ravageurs, la fertilisation azotée et la lutte contre les adventices, tout en limitant l'usage des intrants, ce qui constitue un levier d'amélioration prometteur du système de production, en plus du levier variétal (Jamont et al., 2013 ; Cadoux et al., 2015, Génard et al., 2016). Cependant, dans ces nouvelles pratiques, la gestion de la dynamique de l'azote organique dans le sol est complexifiée par rapport à des apports minéraux. Nos indicateurs, notamment ceux en lien avec le suivi hydrique ou azoté, ne



sont probablement pas adaptés à ce type de conduite. L'identification d'indicateurs pour ces pratiques se complexifie et les modèles de culture ne prennent pas en compte pour le moment les cultures en association. Ainsi c'est l'ensemble des modèles biologiques, statistiques et mathématiques qui doivent être repensés afin de répondre aux enjeux posés par les nouvelles pratiques agricoles.



## IV.2. Les déterminants génétiques de l'interaction

---

Le second objectif de ce travail était d'identifier les déterminants génétiques du rendement et d'analyser comment se structuraient les QTL en termes de stabilité ou d'interactivité. Nous avons pu mettre en évidence, avec les analyses GWAS à l'échelle de chaque essai individuel et à l'échelle du réseau entier, 27 QTL répartis entre le rendement en grains (SY), le nombre de grains (SN) et le poids de mille grains (TSW). Parmi ces QTL, 2 sont qualifiés de QTL stables car ils sont détectés à l'échelle du réseau et 25 sont définis comme interactifs car uniquement identifiés dans des essais individuels. La majorité des QTL interactifs sont spécifiques d'essais. Par rapport aux études de Bouchet et al. (2016a) qui reprennent une partie de notre matériel génétique et de notre réseau, des résultats semblables ont été trouvés avec une majorité de QTL interactifs. Nos QTL « stables » en A09 et C09 ont déjà été identifiés précédemment (Bouchet et al., 2016a) mais le QTL interactif A07a n'avait pas été identifié, montrant ainsi l'intérêt de notre méthode pour découvrir de nouvelles zones du génome en lien avec les G×E et *in fine* les facteurs limitants.

### IV.2.1. Peu de QTL détectés pour le rendement

Nous n'avons pu identifier que 27 QTL tous traits confondus. En comparaison, Bouchet et al. (2014) ont mis en évidence 76 QTL de rendement sur des populations d'haploïdes doublés d'effectifs similaires ou inférieurs. En 2016, les études menées sur deux panels de 69 et 92 génotypes avaient révélé 946 QTL en GWAS (Bouchet et al., 2016a), confirmant la nature hautement polygénique du rendement. La différence observée entre ces études et la nôtre peut être expliquée par plusieurs facteurs. La première explication repose sur le matériel génétique utilisé. Dans notre étude nous avons utilisé un panel de 173 génotypes, majoritairement de type 00 et contenant des variétés élites, tandis que les études menées par Bouchet et collaborateurs reposaient sur des populations en ségrégation et sur des panels différents. Ce fond génétique différent peut expliquer la réduction du nombre de QTL. La deuxième explication repose sur la méthode GWAS employée. Bouchet et al., (2016a) ont utilisé un modèle MLM associé à une matrice d'apparentement pour l'ensemble du génome et défini un SNP associé selon un seuil de False Discovery Rate (FDR) de 15% pour la GWAS. Des modèles MQM (multiple QTL mapping) ont été utilisés pour les analyses de liaison. Dans notre cas, nous avons utilisé un modèle FastLMM associé à une matrice d'apparentement par chromosome et un seuil de Gao de 10%, plus stringent qu'un FDR. De plus, la méthode que nous avons utilisée permet un contrôle efficace des faux positifs par rapport aux modèles plus classiques (Rincent et al.,



2014). Les méthodes, le matériel génétique et les seuils utilisés peuvent expliquer cette différence dans le nombre de QTL identifiés. Afin de confirmer cela, il faudrait envisager de retraiter les données GWAS de Bouchet et al. (2016a) en utilisant le modèle FastLMM.

Enfin l'ensemble des QTL détectés sont majoritairement fixés dans les variétés élites. La diversité génétique du colza d'hiver est étroite, il faut alors raisonner sur les moyens possibles pour l'élargir. Le recours aux variétés semi-synthétiques obtenus par la voie des hybrides triploïdes est un moyen de réintroduire une diversité génétique nouvelle grâce à l'augmentation du taux de recombinaison notamment dans des régions peu recombinantes (Leflon et al., 2006 ; Pelé et al., 2017).

#### IV.2.2. QTL de stabilité vs. QTL interactifs

Parmi les QTL identifiés, deux ont été définis comme stables car ils ont été détectés à l'échelle du réseau et dans plusieurs essais. Cependant lorsque l'on regarde le nombre d'essais où ces QTL ont été détectés, on remarque qu'il est faible par rapport au nombre total d'essais du réseau, d'ailleurs aucun QTL n'a été détecté sur tous les essais. Selon les définitions proposées par Des Marais et al. (2013) ces QTL pourraient être des QTL interactifs à effet 'conditionnellement neutre' car ils ont été détectés dans certains essais et absents dans les autres. Cependant, nous pouvons nous questionner sur ces définitions. En effet, des QTL conditionnellement neutres peuvent être considérés comme un cas particulier des QTL pour lesquels les effets sont différenciés en fonction des environnements (sensibilité différentielle). Sur la base de cette réflexion, on peut alors admettre qu'il sera plus difficile d'identifier un QTL stable lorsque l'on travaille sur un réseau étendu. La notion même de stabilité statique d'un QTL (même effet quel que soit les environnements) peut être remise en question, en partant du postulat que les QTL constitutifs du rendement et qui pourraient correspondre à des QTL stables ont déjà été fixés dans les populations, ceux qui sont détectés relèvent alors de QTL interactifs de type 'sensibilité différentielle'. En reprenant ces définitions, on peut limiter la définition des QTL interactifs aux QTL ayant des effets antagonistes en fonction des environnements.



## IV.3. Utilisation des envirotypes en sélection

---

### IV.3.1. Les QTL détectés à l'échelle des envirotypes mettent en évidence des zones d'intérêt pour la sélection assistée par marqueurs

La dernière question de ce travail de thèse consistait à identifier les déterminants génétiques qui contrôlent l'interaction et voir s'ils pouvaient être caractérisés en fonction de la qualification des envirotypes. Les envirotypes ont permis d'identifier 5 QTL de rendement et de nombre de grains, dont 4 confirment des QTL détectés à l'échelle des essais individuels. Parmi les QTL spécifiques aux envirotypes, nous avons pu mettre en évidence un QTL qui n'était pas fixé dans les lignées élites d'un obtenteur. Par ailleurs, un second QTL, spécifique à l'envirotype le plus stressé, n'est pas fixé dans le matériel récent, contrairement aux QTL stables.

Au sein des programmes de sélection assistée par marqueurs, les QTL spécifiques aux envirotypes sont des cibles à privilégier ou à cumuler aux QTL stables. Pour aller plus loin dans la compréhension de ces QTL interactifs, on peut essayer d'investiguer le lien entre l'effet du QTL et la réponse à une combinaison de facteurs limitants en se basant sur le profil de facteurs limitants de l'envirotype considéré. Dans la suite nous allons discuter de moyens pour tester cette interaction et l'utiliser dans la construction de réseaux permettant de tester la spécificité de QTL déjà identifiés par exemple. Nous discuterons également de l'intérêt d'utiliser les facteurs limitants et les envirotypes en sélection génomique.

### IV.3.2. Relations entre QTL et facteurs limitants

Lorsqu'un QTL est spécifique à un envirotype, il est possible d'intégrer des covariables environnementales dans des modèles de régression factorielle afin d'exprimer l'effet du QTL en fonction des différentes conditions environnementales (Malosetti et al., 2004 ; Malosetti et al., 2013 ; Garin et al., 2020). Cette méthode de régression a notamment été utilisée sur le riz et a mis en évidence des interactions significatives entre des QTL et des covariables environnementales telles que les précipitations, l'humidité ou la température. Ce type d'approche permet de mieux comprendre l'interaction grâce à l'identification de mécanismes biologiques qui sont mis en jeu (Monteverde et al., 2019). L'envirotypage offre la possibilité de prioriser un petit nombre de covariables environnementales candidates susceptibles d'expliquer l'interaction QTL×E. L'interaction



entre les QTL et les facteurs limitants peut être testée afin de mettre en relation le ou les facteurs ayant permis de mettre en évidence ces QTL dans leur envirotypes respectifs. L'identification d'une interaction 'QTL × facteur' limitant significative peut aider à définir des hypothèses biologiques impliquées dans l'interaction. Les environnements exprimant un facteur limitant impliqué dans l'interaction, peuvent alors être incorporés au sein des réseaux expérimentaux et des réseaux d'évaluation des variétés et ainsi permettre d'identifier des allèles favorables dans des conditions climatiques défavorables à l'image du QTL A07.

### IV.3.3. Utilisation des facteurs limitants et des envirotypes en sélection génomique

Nous avons pu voir l'utilisation potentielle des envirotypes en sélection assistée par marqueurs, mais ils peuvent également être employés dans le domaine de la sélection génomique. En effet plusieurs modèles de types gBLUP ont été adaptés en modifiant l'effet de l'environnement par une régression aléatoire sur des covariables environnementales (Jarquin et al., 2014). Ces modèles offrent un gain de la prédiction comprise entre 17% et 34% par rapport aux modèles gBLUP classiques. Un modèle similaire, noté FR-gBLUP décrit par Ly et al. (2018), repose quant à lui sur l'utilisation d'une seule covariable environnementale. Ce modèle, appliqué sur des regroupements d'environnements, permet également de gagner en qualité de prédiction. Cependant les auteurs soulignent que ce résultat est dépendant de la part de l'interaction G×E restante dans les groupes, le modèle étant plus efficace lorsque l'interaction explique une forte part de la variation. Nos envirotypes, ainsi que les facteurs limitants identifiés, pourraient être utilisés dans les modèles de sélection génomique développés sur colza.

Afin de commencer à investiguer l'impact des envirotypes sur les modèles de sélection génomique, nous avons initié des premiers essais lors de la thèse et dans le cadre du projet Rapsodyn. Pour cela, nous avons utilisé des modèles gBLUP calibrés soit sur les effets génétiques estimés à l'échelle du réseau global, soit sur les effets génétiques estimés à l'échelle de chaque envirotype. Dans le cas où ces modèles sont utilisés pour prédire un nouveau génotype (non présent dans la population de calibration) dans un environnement donné, des résultats contrastés ont été observés. Parmi les 5 modèles testés par envirotype (un par envirotype) 50% d'entre eux offrent une meilleure prédiction par rapport à un modèle calibré sur l'ensemble du réseau. Dans la situation où l'on souhaite prédire le comportement d'un génotype connu dans un nouvel environnement, il est nécessaire d'assigner cet environnement à un envirotype en fonction des valeurs des indicateurs environnementaux correspondant aux facteurs limitants identifiés à l'échelle du réseau. On compare ensuite si la



prédiction est meilleure avec l'équation envirotype qu'avec l'équation du réseau complet. Les résultats ont montré que le modèle calibré sur le réseau est plus performant (Retailleau, 2018).

Ainsi, les facteurs limitants ainsi que les envirotypes utilisés en sélection génomique sur le colza offrent la possibilité d'accroître les prédictions et ainsi de favoriser la création de variétés plus adaptées à certaines conditions pédoclimatiques en optimisant les interactions G×E.



# REFERENCES BIBLIOGRAPHIQUES

---



- Agreste, 2019. Pratiques culturales en grandes cultures 2017 : IFT et nombre de traitements [WWW Document]. URL [https://agreste.agriculture.gouv.fr/agreste-web/download/publication/publie/Chd1903/cd2019-3 PK \\_ janvier 2020 v2.pdf](https://agreste.agriculture.gouv.fr/agreste-web/download/publication/publie/Chd1903/cd2019-3 PK _ janvier 2020 v2.pdf)
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. Irrig. Drain. 1–15. <https://doi.org/10.1016/j.eja.2010.12.001>
- Allender, C.J., King, G.J., 2010. Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biol.* 10. <https://doi.org/10.1186/1471-2229-10-54>
- Angadi, S. V., Cutforth, H.W., Miller, P.R., Mcconkey, B.G., Entz, M.H., Brandt, S.A., 2000. Response of three *Brassica* species to high temperature stress during reproductive growth. *Can. J. Plant Sci.* 80, 693–702.
- Annicchiarico, P., 2002. Defining adaptation strategies and yield stability targets in breeding programmes. *Quant. Genet. genomics, plant Breed.* 365–383.
- Annicchiarico, P., Bellah, F., Chiari, T., 2005. Defining Subregions and Estimating Benefits for a Specific-Adaptation Strategy by Breeding Programs: A Case Study. *Crop Sci.* 45, 1741–1749. <https://doi.org/10.2135/cropsci2004.0524>
- Astle, W., Balding, D.J., 2009. Population Structure and Cryptic Relatedness in Genetic Association Studies. *Stat. Sci.* 24, 451–471. <https://doi.org/10.1214/09-STS307>
- Avice, J.C., Etienne, P., 2014. Leaf senescence and nitrogen remobilization efficiency in oilseed rape (*Brassica napus* L.). *J. Exp. Bot.* 65, 3813–3824. <https://doi.org/10.1093/jxb/eru177>
- Bagot, P., Salvi, F., Gombert, J., 2014. Etude préliminaire à la caractérisation du comportement des variétés de colza oléagineux d’hiver dans des itinéraires techniques particuliers sur la base du réseau CTPS existant. *Innov. Agron.* 35, 61–73.
- Baker, R.J., 1988. Tests for crpssover genotype-environmental interactions. *Can. J. Plant Sci.* 68, 405–410.
- Barbottin, A., 2004. Utilisation d ’ un modèle de culture pour évaluer le comportement des génotypes : Pertinence de l ’ utilisation d ’ Azodyn pour analyser la variabilité du rendement et de la teneur en protéines du blé tendre. Ecole Doctorale ABIES.
- Basunanda, P., Radoev, M., Ecke, W., Friedt, W., Becker, H.C., Snowdon, R.J., 2009. Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 120, 271. <https://doi.org/10.1007/s00122-009-1133-z>
- Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Using lme4. *J. Stat. Softw.* 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Baux, A., Wegmüller, J., Holzkämper, A., 2015. Exploring Climatic Impact on Oilseed Rape Yield in Switzerland. *Procedia Environ. Sci.* 29, 123. <https://doi.org/10.1016/j.proenv.2015.07.209>
- Bayer, P.E., Hurgobin, B., Golicz, A.A., Chan, C.K., Yuan, Y., Renton, M., Meng, J., Li, R., Long, Y., Zou, J., Bancroft, I., Chalhoub, B., King, G.J., Batley, J., Edwards, D., 2017. Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnol. J.* 15, 1602–1610. <https://doi.org/10.1111/pbi.12742>
- Becker, H.C., Engqvist, G.M., 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>
- Becker, H.C., Leon, J., 1988. Stability Analysis in Plant Breeding. *Plant Breed.* 101, 1–23. <https://doi.org/10.1111/j.1439-0523.1988.tb00261.x>
- Bell, J.L., Sloan, L.C., Snyder, M.A., 2004. Regional changes in extreme climatic events: A future climate scenario. *J. Clim.* 17, 81–87. [https://doi.org/10.1175/1520-0442\(2004\)017](https://doi.org/10.1175/1520-0442(2004)017)
- BIPE, Sofiprotéol, 2014. La filière des oléoprotéagineux pour répondre aux enjeux globaux à horizon 2030.
- Bouchereau, A., Clossais-Besnard, N., Bensaoud, A., Leport, L., Renard, M., 1996. Water stress effects on



- rapeseed quality. *Eur. J. Agron.* 5, 19–30. [https://doi.org/https://doi.org/10.1016/S1161-0301\(96\)02005-9](https://doi.org/https://doi.org/10.1016/S1161-0301(96)02005-9)
- Bouchet, A.-S., 2015. Architecture génétique du rendement et de ses composantes chez le colza d'hiver (*Brassica napus* L.) cultivé sous contrainte azotée. Université de Rennes 1.
- Bouchet, A.-S., Laperche, A., Bissuel-Belaygue, C., Baron, C., Morice, J., Rousseau-Gueutin, M., Dheu, J.-E., George, P., Pinochet, X., Foubert, T., Maes, O., Dugué, D., Guinot, F., Nesi, N., 2016a. Genetic basis of nitrogen use efficiency and yield stability across environments in winter rapeseed. *BMC Genet.* 17, 131. <https://doi.org/10.1186/s12863-016-0432-z>
- Bouchet, A.-S., Laperche, A., Bissuel-Belaygue, C., Snowdon, R., Nesi, N., Stahl, A., 2016b. Nitrogen use efficiency in rapeseed. A review. *Agron. Sustain. Dev.* 36. <https://doi.org/10.1007/s13593-016-0371-0>
- Bouchet, A.-S., Nesi, N., Bissuel, C., Bregeon, M., Laripe, A., Navier, H., Ribière, N., Orsel, M., Grezes-Besset, B., Renard, M., Laperche, A., 2014. Genetic control of yield and yield components in winter oilseed rape (*Brassica napus* L.) grown under nitrogen limitation. *Euphytica* 199, 183–205. <https://doi.org/10.1007/s10681-014-1130-4>
- Bouffier, B., Derory, J., Murigneux, A., Reynolds, M., Le Gouis, J., 2015. Clustering of environmental parameters discriminates drought and heat stress bread wheat trials. *Agron. J.* 107, 1489–1503. <https://doi.org/10.2134/agronj14.0452>
- Braatz, J., Harloff, H.J., Mascher, M., Stein, N., Himmelbach, A., Jung, C., 2017. CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* 174, 935–942. <https://doi.org/10.1104/pp.17.00426>
- Brancourt-Hulmel, M., Biarnès-Dumoulin, V., Denis, J., 1997. Points de repère dans l'analyse de la stabilité et de l'interaction génotype-milieu en amélioration des plantes. *Agronomie* 17, 219–246.
- Brancourt-Hulmel, M., Denis, J.B., Lecomte, C., 2000. Determining environmental covariates which explain genotype environment interaction in winter wheat through probe genotypes and biadditive factorial regression. *Theor. Appl. Genet.* 100, 285–298. <https://doi.org/10.1007/s001220050038>
- Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., et al., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie* 18, 311–346. <https://doi.org/10.1051/agro:19980501>
- Browning, B.L., Browning, S.R., 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am. J. Hum. Genet.* 84, 210–223. <https://doi.org/https://doi.org/10.1016/j.ajhg.2009.01.005>
- Bruand, A., Duval, O., Cousin, I., 2004. Estimation des propriétés de rétention en eau des sols à partir de la base de données SOLHYDRO: Une première proposition combinant le type d'horizon, sa texture et sa densité apparente. *Étude Gest. des Sols* 11, 323–334.
- Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J.C., Goodman, M.M., Harjes, C., Guill, K., et al., 2009. The Genetic Architecture of Maize Flowering Time. *Science* (80-. ). 325, 714 LP – 718. <https://doi.org/10.1126/science.1174276>
- Bus, A., Körber, N., Snowdon, R.J., 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>
- Butruille, D. V., Guries, R.P., Osborn, T.C., 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>



- Cadoux, S., Sauzet, G., Valantin-Morison, M., Pontet, C., Champolivier, L., Robert, C., Lieven, J., Flénet, F., Mangenot, O., Fauvin, P., Landé, N., 2015. Intercropping frost-sensitive legume crops with winter oilseed rape reduces weed competition, insect damage, and improves nitrogen use efficiency. *OCL* 22.
- Cai, D., Xiao, Y., Yang, W., Ye, W., Wang, B., Younas, M., Wu, J., Liu, K., 2014. Association mapping of six yield-related traits in rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 127, 85–96. <https://doi.org/10.1007/s00122-013-2203-9>
- Cai, G., Yang, Qingyong, Chen, H., Yang, Qian, Zhang, C., Fan, C., Zhou, Y., 2016. Genetic dissection of plant architecture and yield-related traits in *Brassica napus*. *Sci. Rep.* 6, 21625. <https://doi.org/10.1038/srep21625>
- Casadebaig, P., Guilioni, L., Lecoœur, J., Christophe, A., Champolivier, L., Debaeke, P., 2011. SUNFLO, a model to simulate genotype-specific performance of the sunflower crop in contrasting environments. *Agric. For. Meteorol.* 151, 163–178. <https://doi.org/10.1016/j.agrformet.2010.09.012>
- Ceccarelli, S., 1989. Wide adaptation: How wide? *Euphytica* 40, 197–205. <https://doi.org/10.1007/BF00024512>
- Chagné, D., Bianco, L., Lawley, C., Micheletti, D., Jacobs, J.M.E., 2015. Methods for the Design, Implementation, and Analysis of Illumina Infinium SNP Assays in Plants, in: Batley, J. (Ed.), *Plant Genotyping: Methods and Protocols*. Springer New York, New York, NY, pp. 281–298. [https://doi.org/10.1007/978-1-4939-1966-6\\_21](https://doi.org/10.1007/978-1-4939-1966-6_21)
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M., Da Silva, C., Just, J., et al., 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* (80-. ). 345, 950–953. <https://doi.org/10.1126/science.1253435>
- Champolivier, L., Merrien, A., 1996. Effects of water stress applied at different growth stages to *Brassica napus* L. var. *oleifera* on yield, yield components and seed quality. *Eur. J. Agron.* 5, 153–160. [https://doi.org/10.1016/S1161-0301\(96\)02004-7](https://doi.org/10.1016/S1161-0301(96)02004-7)
- Chandler, J., Corbesier, L., Spielmann, P., Dettendorfer, J., Stahl, D., Apel, K., Melzer, S., 2005. Modulating flowering time and prevention of pod shatter in oilseed rape. *Mol. Breed.* 15, 87–94. <https://doi.org/10.1007/s11032-004-2735-4>
- Chao, H., Raboanatahiry, N., Wang, X., Zhao, W., Chen, L., Guo, L., Li, B., Hou, D., Pu, S., Zhang, L., Wang, H., Wang, B., Li, M., 2019. Genetic dissection of harvest index and related traits through genome-wide quantitative trait locus mapping in *Brassica napus* L. *Breed. Sci.* 69, 104–116. <https://doi.org/10.1270/jsbbs.18115>
- Chapman, S.C., 2008. Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. *Euphytica* 161, 195–208. <https://doi.org/10.1007/s10681-007-9623-z>
- Charbonnier, E., Fugerey-Scarbel, A., Lemarié, S., 2019. Rapeseed: how to value varieties with higher nitrogen use efficiency in France. *OCL* 26, 26. <https://doi.org/10.1051/ocl/2019021>
- Chen, G., Geng, J., Rahman, M., Liu, X., Tu, J., Fu, T., Li, G., McVetty, P.B.E., Tahir, M., 2010. Identification of QTL for oil content, seed yield, and flowering time in oilseed rape (*Brassica napus*). *Euphytica* 175, 161–174. <https://doi.org/10.1007/s10681-010-0144-9>
- Cheng, C., Krishnakumar, V., Chan, A.P., Town, C.D., 2016. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89, 789–804. <https://doi.org/10.1111/tpj.13415>
- Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., Bonnema, G., Wang, X., 2012. Biased Gene Fractionation and Dominant Gene Expression among the Subgenomes of *Brassica rapa*. *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0036442>
- Chenu, K., Cooper, M., Hammer, G.L., Mathews, K.L., Dreccer, M.F., Chapman, S.C., 2011. Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. *J. Exp. Bot.* 62, 1743–1755. <https://doi.org/10.1093/jxb/erq459>



- Clarke, W.E., Higgins, E.E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., Batley, J., Edwards, D., Meng, J., Li, R., Lawley, C.T., Pauquet, J., Laga, B., 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>
- Colnenne, C., Meynard, J., Reau, R., Justes, E., Merrien, A., 1998. Determination of a Critical Nitrogen Dilution Curve for Winter Oilseed Rape. *Ann. Bot.* 81, 311–317. <https://doi.org/10.1006/anbo.1997.0557>
- Corlouer, E., Gauffreteau, A., Bouchet, A.S., Bissuel-Bélaygue, C., Nesi, N., Laperche, A., 2019. Envirotypes based on seed yield limiting factors allow to tackle G × E interactions. *Agronomy* 9, 1–18. <https://doi.org/10.3390/agronomy9120798>
- Dalton-Morgan, J., Hayward, A., Alamery, S., Tollenaere, R., Mason, A.S., Campbell, E., Patel, D., Lorenc, M.T., Yi, B., Long, Y., Meng, J., Raman, R., Raman, H., Lawley, C., Edwards, D., Batley, J., 2014. A high-throughput SNP array in the amphidiploid species *Brassica napus* shows diversity in resistance genes. *Funct. Integr. Genomics* 14, 643–655. <https://doi.org/10.1007/s10142-014-0391-2>
- Dejoux, J.-F., Recous, S., Meynard, J.-M., Trinsoutrot, I., Leterme, P., 2000. The fate of nitrogen from winter-frozen rapeseed leaves: mineralization, fluxes to the environment and uptake by rapeseed crop in spring. *Plant Soil* 218, 257–272. <https://doi.org/10.1023/A:1014934924819>
- Delourme, R., Chèvre, A.M., Brun, H., Rouxel, T., Balesdent, M.H., Dias, J.S., Salisbury, P., Renard, M., Rimmer, S.R., 2006. Major Gene and Polygenic Resistance to *Leptosphaeria maculans* in Oilseed Rape (*Brassica napus*). *Eur. J. Plant Pathol.* 114, 41–52. <https://doi.org/10.1007/s10658-005-2108-9>
- Delourme, R., Falentin, C., Fomeju, B.F., Boillot, M., Lassalle, G., André, I., Duarte, J., Gauthier, V., Lucante, N., Marty, A., Pauchon, M., Pichon, J.P., Ribière, N., Trotoux, G., Blanchard, P., Rivière, N., Martinant, J.P., Pauquet, J., 2013. High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC Genomics* 14, 1–18. <https://doi.org/10.1186/1471-2164-14-120>
- Delourme, R., Laperche, A., Bouchet, A., Jubault, M., Paillard, S., Nesi, N., 2018. Genes and Quantitative Trait Loci Mapping for Major Agronomic Traits in *Brassica napus* L., in: *The Brassica Napus Genome. Compendium of Plant Genomes.* Springer International Publishing, pp. 41–85. <https://doi.org/10.1007/978-3-319-43694-4>
- Denis, J.-B., 1988. Two way analysis using covariates. *Stat. A J. Theor. Appl. Stat.* 19, 123–132. <https://doi.org/DOI:10.1080/02331888808802080>
- Des Marais, D.L., Hernandez, K.M., Juenger, T.E., 2013. Genotype-by-environment interaction and plasticity: Exploring genomic responses of plants to the abiotic environment. *Annu. Rev. Ecol. Evol. Syst.* 44, 5–29. <https://doi.org/10.1146/annurev-ecolsys-110512-135806>
- Diepenbrock, W., 2000. Yield analysis of winter oilseed rape (*Brassica napus* L.): A review. *F. Crop. Res.* 67, 35–49. [https://doi.org/10.1016/S0378-4290\(00\)00082-4](https://doi.org/10.1016/S0378-4290(00)00082-4)
- Diers, B.W., Osborn, T.C., 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>
- Ding, G., Zhao, Z., Liao, Y., Hu, Y., Shi, L., Long, Y., Xu, F., 2012. Quantitative trait loci for seed yield and yield-related traits, and their responses to reduced phosphorus supply in *Brassica napus*. *Ann. Bot.* 109, 747–759. <https://doi.org/10.1093/aob/mcr323>
- Ding, G., Zhao, Z., Wang, L., Zhang, D., Shi, L., Xu, F., 2014. Identification and multiple comparisons of QTL and epistatic interaction conferring high yield under boron and phosphorus deprivation in *Brassica napus*. *Euphytica* 198, 337–351. <https://doi.org/10.1007/s10681-014-1110-8>
- Dufey, I., Hiel, M.-P., Hakizimana, P., Draye, X., Lutts, S., Koné, B., Dramé, K.N., Konaté, K.A., Sie, M., Bertin, P., 2012. Multienvironment Quantitative Trait Loci Mapping and Consistency across Environments of Resistance Mechanisms to Ferrous Iron Toxicity in Rice. *Crop Sci.* 52, 539–550. <https://doi.org/10.2135/cropsci2009.09.0544>
- Ecke, W., Clemens, R., Honsdorf, N., Becker, H.C., 2010. Extent and structure of linkage disequilibrium in canola



- quality winter rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 120, 921–931. <https://doi.org/10.1007/s00122-009-1221-0>
- Edwards, D., Batley, J., Snowdon, R.J., 2013. Accessing complex crop genomes with next-generation sequencing. *Theor. Appl. Genet.* 126, 1–11. <https://doi.org/10.1007/s00122-012-1964-x>
- El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M., Aarts, M.G.M., 2014. Genotype × environment interaction QTL mapping in plants: Lessons from Arabidopsis. *Trends Plant Sci.* 19, 390–398. <https://doi.org/10.1016/j.tplants.2014.01.001>
- Elferjani, R., Soolanayakanahally, R., 2018. Canola responses to drought, heat, and combined stress: shared and specific effects on carbon assimilation, seed yield, and oil composition. *Front. Plant Sci.* 9, 1–17. <https://doi.org/10.3389/fpls.2018.01224>
- Falconer, D.S., 1990. Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. *Genet. Res.* 56, 57–70. <https://doi.org/DOI: 10.1017/S0016672300028883>
- Fan, C., Cai, G., Qin, J., Li, Q., Yang, M., Wu, J., Fu, T., Liu, K., Zhou, Y., 2010. Mapping of quantitative trait loci and development of allele-specific markers for seed weight in *Brassica napus*. *Theor. Appl. Genet.* 121, 1289–1301. <https://doi.org/10.1007/s00122-010-1388-4>
- Fan, J., McConkey, B., Wang, H., Janzen, H., 2016. Root distribution by depth for temperate agricultural crops. *F. Crop. Res.* 189, 68–74. <https://doi.org/10.1016/j.fcr.2016.02.013>
- FAOSTAT, 2020. Production/Yield quantities of Rapeseed in World [WWW Document]. URL <http://www.fao.org/faostat/en/?#data/QC/visualize>.
- Fasahat, P., Rajabi, A., Mahmoudi, S.B., Noghabi, M.A., Rad, J.M., 2015. An overview on the use of stability parameters in plant breeding. *Biometrics Biostat. Int. J.* 2, 149–159. <https://doi.org/10.15406/bbij.2015.02.00043>
- Fattahi, F., Fakheri, B.A., Solouki, M., Möllers, C., Rezaizad, A., 2018. Mapping QTL controlling agronomic traits in a doubled haploid population of winter oilseed rape (*Brassica napus* L.). *J. Genet.* 97, 1389–1406. <https://doi.org/10.1007/s12041-018-1044-3>
- Ferreira, M.E., Satagopan, J., Yandell, B.S., Williams, P.H., Osborn, T.C., 1995. Mapping loci controlling vernalization requirement and flowering time in *Brassica napus*. *Theor. Appl. Genet.* 90, 727–732.
- Finlay, K.W., Wilkinson, G.N., 1963. The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.* 14, 742–754. <https://doi.org/10.1071/AR9630742>
- Fischer, R.A., 1985. Number of kernels in wheat crops and the influence of solar radiation and temperature. *J. Agric. Sci.* <https://doi.org/10.1017/CBO9781107415324.004>
- Fletcher, R.S., Mullen, J.L., Heiliger, A., McKay, J.K., 2015. QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in *Brassica napus*. *J. Exp. Bot.* 66, 245–256. <https://doi.org/10.1093/jxb/eru423>
- Friedt, W., Tu, J., Fu, T., 2018. Academic and Economic Importance of *Brassica napus* Rapeseed., in: *The Brassica Napus Genome. Compendium of Plant Genomes.* pp. 1–20. [https://doi.org/10.1007/978-3-319-43694-4\\_1](https://doi.org/10.1007/978-3-319-43694-4_1)
- Fu, D., Mason, A.S., Xiao, M., Yan, H., 2016. Plant Science Effects of genome structure variation , homeologous genes and repetitive DNA on polyploid crop research in the age of genomics. *Plant Sci.* 242, 37–46. <https://doi.org/10.1016/j.plantsci.2015.09.017>
- Gabrielle, B., Denoroy, P., Gosse, G., Justes, E., Andersen, M.N., 1998. Development and evaluation of a CERES-type model for winter oilseed rape. *F. Crop. Res.* 57, 95–111. [https://doi.org/10.1016/S0378-4290\(97\)00120-2](https://doi.org/10.1016/S0378-4290(97)00120-2)
- Gammelvind, L.H., Schjoerring, J.K., Mogensen, V.O., Jensen, C.R., Bock, J.G.H., 1996. Photosynthesis in leaves and siliques of winter oilseed rape (*Brassica napus* L.). *Plant Soil* 186, 227–236. <https://doi.org/10.1007/BF02415518>



- Ganal, M., Polley, A., Plieske, J., 2018. SNP genotyping with optimized molecular marker sets for plant breeding in the Brassica species (*B. napus*, *B. oleracea* and *B. rapa*), in: 21st Crucifer Genetics Conference. Saint-Malo.
- Ganal, M.W., Polley, A., Graner, E.-M., Plieske, J., Wieseke, R., Luerssen, H., Durstewitz, G., 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, E.R., 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>
- Garin, V., Malosetti, M., van Eeuwijk, F., 2020. Multi-parent multi-environment QTL analysis: an illustration with the EU-NAM Flint population. *Theor. Appl. Genet.* 133, 2627–2638. <https://doi.org/10.1007/s00122-020-03621-0>
- Gauch, H.G., 1992. *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*. Elsevier.
- Gauch, H.G., 1988. Model Selection and Validation for Yield Trials with Interaction. *Biometrics* 44, 705. <https://doi.org/10.2307/2531585>
- Gauchi, J.-P., Chagnon, P., 2001. Comparison of selection methods of explanatory variables in PLS regression with application to manufacturing process data. *Chemom. Intell. Lab. Syst.* 58, 171–193. [https://doi.org/10.1016/S0169-7439\(01\)00158-7](https://doi.org/10.1016/S0169-7439(01)00158-7)
- Gehring, A., Snowdon, R., Spiller, T., Basunanda, P., Friedt, W., 2007. New Oilseed Rape (*Brassica napus*) Hybrids with High Levels of Heterosis for Seed Yield under Nutrient-poor Conditions. *Breed. Sci.* 57, 315–320. <https://doi.org/10.1270/jsbbs.57.315>
- Génard, T., Etienne, P., Diquélou, S., Yvin, J.-C., Revellin, C., Laîné, P., 2017. Rapeseed-legume intercroppings: plant growth and nitrogen balance in early stages of growth and development. *Heliyon* 3, e00261. <https://doi.org/10.1016/j.heliyon.2017.e00261>
- Gervois, S., Clément, C., Chabert, T., Valantin-Morison, M., Pinochet, X., Laperche, A., 2019. AZODYN-rapeseed: a dynamic crop model to simulate the performance of rapeseed crop in contrasting environments, in: 15th International Rapeseed Congress. Berlin.
- Gourrion, A., Simon, C., Vallée, P., Delourme, R., Chatre, S., Dheu, J.É., 2020. Enlarging the genetic diversity of winter oilseed rape (WOSR) by crossing with spring oilseed rape (SOSR). *OCL - Oilseeds fats, Crop. Lipids* 27. <https://doi.org/10.1051/ocl/2020013>
- Gül, M.K., Becker, H., Ecke, W., 2003. QTL mapping and analysis of QTL x nitrogen interactions for protein and oil contents in *Brassica napus* L, in: 11th International Rapeseed Congress. Copenhagen.
- Habekotté, B., 1997a. Evaluation of seed yield determining factors of winter oilseed rape (*Brassica napus* L.) by means of crop growth modelling. *F. Crop. Res.* 54, 137–151.
- Habekotté, B., 1997b. Description, parameterization and user guide of LINTUL-BRASNAP 1.1 : a crop growth model of winter oilseed rape (*Brassica napus* L.). AB-DLO, Wageningen.
- Harloff, H.J., Lemcke, S., Mittasch, J., Frolov, A., Wu, J.G., Dreyer, F., Leckband, G., Jung, C., 2012. A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. *Theor. Appl. Genet.* 124, 957–969. <https://doi.org/10.1007/s00122-011-1760-z>
- Hasan, M., Seyis, F., Badani, A.G., Pons-Kühnemann, J., Friedt, W., Lühs, W., Snowdon, R.J., 2005. 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, S. V., Frisch, M., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.H., Leckband, G., Abbadi, A., Snowdon, R.J., 2015. Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. *Front. Plant Sci.* 6, 1–13. <https://doi.org/10.3389/fpls.2015.00221>
- Hayward, A., Mason, A.S., Manuel, J.D., David, Z., Jacqueline, E., 2012. SNP discovery and applications in *Brassica napus*. *J Plant Biotechnol.* 39, 49–61.



- He, D., Wang, E., Wang, J., Lilley, J.M., 2017. Genotype × environment × management interactions of canola across China: A simulation study. *Agric. For. Meteorol.* 247, 424–433. <https://doi.org/10.1016/j.agrformet.2017.08.027>
- Hebinger, H., 2013. *Le Colza*, Editions F. ed.
- Huang, H., Cui, T., Zhang, L., Yang, Q., Yang, Y., Xie, K., Fan, C., Zhou, Y., 2020. Modifications of fatty acid profile through targeted mutation at BnaFAD2 gene with CRISPR/Cas9-mediated gene editing in *Brassica napus*. *Theor. Appl. Genet.* 133, 2401–2411. <https://doi.org/10.1007/s00122-020-03607-y>
- Husson, F., Lê, S., Pagès, J., 2017. *Computer Science and Data Analysis Series Exploratory Multivariate Analysis by Example Using R*.
- Imelfort, M., Duran, C., Batley, J., Edwards, D., 2009. Discovering genetic polymorphisms in next-generation sequencing data. *Plant Biotechnol. J.* 312–317. <https://doi.org/10.1111/j.1467-7652.2009.00406.x>
- 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>
- Jamont, M., Piva, G., Fustec, J., 2013. Sharing N resources in the early growth of rapeseed intercropped with faba bean: does N transfer matter? *Plant Soil* 371, 641–653. <https://doi.org/10.1007/s11104-013-1712-2>
- Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Pérez, P., Calus, M., Burgueño, J., de los Campos, G., 2014. A reaction norm model for genomic selection using high dimensional genomic and environmental data. *Theor. Appl. Genet.* 127, 595–607. <https://doi.org/10.1007/s00122-013-2243-1>
- Jenczewski, E., Chèvre, A.M., 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.ch10>
- Jestin, C., Lodé, M., Vallée, P., Domin, C., Falentin, C., Horvais, R., Coedel, S., Manzanares-Dauleux, M.J., Delourme, R., 2011. Association mapping of quantitative resistance for *Leptosphaeria maculans* in oilseed rape (*Brassica napus* L.). *Mol. Breed.* 27, 271–287. <https://doi.org/10.1007/s11032-010-9429-x>
- Jeuffroy, M.-H., Valantin-Morison, M., Champolivier, L., Reau, R., 2006. Azote, rendement et qualité des graines : mise au point et utilisation du modèle Azodyn-colza pour améliorer les performances du colza vis-à-vis de l'azote. *OCL* 13, 388–392.
- Jeuffroy, M.H., Recous, S., 1999. Azodyn: A simple model simulating the date of nitrogen deficiency for decision support in wheat fertilization. *Eur. J. Agron.* 10, 129–144. [https://doi.org/10.1016/S1161-0301\(98\)00059-8](https://doi.org/10.1016/S1161-0301(98)00059-8)
- Jiang, C., Zeng, Z., 1995. Jiang CJ, Zeng ZB.. Multiple-trait analysis of genetic-mapping for quantitative trait loci. *Genetics* 140: 1111-1127. *Genetics* 140, 1111–1127.
- Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J.E., McKain, M.R., McNeal, J., Rolf, M., Ruzicka, D.R., Wafula, E., Wickett, N.J., 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>
- Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., Dean, C., 2000. Molecular Analysis of *FRIGIDA*, a Major Determinant of Natural Variation in *Arabidopsis* Flowering Time. *Science* (80-. ). 290, 344 LP – 347. <https://doi.org/10.1126/science.290.5490.344>
- Josse, J., Husson, F., 2016. missMDA: A package for handling missing values in multivariate data analysis. *J. Stat. Softw.* 70. <https://doi.org/10.18637/jss.v070.i01>
- Jullien, A., Mathieu, A., Allirand, J.M., Pinet, A., De Reffye, P., Cournède, P.H., Ney, B., 2011. Characterization of the interactions between architecture and sourcesink relationships in winter oilseed rape (*Brassica napus*) using the GreenLab model. *Ann. Bot.* 107, 765–779. <https://doi.org/10.1093/aob/mcq205>
- Körber, N., Bus, A., Li, J., Parkin, I.A.P., Wittkop, B., Snowdon, R.J., Stich, B., 2016. Agronomic and Seed Quality



- Traits Dissected by Genome-Wide Association Mapping in *Brassica napus* *Front. Plant Sci.* 7:386 doi: 10.3389/fpls.2016.00386 .
- Kräling, K., 1987. Utilization of Genetic Variability of Resynthesized Rapeseed. *Plant Breed.* 99, 209–217. <https://doi.org/10.1111/j.1439-0523.1987.tb01174.x>
- Krzanowski, W.J., Lai, Y.T., 1988. A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. *Biometrics* 44, 23. <https://doi.org/10.2307/2531893>
- Kuznetsova, A., Brockhoff, P., Christensen, R., 2017. lmerTest Package : Tests in Linear Mixed Effects Models. *J. Stat. Softw.* 82. <https://doi.org/10.18637/jss.v082.i13>
- Lacoste, M., Mulder, V.L., Richer-De-Forges, A.C., Martin, M.P., Arrouays, D., 2016. Evaluating large-extent spatial modeling approaches: A case study for soil depth for France. *Geoderma Reg.* 7, 137–152. <https://doi.org/10.1016/j.geodrs.2016.02.006>
- Lancashire, P.D., Bleiholder, H., Van Den BOOM, T., Langelüddeke, P., Stauss, R., Weber, E., Witzemberger, A., 1991. A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.* 119, 561–601. <https://doi.org/10.1111/j.1744-7348.1991.tb04895.x>
- Lê, S., Josse, J., Husson, F., 2008. FactoMineR : A Package for Multivariate Analysis. *J. Stat. Softw.* 25, 1–18. <https://doi.org/10.18637/jss.v025.i01>
- Leach, J.E., Darby, R.J., Williams, I.H., Fitt, B.D.L., Rawlinson, C.J., 1994. Factors affecting growth and yield of winter oilseed rape (*Brassica napus*), 1985–89. *J. Agric. Sci.* 122, 405–413. <https://doi.org/10.1017/S0021859600067332>
- Lecomte, C., 2005. L'évaluation expérimentale des innovations variétales. Proposition d'outils d'analyse de l'interaction génotype - milieu adaptés à la diversité des besoins et des contraintes des acteurs de la filière semences . INA Paris-Grignon.
- Lee, H.T., Chawla, H.S., Obermeier, C., Dreyer, F., Abadi, A., Snowdon, R., 2020. Chromosome-Scale Assembly of Winter Oilseed Rape *Brassica napus*. *Front. Plant Sci.* 11, 1–13. <https://doi.org/10.3389/fpls.2020.00496>
- Leflon, M., Eber, F., Letanneur, J.C., Chelysheva, L., Coriton, O., Huteau, V., Ryder, C.D., Barker, G., Jenczewski, E., Chèvre, A.M., 2006. Pairing and recombination at meiosis of *Brassica rapa* (AA) × *Brassica napus* (AACC) hybrids. *Theor. Appl. Genet.* 113, 1467–1480. <https://doi.org/10.1007/s00122-006-0393-0>
- Leterme, P., 1985. Modélisation de la croissance et de la production des siliques chez le colza d'hiver (*Brassica napus* L.). INA Paris-Grignon.
- Leveugle, M., Duborjal, H., West, M., Pothier, A., Clementi, A., Nesi, N., Dyrszka, E., Pichon, J., Riviere, N., 2015. Rapsodyn Whole Exome capture and genotyping in the polyploid *Brassica napus*, in: JOBIM. Clermont-Ferrand.
- Li, N., Shi, J., Wang, X., Liu, G., Wang, H., 2014. A combined linkage and regional association mapping validation and fine mapping of two major pleiotropic QTLs for seed weight and silique length in rapeseed (*Brassica napus* L.). *BMC Plant Biol.* 14, 114. <https://doi.org/10.1186/1471-2229-14-114>
- Lin, C.S., Binns, M.R., Lefkovich, L.P., 1986. Stability Analysis: Where Do We Stand?1. *Crop Sci.* 26, <https://doi.org/10.2135/cropsci1986.0011183X002600050012x>
- Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., Heckerman, D., 2011. FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8, 833–835. <https://doi.org/10.1038/nmeth.1681>
- Liu, P., Zhao, Y., Liu, G., Wang, M., Hu, D., Hu, J., Meng, J., Reif, J.C., Zou, J., 2017. Hybrid Performance of an Immortalized F2 Rapeseed Population Is Driven by Additive, Dominance, and Epistatic Effects . *Front. Plant Sci.* 8:815. doi: 10.3389/fpls.2017.00815.
- Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A.P., Zhao, M., Ma, J., Yu, J., Huang, S., Wang, X., Wang, J., Lu, K., Fang, Z., Bancroft, I., et al., 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5, 3930. <https://doi.org/10.1038/ncomms4930>



- Luo, X., Ma, C., Yue, Y., Hu, K., Li, Y., Duan, Z., Wu, M., Tu, J., Shen, J., Yi, B., Fu, T., 2015. Unravelling the complex trait of harvest index in rapeseed (*Brassica napus* L.) with association mapping. *BMC Genomics* 16, 379. <https://doi.org/10.1186/s12864-015-1607-0>
- Luo, Z., Wang, M., Long, Y., Huang, Y., Shi, L., Zhang, C., Liu, X., Fitt, B.D.L., Xiang, J., Mason, A.S., Snowdon, R.J., Liu, P., Meng, J., Zou, J., 2017. Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. *Theor. Appl. Genet.* 130, 1569–1585. <https://doi.org/10.1007/s00122-017-2911-7>
- Ly, D., 2016. Prédiction génomique des interactions Génotype x Environnement à l'aide d'indicateurs agro-climatiques chez le blé tendre (*Triticum aestivum* L.). Université Blaise Pascal Clermont-Ferrand II.
- Ly, D., Huet, S., Gau, A., Rincet, R., Touzy, G., Mini, A., Jannink, J., Cormier, F., Paux, E., Lafarge, S., Le, J., Charmet, G., 2018. Whole-genome prediction of reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.) by genomic random regression. *F. Crop. Res.* 216, 32–41. <https://doi.org/10.1016/j.fcr.2017.08.020>
- Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., Schubert, I., 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5224–5229. <https://doi.org/10.1073/pnas.0510791103>
- MacDonald, A.J., Poulton, P.R., Powlson, D.S., Jenkinson, D.S., 1997. Effects of season, soil type and cropping on recoveries, residues and losses of 15N-labelled fertilizer applied to arable crops in spring. *J. Agric. Sci.* 129, 125–154. <https://doi.org/10.1017/S0021859697004619>
- MacKay, T.F.C., Stone, E.A., Ayroles, J.F., 2009. The genetics of quantitative traits: Challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. <https://doi.org/10.1038/nrg2612>
- Malabat, C., Atterby, H., Chaudhary, Q., Renard, M., Gueguen, J., 2003. Genetic variability of rapeseed protein composition, in: 11th International Rapeseed Congress. pp. 205–208.
- Malagoli, P., Laine, P., Rossato, L., Ourry, A., 2005. Dynamics of nitrogen uptake and mobilization in field-grown winter oilseed rape (*Brassica napus*) from stem extension to harvest: I. Global N flows between vegetative and reproductive tissues in relation to leaf fall and their residual N. *Ann. Bot.* 95, 853–861. <https://doi.org/10.1093/aob/mci091>
- Malosetti, M., Ribaut, J.M., van Eeuwijk, F.A., 2013. The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. *Front. Physiol.* 4 MAR, 1–17. <https://doi.org/10.3389/fphys.2013.00044>
- Malosetti, M., Voltas, J., Romagosa, I., Ullrich, S.E., Eeuwijk, F.A.V., 2004. Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137, 139–145. <https://doi.org/10.1023/B:EUPH.0000040511.46388.ef>
- Mandakova, T., Lysak, M.A., 2008. Chromosomal Phylogeny and Karyotype Evolution in x = 7 Crucifer Species (*Brassicaceae*). *Plant Cell* 20, 2559–2570. <https://doi.org/10.1105/tpc.108.062166>
- Manzanares-Dauleux, M.J., Delourme, R., Baron, F., Thomas, G., 2000. Mapping of one major gene and of QTLs involved in resistance to clubroot in *Brassica napus*. *Theor. Appl. Genet.* 101, 885–891. <https://doi.org/10.1007/s001220051557>
- Marjanović-Jeromela, A., Marinković, R., Mijić, A., Jankulovska, M., Zdunić, Z., Nagl, N., 2008. Oil Yield Stability of Winter Rapeseed (*Brassica napus* L.) Genotypes. *Agric. Conspec. Sci.* Vol 73, No 4.
- Mason, A.S., Chèvre, A.-M., 2016. Optimization of Recombination in Interspecific Hybrids to Introduce New Genetic Diversity into Oilseed Rape (*Brassica napus* L.), in: Mason, A.S. (Ed.), *Polyploidy and Hybridization for Crop Improvement*. CRC Press, pp. 431–444.
- Mason, A.S., Higgins, E.E., Snowdon, R.J., Batley, J., Stein, A., Werner, C., Parkin, I.A.P., 2017. A user guide to the *Brassica* 60K Illumina Infinium™ SNP genotyping array. *Theor. Appl. Genet.* 130, 621–633. <https://doi.org/10.1007/s00122-016-2849-1>



- McCown, R.L., Hammer, G.L., Hargreaves, J.N.G., Holzworth, D.P., Freebairn, D.M., 1996. APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. *Agric. Syst.* 50, 255–271. [https://doi.org/10.1016/0308-521X\(94\)00055-V](https://doi.org/10.1016/0308-521X(94)00055-V)
- McDonald, B.A., Linde, C., 2002. Pathogen Population Genetics, Evolutionary Potential, and Durable Resistance. *Annu. Rev. Phytopathol.* 40, 349–379. <https://doi.org/10.1146/annurev.phyto.40.120501.101443>
- Météo France, 2016. Données quotidiennes du modèle de simulation des schémas de surface [WWW Document]. URL [https://donneespubliques.meteofrance.fr/?fond=produit&id\\_produit=230&id\\_rubrique=40](https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=230&id_rubrique=40) (accessed 1.1.16).
- Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Mücher, C.A., Watkins, J.W., 2005. A climatic stratification of the environment of Europe. *Glob. Ecol. Biogeogr.* 14, 549–563. <https://doi.org/10.1111/j.1466-822X.2005.00190.x>
- Miersch, S., Gertz, A., Breuer, F., Schierholt, A., Becker, H.C., 2016. Influence of the Semi-dwarf Growth Type on Nitrogen Use Efficiency in Winter Oilseed Rape. *Crop Sci.* 56, 2952–2961. <https://doi.org/10.2135/cropsci2016.01.0044>
- Millet, E.J., Welcker, C., Kruijer, W., Negro, S., Coupel-Ledru, A., Nicolas, S.D., Laborde, J., Bauland, C., Praud, S., Ranc, N., Presterl, T., Tuberosa, R., Bedo, Z., Draye, X., Usadel, B., Charcosset, A., Van Eeuwijk, F., Tardieu, F., 2016. Genome-Wide Analysis of Yield in Europe: Allelic Effects Vary with Drought and Heat Scenarios. *Plant Physiol.* 172, 749 LP – 764. <https://doi.org/10.1104/pp.16.00621>
- Miro, B., 2010. Identification of traits for nitrogen use efficiency in oilseed rape (*Brassica napus* L.). Newcastle University.
- Mittler, R., 2006. Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 11, 15–19. <https://doi.org/10.1016/j.tplants.2005.11.002>
- Moghaddam, M.J., Pourdad, S.S., 2011. Genotype × environment interactions and simultaneous selection for high oil yield and stability in rainfed warm areas rapeseed (*Brassica napus* L.) from Iran. *Euphytica* 180, 321–335. <https://doi.org/10.1007/s10681-011-0371-8>
- Moll, R.H., Kamprath, E.J., Jackson, W. a., 1982. Analysis and Interpretation of Factors Which Contribute to Efficiency of Nitrogen Utilization. *Agron. J.* 74, 562. <https://doi.org/10.2134/agronj1982.00021962007400030037x>
- Monteverde, E., Gutierrez, L., Blanco, P., Pérez de Vida, F., Rosas, J.E., Bonnacarrère, V., Quero, G., McCouch, S., 2019. Integrating Molecular Markers and Environmental Covariates To Interpret Genotype by Environment Interaction in Rice (*Oryza sativa* L.) Grown in Subtropical Areas. *G3 Genes|Genomes|Genetics* 9, 1519 LP – 1531. <https://doi.org/10.1534/g3.119.400064>
- Morrison, M.J., 1993. Heat stress during reproduction in summer rape. *Can. J. Bot.* 71, 303–308.
- Müller, J., Diepenbrock, W., 2006. Measurement and modelling of gas exchange of leaves and pods of oilseed rape. *Agric. For. Meteorol.* 139, 307–322. <https://doi.org/https://doi.org/10.1016/j.agrformet.2006.07.010>
- Murat, F., Louis, A., Maumus, F., Armero, A., Cooke, R., Quesneville, H., Crollius, 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>
- Nesi, N., Delourme, R., Brégeon, M., Falentin, C., Renard, M., 2008. Genetic and molecular approaches to improve nutritional value of *Brassica napus* L. seed. *Comptes Rendus - Biol.* 331, 763–771. <https://doi.org/10.1016/j.crv.2008.07.018>
- Nowosad, K., Liersch, A., Popławska, W., Bocianowski, J., 2016. Genotype by environment interaction for seed yield in rapeseed (*Brassica napus* L.) using additive main effects and multiplicative interaction model. *Euphytica* 208, 187–194. <https://doi.org/10.1007/s10681-015-1620-z>



- Nuttal, W.F., Moulin, A.P., Townley-Smith, L.J., 1992. Yield Response of Canola to Nitrogen, Phosphorus, Precipitation, and Temperature. *Agron. J.* 84, 765–768.
- Obermeier, C., Friedt, W., 2015. Applied oilseed rape marker technology and genomics, in: Poltronieri, P., Hong, Y. (Eds.), *Applied Plant Genomics and Biotechnology*. Elsevier, Heidelberg, pp. 253–295.
- Olatinwo, R., Hoogenboom, G., 2014. Chapter 4 - Weather-based Pest Forecasting for Efficient Crop Protection, in: Abrol, D.P.B.T.-I.P.M. (Ed.), . Academic Press, San Diego, pp. 59–78. <https://doi.org/https://doi.org/10.1016/B978-0-12-398529-3.00005-1>
- ONU, 2019. Deux milliards de personnes de plus sur la Terre en 2050, selon l'ONU [WWW Document]. URL <https://www.un.org/development/desa/fr/news/population/world-population-prospects-2019.html>
- Palermo, G., Piraino, P., Zucht, H.D., 2009. Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. *Adv. Appl. Bioinforma. Chem.* 2, 57–70. <https://doi.org/10.2147/AABC.S3619>
- Pariset-Baril, C., 1992. Etude de la stabilité du rendement chez le ble tendre d'hiver (*triticum aestivum* L. THELL.). Université Paris-Sud.
- Parkin, I.A.P., Gulden, S.M., Sharpe, A.G., Lukens, L., Trick, M., Osborn, T.C., Lydiate, D.J., 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, I.A.P., Koh, C., Tang, H., Robinson, S.J., Kagale, S., Clarke, W.E., Town, C.D., Nixon, J., Krishnakumar, V., Bidwell, S.L., Denoeud, F., Belcram, H., Links, M.G., et al., 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol.* 15, 1–18. <https://doi.org/10.1186/gb-2014-15-6-r77>
- Parnaudeau, V., Jeuffroy, M., Machet, J., Reau, R., Bissuel, C., 2009. Methods for determining the nitrogen fertiliser requirements of some major arable crops in France, in: *International Fertiliser Society*. Cambridge, pp. 1–26.
- Pelé, A., Falque, M., Trotoux, G., Eber, F., Nègre, S., Gilet, M., Huteau, V., Lodé, M., Jousseau, T., Dechaumet, S., Morice, J., Poncet, C., Coriton, O., Martin, O.C., Rousseau-Gueutin, M., Chèvre, A.M., 2017. Amplifying recombination genome-wide and reshaping crossover landscapes in Brassicas, *PLoS Genetics*. <https://doi.org/10.1371/journal.pgen.1006794>
- Picoult-Newberg, L., Ideker, T.E., Pohl, M.G., Taylor, S.L., Donaldson, M.A., Nickerson, D.A., Boyce-jacino, M., 1999. Mining SNPs From EST Databases. *Genome Res.* 9, 167–174.
- Pinochet, X., Renard, M., 2012. Progrès génétique en colza et perspectives. *OCL - Ol. Corps Gras Lipides* 19, 147–154. <https://doi.org/10.1684/ocl.2012.0456>
- Potier, D., 2014. Pesticides et agro-écologie, les champs du possible.
- Pullens, J.W.M., Sharif, B., Trnka, M., Balek, J., Semenov, M.A., Olesen, J.E., 2019. Risk factors for European winter oilseed rape production under climate change. *Agric. For. Meteorol.* 272–273, 30–39. <https://doi.org/https://doi.org/10.1016/j.agrformet.2019.03.023>
- Qi, L., Mao, L., Sun, C., Pu, Y., Fu, T., Ma, C., Shen, J., Tu, J., Yi, B., Wen, J., 2014. Interpreting the genetic basis of siliqua traits in *Brassica napus* using a joint QTL network. *Plant Breed.* 133, 52–60. <https://doi.org/10.1111/pbr.12131>
- Qian, L., Qian, W., Snowdon, R.J., 2014. Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. *BMC Genomics* 15, 1–17. <https://doi.org/10.1186/1471-2164-15-1170>
- Qian, W., Meng, J., Li, M., Frauen, M., Sass, O., Noack, J., Jung, C., 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>



- Quijada, P.A., Udall, J.A., Lambert, B., Osborn, T.C., 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>
- R Development Core Team, 2017. R: A language and environment for statistical computing.
- Rabonatahiry, N., Chao, H., Dalin, H., Pu, S., Yan, W., Yu, L., Wang, B., Li, M., 2018. QTL alignment for seed yield and yield related traits in *Brassica napus*. *Front. Plant Sci.* 9, 1–14. <https://doi.org/10.3389/fpls.2018.01127>
- Radoev, M., Becker, H.C., Ecke, W., 2008. Genetic Analysis of Heterosis for Yield and Yield Components in Rapeseed (*Brassica napus*; L.) by Quantitative Trait Locus Mapping. *Genetics* 179, 1547 LP – 1558. <https://doi.org/10.1534/genetics.108.089680>
- Rahman, H., Bennett, R.A., Kebede, B., 2017. Mapping of days to flower and seed yield in spring oilseed *Brassica napus* carrying genome content introgressed from *Brassica oleracea*. *Mol. Breed.* 37, 5. <https://doi.org/10.1007/s11032-016-0608-2>
- Rahman, H., Peng, G., Yu, F., Falk, K.C., Kulkarni, M., Selvaraj, G., 2014. Genetics and breeding for clubroot resistance in Canadian spring canola (*Brassica napus* L.). *Can. J. Plant Pathol.* 36, 122–134. <https://doi.org/10.1080/07060661.2013.862571>
- Raman, R., Diffey, S., Carling, J., Cowley, R.B., Kilian, A., Luckett, D.J., Raman, H., 2016. Quantitative genetic analysis of grain yield in an Australian *Brassica napus* doubled-haploid population. *Crop Pasture Sci.* 67, 298–307.
- Rathke, G.W., Behrens, T., Diepenbrock, W., 2006. Integrated nitrogen management strategies to improve seed yield, oil content and nitrogen efficiency of winter oilseed rape (*Brassica napus* L.): A review. *Agric. Ecosyst. Environ.* 117, 80–108. <https://doi.org/10.1016/j.agee.2006.04.006>
- Rémy, J., Hébert, J., 1977. Le devenir des engrais azotés dans le sol. *Acad. l'Agriculture Fr.* 63.
- Retailleau, L., 2018. Potentiel de la sélection génomique pour prédire le rendement du colza d'hiver dans un réseau multi local. Rennes.
- Rincent, R., Ababaei, M.M.B., Touzy, G., Bogard, A.M.M., Le, P.M.J., 2019. Using crop growth model stress covariates and AMMI decomposition to better predict genotype - by - environment interactions. *Theor. Appl. Genet.* 132, 3399–3411. <https://doi.org/10.1007/s00122-019-03432-y>
- Rincent, R., Moreau, L., Monod, H., Kuhn, E., Melchinger, A.E., Malvar, R.A., Moreno-Gonzalez, J., Nicolas, S., Madur, D., Combes, V., Dumas, F., Altmann, T., Brunel, D., Ouzunova, M., Flament, P., Dubreuil, P., Charcosset, A., Mary-Huard, T., 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>
- Robertson, M.J., Lilley, J.M., 2016. Simulation of growth, development and yield of canola (*Brassica napus*) in APSIM. *Crop Pasture Sci.* 67, 332–344.
- Rossato, L., Lainé, P., Ourry, A., 2001. Nitrogen storage and remobilization in *Brassica napus* L. during the growth cycle: nitrogen fluxes within the plant and changes in soluble protein patterns. *J. Exp. Bot.* 52, 1655–1663. <https://doi.org/10.1093/jxb/52.361.1655>
- Rousseau-Gueutin, M., Belser, C., Silva, C. Da, Richard, G., Istace, B., Cruaud, C., Falentin, C., Boideau, F., Boutte, J., Delourme, R., Deniot, G., Engelen, S., de Carvalho, J.F., Lemainque, A., Maillet, L., Morice, J., Wincker, P., Denoeud, F., Chèvre, A.-M., Aury, J.-M., 2020. Long-reads assembly of the *Brassica napus* reference genome, Darmor-bzh. *bioRxiv*. <https://doi.org/10.1101/2020.07.22.215749>
- Rouxel, T., Penaud, A., Pinochet, X., Brun, H., Gout, L., Delourme, R., Schmit, J., Balesdent, M.-H., 2003. A 10-year Survey of Populations of *Leptosphaeria maculans* in France Indicates a Rapid Adaptation Towards the Rlm1 Resistance Gene of Oilseed Rape. *Eur. J. Plant Pathol.* 109, 871–881. <https://doi.org/10.1023/A:1026189225466>
- Rusch, A., Valantin-Morison, M., Roger-Estrade, J., Sarthou, J.P., 2012. Using landscape indicators to predict



- high pest infestations and successful natural pest control at the regional scale. *Landsc. Urban Plan.* 105, 62–73. <https://doi.org/https://doi.org/10.1016/j.landurbplan.2011.11.021>
- Samans, B., Chalhouh, B., Snowdon, R.J., 2017. Surviving a Genome Collision: Genomic Signatures of Allopolyploidization in the Recent Crop Species *Brassica napus*. *Plant Genome* 10, plantgenome2017.02.0013. <https://doi.org/10.3835/plantgenome2017.02.0013>
- Sanchez, G., 2012. *Partial Least Squares (PLS) Data Analysis Methods*.
- Sashidhar, N., Harloff, H.J., Potgieter, L., Jung, C., 2020. Gene editing of three BnITPK genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnol. J.* 1, 1–10. <https://doi.org/10.1111/pbi.13380>
- Schatzki, J., Ecke, W., Becker, H.C., Möllers, C., 2014. Mapping of QTL for the seed storage proteins cruciferin and napin in a winter oilseed rape doubled haploid population and their inheritance in relation to other seed traits. *Theor. Appl. Genet.* 127, 1213–1222. <https://doi.org/10.1007/s00122-014-2292-0>
- Schiessl, S., Samans, B., Hüttel, B., Reinhard, R., Snowdon, R.J., 2014. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* 5, 1–14. <https://doi.org/10.3389/fpls.2014.00404>
- Schranz, M.E., Lysak, M.A., Mitchell-olds, T., 2006. The ABC 's of comparative genomics in the Brassicaceae : building blocks of crucifer genomes. *Trends Plant Sci.* 11. <https://doi.org/10.1016/j.tplants.2006.09.002>
- Schürmann, G., Ebert, R.U., Chen, J., Wang, B., Kühne, R., 2008. External validation and prediction employing the predictive squared correlation coefficient - Test set activity mean vs training set activity mean. *J. Chem. Inf. Model.* 48, 2140–2145. <https://doi.org/10.1021/ci800253u>
- Shafii, B., Mahler, K.A., Price, W.J., Auld, D.L., 1992. Genotype x Environment Interaction Effects on Winter Rapeseed Yield and Oil Content. *Crop Sci.* 32, 922. <https://doi.org/10.2135/cropsci1992.0011183X003200040017x>
- Sheldon, C.C., Finnegan, E.J., Rouse, D.T., Tadege, M., Bagnall, D.J., Helliwell, C.A., Peacock, W.J., Dennis, E.S., 2000. Control of Flowering By Vernalization. *Curr. Opin. Plant Biol.* 3, 418–422.
- Shi, J., Li, R., Qiu, D., Jiang, C., Long, Y., Morgan, C., Bancroft, I., Zhao, J., Meng, J., 2009. Unraveling the Complex Trait of Crop Yield With Quantitative Trait Loci Mapping in *Brassica napus* *Genetics* 182, 851 LP – 861. <https://doi.org/10.1534/genetics.109.101642>
- Shi, T., Li, R., Zhao, Z., Ding, G., Long, Y., Meng, J., Xu, F., Shi, L., 2013. QTL for Yield Traits and Their Association with Functional Genes in Response to Phosphorus Deficiency in *Brassica napus*. *PLoS One* 8, e54559.
- Sieling, K., Christen, O., 1997. Effect of preceding crop combination and N fertilization on yield of six oil-seed rape cultivars (*Brassica napus* L.). *Eur. J. Agron.* 7, 301–306. [https://doi.org/https://doi.org/10.1016/S1161-0301\(97\)00009-9](https://doi.org/https://doi.org/10.1016/S1161-0301(97)00009-9)
- Sieling, K., Günther-Borstel, O., Hanus, H., 1997. Effect of slurry application and mineral nitrogen fertilization on N leaching in different crop combinations. *J. Agric. Sci.* 128, 79–86. <https://doi.org/DOI:10.1017/S0021859696004005>
- Snowdon, R., Lühs, W., Friedt, W., 2007. Oilseed Rape, in: Kole, C. (Ed.), *Oilseeds*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 55–114. [https://doi.org/10.1007/978-3-540-34388-2\\_2](https://doi.org/10.1007/978-3-540-34388-2_2)
- Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., Xie, W.Z., Cheng, Y., Zhang, Y., Liu, K., Yang, Q.Y., Chen, L.L., Guo, L., 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>
- Tadege, M., Sheldon, C.C., Helliwell, C.A., Stoutjesdijk, P., Dennis, E.S., Peacock, W.J., 2001. Control of flowering time by FLC orthologues in *Brassica napus*. *Plant J.* 28, 545–553. <https://doi.org/10.1046/j.1365-313X.2001.01182.x>
- Tang, H., Woodhouse, M.R., Cheng, F., Schnable, J.C., Pedersen, B.S., Conant, G., Wang, X., Freeling, M., Pires, J.C.M., 2015. The genome of *Brassica napus* is a product of recent hybridization between *B. rapa* and *B. oleracea*. *Nat. Genet.* 47, 123–129. <https://doi.org/10.1038/ng.3177>



- J.C., 2012. Altered Patterns of Fractionation and Exon Deletions in Brassica rapa Support a Two-Step Model of Paleohexaploidy. *Genetics* 190, 1563 LP – 1574. <https://doi.org/10.1534/genetics.111.137349>
- Tenenhaus, M., 1998. La régression PLS. *Théorie et Pratique*.
- Terres Inovia, 2020. Colza [WWW Document]. URL <https://www.terresinovia.fr/colza>
- Terres Univia, 2019. Chiffres clés oléagineux et plantes riches en protéines 2018.
- The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408, 796–815. <https://doi.org/10.1038/35048692>
- Touzy, G., Rincet, R., Bogard, M., Lafarge, S., Dubreuil, P., Mini, A., Deswarte, J.C., Beauchêne, K., Le Gouis, J., Praud, S., 2019. Using environmental clustering to identify specific drought tolerance QTLs in bread wheat (*T. aestivum* L.). *Theor. Appl. Genet.* 132, 2859–2880. <https://doi.org/10.1007/s00122-019-03393-2>
- U, N., 1935. Genome-analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization.
- Udall, J.A., Quijada, P.A., Lambert, B., Osborn, T.C., 2006. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 2. Identification of alleles from unadapted germplasm. *Theor. Appl. Genet.* 113, 597–609. <https://doi.org/10.1007/s00122-006-0324-0>
- Vaezi, B., Pour-Aboughadareh, A., Mohammadi, R., Mehraban, A., Hossein-Pour, T., Koohkan, E., Ghasemi, S., Moradkhani, H., Siddique, K.H.M., 2019. Integrating different stability models to investigate genotype × environment interactions and identify stable and high-yielding barley genotypes. *Euphytica* 215, 63. <https://doi.org/10.1007/s10681-019-2386-5>
- Van Eeuwijk, F.A., Bustos-Korts, D. V., Malosetti, M., 2016. What should students in plant breeding know about the statistical aspects of genotype × Environment interactions? *Crop Sci.* 56, 2119–2140. <https://doi.org/10.2135/cropsci2015.06.0375>
- van Eeuwijk, F.A., Denis, J.B., Kang, M.S., 1996. Incorporating additional information on genotypes and environments in models for two-way tables, in: Kang, M.S., Gauch, H.G. (Eds.), *Genotype-by-Environment Interaction*. Boca Raton, pp. 19–49.
- Van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance-A review. *F. Crop. Res.* 143, 4–17. <https://doi.org/10.1016/j.fcr.2012.09.009>
- Vargas, M., Crossa, J., Van Eeuwijk, F., Sayre, K.D., Reynolds, M.P., 2001. Interpreting Treatment × Environment Interaction in Agronomy Trials. *Agron. J.* 93, 949–960. <https://doi.org/10.2134/agronj2001.934949x>
- Verbeke, G., Molenberghs, G., 1997. *Linear Mixed Models in Practice*. Springer. [https://doi.org/10.1007/978-1-4612-2294-1\\_3](https://doi.org/10.1007/978-1-4612-2294-1_3)
- von Wricke, G., 1962. Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. *Z. Planzengucht* 47, 92–96.
- Wallace, D.H., Zobel, R.W., Yourstone, K.S., 1993. A whole-system reconsideration of paradigms about photoperiod and temperature control of crop yield. *Theor. Appl. Genet.* 86, 17–26. <https://doi.org/10.1007/BF00223804>
- Wang, H., Yan, M., Xiong, M., Wang, P., Liu, Y., Xin, Q., Wan, L., Yang, G., Hong, D., 2020. Genetic dissection of thousand-seed weight and fine mapping of *cqSW.A03-2* via linkage and association analysis in rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 133, 1321–1335. <https://doi.org/10.1007/s00122-020-03553-9>
- Wang, J., Dun, X., Shi, J., Wang, X., Liu, G., Wang, H., 2017. Genetic Dissection of Root Morphological Traits Related to Nitrogen Use Efficiency in *Brassica napus* L. under Two Contrasting Nitrogen Conditions. *Front. Plant Sci.*
- Wang, N., Wang, Y., Tian, F., King, G.J., Zhang, C., Long, Y., Shi, L., Meng, J., 2008. A functional genomics resource for *Brassica napus*: Development of an EMS mutagenized population and discovery of FAE1



- point mutations by TILLING. *New Phytol.* 180, 751–765. <https://doi.org/10.1111/j.1469-8137.2008.02619.x>
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.-H., Bancroft, I., Cheng, F., Huang, Sanwen, Li, X., Hua, W., et al., Consortium, T.B. rapa G.S.P., 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. <https://doi.org/10.1038/ng.919>
- Wang, Y., Chen, W., Chu, P., Wan, S., Yang, M., Wang, M., Guan, R., 2016. Mapping a major QTL responsible for dwarf architecture in *Brassica napus* using a single-nucleotide polymorphism marker approach. *BMC Plant Biol.* 16, 178. <https://doi.org/10.1186/s12870-016-0865-6>
- Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* <https://doi.org/10.1080/01621459.1963.10500845>
- Weber, E., Bleiholder, H., 1990. Explanations of the BBCH decimal codes for the growth stages of maize, rape, faba beans, sunflowers and peas - with illustrations. *Gesunde Pflanz.* 42, 308–321.
- 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)
- Weymann, W., Böttcher, U., Sieling, K., Kage, H., 2015. Effects of weather conditions during different growth phases on yield formation of winter oilseed rape. *F. Crop. Res.* 173, 41–48. <https://doi.org/10.1016/j.fcr.2015.01.002>
- Wold, S., Albano, C., Dunn III, W.J., Esbensen, K., Hellberg, S., Johansson, E., Sjöström, H., 1983. Pattern Recognition: Finding and Using Regularities in Multivariate Data, in: Martens, J. (Ed.), *IUFOST Conf. "Food Research and Data Analysis."* Applied Science Publications, London, pp. 147–188.
- Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., Wang, H., Chen, X., Wang, Q., Zhu, L., Maodzeka, A., et al., 2019. Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence. *Mol. Plant* 12, 30–43. <https://doi.org/https://doi.org/10.1016/j.molp.2018.11.007>
- Yan, W., 2016. Analysis and Handling of G × E in a Practical Breeding Program. *Crop Sci.* 56, 2106–2118. <https://doi.org/10.2135/cropsci2015.06.0336>
- Yan, W., Hunt, L. a., Sheng, Q., Szlavnic, Z., 2000. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *Crop Sci.* 40, 597. <https://doi.org/10.2135/cropsci2000.403597x>
- Yan, W., Kang, M.S., 2002. *GGE Biplot Analysis*. CRC Press, Boca Raton. <https://doi.org/https://doi.org/10.1201/9781420040371>
- Yang, P., Shu, C., Chen, L., Xu, J., Wu, J., Liu, K., 2012. Identification of a major QTL for silique length and seed weight in oilseed rape (*Brassic napus* L.). *Theor. Appl. Genet.* 125, 285–296. <https://doi.org/10.1007/s00122-012-1833-7>
- Yang, Y., Shen, Y., Li, S., Ge, X., Li, Z., 2017. High Density Linkage Map Construction and QTL Detection for Three Silique-Related Traits in *Orychophragmus violaceus* Derived *Brassica napus* Population. *Front. Plant Sci.* 8:1512. doi: 10.3389/fpls.2017.01512.
- Yang, Y., Shi, J., Wang, X., Liu, G., Wang, H., 2016. Genetic architecture and mechanism of seed number per pod in rapeseed: elucidated through linkage and near-isogenic line analysis. *Sci. Rep.* 6, 24124. <https://doi.org/10.1038/srep24124>
- Young, L.W., Wilen, R.W., Bonham-smith, P.C., 2004. High temperature stress of *Brassica napus* during flowering reduces micro- and megagametophyte fertility, induces fruit abortion, and disrupts seed production. *J. Exp. Bot.* 55, 485–495. <https://doi.org/10.1093/jxb/erh038>
- Zhang, H., Berger, J.D., Milroy, S.P., 2013. Genotype×environment interaction studies highlight the role of phenology in specific adaptation of canola (*Brassica napus*) to contrasting Mediterranean climates. *F. Crop. Res.* 144, 77–88. <https://doi.org/10.1016/j.fcr.2013.01.006>



- Zhang, H., Feng, J., Hwang, S.-F., Strelkov, S.E., Falak, I., Huang, X., Sun, R., 2016. Mapping of clubroot (*Plasmodiophora brassicae*) resistance in canola (*Brassica napus*). *Plant Pathol.* 65, 435–440. <https://doi.org/10.1111/ppa.12422>
- Zhang, L., Yang, G., Liu, P., Hong, D., Li, S., He, Q., 2011. Genetic and correlation analysis of silique-traits in *Brassica napus* L. by quantitative trait locus mapping. *Theor. Appl. Genet.* 122, 21–31. <https://doi.org/10.1007/s00122-010-1419-1>
- Zhao, W., Wang, X., Wang, H., Tian, J., Li, B., Chen, L., Chao, H., Long, Y., Xiang, J., Gan, J., Liang, W., Li, M., 2016. Genome-Wide Identification of QTL for Seed Yield and Yield-Related Traits and Construction of a High-Density Consensus Map for QTL Comparison in *Brassica napus*. *Front. Plant Sci.* 7:17. doi:10.3389/fpls.2016.00017.
- Zhao, W., Zhang, Lina, Chao, H., Wang, H., Ta, N., Li, H., Wang, X., Li, S., Xiang, J., Zhang, Libin, Li, M., 2019. Genome-wide identification of silique-related traits based on high-density genetic linkage map in *Brassica napus*. *Mol. Breed.* 39. <https://doi.org/10.1007/s11032-019-0988-1>
- Zhao, X., Li, B., Zhang, K., Hu, K., Yi, B., Wen, J., Ma, C., Shen, J., Fu, T., Tu, J., 2016. Breeding signature of combining ability improvement revealed by a genomic variation map from recurrent selection population in *Brassica napus*. *Sci. Rep.* 6, 29553. <https://doi.org/10.1038/srep29553>
- Zhao, Z., Wu, L., Nian, F., Ding, G., Shi, T., Zhang, D., Shi, L., Xu, F., Meng, J., 2012. Dissecting Quantitative Trait Loci for Boron Efficiency across Multiple Environments in *Brassica napus*. *PLoS One* 7, e45215.
- Zhou, W., Lin, X., 1995. Effects of waterlogging at different growth stages on physiological characteristics and seed yield of winter rape (*Brassica napus* L.). *F. Crop. Res.* 44, 103–110. [https://doi.org/10.1016/0378-4290\(95\)00075-5](https://doi.org/10.1016/0378-4290(95)00075-5)



ANNEXE A

---

SCRIPT PLS

---



# Script PLS

## Etape 1 : Faire tourner les différentes régressions PLS pour éliminer les variables une à une

```
#### SCRIPT REGRESION PLS UNIVARIEE ASSOCIEE A UNE SELECTION DE
VARIABLES ####

#### LIBRARY
library(plsdepot)

#### DATA

PLS_DATA <- read.table("DATA.txt", sep = "\t", dec = ".", header =
TRUE, na.strings = "NA", row.names = 4)

PLS_DATA_79 <- PLS_DATA_45E[-c(1:3)] # Retrait des métadonnées, ne
garder que les indicateurs et le rendement, 79 fait référence au
nombre d'indicateurs considérés

#### REGRESSION PLS
#### LE PROCESSUS EST SEMI AUTOMATIQUE
#### EXEMPLE SUR 2 RUNS, FAIRE N-1 RUN où N est le nombre
d'indicateurs considérés

#####RUN_01#####
## 79 VAR

BIND_Q2_TABLE_79 <- c("PRESS", "RSS", "Q2", "LimQ2", "Q2cum") #
CREATION DE LA TABLE DE Q2

for (i in 1:10) {
  print(i)
  res.PLS_SY_79 <- plsreg1(predictors = PLS_DATA_79[,c(1:79)],
response = PLS_DATA_79[,80, drop=FALSE], comps = 5, crossval = TRUE)
#PLS
  Q2_table <- as.data.frame(res.PLS_SY_79$Q2)#Extraction du tableau
résultat de crossval
  #Identification du facteur à retirer
  temp <- as.data.frame(res.PLS_SY_79$std.coefs)
  temp$ABS <- abs(temp$`res.PLS_SY_79$std.coefs`)
  supr <- which.min(temp$ABS)
  print(colnames((PLS_DATA_79)[supr])) # DONNE LE NOM DE LA VARIABLE
A RETIRER
  #Compilation des tableaux de crossval
  BIND_Q2_TABLE_79 <- rbind(BIND_Q2_TABLE_79,Q2_table)
  #Préparation du jeu de données pour la prochaine étape
  PLS_DATA_78 <- PLS_DATA_79[,-supr]
  #Reset des variables temporaires
  rm(temp)
  rm(supr)
}
#Affichage du tableau de crossval
```



```
colnames(BIND_Q2_TABLE_79) <- c("PRESS", "RSS", "Q2", "LimQ2", "Q2cum")
BIND_Q2_TABLE_79 <- BIND_Q2_TABLE_79[-1,]
BIND_Q2_TABLE_79$PRESS <- as.factor(BIND_Q2_TABLE_79$PRESS)
BIND_Q2_TABLE_79$RSS <- as.factor(BIND_Q2_TABLE_79$RSS)
BIND_Q2_TABLE_79$Q2 <- as.factor(BIND_Q2_TABLE_79$Q2)
BIND_Q2_TABLE_79$LimQ2 <- as.factor(BIND_Q2_TABLE_79$LimQ2)
BIND_Q2_TABLE_79$Q2cum <- as.factor(BIND_Q2_TABLE_79$Q2cum)
BIND_Q2_TABLE_79 <- as.data.frame(BIND_Q2_TABLE_79)
write.table(BIND_Q2_TABLE_79, file="Q2_Table_79.txt")
```

```
#####RUN_02#####
## 78 VAR
```

```
BIND_Q2_TABLE_78 <- c("PRESS", "RSS", "Q2", "LimQ2", "Q2cum")
```

```
for (i in 1:10) {
  print(i)
  res.PLS_SY_78 <- plsreg1(predictors = PLS_DATA_78[,c(1:78)],
  response = PLS_DATA_78[,79, drop=FALSE], comps = 5, crosval = TRUE)
  Q2_table <- as.data.frame(res.PLS_SY_78$Q2)
  temp <- as.data.frame(res.PLS_SY_78$std.coefs)
  temp$ABS <- abs(temp$res.PLS_SY_78$std.coefs)
  supr <- which.min(temp$ABS)
  print(colnames((PLS_DATA_78)[supr]))
  BIND_Q2_TABLE_78 <- rbind(BIND_Q2_TABLE_78, Q2_table)
  PLS_DATA_77 <- PLS_DATA_78[,-supr]
  rm(temp)
  rm(supr)
}
```

```
colnames(BIND_Q2_TABLE_78) <- c("PRESS", "RSS", "Q2", "LimQ2", "Q2cum")
BIND_Q2_TABLE_78 <- BIND_Q2_TABLE_78[-1,]
BIND_Q2_TABLE_78$PRESS <- as.factor(BIND_Q2_TABLE_78$PRESS)
BIND_Q2_TABLE_78$RSS <- as.factor(BIND_Q2_TABLE_78$RSS)
BIND_Q2_TABLE_78$Q2 <- as.factor(BIND_Q2_TABLE_78$Q2)
BIND_Q2_TABLE_78$LimQ2 <- as.factor(BIND_Q2_TABLE_78$LimQ2)
BIND_Q2_TABLE_78$Q2cum <- as.factor(BIND_Q2_TABLE_78$Q2cum)
BIND_Q2_TABLE_78 <- as.data.frame(BIND_Q2_TABLE_78)
write.table(BIND_Q2_TABLE_78, file="Q2_Table_78.txt")
```

## Etape 2 : Identifier le modèle optimal

- Ouvrir les différents tableaux de Q2 pour chaque étape de régression PLS.
- Identifier le nombre de composantes valides, une composante est valide lorsque la valeur  $Q^2$  est supérieure à la valeur du  $Q^2$  limite. Attention si la composante 2 est valide, mais pas la 3 et que la 4 est valide, seules les 2 premières composantes sont conservées.
- Faire la moyenne de tous les  $Q^2_{cum}$  pour chaque run et identifier le plus fort  $Q^2_{cum}$  pour le plus petit nombre d'indicateurs pédoclimatiques



### Etape 3 : Effectuer le clustering

- Calculer la matrice  $c^*t$  en ayant extrait les données de la PLS pour le nombre de composantes valides
- Calculer la matrice de distance euclidienne, identifier le nombre de groupes le plus probable et clusteriser

```
#CLASSIFICATION SUR LES PLANS DE LA PLS
```

```
## IMPORTATION de la matrice CT
```

```
Matrice_CT <- read.table("MatriceCT.txt", sep = "\t", dec = ".",  
header = TRUE, row.names = 1, na.strings = "NA")
```

```
## Determination du nombre de groupes
```

```
library(NbClust)
```

```
nbclust <- NbClust(data = Matrice_CT_GWAS, diss = NULL, distance =  
"euclidean", method = "ward.D2", min.nc = 2, max.nc = 15, index =  
"all")
```

```
nbclust$All.index
```

```
nbclust$Best.partition
```

```
## CLASSIFICATION
```

```
#Calcul de la matrice de distance euclidienne
```

```
matrice_de_distance <- dist(Matrice_CT, method = "euclidean")
```

```
#Clustering
```

```
res <- hclust(matrice_de_distance, method = "ward.D2")
```

```
#Graphique des gains d'inertie
```

```
inertie <- sort(res$height, decreasing = TRUE)
```

```
plot(inertie[1:20], type = "s", xlab = "No of clusters", ylab =  
"Inertia gain")
```

```
#Dendrogramme
```

```
library(factoextra)
```

```
fviz_dend(res)
```

```
classif <- cutree(res, k=4) #k correspond au nombre de groupes
```

```
fviz_dend(res, k=4, repel = TRUE )
```



## ANNEXE B

---

SEEDPROT (side project)

---



# SEEDPROT

---

## **Teneur et qualité des protéines de la graine de colza : exploration de la diversité génétique en vue d'optimiser les usages en alimentation animale et humaine**

### **Partenariat**

- IGEPP, INRAE, Institut Agro, Université Rennes 1 – Le Rheu
- BIA, INRAE – Nantes
- IJPB) – INRAE, AgroParisTech – Versailles

### **Financement**

- PROMOSOL – 36 mois (01/01/2016 – 31/12/2018)

## **Objectifs du programme**

Le projet SEEDPROT avait pour ambitions i) d'évaluer la variabilité génétique pour la teneur et la composition en protéines de la graine et du tourteau en relation avec le contenu en huile de la graine (tâches 1 et 2), et ii) d'appréhender les propriétés fonctionnelles des protéines de réserve (tâche 3). L'objectif de la tâche 1 (IGEPP et IJPB) était d'identifier les déterminants génétiques et moléculaires qui contrôlent l'accumulation de l'huile et des protéines dans la graine et le tourteau, en prenant en compte la corrélation négative qui existe entre ces composants. Les travaux prévus dans cette tâche se sont appuyés sur les données acquises chez le colza et Arabidopsis et ont conduit à la description de l'architecture des facteurs génétiques impliqués dans l'accumulation des composés de la graine et du tourteau.

Dans la tâche 2 (IGEPP et BIA), il s'agissait d'explorer la variabilité pour la composition en protéines de réserve de type 12S (cruciférines) et 2S (napines) de la graine de colza. Pour cela, il était nécessaire i) de développer des méthodes d'extraction et d'analyse des protéines de réserves adaptées à un débit compatible avec des analyses génétiques tout en conservant au mieux les propriétés de ces protéines, ii) d'appréhender la variabilité génétique de la composition en protéines de réserve dans la graine chez le colza, et iii) d'étudier la variabilité génotypique de gènes candidats. Enfin, la tâche 3 (BIA) était dédiée à l'analyse des fonctionnalités des protéines de colza, en focalisant sur leurs propriétés émulsifiantes et leur digestibilité. Pour cela, on a cherché à estimer les capacités émulsifiantes et la teneur en acide phytique sur une dizaine d'extraits protéiques présentant des compositions en protéines (12S/2S) contrastées.

J'ai travaillé sur ce projet pendant 2 ans (1/10/2016-31/12/2018). J'ai réalisé la totalité des analyses génétiques en lien avec le projet. Cette expérience m'a permis d'intégrer la communauté Promosol. Les principaux résultats que j'ai acquis sont présentés ci-après avec :

- 1) une analyse des déterminants génétiques et moléculaires de la qualité de la graine chez le colza (tâche 1) ;
- 2) l'exploration de la diversité génétique et moléculaire pour la composition en protéines de réserve de la graine de colza et pour le rapport 12S/2S (tâche 2). Ces résultats sont présentés sous la forme d'une affiche qui a reçu le prix du meilleur poster à l'International Conference on Rapeseed (Berlin, Juin 2019).



## 1. Déterminants génétiques et moléculaires de la qualité de la graine chez le colza

Le projet a exploré le déterminisme génétique et moléculaire des variables liées à la qualité de la graine chez le colza. Les variables à l'étude sont la teneur en huile (O), en protéines de la graine (SP) ou du tourteau (MPT), en glucosinolates (GSL), en résidus (RES), le rapport O/SP, la déviation par rapport à la droite de régression O×SP (DSC), et la composition en protéines de réserve (2S, 12S, 12S/2S). Des analyses génétiques par association (GWAS) ont été menées une population de 174 individus maximisant la diversité génétique du colza oléagineux d'hiver (panel RAPSODYN).

- Acquisition des variables

O : teneur en huile dans la graine, estimée par NIRS – données issues du réseau RAPSODYN

SP : teneur en protéines dans la graine, estimée par NIRS – données issues du réseau RAPSODYN

OSP : rapport O sur SP, calculé

MPT : teneur en protéines dans le tourteau, estimée par NIRS – données issues du réseau RAPSODYN

DSC : delta seed composition (déviations par rapport à la droite de régression O/SP), calculé

RES : teneur en résidus dans la graine, calculé :  $100 - (O + SP + WC)$  avec  $WC = 5\%$

GSL : teneur en glucosinolates dans la graine, estimée par NIRS – données issues du réseau RAPSODYN

NAP\_2S : teneur en napines, dosée par chromatographie (L2.1) – données issues du projet SEEDPROT

CRU\_12S : teneur en cruciférines, dosée par chromatographie (L2.1) – données issues du projet SEEDPROT

12S/2S : rapport cruciférines/napines, calculé

- Analyses GWAS pour la détection de QTL

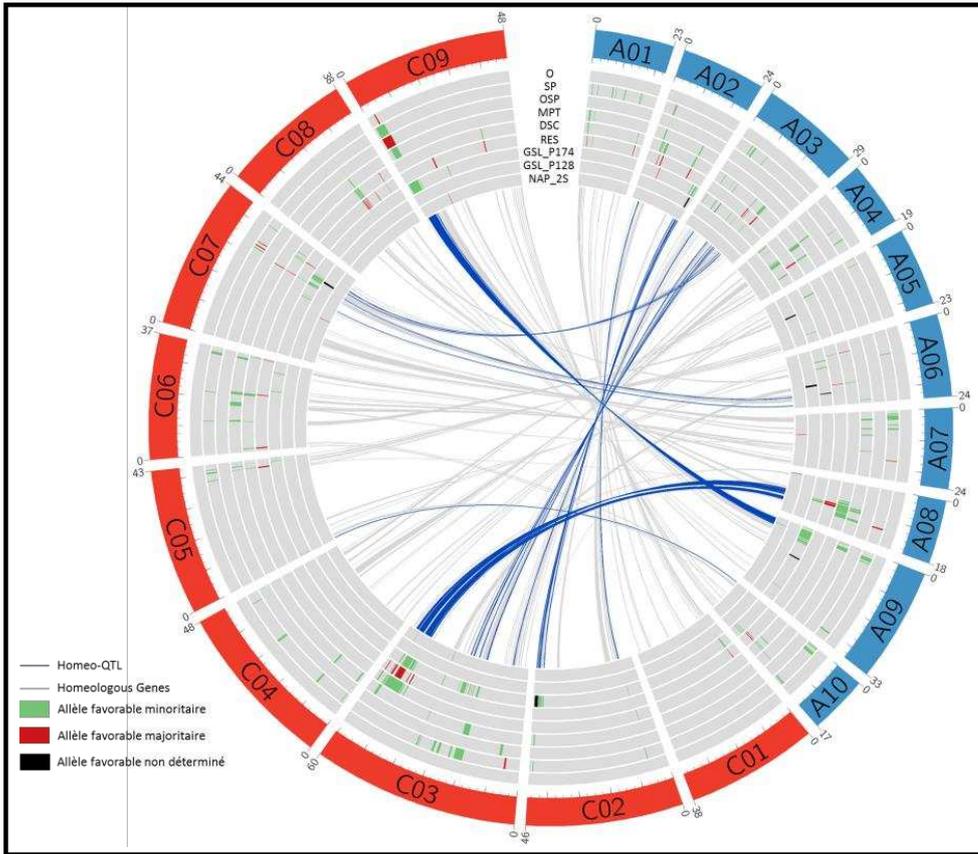
Pour toutes les analyses génétiques, les données de génotypage sont issues du projet RAPSODYN : matrice de 237k SNP (exome capture et puce 60K Brassica).

Variables O, SP, OSP, MP, DSC, RES et GSL : Les données proviennent du projet RAPSODYN. Elles ont été acquises sur 174 génotypes (30 ++, 15 0+, 1+0 et 128 00) mis en essais au champ sur 9 environnements (combinaisons lieu × année) et 2 conditions de nutrition azotée (non limitante/limitante), soit un total de 18 essais (lieu × année × N) avec 2-3 répétitions de chaque génotype dans chaque condition. Un modèle linéaire mixte ( $Y_{ij} = \mu + G_i + R_j + \epsilon_{ij}$ ) a permis de corriger les données par essai et d'extraire l'effet génotypique pour chaque variable d'intérêt. Ces valeurs ont été utilisées comme variables d'entrée des analyses GWAS (modèle linéaire mixte avec Kinship, MLM-K). Deux détections ont été réalisées pour les GSL, en considérant le panel entier (174 génotypes) et uniquement les double zéros (128 génotypes). Les SNP présentant un FDR < 20% (seuil de 5% pour les GSL) ont été considérés comme significativement associés au trait ciblé. Pour synthétiser l'information issue des analyses GWAS, nous avons défini des QTL comme étant des groupes de SNP respectant les critères suivants : i) au minimum 3 SNP associés distants d'au plus une étendue de DL, et ii) seuls les QTL détectés dans au moins deux environnements sont conservés.

Variables NAP\_2S, CRU\_12S et 12S/2S: Les données ont été acquises par chromatographie à BIA Nantes sur le panel SEEDPROT (104 génotypes) provenant d'un seul lot de graines (pas de répétitions biologiques). Par conséquent, les analyses GWAS ont été réalisées sur les médianes des données brutes (absence d'estimation de l'effet génotypique), aussi les résultats GWAS sont à considérer avec précaution du fait de la faible puissance statistique liée au nombre de données.

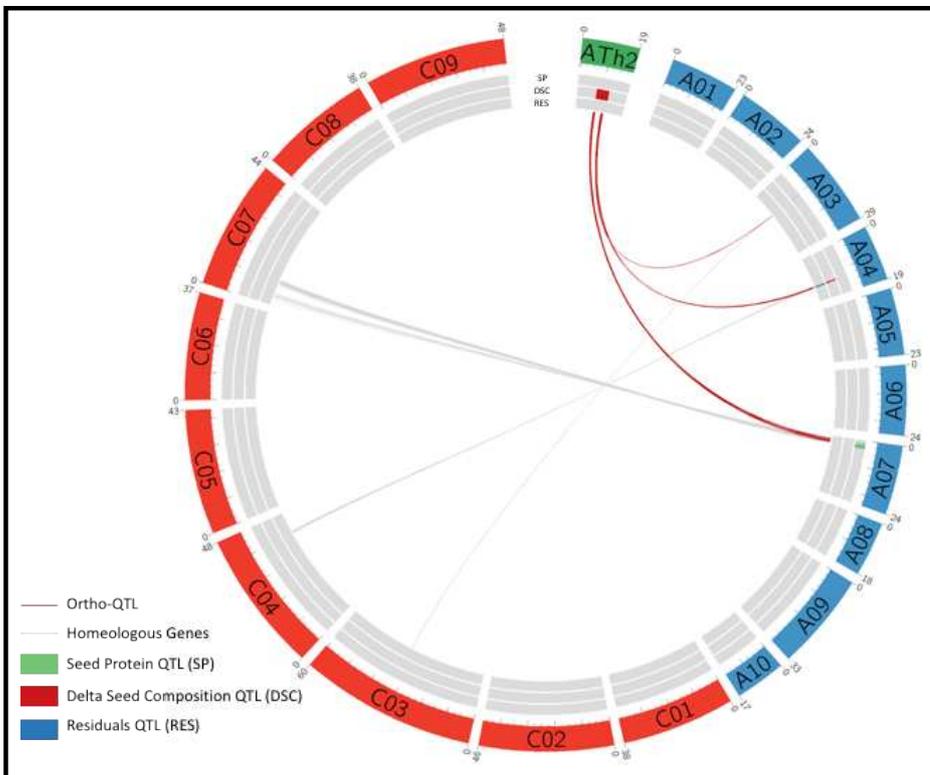
- Résultats

Sur la base de ces critères, nous avons pu identifier un total de **259 QTL** pour les traits analysés (Figure A) : **O (1), SP (57), O/SP (2), MPT (33), DSC (45), RES (39), GSL\_174 (69), GSL\_128 (6), NAP\_2S (7), CRU\_12S (0) et 12S/2S (0)**. De nombreuses co-localisations sont observées notamment entre les



**Figure B.1 : Représentation des QTL de qualité de la graine sur le génome du colza**

Le graphe Circos représente le génome du colza (chromosomes A01-A10 en rouge et C01-C09 en bleu). Les QTL identifiés par GWAS sont indiqués sur les pistes des traits correspondants et avec un code couleur selon l'effet allélique. De nombreuses colocalisations sont observées (notamment pour les loci impliqués dans le contrôles DSC/RES). Par ailleurs, les relations d'homéologie sont indiqués par les traits au centre. Les traits bleus mettent en valeur les couples d'homéo-QTL.



**Figure B.2 : Colinéarité entre les génomes de colza (A-C) et arabidopsis (chromosome Ath2)**

Un QTL spécifique du DSC identifié chez arabidopsis est situé sur une portion du chromosome 2 qui présente de fortes similarités de séquences avec des zones des chromosomes A03, A04, A07 de colza qui portent elles-mêmes des QTL de qualité de la graine (SP, DSC, RES).

traits DSC et RES. Par ailleurs, il existe des relations d'homéologie entre QTL (couples A02/C02, A03/C03-a, A03/C03-b, A03/C07, A08/C03, A09/C09) (Figure B.1).

Par ailleurs, la comparaison des QTL de qualité entre les génomes d'Arabidopsis et de colza a permis d'identifier des régions colinéaires portant des QTL d'intérêt (Figure B.2).

Un travail à venir (projet SEEDQUAL, financement PROMOSOL 2019-2022) permettra de prioriser les QTL les plus intéressants pour pouvoir poursuivre leur caractérisation fine.



# GENETIC DETERMINISM OF SEED STORAGE PROTEINS IN OILSEED RAPE: EVIDENCE FOR A LINKAGE DRAG BETWEEN LOCI CONTROLLING GLUCOSINOLATE OR NAPIN CONTENTS

Erwan CORLOUER<sup>1</sup>, Véronique SOLE-JAMAULT<sup>2</sup>, Aude LE GOFF<sup>2</sup>, Julien NAVARRO<sup>1</sup>, Anne LAPERCHE<sup>1</sup>, Nathalie NESI<sup>1</sup>

<sup>1</sup> IGEPP, INRAE, Institut Agro, Université Rennes, 35650 Le Rheu, France

<sup>2</sup> BIA, INRAE, 44316 Nantes, France

## CONTEXT

Seed storage proteins in rapeseed consist of 12S globulins (cruciferins) and 2S albumins (napins) that stand for around 70% of total seed proteins. The amino acid composition of rapeseed proteins is excellent although slightly deficient in essential amino acids compared to soy proteins. In addition, with 3-4% more sulfur residues rapeseed protein quality would reach FAO's recommendations for human nutrition. Since napins contain more sulfur residues than cruciferins (8,1% Cys + Met vs. 2.7%), they display high value for food or feed usages. To date, genetic diversity of the seed storage protein composition has been poorly documented in rapeseed, but a few studies have reported a cruciferin enrichment at the expense of napins in modern grown double low (00) rapeseed cultivars (Malabat et al., 2003; Schatzki et al., 2014). The aim of the present study was to confirm this latter tendency within a population maximizing the genetic diversity in winter oilseed rape and characterize its genetic and molecular bases. In particular, the potential linkage between the seed glucosinolate content and seed storage protein profile was investigated.

## PLANT MATERIAL

### ✓ Diversity set

- 174 WOSR accessions incl.:
  - 31 “++” varieties, high erucic acid (EA) & glucosinolates (GSL)
  - 15 “0+” varieties, low EA & high GSL
  - 1 “+0” variety, high EA & low GSL
  - 127 “00” varieties, low EA & GSL

### ✓ Field network

- The WOSR panel was grown over 7 environments (location x year) under 2 N nutrition regimes => 14 trials (environment x N) included in the RAPSODYN network (Figure B.3).

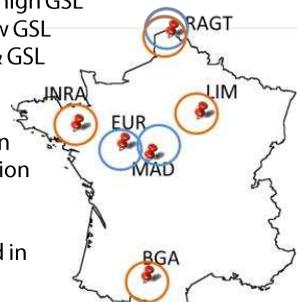


Figure B.3: French field network for seed quality phenotyping in rapeseed

### ✓ Genotyping

- High dense genotyping assays (237k SNP)

## ANALYSES of 12S and 2S CONTENTS

Seed storage proteins were extracted from milled and defatted seeds in Tris 50mM pH8.5 NaCl 750mM EDTA 5mM sodium bisulfite 15mM buffer (ratio 1/10). Pigments were removed by desalting chromatography. Samples were injected onto size exclusion chromatography (Superdex75 Increase, GE Healthcare) with refractometer detection. Linear regression was done between area response and increasing quantity of 2S and 12S standards. After calibration, the 2S and 12S contents of each accession were determined in triplicates.

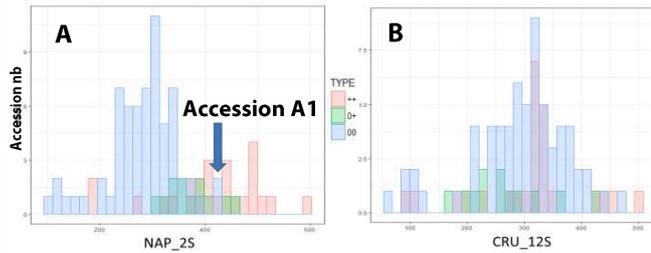


Figure B.4: Distribution of 2S (left) and 12S (right) seed contents after quantification by size exclusion chromatography. Note that only 102 WOSR accessions were considered for quantification due to the low throughput of the method. These accessions were chosen to fit the whole panel composition.

Distributions of 2S and 12S seed contents highlighted significant lower 2S amounts in double low modern grown varieties compared to high GSL (“++” and “0+”) accessions (Figure B.4-A), thus confirming previous studies. However, high 2S content was still retained in some double low accessions (ex. Accession A1). On the contrary, there was no clear effect on the distribution of the 12S contents in the whole population studied (Figure B.4-B).

## GWAS RESULTS

- GWAS was conducted using a mixed linear model including a Kinship matrix (MLM-K). A total of 7 loci controlling the 2S content were identified in the rapeseed genome, located on chromosomes A07, A08, A09, C03 (2), C07 and C09 (Figure B.5 – arrows).
- QTL located on chrom. A07, A08, C07 and C09 were almost fixed in the population (data not shown). On the contrary, for the 3 other loci (chrom. A09 and C03), the favorable allele for 2S content was the minor allele in the population (Figure B.5), mostly carried by the old varieties rich in GSL (Figure B.6).
- In addition, these 3 latter QTL co-localized with QTL for seed GSL content found in the whole panel (Figure B.5) for which the favorable allele was also minor within the population. There was no colocalization with GSL QTL found in the double low accessions. These results pointed out the undeniable relationship between high GSL and 2S respective contents, with loci located on A09 and C03 arising as good candidate regions for controlling both traits.
- Finally, some double low genotypes (ex. Accession 1) showed the favorable allele at QTL A09 (Figure B.6), providing evidence that there is still scope to disconnect 2S and GSL contents for breeding purposes.

Figure B.5: Circos diagram showing the mapping of QTL for 2S and GSL onto the rapeseed genome.

Seven 2S QTL were identified based on protein composition analysis as shown on Fig.2. The GSL QTL were calculated within the whole population (“GSL-ALL” lane) or the double low subset (“GSL-00” lane). Pairs of homoeologous genes between the A (blue) and C (red) sub-genomes are represented by gray lines (homoeologous genes between QTL and regions carrying no other QTL) or blue lines (homoeologous genes between QTL).

- Minor allele is favorable
- Major allele is favorable

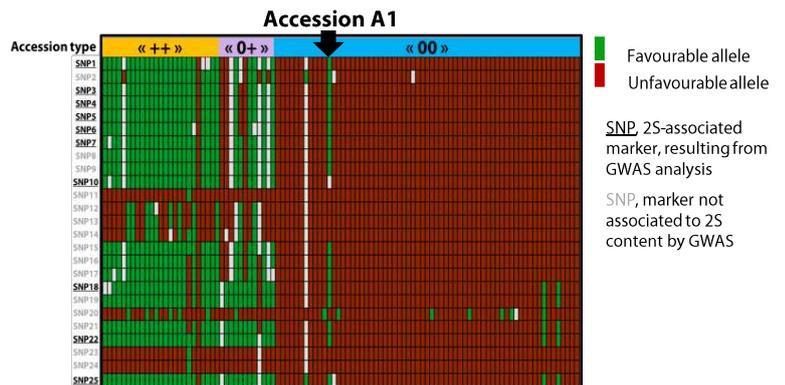
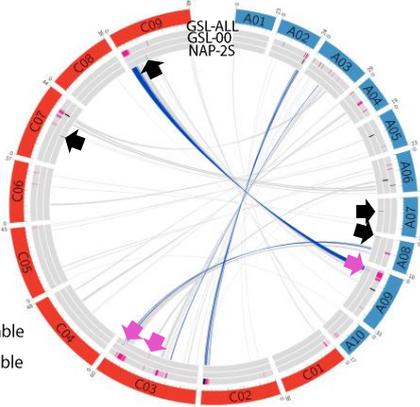


Figure B.6: SNP information according to genotypes: example of the 2S QTL A09.

## CONCLUSION

Our results showed that the glucosinolates and 2S napin contents were positively correlated in winter oilseed rape. Conversion to low glucosinolate varieties seemed to have reduced the napin content in seeds. The sulfur metabolic pathway has probably been affected by selection. These results provided new clues how to improve the seed protein composition in rapeseed.



## ANNEXE C

---

## ACTIVITES SCIENTIFIQUES

---



### Article avec comité de lecture

**Corlouer E.**, Gauffreteau A., Bouchet A-S., Bissuel-Bélaygue C., Nesi N., Laperche A. Envirotypes based on seed yield limiting factors allow to tackle GxE interactions. *Agronomy* 2019, 9, 798. 798; <https://doi.org/10.3390/agronomy9120798>

### Article sans comité de lecture

Pinochet X., Laperche A., **Corlouer E.**, Nesi, N. La recherche se penche sur la valorisation de l'azote. *Perspectives Agricoles* 2020, n°480

### Communications orales avec acte

**Corlouer E.**, Bouchet A-S., Gauffreteau A., Bissuel-Bélaygue C., Nesi N., Laperche A. – Temperature and radiation stresses explained most of the environmental variations of seed yield across a French experimental network, and allowed to tackle GxE interaction issues in WOSR. *Symposium Gis BV inter PIA, Octobre 2019 à Paris (France)*

**Corlouer E.**, Bouchet A-S., Gauffreteau A., Bissuel-Bélaygue C., Nesi N., Laperche A. – Temperature and radiation stresses explain most of the environmental variation of seed yield across a French network, and allow to tackle GxE interaction in winter oilseed rape cultivars. *15<sup>th</sup> International Rapeseed Congress, Juin 2019 à Berlin (Allemagne)*

**Corlouer E.**, Solé-Jamault V., Le Goff A., Navarro J., Laperche A., Nesi N. – La sélection pour la faible teneur en glucosinolates a induit une modification de la composition en protéines de réserve de la graine de colza. *Graines 2019, Mai 2019 à Angers (France)*

**Corlouer E.**, Bouchet A-S., Gauffreteau A., Bissuel-Bélaygue C., Nesi N., Laperche A. – Identification of the main factors limiting seed yield in rapeseed helps dissecting the environment and GxE effects. *Journée Jeunes Chercheurs BAP 2018, Septembre 2018 à Paris (France)*

**Corlouer E.**, Guilbert O., Bouchet A-S., Baron C., Faure S., Leveugle M., Dheu J-E., Simon C., Foubert T., Maes O., Dugué D., Dyrszka E., Moreau J., Laperche A., Nesi N. – Genetic and molecular determinants of nitrogen use efficiency in winter oilseed rape. *Brassica 2018, Juillet 2018 à Saint-Malo (France)*

**Corlouer E.**, Bouchet A-S., Laperche A., Nesi N. – Analysis of QTL stability/sensitivity to the N regime and environmental factors. *International "Plant Abiotic Stresses" Workshop, Novembre 2016 à Rennes (France)*

Laperche A., Le Cahérec F., **Corlouer E.**, Bouchet A-S., Navarro J., Orsel-Baldwin M., Bissuel C., Nesi N., Niogret M-F. – Genetic basis of Nitrogen Use Efficiency and Yield stability across N conditions in rapeseed: combining quantitative genetic and functional genomic approaches. *The Plant and Animal Genome XXV Conference (PAG), Janvier 2017 à San Diego (CA-USA)*

### Communications orales sans acte

**Corlouer E.**, Nesi N., Laperche A. – New indicators revealed by PLS analyses. *6<sup>th</sup> RAPSODYN annual meeting, Novembre 2018 à Rodez (France)*

**Corlouer E.**, Solé-Jamault V., Guerche P., Jasinski S., Nesi N. – SEEDPROT project. *6<sup>th</sup> RAPSODYN annual meeting, Novembre 2018 à Rodez (France)*



Laperche A., Retailleau L., **Corlouer E.** – Genomic prediction assays using environmental indicators. *6<sup>th</sup> RAPSODYN annual meeting, Novembre 2018 à Rodez (France)*

Nesi N., **Corlouer E.**, Guerche P., Lécureuil A., Le Goff A., Anton M., Guilbert O., Navarro J., Rolland S., Baron C., Solé-Jamault V., Jasinski S. – SEEDPROT. Teneur et qualité des protéines de la graine de colza : exploration de la diversité génétique en vue d'optimiser les usages en alimentation animale et humaine. *Carrefour Colza, Janvier 2018 à Laval (France)*

**Corlouer E.**, Bouchet A-S., Gauffreteau A., Bagot P., Bissuel C., Nesi N., Laperche A. – Characterization of the Rapsodyn field network using environmental variables to investigate GxE interaction. *5<sup>th</sup> Annual Meeting RAPSODYN, Décembre 2017 à Mondonville (France)*

Guilbert O., **Corlouer E.**, Nesi N., Laperche A. – Identification and characterization of genetic and molecular determinants of seed yield in winter oilseed rape. *5<sup>th</sup> Annual Meeting RAPSODYN, Décembre 2017 à Mondonville (France)*

### Communications écrites

Virlet G., **Corlouer E.**, Termier A., Laperche A., Nesi N., Lavenier D. – Interaction de SNPs d'ordre N par pattern mining. *Jobim 2019, Juillet 2019 à Nantes (France)*

**Corlouer E.**, Solé-Jamault V., Le Goff A., Navarro J., Laperche A., Nesi N. – Genetic determinism of seed storage proteins in oilseed rape : evidence for a linkage drag between loci controlling glucosinolate or napin contents. *15<sup>th</sup> International Rapeseed Congress, Juin 2019 à Berlin (Allemagne).*

Guerche P., Lécureuil A., **Corlouer E.**, Nesi N., Jasinski S. – The Mitogen-Activated Protein Kinase 12 is involved in Arabidopsis seed filling. *Graines 2019, Mai 2019 à Angers (France)*

**Corlouer E.**, Solé-Jamault V., Le Goff A., Laperche A., Nesi N. – Intense selection for seed quality traits has changed the seed storage protein balance in oilseed rape. *Plant and Animal Genome XXVII Conference, Janvier 2019 à San Diego (USA)*

Solé-Jamault V., Le Goff A., **Corlouer E.**, Guilbert O., Nesi N. – Selection for low glucosinolate content in seeds has modified the seed storage protein profile in rapeseed. *Journée Jeunes Chercheurs BAP 2018, Septembre 2018 à Paris (France)*

Lécureuil A., **Corlouer E.**, Guerche P., Nesi N., Jasinski S. – Genetic architecture of seed quality in Brassicaceae seeds. *Brassica 2018, Juillet 2018 à Saint-Malo (France)*

Solé-Jamault V., Le Goff A., **Corlouer E.**, Guilbert O., Nesi N. – Selection for low glucosinolate content in seeds has modified the seed storage protein profile in rapeseed. *Brassica 2018, Juillet 2018 à Saint-Malo (France)*

**Corlouer E.**, Bouchet A-S., Gauffreteau A., Bissuel-Belaygue C., Nesi N., Laperche A. – Identification of the main factors limiting seed yield in rapeseed helps dissecting the environment and GxE effects. *Brassica 2018, Juillet 2018 à Saint-Malo (France)*

Lécureuil A., Guerche P., **Corlouer E.**, Navarro J., Nesi N., Jasinski S. – Genetic architecture of seed quality in Brassicacea species. *Graines 2017, octobre 2017 à Montpellier (France)*



## Manifestation de médiation scientifique

Participation au concours 'Ma Thèse en 180 secondes', qualification rennaise, Mars 2020.

## Formations

MOOC Ethique de la recherche

Organisation d'une animation scientifique hebdomadaire : Lab Meeting RCA (2018-2019)

Formation : « vulgarisation scientifique » dans le cadre du concours MT180

Formation : « Analyse des réseaux expérimentaux et la méta-analyse »

## Enseignement

Vacation d'enseignement à AgroCampus Ouest – Chargé de TD en génétique quantitative (étudiants de niveaux M2 et 3<sup>ème</sup> année ingénieurs)

## Distinction

Winner Poster Award: 15<sup>th</sup> International Rapeseed Congress (IRC), Berlin – June 2019





**Titre :** Identification des déterminants génétiques et moléculaires contrôlant la stabilité du rendement chez le colza d'hiver

**Mots clés :** *Brassica napus*, Rendement, Interaction G×E, Envirotypage, QTL

**Résumé :** Un des enjeux actuels de l'amélioration des plantes est d'assurer une production optimisée et stable dans des systèmes de culture économes en intrants et soumis à des environnements de plus en plus fluctuants. Le phénotype observé résulte de l'expression des effets du génotype et de l'environnement mais aussi de l'interaction génotype × environnement (G×E). Proposer des variétés stables/adaptées à une large gamme de milieux ou adaptées à des environnements spécifiques nécessite de comprendre finement et d'exploiter les interactions G×E. Dans ce contexte, ce travail de thèse a cherché à répondre aux questions suivantes : quels sont les principaux facteurs abiotiques limitant le rendement chez le colza ? Comment valoriser ces facteurs pour établir une typologie des environnements qui permette de les caractériser et de les regrouper de manière à décomposer l'interaction G×E ?

Cet envirotypage permet-il de caractériser les déterminants génétiques contrôlant le G×E ? Une stratégie basée sur la caractérisation et l'analyse d'un réseau de 47 essais répartis sur les principales zones de culture du colza en France et conduits dans des conditions contrastées de fertilisation azotée a été déployée pour identifier les facteurs limitant le rendement et regrouper les essais en envirotypes. Puis, une analyse génétique par association a été menée sur 173 génotypes de colza d'hiver expérimentés dans ce réseau afin d'identifier les QTL qui contrôlent le rendement et l'interaction G×E. Des QTL stables et spécifiques ont été identifiés en fonction des envirotypes. L'intérêt de cette approche et des résultats obtenus est discuté, notamment au regard de leur valorisation potentielle dans des programmes de sélection

**Title :** Identification of the genetic and molecular determinants controlling seed yield stability in winter oilseed rape

**Keywords :** *Brassica napus*, seed yield, G×E interaction, envirotyping, QTL

**Abstract:** One of the current challenges in plant breeding is to ensure optimized and stable production in low-input cropping systems under fluctuating environments. The observed phenotype results from the expression of the effects of the genotype and the environment in addition to genotype × environment interaction (G×E). Breeding for varieties that are stable/adapted to a wide range of environments or, on the opposite, specifically adapted to targeted environments requires understanding and exploiting the G×E interaction. In this context, this thesis work aimed at answering the following questions: what are the main abiotic factors limiting seed yield in winter oilseed rape. How can these factors be used to define environmental typologies that make it possible to characterize and cluster environments to decompose the G×E interaction?

Finally, is this envirotyping relevant to characterize the genetic determinants controlling the G×E ? A strategy based on the characterization and analysis of a network of 47 trials spread over the main rapeseed growing areas in France and conducted under contrasting nitrogen nutrition regimes, was used to identify the factors limiting seed yield in rapeseed and to cluster the trials into envirotypes. Subsequently, a genome wide association study was conducted on 173 winter oilseed rape genotypes tested within this network in order to identify the QTL that control seed yield and the G×E interaction. Stable and specific QTL were identified according to envirotypes. The interest of this approach and the results obtained are discussed, particularly with the regards to their potential use in breeding programs.