

Specialized metabolism in Brassica napus: Characterization of phytochemical diversity, its genetic determinants, and its regulation by pathogen infection

Anani Amegan Missinou

▶ To cite this version:

Anani Amegan Missinou. Specialized metabolism in Brassica napus : Characterization of phytochemical diversity, its genetic determinants, and its regulation by pathogen infection. Agricultural sciences. Agrocampus Ouest, 2022. English. NNT: 2022NSARC162. tel-04382865

HAL Id: tel-04382865 https://theses.hal.science/tel-04382865

Submitted on 9 Jan2024

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.



Thèse de doctorat de

AGROCAMPUS OUEST

ECOLE DOCTORALE N° 600 Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation Spécialité : « Génétique, génomique et bio-informatique »

Par Anani Amegan MISSINOU

Specialized metabolism in *Brassica napus*: Characterization of phytochemical diversity, its genetic determinants, and its regulation by pathogen infection

Thèse présentée et soutenue à Rennes, le 04 Juillet 2022

Unité de recherche : Institut de Génétique, Environnement et Protection des Plantes

Thèse N° : C-162 2022-13

Rapporteurs :

Emmanuel Gaquerel Sylvie Baudino Professeur, Université Strasbourg Professeure, Université Jean Monnet Saint Etienne

Composition du Jury :

Président : Joël Boustie

Examinateurs :

Charles-Eric Durel Marie-Hélène Balesdent Sandrine Lagarrigue

Dir. de thèse : Antoine Gravot **Co-dir. de thèse :** Maria Manzanares-Dauleux Professeur, Université Rennes 1

Directeur de recherche, INRAE, IRHS Directrice de recherche, INRAE, UMR 1290 BIOGER Professeur, Institut Agro Rennes-Angers, PEGASE

Professeur, Université Rennes 1, IGEPP Professeure, Institut Agro Rennes-Angers, IGEPP

Remerciements

Toute la louange revient à Allah, le Tout Puissant qui améliore ma situation en tout point et ne m'a jamais délaissé à mon propre sort ne serait-ce que le temps d'un clin d'œil, et m'a octroyé la force, le courage et la patience d'accomplir ce modeste travail.

Fruit d'un travail de recherche triennale et demi, ce manuscrit résulte d'un concours d'efforts solidaires de plusieurs personnes envers qui je suis reconnaissant et que je ne pourrai tous citer.

En première ligne, Antoine GRAVOT et Maria MANZANARES-DAULEUX m'ont fait vivre plus qu'une direction de thèse, comme si un directeur de thèse pouvait aussi adopter son doctorant. Vous avez été à l'écoute de mes questions et intéressés par l'avancée de mes travaux. Votre capacité d'analyse et votre enthousiasme m'ont montré que le monde de la recherche pouvait être un univers passionnant. Les nombreuses discussions que nous avons eues ainsi que vos encouragements et conseils ont, d'une part guidé mes pas et donné un ton initiatique à ces années de recherche et d'autre part sont pour beaucoup dans le résultat final de ce travail. Veuillez trouver ici, l'expression de mes respectueux hommages.

En seconde ligne, ma profonde gratitude va à Régine DELOURME, Anne-Marie CHÊVRE, Anne LAPERCHE et Alain BOUCHEREAU, pour leurs accueils chaleureux, leurs conseils avisés, leurs franchises, leurs corrections si éclairantes. Mention spéciale à Régine D., pour sa protection maternelle, qu'elle en soit aussi remerciée pour sa gentillesse, sa disponibilité permanente et pour ses nombreux encouragements et aussi à Alain B. pour ses implications conséquentes qui ont permis de rendre les différentes collaborations plus agréables et d'enrichir les travaux présentés dans ce manuscrit et pour les opportunités qu'il m'a offerte de faire connaître mes travaux en m'encourageant à participer aux conférences.

En troisième ligne, je remercie grandement Nathalie MARNET pour toute ton aide. Je suis ravi d'avoir travaillé en sa compagnie car outre son appui scientifique et les nombreux conseils qu'elle m'a prodigué, non seulement elle m'a initié à la théorie et pratique en spectrométrie de masse en début de ma thèse, mais elle m'a aussi inculqué une rigueur et précision dans l'interprétation des spectres de masse. Je remercie Thomas DELHAYE pour son soutien indéfectible, nos discussions techniques autour des spectres de fragmentation, son regard différent et plus pragmatique en tant qu'expert du domaine et ces commentaires très enrichissants. Merci également à Clément ORIONE, Philippe JÉHAN, et Régis LAVIGNE pour leur participation scientifique ainsi que le temps qu'ils ont consacré à ma recherche, remerciements auxquels j'associe Sylvain CHEREAU pour sa bonne humeur permanente et pour son aide, en particulier sur la fin de la thèse. Tous répondirent avec calme, patience enthousiasme et sympathie aux nombreuses questions dont je les accablais. Ce fut un réel plaisir de travailler avec vous et merci encore pour vos contributions analytiques.

En quatrième ligne, j'adresse mes sincères remerciements à tous les membres de mon comité de thèse (Catherine LEBLANC, Hugues RENAULT, Mathieu ROUSSEAU-GUEUTIN et Mélanie LAGARRIGUE-REBOUTIER) pour les discussions, les encouragements et leur regard extérieur sur mon travail, mais également à tous les membres du jury (Emmanuel GAQUEREL, Joël BOUSTIE, Sylvie BAUDINO, Marie-Hélène BALESDENT, Charles-Éric DUREL et Sandrine LAGARRIGUE) pour l'honneur qu'ils m'ont fait en acceptant de siéger dans le jury de ma soutenance en vue d'examiner, de juger et délibérer sur ce travail et pour avoir eu avec eux des échanges agréables et intéressants lors de la soutenance.

En outre, je témoigne un sincère remerciement à tous mes collaborat.eur.rice.s et collègues chercheur(e)s, postdoctorant(e)s, doctorant(e)s, ingénieur(e)s, stagiaires, technicien(ne)s, et du service commun administratif et informatique, plus particulièrement à Anne L., Cyril F., Christine L., David A.S., Erwan C., Franz B., Géraldine B., Grégoire B., Gautier R., Jocelyne L., Julien B., Julie F.d.C., Loeiz M., Marianne L., Mathilde P., Magali E., Maude F., Olivier F., Pascale I.N., Philippe D., Pierre-Nicolas B., Yann G., Youssef A.H., Younès D., Sandra W., Solenne B., Sophie C., Thelma J., ... pour leurs aides précieuses, leurs sollicitudes, leurs encouragements, leurs savoirs scientifiques et leurs échanges amicaux pendant cette période de labeur.

Merci à Youssef B. pour sa bonne humeur, ses tajines, son soutien sans faille, sa gentillesse rare et pour tous les bons moments partagés que nous ne manquerons pas de renouveler en France, au Togo, au Maroc ou ailleurs. Merci également à Youcef H. avec qui j'ai pris beaucoup de plaisir à travailler et qui après avoir été un stagiaire très sympathique, est devenu un ami et j'espère le restera.

Je voudrais ici avoir une pensée particulière à mes enseignant(e)s, professeur(e)s et encadrant(e)s de stage, en particulier Xavier DAUVERGNE qui m'a fait découvrir le monde palpitant de la phytochimie en Master et m'a recommandé pour cette thèse ainsi à Sophie GALLINA pour m'avoir prodigué les ficelles de la bio-informatique. Qu'ils reçoivent ici mon plus profond respect et ma reconnaissance.

Une série de remerciements à toutes les personnes avec qui j'ai partagé mes études et notamment en Côte D'ivoire, au Togo et en France, à mes amis et compatriotes rencontrés à l'université de Lomé (Adjélevi, T.L, Amétépé H., Bila, Charles, Carine A., Anty D., Kpoti T.L, Nam-pan S., Richel A., ...) et à ceux avec qui le chemin s'est poursuivi en France (Dedry K., Anani K., Pascal N. et Venceslas N.), à Guendehou D., Hoang N., Landy R., Sena A., Didier K., Selom S., Tatiana T., Tilate K., et Nicole R. pour leur proximité permanente et illimitée depuis notre rencontre à Brest, à l'ensemble des inconditionnels soutiens des frères en Allah (Aboubakr S., Asbath B., Abdallah K., Ismaïl M., Mehdi L., Sékou D., Youssef B., Mehdi B., ...), je vous exprime ma grande reconnaissance.

Enfin, mes chaleureux remerciements aux membres de la famille MISSINOU pour leurs soins, attentions et espérances perpétuels pour conduire leur petit-fils, fils, neveu et frère, à l'accomplissement de ce travail, en particulier à mon Père et ma Mère, qui ont tout fait pour me soutenir et surtout me supporter dans tout ce que j'ai entrepris.

Table des matières

Foreword	11
Chapter I: Bibliographic introduction	13
Introduction	14
I.A. The specialized metabolism in vascular plants	15
I.A.1. The paradigm of the multifunctionality of specialized plant metabolism	16
I.A.2. Structural and biosynthetic diversity of the plant-specialized metabolites	21
I.A.2.1. Structural diversity of plant-specialized metabolites	21
I.A.2.2. Metabolic origin of plant specialized metabolites	27
I.A.2.3. Biosynthetic diversity of plant-specialized metabolites	30
I.A.3. Factors involved in the evolutionary diversification of plant specialized metabolites	31
I.A.3.1. Genomic processes involved in the diversification of SM-related biosynthetic enzymes I.A.3.1.1. Gene duplication	31 <i>32</i>
I.A.3.1.2. Convergent evolution	33
I.A.3.1.3. Enzyme promiscuity	35
I.A.3.2. Environmental drivers of plant SM diversification over geologic time-scale	36
I.A.4. Factors involved in the regulation of plant-specialized metabolite biosynthesis	38
I.A.4.1. Transcription factors (TF)	39
I.A.4.2. Ontogenic factors	41
I.A.5. Plant Immunity: Qualitative and quantitative resistance	42
I.A.5.1. Qualitative resistance	43
I.A.5.2. Quantitative resistance	45
I.A.6. Contribution of specialized metabolism to plant immunity	46
I.A.6.1. Constitutive chemical defenses	46
I.A.6.2. Induced chemical defenses	49
I.A.7. Oilseed rape (<i>B. napus</i>)	51
I.A.7.1. Origin and genomic evolution of oilseed rape	51
I.A.7.2. Economic importance and breeding of <i>B. napus</i>	53
I.A.7.3. <i>Brassica</i> Genomic Resources	54
I.A. 7.4. Brassica napus - Leptosphaeria maculans interaction	55 57
1.A.7.4.1. The life cycle of Leptosphaeria maculans in Europe	57
1.A.7.4.2. Biochamical bases of constitution and indusible chamical defenses of Basesian	00
I.A.7.4.3. Biochemical bases of constitutive and inducible chemical defenses of Brassicas	63
I.B. Presentation of the Ph.D	68
I.B.1 Objectives of the Ph.D. and research questions	69
I.B.2. Thesis outlines	69
Chapter II: Characterization of the diversity of constitutive specialized metabolites and	d its
genetic determinants in <i>Brassica</i> species	73
Introduction du chapitre II	75

Paper I: Identification and Quantification of Glucosinolates and Phenolics in a Large Panel of	of
Brassica napus Highlight Valuable Genetic Resources for Chemical Ecology and Breeding	81
Paper II: GWAS reveals genetic networks involved in the control of phenolics and glucosino	lates,
and specifies the metabolic consequences of breeding history in <i>Brassica napus</i> cultigroups	101
1 Introduction	103
2 Resultats and Discussion	106
2.1. Glucosinolates and Phenolic compounds identified in <i>B. nanus</i>	106
2.2. Multivariate analyses to identify discriminating molecules between <i>B</i> nanus cron types	110
2.3. GWA analysis to decinher the genetic control of GLS and PHL content in leaves and roots	110
3 Experimental procedures	121
3.1. Plant material	121
3.2 Genotyping	123
3.3. Statistical analyses	123
3.4. Genome-Wide Association analyses	124
3.5. Linkage disequilibrium approach to identify OTLs	124
3.6. Network analyses	125
Paper III: Genetical phytochemistry of <i>Brassica napus</i> roots revealed OTLs involved in the	_
control of two pteridine-like aminoacid conjugates	139
Abstract	141
1 Introduction	141
1. Introduction	1 4 2
2.1 Plant restarial and severalize	145
2.1. Plant material and sampling	145
2.2. Chemicals	145
2.3. Metabolite extraction procedure	140
2.4. Chromatography, semi-quantitative proming	140
2.5. Multiveriate statistical analysis	147
2.6. Multivariate statistical analysis	147
2.7. Metabolic genome-wide association analysis	140
2.0. Metabolito nurification	149
2.5. Metabolice purification	149 1/0
Reversed-phase SFL fractionation procedure	149
Semi-preparative HPLC	150
2.10. Structural analysis of purified molecules	151
Q-Exactive Orbitrap MS/MS experiments	151
NMR experiments	151
2 Posults	152
2.1 LC MS/MS fragmentation of eighteen unidentified LW absorbing compounds extracted from	152
S.1. EC-Wis/Wis/Hagmentation of eighteen unidentified OV-absorbing compounds extracted from	150
2.2. Targeted metabolite profiling of UV absorbing at 220 pm in <i>Practice</i> roots highlights wide	152
5.2. Targeted metabolite proming of 00-absorbing at X550 mm in <i>Brassica</i> roots nightights wide	155
3.3. Chemotypes pointed out through multivariate analyses	ررب ۲۲۵
3.4 mGWAS identified genomic regions (mOTLs) associated with novel chemical traits in the roo	100
Brassica	163
3.5. Analysis of potential candidate genes	166

3.6. Combining NMR and MS allowed a simultaneous assignment and structural elucidation	on of two
novel root compounds in <i>Brassica</i>	
4. Discussion	
5. Conclusion	
Author contributions	
Funding sources	175
Associated content	
Supporting Figures	
List of supporting Figures	180
List of tables	
List of Figures	181
References	182
Chapter III: Metabolic responses of Brassica napus after Leptosphaeria maculan	s infection
	185
Introduction du chapitre III	187
1. Introduction	189
2. Materials and Method	191
2.1. Chemicals	191
2.2. Plant growth	191
2.3. Inoculum preparation and petiole inoculations	
2.4. Metabolomic analysis	192
2.4.1. Targeted metabolomic profiling	192
2.4.2. Untargeted metabolomic profiling procedures : Sampling, GLSs extraction, and que	Intification
	192
2.4.3. Mass spectrometry imaging (MSI) by MALDI (Matrix-Assisted Laser Desorption-Ioni	sation)
procedures	193
2.4.3.1. Preparation of stem cross-sections	193
2.4.3.2. High-resolution MALDI FT-ICR mass spectrometry imaging measurement	193
3. Results and Discussion	194
3.1. Regulation of GLS content in oilseed rape stems infected by <i>L. maculans</i>	194
3.2. Mass spectrometry imaging reveals the spatial distribution of metabolic respon	ses in stems
of B. napus after L. maculans infection	198
3.2.1. Metabolic responses to <i>L. maculans</i> infection in the two genotypes of <i>B. napus</i> 'Da (Bristol'	rmor' and
3.2.2 Genotype-specific metabolic responses to L maculans	
A Conclusion and perspectives	201
References	203 201
Chapitre IV : Discussion générale et Perspectives	204
IV. A. Discussion et conclusion générale	209
IV.A.1. Les GLSs de Brassica issues de voies de biosvnthèses anciennes	
IV.A.2. Les GLSs issues des voies de biosvnthèses plus récentes	
IV.A.3. Redondance structurelle dans la biodiversité GLSs	

IV.B. Perspectives	214
IV.B.1. Étendre l'inventaire des métabolites spécialisés chez <i>B. napus</i>	. 214
IV.B.2. Profilage métabolomique non ciblé de <i>B. napus</i> et annotation des métabolites à l'aic	le de
réseaux moléculaires	. 216
IV.B.3. Affiner la puissance de détection de gènes candidats	. 217
IV.B.4. Validation fonctionnelle des gènes candidats	. 218
IV.B.5. Investiguer les clusters de gènes biosynthétiques (BGCs) et leur évolution chez les	
Brassicas	. 219
References	.221

Líste des fígures

Figure 1: Similarities between the chemical structures of phytohormones and primary and specialized metabolites.

Figure 2: Proposed model for the role of Pipecolic acid (Pip (02)) during activation of local resistance, systemic acquired resistance, and defense priming in *Arabidopsis thaliana*.

Figure 3: A New paradigm of the functional versatility/polyvalence of the plant metabolites.

Figure 4: Core structural diversity of specialized metabolites derived from the central metabolism in the plant kingdom.

Figure 5: Simplified biosynthesis route showing the unexhaustive origin of most plant-specialized metabolites from (a) Calvin cycle and (b) glycolysis.

Figure 6: Diverse routes of metabolite precursors in the conceptual metabolic grid schema

Figure 7: Potential fates of duplicated genes.

Figure 8: Independently evolved O-methylation of benzoxazinoids and glucosinolates regulates defense and resistance

Figure 9: Substrate and donor promiscuity of Sb3GT1 from Scutellaria baicalensis.

Figure 10: Origin and expression of specialized metabolites during green plant phylogeny lineage evolution

Figure 11: Conserved structures and domains of transcription factors regulate plant stress responses through mediating specialized metabolism.

Figure 12: Overexpression of the MYB transcription factor gene *PAP1* leads to enhanced phenylpropanoid production in *Arabidopsis thaliana* (left) and *Nicotiana tabacum cv xanthi* (right).

Figure 13: Prediction of ontogenetic trajectories of some plant defenses

Figure 14: Qualitative resistances and cellular actors underlying the characterized R genes.

Figure 15: Quantitative resistances and cellular actors underlying the characterized resistance QTLs

Figure 16: Localization of antifungal compounds in plants. (A) UV light autofluorescence of the saponin avenacin A1.

Figure 17: The "triangle of U" diagram (Nagaharu U, 1935) shows the genetic relationships between six genus species of *Brassica*.

Figure 18: Genomic origin of oilseed rape

Ph D. manuscript | Anani Amegan MISSINOU

Figure 19: A) Global distribution of oilseed rape production measured in tons in 2017 and B) Evolution of rapeseed production measured in tons from 1961 to 2017.

Figure 20: Presence of important diseases on oilseed rape worldwide (A) and in Europe (B).

Figure 21: The pathogenic life cycle of *Leptosphaeria maculans* in winter oilseed rape in Europe.

Figure 22: Genetic resistance associated with the pathogenic life cycle of *Leptosphaeria maculans*.

Figure 23: Representation of the main biosynthetic pathways of methionine-derived glucosinolates in *Brassica* species and *Arabidopsis*.

Figure 24: Biosynthesis pathways of auxins, phytoanticipins, and tryptophan-derived phytoalexins characterized (at least in part) in *Arabidopsis* and some *Brassica species*.

Foreword

Plants synthesize a vast diversity of low molecular weight compounds (<2000 Da), which exhibit a tremendous structural diversity and are highly specific to a taxon, a cell type, or a developmental stage. These compounds generally termed «secondary metabolites» (SM), have long been widely used in medicine, cosmetics, and agriculture, like spices, perfumes, flavors, fragrances, drugs, insecticides, and herbicides (Traka and Mithen, 2009; Dinkova-Kostova and Kostov, 2012; Zachariah and Leela, 2018; Faccio, 2020). They also include pigments or volatile compounds accumulated in flower petals to attract pollinators or respond to abiotic stress, bitter or toxic chemicals accumulated or triggered by herbivore injuries that serve as antifeedants, and compounds released by roots into the rhizosphere to modulate soil microorganisms. These metabolites are now considered «specialized» due to their predominant role in plant physiology and ecology (Bentley, 1999; Wink, 2003; Wink, 2010; Spinelli et al., 2011; Tissier et al., 2015), thus reconciling functional specialization and phylogenetic specificity. Consequently, they have commonly been considered the principal actors of ecological interaction processes that shape the species diversity, their coexistence, and trophic networks (Weng, 2014; Corwin and Kliebenstein, 2017; Vleminckx et al., 2018; Katz et al., 2021; Uckele et al., 2021).

The study of specialized metabolites is crucial in agriculture because of their ecological role. However, our understanding of the plant's specialized metabolome is yet insufficient. Indeed, a considerable part of the enzymes involved in specialized metabolism remains to be discovered or characterized (Milo and Last, 2012). The specific roles of most known specialized metabolites in physiological processes or responses to environmental stresses have not been conclusively established. The

Ph D. manuscript | Anani Amegan MISSINOU

recent development of high-throughput, spatial, or genetical metabolomics has opened new perspectives to assess the phytochemical diversity in large panels of genotypes, investigate their regulation under stress conditions, and identify the genetic determinants controlling their biosynthesis. Considering the central role of specialized metabolism in biotic interactions, this knowledge can support plant breeding strategies (Yang et al., 2021).

In this project, oilseed rape (*Brassica napus*) was chosen as a study model to analyze the diversity of the specialized metabolism and its relationship with the plant's response to biotic stresses. Today, oilseed rape/canola is the worldwide second most-produced oilseed, behind soybeans (Mittaine and Mielke, 2012), after being an auxiliary crop for food and industrial purposes in the last half-century. In France, its total cultivated area is 1.17 million hectares (Agreste: Grande cultures - n°2022-15 de Février 2022). The significant oilseed rape cultivation extension brought setbacks, such as a negative fingerprint on the environment, and the emergence of harmful pathogens and pests (i.e., blackleg, clubroot, broomrape, insects, etc.), endangering and compromising oilseed rape production. However, although taxonomic closeness with *Arabidopsis thaliana*, the diversity, regulation, and genetic determinism of the specialized metabolism in *Brassica napus* is poorly understood. This lack of effort limits scientific understanding of key determinants of biochemical and agroecological processes affecting the tolerance to biotic constraints in this crop. This was the main challenge we set ourselves during this thesis work.

This manuscript is built according to an "articles-based theses" format. It is structured around four chapters which first one will be devoted to the bibliographic introduction.

Chapter I: Bibliographic introduction

Introduction

For more than two centuries, scientific research around metabolism has shaped the perception of biologists across time, leading them to distinguish metabolites into three major classes based on approximations of their functions: primary metabolites, secondary/specialized metabolites (SM), and regulating metabolites like hormones or vitamins (Wink, 2010). Moreover, the terms "goal-directed metabolites" or "chemical defensive means" were suggested by the pioneers of chemical ecology (Anton Kerner von Marilaun, Léo Errera, and Ernst Stahl). In particular, Stahl uses these terms to designate, among other things, "the chemical defenses that plants produce to protect themselves" (Hartmann, 2007; Hartmann, 2008). However, for a long time, this idea was firmly rejected by phytochemists and plant physiologists, who considered these compounds as "metabolic wastes" or "detoxification products". It was only at the end of the XXth century-beginning of the XXIst that the functional studies focused on the role of the plant-specialized metabolism in the communication with other plant and nonplant species started (such as allelopathy and legume-rhizobia interactions, defense against pathogens, herbivores, and beneficial interactions with insects). Results obtained over the last two decades clearly highlight the function of specialized metabolites as mediators between plants and the environment (Weng et al. 2021).

Among the pressing questions arising from the field of specialized metabolism and chemical ecology are the assession of their structural and biosynthetic diversity, their adaptive significance, and the identification of the genetic determinants controlling their content. The structural investigation of the compounds involved in both abiotic and biotic interactions has allowed the inventory, for example, of 57,000 metabolites entries in the KNApSAcK database (Afendi et al., 2012). The expected number of metabolites in the plant kingdom is 200,000 to one million (Dixon and Strack, 2003; Rai et al., 2017). However, despite the significant scientific efforts attested by abundant literature in this field, it is rare to conclusively identify more than two hundred metabolites in a single taxon while the contents are probably much higher.

Due to their sessile nature, plants modify their morphology, physiology, and metabolism to acclimatize and adapt to their dynamic environment. While some subclasses of SM appear widely present in the plant kingdom (e.g., terpenoids and phenolics, including flavonoids), others appear restricted to some taxonomic groups (e.g., glucosinolates, cyanogenic glycosides, and alkaloids), indicating both ancestral and recent metabolic pathways. Recent advances in whole-genome sequencing suggest that all land plants have duplicated genomes, the fate of which several factors, including environmental stress, dictate enzymatic gene copies. Some biosynthetic similarities (enzymes, chemical structure, and chemical function) may appear independently in distant taxa, raising huge evolutionary questions about plant-specialized metabolism.

In this chapter, by focusing on specific examples, I will briefly review the metabolism so-called specialized of plants, focusing on the metabolism of the polyploid species under study (Section I.A.) before introducing the thesis objectives (Section I.B.).

I.A. The specialized metabolism in vascular plants

Firstly, is reported the progress in understanding the roles of metabolites, especially those considered specialized, in-plant/environment interactions (section I.A.1). Secondly, is provided a global overview of the structural and functional diversity of the

SM in plants, including metabolic pathways and specific examples of enzymes that govern the structural diversity of metabolites universally distributed in land plants or restricted to some taxa (section I.A.2). Factors shaping the SM of plants are discussed in section I.A.3 before presenting some approaches used to investigate the specialized metabolism (section I.A.4). Next, will be demonstrated how SM contribute to plants' immune systems in response to bioaggressors (section I.A.5). The biochemical and genetic basis of specialized metabolites in *Brassica napus* are finally presented in section I.A.6.

I.A.1. The paradigm of the multifunctionality of specialized plant metabolism

For the following, numbers in bold after each metabolite name correspond to the number of the illustration of the structures of compounds 01 to 06 shown in Figure 1. These numbering will be used and incremented throughout this chapter.

Historically, the distinction between primary and secondary (or specialized) metabolites was based on their functions: Primary metabolites are crucial to the immediate growth and development of plants when natural (secondary or specialized) products are required to interact with the biotic and abiotic factors of environment (Hartmann, 2007). Besides, phytohormones or vitamins regulate the growth, development, and reproductive processes (Van Overbeek, 1966; Davies, 2015). However, more recently, some evidence that compounds that have been classically assigned as primary or secondary metabolites can also play roles in plant cell signaling, thus suggesting that the rough distinction into three distinct classes of compounds might be obsolete. This has led to consider the "multifunctionality" of the metabolism, thus marking a new historical landmark in the functions of plant

metabolism (Návarová et al., 2013; Erb and Kliebenstein, 2020; Durán-Medina et al., 2021; Li et al., 2021). Some nitrogen-containing organic acid compounds encompassed primary or specialized roles, even phytohormones, e.g., L-proline (01), pipecolic acid (02), and heteroauxin (03) (Figure 1). Thus, pipecolic acid (02), an alkaloid analog to proteinogenic amino acid L-proline (01)), is a cyclic non-proteinogenic amino acid derived from L-lysine catabolism, whereas heteroauxin (03) derived from L-tryptophan is a phytohormone alkaloid. Similarly, diterpenes (C20) often encompassed the three functional categories (primary, secondary, and hormones). For instance, kaurenoic acid (04), abietic acid (05), and gibberellin A12 (GA12) (06) are diterpenoid carboxylic acids (Figure 1).



Figure 1: Similarities between the chemical structures of phytohormones and primary and specialized metabolites. L-proline (01), pipecolic acid (02), heteroauxin (IAA: indole-3-acetic acid) (03), ent-kaurenoic acid (04), abietic acid (05), and gibberellin A12 (06). Proline (01) and kaurenoic acid (04) are admitted as primary metabolites, while they are structural neighbors of the specialized metabolites: pipecolic acid (02) and abietic acid (05). Gibberellins A12 (06) and heterauxin (03) are phytohormones (adapted from Croteau et al., 2019).

These three compounds (04, 05, 06) exhibit close structures and derive from geranylgeranyl pyrophosphate. Abietic acid (05) has been described as a significant component of the secreted resin and a defense phytochemical against insect and pathogen attacks found in Fabaceae, Pinaceae, and other conifer species (Funk and Croteau, 1994). Ent-kaurenoic acid (04) is also oxidized into the phytohormone GA12 (06) by an NADPH-dependent, membrane-bound CYP701A (ent-kaurene oxidase) (Helliwell et al., 2001). These structural overlaps of metabolite types (primary, specialized, and phytohormone) illustrate the blurred boundaries when applying a functional classification. Moreover, current research suggests that this functional separation is becoming obsolete such as the works characterizing for instance the accumulation of the proteinogenic aminoacid proline (01) in the context of incompatible interaction Arabidopsis thaliana-bacteria (Fabro et al., 2004) and those detailing the contribution of hormone signalization (heteroauxin (03), JA, ethylene) to the plant defense against the attacks of pathogens and herbivores (So et al., 2006; Chen et al., 2007; Wang et al., 2007; Grant and Jones, 2009). Similarly, GC/MS-based profiling of mutants of A. thaliana revealed how pipecolic acid (Pip) (02) orchestrates systemic immunity (Návarová et al., 2013) (Figure 2). Pip is essential to establishing the local and systemic immune response (SAR¹) to pathogens. Indeed, the accumulation and the signaling contribution of Pip in the distal part of the infected leaf site of A. thaliana by Pseudomonas syringae have been pointed out. These few examples represent some breakthroughs in our understanding of the interconnection between specialized metabolites and hormones in plant defense and primary metabolites. Besides, several

¹ SAR is an integral immune-activating pathway of plants, also called systemic immune response of plants, triggered by localized infection by pathogens from primary (local) tissues to uninfected secondary (distal) tissues.

reviews have highlighted a more direct contribution of primary metabolism in plant defense responses against bioaggressors (Drew and Demain, 1977; Bolton, 2009; Morkunas and Ratajczak, 2014; Rojas et al., 2014).



Figure 2: Proposed model for the role of Pipecolic acid (Pip (02)) during activation of local resistance, systemic acquired resistance, and defense priming in *Arabidopsis thaliana.* **1°, pathogen-inoculated leaf; 2°, distal leaf. Leaf infection by** *P. syringae* **induces Pip production through lysine catabolism. (1) Pip activates the amplification and defense priming and effectively synthesizes the phytoalexin camalexin. (2) Pip activates the feedback amplification cycle facilitating the SAR response in the distal leaves, involving the increased Pip biosynthesis, defense hormones salicylic acid (S.A.), and overexpression of defense enzymes (PAD4, ISC1, and ALD1). PAD4: phytoalexins-deficient4, ISC1: isochorismate synthase1, ALD1: AGD2-like defense response protein (Návarová et al., 2013).**

Recently, Erb and Kliebenstein (2020) demonstrated that all the metabolites (primary, specialized, and hormone) would be endowed with functional versatility (growth, development, defense, and regulation processes), which breaks the functional trichotomy paradigm. This functional versatility suggests that SMs interact with other

Ph D. manuscript | Anani Amegan MISSINOU

classes of compounds (Figure 3). This new concept stimulates scientific thought and provides some working hypotheses for future research on the ecological roles of specialized metabolites. The questions that arise now are, for example: how do the metabolites that perform regulatory functions work? Do they act by binding to protein receptors? or by altering specific non-receptor proteins? or through general autotoxicity/reactivity? (Erb and Kliebenstein, 2020). Other authors also pointed out the functional overlap between primary/specialized metabolites and hormones, thus supporting functional multiplicity, which is becoming a norm rather than an exception (Durán-Medina et al., 2021; Li et al., 2021).



Figure 3: A New paradigm of the functional versatility/polyvalence of the plant metabolites. The overlapping zones illustrate the inter-regulation of phytohormone, primary, and specialized metabolites. (Erb and Kliebenstein, 2020).

I.A.2. Structural and biosynthetic diversity of the plant-specialized metabolites

I.A.2.1. Structural diversity of plant-specialized metabolites

The structural patterns of SMs can be divided into major classes: alkaloids, cyanogenic glycosides, glucosinolates, tannins, coumarins, quinones, phenylpropanoids, flavonoids, polyketides, terpenoids, and benzoxazinoids. These categories of compounds typically display strong biological activities that largely contribute to the structuration of plants used by humans for food and non-food applications. The biosynthesis of terpenoids, alkaloids, phenolics, quinones, and polyketides occurs in the whole plant kingdom. Terpenoids, alkaloids, and phenolics regroup the highest number of already described molecules with \geq 25,000 structures for terpenoids, \geq 21,000 structures for alkaloids, and \geq 10,000 structures for phenolic compounds (Wink, 2010; Christianson, 2017).



Figure 4: Core structural diversity of specialized metabolites derived from the central metabolism in the plant kingdom. Photosynthesis (center), central metabolism (2nd circumscribed layer), tailored enzymes-catalyzed modifications (3rd circumscribed layer), specialized metabolites (4th circumscribed layer). Structure of <u>terpenoids</u>: 5,9-epi-penstemoside (07), 2-ketoepimanool (08), capsidiol (09), lubimin (10), oryzalide A (11), artemisinin (12), kauralexin B2 (13), β-caryophyllene (14); <u>phenolic acids</u>: gallic acid (15), sinapoyl-O-glucose (16); <u>stilbenoids</u>: trans-resveratrol; <u>coumarins</u>: (17), scopolin (18), isopimpinellin (19); <u>flavonoids</u>: 3-deoxyanthocyanidin (20), kaempferol-3-O-glucose (21); <u>lignans</u>: pinoresinol-glucose (22), lariciresinol (23); <u>tannins</u>: gallotanin (24); <u>alkaloids</u>: nicotine (25), camalexin (26), spirobrassinin (27), 6-GlcO-ICOOGlc (28), caffeine (29), lumazine (30), gramine (31), anagyrine (32), <u>benzoxazinoids</u>: BOA-glucose (33), DIBOA-glucose (34), DIMBOA-glucose (35), HDMBOA-glucose (36), HMBOA-glucose (37), <u>cyanogenic glycosides</u>: lotaustralin (38), linamarin (39), amydalin (40), dhurrin (41), <u>glucosinolates</u>: neoglucobrassicin (42), (epi)glucobarbarin (43), (epi)progoitrin (44) are displayed. Adapted from (Hartmann, 2007; Wang et al., 2019).

Terpenoids (TPNs), so-called terpenes or isoprenoids, are ubiquitous in the plant's kingdom. However, some of the TPN subgroups (mono- (C10), sesqui- (C15), di-(C20), tri- (C30), and tetraterpenoids (C40)) have been widely used for chemotaxonomic studies due to the restriction of specific molecules to specific taxons. For example, in *Lamiaceae*², glycosylated monoterpenes (C10) (called iridoids), such as 5,9-epi-penstemoside **(07)** (Figure 4), are specific to certain plant species. This allows subdividing the *Lamiaceae* family (3500 species) into two sub-families: the *Nepetoideae* members (that are potent producers of essential oils and are lacking iridoids), and the *Lamioideae*, which are weak producers of essential oil but have a high content of 5,9-epi-penstemoside **(7)** (Delazar et al., 2004) (Figure 4).

Most sesquiterpenes and diterpenes have antimicrobial activity, e.g., 2-ketoepimanool (08), capsidiol (09), and lubimin (10) in tobacco *Nicotiana spp* (Cohen et al., 1983; Threlfall and Whitehead, 1988; Ahmed et al., 2000), oryzalide A (11) in rice (Kono et al., 2004), artemisinin (12) in *Artemisia annua* (Klayman, 1985; Li and Wu, 2010), kauralexin B2 (13) in maize, and β -caryophyllene (14) in *A. thaliana* (Schmelz et al., 2011; Poloni et al., 2014). Triterpenoids (C30) include several classes of compounds, many of which have potent deterrent and antifeedant effects on herbivores and insects (e.g., phytosterols, saponins, sapogenins, and pentacyclic triterpenes) (Kuzina et al., 2009; González-Coloma et al., 2011; Thimmappa et al., 2014). Finally, tetraterpenoids (C40) are carotenoids that determine the red color of flower petals, and polyterpenoids that are non-polar tackifiers, such as rubber.

² Lamiaceae family, formerly the mint family, with 230/7000 genera/species, is one of the most investigated families of angiosperms (Zhao et al., 2021).

Ph D. manuscript | Anani Amegan MISSINOU

Phenolic compounds (PNLs) are among the molecules amply used in chemotaxonomic studies. They are divided into different classes, from simple C6-C1-phenolics (phenolic acids³, e.g., gallic acid, **15**), C6-C3-phenolics (including esters of cinnamic acids, or also coumarins, e.g., scopoletin, **18** and isopimpinellin **19**), to more complex structures called polyphenols (flavonoids⁴, stilbenoids (trans-resveratrol, **17**), lignans⁵, tannins⁶) (Figure 4). Cinnamic acid derivatives (with acyl groups) are crucial as metabolic crossroads for building phenolic compounds. Among flavonoids, some have no chemical modification (e.g., 3-deoxyanthocyanidin, **20**), while others can be acylated by one or more glycosylations (Figure 4). The glycosylation of flavonoids is favored in 3-OH (e.g., kaempferol-3-O-glucose, **21**) and 7-OH groups of the flavone aglycone skeleton (Figure 4).

Some plant SMs contain one or more nitrogen atoms in their carbon skeletons. The most important classes are alkaloids, cyanogenic glycosides, benzoxazinoids, and glucosinolates (sulfur-containing compounds) (Figure 4). Globally, the nitrogen atom is derived from the amine function of primary amino acids.

Alkaloids are nitrogenous compounds with basic properties, most often harboring a heterocyclic ring system (Figure 4), and they are reported in more than 150 plant families (Yamane et al., 2010). These compounds are synthesized from the decarboxylation of amino acids or transamination followed by structural modifications and are classed according to the amino acid (often nucleic acids) precursor and/or

³ Phenolic acids: e.g., (hydro)benzoic, salicyclic, caffeic, coumaric, ferulic, sinapic, syringic, m/p-cinnamic, coumaric gallic, protocatechuic, gentisic, vanillic acids and etc.

 ⁴ Flavonoids skeleton consists of rings A, B, and C of the C6-C3-C6 skeleton of flavone, on which variable substitutions (ketone, hydroxy, O-methyl) allow their classification into main classes: flavonols, flavone, isoflavone, flavan-3-ol, flavanone, anthocyanidin.
⁵ Lignans are a large group of low molecular weight polyphenols that are precursors to phytoestrogens.

⁶ Tannins contain three or more phenol subunits subdivided into tannins and their polymers, hydrolyzable or non-hydrolyzable.

according to their chemical skeleton (e.g., the indole skeleton of tryptophan in camalexin (26), spirobrassinin (27), and 6-GlcO-ICOOGlc⁷ (28)). As a result, some alkaloids with similar skeletons, or even the same precursors, can come from different biosynthetic pathways (e.g., the two phytoalexins camalexin (26) and 6-GlcO-ICOOGlc (28)) or may present contrasting biological activities (e.g., the hormone IAA (03) and the phytoalexin 6-GlcO-ICOOGlc (28)) (Böttcher et al., 2014). Thus, based on their underlying heterocyclic nucleus and biosynthetic precursor, alkaloids can be grouped into pyrrole, pyridine, piperidine, purine, pyrimidine, imidazole, and indole, among others ⁸ (Ponikvar-Svet et al., 2021).

Benzoxazinoids (BXZ) are tryptophan-derived specialized metabolites of the *Poaceae* family, including wheat, rye, maize, and some dicots. They are often methylated and glycosylated which monoglycosylated benzoxazinoids being the most frequent. Some unusual di-, tri-, tetra-hexoxylated, and acetyl-hexosylated derivatives have recently been reported (De Bruijn et al., 2016; Pihlava and Kurtelius, 2016). These SMs are effective in plant defense against pathogens (Villagrasa et al., 2006; Macías et al., 2009; Tanwir et al., 2013). Non-glycosylated benzoxazinoids are toxic, and plants or insects detoxify them using various mechanisms such as hydroxylation and O- or N-glycosylation (de Bruijn et al., 2018). These inactive conjugated forms stored in vacuoles and commonly termed phytoanticipins are enzymatically released when needed to respond to a microbial attack or other environmental stimuli (Cambier et al., 1999; Niemeyer, 2009). They can be divided into three structural classes: (A) benzoxazolinone glycosides, for example, 1,3-benzoxazol-2-one (BOA-glucose, **33**).

Ph D. manuscript | Anani Amegan MISSINOU

⁷ 6-GlcO-ICOOGlc stand for b-D-glucosyl 6-hydroxyindole- 3-carboxylate 6-O-b-D-glucoside.

⁸ Other heterocycle of alkaloids: pyrroline, pyrrolidine, piperideine, pyrrolizidine, triazine, tropane, (benyl)(iso)quinoline.

(B) hydroxamic acids glycoside, for example, 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA-glucose, **34**) and 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3- one (DIMBOA-glucose, **35**); and (C) lactams glycosides, for example, 2-hydroxy-7-methoxy-1,4-benzoxazin-3-one (HMBOA-glucose, **37**) (Figure 4).

Cyanogenic glycosides (CNglcs) are nitrogen- compounds belonging to cyanogenic compounds ⁹ composed of cyanide aglycone (from amino acids) attached to a sugar moiety (mostly β-D-monoglucose), which sometimes gets modified to a second monosaccharide, a sugar ester (e.g., Amygdalin (40)) (Francisco and Pinotti, 2000) (Figure 4). Cyanogenic glycosides have been isolated from 2650 plant species belonging to 130 families, including *Poaceae*, *Fabaceae*, *Euphorbiaceae*, *Rosaceae*, *Asteraceae*, etc. The most commonly reported cyanogenic plant glycosides are lotaustralin (38), linamarin (39), and amygdalin (40) from the *Rosaceae* family and dhurrin (41) of *Poaceae* (Yamane et al., 2010).

Cyanogenic glycosides are phytoanticipins activated into toxic HCN derivatives. The enzymes (O- β -glycoside) catalyze the production of HCNs in different cellular compartments from which the substrates are present (Poulton and Li, 1994). For example, in *Sorghum bicolor*, the enzyme responsible for the hydrolysis of dhurrin **(41)** is present in the mesophyll cells. At the same time, its substrate is stored in the vacuole of the leaf epidermal cells (Nielsen et al., 2016). The HCNs derivatives released in plant cells are detoxified by mitochondrial β -cyanoalanine synthase and rhodanese (i.e., sulfurtransferase) (Machingura et al., 2016).

⁹ Cyanogenic compounds encompass, cyanogenic lipids (from Sapindaceae) cyanohydrin, cyanogenic glycosides

Glucosinolates (GLSs) are nitrogen- sulfur-containing compounds consisting of a β-D-glucopyranose moiety linked to sulfated oxime *via* the isothiocyanate group and variable aglycone. GLSs can be phenylalkyl, indolic, aliphatic linear, or branched/linear derivatives of amino acid precursors such as phenylalanine, tryptophan, methionine, or branched-chain amino acids, respectively (Fahey et al., 2001; Blažević et al., 2020; Missinou et al., 2022). GLSs are found mainly in the order of Brassicales, whose evolution and radiation began 103 million years ago (Cardinal-McTeague et al., 2016) (Figure 4). GLSs are hydrolyzed by β-thioglucosidases, also known as myrosinases. Like benzoxazinoids, myrosinase, and glucosinolates are stored in separate subcellular compartments and come together only under conditions of tissue injury (e.g., cell disruption due to biting or chewing by herbivores), and the resulting products (nitriles, epithionitrile, thiocyanates, isothiocyanates, oxazolidinethiones) are bioactive against generalist herbivores.

The contribution of these antimicrobial compounds to plant immunity will be described in section I.A.5.

I.A.2.2. Metabolic origin of plant specialized metabolites

All SM categories are derived from the primary metabolism through a series of key enzymes. Geranyl diphosphate synthase (GPPS), geranylgeranyl diphosphate synthase (GGPPS), farnesyl pyrophosphate synthase (FPPS), and chalcone synthase (CHS) are essentially ubiquitous enzymes in higher plants. The active forms of GGPPS, FPPS PAL/TAL, and CHS are the first committed enzyme in phenylpropanoids and flavonoid biosynthesis (Figures 5.A and 5.B). Phenylalanine ammonia-lyase (PAL), tyrosine ammonia-lyase (TAL), and chalcone synthase (CHS)

Ph D. manuscript | Anani Amegan MISSINOU

are the first committed enzymes in the pathway that diverts the central flux of carbon from the primary metabolism (i.e., phenylalanine et tyrosine) to initiate the synthesis of various phenolics compounds (Figure 5). These series of enzymes, at the interface between primary and specialized metabolism, are conserved in the plant kingdom.

The resulting products of these reactions become the fundamental precursors and biochemical scaffolds of the whole phytochemical diversity by entering into more specific biosynthetic pathways involving taxon-specific enzymes from SM-related large multigenic families. These notably include terpene synthases, polyketide synthases for the biosynthesis of the diversity of SM 'molecular scaffolds', and for example, cytochrome P450s, glycosyltransferases, methyltransferases, or flavin monooxygenases, often involved in the subsequent 'decoration' or 'tailoring' of SMs which further contribute to their structural and functional diversification (Anarat-Cappillino and Sattely, 2014)



Figure 5: Simplified biosynthesis route showing the unexhaustive origin of most plantspecialized metabolites from (A) Calvin cycle and (B) glycolysis. Some specialized metabolites classes are derived from several amino acid precursors, such as glucosinolate subclasses. PEP: phosphoenolpyruvate, Phe: phenylalanine, Tyr: tyrosine, Trp: tryptophan, TDC: tryptophan decarboxylase, PAL: phenylalanine ammonia-lyase, TAL: tyrosine ammonia-lyase, CHS: chalcone synthase, STS: strictosidine synthases, DMAPP: dimethylallyl diphosphate, IPP: isopentenyl diphosphate, GPPS: geranyl diphosphate synthase, GGPPS: geranylgeranyl diphosphate synthase, FPPS: farnesyl pyrophosphate synthase. The mevalonate-independent pathway for isoprenoid biosynthesis is not represented. The figure was adapted from Tissier et al., 2015.

I.A.2.3. Biosynthetic diversity of plant-specialized metabolites

The metabolic grid concept ¹⁰ is based on a theory proposed in the 1990s, stating that the enzymes catalyzing a series of transformations are not substrate-specific (Bu'lock, 1973). Today, this enzymatic property referred to as enzyme-substrate promiscuity (discussed in section I.A.3.1.3.), displayed by some secondary metabolism enzymes, is one of the drivers providing varied chemical structures from a conserved limited number of primary precursors.





The different routes that metabolic precursors can take are illustrated in Figure 6. In this schema, six enzymes (numbered E1 to E6) could theoretically be involved in the biosynthesis of 27 different products from a single precursor depending on their relative expression level (Figure 6). The relative amounts of each metabolite in the grid will

¹⁰ Bu'Lock has largely contributed to the development of the metabolic grid concept in relation to the biosynthesis of aromatic polyketides in *Penicillium urtice* (Bu'lock, 1973).

depend on the relative level of expression of the different enzymes, the availability of these enzymes in the various tissues or cellular compartments, and the subsequent modification or transport of the metabolites. Additionally, enzyme isoforms can coexist with slightly different affinities for molecules within a set of related metabolites (Tissier et al., 2015).

This model makes it possible to speculate on uncharacterized metabolic pathways or even produce a distinct set of compounds with specific biological functions by metabolic engineering. This flexible pathway architecture has allowed the characterization of the biosynthetic pathways of the *Leguminosae* quinolizidine alkaloids (e.g., anagyrine, **32**) (Nowacki and Waller, 1975), isoflavones (Barz and Welle Roland, 1992; Barz and Mackenbrock, 1994), and phenolic glycosides from phenylpropanoid and benzenoid precursors (Babst et al., 2010).

I.A.3. Factors involved in the evolutionary diversification of plant specialized metabolites

The evolutionary diversification of plant SM relies, on one hand, on a series of genomic processes involved in the diversification of SM biosynthetic enzymes and, on the other hand, on the driving selection exerted by biotic and abiotic environmental pressures.

I.A.3.1. Genomic processes involved in the diversification of SM-related biosynthetic enzymes

The genomes of terrestrial land plants are estimated to contain between 15 and 28% of genes coding for biosynthetic enzymes involved in the biosynthesis of specialized metabolites (Schläpfer et al., 2017; Moore et al., 2019). Many mechanisms, including

gene duplication, enzyme promiscuity, changes in substrate specificity/affinity, and gene expression modifications, can affect the diversity and the content of specialized metabolites.

I.A.3.1.1. Gene duplication

Polyploidy, a unique whole-genome duplication mechanism providing organisms with more than two complete sets of homologous chromosomes is favorable to plant species diversification because duplicated genes can undergo distinct evolutionary fates. It is today accepted that all the higher plants are paleopolyploids because they descend from an ancient polyploid ancestor. Depending on the direction and intensity of evolutionary forces operating after duplication, the resulting gene copies (paralogs) may suffer distinct fates. One of the paralogous copies can maintain the parental function; other paralogs are redundant and could undergo purifying selection, which would promote neutral evolution and the accumulation of random mutations, if deleterious, resulting in a dysfunctional copy (pseudogenization); eventually, positive selection would fix the daughter paralogue in the population. In contrast, random mutations can alter different cis-regulatory and protein-coding modules, thus modifying parental function (sub-functionalization). Occasionally, if these random mutations are advantageous, the paralog obtain а functional innovation can new (neofunctionalization) (Figure 7).



Figure 7: Potential fates of duplicated genes. (a) One paralog becomes a non-functional pseudogene by acquiring deleterious mutations, while the other retains the ancestral functions. (b) Both genes acquire deleterious mutations in cis-regulatory and protein-coding modules, thus sub-functionalizing the ancestral gene function. (c) One paralog acquires a novel advantageous mutation (orange) that can lead to evolutionary innovation in plant form, whereas the other paralog retains the ancestral functions (Preston, 2016).

I.A.3.1.2. Convergent evolution

Convergent evolution may be the independent occurrence of similar or identical features in unrelated (more or less distant lineages) organisms. In specialized plant metabolism, multiple cases of convergent evolution have been reported. Two main classes of convergent evolution may be distinguished: (A) the compounds obtained are structurally very different but are involved in identical ecological functions, and (B) the compounds or reactions produced are identical but obtained by distinct evolutionary enzymes (Pichersky and Lewinsohn, 2011).

Concerning the first class, a surprising co-occurrence of cyanogenic glycosides and glucosinolates has been reported in *Carica papaya* (*Caricaceae,* Brassicales) (Williams et al., 2013) and of indolic- CNglc and indolic-GLS in *A. thaliana* (Rajniak et al., 2017). This co-occurrence of cyanogenic glycosides and glucosinolates for critical

Ph D. manuscript | Anani Amegan MISSINOU

enzymes such as CYP82C2 (indolic-CNglc) and CYP79B2/3 (of indolic-GLSs) can be traced back to ancestral plant forms.



Figure 8: Independently evolved O-methylation of benzoxazinoids and glucosinolates regulates defense and resistance (Li et al., 2018).

Other examples include the couples of anthocyanins/betacyanins that play essential roles in pollination (Brockington et al., 2011) and indolic-GLSs/benzoxazinoids in insect defense (Figure 8) (Li et al., 2018). This is also observed in *Brassica* pathogens such as *Leptosphaeria maculans,* which produce sulfur metabolites like its host, making blur/confound attributions in metabolomics studies. Similarly, cyanogenic glycosides (CNglcs) and glucosinolates (GLSs) are both activated by plant enzymes in response to tissue damage, producing cyanide and thiocyanates derivatives that target host-plant protection against insect feeding (Morrissey and Osbourn, 1999). A chemical

convergence between plants and insects is associated with plants producing phenyl-CNglc and benzyl-GLS (Beran et al., 2019).

An example of the second type of convergence evolution is the case of caffeine **(29)**, whose biosynthetic pathways (nearly 12 potential pathways) have evolved independently in many quite distant species, such as the lineage of coffee (*Coffea arabica* (Gentianales), tea (*Camellia sinensis*, Ericales), Citrus spp, cacao (*Theobroma cacao*, Malvaceae) and Guarana (*Paullinia cupana*) (Huang et al., 2016).

I.A.3.1.3. Enzyme promiscuity

The promiscuity of an enzyme is the ability to catalyze unexpected reactions, so-called promiscuous reactions distinct from predictable (primary) reactions; this promiscuity is due to nucleotide mutations impacting one or fewer amino acids. Two main promiscuities can be distinguished for enzyme specificities: substrate promiscuity and product promiscuity. The ability of an enzyme to catalyze different products from a single substrate refers to the **enzyme-product promiscuity**, such as aureusidin synthase of snapdragon (*Antirrhinum majus*), which provides unequivocal oxidation of polyphenol, giving the yellow color (Kusumi et al., 2000) or even the oxidative rearrangement of (+)-sesamin by sesaminol synthase (CYP91C) in sesame seeds (*Sesamum indicum*) (Murata et al., 2017). By contrast, **enzyme-substrate promiscuity** is an ability of an enzyme to use multiple substrates with relaxed or broad substrate specificity, which is valuable for environmental adaptation. For example, 4-coumarate-CoA-ligase, which can catalyze the binding of coenzyme A on different hydroxycinnamate substrates (e.g., coumaric acid, caffeic acid, ferulic acid, etc.), and SABATH methyltransferases, which catalyze methyl group transfer to functional
groups: hydroxyl and sulfhydryl, so including broad aromatic and aliphatic thiols substrates, (Zhao et al., 2012; Leong and Last, 2017). Similarly, the enzymes coded by paralogs flavonoid-3'-hydroxylase and flavonoid-3,5-hydroxylase have broad substrate specificity subject to hydroxylation of the B-ring of hydroxylated flavanones, dihydroflavonols, flavonols, and flavones determining red cyanidin-/blue delphinidin-based anthocyanins flower of various plant species (Castellarin et al., 2006; Tanaka and Brugliera, 2013). Furthermore, the substrate and donor of flavonoid-3-O-glycosyltransferase from *Scutellaria baicalensis* (Sb3GT1) could accept five types of sugars like glucose, galactose, N-acetylglucosamine, rhamnose, xylose, arabinose brought by uridine 5'-diphosphate (UDP)-sugar as donors (Figure 9) (Wang-Zilong et al., 2019).



Figure 9: Substrate and donor promiscuity of Sb3GT1 from *Scutellaria baicalensis.* 1a-1e: type of substrate. Glc: glucose, Gal: galactose, GlcNAs: N-acetylglucosamine, Xyl: xylose, Ara: arabinose (Wang-Zilong et al., 2019).

I.A.3.2. Environmental drivers of plant SM diversification over geologic timescale

Establishing the environmental factors that participated in the diversification of specialized metabolites is fundamental to understanding how these metabolic

pathways co-evolved with environmental variation. Indeed, the taxonomic distribution of major plant-specific metabolic pathways suggests a very early diversification of central biosynthetic blocks since the adaptation of plants to the terrestrial environment (Figure 10). The biogeochemical-based models estimate that the land plants (e.g., charophytes) emerged in the middle of the Cambrian and early Ordovician eras (~512-443 million years ago). Weng et al. (2014) reported that the closest primitive plant form to Charophytes (green algae, *Chlamydomonas reinhardtii*) already had important enzyme families known to participate in specialized plant metabolism and primary metabolism. These enzymes potentially contributed to redox homeostasis and the emergence of natural antioxidants in the aquatic environment of the early Ordovician age era (~500 million years ago) (Weng, 2014).



Figure 10: Origin and expression of specialized metabolites during green plant phylogeny lineage evolution (Delgoda and Murray, 2017).

Ph D. manuscript | Anani Amegan MISSINOU

For instance, the core biosynthesis pathway of phenylpropanoids and flavonoids probably emerged in response to UV-induced stress. The clade of vascular plants appeared in the late Ordovician and Devonian interval era (~485-358 million years ago); at the end of the Paleozoic period, these vascularization preconditioned plant body innovations (e.g., leaves, trichomes, roots), associated with the biosynthesis of terpenes, phenolics, and alkaloids as antiherbivore defense (Labandeira et al., 2014; Delgoda and Murray, 2017). The seeds appear in the early Mesozoic with the expression of tannins that protect them from biotic and abiotic stresses (Figure 10). The appearance of fruits is more recent in the middle of the Cenozoic era (~40 million years), following the production of volatile compounds necessary for the attraction of pollinators (Delgoda and Murray, 2017).

I.A.4. Factors involved in the regulation of plant-specialized metabolite biosynthesis

Specialized metabolites typically accumulate in concentrations ranging from nmol to tens of µmol.g⁻¹ DW, and usually are restricted to the cellular production site. Sometimes, they are stored in vacuoles under glycosylated forms, often in specialized secretory structures such as trichomes, laticifers, ducts, and canals. Two types of factors condition the metabolite content: intrinsic (genetic, ontogenic, and morphogenetic) and environmental or extrinsic factors (temperature, salinity, drought, UV, altitude, injuries, and nutrient deficiencies)

I.A.4.1. Transcription factors (TF)

Transcription factors affect the biosynthesis of specialized metabolites by regulating gene expression. For example, *MYC2* transcripts are accumulated in leaves of *Nicotiana attenuate* in response to herbivore attack (*Manduca sexta*); this TF regulates the expression of nicotine and phenolamides, jasmonates, and some JA-/NaCOI1-dependent metabolites (e.g., 17-hydroxy-geranyl-linalool diterpene glycosides) (Woldemariam et al., 2013). Many other studies report the regulation of many biosynthetic pathways such as alkaloids, terpenoids, glucosinolates, and saponins by TFs belonging to AP2/ERF, WRKY, bHLH, bZIP, and NAC FTs families (Figure 11) in different plant species such as *A. thaliana*, *Artemisia annua*, *Vitis vinifera*, *Salvia sclarea*, *Withania somnifera*, *Scutellaria baicalensis*, *Camellia sinensis*, *Medicago truncatula*, etc. (Jan et al., 2021).



Figure 11: Conserved structures and domains of transcription factors regulate plant stress responses through mediating specialized metabolism. The AP2/ERF TF family (APETALA2 (AP2)/Ethylene Responsive Element Binding Factor (EREB)) has a domain of 60 amino acids involved in DNA binding and structurally contains three β -sheets before the α -helix motif. WRKY TF family transcription factor includes a 60-amino-acid well-conserved in the WRKY domain composed of a conserved WRKYGQK motif (N-terminal) and a zinc-finger motif at the C-terminal. bHLH TF (basic helix-

loop-helix) 60 conserved amino acid residues and consist of bipartite to the basic region and helix-loophelix region. The members of the bZIP TF family comprise a conserved leucine zipper and a positively charged DNA binding site. Divided into four groups, 1R-, R2R3-, 3R-, and 4RMYB, the MyB TF R2/R3 family of MYB TF significantly regulates plant-specialized metabolite pathways. NAC TF family has five subdomains (A-E), with A necessary for dimerization, B and E responsible for functional diversity, and C and D highly conserved and accountable for DNA binding (Meraj et al., 2020; Jan et al., 2021).

AP2/ERF TFs family members bind to the jasmonate- and elicitor-responsive element (JERE) in the promoter site of genes involved in the biosynthesis of terpenoid indole alkaloids, steroidal glycoalkaloids, nicotine, saponins, taxol, and resveratrol following biotic stress (Van Der Fits and Memelink, 2000; Meraj et al., 2020; Jan et al., 2021) (Figure 10). WRKY TFs regulate the expression of the genes involved in the biosynthetic pathway responsible for hydroxycinnamic acid amides (HCAAs), alkaloids, terpenoids, terpenes, benzylisoquinoline, and alkaloid production (Meraj et al., 2020; Jan et al., 2021). bHLH TFs are associated with anthocyanin, alkaloids, glucosinolates, diterpenoid phytoalexins, and saponins biosynthesis (Meraj et al., 2020; Jan et al., 2021). bZIP TFs are involved in stress responses to ROS, osmotic, nutrient, drug, and pathogens stresses by regulating anthocyanin, tanshinone, artemisinin, terpenoid phytoalexins production (Meraj et al., 2020; Jan et al., 2021). MYB TFs regulate genes involved in the biosynthesis of flavonoids, HCAAs, and proanthocyanins in response to attacks by fungi and herbivores (Meraj et al., 2020; Jan et al., 2021) (Figure 12). Finally, the biotic and abiotic stress-responsive NAC TFs are involved in the biosynthetic pathway of flavonoid, putrescine, and camalexin (Meraj et al., 2020; Jan et al., 2021).



Figure 12: Overexpression of the MYB transcription factor gene *PAP1* leads to enhanced phenylpropanoid production in *Arabidopsis thaliana* (left) and *Nicotiana tabacum cv xanthi* (right). Flowers on 12-week-old plants control transformants (A and C) and transformants overexpressing *Pap1* (D and B) (Borevitz et al., 2000).

I.A.4.2. Ontogenic factors

Some chemical traits may vary or even be specific to some plant developmental stages: seeds, seedlings, and juvenile or mature reproductive individuals. Plant ontogeny can influence the composition and content of specialized metabolites in the whole organism and/or in specific organs (leaves, roots, stems, glands, laticifers, trichomes,...). The content of SM declines during ontogeny in the case of cyanogenic glycosides (Neilson et al., 2011; Ochoa-Lopez et al., 2019), glucosinolates (Valente Pereira et al., 2002; Palaniswamy et al., 2003; Schreiner et al., 2009; Hanschen and Schreiner, 2017), sesquiterpenes, and triketones (Goodger et al., 2018), alkaloids (Liu et al., 1998; Eunice Karinho-Betancourt et al., 2015). Ontogeny can also affect in opposite ways the content of several plant defense metabolites; for instance, phenolics decrease, and terpenoids increase throughout ontogeny (Goodger et al., 2013). These studies highlighted the role of ontogeny as a source of variation of the SM composition and content, contributing thus to the adaptive value of the plant to its environment. Otherwise, some compounds may exhibit different ontogenic patterns due to the

existence of opposite ontogenic trajectories, as shown in figure 13. This suggests a "Janus-faced" ontogeny impact on specialized metabolism, which prompts careful extrapolation of metabolic observations from the juvenile stage to the reproductive stage of plants.



Figure 13: Prediction of ontogenetic trajectories of some plant defenses (Ochoa-Lopez et al., 2019).

I.A.5. Plant Immunity: Qualitative and quantitative resistance

Depending upon the nature of plant-pathogen interactions (non-host¹¹ and host¹² relationship), plants generally have three main defense mechanisms: non-host resistance (which will not be discussed), qualitative resistance, and quantitative resistance (Figure 14a).

¹¹ Non-host relationship is characterized by an incompatibility interaction between all cultivars of the species and all biotypes of the pathogen agent.

¹² Host relationship is described by a compatible or incompatible interaction conditioned by cultivar or biotype.

I.A.5.1. Qualitative resistance

Qualitative resistance prevents symptom appearance by rapidly stopping the progression and multiplication of pathogens in the host plant tissues. This type of resistance is mainly controlled by major genes¹³ (R genes) (Figure 14b). Of the 314 R genes cloned so far, more than half code for proteins belonging to the NLR (191 genes), Receptor-like protein (RLP), and receptor-like kinase (RLK) (60 genes) families (Kourelis and van der Hoorn, 2018) (Figure 14c,e).



Figure 14: Qualitative resistances and cellular actors underlying the characterized R genes. (a) Mendelian ratio phenotypes of total resistance contrast sharply with susceptible plants and **(b)** monogenic determinism (Corwin and Kliebenstein, 2017), **(c)** perception of 'non-self' by receptors (NBS-LRR), and rapid initiation of the qualitative resistance response through a network implying a small number of co-regulated genes (Roux et al., 2014), **(d)** Qualitative resistance is essentially determined by genes of the RLK family, NBS-LRR, and transcription factors, **(e)** 191 R genes code for NLR or NLR-ID (Integrated Decoy) proteins, involved in the direct recognition or direct influence of intracellular

¹³ Major gene is a gene with variation at one locus explains the entire phenotype

Ph D. manuscript | Anani Amegan MISSINOU

effectors. 60 R genes encode transmembrane proteins of the RLP/RLK type involved indirectly or directly recognizing extracellular effectors (Kourelis and van der Hoorn, 2018).

NLRs are cytosolic proteins involved in the direct or indirect perception of certain effectors ("avirulence proteins") delivered by microbes into plant cells. RLP/RLK are transmembrane proteins with an extracellular domain capable of detecting specific effectors or PAMPs¹⁴ secreted by microorganisms in the apoplast (Figure 14d). This suggests that qualitative resistance is essentially based on the perception of "non-self" and "self-modification by non-self ", which allows resistant plants to quickly initiate effective defense responses to stop the progression of the pathogen (Dangl and Jones, 2001; Cui et al., 2015). Ultimately, activation of *R* genes leads to two types of responses: resistance-related (RR) metabolites (RRMs) and proteins (RRPs). The *R* genes regulate the amount of specialized metabolites (e.g., phenylpropanoids, flavonoids, and some alkaloids) biosynthesized, either by metabolic-type regulators (hormones) or by activating protein-type regulators (transcription factors). These strengthen the cell walls or directly suppress the pathogen depending on their antimicrobial properties (Kushalappa et al., 2016).

However, this resistance response is generally effective against only a limited number of strains (so-called specific resistance). The breakdown of qualitative resistance is usually very rapid after their deployment due to the evolution of the pathogen populations and the inability of a deployed resistant variety (resistant to a specific strain) to recognize the effectors of other strains. This type of resistance has been the most used in plant breeding for the development of resistant varieties during the last century.

¹⁴ Pathogen-associated molecular patterns (PAMPs) are recognized by pattern-recognition receptors (PRRs),

I.A.5.2. Quantitative resistance

Faced with the breakdown of qualitative resistance in several major crops, the breeding process today aims to build varieties with more durable resistance. Indeed, other plant pathosystems display diverse phenotypes or resistance continuously ranging from partially resistant individuals to partially susceptible ones. Genetic analyses of this type of resistance in segregating populations show a continuous distribution of phenotypes and a polygenic control of this trait (Figure 15). This often corresponds to a more complex genetic architecture involving different genetic factors of resistance (Quantitative Trait Loci-QTLs) each with a small effect on the phenotype.



Figure 15: Quantitative resistances and cellular actors underlying the characterized resistance QTLs. (a) continuous distribution of the resistance trait of individuals ranging from partially resistant to partially susceptible **(b)** polygenic determinism involving several minor QTLs contributing to the final phenotype. The red line indicates the significance threshold chosen for the genetic analysis (Corwin and Kliebenstein, 2017). **(c)** the small effect of 'non-self' perception by receptors and multiplicity initiating quantitative resistances are determined by genes coding for enzymes of the metabolism of organic chemical defenses (phytoalexins, antimicrobial peptides/proteins), hormonal signaling, and oxidative stress.

Quantitative resistance (QR) is often effective against a large number of pathogen strains ("broad spectrum" resistance) (Nelson et al., 2018) and appears to be significantly more durable than resistance controlled by major resistance genes. It has

been shown that the combination of quantitative and qualitative resistance could increase the durability of major resistance genes (Delourme et al., 2014; Pilet-Nayel et al., 2017). Genes underlying the resistance QTLs are diverse in functions, ranging from non-self recognition (NLR or RLK) (Diener and Ausubel, 2005; Huard-Chauveau et al., 2013; Hurni et al., 2015; Debieu et al., 2016), signaling-associated protein families, to primary metabolism and very frequently to specialized metabolism. This suggests a non-negligible contribution of specialized metabolism to the quantitative resistance response. For instance, overexpression of biosynthetic genes of GLSs, BnMAM1, BnCYP83A1, and BnUGT74B1 enhanced the resistance of *B. napus* to *Sclerotinia sclerotiorum* (Zhang et al., 2015; Wu et al., 2016). Functional categories of chemical defenses used in a process underlying quantitative resistance are discussed in the following subsection.

I.A.6. Contribution of specialized metabolism to plant immunity

Plants synthesize and store many specialized defense metabolites to protect themselves from foreign bioaggressors. These chemical defenses can be classified according to the control of biosynthesis: **constitutive chemical defenses** and **inducible chemical defenses**.

I.A.6.1. Constitutive chemical defenses

During evolution, plants have established a sophisticated chemical arsenal to prevent pest attacks. This arsenal encompassed both inorganic and organic chemical defenses.

"Metal hyperaccumulating plants" (e.g., the *Brassicaceae* family, which accumulates cadmium) accumulate inorganic elements from the soil such as silicon, nickel, zinc,

cadmium, and lead in vacuoles or trichomes, and these elements contribute to chemical defense against pests. This strategy, so-called elemental defenses or **inorganic chemical defenses**, allows plants to eliminate these harmful/toxic elements from the cellular machinery and to release them only upon cell death or injury, mainly caused by herbivores (Poschenrieder et al., 2006; Boyd, 2012; Cooke and Leishman, 2012).

Organic chemical defenses include constitutive polymers (callose, lignin, suberin, etc.), defense proteins (e.g., ricin), fatty acid derivatives, nonprotein amino acids, and antimicrobial metabolites (phytoanticipins) essentially.

Phytoanticipins¹⁵, including saponins, are an essential source of constitutive antimicrobials, setting chemical barriers against pathogens. They are biosynthesized in healthy plants, accumulated in dead cells, excreted into the external environment (e.g., rhizosphere), or stored in vacuoles in an inactive form even before the pathogen attack (VanEtten et al., 1994; Osbourn, 1996). Phytoanticipins encompass phenolics, nitrogen-containing compounds (alkaloids, glucosinolates, benzoxazinoids, cyanogenic glycosides, polyketides, pyrrolizidine alkaloids), or benzylisoquinolin (Pedras and Yaya, 2015). They also show insecticidal, piscicidal, molluscicidal, allelopathic action, and anti-nutritional effects. Phytoanticipins have soap-like properties (wetting agent with surface-active properties) on the sterols constituting the microbial membrane, which induces pore formation and even cell lysis. The most striking example is avenacin A1 (a glycosylated triterpenoid), which is localized in the

¹⁵ The term of Phytoanticipins coined by John W. Mansfield at a symposium on "Phytoalexin Hypothesis and Beyond" held in Dannenfels, Germany, to honor the 50-year anniversary of the phytoalexin concept and define as « are low molecular weight, antimicrobial compounds that are present in plants before the challenge by microorganisms or are produced after infection solely from preexisting constituents » (VanEtten et al., 1994).

Ph D. manuscript | Anani Amegan MISSINOU

epidermal cells of *Gramineae* roots (Figure 16A) (Osbourn, 1996; Morrissey and Osbourn, 1999) and the α-tomatine (a steroidal glycoalkaloid) of *Solanaceae* species (Roddick, 1977; Keukens et al., 1995). Similarly, tannins of the terpenoid class, 2-ketoepimanool **(08)** of *Nicotiana spp* (Solanaceae), oryzalide **A (11)** of rice, and artemisinin **(12)** of *A. annua*, have been identified as constitutive antimicrobial compounds in leaves where their content increases after fungal infection (Cohen et al., 1983; Kono et al., 2004; Bora and Sharma, 2011).



Figure 16: Localization of antifungal compounds in plants. (A) UV light autofluorescence of the saponin avenacin A1. This compound is localized in the epidermal cell layer of oat roots. **(B)** A resistant sorghum line responding to an attack by *Colletotrichum graminicola*, showing the formation and mobilization of vesicles containing pigmented 3-deoxyanthocyanidin phytoalexins toward the site of attempted penetration by an appressorium (visualized by light microscopy) (Morrissey and Osbourn, 1999).

Many nitrogen-containing compounds have constitutive antimicrobial activity, like gramine alkaloids (**31**) of Solanaceae, quinolizidine alkaloids (anagyrine) (**32**) of Fabaceae, all benzoxazinoids of Poaceae (Freezman, 1997), breakdown product glucosinolates of Brassicales (e.g., neoglucobrassicin (**42**), epiglucobarbarin (**43**), (epi)progoitrin (**44**)). Cyanogenic glycosides¹⁶ (e.g., lotaustralin (**38**), amydalin (**40**))

¹⁶ Cyanogenic glycosides found in *Fabaceae, Rosaceae, Leguminosae, Linaceae, and Compositae*.

release toxic HCN, which inhibits the activity of the cytochrome c oxidase and metalloenzymes of herbivores/pathogens (Gleadow and Møller, 2014).

I.A.6.2. Induced chemical defenses

In response to the plethora of biotic stress factors, plants have developed an innate immune system that detects and/or represses bioagressors. Plant immunity depends on perception, signal transduction, and the defense response (Spoel and Dong, 2012).

The **induced chemical defense** is triggered once the physical defenses have been breached. This starts with the detection of non-self by surface-anchored recognition receptors (PRR¹⁷). These receptors are highly conserved across the plant kingdom and perceive damage/herbivore/microbe/pathogen-associated molecular patterns (DAMPs/HAMPs/MAMPs/PAMPs), then triggering the innate immune response, DAMPs-triggered immunity (DTI), HAMPs-triggered immunity (HTI), (MAMP)-triggered immunity (MTI), and pattern-triggered immunity (PTI) (Spoel and Dong, 2012; Zipfel, 2014). However, some pathogens can introduce or release effector proteins in the host cell to inhibit or degrade host immune factors. In turn, plants rely on versatile symplastic sentinels, namely nucleotide-binding site and leucine-rich repeat receptors (NBS-LRR, also termed NLR), to identify specific pathogen-derived effector proteins and trigger effector-triggered immunity (ETI). These receptors activate signal transduction networks, including phytohormone-dependent defense signaling (salicylic, jasmonic, gibberellic, indole acetic acid derivatives, and ethylene). These hormones regulate plant defense responses through highly interconnected signaling networks (Glazebrook, 2005).

¹⁷ Plant PRRs are either receptor-like kinases (RLKs) or receptor-like proteins (RLPs).

Ph D. manuscript | Anani Amegan MISSINOU

This detection system provides broad-spectrum surveillance of various pathogens and contributes to the evolution of disease resistance specificity across the plant kingdom. Nevertheless, triggered innate immune signaling pathways differ among species.

Research on the integration of growth and defense and their involvement in the induction of antimicrobial defense compounds is currently intense (Chen et al., 2018; Hartmann et al., 2018; Hartmann and Zeier, 2018; Hartmann and Zeier, 2019; Yildiz et al., 2021) because these compounds regulate the expression of genes involved in the biosynthetic pathways of antimicrobial molecules often limited to one species (Ahuja et al., 2012). This synthesis concerns: antimicrobial polymers (callose and lignin (Pei et al., 2020)), antimicrobial proteins¹⁸ Ace-AMP1 of onion *Allium cepa L*. (Cammue et al., 1995), antimicrobial peptides¹⁹ (Campos et al., 2018), and **induced antimicrobial metabolites** (phytoalexins) (Pedras and Yaya, 2015).

Phytoalexins are antimicrobial compounds produced by plants in response to biotic and abiotic stresses (VanEtten et al., 1994). For instance, 3-deoxyanthocyanidin (**20**, Figure 4, Figure 16B) of sorghum, lubimin (**10**) of potato, kauralexin B2 (**13**) of maize, pinoresinol-4-O-glucose (**22**) of pea pods, camalexin (**26**) of *A. thaliana*, spirobrassinin (**27**) of *Brassica* crops, all cyanides released from cyanogenic glycosides and deglycosylated benzoxazinoids (e.g., **33** to **40**) (Figure 4) are classic phytoalexins associated with fungus infection.

¹⁸ Plant antimicrobial proteins: e.g., PR: pathogenesis-related

 $^{^{19}}$ Plant antimicrobial peptide contains 10 and 60 amino acids thionins, defensins, hevein-like peptides, knottin-type peptides, α -hairpinin family, lipid transfer proteins, snakins (Campos et al., 2018).

I.A.7. Oilseed rape (B. napus)

I.A.7.1. Origin and genomic evolution of oilseed rape

Oilseed rape (*Brassica napus L.*) in the genus *Brassica* is a very complex member of the *Brassicaceae* family that contains many cultivated plants and wild species. The cytogenetic relationships between *B. napus* and their closest relatives were first systematically explained more than 80 years ago (Nagaharu U, 1935) (Figure 18). This allopolyploid species ('AACC', 2n = 4X = 38) is the result of interspecific crosses whose estimate dates back to about 0.075 Ma (Chalhoub et al., 2014), between a diploid turnip (*Brassica rapa*, 'AA', 2n=2x=20) and a diploid cabbage (*Brassica oleracea*, 'CC', 2n=2x=18) (Nagaharu U, 1935) (Figure 17). Three subspecies are recognized in *B. napus: subsp. oleifera* (oilseed rape or rape), *subsp. rapifera* (rutabaga or swede), and *subsp. pabularia* (leaf rape).



Figure 17: The "triangle of U" diagram (Nagaharu U, 1935) **shows the genetic relationships between six genus species of** *Brassica*. Different colors represent the chromosomes of each of the A, B, and C genomes.

Ph D. manuscript | Anani Amegan MISSINOU

Two triplication events and two genome duplication events occurred during the evolution of *Brassicas*. The oldest triplication event allowed the separation between monocots and dicots, and the most recent caused the hexaploid ancestor of Brassicas (Figure 18). Phylogenetic approaches have well-defined the evolutionary relationships between most of these genomes. The diploid progenitors, each derived by ancestral hexaploidy from a common origin, diverged from the *A. thaliana* pattern 17-20 Mya ago (Waminal et al., 2016).



Figure 18: Genomic origin of oilseed rape. Different authors provide different estimates of divergence time and WGD/WGT time. WGD: whole-genome duplications (polyploidization), WGT: whole-genome triplication. Mya: millions of years ago (Waminal et al., 2016).

Originally, oilseed rape was a winter crop that underwent adaptive changes under a combination of natural and artificial selection facing different geographical and climatic environments. So far, spontaneous populations of *B. napus* morphotypes that were initially hybridized have not been identified. However, recent studies have identified

that portions of the A and C subgenomes come from European *B. napus* (An et al., 2019; Song et al., 2020). *B. napus* ecotypes are widely distributed in Europe, Asia, Australia, and North America (Zou et al., 2019), with winter and spring cultigroups.

I.A.7.2. Economic importance and breeding of *B. napus*

Oilseed rape is the second most widely grown oilseed crop in the world, behind soybeans (Mittaine and Mielke, 2012), whose nutritional intake is based on oil (44% of seed weight) and proteins (23% of seed weight) (Balalić et al., 2017), with a world contribution extending to around 13-16% of vegetable oil (Wang et al., 2018). In Western Europe and Asia, mainly winter oilseed rape accessions are grown, while spring oilseed rape accessions are primarily grown in Canada, Australia, and northern Europe. Winter and spring oilseed rape have been the subject of specific breeding to improve their production, resistance to pathogens, and seed quality. By decreasing their GLS and erucic acid contents in the seeds, winter and spring *B. napus* accessions with low GLS content in the seeds have emerged (Pellan-Delourme and Renard, 1988; Denis et al., 1993; Delourme et al., 1994; Nesi et al., 2008).

First, "0" cultivars (the "Primor" cultivar) with low levels of unhealthy erucic acid (a monounsaturated fatty acid) in the oil were obtained (Morice, 1979). Then, the level of GLS was reduced in the meal of these cultivars, giving rise to "00" cultivars. This double innovation allowed the production of high-quality edible vegetable oil and meal proteins (Pinochet and Renard, 2012), which spread throughout Europe. Then, high-yielding fertile hybrid cultivars were produced and are currently grown in most regions dedicated to oilseed rape cultivation in France (Pinochet and Renard, 2012). Today,

oilseed rape/ canola²⁰ after having been an auxiliary crop for food and industrial purposes for the last half-century, the total cultivated area in France is 1.4 million hectares, yielding more than 5.4 million tons (figure 19) (Charbonnier et al., 2019).



Figure 19: A) Global distribution of oilseed rape production measured in tons in 2017 and B) Evolution of rapeseed production measured in tons from 1961 to 2017. The color gradient illustrates the growing regions from the FAO database). (https://ourworldindata.org/grapher/rapeseed-production).

I.A.7.3. Brassica Genomic Resources

In 2018, at the beginning of my research work, the first *B. napus* genome (NCBI taxonomic identifier: txid3708) was already assembled from the variety "Darmor-*bzh*" that possesses a dwarf gene compared to its parent, 'Darmor' (Chalhoub et al., 2014). This accession exhibits phenotypic resistance to *Leptosphaeria maculans*, the causative agent of stem canker (or Phoma), whose cycle and impact on *Brassica* crops will be discussed in the following subsections. Today, there are nine other genomes assembled from other *B. napus* accessions (Song et al., 2020). Genomes of good quality are also available for the progenitors, with two *B. oleracea* genomes (Parkin et

²⁰ Canola is an annual winter or spring oil crop in the Brassicaceae family, created by plant breeding in Canada in the 1970s.

al., 2014; Belser et al., 2018; Lv et al., 2020) and one *B. rapa* genome (Wang et al., 2011; Belser et al., 2018).

The 'Darmor-*bzh*' genome has undergone updates (Bayer et al., 2017; Lee et al., 2020), the most recent of which (Rousseau-Gueutin et al 2020) brings new data and offers, among other things, better assembly and annotation of genes shifting from 89% to 97%. The improved quality of this genomic resource is a key point for every functional genomic approach (Rousseau-Gueutin et al., 2020).

I.A.7.4. Brassica napus - Leptosphaeria maculans interaction

The significant oilseed rape cultivation extension brought setbacks, such as a negative fingerprint on the environment, the emergence of harmful pathogens, and pests endangering oilseed rape cultivation (Carré and Pouzet, 2014). Several major oilseed rape diseases are caused by pathogens that affect the stems (*Sclerotinia sclerotiorum, Leptosphaeria maculans, L. biglobosa, Alternaria spp., Pseudocercosporella capsellae, Pyrenopeziza brassicae*) or roots (*Plasmodiophora brassicae, Verticillium longisporum*) (Figure 20). The most damaging/dangerous diseases to final yield in all oilseed rape growing regions are clubroot, sclerotinia stalk rot, and stem canker (Phoma); light leaf spot and ringspot caused by *Mycosphaerella* are only reported in Europe (Figure 20) (Zheng et al., 2020).



Figure 20: Presence of important diseases on oilseed rape worldwide (A) and in Europe (B). The top 10 diseases worldwide are shown in bold. *In China, only *Leptosphaeria biglobosa* has been reported (absence of *L. maculans*) according to Zheng et al., 2020.

Oilseed rape production in many countries is affected by the pathogenic fungus *L. maculans* (commonly known as phoma), which causes crown stem cankers, known as blackleg disease or phoma (Fitt et *al.*, 2006; Dukhnytskyi, 2019). The disease results in production losses of 90%, equivalent to approximately \$7.3 million, leading to severe damage to the oilseed rape oil industry (Sprague et al., 2006a; van de Wouw et al., 2010). Phoma also affects other *Brassica* crops (oilseed rape, canola, mustard, turnip, cabbage) worldwide.

I.A.7.4.1. The life cycle of Leptosphaeria maculans in Europe

L. maculans (anamorph *Phoma lingam* Tode ex Fr.) belongs to the Phylum *Ascomycetes*, Class *Dothideomycetes* (*Loculoascomycetes*), and Order *Pleosporales*. This fungus is a hemibiotrophic²¹ pathogen with a complex life cycle, alternating saprophytic and long parasitic phases. Its life cycle is closely linked to its host plant, oilseed rape (Rouxel and Balesdent, 2005; Bousset et al., 2018).

<u>Conservation and primary infection</u>: infection starts in summer; the strains of *L. maculans*, having infected plants the previous season, ensure their long-term survival by saprophytism on the stem residues (**stage 1: summer**, Figure 21). These strains of *L. maculans* are of the obligate heterothallic²² type, and their sexual reproduction forms pseudothecia containing ascospores after sporulation²³. The ascospores contained in the pseudothecia²⁴ serve as inoculum for primary infection, and their maturation requires a temperature varying between 15°C and 20°C (Naseri et al., 2009). This coincides with the young plant stage of the developmental cycle. Mature ascospores are ejected from the pseudothecia during rainfall events or even dew.

Ph D. manuscript | Anani Amegan MISSINOU

²¹ Hemibiotrophic pathogens keep their host alive while establishing themselves in the host tissue, absorbing nutrients with a brief biotrophic-like phase before having a necrotrphic phase.

²² An obligate heterothallic fungus produces ascospores only if mycelia from strains with opposite sexual signs (mating types) occur on the same plant.

²³ Sporulation is the phenomenon of differentiation which leads from the vegetative form to the spore (form of resistance).

²⁴ Pseudothecia arise from the encounter between two strains with opposite sexual signs.



Figure 21: The pathogenic life cycle of *Leptosphaeria maculans* in winter oilseed rape in Europe. (1) saprophytism on stem residues (long-term survival); (2) sexual reproduction/construction of primary inoculum; (3 to 4) a short phase of asymptomatic necrotrophy on the leaves of the plants after the establishment of the first infections; (5) a long phase of endophyte (biotroph) with asymptomatic colonization in the tissues of the leaf tissues at the stem and the crown stem (late Autumn to late Winter); (6) second phase of symptomatic necrotrophy, formation of basal stem cankers (severe necrosis) and lodging phenomena. (Rouxel and Balesdent, 2005).

After the precipitation decreases, it is disseminated over several meters or even a few hundred meters (Marcroft et al., 2004; Lô-Pelzer et al., 2009). Ascospores initiate primary infection by plant penetration through leaf wounds or stomata, causing the formation of leaf lesions or leaf spots. The effectiveness of this contamination requires leaves moistened by rain or dew, an optimum temperature (15-20°C), and darkness (**step 2: mid-autumn,** Figure 21). From then on, *L. maculans* colonizes the tissues without expression of visible symptoms and grow exclusively in the apoplast for a few days before becoming necrotrophic over a short period (Gout et al., 2006). This change

in the mode of nutrition marks the induction of the death of infected cells and the formation of leaf lesions. Although not detrimental to the final yield, these lesions contain the spores which, by asexual reproduction, will give rise to pycnidia having conidia, sources of the secondary/alternative inoculum (**stage 3: in mid-autumn,** Figure 21).

The secondary phase of the cycle is dispersed locally by the wind and splashes onto nearby plants and leaves. Thus, the infection follows curves close to those defined for ascospores (stage 4: mid-autumn, Figure 21). At this time, L. maculans adopts a biotrophic²⁵ mode of nutrition and enters an endophytic phase. The fungus continues its development by migrating through the leaf veins, then the petiole, and finally reaching the stem. From the stem, *L. maculans* overwinters and continues its journey systemically colonizing the plant until it reaches the crown of the stem by migrating into the intercellular spaces through the plant tissues and completing its journey at the crown level of the stem. This whole phase of migration from the leaves to the crown is asymptomatic and can last a few months in the case of winter oilseed rape in Europe (stage 5, Figure 21). In spring, *L. maculans* becomes necrotrophic again and destroys the cortex cells at the base of the stem (crown of the stem) by altering the water and mineral nutrition of the plant, eventually inducing severe necrosis, which weakens the bottom of the stem. These necroses lead to lodging phenomena and yield losses (stage 6: from the middle of spring, Figure 21). (West et al., 2001). The disease epidemics are favored by mild and wet autumns and high temperatures in the spring (Evans et al., 2008).

 $^{^{\}rm 25}$ A biotroph is an organism that feeds off another organism.

Ph D. manuscript | Anani Amegan MISSINOU

After the cultivation period and stem lodging, the sexual reproduction phase sets in, and the cycle can begin again. The amount of pathogen in the spring is a determining factor in the inoculum level the following fall.

Several methods are used to control phoma in *Brassica* crops (prophylactic, chemical, biological, genetic) (Aubertot et al., 2006). Genetic control by improving host resistance traits is the most economical and ecological method.

I.A.7.4.2. Genetic basis of B. napus resistance to L. maculans infection

There are two types of resistance to *L. maculans* in oilseed rape occurring at different ontogenetic stages of oilseed rape. The qualitative resistance has a monogenic determinism (major gene R), which confers total resistance to a specific strain of the pathogen, which is a total resistance of seedlings expressed from the cotyledon stage onwards (figure 21). The second type is quantitative resistance, governed by several minor genes; it has a polygenic determinism and confers active partial resistance on a wide range of strains of *L. maculans* (in other words, "broad-spectrum resistance"). It is mainly expressed at the stem stage of infection in adult plants (Figure 22). (Delourme et al., 2006; Huang et al., 2009; Levrel et al 2022).



Figure 22: Genetic resistance associated with the pathogenic life cycle of *Leptosphaeria maculans.* (A) qualitative/total and monogenic resistance (B) partial and polygenic quantitative resistance (Adapted from Huang et al., 2014).

There are different sources of resistance to *L. maculans*. Many studies on the hereditary transmission of resistance have been carried out in the growth stages of seedlings and adult plants. More than fifteen specific genes (*Rlm1* to *Rlm14*; *LepR1* to *LepR4*) have been identified. Two of these genes have been cloned into *B. napus* cultivars, including *LepR3/Rlm2* (BnaA10g20720D), and *Rlm9* (BnaA07g20220D). *Rlm2/LepR3* (allelic variants) and *Rlm9* code for a receptor-like protein (LRR-RLP) and a wall-associated kinase-like protein (WAKL), respectively. *LepR3* and *Rlm2* mediate specific resistance to *L. maculans* isolates harboring the avirulence proteins *AvrLm1*, *AvrLm2*, and *Rlm9* to those harboring *AvrLm5-9*, respectively (Larkan et al 2012.; Larkan et al., 2014; Larkan et al., 2020). While *R* genes cause complete resistance, some *R* genes lose their effectiveness after deployment in the field. Indeed, many cultivars recorded as being resistant show shreds of evidence of specific resistance breakdown in response to rapidly changing populations of *L. maculans*. These include

Rlm1, Rlm3, Rlm6, Rlm7, LepR1, and LepR3 genes (Brun et al., 2000; Rouxel et al., 2003; Sprague et al., 2006a; Sprague et al., 2006b; Brun et al., 2010; Winter and Koopmann, 2016; Zhang et al., 2016; Van de Wouw et al., 2022).

Faced with this problem, it is crucial to adopt a new approach for sustainable management of oilseed rape cultivation against phoma disease and resistance breakdown. Chemical control is difficult to implement and has a high environmental footprint. Interestingly, quantitative resistance mediated by multiple genetic factors was shown to increase the potential for the durability of major resistance genes (Brun et al. 2010; Delourme et al. 2014). Breeding cultivars with quantitative resistance or with combined gualitative and guantitative resistance can thus effectively contribute to stem canker control in oilseed rape. Quantitative resistance of *B. napus* to L. maculans is highly polygenic and controlled by many QTL (Pilet et al. 1998; 2001; Fopa Fomeju et al. 2014; Jestin et al. 2015; Raman et al. 2018; 2020; Kumar et al. 2018; Fikere et al. 2020; Fu et al. 2020; Vollrath et al. 2021). However, little is known about quantitative resistance. Stem infection processes and histological changes were studied in wintertype B. napus by Hammond and Lewis (1987) and Li et al. (2007). Hammond and Lewis (1987) observed a sequence of host reactions, including lignification, cambium formation, and callose deposition at the perimeter of the lesion in the stems, with variation in the timing, the spatial distribution, and the intensity of lignification according to cultivars and isolates. Li et al. (2007) showed the variability in location, timing, and histochemistry of stem responses between compatible and incompatible interactions. Post-penetration defense reactions such as lignification, suberization, and additional cambium formation occurred in resistant cultivars but four to five days earlier in cultivars with monogenic resistance than quantitative resistance, resulting in

hypersensitive reaction in incompatible interactions and limitation of the necrosis in cultivars with quantitative resistance. Significant advances were made in deciphering the pathogen's life cycle and especially in identifying effectors that correspond to small secreted proteins (Sánchez-Vallet et al., 2018). Eight waves of genes specifically expressed during the interaction with the plant were discriminated, all being enriched in genes encoding effectors (Gay et al., 2021). Of these, *L. maculans* expresses several genes encoding late effectors during the systemic colonization of rapeseed and are thought to interfere with plant defenses in the stem (Gervais et al., 2017). The identification of these plant defenses would help in understanding the mechanisms underlying quantitative resistance and its further use in plant breeding.

I.A.7.4.3. Biochemical bases of constitutive and inducible chemical defenses of Brassicas

The ambition here is not to (re)conceptualize "the evolution of metabolic pathways in *Brassicaceae*". Above all, with the proven risk of being partial, given the difficulty of not being able to list exhaustively the articles on the subject of shared or specific specialized metabolic pathways between *Arabidopsis* and *Brassica*, the interest will be focused on those involved in the plant-pathogen interactions.

Given that, the diverse origin of these compounds was discussed in the section dealing with structural diversity (I.A.2.1). In what follows, I will briefly present the biosynthetic pathways of the most characterized GLSs: aliphatic glucosinolates derived from methionine and indole GLSs derived from tryptophan. The biosynthetic pathway of aliphatic GLSs derived from methionine is initiated by BCAT4²⁶, which operates an elongation of the methionine-C3 chain into a C3 precursor at the crossroads of the biosynthetic blocks of C3-, to C10- derived from methionine, by the activity of MAM1²⁷, IPMI, IPMDH1. These hammers of each contingent are used for the biosynthesis of methylsulfanyl-GLS derivatives by concerted actions of several enzymes, including CYP79F1/F2, CYP83A1, GSTF11/U20, GGP1, SUR1, UGT74C1, SOT17/18. GS-OX supports methylsulfanyl-GLSs derivatives for methylsulfinyl-GLSs derivative biosynthesis (figure 22). By the action of AOP2 and AOP3 enzymes, alkenyl-GLSs, and hydroxyalkyl-GLSs. The alkenyl-GLSs are then taken up by GLS-OH to give hydroxyalkenyl-GLSs, including the famous (epi)progoitrin. Only C3- to C5- GLSs derived from methionine are conserved between *Arabidopsis* and *Brassica* (Figure 23).

²⁶ BCAT: branched-chain aminotransferase

²⁷ MAM : methylthioalkylmalate synthase



Figure 23: Representation of the main biosynthetic pathways of methionine-derived glucosinolates in *Brassica* species and Arabidopsis. Allelic variability on MAM genes contributes to the diversity of side-chain lengths. Allelic variability in the genes encoding the enzymes AOP2, AOP3, GS-OH, and BZO1 contributes to the variety of side-chain chemical decorations. Esterified forms of benzoates (3BOP and 4BOB) have never been described in *Brassica*, despite copies of BZO1 in the *B. napus* genome. The C11-aliphatic compounds of Camelina are not shown. (Gravot, 2019).

The biosynthetic pathway of indole glucosinolates derived from tryptophan is conserved between *Arabidopsis* and *Brassica*. It is dependent on the concerted action of cytochromes P450 (CYP79B2/B3 and CYP83B1), providing the precursors necessary to produce glucobrassicin (3IMG). Some enzymes degrade/convert GLSs into compounds that can be classified as nitriles (Figure 24, IAN in panel), phytoalexins (Figure 24, all combinations in Box 7), thiazolidinediones (Figure 24, raphanusamic Acid, Box 6), carbinol (Figure 24, 1MI3C and 4MI3C box 2). The release of these

products repels herbivorous insects, birds, nematodes, and fungi. However, the enzymes downstream of glucobrassicin and IAN synthesis are primarily taxon-specific, so many of them remain partially understood in *B. napus*. A number of these metabolites are thought to be absent in *Brassica* species. For example, raphanusamic acid, an inducible insecticide in *Arabidopsis*, results from the activity of atypical myrosinase in intact cells of infected plants (Bednarek et al., 2009). More than 40 *Brassica*-specific indole phytoalexins have been reported, and for some of them, the biosynthetic pathways are poorly characterized (Pedras and Yaya, 2010).

On the indole core of 3IMG, CYP81F4, and CYP81F1-3 form hydroxylated derivatives of glucobrassicin (1-hydroxylucobrassicin and 4-hydroxylucobrassicin), which receive methyl groups to produce neoglucobrassicin and 4-methoxylucobrassicin by the action of IGMT5 and IGMT1-4 (Fig. 24). Assuming an absence of mutual exclusion of the origin of these isomeric forms. there enzymes at the may be that dimethoxyglucobrassicin as reported in Barbarea vulgaris (Blažević et al., 2020), or dihydroxyglucobrassicin (never registered), or even chimeric forms such as hydroxymethoxy-glucobrassicin, can exist in oilseed rape (figure 24, box 2).



Figure 24: Biosynthesis pathways of auxins, phytoanticipins, and tryptophan-derived phytoalexins characterized (at least in part) in Arabidopsis and some Brassica species. (1) The multiplicity of auxin biosynthesis pathways described from tryptophan. A recent consensus attributes a significant role of IAA synthesis to the TAA / YUCCA pathway under non-pathological conditions (Won et al., 2011; Lehmann et al., 2017). (2) Synthesis of indolic glucosinolates in all Brassicaceae: I3MG (glucobrassicin) and its methoxylated derivatives (1M-I3MG and 4M-I3MG) can be metabolized by the activity of myrosinases (TGGs) to give I3C (indole-3-carbinol or its methoxylated derivatives 1MI3C and 4MI3C). Based on allelic combinations on several loci, the presence of Epithiospecifier proteins (ESP / ESM1) conditions the metabolism of **I3MG** to IAN, which can be a source of biosynthesis for auxin (IAA) via the activity of nitrilases (NIT1-3) (Pfalz et al., 2016; Lehmann et al., 2017) (Lehmann et al. 2017). (3) The synthesis of 3-carboxylic acid Indole (I3CA) and its derivatives from IAN as described by Erich Glawischnig's team (Böttcher et al., 2014). (4) In Arabidopsis, the biosynthetic pathway (absent in the Brassica genus) results in the synthesis of what has long been considered the only phytoalexin in Arabidopsis: camalexin (Iven et al., 2012). (5) In parallel with camalexin, IAN is the precursor of a procession of phytoalexins derived from indole 3carboxynitrile, whose pathway was characterized recently in Arabidopsis, appears to be absent from the Brassica genus (Rajniak et al., 2015). (6) Also, in Arabidopsis, I3MG and its methoxylated derivative 4M-I3MG can be converted by atypical myrosinase into intact cells of infected plants, leading to the synthesis of raphanusamic acid (Bednarek et al., 2009; Clay et al., 2009). (7) Of the nearly 40 indolic phytoalexins described in Brassica, significant steps have recently been characterized (Pedras and Yaya, 2010). Brassinin is synthesized from the I3MG-derived isothiocyanate via the action of atypical BABGa/b myrosinases (Brassinin Associated Beta Glucosidases) and can lead to the synthesis of spirobrassinin and its derivatives spirobrassinin in B. rapa, or 1-methoxybrassenin B in B. oleracea (Klein and Sattely, 2017). (Figure was adapted from Gravot, 2019).

The biosynthetic pathways of GLSs derived from phenylalanine, tyrosine, and linear and branched amino acids are even less described in all species of *Brassicales*, and these compounds are rarely reported in the genera *Arabidopsis* and *Brassica*, except for gluconasturtiin (Agerbirk et al., 2015; Liu et al., 2019).

Concerning phenolics, the biosynthetic pathways are poorly described, and most of the studies are based on corroborations with *Arabidopsis*. However, there is quite a dazzling literature on polyglycosylated and polyacylated forms of *Brassica* phenolic compounds (Nielsen et al., 1998; Llorach et al., 2003; Ferreres et al., 2005; Harbaum et al., 2007; Francisco et al., 2009; Olsen et al., 2009; Schmidt et al., 2010; Velasco et al., 2011; Farag et al., 2013; Shao et al., 2014). The characterization of these pathways highlights the enzymes involved in plant-pathogen interactions. The genetic determinism of a large part of these phytoanticipins is unknown.

I.B. Presentation of the Ph.D.

I carried out my thesis²⁸ at IGEPP (Institute of Genetics, Environment and Plant Protection, UMR INRAE/Institut Agro/Université Rennes 1) in the DEBI team (<u>D</u>iversity, <u>E</u>volution, and genomics of <u>BI</u>otic interactions). This group, led by Pr. Antoine Gravot and Dr. Mathieu Rousseau-Gueutin, works on the genetics/genomics of biotic interactions in *Brassica*, including blackleg and clubroot. The objectives of the team are (1) to identify/diversify sources of resistance in the *Brassica* gene pool, (2) to unravel the genetic/evolutive mechanisms of quantitative resistance, (3) to understand the cellular/molecular processes involved in quantitative resistance in this species, and

²⁸ With a 3.5 years scholarship from INRAE and the Region of Brittany (starting in November 2018)

(4) to better characterize the phytochemical diversity of oilseed rape and its modulation in the presence of environmental stress.

I.B.1 Objectives of the Ph.D. and research questions

The complex SM repertoire of *B. napus* and its genomic components, despite the close phylogenetic proximity with the species *Arabidopsis thaliana*, exhibits a range of specificities. This leads to still incomplete knowledge of this repertoire, its diversity, and its regulation, which impairs ambitious functional genomics works on *Brassica* defense chemistry. Hence, the characterization of the specialized metabolism diversity in *Brassica* crops is of dual importance: first, from a fundamental perspective, this represents a direct or indirect bridge between genotypes and phenotypes of these crops; second, from a breeding perspective, it is a first step to identify candidate metabolites/genes and to elucidate the relevant metabolic pathways in ecological interactions.

The objective of my thesis was to characterize the specialized metabolism of *B. napus* in relation to the pathogen resistance response. This work was subdivided into two parts: (1) a characterization of the phytochemical diversity in a panel of *Brassica*, and the genetic control of SMs constitutively biosynthesized in roots and leaves, and (2) a study on the regulation of SMs in response to fungal infection by *L. maculans* in the stems of *B. napus*.

I.B.2. Thesis outlines

The second chapter highlights our work characterizing the phytochemical diversity of *B. napus* and its genetic determinants using metabolomics and quantitative genetics

approaches. We have developed a metabolic profiling method to simultaneously identify ninety-two oilseed rape compounds by liquid chromatography coupled in tandem with detection by high-resolution mass spectrometry (HPLC-Orbitrap LTQ-XL). Eighty-six compounds resolved by chromatography were selected and quantified by UPLC-UV-TQD. Of these compounds, 18 UV-absorbing root compounds were detected for the first time herein. The results have been described in depth, resulting in one published article and two manuscripts in preparation.

The first, recently published in "*Journal of Agricultural and Food Chemistry*" (IF 5.279), describes the identification and quantification of 36 glucosinolates and 32 phenolic compounds in a large panel of *B. napus* and highlights valuable genetic resources for chemical ecology and breeding (Missinou et al., 2022). Furthermore, the second manuscript concerns genome-wide association studies that provide insights into the genetic basis of specialized metabolism and reveal seed glucosinolate content's breeding history in *B. napus*. Its submission is planned to "*Plant Biotechnology Journal*" (IF 9.803). Thirdly, we exploited the absorbance variation of 18 UV-absorbing compounds in the genetic investigation of novel root UV-absorbing and used NMR to initiate de novo structural elucidation of the most relevant metabolites. This article is planned to submit to the "*Phytochemistry*" journal (IF 4.072).

The third chapter was devoted to the characterization of metabolic responses of *B. napus* to the fungal *Leptosphaeria maculans*. This chapter investigated previously unknown phytoalexins involved in the defense response of *B. napus* to *L. maculans* infection by UPLC-Orbitrap-Q-Exactive and molecular network-based (GNPS) annotation. The fifth manuscript deals with the spatialization of the metabolic response to infection using mass spectrometry imaging approaches. Both submissions are planned in "*Journal of Experimental Botany*," IF 5.906).

The first article is associated with so-called additional information documents, presenting the details of the experimental data. These documents, which can be consulted on the publisher's websites, have not been included in this thesis. The link is specified at the beginning of the article. As for the additional information, documents of the other manuscripts are presented in the appendix of each study.

In addition to these results, I participated in MetaPhor project, another project of the team. I work with Regine Delourme (Researcher within the team) on transcriptomic and metabolomic modulation of oilseed rape during *L. maculans* infection. My contribution to validating the metabolomic variation presaged by the variation in gene expression associated with the metabolism of aliphatic and indole glucosinolates (Delourme et al., in preparation).

The fourth chapter leveraged a general discussion to summarize the significance of the results and identifies the boundaries of our research to offer directions for future works.
Chapter II: Characterization of the diversity of constitutive specialized metabolites and its genetic determinants in *Brassica* species

Introduction du chapitre II

La diversité structurelle des phytoanticipines correspond souvent à l'adaptation évolutive des organismes confrontés à des changements environnementaux spécifiques. Du fait de leur emplacement apical²⁹ dans l'arbre phylogénétique des *Brassicales*, les espèces du genre *Brassica* constituent un modèle idoine pour l'identification et la caractérisation des gènes associés aux étapes biosynthétiques menant à de nouvelles structures de défense chimiques propres aux espèces de la famille des *Brassicacées* (Edger et al., 2018).

L'objectif global de ce chapitre est de caractériser la diversité phytochimique du colza et de déterminer les déterminants génétique contrôlant leur teneur afin de contribuer à une meilleure connaissance du métabolisme des phytoanticipines. Nous avons incorporé *B. napus* et les espèces progénitrices (*B. rapa* et *B. oleracea*), y compris divers cultigroupes englobant les accessions WOSR (winter-type oilseed rape), SOSR (spring-type oilseed rape), rutabaga, fourrager, qui avaient le statut "+ /+", "0/+", "+/0" et "0/0" pour l'acide érucique et les GLSs dans les graines afin de maximiser la diversité phytochimique potentielle (combinant ainsi les génotypes de printemps, hiver, fourrager et rutabaga). L'ambition ici était : (i) d'aller au-delà des affirmations empiriques de l'absence/présence de certains composés rarement décrits chez le colza, (ii) de mettre les données en perspective avec des génotypes représentatifs de la diversité des deux espèces progénitrices, (iii) de fournir une étape importante de l'architecture génétique de composés rarement rapportés chez *Brassica*, (iv) d'étudier les conséquences de la sélection du colza pour la teneur en GLSs dans les graines,

²⁹ Le genre *Brassica* est issu du dernier événement de triplication des génomes chez les plantes.

Ph D. manuscript | Anani Amegan MISSINOU

et (v) d'enrichir le catalogue de la diversité phytochimique des composés racinaires absorbant dans les UV et leurs déterminants génétiques. Le chapitre est structuré autour de trois articles complémentaires, un publié et deux en cours de finalisation.

Le premier article, publié dans Journal of Agricultural and Food Chemistry (AA Missinou, J Ferreira de Carvalho, N Marnet, T Delhaye, OHamzaoui, D Abdel Sayed, Y Guitton, LLebreton, C Langrume, A Laperche, R Delourme, MJ Manzanares-Dauleux, A Bouchereau, A Gravot. 2022. Identification and Quantification of Glucosinolates and Phenolics in a Large Panel of Brassica napus Highlight Valuable Genetic Resources for Chemical Ecology and Breeding. J. Agric. Food Chem https://doi.org/10.1021/acs.jafc.1c08118), vise à décrire la diversité phytochimique des glucosinolates et des composés phénoliques dans un large éventail d'accessions de Brassica. Pour sa réalisation, j'ai collaboré avec des ingénieur(e)s chimistes, dont Nathalie Marnet de plateforme P2M2 (Corsaire, Biogenouest), Yann Guitton plateforme MELISA³⁰ (Corsaire, Biogenouest) et Thomas Delhaye de la plateforme de spectrométrie de masse et métabolomique de l'UMR 6164. Nous avons développé une méthode de profilage simultanée des glucosinolates et des composés phénoliques. Un atlas phytochimique de 36 glucosinolates (GLS) foliaires/racinaires et 32 composés phénoliques foliaires a été établi, dont certains sont rarement décrits chez les Brassica. La quantification de ces composés dans un panel de 304 accessions de Brassica a révélé de grandes variations qualitatives et quantitatives, notamment pour la glucocochlearine et les dérivés kaempferol tetra- et penta-hexosides. Deux principaux

³⁰ La plateforme MELISA (MEtabolomics Llpidomics Steroidomics Analysis) du Laboratoire d'études des résidus et contaminants dans les aliments (LABERCA) est membre de la composante "Analyses structurale et métabolomique" (Corsaire) du réseau Biogenouest.

chémotypes se dégagent en fonction des proportions de sinapoyl-O-hexoside et de kaempferol-O-trihexoside. Eu égard à ces phénotypes variables, nous avons émis le postulat que des polymorphismes génétiques seraient responsables des variations qualitatives et quantitatives observées entre les accessions et cultivars, à l'instar des résultats obtenus dans plusieurs études précédentes sur les glucosinolates (Kliebenstein et al., 2001b; Kliebenstein et al., 2001a; Rowe et al., 2008; Kittipol et al., 2019; Liu et al., 2020) et les composés phénoliques (Routaboul et al., 2012; Francisco et al., 2016; Ishihara et al., 2016; Tohge et al., 2016).

Le deuxième article (en préparation J Ferreira de Carvalho*, AA Missinou*, N Marnet, L Lebreton, C Langrume, A Bouchereau, M Manzanares-Dauleux, R Delourme, A Laperche, A Gravot. Genome-wide association studies provide insights into the genetic basis of specialized metabolism and reveal the breeding history of seed glucosinolate content in Brassica napus. For submission to Plant Biotechnology Journal) a pour objectifs : (1) d'identifier des QTLs associés à des variations métaboliques, (2) de proposer des gènes candidats sous-jacents aux QTL détectés pour des études de génomique fonctionnelle, et (3) d'appréhender l'impact de la sélection pour une faible teneur en GLSs aliphatiques dérivés de la méthionine dans les graines de colza sur les GLSs racinaires et foliaires (incluant les GLSs rares des Brassica) et sur les composés phénoliques foliaires. L'objectif à moyen terme est de mettre en relation les QTL métaboliques et les QTL associés à la résistance à des agents pathogènes afin d'établir des hypothèses sur les mécanismes sous-jacents à ces résistances. Pour atteindre ces objectifs, j'ai collaboré avec Julie Ferreira de Carvalho qui a réalisé les analyses de génétique d'association (GWAS) pour identifier les QTLs contrôlant le contenu de ces métabolites (QTL métaboliques ou mQTLs) dans les différents

Ph D. manuscript | Anani Amegan MISSINOU

organes. Le défi majeur de cette approche est de construire un réseau à partir des mQTLs identifiés et de l'analyser en regard de la diversité structurelle et des voies de biosynthèse des phytoanticipines. A l'instar du travail initié précédemment au laboratoire (Wagner et al., 2019), j'ai construit des réseaux intégrant les associations mQTL-métabolites afin de formaliser l'architecture génétique associée aux variations qualitatives et quantitatives du phénotype métabolique de *B. napus*. Nous avons identifié 105 mQTLs impliqués dans le contrôle des profils phytochimiques racinaires et foliaires, soulignant une diversité de modules génétiques contrôlant indépendamment les différentes sous-catégories biochimiques.

Lors de l'étude de la diversité phytochimique dans un large éventail d'accessions de *Brassica*, nous avons repéré 18 métabolites racinaires absorbant dans les UV et qu'il n'a pas été possible d'annoter car vraisemblablement absents des bases de données métabolomiques. Parmi ces molécules, deux métabolites de structure inconnue avaient été suggérés par mon laboratoire d'accueil comme étant impliqués dans la résistance à la hernie (Wagner et al., 2019). L'objectif de cette troisième partie, présentée sous forme d'article en préparation (*AA Missinou, C Orione, J Ferreira de Carvalho, Y Haddad, N Marnet, P Jéhan, A Laperche, R Delourme, A Bouchereau, MJ Manzanares-Dauleux, A Gravot. Structural and genetic investigation of novel root UV-absorbing compounds reveals major QTLs controlling the content of two pteridine-like amino acid conjugates in Brassica. To be submitted to Phytochemistry) est de caractériser la chimie structurale et le contrôle génétique de ces nouveaux métabolites racinaires. Un effort de purification a conduit à une annotation structurale par RMN concluante pour certaines des molécules. Les résultats préliminaires suggèrent qu'il s'agirait d'une petite famille de composés nouveaux, dont deux ont précédemment été*

rapporté comme étant associés à la résistance à la hernie (Wagner et al., 2019). L'analyse GWAS a identifié 11 mQTLs contrôlant la teneur en ces composés, en majorité structurellement distincts les uns des autres. Dans cette partie, j'ai co-encadré avec Clément Orione (Ingénieur RMN à la plateforme CRMPO³¹), le stage de Master 2 de Youcef Haddad portant sur la purification et la caractérisation structural des molécules racinaires par raisonnance magnétique nucléaire (RMN) combiné à la spectrométrie de masse à haute resolution.

L'ensemble de ces travaux fournit une ressource utile pour l'écologie chimique et l'amélioration des *Brassica*.

³¹ Centre régional de mesures physiques de l'Ouest, Rennes, France

Paper I: Identification and Quantification of Glucosinolates and Phenolics in a Large Panel of *Brassica napus* Highlight Valuable Genetic Resources for Chemical Ecology and Breeding

Link for supplementary files

https://pubs.acs.org/doi/10.1021/acs.jafc.1c08118?goto=supporting-info

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

pubs.acs.org/JAFC

Identification and Quantification of Glucosinolates and Phenolics in a Large Panel of Brassica napus Highlight Valuable Genetic **Resources for Chemical Ecology and Breeding**

Anani Amegan Missinou,^{\perp} Julie Ferreira de Carvalho,^{\perp} Nathalie Marnet, Thomas Delhaye, Oumayma Hamzaoui, David Abdel Sayed, Yann Guitton, Lionel Lebreton, Christophe Langrume, Anne Laperche, Régine Delourme, Maria J. Manzanares-Dauleux, Alain Bouchereau, and Antoine Gravot*

Cite This: J. Ag	ric. Food Chem. 2022, 70, 5245-	-5261	Read Online	
ACCESS	III Metrics & More		E Article Recommendations	s Supporting Information

ABSTRACT: Glucosinolate (GLS) and phenolic contents in Brassicaceae contribute to biotic and abiotic stress responses. Breeding crop accessions harboring agroecologically relevant metabolic profiles require a characterization of the chemical diversity in Brassica germplasm. This work investigates the diversity of specialized metabolites in 281 accessions of B. napus. First, an LC-HRMS2-based approach allowed the annotation of 32 phenolics and 36 GLSs, revealing 13 branched and linear alkyl-GLSs and 4 isomers of hydroxyphenylalkyl-GLSs, many of which have been rarely reported in Brassica. Then, quantitative UPLC-UV-MS-based profiling was performed in leaves and roots for the whole panel. This revealed striking variations in the content of 1-methylpropyl-GLS (glucocochlearin) and a large variation of tetra- and penta-glucosyl kaempferol derivatives among accessions. It also highlighted two main chemotypes related to sinapoyl-O-hexoside and kaempferol-O-trihexoside contents. By offering an unprecedented overview of the phytochemical diversity in B. napus, this work provides a useful resource for chemical ecology and breeding.

KEYWORDS: oilseed rape, specialized metabolism, chemical diversity, BCAA-derived glucosinolates, flavonols, phenylpropanoids, chemotypes

INTRODUCTION

Brassicaceae synthesize a large diversity of glucosinolates (GLSs) and phenolic compounds as bioactive so-called secondary compounds or specialized metabolites. Their profile signatures are often distinctive of subtaxonomic groups¹ and are involved in both the diverse food and feed usages^{2,3} and chemical ecology.^{4,5} For instance, contrasted GLS profiles among 96 ecotypes of Arabidopsis thaliana were associated with phytophagous insect resistance.⁶ The ability to biosynthesize certain acylated flavonoids is a major determinant of UV tolerance among ecotypes of A. thaliana.' Such large-scale studies on the biological importance of intraspecific chemical variations have been mostly investigated in wild *Brassicaceae* species^{8–10} and reported by other authors^{11–13} including wild Brassica oleracea.^{14,15} Available data are however still scarce for cultivated Brassica species.¹⁶⁻¹⁸

Brassica napus is a cultivated plant species of major agronomical importance, encompassing various forms of human and animal feed: oilseed rape (spp. oleifera), swede (ssp. rapifera), and fodder rapeseed. This allopolyploid species arose from a postneolithic interspecific hybridization between B. rapa and B. oleracea, 19,20 which occurred spontaneously. Since 1970, first modification of low erucic acid and low GSLs for animal feeding, modern oilseed rape cultivars have been selected to lower the amount of GLS in seeds (<30 μ mol g⁻¹ DW) from Bronowski accession.²¹ This selection had some side-effect impact on the general lowering of a series of GLS in

leaves of modern cultivars,²² although the extent of this phenomenon may have been mitigated by the counter selection exerted by herbivory pressure during breeding processes.²³ The phytochemical composition in vegetative organs of *B. napus, B. oleracea,* or *B. rapa* has been well documented²⁴⁻³⁰ and put in line with its possible implications in human health and animal nutrition^{2,3} as well as in ecological interactions.³¹⁻³⁵ Remarkable chemical features in leaves of Brassica species are a series of multihexosylated and multiacylated kaempferol, quercetin, and isorhamnetin derivatives, hexosylated or quinic-acid conjugated phenylpropa-noids,^{24,26,27,30,36-41} and glucosinolates. GSL diversity in the genus is mainly derived from tryptophan and chain elongated homologues of methionine and phenylalanine, but the major part of leaf glucosinolates of Brassica napus belong to the methionine-derived aliphatic subfamily.^{26,29,37} By contrast, roots typically accumulate much lower levels of phenolic compounds,⁴² and a broader diversity of glucosinolate structures, including abundant levels of the phenylalanine-

Received: December 23, 2021 **Revised:** March 18, 2022 Accepted: March 21, 2022 Published: April 14, 2022





derived phenylalkyl-GLS gluconasturtiin²⁹ and tryptophanderived indolic-GLS.⁴³ A surge of phytochemistry research has explored the structural diversity of glucosinolates and/or phenolics in different organs and environmental conditions yet often across a limited number of accessions of *Brassica*.^{24–26,29,30,37,43–47} More recently, main glucosinolates have been reported in large panels of *B. napus* accessions.^{22,48} There is still, however, a need for more exhaustive characterization of GLS and phenolic profiles within leaves and roots of *B. napus* as well as the main patterns of accumulation of these compounds across the available varietal diversity.

Hence, the objective of the present work was to get an extensive overview of the diversity of chemical profiles of glucosinolates and phenolic compounds in roots and leaves of a large panel of Brassica accessions. For this, MS/MS annotation of the largest possible range of glucosinolates and phenolic compounds was performed using HPLC-HRMS in selected sets of root and leaf samples reflecting the varietal diversity. From this, the quantitative analysis of 36 glucosinolates and 32 phenolic compounds (in leaves and roots) was performed (UPLC-UV-MS) on a large panel of 304 Brassica accessions. The resulting data set was then analyzed (i) to investigate the chemical diversity among B. napus accessions, (ii) to assess the range of accumulation for each compound, and (iii) to search for the existence of chemotypes. We have searched, in particular, for chemicals showing notable differences between B. napus accessions and the accessions of both parental species B. rapa and B. oleracea. Altogether, the results of this work provides a valuable resource representative of the Brassica chemical diversity for potential use in chemical ecology and breeding.

MATERIALS AND METHODS

Chemicals. HPLC-grade acetonitrile, methanol, and formic acid were from Carlo Erba. Authentic glucosinolate standards (Supporting Table S2) were acquired from PHYTOPLAN Diehm & and Neuberger GmbH. Authentic phenolic standards were provided by Sigma-Aldrich. The Marvin suite was used for drawing and displaying chemical structures and substructures with Marvin version 21.14 ChemAxon (https://www.chemaxon.com).

Plant Growth Conditions and Experimental Design. A total of 281 accessions of Brassica napus (261 ssp oleifera, 10 ssp rapifera, and 10 fodder) were analyzed in this project. The set of spp oleifera encompassed 130 accessions of winter oilseed rape (WOSR), 120 spring oilseed rape (SOSR), and 11 elite WOSR hybrid accessions. WOSR and SOSR included both ancient and "00" accessions. In addition, phytochemical analyses were performed in two small genetic core collections reflecting the genetic diversity of B. rapa (12 accessions) and B. oleracea (11 accessions), thus leading to a whole panel of 304 Brassica accessions. All seed stocks were produced, at the same site, in two batches during the last 2 years before the experiment. Seeds were germinated in a glasshouse in pots, using a mixture of soil and sand (ratio 2:1). Plants were cultivated in a phytotron (photoperiod 16 h light/8 h dark, 20/18 °C) and were fertilized twice with a half-strength Hoagland solution. Additional watering with tap water was automatically brought under the control of soil moisture sensors to maintain soil hygrometry at about 70%. The experiment consisted of three independent biological replicates, performed at 6-week intervals. The 304 accessions were studied in each biological replicate: 8 plants from each accession were cultivated in two spatially randomized blocks of 4 plants. Four weeks after sowing, root and leaf tissues of the 8 plants were removed and pooled per tissue. Roots were rapidly washed with tap water, then immediately frozen in liquid nitrogen, then stored at -80 °C before lyophilization. Dried samples were ground using a TissueLyser (Qiagen, Hiden Germany). The resulting powders were stored in the dark and protected from humidity up to the extraction.

Four weeks was chosen because the root biomass at this stage is sufficient for metabolic analyses and the relatively short experiment time allows performing three independent replicates in 1 year. Morever, many biotic interactions can typically be studied at this stage (clubroot, phoma, insect susceptibility,...).

Metabolites Extraction Procedure. Samples (30 mg) of freezedried powders were extracted with 1 mL of methanol/formic acid (99/1, v/v). The tubes were vortexed and then sonicated in a water bath at 25 °C for 10 min. After centrifugation at 7000 rpm for 10 min, the supernatants were collected and filtered using a 0.22 μ m PTFE membrane. Extracts were stored at 4 °C and analyzed by LC-MS less than 24 h after extraction. This extraction procedure was conducted on each sample for quantitative analyses. In addition, extraction was performed on pools from several roots or leaf powder samples from different accessions (five randomly chosen accessions for each of the following groups: *B. napus* winter, spring, fodder, swede, *B. oleracea*, and *B. rapa*). These two bulks were thereafter used for HPLC-Orbitrap LTQ-XL analysis and annotation of GLS and phenolics (see below).

HPLC-Orbitrap LTQ-XL Analyses for the Annotation of GLS and Phenolics. The following chromatographic conditions were designed to maximize the simultaneous separation of both glucosinolates and phenolics but also to separate isobaric compounds which presented identical fragmentation and therefore an equivocal isomerization. HPLC analysis was performed using an Xselect column (CSH C18, 2.5 μ m 2.1 mm × 150 mm; Waters Corp) maintained at 30 °C using Agilent 1200 series LC system. The mobile phase consisted of solvent A (water with 0.1% formic acid, v/v) and solvent B (acetonitrile with 0.1% formic acid, v/v). The elution gradient was 0-3 min (2-9% of B), 3-12 min (9-17% of B), 12-13 min (17-20% of B), 13-24 min (20-32% of B), 24-29 min (32-90% of B), 29-30 min (90-2% of B), and held for column equilibration for the next analysis. The flow rate of the mobile phase was 0.4 mL min⁻¹, and the injection volume was 2.0 μ L. Mass spectrometry was performed using the LTQ-Orbitrap XL equipped with an ESI source operating in negative mode. Experimental settings were optimized, and the following parameters were applied: 3.0 kV, electrospray ionization spray voltage; 20 V, capillary voltage; 55 arbitrary units, sheath gas flow; 9 a.u., auxiliary gas pressure; capillary temperature, 350 °C; sweep gas flow rate, 0.0,; vaporizer and drying-gas temperature, 350 °C. The MS1 spectra were acquired as a profile in full scan mode and used for recording an accurate mass ranging from 250 to 1500 m/z, with 500 ms of scan time (with a resolution of 30 000 at m/z 200). The MS/MS scanning mode was set as a datadependent ms2 scan (dd-ms2). The high energy collision dissociation was 30 eV, and the cone voltage was at 35 V. The available spectral and chromatographic information were used for an extensive annotation of GLS and phenolic compounds using the software Xcalibur Qual Browser 2.2 SP1.48 (Thermo Fisher Scientific, see details in the Results and Discussion).

UPLC-UV-TQD for the Quantification of GLS and Phenolics. The quantification of 68 metabolites (including 36 glucosinolates and 32 phenolic compounds) was performed in the 1820 samples. Samples were extracted, and 2 μ L was injected in a UPLC ACQUITY system equipped with a column Acquity UPLC CSH C18, 1.7 μ m $(2.1 \text{ mm} \times 150 \text{ mm}; \text{Waters Corporation})$, coupled to a photodiode array detection and a triple quadrupole detector (TQD, Waters Corporation). The mobile phase consisted of a mixture of solvent A (water with 0.1% formic acid, v/v) and solvent B (acetonitrile with 0.1% formic acid, v/v) at a flow rate of 0.45 mL min⁻¹. The gradient program was as follow with 0.45 mL min⁻¹ flow rate: initially, 2–9% B from 0 to 3 min, 9-17% B from 3 to 12 min, 17-20% B from 12 to 13 min, 20-32% from 13 to 24 min, 32-90% from 24 to 29 min for washing the column, 90-2% B from 29 to 30 min, and held for column equilibration for the next analysis. The sample holder was kept at 4 °C during the analysis and the column temperature was 30 °C. The ESI source of the TQD was set in the negative mode to perform glucosinolate detection with conditions as follows: the

					Diagn	ostic ions [M-H] [_] M	S2 (m/z) (%	relative	ntensity)	Identifica	ion
		H-M	VIS1 (m/z)		Tunioral ione of GI S	-SGIc -(SO3) -GIy((+OH) -(SO3) -GIy	-GlycSGI SO3	-SGIc-		Comi evetomotio nomo	Trivial namo
₽	RT min (HPLC)	n RT mi) (UPLO	n Exp. Theo.	Err. (ppm)	i ypicar ioris or GL3	-178 -80 -162	: -242 -196	-223		Jeillesystematic name	
G01**	11.95	6.6	406,029 406,031	-4,8	259(100), 275(36), 195(9), 227(5), 291(2)	228(17) 326(14)	164(8) 210(8		388(6)[M-H-18]	3-methysulfanylpropyl-GLS**	Glucoiberverin**
G02	5.64	2.74	422,025 422,025	-2,2	259(9), 275(2), 291(1), 195(1), 227(<1), 241(<1)	244(<1) 342(1) 260(<	1) 180(<1)		358(100) M-H-64], 407(3)[M-H-15], 404(1)[M-H-18]	3-methysulfinylpropyl-GLS	Glucoiberin
G03	7.42	3.82	358,027 358,027	-2,0	259(100), 275(34), 195(9), 227(8), 241(4)	180(21) 278(8) 196(3) 116(4) 162(65) 135(3)	340(2)[M-H-18]	2-propenyl-GLS	Sinigrin
G04	15.8	9.17	420,043 420,046	-7,8	259(100), 275(53), 291(17), 227(10), 195(3)	242(22) 340(19)	178(10) 224(2	197(2)	402(4)[M-H-18]	4-methysulfanylbutyl-GLS	Glucoerucin
G05	6.69	3.18	436,039 436,041	-5,7	372(100), 259(4), 291(1), 275(<1), 195(<1), 227(<1)	258(2) 356(<1)	194(<1)		372(100)[M-H-64], 421(1)[M-H-15]*	4-methylsulfinylbutyl-GLS	Glucoraphanin
G06	10.56	5.52	372,041 372,043	-5,8	259(100), 275(46), 195(19), 241(7), 227(5), 291(3)	194(18) 292(11) 210(2) 130(7) 176(8	_	354(4)[M-H-18]	3-butenyl-GLS	Gluconapin
G07	6.79	3.4	388,035 388,038	-6,5	259(100), 275(18), 195(16), 291(12), 241(7)	210(22) 308(18) 226(3) 146(4) 192(5		332(99)[M-H-56]; 136(35)[M-H-252]; 370(14)[M-H-18] , 154(12)[M-H-234]; 388(9)[M-H]; 170(7)[M-H-218]	(R)-2-hydroxy-3-butenyl-GLS	Progoitrin
G08**	6.79	3.62	388,036 388,038	4,0						(S)-2-hydroxy-3-butenyl-GLS**	Epiprogoitrin**
G09**	19.24	12.87	434,060 434,062	-5,1	259(100), 275(59), 241(8), 195(6), 227(4)	256(14) 354(24)	192(26) 238(14) 211(10)		5-methysulfanylpentyl-GLS**	Glucoberteroin**
G10	8.21	3.98	450,054 450,057	-6,5	259(2), 227(1), 291(1), 195(<1)	272(<1)			386(100)[M-H-64], 435(3)[M-H-15], 432(2)[M-H-18]	5-methylsulfinylpentyl-GLS	Glucoalyssin
G11	14.48	8.36	386,056 386,059	-5,4	259(100), 275(24), 227(3),291(1)	208(13) 306(4)	144(5) 190(1	163(8)		4-pentenyl-GLS	Glucobrassicanapin
G12**	9.34	4.75	402,051 402,053	-6,0	259(100), 275(41), 241(7), 195(7), 291(4)	224(49) 322(12) 240(2) 160(8) 206(6	179(1)	332(36)[M-H-70], 136(14)[M-H-266], 170(2)[M-H-232] , 154(3)[M-H-248], 384(1)[M-H-18]	hydroxy-pentenyl-GLS**	Gluconapoleiferin**
G13**	24.56	16.93	448,077 448,078	-0,8	-		-	I	-	6-methysulfanylhexyl-GLS**	Glucoesquerellin**
G14	19.74	13.15	422,056 422,059	-6,6	259(100), 275(50), 227(8), 195(5)	244(18) 342(23) 260(6) 180(4) 226(9		404(24)[M-H-18], 163(16)[M-H-259], 332(15)[M-H-90] , 422(2)[M-H]	2-phenylethyl-GLS	Gluconasturtiin
G15*	12.44	7.44	438,052 438,053	4,0	259(81), 275(60), 291(20), 195(19)	260(45) 358(21)	196(11) 242(45	~	420(47)[M-H-18], 242(45)[M-H-178-18]	p-hydroxyphenylethyl-GLS*	Homosinalbin*
G16*	13.9	ø	438,052 438,053	-3,6	259(69), 275(13), 241(10), 227(6)	260(43) 358(8)	196(8) 242(9	215(6)	420(20)[M-H-18]; 136(11)[M-H-178-106-18]; 154(9)[M H-284]; 332(82)[M-H-106]; 242(9)[M-H-178-18]	(R)-2-hydroxy-2-phenetyl-GLS*	Epiglucobarbarin*
G17	14.37	8.32	438,052 438,053	-3,6	I	1	ı ı	I	I	(S)-2-hydroxy-2-phenetyl-GLS	Glucobarbarin
G18**	17.9	11.24	438,051 438,053	-4,9	259(100), 275(16), 291(3), 241(2)	260(10) 358(17) 276(2) 196(10) 242(2		420(27)[M-H-18]	m-hydroxyphenylethyl-GLS**	m-hydroxygluconasturtiin**
G19	18.08	11.5	447,051 447,054	-6,5	259(100), 275(30), 291(7), 241(3)	269(9) 367(9) 285(3) 205(23)			3-indolyImethyI-GLS	Glucobrassicin
G20*	24.03	16.89	477,062 477,064	4,2	259(4)	299(2)			446(100)[M-H-31]*, 447(55)[M-H-30]*, 459(9)[M-H-18]	1-methoxyindol-3-ylmethyl-GLS*	Neoglucobrassicin*
G21*	20.68	14.15	477,061 477,064	-5,9	259(100), 275(67), 195(7)	299(23) 397(13)	235(16) 281(4	254(4)	477(9)[M-H] ⁻ , 459(8)[M-H-18] ⁻ , 284(6)[M-H-162-31] ⁻	4-methoxyindol-3-ylmethyl-GLS*	4-Methoxyglucobrassicin*
G22**	13.73	7.75	463,047 463,049	4,7	259(10), 275(2)	285(100) 383(5) 301(1) 221(4) 267(9) 240(19)	160(8)[M-H-196-27-80]	4-hydroxyindol-3-ylmethyl-GLS**	4-Hydroxyglucobrassicin**
G23**	14.66	8.22	463,046 463,049	-6,0	259(36)	285(26.4) 383(7.4)	221(6.3)		445(17)[M-H-18], 446(7)[M-H-17]	1-hydroxyindol-3-ylmethyl-GLS**	1-Hydroxyglucobrassicin**
G24*	8.96	4.47	360,042 360,043	-3,2	259(17), 275(13)				342(100)[M-H-18]-, 358(4)[M-H-2H]-	n-propyl-GLS*	1
G25*	9.19	4.54	360,042 360,043	-3,2	259(87), 275(14)	280(35) 198(8	~		360(100)[M-H]-, 342(22)[M-H-18]-	1-methylethyl-GLS*	Glucoputranjivin*
G26*	12.57	6.62	374,057 374,059	-3,2	259(83), 227(6), 275(4)	196(15) 294(30) 212(4) 132(7)		356(17)[M-H-18]-, 331(4)[M-H-43]-•	2-methylpropyl-GLS* /lsobutyl-GLS*	Glucoconringianin*
G27*	12.89	6.72	374,057 374,059	-3,2	259(100), 275(43), 227(19), 195(9), 241(2)	196(7) 294(13)	132(9) 178(1	_		1-methylpropyl-GLS*	Glucocochlearin*
G28**	8.11	3.92	390,052 390,053	4,7	259(76), 275(57), 291(2)	310(10)	148(9)	167(2)	344(100)[M-H-46]-, 372(11)[M-H-18]-, 390(5)[M-H]-	2-hydroxy-2-methylpropyl-GLS**	Glucoconringiin**
G29**	8.9	4.02	390,052 390,053	4,1	259(100), 275(42), 291(9), 227(8)	212(18) 310(18)	148(14)	167(2)	344(64)[M-H-46]-, 372(18)[M-H-18]-, 390(2)[M-H]-, 354(2)[M-H-36]-	1-ethyl-2-hydroxyethyl-GLS**	Glucosisautricin**
G30**	17.19	10.6	388,073 388,074	-2,8	259(100), 275(41), 195(9), 241(9)	210(15) 308(8)	146(3) 192(15		370(16)[M-H-18]-, 388(8)[M-H]-	pentyl-GLS isomer**	I
G31*	21.79	14.83	402,088 402,090	4,6	259(100), 275(22), 227(9), 195(7)	224(14) 322(7)	160(2) 206(2	179(7)	384(20)[M-H-18]-, 402(3)[M-H]-	3-methylpentyl-GLS*	I
G32**	22.21	15.46	402,088 402,090	-3,4	259(100), 275(67), 241(10), 195(3)	224(34) 322(14)	206(3	_	384(37)[M-H-18]-, 402(7)[M-H]-	hexyl-GLS iso. I**	I
G33**	22.95	15.75	402,088 402,090	-3,4	259(62), 275(32), 241(14),195(12)	224(14) 322(2)	206(2	_	384(18)[M-H-18]-	hexyl-GLS iso. II**	1
G34**	26.7	18.93	416,104 416,105	4,6	259(100), 195(16), 227(11)	336(21)			398(11)[M-H-18]-	heptyl-GLS iso. I**	1
G35**	27.02	19.3 22	416,104 416,105	, 3,2 9	259(100), 275(32), 227(8), 291(6), 195(4), 241(4)	238(26) 336(9)	174(9) 220(24) 193(5)	398(11)[M-H-18]-	heptyl-GLS iso. II**	1
G36	21.24	ZU	416,104 416,100	-3,2	259(100), 275(36), 227(9), 195(9), 291(b)	238(11)	1/4(6) 220(0		416(6)[M-H]-, 398(3)[M-H-18]-	heptyl-GLS iso. III**	1

Table 1. continued

stands for the neutral structure. GLSs annotated with two asterisks (**) were tentatively identified only from daughter anions and accurate parent mass; those with one asterisk (*) were further validated with certified reference plant materials; and those remaining without an asterisk were confirmed with authentic reference compounds.



Figure 1. Structures of glucosinolates found in *Brassica napus*. Numeration of GLSs follows subcategories G01–G13, Met-derived aliphatic-GLSs; G14–G17, Phe/Tyr-derived phenylalkyl-GLSs; G18–G23, tryptophan-derived indolic-GLSs; and G24–G36, branched and linear alkyl-GLSs. Alkyl-GLSs derived from unknown amino acids are not represented due to their unknown structure and multiple possibilities of branched isomers. Only m/z for molecular anions are represented. Please see Table 1 and Supporting Results 1.

capillary voltage was maintained at 2.9 kV, source and desolvation temperatures were set at 150 and 400 °C, respectively. Pure analytical standards of glucosinolates were used to prepare the calibration standard stock methanol solution, stored at -25 °C. This stock solution was diluted as needed each day for making a 5-point calibration curve. This external calibration gives a linear regression for each glucosinolate and was checked before the quantification of the samples. Total data acquisition was controlled using Masslynx (version 4.1) software and preprocessing with the Quanlynx program (Waters Corporation). Each of the 32 GLS analyzed was quantified using the calibration curve of the authentic GLS standard with the

closest chemical structure (details in Supporting Table S2). In parallel to GLS analysis, the phenolic compounds were measured through UV absorbance at 330 nm and quantified using external calibration with the following commercially available flavonols and hydroxycinnamic acids (HCA) (kaempferol glucose, quercetin glucose, isorhamnetin glucose, sinapic acid, coumaric acid, and chlorogenic acid). From all UV-chromatograms, peak deconvolution was preprocessed using Empower (version 2). The quantification of each 36 phenolic compounds was based on the authentic standard with the most closely related aglycone structure. To avoid an overestimated quantification of multiacylated molecules, a theoretical UV-response coefficient was calculated for each compound by adding the UV-responses of authentic standards corresponding to flavonol and acyl building moieties of each molecule (*i.e.*, kaem-O-Disin-O-dihex was calibrated using the addition of response coefficients of kaem-O-glucose and twice those of sinapic acid) (Supporting Table S2). This approach was validated using a calibration curve of the acylated kaempferol derivative tiliroside (from Sigma).

Multivariate Analyses. Analyses were performed using software R and interface RStudio (R Core Team, 2021). Variables with unquantified values representing more than 5% of the data set were removed from the following analyses. Normalized data sets were mean-centered using the least-squares method and a linear model including biological replicates using the "emmeans" R package.⁴⁹ Mean-centered values of each compound were visualized on boxplots and comparisons between species were evaluated using Tukey's tests. Multivariate statistical analyses were performed to describe metabolic profile diversity. PCAs were made with "gridextra" and "ade4" R packages, ⁵⁰ heatmaps were performed with the "pheatmap" R package, and hierarchical clustering were achieved using the "NbClust" R package.⁵¹ All visualization plots were realized using the "ggplot2" R package.⁵²

RESULTS AND DISCUSSION

Identification of 36 Glucosinolates in Brassica Leaves and Roots. GLS annotation was based on the corpus of analytical information about all GLS previously reported in Brassicaceae species,⁵³⁻⁵⁵ including high-resolution mass spectrometry (HRMS) and mass fragmentation patterns.⁵⁶⁻⁵ In Table 1 are listed 36 ion features (most of which were observed in both root and leaf extracts) that matched (<10 ppm) with theoretical m/z from published GLS data and with the spectral signature expected from compounds containing one or two nitrogen atoms and two or three sulfur atoms (Supporting Table S1). The GLS structure of those compounds was then corroborated (Table 1) with typical GLS-related diagnostic ions (at m/z 195, 227, 241, 259, 275, and 291).⁵⁶⁻⁵⁹ Neutral losses (80, 162, 178, 196, 242, and 223 Da), which correspond to sulfur-trioxide, anhydroglucose, gluconolactone, thioglucose, sulfur trioxide + anhydroglucose, hydrogen-cyanide + thioglucose, respectively, gave further indications on the fragmentation and structure of R-groups (Table 1). The main results from these annotations are detailed below, and extensive explanations for each 36 individual GLS are given in Supporting Results S1. Eleven of these compounds were confirmed by the retention time of authentic standards. An additional series of GLSs were further validated by identical retention time with compounds from reference plant materials and were indicated with one asterisk (*) (Supporting Results S3). For the rest of GLSs (with two asterisks), the proposed annotation was based on daughter anions and accurate parent mass (Table 1, see also Supporting Results S1).

Thirteen aliphatic methionine-derived GLS were found (from peak G01** to peak G13**, 3-methysulfanylpropyl-GLS**, 3-methysulfinylpropyl-GLS, 2-propenyl-GLS, 4-methysulfanylbutyl-GLS, 4-ethylsulfinylbutyl-GLS, 3-butenyl-GLS, (R)-2-hydroxy-3-butenyl-GLS, (S)-2-hydroxy-3-butenyl-GLS**, 5-methysulfanylpentyl-GLS**, 5-methylsulfinylpentyl-GLS, 4-pentenyl-GLS, hydroxy-pentenyl-GLS**, and 6-methysulfanylhexyl-GLS** (Figure 1). This series of glucosinolates has been widely reported in *Brassicaceae* and was especially expected in leaves and roots of *Brassica* species.

Five phenylalanine/tyrosine-derived phenylalkyl-GLSs (peak G14–G18**) were identified (Figure 1). Peak G14 (m/z 422.056 $[M - H]^-$) was identified as gluconasturtiin, and

confirmed with an authentic reference compound. The peaks G15*-G18** were four isobaric anions of m/z 438.053 [M -H]⁻, putatively corresponding to four hydroxyl-derivatives of phenylalkyl-GLSs. Peak G15* was annotated as parahydroxyphenethyl-GLS* (= p-hydroxygluconasturtiin* = homosinalbin*) and confirmed using an extract of Arabis soyeri, a documented source for this compound⁶⁰ (Supporting Results S3). For peak G16*, the chromatographic elution pattern of root and seed extracts of Barbarea vulgaris P-type⁶² (classic source of epiglucobarbarin*) allowed the validation of this compound as (R)-2-hydroxy-2-phenethyl-GLS* (Table 1, Supporting Results S1 and Supporting Figure S3). Peak G17 was proposed as (S)-2-hydroxy-2-phenethyl-GLS and validated by the retention time of an authentic reference of Glucobarbarin and an extract from Barbarea vulgaris G-type (Table 1, Supporting Results S1 and Supporting Figure S3). The presence of this compound in B. napus has been already observed in at least one previous work on B. rapa and B. oleracea.45 Peak G18** was proposed as meta-hydroxyphenethyl-GLS** (= m-hydroxygluconasturtiin**) (Figure 1, Table 1, Supporting Results S1). Although Wang and coworkers⁶¹ recently suggested the presence of *m*-methoxybenzyl-GLS in B. napus, we did not find this compound in our samples compared to an extract of Aubrieta deltoidei, which is a classic source (Supporting Result S3). We can however not fully rule out the hypothesis proposed by Blazevic et al., that G18** in B. napus could be an isoGLS, i.e., an isobaric GLS derived from "exotic" hexoses. To our knowledge, this is the first time that these four hydroxyl derivatives of phenylalkyl-GLSs are described in Brassica species, although they have been previously reported in Arabis (tribe Arabideae)⁶⁰ and Barbarea (tribe Cardamineae) genus.

Five tryptophan-derived indolic-GLSs (peak G19-G23**) were recorded (Figure 1, Table 1, Supporting Results S1). The identification of peak G19 as 3-indolylmethyl-GLS (glucobrassicin, m/z 447.051 [M - H]⁻) was validated by an authentic standard. The peaks G20* and G21* were isobaric anions of m/z 477.062 $[M - H]^-$ and were putatively assigned as neoglucobrassicin* and 4-methoxyglucobrassicin* based on mass fragmentation profiles (Supporting Results S1). This was confirmed (Supporting Results S3) by the comparison with a root extract of Barbaria vulgaris P-type.⁶² Both compounds have been abundantly reported in Brassica species.⁶³⁻⁶⁷ Another pair of isobaric anions of m/z 463.047 $[M - H]^{-}$ was putatively annotated as 4-hydroxyglucobrassicin** (G22**) and 1-hydroxyglucobrassicin** (G23**), respectively (see details in Supporting Results S1). Those indolic-GLS have been commonly reported in Brassica species. 63,65

Thirteen branched and linear alkyl-GLSs (G24-G36) were annotated. It contained one alanine-derived alkyl-GLS ($G24^{**}$), six BCAA-derived alkyl-GLSs ($G25^*-G29^{**}$ and $G31^*$), six alkyl-GLSs derived from unknown amino acids ($G30^{**}$, $G32^{**}-G36^{**}$) (Table 1, details of the annotations are in Supporting Results S1).

G24** was annotated as *n*-propyl-GLS** (Table 1, details of the annotations are in Supporting Results S1), unprecedentedly reported in any plant.⁶⁸ This GLS might be derived from alanine with two chain elongations or possibly methionine derived in a complex way.

Peak **G25**^{*} was putatively identified as isopropyl-GLS^{*} (glucoputranjivin^{*}), a valine derivative, and this annotation was validated by comparison with a *Cardamine diphylla* seed extract as reference material.⁶⁹ Peaks **G26**^{**} and **G27**^{*} were

Article



Figure 2. UPLC-UV chromatogram of a leaf extract (QC). Annotated phenolic compounds are numbered according to the different phenolic subcategories (see in the text, and Tables 2, 3, and 4).

Table	2.	Identification	of Pheny	vlpro	panoids [•]
I WOIC	~.	rachtenteuton	or r nen	, , , , , , , , , , , , , , , , , , , ,	pulloiuo

						D	iagnostic	ions [M	-H] ⁻ MS2	(m/z) (%	relative intensity)	
	[M-	Hjm	51 (m/z)		-Hexosyl	-diHexosyl	-Feruloyl	-Ferulic acid	-Sinapoyl	-Sinapic acid	Otherions	Identification
ID	RT min (HPLC)	RT min (UPLC)	Exp. Theor.	Err. (ppm	-162	-324	-176	-194	-206	-224	Outer tons	
PL01	10.22	5.03	353.086 353.088	-5.7	-	-	-	-	-	-	-	Caffeoylquinic acid
PL02**	13.67	7.48	367.102 367.103	-3.7			191(1)	173(4)			193(100)[M-H-174] ⁺ , 134(4)[M-H-174-44-15] ⁺ , 149(1)[M-H-174-44] ⁻	Feruloylquinic acid**
PL03**	12.97	7.03	355.102 355.103	-5.5	193(100)						175(31)[M-H-162-18] [*] , 337(3)[M-H-H2O] [*] , 178(1)[M- H-162-15] [*] , 149(1)[M-H-162-44] ⁻	Feruloyl-O-hexoside**
PL04**	13.36	7.26	385.112 385.114	-6.0	223(100)				179(<1)		205(45)[M-H-162-18] ⁻ , 190(2)[M-H-162-18-15] ^{-*} , 367(1)[M-H-H2O] ⁻ , 208(<1)[M-H-162-15] ^{-*}	Sinapoyl-O-hexoside**
PL05**	11.88	6.04	547.165 547.167	-3.4		223(100)				323(9)	208(9)[M-H-162-162-15] [•] , 190(1)[M-H-324-18-15] [•]	Sinapoyl-O-dihexoside**
PL06**	24.86	17.63	693.201 693.204	-4.5				499(100)			675(<1)[M-H-H2O] ⁻	Diferuloyl-O-dihexoside**
PL07**	23.32	15.87	753.221 753.225	-5.2					547(1)	529(100)	223(3)[M-H-162-224-162] ⁻ , 735(<1)[M-H-H2O] ⁻	Disinapoyl-O-dihexoside, isomer I**
PL08**	23.82	16.99	753.223 753.225	-3.0					547(<1)	529(100)	223(2)[M-H-162-224-162] ⁻ , 735(<1)[M-H-H2O] ⁻	Disinapoyl-O-dihexoside, isomer II**
PL09**	24.03	17.02	723.210 723.214	-5.2			547(<1)	529(100)		499(1)	223(2)[M-H-176-206-162] ⁻	Feruloyl-O-sinapoyl-O-dihexoside, isomer I**
PL10**	24.24	17.17	723.210 723.214	-5.9				529(<1)	517(<1)	499(100)	223(1)[M-H-176-162-162] ⁻ , 385(<1)[M-H-176-162] ⁻ , 705(<1)[M-H-H2O] ⁻	Feruloyl-O-sinapoyl-O-dihexoside, isomer II**
PL11**	24.67	17.58	723.211 723.214	-4.6	561(1)			529(2)	517(2)	499(100)	223(1)[M-H-176-162-162] ⁻	Feruloyl-O-sinapoyl-O-dihexoside, isomer III**
PL12**	26.27	18.7	959.280 959.283	-3.2					753(3)	735(100)	511(8)[M-H-224-224] ⁻ , 529(6)[M-H-224-206] ⁻ , 941(2)[M-H-H2O] ⁻ , 385(1)[M-H-162-206-206] ⁻	Trisinapoyl-O-dihexoside**
PL13**	26.97	19.55	929.269 929.272	-3.8			753(1)	735(1)	723(5)	705(100)	511(7)[M-H-224-194], 481(2)[M-H-224-224], 385(1)[M-H-162-206-176], 499(1)[M-H-224-206]	Feruloyl-O-disinapoyl-O-dihexoside**

 a^{a} [M] stands for the neutral structure. phenolics annotated with two asterisks (**) were tentatively identified only from daughter anions and accurate parent mass, and the phenolics remaining without an asterisk were confirmed with authentic reference compounds. Additional details are given in Supporting Table S1.

putatively identified as 2-methylpropyl-GLS** (glucoconringianin** = isobutyl-GLS**) and 1-methylpropyl-GLS* (glucocochlearin*), respectively (Table 1, Supporting Results S1). 2-Methylpropyl-GLS** was previously convincingly characterized in other *Brassicales* based on spectroscopic data (NMR and MS).^{54,55,70,71} Peak G27* was confirmed by comparison with a leaf extract from *Cochlearia officinalis*, the species where it was initially characterized.⁷² This compound was also detected in leaves of *B. rapa*¹⁶ and *B. napus*.⁷³ Peaks G28** and G29** were annotated as 2-hydroxy-2-methylpropyl-GLS** (glucoconringiin**) and 1-ethyl-2-hydroxyethyl-GLS** (glucosisautricin**), previously reported in *B. juncea.*⁴⁶ Peak **G31*** was annotated as 3-methylpentyl-GLS* by comparison with an extract from *Cardamine pratensis.*⁷⁴ Peak **G30**** was tentatively annotated as pentyl-GLSisomer** (ramification unknown). Peaks **G32**** and **G33**** were putatively annotated as hexyl-GLS** isomers (ramification unknown) and peaks **G34**–G36**** as heptyl-GLS** (ramification unknown). Interestingly, a series of 3-methylbutyl-GLS, 4-methylpentyl-GLS, *n*-pentyl-GLS, *n*-hexyl-GLS,

Table 3. Identification of Nonacylated Hexosylated Flavonols^a

	[]	и-н]. м	IS1 (m/	z)		D	iagnostic i	ons [M-H] ⁻	MS2 (m/z) (% relative intensity)	
						-Hexosyl	-diHexosyl	-triHexosyl		Identification
ID	RT min (HPLC)	RT min (UPLC)	Exp.	Theor.	Err. (ppm)	-162	-324	-486	Other ions	
PL14**	19.22	13.23	609.144	609.146	-3.4	447(16)	285(100)		429(98)[M-H-180] [*] , 284(67)[Y0-H] [*] , 327(7)M-H- 120-162] [*] , 489(7)[M-H-120] [*]	Kaempferol-O-dihexoside**
PL16**	11.26	5.25	771.196	771.199	-4.5	609(100)	447(<1)		753(<1)[M-H-H2O] ⁻	Kaempferol-O-trihexoside, isomer I**
PL17**	14.69	8.04	771.197	771.199	-2.8	609(100)	447(4)	285(4)	651(2)[M-H-120] ⁻ , 753(<1)[M-H-H2O] ⁻	Kaempferol-O-trihexoside, isomer II**
PL28**	10.37	4.86	787.189	787.194	-5.1	625(100)	463(1)	301(<1)	300(<1)[Y0-H] , 769(<1)[M-H-H2O]	Quercetin-O-trihexosie**
PL32**	15.22	8.52	801.207	801.209	-3.4	639(100)	477(5)	315(3)	783(<)[M-H-H2O] ⁻ , 300(<)[Y0-H-15] ⁻	Isorhamnetin-O-trihexoside**

 a [M] stands for the neutral structure. Phenolics annotated with two asterisks (**) were tentatively identified only from daughter anions and accurate parent mass, and Phenolics remaining without an asterisk were confirmed with authentic reference compounds. Additional details are given in Supporting Table S1.

Table 4. Identification of Acylated Flavonoid Hexosides in Leaves of 304 Accessions of Brass
--

								Diagnos	tic ions [l	M-H] ⁻ MS2	(m/z) (% r	elative intensity)	
			[M-H] ⁻ (m/z)			-Hexosyl	-diHexosyl	-triHexosyl	-Sinapoyl	-Feruloyl	-OHFeruloyl	Other ions	Identification
ID	RT min (HPLC)	RT min (UPLC)	Exp.	Theor.	Err. (ppm)	-162	-324	-486	-206	-176	-192		
PL15**	21.6	15.03	785.190	785.193	-4.5	623(87)	461(2)			609(100)		591(49)[M-H-194]", 284(3)[Y0-H]", 592(3)[M-H-193]", 285(2)[Y0]", 443(1)[M-H-162-180]"	Kaempferol-O-feruloyl-O-dihexoside**
PL18**	19.95	13.9	771.174	771.178	-4.3	609(100)	447(<1)					429(3)[M-H-162-162-18]`, 285(2)[Y0]`, 591(2)[M-H-180]`, 489(<1)[M-H-162-120]`, 284(<1)[Y0-H]`	Kaempferol-O-caffeoyl-O-dihexoside**
PL19**	19.28	12.79	801.186	801.188	-3.0						609(100)	623(7)[M-H-178]7, 610(7)[M-H-191]7, 285(2)[Y0]7, 284(1)[Y0 H]*, 447(<1)[M-H-162-194] ⁻	Kaempferol-O-hydroxyferuloyl-O-dihexoside**
PL20**	20.3	14.07	815.201	815.204	-3.0				609(100)			591(38)[M-H-224] ⁻ , 429(5)[M-H-162-224] ⁻ , 285(3)[Y0] ⁻ , 284(2)[Y0-H] ^{-*} , 447(1)[M-H-162-206] ⁻	Kaempferol-O-sinapoyl-O-dihexoside**
PL21**	12.61	6.28	963.238	963.241	-3.5	801(100)						609(1)[M-H-192-162] ⁻ , 285(<1)[Y0] ⁻	Kaempferol-O-hydroxyferuloyl-O-trihexoside**
PL22**	15.08	8.38	947.242	947.246	-5.1	785(100)	623(<1)					609(2)[M-H-176-162], 929(0.27)[M-H-H2O]	Kaempferol-O-feruloyl-O-trihexoside**
PL23**	14.41	7.85	977.252	977.257	-5.5	815(100)			771(<1)			609(2)[M-H-206-162], 959(<1)[M-H-H2O], 753(<1)[M-H- 224], 591(<1)[M-H-224-162],	Kaempferol-O-sinapoyl-O-trihexoside**
PL24**	15.02	8.25	1 109.295	1109.299	-4.4	947(2)	785(100)	623(2)		933(<1)		609(22)(M-H-176-162-162], 1109(21)(M-HJ, 915(1)(M-H- 194], 429(1)(M-H-176-162-162-162-18], 1091(<1)(M-H- H2OJ, 771(<1)(M-H-176-162]	Kaempferol-O-feruloyl-O-tetrahexoside**
PL25**	14.34	7.7	1 139.305	1139.310	-4.0	977(1)	815(100)		933(<1)			609(21)[M-H-206-162-162], 591(3)[M-H-224-162-162], 915(1)[M-H-224], 429(1)[M-H-224-162-162-162], 1121(<1)[M-H-H2O], 447(<1)[M-H-206-162-162-162],	Kaempferol-O-sinapoyl-O-tetrahexoside**
PL26**	14.05	7.48	650.176 [M-2H] ²⁻	650.177 [M-2H] ²⁻	-2.4	569(7)[M-2H] ²⁻	488(86)[M-2H] ²⁻					977(100)[M-H-324] ⁻ , 245(3)[M-2H-(5*162)] ² , 323(2)[M-H- 285-162-162-206-162] ⁻	Kaempferol-O-sinapoyl-O-pentahexoside**
PL27**	20.22	14.26	753.203 [M+2H] ²⁻	753.206 [M-2H] ²⁻	-3.4		591(100)[M-2H] ²					1183(26)[M-H-324] ⁻ , 977(7)[M-H-206-324] ⁻ , 650(6)[M-2H- 206] ² , 1301(5)[M-H-206] ⁻ , 488(1)[M-2H-206-162-162] ² , 245(1)[M-2H-206-(162 x5)] ² , 285(1)[Y0] ⁻	Kaempferol-O-disinapoyl-O-pentahexoside**
PL29**	11.99	5.85	949.222	949.226	-3.4	787(100)	625(9)					769(<1)[M-H-180] ⁻ , 607(<1)[M-H-162-180] ⁻ , 949(<1)[M-H] ⁻	Quercetin-O-caffeoyl-O-trihexoside**
PL30**	11.72	5.63	979.255	979.236	-3.9	817(100)					787(87)	625(27)[M-H-192-162] ⁻ , 639(2)[M-H-162-162] ⁻	Quercetin-O-hydroxyferuloyl-O-trihexoside**
PL31**	20.79	14.19	1 361.359	1361.362	-2.3	1199(<1)						1361(100)[M-H] ⁻	Quercetin-O-disinapoyl-O-tetrahexoside**

"[M] stands for the neutral structure. Phenolics annotated with two asterisks (**) were tentatively identified only from daughter anions and accurate parent mass, and Phenolics remaining without an asterisk were confirmed with authentic reference compounds. Additional details are given in Supporting Table S1.

and *n*-heptyl-GLS has already been detected in *B. rapa* based on the detection of their isothiocyanate derivatives.⁷⁵

Based on these annotations, a UPLC-PDA-TQD method was developed for the quantification of individual GLS, using an optimized gradient of 30 min enabling the simultaneous resolution of phenolic compounds (see the Materials and Methods).

Thorough Annotation of 32 Compounds as Main Phenolics in Leaf Extracts. UPLC-based UV-chromatograms from leaves displayed well-resolved peaks in leaves (Figure 2). UV-spectra, MS, and MS/MS spectra underlying each of those peaks were combined with high-resolution data from HPLC-ESI LTQ-Orbitrap, thus leading to the identification of 32 well-resolved phenolics (Figure 2) which were numbered following their categorization into different phenolic subclasses. The occurrence of these compounds in *Brassica* species and their fragmentation patterns were often confirmed by the literature. UV-based quantification was not possible for

pubs.acs.org/JAFC



Figure 3. Common aglycone structures of phenolics found in Brassica species: (A) phenylpropanoids (B) and flavonols.

a few UV peaks for which m/z spectra revealed multiple chromatographically unresolved phenolic compounds. The numbered peaks including acyl-quinic acids, hydroxycinnamic acid hexoside esters, and hexosylated and acylated flavonols are summarized in Tables 2-4 and Figure 3. The main flavonoids present in this Brassica panel were O-hexosides of kaempferol, quercetin, and isorhamnetin, which were consistent with previous works.^{26,36,44,76,77} These studies have also reported many Brassica phenolic compounds and their fragmentation pattern, including multihexosylated and acylated. The hexosylation was characterized by the multiple neutral loss of m/z 162 (loss of anhydrohexose). Phenolic compounds yielded characteristic neutral losses of acyl groups, the loss of m/z 162, 146, 192, 194, 206, and 204 were assigned to caffeoyl, coumaroyl, feruloyl, hydroxyferuloyl, ferulic acid, sinapoyl, and sinapic acid groups (Tables 2-4 and Supporting Results S2). In a similar way to GLS, we also frequently observed dehydration of phenolic compounds in fragmentation patterns. For each analyte, fragmentation propensities were deduced from the daughter ion's profiles and applied for structural identification using previously reported patterns of phenolic fragmentations.^{26,36,44,76,77} Three propensities of phenolics fragmentation were deduced from (1) the possible occurrence of deacetylated and/or dehexosylated moieties of acylated and/or hexosylated phenolics frequently associated with dehydration, (2) possible dehydration of phenolics due to the hydroxyl group, (3) then demethylation followed by a decarboxylation or vice versa. The main results from these annotations are detailed below, in Tables 2-4, and extensive explanations for all 32 phenolic compounds are given in Supporting Results S2.

Phenylpropanoids. Thirteen phenylpropanoids belonging to acyl-quinic acids (peaks **PL01** and **PL02****) and hydroxycinnamic acid hexoside esters (peaks **PL03** to **PL13**) were annotated (Table 2, Supporting Results S2). Among these compounds, we found some isomers chromatographically resolved (i.e., duo peaks, **PL07**-PL08****, and trio peaks, **PL09**-PL10**-PL11****). Succinctly, **PL01** was authentified by the standard 5-caffeoylquinic acid (5-CQA) and **PL02****, as feruloyl-O-quinic acid** (Table 2). The presence of these two acyl-quinic acids is consistent with earlier research in *Brassica* species.^{38,47,78} Peak **PL03**** and peak **PL06**** were identified as feruloyl-O-hex** and diferuloyl-O-dihex** (dimers of feruloyl-O-hex), respectively (Table 2), already reported in *Brassica.*^{38,47} Peak **PL04**** and peak **PL05**** were designated as sin-O-hex** and sin-O- dihex**, respectively. Unlike **PL05****, **PL04**** was reported in *B. rapa*³⁷ and described as being among the most abundant hydroxycinnamic acids of *B. oleracea* and *B. napus* leaves.²⁶ Disinapoyl-O-dihex** and trisinapoyl-O-dihex** were assigned to peak **PL07**** and **PL08****, respectively, which have been previously reported in cabbage and turnip.^{26,37,38,47,78} Peaks **PL09****, **PL10****, and **PL11**** were nominated as feruloyl-O-sinapoyl-O-dihex**. Although this structure has been identified earlier in cabbage and oilseed rape,^{24–26,38,47} heretofore no structural isomers have been reported. Peak **PL12**** was characterized as trisin-O-dihex**, consistent with a previous study on *Brassica*.^{24,26,37,38,47,78} Peak **PL13**** was earmarked as feruloyl-di-O-sin-O-dihex** and already recognized in *Brassica* species.^{24,26,37,38,47,78}

Flavonols. We detected five nonacylated hexosylated flavonols (Table 3, Supporting Results S2). Peak PL14** was identified as kaem-O-dihex**, as previously detected in B. rapa.⁴⁰ Peak PL16** and PL17** were proposed as kaem-Otrihex isomers I** and II**, respectively, consistent with previous reports on B. oleracea var. botrytis and var. sabellica.^{27,36,40} Peak PL28** was annotated as quer-Otrihex**, already described in some Brassica species.^{26,27,36,40} PL32** was tentatively proposed as isor-O-trihex** described in Brassica species.^{78,79'} Fourteen flavonols (peak PL15, PL19-PL27, and PL29-PL31) were detected, and almost all were monocharged except peaks PL26** and PL27** (Table 3 and Supporting Results S2). The peaks PL29** and PL31** were acylated hexosylated quercetin derivatives and the remaining were exclusively kaempferol derivatives. Briefly, peaks PL15**, PL22**, and PL24** were provisionally assigned as kaem-Ofer-O-dihex**, kaem-O-fer-O-trihex**, and kaem-O-fer-Otetrahex**, respectively, as described by previous studies in *B. napus*^{25,30} and *B. oleracea* and *B. rapa*.^{24,27,36–38,40,41,47} Peaks PL19** and PL21** exhibited aglycon ions of kaempferol $(285 [Y0]^{-} and 284 [Y0 - H]^{-})$ and hydroxyferuloyl moieties, allowing their provisional identification as kaem-O-hydroxyferuloyl-O-dihex** and kaem-O-hydroxyferuloyl-O-trihex**, respectively, which was in line with the reported literature of Brassica crops.^{24,27,30,36,38,47} Peak PL18** was tentatively defined as kaem-O-caf-O-dihex** comparably with anterior studies.^{26,27,30,37,47} Peaks PL20**, PL23**, and PL25** were annotated as kaem-O-sin-O-dihex**, kaem-O-sin-O-trihex**, and kaem-O-sin-O-tetrahex**, as supported by previous *Brassica* works.^{25-27,30,36-38,41,47} Peaks PL26** and PL27** yielded double-charged parents which were annotated as kaem-O-sin-O-pentahex** and kaem-O-Disin-O-pentahex**.



Figure 4. Histograms of total glucosinolate concentrations in roots (A) and leaves (B) and phenolics in leaves (C) in the panel of 304 *Brassica* accessions. The *x*-axis represents all varieties ordered decreasingly according to total glucosinolates and phenolic compounds concentrations for each species. Each bar represents the mean concentration for one variety, from three independent biological replicates.

Although **PL26**^{**} was to our knowledge never reported in *Brassicaceae*, **PL27**^{**} was previously detected in leaves of *B. oleracea* curly kale.²⁴ The remaining peaks **PL29**^{**}, **PL30**^{**}, and **PL31**^{**} were characterized as quer-O-caf-O-trihex^{**}, quer-O-hydroxyferuloyl-trihex^{**}, and quer-O-Disin-O-tetrahex^{**}, all of which have been previously described in *Brassica* species.^{24,26,36,37,47}

Phytochemical Diversity and Concentration Ranges in 304 Brassica Accessions. Quantitative profiling of the above detailed 36 GLS and 32 phenolics was performed on a set of 1820 root and leaf samples from 304 accessions (3 biological replicates). The whole data set is given in Supporting Tables S3, S4, and S5 for root GLS, leaf GLS, and leaf phenolics, respectively. Supporting Tables S3a, S4a, and S5a show data of the three replicates. Supporting Tables S3b, S4b, and S5b give mean values, min, max, and percentiles S% and 95% for each accession. Supporting Tables S3c, S4c, and S5c allow, for any metabolic compound, to screen for genotypes harboring among the 5% lowest or highest mean values. These tables also allow rapid visualization of the peculiar metabolic features of any given genotype. This constitutes a unique and easily accessible phytochemical atlas for *Brassica* scientists. In the following parts, the most striking features are presented.

Glucosinolates. Figure 4 and Table 5 give key quantitative results and propose a global overview of phytochemical profiles among accessions of the three species. *B. napus* contained on average 2.5 and 26 μ mol g⁻¹ DW of glucosinolates in leaves and roots, respectively. Total GLS in leaves were mainly composed of Met-derived aliphatic-GLS (on average 74%), as seen in Figure 4 and as detailed in Table 5. This mean value encompasses striking interaccession variations. This diversity partly reflected the presence in the panel of both ancient and recent accessions which have been selected for a low aliphatic

Table 5. Descriptive Statistics of Major Categories of Glucosinolates and Phenolic Compounds Quantified in Leaves and Roots of *Brassica* Accessions^a

			Lea	f GLS		
	B. n.	apus	В.	rapa	B. ol	leracea
	mean	min - max	mean	min - max	mean	min - max
Leaf GLS (nmol.g ⁻¹ DW)	2 508.6	238 - 7384	6 905	320 - 16455	4 420	1425 - 9706
% Aliphatic-GLS / Total GLS	73.9	11 - 97	81.2	25 - 99	85.7	74 - 97
% Phenylalkyl-GLS / Total GLS	12.0	0 - 58	10.6	0.7 - 38	5.7	1.6 - 12
% Indolic-GLS / Total GLS	11.1	0 - 63	7.3	0.4 - 41	4.0	0.4 - 12
% Alkyl-GLS / Total GLS	3.0	0 - 41	0.9	0.1 - 3.4	4.5	0.3 - 16
% C3 / Aliphatic-GLS	0.2	0 - 4	0.1	0 - 0.2	44.1	0.3 - 89
% C4 / Aliphatic-GLS	45.7	22 - 98	63.9	23 - 99	55.6	10 - 99
% C5 / Aliphatic-GLS	54.0	1 - 78	36.0	1.1 - 77	0.3	0 - 1.1
% C6 / Aliphatic-GLS	0.1	0 - 3	0.0	0 - 0.2	0.0	0 - 0.1
% Methylsulfanyl- / Aliphatic-GLS	0.1	0 - 6	0.1	0 - 0.3	0.0	0 - 0.1
%Methylsulfinyl- / Aliphatic-GLS	21.2	1 - 65	7.2	0.1 - 30	54.6	5 - 86
%Alkenyl- / Aliphatic-GLS	43.8	1 - 80	77.2	15 - 100	21.2	0 - 64
% Hydroxyalkenyl- / Aliphatic-GLS	34.9	10 - 71	15.5	0.2 - 60	24.1	5 - 43
% Gluconasturtiin / Phenylalkyl-GLS	92.1	41 - 99	89.1	76 - 96	93.0	80 - 99
% Epiglucobarbarin / Phenylalkyl-GLS	2.3	0 - 10	2.8	1.1 - 4.5	1.4	0.3 - 2.7
% Neoglucobrassicin / Indolic-GLS	4.5	0 - 24	4.1	1.1 - 11	8.2	1.8 - 26
% 4M-I3MG / Indolic-GLS	4.8	0 - 45	2.5	1 - 5.5	10.4	2.1 - 51

			Root	GLS		
	B. n	apus	B. I	rapa	B. ol	eracea
	mean	min - max	mean	min - max	mean	min - max
Leaf GLS (nmol.g ⁻¹ DW)	26 023.7	5746 - 49463	24 861.8	13545 - 39942	21 611.4	13328 - 41869
% Aliphatic-GLS / Total GLS	37.2	7 - 80	31.1	8 - 60	62.2	51 - 87
% Phenylalkyl-GLS / Total GLS	46.9	2 - 77	53.1	29 - 77	23.2	1 - 36
% Indolic-GLS / Total GLS	15.1	5 - 65	14.9	8 - 22	13.7	9 - 20
% Alkyl-GLS / Total GLS	0.8	0 - 6	0.9	0 - 3	0.9	0 - 4
% C3 / Aliphatic-GLS	0.2	0 - 1	0.1	0 - 0	35.7	0 - 83
% C4 / Aliphatic-GLS	62.2	36 - 97	65.0	34 - 98	62.9	16 - 98
% C5 / Aliphatic-GLS	35.6	1 - 62	32.2	2 - 64	0.4	0 - 1
% C6 / Aliphatic-GLS	2.0	0 - 29	2.7	1 - 8	1.0	0 - 2
% Methylsufanyl- / Aliphatic-GLS	28.9	5 - 59	13.8	6 - 26	23.1	3 - 47
%Methylsulfinyl- / Aliphatic-GLS	39.8	23 - 65	31.3	15 - 60	38.4	23 - 58
%Alkenyl- / Aliphatic-GLS	7.1	0 - 23	35.6	5 - 78	25.8	1 - 59
% Hydroxyalkenyl- / Aliphatic-GLS	24.2	2 - 50	19.3	0 - 55	12.7	0 - 32
% Gluconasturtiin / Phenylalkyl-GLS	99.8	98.9-99.9	99.7	99.7-99.8	99.8	99.7-99.9
% Epiglucobarbarin / Phenylalkyl-GLS	0.01	0.0-0.18	0.01	0.01-0.02	0.02	0.0-0.09
% Neoglucobrassicin / Indolic-GLS	37.2	7 - 63	39.7	29 - 52	44.5	25 - 66
% 4M-I3MG / Indolic-GLS	19.2	6 - 37	16.9	12 - 23	28.5	17 - 42

			Leaf Ph	enolics		
	B. na	apus	B. r	apa	B. ol	eracea
	mean	min - max	mean	min - max	mean	min - max
Leaf PHENOLICS (nmol.g ⁻¹ DW)	14 909.3	7409 - 23469	10 985.6	5606 - 14711	13 585.0	9804 - 22340
% PPN / Phenolics	36.1	18 - 54	36.1	29 - 42	47.7	38 - 62
% Flavonol / Phenolics	63.9	46 - 83	63.9	58 - 71	52.3	38 - 62
%AQAs / PPN	13.0	3 - 31	20.9	5 - 48	20.3	13 - 31
%HAGEs / PPN	87.0	69 - 97	79.1	52 - 96	79.7	69 - 88
%PPN-monoAcyl / PPN	45.2	29 - 63	55.6	41 - 70	39.4	25 - 57
%PPN-diAcy1/ PPN	50.0	30 - 66	42.1	27 - 58	53.0	35 - 66
%PPN-triAcyl / PPN	4.7	0 - 13	2.3	1 - 3	7.6	5 - 12
%Flavonol-diHex / Flavonols	5.0	1 - 42	6.4	1 - 12	11.1	5 - 17
%Flavonol-triHex / Flavonols	85.5	51 - 94	77.8	64 - 99	72.3	59 - 80
%Flavonol-tetraHex/ Flavonols	5.5	1 - 22	9.5	0 - 19	9.9	5 - 18
%Flavonol-pentaHex/ Flavonols	3.9	1 - 12	6.3	0 - 13	6.8	4 - 12
% Flavonol-monoacyl / Flavonols	65.6	48-83	84.9	71 - 92	68.6	60 - 84
% Flavonol-diacyl / Flavonols	4.1	0 - 12	3.5	0 - 9	6.8	3 - 9
% Kaempferol / Flavonols	80.5	63 - 97	92.1	89 - 96	83.8	76 - 91
% Quercetin / Flavonols	12.6	2 - 28	7.2	4 - 10	13.8	8 - 22
% Isorhamnetin / Flavonols	6.9	0 - 20	0.8	0 - 2	2.4	0 - 6

^aData included 291 accessions of *B. napus*, 12 accessions of *B. rapa*, and 11 accessions of *B. oleracea*. Details for each accession and compound are given in Supporting Tables S2, S3, and S4. The percentile 5 and percentile 95 of metabolic trait in each per species are given per species in Supporting Tables S2d, S3d, and S4d.

GLS content in seeds. Root GLS content also displayed interaccession variations, from 5.7 to 49.4 μ mol g⁻¹ DW, but included a larger proportion of Phe-derived (~47% on average), besides Met-derived aliphatics (37% on average). In both roots and leaves, several accessions displayed strikingly consistent accumulations of branched and linear alkyl-GLSs, with few exceptions (detailed below).

Met-derived aliphatic GLS exhibited quasi exclusively 4 and 5-carbon chains. Relative proportions between those C4- and C5-aliphatic GLS displayed great variations among accessions, with C4-aliphatic GLS representing between 22% and 98% of total Met-derived aliphatic GLS (Table 5). As expected, Met-derived aliphatic glucosinolates with 3-carbon radical chains (C3-aliphatics) were found very low in accessions of both *B*.

napus and B. rapa but ranged from 0.3 up to 89% in accessions of B. oleracea (Table 5). The R-chain length of Met-derived aliphatic glucosinolates influences their biological activities, as exemplified by the consequences of allelic variations at the GS-ELONG locus on plant-insect interactions in Arabidopsis^{80,81} and wild Brassica oleracea.⁵ Thus, the diversity of C4 to C5aliphatic GLS ratio in *B. napus* accessions from our panel may have some interesting consequences on resistance to insects that would deserve further explorations. Beyond chain elongation, the level of maturation of the R-chain of Metderived is also relevant to the biological activity.^{80,81} From this perspective, GLS profiles were qualitatively distinct in roots (mainly alkenyl-, hydroxyalkenyl-, and methylsulfinyl-GLS) and leaves (mostly methylsulfinyl-, methylthio-, and hydroxvalkenyl-GLS). Most remarkable was the large interaccession range of ratios between those GLS subcategories (Table 5). For example, in leaves, the proportion of alkenyl-GLS toward total Met-derived aliphatic-GLS varied between 0.5% (accession no. 271-WIL) to 80% (accession no. 001-ADR).

Consistently with the literature, gluconasturtiin was by far the most abundant phenylalkyl-GLS in roots and leaves (92.1% and 99.8% of total phenylalkyl-GLS, respectively). Interestingly, a series of hydroxylated derivatives including glucobarbarin were observed and reached up to 10% of total phenylalkyl-GLSs (in leaves of no. 152-C34). Such a concentration can be seen as elevated considering that only a few studies have reported glucobarbarin in Brassica crops,⁴⁶ and this compound was even used as an internal standard in some Brassica studies.⁸² Glucobrassicin was the most abundant indolic-GLS in leaves. However, in roots, a wide range of contents of its methoxy-derivatives 1-methoxyglucobrassicin (from 7 to 63% of total indolic-GLS) and 4-methoxyglucobrassicin (from 7 to 63% of total indolic-GLS) was observed. The primary role of the latter as a precursor of various phytoalexins⁸³ suggests that this range of variations might be worthy of further investigation in the context of response to pathogen attacks.

Although the biosynthesis of BCAA-derived alkyl-GLSs is based, from an evolutionary point of view, on an ancient biosynthetic pathway among Brassicaceae species,⁵³ those compounds are only occasionally reported in Brassica species. Our work highlighted the occurrence in B. napus of understudied isomers from propyl to heptyl-GLSs, accounting from 0% up to 41% of total GLS in leaves. Isomeric butyl-GLSs (particularly glucochlearin**, G27**), although accumulated mostly in a small subset of accessions, were found up to 0.9 μ mol g⁻¹ DW (in leaves of no. 088-Oase) and up to 2 μ mol g^{-1} DW (in roots of no. 147-C25). Noteworthy, seven swede accessions were in the top percentile 95% for the highest butyl-GLSs concentration in roots, with above 0.5 μ mol g⁻¹ DW. To temperate this statement, those values have to be compared to higher BCAA-derived alkyl-GLSs contents found in some wild accessions of *Brassicaceae* with for instance up to 29 μ mol g⁻¹ DW isopropyl-GLS in Cardamine diphylla.84 Nevertheless, as branched-chained and linear alkyl-GLSs have been pinpointed for their volatility and possible role in plant-insect interactions,⁸⁵ their observed content variations among B. napus accessions may be worthy of interest for breeding purposes.

Phenolic Compounds. Using our methodology, we did not detect quantifiable phenolic compounds in root extracts, which was consistent with previously reported low levels of phenolic compounds in *Brassica* roots.³⁰ By contrast, the total molar

concentration of phenolics was on average at 15 μ mol g⁻¹ DW in leaves, a magnitude which was consistent with Neugart et al.⁸⁶ and which was 6-fold higher compared to glucosinolates. Phenylpropanoids and flavonol derivatives accounted for about $^{1}/_{3}$ and $^{2}/_{3}$ (respectively) of total leaf phenolics detected by $UV_{330 \text{ nm}}$ in our analyses in *B. napus* and *B. rapa*, and about 1/2in B. oleracea (Table 5). Equivalent ratios were previously reported for *B. rapa.*⁸⁷ Globally, total concentrations of flavonoid derivatives were higher in the panel of B. napus and B. oleracea compared to B. rapa. Accessions of B. napus with the highest concentrations of flavonol derivatives (percentile 95% > 13.5 μ mol g⁻¹ DW) belonged to the spring type, while most of the accessions with the highest concentrations of phenylpropanoids (percentile 5% >7.5 μ mol g⁻¹ DW) belonged to the winter type (Supporting Table S5c). Overall, 80.5% of flavonols in B. napus were derived from the aglycone kaempferol (68% of flavonols being trihexosylated kaempferol derivatives). This proportion varied from 63% (accession no. 150-C30) to 97% (accession no. 234-TOK). The data set also highlighted noticeable interaccession variations for quercetin (from 2% in no. 234-TOK to 28% in no. 171-GEN) and isorhamnetin derivatives (from 0% in no. 148-C28 to 20% in no. 138-BSC). These relatively low concentrations of isorhamnetin derivatives in leaves are consistent with previous works in Brassica, which reports larger accumulations in inflorescences and seeds.³⁰ In B. rapa, they contribute to the visual attractiveness of flowers to insect pollinators.⁸⁸ On average, $^2/_3$ of flavonols harbored one acylation in B. napus (with variations from 48% in no. 119-SWG to 83% in no. 164-DRK), and only 4% of flavonols harbored two acylations (with the highest values in swede accessions, i.e., 12% in the swede variety no. 264-BRH). In line with ref 7, accession-dependent variations in the concentrations of these high UV-absorbing multiacylated compounds may play a role in the tolerance of oilseed rape to abiotic constraints.

In B. napus accessions, leaf phenylpropanoids were composed mainly of (87%) hydroxycinnamic acid hexosylated esters and more precisely of monoacyl- (25-70%) and diacylderived (27-66%) compounds. The prominence of Disin-Odihex** (PL07**) and feruloyl-O-sin-O-dihex** (PL10**) was consistent with previous reports in kale.⁸⁶ Remarkably in B. napus, the highest accumulations for those compounds were found in winter accessions. Triacyl-derived compounds were found at globally low concentrations in B. napus leaves, although they reached up to 12% of total phenylpropanoids in some accessions (no. 088-OAS). In line with Francisco et al., accession-dependent variations in the concentrations of those high UV-absorbing multiacylated compounds may play a role in the tolerance of oilseed rape to abiotic constraints. High levels of hexosylation are a key signature of Brassica phytochemistry.⁸⁹ Indeed, our data set revealed that 48% to 98% of flavonols in B. napus are trihexosylated derivatives, the highest levels being typically gained in spring accessions. Tetrahexoside and penta-hexoside kaempferol derivatives were also found to reach up to 12 and 20% (respectively), and the largest accumulation of the two latter categories was typically found in swede accessions. In our data set, all hexosylation residues were hexosides, which was consistent with previous reports on Brassica species⁴ and significantly differed from the presence of flavonol rhamnoside derivatives in many other Brassicaceae species.^{4,90} Finally, the present data set allowed the identification of varietal accession contrasts for phenolics



Figure 5. Multivariate analyses of root glucosinolate profiles of the 304-accession panel of *Brassica* varieties. (A) Heatmap and clustering of accessions (Y-axis) with the 36 glucosinolates quantified in this work (X-axis). Data were mean-centered for each accession. (B) Total concentration (μ mol g⁻¹ DW) of root glucosinolates for each accession. (C) PCA analysis. Data were mean-centered for each compound. Representation of accessions on the score plot and variables on the loading plot as defined by the first three principal components. For visualization clarity, the G14 compound as the most abundant molecule in the root panel was not shown.

with established ecological functions. This is illustrated by a range between 0.4 μ mol g⁻¹ DW (in the swede accession no. 268-LOR) and 3.2 μ mol g⁻¹ DW (in the SOSR accession no. 177-GUZ) of kaem-O-sin-O-trihex** (PL23**), a flavonol which has been reported to play a role in resistance toward cabbage seedpod weevil.⁹¹⁹²

Chemotypes Pointed out through Multivariate Analyses. Unsupervised multivariate analyses were conducted to identify the prevalence of accession clusters and the most relevant variables describing the phytochemical specificity (Figure 5 for root glucosinolates, Figure 6 for leaf phenolics). Although the diversity of phytochemical patterns was relatively modest, as illustrated by short branches among accession clusters in roots and leaves (Figure 5A, Figure 6A), this approach nevertheless highlighted phytochemically defined clusters of *B. napus* accessions and highlighted their respective



Figure 6. Multivariate analyses of leaf phenolic profiles among the 304-accession panel of *Brassica* varieties. (A) Clustering of accessions (*Y*-axis) according to the concentration of 32 phenolics (*X*-axis). Data were mean-centered for each accession. (B) Total concentration (μ mol g⁻¹ DW) of leaf phenolics in each accession and (C) PCA analysis. Data were mean-centered for each compound. Representation of accessions on the score plot and variables on the loading plot as defined by the first three principal components (AQA, acyl-quinic acids; HAGE, hydroxycinnamic acid hexoside esters; Quer, quercetin; Isor, isorhamnetin).

closeness and distance with *B. oleracea* and *B. rapa* accessions. Using root glucosinolates, three main groups were observed (Figure 5). One group including *B. napus* and *B. rapa* accessions was characterized by higher relative concentrations of indolic compounds (especially **G19** and **G20***). A second group containing only *B. napus* accessions were characterized by elevated relative levels of C4 and C5 aliphatic methioninederived molecules (especially progoitrin (**G07**), glucoberteroin^{**} (G09^{**}), and glucoalyssin (G10). Finally, the third group included all *B. oleracea* accessions and a few accessions from the two other species. PCA analysis highlighted that *B. oleracea* harbored elevated levels of three C3-type aliphatic glucosinolates typically accumulated in cabbages (glucoiberverin^{**} G01^{**}, glucoiberin G02, and sinigrin G03) and the pentyl-GLS-isomer^{**} (G30^{**}). Leaf phenolic-based clustering permitted the identification of two main chemotypes among

Brassica accessions (Figure 6). A first large cluster 1 (bottom) with all B. oleracea and B. rapa as well as a large proportion of B. napus accessions, harbored high levels of sin-O-hex** (PL04**). In this cluster, all B. rapa were grouped together and were discriminated by high concentrations of PL22** (kaem-O-fer-O-trihex**). A smaller second cluster (above) consisting solely of B. napus accessions was characterized by high concentrations of kaem-O-trihex** (PL16**). This clustering was driven by compounds with the highest concentrations in leaves. As a complement, the PCA approach highlighted compounds that, despite low concentrations in leaves, also contributed to the structuration of the chemical diversity. These included three isomers of feruloyl-O-sinapoyl-O-dihex** (PL09**, PL10**, and PL11**) whose concentrations well discriminated B. oleracea from B. rapa accessions (PC1 axis). For each phenolic compound, a comparison was made between leaf content in the two diploid species and B. napus (Supporting Results 2). Interestingly, molecules PL02**, PL06**, PL10**, PL11**, PL13**, PL18**, PL20**, PL22**, PL24**, and PL25** were found accumulated at significantly higher levels in B. rapa and B. oleracea, respectively, compared to B. napus. Some of these candidates with agronomically relevant biological traits could be easily introgressed from progenitors to B. napus.

In summary, this work provides enriching information about the chemical diversity in B. napus. The number of compounds (36 GLS and 32 phenolics) and the number of accessions analyzed (304) represent one of the largest set described in a single study until now.^{22,93} Kaempferol-O-sin-O-pentahex** (PL26**) is reported here for the first time in Brassicaceae. To our knowledge, this is also the first report in Brassica species of four isomers of hydroxyphenylethyl-GLS (G15*, G16*, G17, G18**) and branched chain or linear alkyl-GLS (G25*, G26*, G28**, G30**, G31*, G32**-G36**) in B. napus. The large range of glucocochlearin (G27*) concentrations (from zero up to 26% of leaf GLS) over the diversity of B. napus accessions was unexpected, suggesting possible ecological consequences to be explored. The content of phenolics in B. napus also greatly varied among genotypes, offering a basis for further works in chemical ecology and breeding. This metabolic data set will be exploited in our future work to investigate the genetic architecture of metabolic traits, and identify underlying genes involved in the control of these metabolic variations, and explore their roles in biotic interactions.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jafc.1c08118.

Supporting Results S1, detailed explanations for the annotation of GLS in this study; Supporting Results S2, detailed explanations for the annotation of phenolics in this study; Supporting Results S3, chromatograms of isobaric ions from rare GLSs observed in *Brassica* extracts and chromatograms from reference plant materials; and Supporting Results S4, box plots comparing the diversity of leaf content for every phenolics and GLS among the accessions for the three species *B. napus, B. rapa,* and *B. oleracea* (PDF)

Supporting Table S1, detailed HRMS and MS/MS data, elution times, and proposed annotation of 36

glucosinolates and 32 phenolic compounds detected in the panel of 304 *Brassica* accessions (XLSX)

Supporting Table S2, external standards used for quantification (XLSX)

Supporting Table S3, leaf GLS content in a panel of 304 *Brassica* accessions; Supporting Table S3a, leaf GLS content in each biological replicate (nmol g^{-1} DW); Supporting Table S3b, mean leaf GLS content, min, max, percentile 5, and percentile 95; Supporting Table S3c, selection of accessions belonging to percentile 5 and percentile 95 for each GLS (XLSX)

Supporting Table S4, root GLS content in a panel of 304 *Brassica* accessions; Supporting Table S4a, root GLS content in each biological replicate (nmol g^{-1} DW); Supporting Table S4b, mean root GLS content, min, max, percentile 5, and percentile 95; Supporting Table S4c, selection of accessions belonging to percentile 5 and percentile 95 for each GLS (XLSX)

Supporting Table S5, leaf phenolic content in a panel of 304 *Brassica* accessions; Supporting Table S5a, leaf phenolic content in each biological replicate (nmol g^{-1} DW); Supporting Table S5b, mean leaf phenolic content, min, max, percentile 5, and percentile 95; Supporting Table S5c, selection of accessions belonging to percentile 5 and percentile 95 for each phenolic compound (XLSX)

AUTHOR INFORMATION

Corresponding Author

Antoine Gravot – Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France; Occid.org/0000-0001-9125-1494; Email: antoine.gravot@univ-rennes1.fr

Authors

- Anani Amegan Missinou Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- Julie Ferreira de Carvalho Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- Nathalie Marnet INRAE, BIA, 35653 Le Rheu, France
- **Thomas Delhaye** Univ Rennes, CNRS, IETR, 35000 Rennes, France
- Oumayma Hamzaoui Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- David Abdel Sayed Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- Yann Guitton Oniris, INRAE, LABERCA, 44300 Nantes, France
- Lionel Lebreton Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- Christophe Langrume Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- Anne Laperche Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- **Régine Delourme** Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- Maria J. Manzanares-Dauleux Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France
- Alain Bouchereau Univ Rennes, Institut Agro, INRAE, IGEPP, 35653 Le Rheu, France

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jafc.1c08118

Author Contributions

¹Anani Amegan Missinou and Julie Ferreira de Carvalho equally contributed. M.J.M.-D., A.B., and A.G. obtained supporting funding. A.L., R.D., M.J.M.-D., L.L., C.L., and A.G. defined the *Brassica* panel and the sampling experimental design. A.A.M., N.M., O.H., Y.G., T.D., and A.G. worked on the extensive research and identification of glucosinolates and phenolics, in a series of pooled samples from *Brassica* accessions, using high-resolution MS/MS data. A.A.M., N.M., O.H., and D.A.S. have set up the UPLC-MS method for quantitative glucosinolate and phenolic profiling and then did a chromatographic analysis and peak integrations to generate quantitative data. A.A.M. and J.F.d.C. worked on the statistical analysis of data. A.A.M., J.F.d.C., and A.G. wrote the manuscript.

Funding

The salaries of A.A.M. and J.F.d.C. were provided by INRAE and Région Bretagne. Operating funding was supported by INRAE, Agrocampus Ouest, Université Rennes 1, and PROMOSOL. Salaries of O.H. and D.A.S. were provided by PROMOSOL (Project DESCRIBE).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge the two facilities P2M2 and MELISA from the CORSAIRE metabolomic and fluxomic platform of Biogenouest. The greenhouse experimental team of IGEPP is acknowledged for contributing to the experiments in the phytotron. The authors would like to thank the BrACySol Biological Resource Center (INRAE Ploudaniel, France) for providing seeds from SOSR accessions, breeder companies from the PROMOSOL Association for providing seeds from elite WOSR accessions, and Professor Rod Sowndon for providing the seeds of rutabaga and the WFR genotypes. Many people from IGEPP have also been mobilized as technical support during the 3 weeks of the leaf and root sampling stage. May they all be here sincerely thanked. We also thank the reviewers for their constructive corrections and comments which helped us to improve the quality of this manuscript.

ABBREVIATIONS USED

AQAs, acyl-quinic acids; BCAAs, branched-chain amino acids; Caf, caffeoyl; DW, dry weight; Fer, feruloyl; GLS, glucosinolate; Hex, hexoside; HAGE, hydroxycinnamic acid hexoside esters; HPLC-HRMS2, liquid chromatography coupled to high-resolution tandem mass spectrometry; Isor, isorhamnetin; Kaem, kaempferol; SGlu, thioglucose; Sin, sinapoyl; SORS, spring oilseed rape; OH-Fer, hydroxyferuloyl; PCA, principal component analysis; Quer, quercetin; UPLC-UV-MS, ultrahigh-performance chromatography with ultraviolet detector coupled in-line to a triple quadrupole detector mass spectrometry; WOSR, winter oilseed rape

REFERENCES

(1) Windsor, A. J.; Reichelt, M.; Figuth, A.; Svatoš, A.; Kroymann, J.; Kliebenstein, D. J.; Gershenzon, J.; Mitchell-Olds, T. Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (*Brassicaceae*). *Phytochemistry* **2005**, *66* (11), 1321–1333.

(2) Dinkova-Kostova, A. T.; Kostov, R. V. Glucosinolates and isothiocyanates in health and disease. *Trends Mol. Med.* **2012**, *18* (6), 337–347.

(3) Traka, M.; Mithen, R. Glucosinolates, isothiocyanates, and human health. *Phytochem. Rev.* **2009**, *8* (1), 269–282.

(4) Cartea, M. E.; Francisco, M.; Soengas, P.; Velasco, P. Phenolic compounds in *Brassica* vegetables. *Molecules* 2011, 16 (1), 251–280.
(5) Bradburne, R. P.; Mithen, R. Glucosinolate genetics and the attraction of the aphid parasitoid *Diaeretiella rapae* to *Brassica. Proc. R. Soc. B Biol. Sci.* 2000, 267 (1438), 89–95.

(6) Francisco, M.; Joseph, B.; Caligagan, H.; Li, B.; Corwin, J. A.; Lin, C.; Kerwin, R. E.; Burow, M.; Kliebenstein, D. J. Genome wide association mapping in *Arabidopsis thaliana* Identifies novel genes involved in linking allyl glucosinolate to altered biomass and defense. *Front. Plant Sci.* **2016**, DOI: 10.3389/fpls.2016.01010.

(7) Schulz, E.; Tohge, T.; Winkler, J. B.; Albert, A.; Schaffner, A. R.; Fernie, A. R.; Zuther, E.; Hincha, D. K. Natural variation among *Arabidopsis* accessions in the regulation of flavonoid metabolism and stress gene expression by combined UV radiation and cold. *Plant Cell Physiol.* **2021**, *62* (3), 502–514.

(8) Mauricio, R. Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *Am. Nat.* **1998**, 151 (1), 20–28.

(9) Mithen, R.; Campos, H. Genetic variation of aliphatic glucosinolates in *Arabidopsis thaliana* and prospects for map based gene cloning. *Entomol. Exp. Appl.* **1996**, *80* (1), 202–205.

(10) Rodman, J. E. Population variation and hybridization in searockets (*Cakile, Cruciferae*): seed glucosinolate characters. *Am. J. Bot.* **1980**, 67 (8), 1145.

(11) Katz, E.; Li, J. J.; Jaegle, B.; Ashkenazy, H.; Abrahams, S. R.; Bagaza, C.; Holden, S.; Pires, C. J.; Angelovici, R.; Kliebenstein, D. J. Genetic variation, environment and demography intersect to shape *Arabidopsis* defense metabolite variation across Europe. *Elife* **2021**, *10*, 1–25.

(12) Pfalz, M.; Vogel, H.; Kroymann, J. The Gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in *Arabidopsis*. *Plant Cell* **2009**, *21* (3), 985–999.

(13) Raybould, A. F.; Moyes, C. L. The Ecological genetics of aliphatic glucosinolates. *Heredity* (*Edinb*). **2001**, 87 (4), 383–391.

(14) Newton, E.; Bullock, J. M.; Hodgson, D. Bottom-up effects of glucosinolate variation on aphid colony dynamics in wild cabbage populations. *Ecol. Entomol.* **2009**, *34* (5), 614–623.

(15) Moyes, C. L.; Collin, H. A.; Britton, G.; Raybould, A. F. Glucosinolates and differential herbivory in wild populations of *Brassica oleracea. J. Chem. Ecol.* **2000**, *26* (11), 2625–2641.

(16) Klopsch, R.; Witzel, K.; Artemyeva, A.; Ruppel, S.; Hanschen, F. S. Genotypic variation of glucosinolates and their breakdown products in leaves of *Brassica rapa*. J. Agric. Food Chem. **2018**, 66 (22), 5481–5490.

(17) Lee, J. G.; Bonnema, G.; Zhang, N.; Kwak, J. H.; De Vos, R. C. H.; Beekwilder, J. Evaluation of glucosinolate variation in a collection of turnip (*Brassica rapa*) germplasm by the analysis of intact and desulfo glucosinolates. *J. Agric. Food Chem.* **2013**, *61* (16), 3984–3993.

(18) Branca, F.; Li, G.; Goyal, S.; Quiros, C. F. Survey of aliphatic glucosinolates in sicilian wild and cultivated *Brassicaceae*. *Phytochemistry* **2002**, *59* (7), 717–724.

(19) Nagaharu, U. Genome-analysis in Brassica with special reference to the experimental formation of B. napus and particular mode of fertilization. Imperial Agricultural Experimental Station, 1935.

(20) Chalhoub, B.; et al. Early allopolyploid evolution in the postneolithic *Brassica napus* oilseed genome. *Science* (80-.) **2014**, 345 (6199), 950–953.

(21) Stefansson, B. R.; Kondra, Z. P. Tower summer rape. Can. J. Plant Sci. 1975, 55, 343–344.

(22) Kittipol, V.; He, Z.; Wang, L.; Doheny-Adams, T.; Langer, S.; Bancroft, I. Genetic architecture of glucosinolate variation in *Brassica napus. J. Plant Physiol.* **2019**, *240* (June), 152988.

(23) Mithen, R. Leaf glucosinolate profiles and their relationship to pest and disease resistance in oilseed rape. *Euphytica* **1992**, 63 (1–2), 71–83.

(24) Olsen, H.; Aaby, K.; Borge, G. I. A. Characterization and quantification of flavonoids and hydroxycinnamic acids in curly kale (*Brassica oleracea L. convar. acephala var. sabellica*) by HPLC-DAD-ESI-MSn. J. Agric. Food Chem. **2009**, 57 (7), 2816–2825.

(25) Shao, Y.; Jiang, J.; Ran, L.; Lu, C.; Wei, C.; Wang, Y. Analysis of flavonoids and hydroxycinnamic acid derivatives in rapeseeds (*Brassica napus L. var. napus*) by HPLC-PDA-ESI(–)-MSn/HRMS. J. Agric. Food Chem. **2014**, 62 (13), 2935–2945.

(26) Velasco, P.; Francisco, M.; Moreno, D. A.; Ferreres, F.; García-Viguera, C.; Cartea, M. E. Phytochemical fingerprinting of vegetable *Brassica oleracea* and *Brassica napus* by simultaneous identification of glucosinolates and phenolics. *Phytochem. Anal.* **2011**, 22 (2), 144– 152.

(27) Llorach, R.; Gil-Izquierdo, A.; Ferreres, F.; Tomás-Barberán, F. A. HPLC-DAD-MS/MS ESI Characterization of unusual highly glycosylated acylated flavonoids from cauliflower (*Brassica oleracea L. var. botrytis*) agroindustrial byproducts. J. Agric. Food Chem. 2003, 51 (13), 3895–3899.

(28) Lee, K.-C.; Chan, W.; Liang, Z.; Liu, N.; Zhao, Z.; Lee, A. W.-M.; Cai, Z. Rapid screening method for intact glucosinolates in chinese medicinal herbs by using liquid chromatography coupled with electrospray ionization ion trap mass spectrometry in negative ion mode. *Rapid Commun. Mass Spectrom.* **2008**, *22* (18), 2825–2834.

(29) Bhandari, S. R.; Jo, J. S.; Lee, J. G. Comparison of glucosinolate profiles in different tissues of nine Brassica crops. *Molecules* **2015**, *20* (9), 15827–15841.

(30) Farag, M. A.; Sharaf Eldin, M. G.; Kassem, H.; Abou El Fetouh, M. Metabolome classification of *Brassica napus L. organs* via UPLC-QTOF-PDA-MS and their anti-oxidant potential. *Phytochem. Anal.* **2013**, *24* (3), 277–287.

(31) Sun, R.; Jiang, X.; Reichelt, M.; Gershenzon, J.; Vassão, D. G. The selective sequestration of glucosinolates by the cabbage aphid severely impacts a predatory lacewing. *J. Pest Sci.* (2004) **2021**, 94 (4), 1147–1160.

(32) Gols, R.; van Dam, N. M.; Reichelt, M.; Gershenzon, J.; Raaijmakers, C. E.; Bullock, J. M.; Harvey, J. A. Seasonal and herbivore-induced dynamics of foliar glucosinolates in wild cabbage (*Brassica oleracea*). *Chemoecology* **2018**, *28* (3), 77–89.

(33) Bruce, T. J. A. Glucosinolates in Oilseed rape: Secondary metabolites that influence interactions with herbivores and their natural enemies. *Ann. Appl. Biol.* **2014**, *164* (3), 348–353.

(34) Giamoustaris, A.; Mithen, R. The effect of modifying the glucosinolate content of leaves of oilseed rape (. *Ann. Appl. Biol.* **1995**, *126*, 347–363.

(35) Renwick, J. A. A.; Radke, C. D.; Sachdev-Gupta, K.; Städler, E. Leaf surface chemicals stimulating oviposition by *Pieris rapae* (*Lepidoptera: pieridae*) on cabbage. *Chemoecology* **1992**, *3* (1), 33–38. (36) Schmidt, S.; Zietz, M.; Schreiner, M.; Rohn, S.; Kroh, L. W.; Krumbein, A. Identification of complex, naturally occurring favonoid glycosides in kale (*Brassica oleracea var. sabellica*) by high-performance liquid chromatography diode-array detection/electrospray ionization multi-stage mass spectrometry. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 2009–2022.

(37) Francisco, M.; Moreno, D. A.; Cartea, M. E.; Ferreres, F.; García-Viguera, C.; Velasco, P. Simultaneous identification of glucosinolates and phenolic compounds in a representative collection of vegetable *Brassica Rapa. J. Chromatogr. A* **2009**, *1216* (38), 6611–6619.

(38) Harbaum, B.; Hubbermann, E. M.; Wolff, C.; Herges, R.; Zhu, Z.; Schwarz, K. Identification of flavonoids and hydroxycinnamic acids in pak choi varieties (*Brassica campestris L. ssp. chinensis var. communis*) by HPLC-ESI-MSn and NMR and their quantification by HPLC-DAD. J. Agric. Food Chem. **2007**, 55 (20), 8251–8260.

(39) Rochfort, S. J.; Trenerry, V. C.; Imsic, M.; Panozzo, J.; Jones, R. Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MSn fragmentation. *Phytochemistry* **2008**, *69* (8), 1671–1679.

(40) Ferreres, F.; Valentão, P.; Llorach, R.; Pinheiro, C.; Cardoso, L.; Pereira, J. A.; Sousa, C.; Seabra, R. M.; Andrade, P. B. Phenolic

compounds in external leaves of tronchuda cabbage (*Brassica oleracea L. var. costata DC*). *J. Agric. Food Chem.* **2005**, 53 (8), 2901–2907.

(41) Nielsen, J. K.; Nørbæk, R.; Olsen, C. E. Kaempferol tetraglucosides from cabbage leaves. *Phytochemistry* **1998**, 49 (7), 2171–2176.

(42) Fernandes, F.; Valentão, P.; Sousa, C.; Pereira, J. A.; Seabra, R. M.; Andrade, P. B. Chemical and antioxidative assessment of dietary turnip (*Brassica rapa var. rapa L.*). *Food Chem.* **2007**, *105* (3), 1003–1010.

(43) Li, Z.; Zheng, S.; Liu, Y.; Fang, Z.; Yang, L.; Zhuang, M.; Zhang, Y.; Lv, H.; Wang, Y.; Xu, D. Characterization of glucosinolates in 80 broccoli genotypes and different organs using UHPLC-triple-TOF-MS method. *Food Chem.* **2021**, *334*, 127519.

(44) Lin, L. Z.; Harnly, J. M. Identification of the phenolic components of collard greens, kale, and chinese broccoli. *J. Agric. Food Chem.* **2009**, *57* (16), 7401–7408.

(45) Liang, X.; Lee, H. W.; Li, Z.; Lu, Y.; Zou, L.; Ong, C. N. Simultaneous quantification of 22 glucosinolates in 12 *Brassicaceae* vegetables by hydrophilic interaction chromatography-tandem mass spectrometry. *ACS Omega* **2018**, *3* (11), 15546–15553.

(46) Kim, H. W.; Ko, H. C.; Baek, H. J.; Cho, S. M.; Jang, H. H.; Lee, Y. M.; Kim, J. B. Identification and quantification of glucosinolates in korean leaf mustard germplasm (*Brassica Juncea var. integrifolia*) by Liquid Chromatography-Electrospray Ionization/ Tandem Mass Spectrometry. *Eur. Food Res. Technol.* **2016**, 242 (9), 1479–1484.

(47) Lin, L. Z.; Harnly, J. M. Phenolic component profiles of mustard greens, yu choy, and 15 other *Brassica* vegetables. J. Agric. Food Chem. 2010, 58 (11), 6850–6857.

(48) Liu, S.; Huang, H.; Yi, X.; Zhang, Y.; Yang, Q.; Zhang, C.; Fan, C.; Zhou, Y. Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome-wide association study. *Plant Biotechnol. J.* **2020**, *18* (6), 1472–1484.

(49) Lenth, R. V.; Buerkner, P.; Herve, M.; Love, J.; Hannes Riebl, H. S. Emmeans: estimated marginal means, aka least-squares means [R package emmeans version 1.7.0]. *Am. Stat.* **2021**, *34* (4), 216–221. (50) Dray, S.; Dufour, A. B. The Ade4 Package: Implementing the

duality diagram for ecologists. J. Stat. Softw. 2007, 22 (4), 1–20.

(51) Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A. Nbclust: An R package for determining the relevant number of clusters in a data Set. J. Stat. Softw. **2014**, *61* (6), 1–36.

(52) Wickham, H. Ggplot2: Elegant graphics for data analysis, 2nd ed.; 2016; Vol. 174. DOI: 10.1111/j.1467-985x.2010.00676 9.x.

(53) Agerbirk, N.; Hansen, C. C.; Kiefer, C.; Hauser, T. P.; Ørgaard, M.; Asmussen Lange, C. B.; Cipollini, D.; Koch, M. A. Comparison of glucosinolate diversity in the crucifer tribe *Cardamineae* and the remaining order *Brassicales* highlights repetitive evolutionary loss and gain of biosynthetic steps. *Phytochemistry* **2021**, *185*, 112668.

(54) Blažević, I.; Montaut, S.; Burčul, F.; Olsen, C. E.; Burow, M.; Rollin, P.; Agerbirk, N. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry* **2020**, *169*, 112100.

(55) Fahey, J. W.; Zalcmann, A. T.; Talalay, P. The Chemical diversity and distribution of glucosinolates and isothiocyanates among Plants. *Phytochemistry* **2001**, *56* (1), 5-51.

(56) Guo, D.; Föll, M. C.; Volkmann, V.; Enderle-Ammour, K.; Bronsert, P.; Schilling, O.; Vitek, O. Deep Multiple instance learning classifies subtissue locations in mass spectrometry images from Tissue-Level Annotations. *Bioinformatics* **2020**, *36* (1), i300–i308.

(57) Kokkonen, P. S.; van der Greef, J.; Niessen, W. M. A.; Tjaden, U. R.; ten Hove, G. J.; van de Werken, G. Identification of intact glucosinolates using direct coupling of high-performance liquid chromatography with continuous-flow frit fast atom bombardment tandem mass spectrometry. *Biol. Mass Spectrom.* **1991**, 20 (5), 259–267.

(58) Cataldi, T. R. I.; Rubino, A.; Lelario, F.; Bufo, S. A. Naturally occurring glucosinolates in plant extracts of rocket salad (*Eruca sativa L.*) identified by liquid chromatography coupled with negative ion

pubs.acs.org/JAFC

electrospray ionization and quadrupole ion-trap mass spectrometry. Rapid Commun. Mass Spectrom. 2007, 21 (14), 2374–2388.

(59) Clarke, D. B. Glucosinolates, structures and analysis in food. *Anal. Methods* **2010**, 2 (4), 310–325.

(60) Agerbirk, N.; Olsen, C. E.; Poulsen, E.; Jacobsen, N.; Hansen, P. R. Complex metabolism of aromatic glucosinolates in *Pieris rapae* caterpillars involving nitrile formation, hydroxylation, demethylation, sulfation, and host plant dependent carboxylic acid formation. *Insect Biochem. Mol. Biol.* **2010**, 40 (2), 126–137.

(61) Wang, C.; Li, Z.; Zhang, L.; Gao, Y.; Cai, X.; Wu, W. Identifying key metabolites associated with glucosinolate biosynthesis in response to nitrogen management strategies in two rapeseed (*Brassica napus*) varieties. *J. Agric. Food Chem.* **2022**, *70* (2), 634–645.

(62) Agerbirk, N.; Olsen, C. E.; Heimes, C.; Christensen, S.; Bak, S.; Hauser, T. P. Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer barbarea vulgaris. *Phytochemistry* **2015**, *115* (1), 130–142.

(63) Hanley, A. B.; Belton, P. S.; Fenwick, G. R.; Janes, N. F. Ring oxygenated indole glucosinolates of *Brassica* species. *Phytochemistry* **1985**, 24 (3), 598–600.

(64) Truscott, R. J. W.; Burke, D. G.; Minchinton, I. R. The characterisation of a novel hydroxindole glucosinolate. 1982, 107 (4), 1258–1264.

(65) Truscott, R. J. W.; Johnstone, P. K.; Minchinton, I. R.; Sang, J. P. Indole glucosinolates in swede (*Brassica napobrassica L. mill*). J. Agric. Food Chem. **1983**, 31 (1964), 863–867.

(66) Kutáček, M. Glucobrassicin a potential inhibitor of unusual type affecting the germination and growth of plants; mechanism of its action. *Biol. Plant.* **1964**, 6 (2), 88–98.

(67) Gmelin, Rolf; Virtanen, A. I. Neoglucobrassicin, einzweiter SCN-precursor vom indoltyp in *Brassica-Arten*. 1962.

(68) Blažević, I.; Montaut, S.; Burčul, F.; Olsen, C. E.; Burow, M.; Rollin, P.; Agerbirk, N. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry* **2020**, *169*, 112100.

(69) Montaut, S.; Bleeker, R. S.; Jacques, C. Phytochemical constituents of *Cardamine diphylla. Can. J. Chem.* **2010**, 88 (1), 50–55.

(70) Mithen, R.; Bennett, R.; Marquez, J. Glucosinolate biochemical diversity and innovation in the *Brassicales*. *Phytochemistry* **2010**, *71* (17–18), 2074–2086.

(71) Daxenbichler, M. E.; Spencer, G. F.; Carlson, D. G.; Rose, G. B.; Brinker, A. M.; Powell, R. G. Composition of seeds from 297 species of Wild Plants. *Phytochemistry* **1991**, *30* (8), 2623–2638.

(72) Dauvergne, X.; Cérantola, S.; Salaün, S.; Magné, C.; Kervarec, N.; Bessières, M. A.; Deslandes, E. General occurrence of the glucosinolate glucocochlearin within the *Cochlearia* genus. *Carbohydr. Res.* **2006**, *341* (12), 2166–2169.

(73) Kim, S.-J.; Fujii, K.; Mohamed, Z. I. S.; Kim, H.-W.; Yamauchi, H.; Ishii, G. Identification and quantitative determination of glucosinolates in *Brassica napus cv. Food Sci. Biotechnol.* **2008**, 17 (5), 1097–1101.

(74) Agerbirk, N.; Olsen, C. E.; Chew, F. S.; Ørgaard, M. Variable Glucosinolate profiles of cardamine pratensis (*Brassicaceae*) with equal chromosome numbers. *J. Agric. Food Chem.* **2010**, *58* (8), 4693–4700.

(75) Taveira, M.; Fernandes, F.; Guedes de Pinho, P.; Andrade, P. B.; Pereira, J. A.; Valentão, P. Evolution of *Brassica rapa var. rapa L.* volatile composition by HS-SPME and GC/IT-MS. *Microchem. J.* **2009**, 93 (2), 140–146.

(76) Fiol, M.; Adermann, S.; Neugart, S.; Rohn, S.; Mügge, C.; Schreiner, M.; Krumbein, A.; Kroh, L. W. Highly glycosylated and acylated flavonols isolated from kale (*Brassica oleracea var. sabellica*) - structure-antioxidant activity relationship. *Food Res. Int.* **2012**, 47 (1), 80–89.

(77) Ferreres, F.; Llorach, R.; Gil-Izquierdo, A. Characterization of the interglycosidic linkage in di-, tri-, tetra- and pentaglycosylated flavonoids and differentiation of positional isomers by liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry. J. Mass Spectrom. 2004, 39 (3), 312–321.

(78) Vallejo, F.; Tomás-Barberán, F. A.; Ferreres, F. Characterisation of Flavonols in Broccoli (*Brassica oleracea L. var. Italica*) by Liquid Chromatography-UV Diode-Array Detection-Electrospray Ionisation Mass Spectrometry. J. Chromatogr. A **2004**, 1054 (1–2), 181–193.

(79) Rochfort, S. J.; Imsic, M.; Jones, R.; Trenerry, V. C.; Tomkins, B. Characterization of flavonol conjugates in immature leaves of pak choi [*Brassica rapa L. ssp. chinensis L.* (Hanelt.)] by HPLC-DAD and LC-MS/MS. J. Agric. Food Chem. **2006**, *54* (13), 4855–4860.

(80) Brachi, B.; Meyer, C. G.; Villoutreix, R.; Platt, A.; Morton, T. C.; Roux, F.; Bergelson, J. Coselected Genes determine adaptive variation in herbivore resistance throughout the native range of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (13), 4032–4037.

(81) Kroymann, J.; Donnerhacke, S.; Schnabelrauch, D.; Mitchell-Olds, T. Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (24), 14587–14592.

(82) Jensen, L. M.; Jepsen, H. S. K.; Halkier, B. A.; Kliebenstein, D. J.; Burow, M. Natural variation in cross-talk between glucosinolates and onset of flowering in *Arabidopsis. Front. Plant Sci.* **2015**, *6*, 1–10.

(83) Hansen, M.; Møller, P.; Sørensen, H.; de Trejo, M. C. Glucosinolates in *broccoli* stored under controlled atmosphere. *J. Am. Soc. Hortic. Sci.* **1995**, 120 (6), 1069–1074.

(84) Klein, A. P.; Sattely, E. S. Biosynthesis of cabbage phytoalexins from indole glucosinolate. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (8), 1910–1915.

(85) Olsen, C. E.; Huang, X.-C.; Hansen, C. I. C.; Cipollini, D.; Ørgaard, M.; Matthes, A.; Geu-Flores, F.; Koch, M. A.; Agerbirk, N. Glucosinolate diversity within a phylogenetic framework of the tribe *Cardamineae* (*Brassicaceae*) unraveled with HPLC-MS/MS and NMR-based analytical distinction of 70 desulfoglucosinolates. *Phytochemistry* **2016**, *132*, 33–56.

(86) Schranz, M. E.; Manzaneda, A. J.; Windsor, A. J.; Clauss, M. J.; Mitchell-Olds, T. Ecological genomics of *Boechera Stricta*: Identification of a QTL controlling the allocation of methionine-vs branched-chain amino acid-derived glucosinolates and levels of insect herbivory. *Heredity (Edinb)*. **2009**, *102* (5), 465–474.

(87) Neugart, S.; Fiol, M.; Schreiner, M.; Rohn, S.; Zrenner, R.; Kroh, L. W.; Krumbein, A. Low and moderate photosynthetically active radiation affects the flavonol glycosides and hydroxycinnamic acid derivatives in kale (*Brassica oleracea var. sabellica*) dependent on two low temperatures. *Plant Physiol. Biochem.* **2013**, *72*, 161–168.

(88) Romani, A.; Vignolini, P.; Isolani, L.; Ieri, F.; Heimler, D. HPLC-DAD/MS characterization of flavonoids and hydroxycinnamic derivatives in turnip tops (*Brassica rapa L. subsp. sylvestris L.*). *J. Agric. Food Chem.* **2006**, *54* (4), 1342–1346.

(89) Sasaki, K.; Takahashi, T. A flavonoid from *Brassica Rapa* Flower as the UV-absorbing nectar guide. *Phytochemistry* **2002**, *61* (3), 339–343.

(90) Durkeet, A. B.; Harborne, J. B. Flavonol glycosides in *Brassica* and *Sinapis*. *Phytochemistry* **1973**, *12* (5), 1085–1089.

(91) Routaboul, J. M.; Kerhoas, L.; Debeaujon, I.; Pourcel, L.; Caboche, M.; Einhorn, J.; Lepiniec, L. Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. *Planta* **2006**, 224 (1), 96–107.

(92) Lee, R. W. H.; Malchev, I. T.; Rajcan, I.; Kott, L. S. Identification of putative quantitative trait loci associated with a flavonoid related to resistance to cabbage seedpod weevil (*Ceutorhynchus Obstrictus*) in sanola derived from an intergeneric cross, *Sinapis alba* × *Brassica napus. Theor. Appl. Genet.* **2014**, *127* (2), 419–428.

(93) Wang, C.; Li, Z.; Zhang, L.; Gao, Y.; Cai, X.; Wu, W. Identifying key metabolites associated with glucosinolate biosynthesis in response to nitrogen management strategies in two rapeseed (*Brassica napus*) varieties. *J. Agric. Food Chem.* **2022**, *70* (2), 634–645.

Paper II: GWAS reveals genetic networks involved in the control of phenolics and glucosinolates, and specifies the metabolic consequences of breeding history in *Brassica napus* cultigroups

GWAS reveals genetic networks involved in the control of phenolics and glucosinolates, and specifies the metabolic consequences of breeding history in *Brassica napus* cultigroups

Julie Ferreira de Carvalho^{#,a}, **Anani Amegan Missinou^{#,a}**, Philippe Duffé^a, Christophe Mougel^a, Alain Bouchereau^{ab}, Maria Manzanares-Dauleux^a, Régine Delourme^a, Anne Laperche^a, Antoine Gravot *,^{a,b}

^a Univ Rennes, Institut Agro, INRAE, IGEPP, 35653, Le Rheu, France
 ^b INRAE, BIA, 35653, Le Rheu, France
 [#] First authors and * corresponding author, <u>Antoine.Gravot@univ-rennes1.fr</u>

1. Introduction

Specialized metabolites play major roles in the interactions between plants and their abiotic and biotic environment. This notably includes resistance to phytophagous insects, tolerance to UV, or the assembling of rhizosphere microbial communities (Abedini et al., 2021). A better understanding of the genetic determinants influencing the diversity of these metabolic compositions among crop varieties is therefore key to define breeding strategies and design varieties adapted to fluctuating environments, as required to face climate change and to promote agroecological systems.

Brassica napus is a crop of major importance, as it is one of the few oilseed species in European cropping systems. Its cultivation is nowadays threatened by abiotic and biotic stresses. These notably include an increasing challenge with insects such as cabbage-stem weevil or flea beetle, but also several microbial pathogens and the parasitic plant broomrape. Therefore, the exploitation of the diversity of specialized

Ph D. manuscript | Anani Amegan MISSINOU

metabolites is a real challenge for this species. Like in all Brassica crops, its specialized metabolite repertoire primarily encompasses glucosinolates (GLS) and phenolic compounds (PHL), including phenylpropanoids and flavonoids. Previous studies on oilseed rape have stated the relevance and complex ecological functions played by GLSs in resistance or stimulation of oilseed rape toward generalist and specialist phytophagous insects (Mithen, 1992; Giamoustaris and Mithen, 1995) as well as their involvement in improving plant resistance to fungi (Zhang et al., 2015; Wu et al., 2016). Phenolic compounds can be also involved in a wide range of functions, such as developmental processes, symbioses, and responses to pathogen attacks (Bhattacharya et al., 2010; Mandal et al., 2017). Metabolite profiles in *B. napus* depend on genetic factors, but also on plant developmental stages, plant organs (Obermeier et al., 2013; Francisco et al., 2016), and environmental conditions. The quantitative genetic control of GLSs has been well documented for seed GLS content (e.g. Wang et al., 2018). A small series of QTLs regulating total seed GLS content, located on chromosomes A09, C02, C07, and C09, may correspond to different copies of MYB28 as suggested by Li et al (2014). More recently, advances have been made to investigate GLS in other organs, leaves, and roots, and their genetic architecture (Kittipol et al., 2019), highlighting A02, A03, A09, C02, and C09 regions. These regions were also found associated with total Met-derived aliphatic-GLSs (Kittipol et al., 2019). Concerning phenolics, few genetic studies, within general unannotated PHL molecules, were undertaken in Brassica sp. so far (Obermeier et al., 2013; Francisco et al., 2016). Interestingly, the focus of attention in phenolic compounds has been the molecular regulation of sinapine (=sinapoylcholine), an anti-nutritional compound accumulated in *B. napus* seeds (Milkowski et al., 2004; Baumert et al., 2005).

Brassica napus specie was formed approximately 7500 years ago by hybridization between B. rapa and B. oleracea followed by spontaneous whole genome doubling (Chalhoub et al., 2014). Since then, diverse vegetable and oilseed crops were selected and grown worldwide. Diversification led to two main subspecies whose origin was estimated ca 3,000 years ago (Hu et al., 2021). One includes swede forms (ssp rapifera) while the other one, ssp oleifera, comprises fodder, winter oilseed rape (WOSR), and spring oilseed rape (SOSR) forms (An et al., 2019). During the 1970s, to compel with animal feed purposes, total glucosinolate content has been drastically reduced in the seeds of modern oilseed rape varieties by the introgression of alleles from the Polish cultivar 'Bronowski' (Kondra and Stefansson, 1970; Rosa et al., 2010). This rapid reduction resulted both in an erosion of genetic diversity within the oilseed rape genetic pool (Bus et al., 2011) and in a concomitant decrease of some glucosinolates (GLSs) in other organs such as leaves and roots (Kittipol et al., 2019; Liu et al., 2019; Liu et al., 2020). The potential impact of this seed GLS reduction on phenolics (PHL) and GLSs in leaves and roots in WOSR and SOSR has been however until now only insufficiently investigated. We also previously identified two major chemotypes in a large panel of *B. napus* accessions, a distinction that was partly explained by the balance between the relative contents of flavonols and phenylpropanoids (Missinou et al., 2022), suggesting that other factors than GLS reduction have shaped the genetic and metabolomic diversity in *B. napus*. All these elements are leading to a need to improve the knowledge of the genetic diversity of these specialized metabolites in a large panel of oilseed rape and their genetic control to better exploit the *B. napus* germplasm for breeding.

In our previous work we have reported the identification and annotation of GLSs and phenolics in a large diversity panel (Missinou et al., 2022) allowing us to document the metabolic diversity in 270 accessions of *B. napus* representing the different cultigroups grown worldwide and representing the history of selection (genotypes presenting low and high seed GLS content). To decipher the genetic bases of metabolic diversity of GLS and phenolics in oilseed rape metabolism, we performed a metabolic genome-wide association study (mGWAS). The present work provides an overview of the genetic control of these two principal families of phytochemicals in roots and leaves of oilseed rape and addresses the question of the impact of breeding for low seed GLS content by comparing the allele frequencies between "low-GLS" and "high-GLS" genotypes both for winter and spring types. Finally, we discussed the results in light of this genetic architecture and diversity in the different *B. napus* cultigroups, which provide avenues to design metabolypes of these specialized metabolites according to breeding needs.

2. Resultats and Discussion

2.1. Glucosinolates and Phenolic compounds identified in *B. napus*

The phytochemical dataset from (Missinou et al., 2022) was investigated to address the phytochemical contrasts between *B. napus* cultigroups and investigate the genetic architecture of GLS and PHL specialized metabolites. Table 1 details the biochemical nomenclature adopted for this dataset by Missinou et al. 2022. GLS molecules may be divided into four main types: Met-derived aliphatic-GLSs (G1 to G13), Phe/Tyr-derived phenylalkyl-GLSs (G14 to G18), Trp-derived indolic-GLSs (G19 to G23) and Branched and linear alkyl-GLSs (G24 to G36) (Missinou et al. 2022). Quantitative profiles of those four biochemical categories in leaves and roots of *B. napus* accessions are represented according to the six identified crop types (Figure 1).

Table 1: Nomenclature and sub-categories of glucosinolates and phenolics used in this work. Met-derived aliphatic GLSs are ranked according to the length (C3 to C6), then the decoration (sulfanyl-, sulfinyl-, alkenyl-, hydroxyalkenyl-) of their lateral chain, and the decoration. Branched and linear alkyl-GLSs are ranked according to the length (C3 to C7) of their lateral chain. Phenylpropanoids are ranked according to the number of acyl and hexose groups. Flavonols are ranked according to the nature of their aglycone (K=Kaempferol, Q=Quercetin, I=Isorhamnetin) and the number of acyl and hexose decorations.

			Glucosinolates				I	Phenolics
	ID	Abbreviated name	Semi systematic name	Trivial name		ID	Abbreviated name	Semi systematic name
	G01	3mSp	3-methylsulfanylpropyl-GLS	Glucoiberverin		PL01	CQA	Caffeoylquinic acid
	G02	3mSOp	3-methylsulfinylpropyl-GLS	Glucoiberin		PL02	FQA	Feruloylquinic acid
	G03	2Prop	2-propenyl-GLS	Sinigrin		PL03	Fer-Hx	Feruloyl-O-hexoside
	G04	4mSb	4-methylsulfanylbutyl-GLS	Glucoerucin	ids	PL04	Sin-Hx	Sinapoyl-O-hexoside
S	G05	4mSOb	4-methylsulfinylbutyl-GLS	Glucoraphanin	ou	PL05	Sin-diHx	Sinapoyl-O-dihexose
Ĕ	G06	3Ben	3-butenyl-GLS	Gluconapin	pa	PL06	diFer-diHx	Diferuloyl-O-dihexoside
ha	G07	(R)-2h3Ben	(R)-2-hydroxy-3-butenyl-GLS	Progoitrin	LO	PL07	diSin-diHx, i1	Disinapoyl-O-dihexoside, isomer I
i	G08	(S)-2h3Ben	(S)-2-hydroxy-3-butenyl-GLS	Epiprogoitrin	vlp	PL08	diSin-diHx, i2	Disinapoyl-O-dihexoside, isomer II
⋖	G09	5mSp	5-methylsulfanylpentyl-GLS	Glucoberteroin	en	PL09	Fer-Sin-diHx, i1	Feruloyl-O-sinapoyl-O-dihexoside, isomer I
	G10	5mSOp	5-methylsulfinylpentyl-GLS	Glucoalyssin	Å	PL10	Fer-Sin-diHx, i2	Feruloyl-O-sinapoyl-O-dihexoside, isomer II
	G11	4Pen	4-pentenyl-GLS	Glucobrassicanapin	_	PL11	Fer-Sin-diHx, i3	Feruloyl-O-sinapoyl-O-dihexoside, isomer III
	G12	2h4Pen	hydroxy-pentenyl-GLS	Gluconapoleiferin		PL12	triSin-diHx	Trisinapoyl-O-dihexoside
	G13	6mSh	6-methylsulfanylhexyl-GLS	Glucoesquerellin		PL13	Fer-diSin-diHx	Feruloyl-O-disinapoyl-O-dihexoside
÷	G14	2PE	2-phenylethyl-GLS	Gluconasturtiin		PL14	K-diHx	Kaempferol-O-dihexoside
la k	G15	p-4hPE	p-hydroxyphenylethyl-GLS	p-hydroxygluconasturtiin		PL15	K-F-diHx	Kaempferol-O-feruloyl-O-dihexoside
ž	G16	(R)-2hPE	(R)-2-hydroxy-2-phenetyl-GLS	Epiglucobarbarin		PL16	K-triHx, i1	Kaempferol-O-trihexoside, isomer I
ē	G17	(S)-2hPE	(S)-2-hydroxy-2-phenetyl-GLS	Glucobarbarin		PL17	K-triHx, i2	Kaempferol-O-trihexoside, isomer II
ᅕ	G18	m-3hPE	m-hydroxyphenylethyl-GLS	m-hydroxygluconasturtiin		PL18	K-C-diHx	Kaempferol-O-caffeoyl-O-dihexoside
	G19	3IM	3-indolylmethyl-GLS	Glucobrassicin		PL19	K-hF-diHx	Kaempferol-O-hydroxyferuloyl-O-dihexoside
ics	G20	1M-3IM	1-methoxyindol-3-ylmethyl-GLS	Neoglucobrassicin		PL20	K-S-diHx	Kaempferol-O-sinapoyl-O-dihexoside
9	G21	4M-3IM	4-methoxyindol-3-ylmethyl-GLS	4-Methoxylucobrassicin	S	PL21	K-hF-triHx	Kaempferol-O-hydroxyferuloyl-O-trihexoside
<u> </u>	G22	4h-3IM	4-hydroxyindol-3-ylmethyl-GLS	4-Hydroxylucobrassicin	lo	PL22	K-F-triHx	Kaempferol-O-feruloyl-O-trihexoside
	G23	1h-3IM	1-hydroxyindol-3-ylmethyl-GLS	1-Hydroxylucobrassicin	, or	PL23	K-S-triHx	Kaempferol-O-sinapoyl-O-trihexoside
	G24	nPr	n-propyl-GLS	-	lav	PL24	K-F-tetraHx	Kaempferol-O-feruloyl-tetrahexoside
÷	G25	iPr	1-methylethyl-GLS	Glucoputranjivin	ш.	PL25	K-S-tetraHx	Kaempferol-O-sinapoyl-O-tetrahexose
Ξ.	G26	2mPr	2-methylpropyl-GLS	Glucoconringianin		PL26	K-S-pentaHx	Kaempferol-O-sinapoyl-O-pentahexoside
r a	G27	1mPr	1-methylpropyl-GLS	Glucocochlearin		PL27	K-diSin-pentaHx	Kaempferol-O-disinapoyl-O-pentahexoside
ea	G28	2h2mPr	2-hydroxy-2-methylpropyl-GLS	Glucoconringiin		PL28	Q-triHx	Quercetin-O-trihexoside
. <mark>E</mark>	G29	1Et2hEt	1-ethyl-2-hydroxyethyl-GLS	Glucosisautricin		PL29	Q-C-triHx	Quercetin-O-caffeoyl-O-trihexoside
p	G30	Pent i	pentyl-GLS-isomer	-		PL30	Q-hF-triHx	Quercetin-O-hydroxyferuloyl-O-trihexoside
ar	G31	3mPent	3-methylpentyl-GLS	-		PL31	Q-diSin-tetraHx	Quercetin-O-disinapoyl-O-tretrahexose
ed	G32	Hex i1	hexyl-GLS iso. I	-		PL32	I-triHx	Isorhamnetin-O-trihexoside
Ъ.	G33	Hex i2	hexyl-GLS iso. II	-				
an	G34	Hept i1	heptyl-GLS iso. I	-				
8	G35	Hept i2	heptyl-GLS iso. II	-				
	G36	Hept i3	heptyl-GLS iso. III	-				

In leaves, Met-derived aliphatic-GLSs are the most represented GLS molecules, and their global content exhibits a wide quantitative variability among accessions. In roots, overall GLS content was higher compared to leaves (by approximately 8 to 10-fold),

Ph D. manuscript | Anani Amegan MISSINOU
with a main proportion of Met-derived aliphatic-GLSs and Phenylalkyl-GLS (= 'aromatics') (Figure 1B). In leaves, WOSR '0' accessions displayed about two-fold lower mean GLS contents compared to WOSR '+' accessions, while SOSR '0' displayed about four-fold lower mean GLS contents compared to '+' types (Figure 1A, Supplementary Table S1), suggesting that breeding of '0' varieties had a stronger impact on the content of Met-derived aliphatic-GLS in SOSR than in WOSR varieties.

In roots also, a similar stronger depressive impact of '0' breeding was observed in SOSR varieties on Met-derived aliphatic-GLS, compared to WOSR varieties (Figure 1B and Supplementary Table S1). This phenomenon had however few consequences on the total root GLS content, due to a higher proportion of the other GLS categories. Met-derived aliphatic-GLS content in swede and fodder types, although represented by fewer individuals, was in the same range as modern oilseed '0' varieties in both leaves and roots.

Met-derived aliphatic GLSs G1 to G13 compounds were found positively correlated between leaves and roots as well as with global GLS content in seeds (Figure S1). In addition, most aromatic phenylethyl-GLSs (G14-G18) were also positively correlated with branched and linear alkyl-GLSs (G30-G37) in leaves but not in roots. On the other hand, in roots, Met-derived aliphatic-GLSs (G1 to G13) were positively correlated, for the most part, with branched and linear alkyl-GLSs (G30-G37) (Figure S2). PHL molecules did not seem strongly correlated with either GLS in seeds, leaves nor roots (Figure S3). Within phenolics, the highest positive correlations are found between phenylpropanoid (PPN) diHexosides and to a lesser extent between Quercetin derivatives (Figure S4).



Figure 1. GLS content in leaves (A) and roots (B) of *B. napus* accessions from the different cultigroups. WOSR+ and WOSR 0: high- and low-GLS varieties of Winter Oilseed Rape, respectively. SOSR+ and SOSR 0: high- and low-GLS varieties of Spring Oilseed Rape.

2.2. Multivariate analyses to identify discriminating molecules between *B*.

napus crop types

For both GLS roots and leaves, Principal Component Analysis revealed weak clustering between crop types. To identify more precisely molecules discriminating the different *B. napus* types, we performed Linear Discriminant Analyses revealing specific molecules distinctive of each type (Figure 2). As expected, several Met-derived aliphatic-GLS (G7=progoitrin, G8=epiprogoitrin, G10=glucoalyssin, and G11= glucobrassicanapin) were preferentially accumulated in leaves of WOSR and SOSR '+' compared to '0' types. Higher leaf concentrations of branched and linear C6/C7 alkyl-GLSs were characteristics of WOSR varieties (Figure 2A). C3/C4 alkyl-GLS were typically accumulated at higher concentrations in the roots of swede varieties (Figure **2B**). Finally, PHL in leaves also discriminates swedes from the other *B. napus* types (Figure 2C) and WOSR from SOSR varieties. Swede accessions are characterized by high amounts of PL5 (Sin-diHex) and PL24 (Kaempferol-Feruloyl-tetraHex). PL21 (Kamepferol-hydroxyFeruloyl-triHex) characterizes SOSR while a series of Hydroxycinnamic Acid Glucosyl Esters (HAGE) were found distinctive of WOSR (Figure 2C).



Figure 2. PLS-DA of phytochemical diversity among *B. napus* **cultigroups.** PLS-DA highlighted discriminating ions, **(A)** aliphatic-GLSs between '+' and '0' accessions in leaves, **(B)** aliphatic-GLSs between '+' and '0' accessions in roots, **(C)** phenolics between WOSR and SOSR accessions in leaves. The aliphatic Met-derived GLSs discriminate WORS+/SORS+ accessions compared to their counterpart with WORS00/SORS00 in leaves (A) and roots (B), illustrating the selection effect of these compounds

in seeds. Branched/linear alkyl GLSs forms discriminates the subsp. *rapifera* (rutabaga) compared to the subsp. *oleifera* in leaves (*B*), with a propensity to compensate the aliphatic Met-derived GLSs in leaves of accessions with '0' status (A). HAGEs and an acylated kaempferol derivative with tetrahexosylated discriminate WORS from SORS and the acylated kaempferol highly hexosylated discriminates between the subsp. *rapifera* (rutabaga) and the subsp. *oleifera* (C).

2.3. GWA analysis to decipher the genetic control of GLS and PHL content in leaves and roots

Heritabilities were quite high on average (equal to 0.65, 0.81, and 0.75 for Leaf GLS, root GLs, and root PHL, respectively) with some variation between traits and organs (Table S2). Using 28,383 genome-wide SNPs with a mixed linear model, we dissected the genetic basis of the phenotypic variation present in 36 GLS and 32 PHL molecules, as well as for additional descriptive variables (sums of some sub-categories and total GLS/PHL content). Among the 270 *B. napus* accessions, a total of 203 individual QTLs corresponding to 104 genomic regions (hereafter referred to as mQTLs) were significantly associated with one or more phenotypic variables at a threshold of -log10 equal to 5.296.

Numerous mQTL, 63 and 41, respectively, were found on both subgenomes A and C. 62, 90, and 51 QTLs were detected for leaf GLS, root GLS, and root PHL, respectively, which correspond to 31, 33, and 54 mQTLs in the *B. napus* genome (Table S3a,b,c). Only a few mQTL were found for most of the compounds, and each of these mQTL displayed R² in a range between 3.1 and 37.5%. A network analysis of the multiple links between mQTL and GLSs and PHLs was drawn using the Cytoscape software, allowing a global overview of the genetic architecture of root and leaf phytochemistry in *B. napus* (Figure 3).



Figure 3: Genetic Networks of QTL controlling GLSs and PHLs in leaves (A) and roots (B). mQTLs and metabolites (nodes) are shown as diamonds and circles, respectively, and links (edges) indicate

the association between specialized metabolites and QTLs, with the thickness of the link proportional to the correlation. Distinct genomic architecture underlies independent control of SMs, especially different subcategories of PHLs (phenylpropanoids, non-acylated hexosylated flavonols, and acylated and hexosylated flavonols) and GLSs (aliphatic-GLSs, phenylalkyl-GLSs, indolic-GLSs, branched and linear alkyl-GLSs). QTLs historically associated with the content of Met-derived aliphatic-GLSs in the seeds are highlighted in red diamonds. Interestingly, a single connection between the foliar phenolic and glucosinolate subnetwork.

This analysis did not identify any master mQTL that would have controlled a large part of the specialized metabolites. Instead, the Figure 3 highlights that the genetic architecture of constitutive phytochemicals in *B. napus* consists in small networks of mQTL which independently control the different subcategories of GLSs and phenolics. No link except for one QTL was observed between GLS and PHL leaf networks, which is in accordance with the absence of correlation between GLS and PHL compound content in the leaves.

Four mQTL controlling total seed GLS content (indicated by red diamonds in the Figure 3) were also strongly involved in the control of a series of Met-derived aliphatic GLSs in both leaves and roots but did not display any genetic link with other categories of GLSs or phenolics. A complementary GWAS analysis conducted with a sub-panel of 00 accessions also identified an additional QTL on the chromosome A03 in the region previously reported by (Liu et al. 2020). Two abundant aliphatic-GLSs 4-methylsufinylbutyl (G5=glucoraphanin) and 3-butenyl-GLS (G6=gluconapin) were controlled by mQTLs that were not found involved in the control of seed total GLSs, highlighting that the genetic control of leaf and root Met-derived GLSs does not fully overlap with the QTL involved in total seed GLS content.

As reported in the previous work of (Missinou et al., 2022), short-chain Met-aliphatic-GLSs (G1=glucoiberverin, G2=glucoiberin, and G3=sinigrin) are accumulated at relatively low concentrations in leaves and roots of *B. napus* (contrasting with high abundance in some accessions of *B. oleracea*). In the present work, however, the genetic analysis identified a series of QTL involved in the control of their variations among the *B. napus* accessions. These mQTL network involved in the control of C3-aliphatic-GLSs was found independent from the above evoked mQTL networks involved in the control of C4- and C5-Met-derived aliphatic-GLSs. This independent genetic control on C3-aliphatic GLSs might have some biological significance in plant-biotic interactions, keeping in mind the strong affinity to sinigrin of gustatory receptors the *Brassicaceae* specialist *Pieris rapae* (Yang et al., 2021).

Branched and linear alkyl GLSs compose only a few percent of the total content of GLSs in most *B. napus* varieties (Missinou et al. 2022³²) but have occasionally been found to accumulate up to 1 µmol.g⁻¹ DW in leaves or 2.4 µmol.g⁻¹ DW in roots of some varieties. Figure 3 and Supplementary Table 1 show that different mQTL networks were found to distinctly control subgroups of those alkyl GLSs in both leaves and roots, apparently in the function of the length and decoration of their lateral chain. Similarly, distinct mQTL were found to control the series of rarely reported hydroxylated derivatives of phenylethyl-GLSs in roots (Supplementary Table 1), depending on if the hydroxylation was positioned on the ethyl (G16 and G17), or the position para-/meta of the phenyl (G15 and G18).

The genetic control of phenolics in leaves was also organized in several genetic subnetworks associated with the control of different biochemical flavonol and phenylpropanoid sub-categories. For flavonols, three genetic networks were involved

³² Cf. Chapitre II, article 1

Ph D. manuscript | Anani Amegan MISSINOU

in the control of flavonol diHexose, two networks were involved in the control of flavonol triHexose, and one QTL was involved in the control of PL24 = Kaempferol-tetraHexose. For hydroxycinnamic acid glycosyl esters (HAGE), different QTL were involved in the control of Acyl-monoHexosides, Acyl-diHexosides, Diacyl-diHexosides, and Diacyl-triHexosides. Finally, the two close compounds feruloyl-quinic acid and coumaroyl-quinic acid were controlled each by two distinct mQTL. Few mQTL were found connected to different biochemical subcategories. For one of them, mQTL A08_2.4, the allele variation had opposite effects on two of the most abundant leaf phenolics PL16 (Kaempferol-triHexose, i1) and PL04 (Sinapoyl-Hexose), whose relative proportions consistently contributed to the distinction of two main chemotypes in unsupervised analyses of *Brassica* accessions reported by Missinou et al. (2022).

GWAS was also conducted to analyze the proportions of metabolite subcategories. Results are indicated in Table S4. This approach allowed identifying the QTL C01_7.3 which was found involved in the control of the length of the lateral chain of Met-derived aliphatic GLS, the two alleles exerting opposite effects on the % of C4-aliphatic GLSs, and the % of C5-aliphatic GLSs (compared to total aliphatic GLSs). These diverging effects of the allelic variation at this QTL were found in both roots and leaves. In line with these results, two QTLs were also found in the control of the % of C3-aliphatic GLSs in leaves, and six QTLs were involved in the control of the % of C6-aliphatic GLSs in roots.

Consistently with its above-evoked opposite effects on PL16 and PL4, the allele variation at the A08_2.4 was found to also exert opposite effects on the relative proportions of phenylpropanoid and flavonol derivatives (expressed as % of total

Ph D. manuscript | Anani Amegan MISSINOU

phenolics). A08_2.4 thus seems to play a significant role in the higher proportion of HAGE in the leaves of winter varieties that have been highlighted by PLS-DA analyses (Figure 4).



Figure 4. Leaf concentrations of HAGE in the varieties from the four-oilseed rape cultigroups of *B. napus*.

Finally, GWA analysis identified a series of QTLs involved in the control of the proportion of the three flavonol aglycones kaempferol, quercetin, and isorhamnetin. Noteworthy, the allele variation on the mQTL A10_10.0 exerted opposite effects on the proportions of Kaempferol and Quercetin derivatives (over the total concentration of flavonoids), and the allele variation on the mQTL C3_7.8 exerted opposite effects on the proportions of quercetin and isorhamnetin derivatives. These results might suggest that those two QTLs could play specific roles in the regulation of the final biosynthetic steps of those three aglycones.

Allele frequencies

To further explore the possible impact of low seed-GLS breeding on the genomic control of *B. napus* phytochemistry, allele frequencies in each oilseed rape cultigroup

(WOSR+, WOSR0, SOSR+, and SOSR0) were mapped on the edges of the mQTL network (Figures 5 and 6).



Figure 5. Mapping of allele frequencies on the genetic networks controlling PHL and GLS in leaves of the four-oilseed rape cultigroups of *B. napus*. Diamonds represent mQTL, circles represent the metabolites. % of the allele controlling high concentration is indicated on each arrow. Red arrows indicate that % of high-concentration alleles is above 90%. Blue arrows indicate that the % of high-concentration alleles is below 10%.



Figure 6. Mapping of allele frequencies on the genetic networks controlling and GLS in roots of the four-oilseed rape cultigroups of *B. napus*. Diamonds represent mQTL, circles represent the metabolites. % of the allele controlling high concentration is indicated on each arrow. Red arrows indicate that % of high-concentration alleles is above 90%. Blue arrows indicate that the % of high-concentration alleles is below 10%.

As expected, the history of selection led to the fixation of alleles controlling lowglucosinolate at the four main mQTL involved in the control of C4/C5 Met-derived aliphatic GLSs, in both leaves and roots, and in both WOSR and SOSR accessions. SOSR 0 accessions were found to display a strong diminution, compared to SOSR+, of the proportion of low-concentration alleles at the QTL A03_1.4, which is involved in the control of root gluconasturtiin concentration (G14). This phenomenon was not observed in WOSR accessions and is consistent with the significant reduction of gluconasturtiin in the roots of SOSR0 accessions compared to other cultigroups (Figure 7).





Beyond this, the breeding of modern low seed-GLSs varieties appeared to have little impact on allele frequencies at almost every other mQTL sub-networks. Interestingly, the QTL A09_10.3 involved in the control of PL24 (Kampferol-FeruloyI-tetraHex) displayed almost exclusively the 'high-concentration allele' in SOSR accessions, but the 'low-concentration allele' in WOSR. This contrast was consistent with the globally higher concentrations observed in the SOSR accessions compared to WOSR accessions (Figure 8). Most mQTL displayed a medium range of allele frequencies, suggesting interesting perspectives for the breeding of varieties with targeted phytochemical profiles.



Figure 8. Leaf Kaempferol-feruloyl-tetraHex content in varieties from the four-oilseed rape cultigroups of *B. napus*

3. Experimental procedures

3.1. Plant material

A total of 270 rapeseed cultivars were chosen from major production areas worldwide to represent the genetic diversity of *B. napus*. This panel includes old and modern varieties, as well as different cultigroups: 130 accessions of WOSR, 120 SOSR, 10 fodder varieties, and 10 accessions of *B. napus* ssp. rapifera (i.e rutabaga/swede). In addition, WOSR and SOSR types were divided into low GLS in seeds (or '0') and high GLS (or '+') types. This information was deduced from pedigree information and validated with quantification of total GLS in seeds by Near-infrared Spectroscopy. Our dataset includes 63 '+' and 67 '0' WOSR as well as 74 '+' and 46 '0' SOSR (Supplementary Table 1). Details of growing conditions and sampling were previously

Ph D. manuscript | Anani Amegan MISSINOU

described in Missinou et al. (202X). Briefly, plants were cultivated in a phytotron with fertilized Hoagland solution. Three independent biological replicates were performed at 1.5-month intervals. Each replicate is a pool of eight plants grown in spatially randomized blocks. Root and leaf tissues of these plants were harvested, rapidly washed with tap water, frozen in liquid nitrogen, and stored at -80°C until lyophilization and further extraction procedure.

Glucosinolate and Phenolic compounds' quantification

A complete description of the extraction protocols, annotation methodology, and descriptive analyses was presented in Missinou et al. (2022). Briefly, glucosinolate extraction was performed using methanol/formic acid on freeze-dried powder samples. After centrifugation, the supernatant was collected and filtered to be further analyzed by mass spectrometry. Correct annotation of both glucosinolate and phenolic compounds was realized with HPLC-Orbitrap LTQ-XL analyses and software Xcalibur Qual Browser 2.2 SP1.48 (Thermo Fisher Scientific). Precise quantification of all 68 metabolites (including 36 glucosinolates and 32 phenolic compounds) was performed on UPLC-UV-TQD. Total data acquisition was controlled using Masslynx[™] (version 4.1) software and preprocessing with the Quanlynx[™] program (Waters Corporation).

Seed glucosinolate content quantification

All the accessions were grown under small cages for seed glucosinolate content quantification under the same environmental conditions per oilseed rape type. Spring accessions were grown in spring 2017 and winter accessions in the 2017-2018 season. Total seed glucosinolate content was assessed through Near-Infrared Reflectance (NIR). Acquisition of NIR spectra was performed using an FT-NIR

spectrometer (MPA, Bruker Optik GmbH, Germany). The range used for scanning was between 12 500 and 4 000 cm⁻¹ with a step of 8 cm⁻¹. Intact seeds (~5g) were placed in a standard ring cup and scanned. Each spectrum was the mean of 64 scans per sample. The results were predicted from an in-house calibration established for total glucosinolate content. The results were given as in µmol of total glucosinolates per seed at 9% humidity.

3.2. Genotyping

The whole panel was genotyped using the *Brassica* 60K Illumina Infinium array (Clarke et al. 2016). A total of 30,817 SNPs were scored and validated using a threshold of 5% for the minor allele frequency (MAF) and 30% for the missing values. The missing genotyping data were inferred using Beagle v3 software (Browning and Browning, 2009). All SNPs were physically anchored by BLASTing context sequences on the *Brassica napus* reference genome Darmor-bzh V10 (Rousseau-Gueutin et al., 2020). Only uniquely mapped context SNPs were included in the analysis.

3.3. Statistical analyses

Various multivariate analyses were performed to describe the phenotypic datasets. Principal component analysis (PCA) was achieved on the data using the R package FactoMineR (Lê et al., 2008) while the PLS-DA (Discriminant Analysis) was performed using the package MASS (Venables and Ripley, 2002). Spearman correlation coefficients (α =0.05) were calculated between metabolic compounds with the package corrplot (Wei T). Broad-sense heritability (h²) was calculated for each metabolic molecule as h²= σ g²/(σ g²+ σ e²) where σ g² is the genetic variance, σ e² is the residual error variance. The estimates of σ g² and σ e² were obtained from mixed linear models

Ph D. manuscript | Anani Amegan MISSINOU

fitted using the Imer function in Ime4 package (Bates et al., 2015) in R environment in R program (R Core Team, 2020).

3.4. Genome-Wide Association analyses

Adjusted quantitative values using the least-squares method were further processed through GWA analyses. These genetic analyses were conducted using the FastLMM algorithm (Lippert et al., 2011) with a kinship calculated using the Van Raden algorithm (Van Raden et al., 2008) for each chromosome as described by Rincent and coworkers (2014). In addition to the kinship, the three main axes of the PCoA explaining respectively 13.0, 9.1, and 6.0% of the variance were also added in the FastLMM analysis added as covariables. To run these different steps the R script developed by Negro and co-authors (2019) was adapted. The significance threshold was set based on the Bonferroni threshold of 5% calculated on a corrected SNP population as proposed by the simple M method developed by (Spinelli et al., 2011) (Gao et al., 2008; Spinelli et al., 2011). This method is based on the composite linkage disequilibrium (CLD) correlation between SNPs. The CLD was used to calculate the effective number of independent tests (M eff). Here, Meff was calculated and fixed to 5,544 SNPs. The threshold -log10(p) was thus defined at 5.296 for the *B. napus* panel.

3.5. Linkage disequilibrium approach to identify QTLs

From significant SNPs, QTL regions were defined as hereafter. First, for each significant SNP, Plink software was used to calculate the local LD decay (ré=0.25) at the chromosome scale. Physical intervals were deduced from LD borders around every significant SNP at each QTL, with a LD decay of 0.2. Then, a custom script (available on github/jferreira) was developed to define the confidence interval for each QTL. It

merged overlapping physical intervals for a single trait, which defined mQTL with its confidence interval. mQTL were named according to the physical position of the peak SNP.

3.6. Network analyses

A first visualization approach using MapChart (Voorrips, 2002 J of Heredity) was performed to identify co-localizing QTLs. Overlapping mQTLs for different metabolites and organs were renamed similarly to identify genomic regions controlling multiple metabolites. So, for each mQTL, all associated metabolites and the maximum R² were retrieved to be further included in a network analysis using the software Cytoscape 3.9.0 (Shannon et al. 2003). In this network, nodes represent QTL (source) or traits (target), and edges represent statistically significant links between loci and phenotypic traits. Edge widths were used to represent R^2 , which is, the percentage of phenotypic variance explained by each QTL for a given metabolite. In addition, the frequency of the allele contributing to an increase of the metabolite was calculated within each subset of the panel *i.e.* within WOSR+, WOSR0, SOSR+ and SOSR0.

Supplementary Table S1: Proportions of GLS and PHL sub-categories in the four-oilseed rape cultigroups

Leaf GLS

	SO	SR +	SO	SR -	WO	SR +	wo	SR -
	mean	min -	mean	min -	mean	min -	mean	min -
		max		max		max		max
Leaf GLS (nmol.g ⁻¹ DW)	3 028.8	955.5 - 6414	713.2	238.4 - 2039.4	4 213.0	1331.1 - 7384.4	1 944.9	674.2 - 3777.5
% Aliphatic-GLS / Total GLS	85.5	57.5 - 96.8	46.1	10.5 - 92.5	84.2	65.8 - 94	69.6	27.3 - 89
% Phenylalkyl-GLS / Total GLS	6.9	0.3 - 21.8	26.8	2.8 - 57.6	6.4	0.3 - 16.9	13.5	1.7 - 37
% Indolic-GLS / Total GLS	6.0	0.4 - 41.7	20.7	4.2 - 55.4	7.2	1.6 - 29.5	13.9	2.5 - 63.4
% Alkyl-GLS / Total GLS	1.5	0.1 - 11.8	6.4	0.3 - 40.7	2.3	0.2 - 19.3	3.0	0.4 - 29.5
% C3 / Aliphatic-GLS	0.2	0 - 0.6	0.4	0 - 3.8	0.2	0 - 0.5	0.2	0 - 0.3
% C4 / Aliphatic-GLS	46.5	21.7 - 98.4	39.9	24.5 - 77.5	44.9	30.4 - 79.9	45.8	31.7 - 66
% C5 / Aliphatic-GLS	53.3	1.4 - 78.3	59.4	21.5 - 75.3	54.9	19.9 - 69.6	54.0	33.8 - 67.8
% C6 / Aliphatic-GLS	0.0	0 - 0.4	0.3	0 - 2.9	0.0	0 - 0.5	0.1	0 - 1.6
% Methylsulfanyl- / Aliphatic-GLS	0.0	0 - 0.4	0.5	0 - 6.3	0.1	0 - 0.5	0.1	0 - 1.6
%Methylsulfinyl- / Aliphatic-GLS	26.1	0.9 - 62.9	27.7	2.2 - 64.4	13.5	2.7 - 40.1	15.7	2.7 - 52.6
%Alkenyl- / Aliphatic-GLS	38.7	6.5 - 73.3	36.3	1 - 79.4	50.1	15.6 - 68	53.1	14.7 - 79.7
% Hydroxyalkenyl- / Aliphatic-GLS	35.1	14 - 58	35.4	9.5 - 53.8	36.3	20 - 61	31.1	16.6 - 48.6
% Gluconasturtiin / Phenylalkyl-GLS	92.1	52.4 - 98.5	92.4	62.7 - 99.1	92.5	44.5 - 99.2	93.1	55.1 - 98.6
% Epiglucobarbarin / Phenylalkyl-GLS	2.4	0.3 - 10.1	2.5	0.4 - 9.1	2.0	0 - 5.8	1.9	0 - 5.6
% Neoglucobrassicin / Indolic-GLS	5.1	0.3 - 22.4	5.4	1.4 - 23.5	4.0	0.3 - 22.1	3.7	0.5 - 15.8
% 4M-I3MG / Indolic-GLS	5.8	0.2 - 44.6	5.0	1.1 - 31.9	4.6	1 - 28.4	4.8	0.7 - 27.3

Root GLS

	SO	SR +	SO	SR -	WO	SR +	wo	SR -
	mean	min -	mean	min -	mean	min -	mean	min -
Leaf GLS (nmol.g ⁻¹ DW)	31 855.0	14785.7 - 44287.9	17 329.1	5746.3 - 31507.5	30 848.9	20722.8 - 49463.3	22 328.6	9203.2 - 34355.4
% Aliphatic-GLS / Total GLS	44.5	21.9 - 80	21.0	7 - 50.4	44.9	30.4 - 60.1	32.9	12 - 49.9
% Phenylalkyl-GLS / Total GLS	40.3	2 - 58.2	49.4	1.7 - 77.3	45.5	24 - 61.4	54.6	31.8 - 74.6
% Indolic-GLS / Total GLS	14.3	6.8 - 34.4	29.0	7 - 64.5	9.0	4.5 - 22.3	11.9	5.3 - 31.5
% Alkyl-GLS / Total GLS	0.9	0.2 - 6.1	0.6	0.1 - 2.9	0.6	0.2 - 3.8	0.6	0.3 - 1.5
% C3 / Aliphatic-GLS	0.2	0 - 0.7	0.3	0.1 - 1	0.1	0 - 0.8	0.2	0 - 0.4
% C4 / Aliphatic-GLS	57.6	36 - 96.9	67.7	44.4 - 86.6	59.3	42.1 - 81.5	64.2	49.3 - 75.1
% C5 / Aliphatic-GLS	40.9	1.5 - 62.2	25.2	9 - 40.9	39.9	17.7 - 56.7	34.6	23.3 - 48.9
% C6 / Aliphatic-GLS	1.4	0.4 - 3.6	6.8	0.6 - 29.1	0.6	0.2 - 1.8	1.1	0.3 - 3.7
% Methylsufanyl- / Aliphatic-GLS	33.3	8.8 - 59	32.1	17.7 - 53.3	27.0	14.2 - 46.9	26.4	12.8 - 37.7
%Methylsulfinyl- / Aliphatic-GLS	34.3	23.2 - 49.6	50.8	30.4 - 63.3	33.4	23.3 - 46	41.7	27 - 64.6
%Alkenyl- / Aliphatic-GLS	7.3	0.6 - 22.6	2.2	0.2 - 7.3	10.4	4.9 - 18.1	7.9	2.1 - 14.5
% Hydroxyalkenyl- / Aliphatic-GLS	25.2	6 - 44.4	14.9	1.9 - 30.4	29.2	14.5 - 46.9	24.0	9.8 - 42.8
% Gluconasturtiin / Phenylalkyl-GLS	99.8	99.7 - 99.9	99.8	99.7 - 99.9	99.8	99.7 - 99.9	99.8	98.9 - 99.9
% Epiglucobarbarin / Phenylalkyl-GLS	0.0	0 - 0	0.0	0 - 0.2	0.0	0 - 0	0.0	0 - 0
% Neoglucobrassicin / Indolic-GLS	42.5	24.9 - 62.1	39.9	7.3 - 62.2	35.0	13.9 - 57.4	32.0	14.5 - 47
% 4M-I3MG / Indolic-GLS	15.0	7.2 - 27.8	14.8	6.3 - 31.8	22.5	13.2 - 36.8	21.5	13.3 - 31.5

Leaf Phenolics

	SOSR +		SO	SR -	WO	SR +	wo	SR -
	mean	min -						
Leaf PHENOLICS (nmol.g ⁻¹ DW)	14 794.9	9743.8 - 23065.9	16 096.7	9976.1 - 23468.9	14 190.2	7408.6 - 19019.3	15 064.5	10683.8 - 19294.2
% PPN / Phenolics	31.5	19.5 - 54.2	29.5	17.5 - 44.5	38.1	27.9 - 47.6	42.1	30 - 53.6
% Flavonol / Phenolics	68.5	45.8 - 80.5	70.5	55.5 - 82.5	61.9	52.4 - 72.1	57.9	46.4 - 70
%HQAs / PPN	15.7	5.9 - 31.1	14.3	6.6 - 27.5	11.3	5.3 - 18.3	10.7	3.2 - 18
%HAGEs / PPN	84.3	68.9 - 94.1	85.7	72.5 - 93.4	88.7	81.7 - 94.7	89.3	82 - 96.8
%PPN-monoAcyl / PPN	48.3	37.2 - 59.8	49.6	37.6 - 57.7	43.8	33.6 - 63.1	41.7	28.9 - 53.9
%PPN-diAcy I / PPN	47.3	35 - 57.2	46.2	38.2 - 58.7	51.2	30.2 - 62.8	52.9	39.7 - 65
%PPN-triAcyl / PPN	4.3	0.6 - 9.8	4.1	2.3 - 6.4	5.0	0.3 - 10.8	5.4	2 - 12.9
%Flavonol-diHex / Flavonols	5.3	1.9 - 42.2	3.4	1.2 - 10.9	5.3	2.3 - 13.5	5.2	2.5 - 13.1
%Flavonol-triHex / Flavonols	86.7	50.9 - 94.2	88.5	81 - 94.2	85.0	78.2 - 89.4	84.7	74 - 91.2
%Flavonol-tetraHex/ Flavonols	4.2	1.5 - 11.3	4.6	1.4 - 9.1	6.0	1.3 - 10.7	6.1	0.6 - 14.1
%Flavonol-pentaHex/ Flavonols	3.8	1.4 - 9.7	3.5	1.6 - 8.3	3.6	1.8 - 7.8	3.9	1.6 - 6.5
% Flavonol-monoacyl / Flavonols	65.1	50 - 82.7	64.4	51.3 - 78.9	64.4	52.1 - 75.8	67.5	47.5 - 81.8
% Flavonol-diacyl / Flavonols	3.7	0.2 - 7.8	3.4	1.4 - 5.9	4.2	0.4 - 9.5	4.2	1.6 - 8
% Kaempferol / Flavonols	81.3	62.8 - 97.1	81.0	66 - 91.6	78.7	68.9 - 90.8	80.7	70 - 88.5
% Quercetin / Flavonols	13.6	2 - 28.1	13.1	7.2 - 22.2	12.3	3.9 - 19.2	11.4	5.9 - 21.2
% Isorhamnetin / Flavonols	5.1	0.4 - 19.8	5.9	1 - 15.9	9.0	3.2 - 14.6	7.9	3.7 - 13.7

Ph D. manuscript | Anani Amegan MISSINOU

Supplementary Table S2: Heritabilities for the GLS and PHL compounds in the leaves and roots. GLSs annotated with two asterisks (**) were tentatively identified only from daughter anions and accurate parent mass; those with one asterisk (*) were further validated with certified reference plant materials; and those remaining without an asterisk were confirmed with authentic reference compounds.

	GLS compounds	Leaf	Root]		PHL compounds	Leaf
G1	Glucoiberverin**	0.675	0.669		PL1	CQA	0.758
G2	Glucoiberin	0.857	0.873]	PL2	FQA**	0.574
G3	Sinigrin	0.897	0.924]	PL3	Fer-Hex**	0.772
G4	Glucoerucin	0.000	0.760]	PL4	Sin-Hex**	0.706
G5	Glucoraphanin	0.880	0.885]	PL5	Sin-diHex**	0.802
G6	Gluconapin	0.931	0.918		PL6	diFer-diHex**	0.780
G7	Progoitrin	0.866	0.839]	PL7	diSin-diHex**	0.902
G8	Epiprogoitrin**	0.878	0.863]	PL8	diSin-diHex**	0.281
G9	Glucoberteroin**	0.579	0.834]	PL9	Fer-Sin-diHex**	0.668
G10	Glucoalyssin	0.855	0.917]	PL10	Fer-Sin-diHex**	0.916
G11	Glucobrassicanapin	0.899	0.939		PL11	Fer-Sin-diHex**	0.640
G12	Gluconapoleiferin**	0.873	0.946]	PL12	triSin-diHex**	0.929
G13	Glucoesquerellin**	0.398	0.867]	PL13	Fer-diSin-diHex**	0.951
G14	Gluconasturtiin	0.332	0.776]	PL14	Kae-diHex**	0.945
G15	p-Hydroxygluconasturtiin* / Homosinalbin*	0.806	0.273		PL15	Kae-Fer-diHex**	0.684
G16	Epiglucobarbarin*	0.782	0.839		PL16	Kae-triHex**	0.965
G17	Glucobarbarin	0.903	0.858		PL17	Kae-triHex**	0.865
G18	m-Hydroxygluconasturtiin**	0.551	0.283		PL18	Kae-Caf-diHex**	0.727
G19	Glucobrassicin	0.771	0.612]	PL19	Kae-OHFer-diHex**	0.924
G20	Neoglucobrassicin*	0.376	0.770]	PL20	Kae-Sin-diHex**	0.680
G21	4-Methoxyglucobrassicin*	0.627	0.304		PL21	Kae-OHFer-triHex**	0.896
G22	4-Hydroxyglucobrassicin**	0.751	0.658		PL22	Kae-Fer-triHex**	0.873
G23	1-Hydroxyglucobrassicin**	0.934	0.855]	PL23	Kae-Sin-triHex**	0.740
G24	Propyl-GLS*	0.665	0.941		PL24	Kae-Fer-tetraHex**	0.587
G25	Glucoputranjivin*	0.636	0.950		PL25	Kae-Sin-tetraHex**	0.306
G26	Glucoconringianin*	0.201	0.988		PL26	Kae-Sin-pentaHex**	0.477
G27	Glucocochlearin*	0.612	0.969		PL27	Kae-diSin-pentaHex**	0.593
G28	Glucoconringiin**	0.947	0.925		PL28	Quer-triHex**	0.824
G29	Glucosisautricin**	0.898	0.943		PL29	Quer-Caf-triHex**	0.857
G30	Pentyl-GLS	0.440	0.775		PL30	Quer-OHFer-triHex**	0.713
G31	3-methylpentyl-GLS *	0.551	0.831		PL31	Quer-diSin-tetraHex**	0.708
G32	Hexyl-GLS isomer I**	0.317	0.890	J	PL32	Iso-triHex**	0.729
G33	Hexyl-GLS isomer II**	0.318	0.749]			
G34	Heptyl-GLS isomer I**	0.346	0.859				
G35	Heptyl-GLS isomer II**	0.244	0.939				
G36	Heptyl-GLS isomer III**	0.654	0.831				

Supplementary Table S3a: QTL identified for seed total GLS, total GLS, and total Aliphatic GLS in the leaves and for individual leaf GLS compounds.

The number of significant SNP per mQTL, the position of the mQTL, and the peak SNP on the chromosome of the DarmorV10 reference sequence are indicated as well as the confidence interval of the mQTL, the p-value, the minor allele effect, and the R² of the peak SNP,

				Chro	nb		Desition of				CNID	
Panel	Organ	Pheno_trait	Nom_QTL	moso	SNP	Peak SNP	Position of	QTL_start	QTL_end	Pvalue	SINP	R ²
	0	-		me	S*		peak SNP				Weight	
Δ11	Seeds	S GIS	mOA09_4_1	Δ <u>09</u>	35	Bn Δ09 n2733282	4 101 935	1 902 968	4 697 411	1 79E-30	-16 473	0 375
	Seeds	S GIS	mQA09_5.7	A 09	8	Bn A09 p3940779	5 738 804	5 270 169	6 320 250	1.75E-10	8 316	0.575
	Soods	5_0L5	m0C02_60_4	C02	11	Bn C2 n51242096	60 470 827	50 555 206	61 602 706	5 00E 10	0.027	0.150
All	Seeus	5_0L3	mQC07_45_5	C02	11	Bii.C2.p31343360	45 520 007	45 530 510	45 526 220	3.900-10	5.557	0.130
All	Seeds	S_GLS	mQC07_45.5	C07	1	Bn.C7.p38813167	45 529 987	45 530 516	45 526 320	3.80E-00	-0.059	0.090
All	Seeds	S_GLS	mQC09_2.8	C09	2	Bn.C9.p2544007	2 741 686	1 410 217	2 939 083	1.01E-06	-7.744	0.095
All	Leaves	L_TOTAL_GLS	mQA09_4.1	A09	20	Bn.A09.p2733282	4 101 935	2 168 403	4 824 244	6.19E-15	-880.952	0.218
All	Leaves	L_TOTAL_GLS	mQA09_5.7	A09	3	Bn.A09.p4433127	6 123 496	5 724 361	6 123 621	5.59E-10	711.158	0.162
All	Leaves	L_TOTAL_GLS	mQC02_60.4	C02	2	Bn.C2.p50557457	59 747 993	59 583 978	60 798 255	1.46E-06	462.845	0.098
All	Leaves	L_TOTAL_GLS	mQC07_45.2	C07	1	Bn.C7.p38504198	45 188 119	45 150 595	45 391 377	3.20E-09	511.106	0.139
All	Leaves	L_Ali_GLS	mQA09_4.1	A09	20	Bn.A09.p3064935	4 416 190	2 168 403	4 824 244	2.83E-09	600.172	0.158
All	Leaves	L Ali GLS	mQA09 5.7	A09	3	Bn.A09.p3940779	5 738 804	5 724 361	6 123 621	1.17E-09	576.568	0.162
All	Leaves	L Ali GLS	mQC02_60.4	C02	1	Bn.C2.p51343986	60 479 827	59 975 293	60 798 255	2.37E-06	571.285	0.096
All	Leaves		mOC07_45_2	C07	1	Bn C7 n38504198	45 188 119	45 150 595	45 391 377	5 44F-10	509 044	0 150
All	Leaves	L_G01	m0C04_63.2	C04	6	Bn C4 n51439576	63 /3/ 171	63 196 9/8	64 242 879	1 51E-11	0 101	0.144
	Leaves	L_001	mQC09_03.2	0.00	1	Bn A00 p36008549	41 624 522	41 271 075	42 052 410	2.275.06	0.101	0.144
All	Leaves	L_002	mQA09_41.4	A09	1	Bii.A09.µ20006546	41 024 522	41 5/1 9/5	42 052 410	2.372-00	-0.217	0.051
All	Leaves	L_G02	mQC08_27.2	008	1	Bn.A09.p26246760	27 243 802	26 943 368	27 408 055	1.12E-11	-0.393	0.117
All	Leaves	L_G05	mQA02_7.4	A02	1	Bn.A02.p7935221	7 434 310	7 425 911	7 451 560	2.55E-07	25.059	0.114
All	Leaves	L_G05	mQA03_15.1	A03	1	Bn.A03.p14423477	15 084 497	15 079 670	15 109 389	3.27E-06	29.456	0.062
All	Leaves	L_G05	mQA03_21.5	A03	1	Bn.A03.p21075664	21 777 968	21 695 753	21 905 632	3.95E-06	27.089	0.060
All	Leaves	L_G05	mQC03_69.1	C03	1	Bn.C3.p60779962	69 096 367	69 090 146	69 097 145	2.05E-08	-40.312	0.133
All	Leaves	L_G05	mQC05_15.1	C05	4	Bn.C5.p13375414	15 108 846	14 907 326	16 240 199	1.03E-06	40.773	0.072
All	Leaves	L G05	mQC08 27.2	C08	1	Bn.A09.p26246760	27 243 802	26 943 368	27 408 055	3.03E-06	-49.596	0.066
All	Leaves	L G06	mQA06_40.8	A06	1	Bn.A06.p22397061	40 835 922	40 822 535	40 940 457	6.82E-07	85.289	0.115
All	Leaves	L G07	mOA09_4_1	A09	10	Bn A09 n2733282	4 101 935	2 757 086	4 824 244	7 16E-13	-254 552	0 190
	Leaves	L_G07	mQA09_5.7	A 09	2	Bn A09 p//33127	6 123 /96	5 724 361	6 123 621	2 10F-08	196 380	0.134
	Leaves	L_007	mQA10_15_5	A09	1	Bn.A05.p4455127	15 519 740	15 409 010	15 5 47 970	2.191-00	130.300	0.134
All	Leaves	L_G07	mQA10_15.5	A10	1	Bn.A10.p11332256	15 518 740	15 498 919	15 547 870	2.39E-06	-125.022	0.084
All	Leaves	L_G07	mQC02_60.4	C02	1	Bn.C2.p51343986	60 4 / 9 82 /	59 975 293	60 /98 255	6.32E-07	191.448	0.093
All	Leaves	L_G07	mQC07_45.2	C07	1	Bn.C7.p38504198	45 188 119	45 150 595	45 391 377	1.50E-08	146.886	0.119
All	Leaves	L_G08	mQA09_4.1	A09	11	Bn.A09.p2733282	4 101 935	2 757 086	4 824 244	3.26E-13	-4.621	0.166
All	Leaves	L_G08	mQA09_5.7	A09	3	Bn.A09.p3940779	5 738 804	5 724 361	6 123 621	6.68E-09	3.147	0.120
All	Leaves	L_G08	mQA10_15.5	A10	1	Bn.A10.p11332256	15 518 740	15 498 919	15 547 870	1.48E-06	-2.290	0.059
All	Leaves	L_G08	mQC02_60.4	C02	1	Bn.C2.p51343986	60 479 827	59 975 293	60 798 255	9.72E-08	3.671	0.077
All	Leaves	L G08	mQC07 45.2	C07	2	Bn.C7.p38504198	45 188 119	45 150 595	45 391 377	2.37E-09	2.783	0.129
All	Leaves	L G09	mOA02_25.8	A02	1	Bn.A02.p21256748	25 782 461	25 408 743	26 143 458	4.86E-06	0.121	0.057
	Leaves	L G10	mQA09_4_1	A09	2	Bn Δ09 n2733282	4 101 935	3 928 026	4 564 748	3 12E-08	-131 894	0.100
	Leaves	L G11	mQA03_14_6	A03	5	Bn A03 n13925731	14 556 259	1/ 531 159	14 619 769	3 23E-07	-172 869	0.104
	Leaves	L_011	mQA00_14.0	A00	15	Bn A00 p2722292	4 449 572	2 169 402	4 924 244	5.252-07	274 766	0.142
All	Leaves	L_011	IIIQA09_4.1	A09	15	BII.A09.µ2755262	4 446 575	2 106 405	4 624 244	3.80E-09	274.700	0.145
All	Leaves	L_G11	mQA09_5.7	A09	3	Bn.A09.p3940779	5 738 804	5 724 361	6 123 621	2.39E-08	231.566	0.112
All	Leaves	L_G11	mQC07_45.2	C07	1	Bn.C7.p38504198	45 188 119	45 150 595	45 391 377	4.46E-10	225.797	0.147
All	Leaves	L_G12	mQA03_14.6	A03	1	Bn.A03.p13925731	14 556 259	14 531 159	14 619 769	7.45E-07	-28.038	0.085
All	Leaves	L_G12	mQA09_4.1	A09	20	Bn.A09.p2733282	4 448 573	2 168 403	4 824 244	1.52E-10	50.295	0.153
All	Leaves	L_G12	mQA09_5.7	A09	3	Bn.A09.p3940779	5 738 804	5 724 361	6 123 621	1.30E-07	36.823	0.093
All	Leaves	L_G12	mQC07_45.2	C07	2	Bn.C7.p38504198	45 188 119	45 150 595	45 521 388	3.32E-12	41.478	0.165
All	Leaves	L G12	mQC09 2.8	C09	1	Bn.C9.p2639171	2 830 804	2 717 239	2 939 083	3.98E-07	-36.433	0.105
All	Leaves	L G13	mOA07_15.5	A07	2	Bn.A07.p10451411	15 508 523	15 466 683	15 512 343	1.42E-07	0.513	0.098
	Leaves	L G16	mOA10_18_3	Δ10	1	Bn A10 n16425155	18 318 219	18 312 553	18 318 458	3 02E-06	0.175	0.062
	Leaves	L_010	m0C06 21 2	C06	2	Bn.C6 p16402569	21 224 242	20 697 412	21 650 292	1 69E 07	0.175	0.002
All	Leaves	L_010	mQA07_15_5	007	2	Bii.C0.p10403508	15 407 200	15 404 902	15 510 342	4.082-07	-0.195	0.033
All	Leaves	L_G20	mQA07_15.5	A07	2	Bn.A07.p10430301	15 487 368	15 494 802	15 512 343	1.73E-07	3.530	0.103
All	Leaves	L_G24	mQA08_2.4	A08	1	Bn.A05.p19362820	2 430 607	2 426 678	2 436 637	1.20E-06	0.747	0.086
All	Leaves	L_G26	mQA01_28.6	A01	1	Bn.A01.p27220075	28 665 028	28 495 970	28 666 611	1.81E-06	0.333	0.131
All	Leaves	L_G26	mQA03_14.6	A03	1	Bn.A03.p14009871	14 655 406	14 647 804	14 658 827	2.42E-06	0.334	0.121
All	Leaves	L_G28	mQA01_27.0	A01	2	Bn.A01.p24320794	27 002 121	26 784 730	27 017 593	2.29E-07	3.071	0.062
All	Leaves	L_G28	mQA02_11.6	A02	1	Bn.A02.p11939983	11 635 301	11 600 706	11 656 615	1.21E-06	2.914	0.058
All	Leaves	L G28	mQA07 22.9	A07	1	Bn.A07.p16557893	22 874 120	22 869 635	22 883 447	4.93E-07	2.258	0.093
All	Leaves	L G28	mQA08 2.4	A08	1	Bn.A05.p19362820	2 430 607	2 426 678	2 436 637	9.77E-07	2.138	0.091
All	Leaves	L G28	mOA10 9.0	A10	6	Bn.A10.p5135356	8 980 780	8 646 149	9 659 628	1.64F-06	2.697	0.035
	Leaves	L G28	mOC04_4_2	C04	2	Bn C/ n2000007	A 220 / 20	1 227 054	A A 20 11 4	1 225 06	2.007	0.035
	Leaves	L_020	mQC0C 19 C	C04	 	Bn C6 n20107002	4 330 437	4 337 330	10 050 004	4.220-00	2.407	0.045
All	Leaves	L_G28	mucub_18.6	CU6	1	DII.CO.D3019/083	18 091 11/	19 021 002	18 828 084	1.90E-06	2.142	0.073
All	Leaves	L_G29	mQA07_22.9	A07	1	вп.А07.р16557893	22 874 120	22 869 635	22 883 447	4.09E-08	2.707	0.109
All	Leaves	L_G29	mQA08_2.4	A08	1	Bn.A05.p19362820	2 430 607	2 426 678	2 436 637	3.81E-06	2.214	0.082
All	Leaves	L_G29	mQC06_18.6	C06	2	Bn.C6.p33128511	19 128 118	18 691 069	19 191 048	6.61E-07	2.362	0.092
All	Leaves	L_G35	mQC01_22.6	C01	2	Bn.C1.p23580781	22 686 549	21 427 588	23 366 240	1.13E-06	1.599	0.077
Subpanel '0'	Leaves	L_Ali_GLS	mQA03_3.3	A03	2	Bn.C3.p4119233	3 333 800	3 261 241	3 436 286	1.13E-07	244.153	0.241
Subnanel '0'	Leaves		m0403_21_5	A03	1	Bn 403 n21294836	22,009,685	22 001 204	22 064 566	1 51E-07	-446 062	0 254

Ph D. manuscript | Anani Amegan MISSINOU

Supplementary Table S3b: QTL identified for total GLS and total Aliphatic GLS in the roots and individual root GLS compounds.

The number of significant SNP per mQTL, the position of the mQTL, and the peak SNP on the chromosome of the DarmorV10 reference sequence are indicated as well as the confidence interval of the mQTL, the p-value, the minor allele effect, and the R² of the peak SNP,

				Chro	nb		Desition of				CND	
Panel	Organ	Pheno_trait	Nom_QTL	moso	SNP	Peak SNP	Position of	QTL_start	QTL_end	Pvalue	SINP	R ²
				me	S*		реак SNP				weight	
All	Roots	R TOTAL GLS	mQA09 4.1	A09	1	Bn.A09.p2733282	4 101 935	3 928 026	4 564 748	9.61E-07	-3261.432	0.135
All	Roots	R Ali GLS	mQA09 4.1	A09	8	Bn.A09.p2733282	4 101 935	3 305 420	4 564 748	8.60E-09	-2344.994	0.164
All	Roots	R Ali GLS	mQC03 6.3	C03	1	Bn.C3.p5694509	6 263 633	6 251 497	6 309 512	1.90E-06	959.718	0.111
All	Roots	R Alkyl GLS	mQA01 0.2	A01	1	Bn.A01.p195253	206 067	171 443	217 424	2.47E-06	-97.418	0.109
All	Roots	R Alkyl GIS	mOA02_10_7	A02	1	Bn A02 n11086855	10 724 754	10 715 108	10 734 166	4 08F-06	-80 447	0 104
All	Roots	R Alkyl GIS	mOA09 41 4	A09	1	Bn A09 n25747213	41 371 883	41 352 833	41 371 975	4 66F-06	-109 126	0.106
	Roots	R Alkyl GIS	mOC01_41_1	C01	1	Bn Scaffold000587 n	41 231 035	41 107 680	41 265 287	3 42E-06	69 961	0 104
	Roots	R_Aro_GLS	mQ402_14	A02	0	Bn.302 n125/275	1 200 000	925 510	1 521 067	7.405.00	1970 505	0.162
	Roots	R_AI0_0L3	mQA05_1.4	AUS	1	Bn A05 p20222090	25 674 052	25 604 222	25 607 649	1 725 06	1 201	0.102
All	ROUIS	R_001	mQA04_0.2	A03	1	BII.A05.p20252969	35 074 055	35 004 225	55 097 046	4.75E-00	1.391	0.075
All	ROOLS	R_G02	mQA04_0.3	A04	1	Bri.A04.p326007	305 / 54	267 447	595 806	3.73E-00	1.839	0.078
All	ROOTS	R_GO2	mQA05_35.6	A05	1	Bn.A05.p20232989	35 674 053	35 604 223	35 697 648	2.64E-07	1.331	0.093
All	Roots	R_G02	mQA06_6.0	A06	1	Bn.A06.p5604858	6 001 197	5 932 485	6 032 185	4.41E-06	1.214	0.074
All	Roots	R_G02	mQA07_29.0	A07	1	Bn.A07.p22198393	28 973 208	28 959 008	28 986 262	3.70E-06	1.309	0.074
All	Roots	R_G03	mQA05_4.7	A05	1	Bn.A05.p4376029	4 682 698	4 680 330	4 701 008	3.30E-06	1.353	0.077
All	Roots	R_G03	mQA05_35.6	A05	1	Bn.A05.p20232989	35 674 053	35 604 223	35 697 648	3.28E-06	1.222	0.076
All	Roots	R_G04	mQA08_16.9	A08	1	Bn.A08.p12828472	16 877 051	16 869 042	16 959 305	2.46E-06	-325.496	0.117
All	Roots	R_G05	mQA03_1.4	A03	3	Bn.A03.p1353309	1 387 829	1 057 128	1 518 204	1.27E-07	-293.995	0.134
All	Roots	R_G07	mQA09_4.1	A09	17	Bn.A09.p2733282	4 101 935	2 757 086	4 564 748	9.81E-11	-709.243	0.179
All	Roots	R_G07	mQC07_45.2	C07	1	Bn.C7.p38504198	45 188 119	45 150 595	45 391 377	5.73E-07	385.612	0.127
All	Roots	R_G08	mQA09_4.1	A09	10	Bn.A09.p2733282	4 101 935	3 305 420	4 564 748	3.93E-09	-10.691	0.132
All	Roots	R_G10	mQA03_14.6	A03	3	Bn.A03.p13943992	14 576 370	14 531 159	14 619 769	3.92E-06	234.425	0.112
All	Roots	R_G10	mQC01_9.3	C01	4	Bn.C1.p8702563	9 347 634	8 952 014	9 409 553	4.04E-06	312.443	0.113
All	Roots	R G11	mQA09 4.1	A09	24	Bn.A09.p2733282	4 101 935	3 305 420	4 564 748	4.62E-11	-264.192	0.172
All	Roots	R G11	mQA09 5.7	A09	3	Bn.A09.p3940779	5 738 804	5 724 361	6 123 621	8.10E-07	168.560	0.111
All	Roots	R G11	mQC02 60.4	C02	1	Bn.C2.p51343986	60 479 827	59 975 293	60 798 255	8.51E-07	214.048	0.121
All	Roots	R G12	mOA03_6.7	A03	1	Bn.A03.p6484137	6 708 315	6 701 237	6 735 300	1.26E-06	81.842	0.110
	Roots	R_G12	mOA09_4.1	A09	30	Bn Δ09 n2733282	4 101 935	2 168 403	4 564 748	3 29E-15	-150.015	0.215
	Roots	R_G12	mQA09_5.7	A09	30	Bn A09 n3940779	5 738 804	5 724 361	6 123 621	9.66F-08	87 879	0.125
	Roots	R_G12	mQC07_45_2	C07	1	Bn (7 n3850/1198	15 188 119	15 150 595	45 391 377	4 08F-07	71 675	0.118
	Roots	R_G13	mQA03_21_5	A03	1	Bn Scaffold017///7 n	21 922 805	21 777 968	21 924 285	1.05E-06	18.450	0.102
	Roots	R_013	mQA03_21.5	A03	0	Bn. A02 p1254275	1 399 900	21777 508	1 521 067	7 565 00	1967 270	0.102
All	ROUIS	R_014	mQA10_11_2	AU5	0	BII.AUS.P1554575	1 245 715	11 210 5 62	11 277 702	2.025.00	-1007.370	0.102
All	ROOLS	R_G15	mQA10_11.2	A10	1	Bn.A10.p7278188	11 245 715	11 219 562	11 277 702	3.02E-00	-0.059	0.054
All	ROOTS	R_GIS	mQC07_8.1	07	2	Bn.C7.p5758601	8 081 200	7 903 546	9 946 152	1.78E-14	0.074	0.185
All	Roots	R_G15	mQC07_25.0	C07	4	Bn.C7.p20472212	25 068 904	24 465 514	24 959 521	3.08E-07	0.070	0.074
All	Roots	R_G16	mQA03_1.4	A03	9	Bn.A03.p1354375	1 388 899	825 519	1 531 067	1.34E-09	-2.261	0.130
All	Roots	R_G16	mQC03_1.5	C03	2	Bn.C3.p1657897	1 742 948	1 733 328	1 775 016	7.32E-07	-1.608	0.088
All	Roots	R_G17	mQA03_1.4	A03	7	Bn.A03.p1352996	1 387 520	1 022 324	1 531 067	8.56E-09	-1.110	0.113
All	Roots	R_G18	mQC08_33.0	C08	3	Bn.C8.p33909080	33 386 967	33 353 812	33 387 491	8.00E-07	0.280	0.074
All	Roots	R_G18	mQC08_39.3	C08	1	Bn.C8.p39594858	39 315 906	39 312 134	39 590 230	2.75E-06	-0.217	0.062
All	Roots	R_G19	mQC02_25.0	C02	23	Bn.C2.p22737175	25 596 414	24 358 921	25 678 957	6.70E-07	257.055	0.122
All	Roots	R_G20	mQA03_21.5	A03	1	Bn.Scaffold017447.p	21 922 805	21 777 968	21 924 285	9.17E-07	218.535	0.119
All	Roots	R_G21	mQC02_1.9	C02	7	Bn.C2.p1671010	1 922 348	1 899 577	2 015 753	6.43E-07	69.661	0.115
All	Roots	R_G23	mQC02_25.0	C02	21	Bn.C2.p22508661	24 991 589	24 358 921	25 678 957	5.35E-07	0.791	0.083
All	Roots	R_G23	mQC04_6.3	C04	1	Bn.C4.p4782065	6 325 136	6 323 669	6 334 609	3.11E-06	0.860	0.074
All	Roots	R_G24	mQA01_0.2	A01	1	Bn.A01.p195253	206 067	171 443	217 424	4.67E-06	-1.301	0.076
All	Roots	R_G25	mQA03_15.1	A03	1	Bn.C3.p21860069	15 080 034	15 079 670	15 096 712	1.51E-06	2.049	0.082
All	Roots	R G26	mQA01 0.2	A01	2	Bn.A01.p289703	292 387	171 443	603 453	2.53E-07	-90.658	0.131
All	Roots	R G26	mQA01 4.0	A01	1	Bn.A01.p3939508	3 999 067	3 997 800	4 001 202	1.04E-06	55.193	0.118
All	Roots	R G26	mQA01 7.8	A01	1	Bn.A01.p7512557	7 844 022	7 790 698	7 866 543	8.34E-08	-70.394	0.137
All	Roots	R G26	mQA02 10.7	A02	1	Bn.A02.p11086855	10 724 754	10 715 108	10 734 166	3.73E-07	-63,779	0.126
All	Roots	 R G26	mOA06 3 5	A06	-	Bn.A06.p3146367	3 592 077	3 531 508	3 592 432	8.50F-09	61.345	0.151
	Roots	R_G26	mOA08 21 4	A08	1	Bn Δ08 n17186700	21 282 572	21 358 550	21 386 9/2	1 495-06	58 152	0.116
	Roots	R_G26	mOA09 24 0	A00	1	Bn A05 n8220506	21 303 373	21 330 339	21 000 042	1.455-07	5/ /02	0.110
	Roote	P_G26	mQA10_24.9	AU0	1	Bn A10 n11211425	24 000 303	15 /00 000	15 500 422	1 1/15 00	50.040	0.125
	Roots	R_020	mQC01_20.0	A10	1	Bn.A10.p11311425	10 05 4 900	10 498 823	20.066.300	1.14E-00	39.049 45.734	0.118
	ROUTS	n_020	mQC01_30.9	C01	1	Dii.AUI.0100/93/0	30 954 899	30 931 532	30 906 289	2.58E-Ub	45.724	0.112
All	ROOTS	K_G26	mQC01_41.1	CU1	1	вп.5сатто10000587.р	41 231 035	41 10/ 680	41 265 287	2.43E-06	48.301	0.112
All	Roots	к_G26	mQC04_20.7	C04	1	вп.с4.р16395955	20 /42 008	20 467 125	215/4015	4.66E-07	105.486	0.128
AII	Roots	R_G26	mQC08_18.9	C08	2	Bn.C8.p21372460	19 660 415	18 855 419	19 681 529	8.68E-07	46.829	0.118

Supplementary Table S3b (end)

Panel	Organ	Pheno_trait	Nom_QTL	Chro moso me	nb SNP S*	Peak SNP	Position of peak SNP	QTL_start	QTL_end	Pvalue	SNP Weight	R²
All	Roots	R_G28	mQA01_2.4	A01	1	Bn.A01.p2360909	2 403 037	2 376 823	2 443 091	3.20E-06	1.518	0.083
All	Roots	R_G28	mQA03_8.3	A03	5	Bn.A03.p7965237	8 289 745	8 261 376	8 299 818	2.37E-10	2.331	0.149
All	Roots	R_G28	mQA03_15.1	A03	1	Bn.C3.p21860069	15 080 034	15 079 670	15 096 712	2.96E-08	1.447	0.114
All	Roots	R_G28	mQA07_2.0	A07	1	Bn.A07.p1658486	1 987 574	1 987 096	2 002 193	1.31E-07	-2.137	0.107
All	Roots	R_G28	mQC03_17.3	C03	1	Bn.C3.p16251571	17 290 744	17 290 676	17 294 237	3.47E-08	1.462	0.113
All	Roots	R_G28	mQC05_48.2	C05	2	Bn.Scaffold00956.p3	48 265 570	47 301 809	48 646 685	2.01E-06	-2.946	0.091
All	Roots	R_G29	mQA01_2.4	A01	1	Bn.A01.p2360909	2 403 037	2 376 823	2 443 091	4.17E-07	2.490	0.097
All	Roots	R_G29	mQA03_8.3	A03	5	Bn.A03.p7965237	8 289 745	8 261 376	8 299 818	1.15E-10	3.475	0.152
All	Roots	R_G29	mQA03_15.1	A03	1	Bn.C3.p21860069	15 080 034	15 079 670	15 096 712	4.49E-09	2.210	0.125
All	Roots	R_G29	mQA06_1.6	A06	1	Bn.A06.p1312619	1 645 125	1 618 247	1 680 171	2.13E-06	1.774	0.084
All	Roots	R_G29	mQA07_2.0	A07	1	Bn.A07.p1658486	1 987 574	1 987 096	2 002 193	1.02E-07	-3.188	0.108
All	Roots	R_G29	mQC03_17.3	C03	1	Bn.C3.p16251571	17 290 744	17 290 676	17 294 237	1.00E-06	1.882	0.089
All	Roots	R_G29	mQC05_48.2	C05	1	Bn.Scaffold00956.p3	48 265 570	47 301 809	48 646 685	1.09E-06	-4.535	0.095
All	Roots	R_G30	mQA04_2.5	A04	1	Bn.A04.p2542000	2 540 344	2 490 530	2 645 854	4.57E-07	0.980	0.088
All	Roots	R_G30	mQC04_44.1	C04	1	Bn.C4.p35007222	44 134 223	44 029 562	44 134 763	9.87E-07	0.897	0.082
All	Roots	R_G31	mQC03_1.5	C03	1	Bn.C3.p1524306	1 589 608	1 573 247	1 648 221	3.67E-06	5.916	0.087
All	Roots	R_G31	mQC04_44.1	C04	1	Bn.C4.p35007222	44 134 223	44 029 562	44 134 763	1.10E-06	5.854	0.090
All	Roots	R_G32	mQA04_2.5	A04	1	Bn.A04.p2542000	2 540 344	2 490 530	2 645 854	3.72E-06	11.708	0.092
All	Roots	R_G32	mQC03_1.5	C03	6	Bn.C3.p1524306	1 589 608	1 565 446	2 166 615	1.22E-08	14.304	0.129
All	Roots	R_G32	mQC03_3.4	C03	2	Bn.C3.p3111862	3 463 918	3 279 879	3 503 807	3.19E-07	-11.890	0.065
All	Roots	R_G32	mQC03_7.2	C03	1	Bn.C3.p6640991	7 223 143	7 178 768	7 297 085	3.48E-06	-10.893	0.048
All	Roots	R_G32	mQC04_44.1	C04	1	Bn.C4.p35007222	44 134 223	44 029 562	44 134 763	1.78E-06	11.341	0.094
All	Roots	R_G33	mQA06_10.1	A06	3	Bn.A06.p9292257	10 134 768	9 956 035	10 250 775	2.13E-06	1.535	0.081
All	Roots	R_G33	mQA07_29.0	A07	1	Bn.A07.p22251229	29 031 875	29 019 013	29 079 529	2.71E-06	1.325	0.078
All	Roots	R_G34	mQC04_44.1	C04	1	Bn.C4.p35007222	44 134 223	44 029 562	44 134 763	4.70E-08	3.067	0.110
All	Roots	R_G36	mQA10_15.8	A10	2	Bn.A10.p12900258	15 878 552	15 833 547	15 911 696	3.07E-06	0.488	0.071
Subpanel '0'	Roots	R_TOTAL_GLS	mQA04_10.5	A04	2	Bn.A04.p7541762	10 511 503	10 511 066	10 570 875	5.24E-08	3937.245	0.277
Subpanel '0'	Roots	R_TOTAL_GLS	mQA04_3.3	A04	3	Bn.A04.p3357700	3 369 708	3 230 824	3 396 425	2.19E-06	3146.280	0.253

Supplementary Table S3c: QTL identified for total PHL, subcategories in the leaves, and for individual leaf PHL compounds.

The number of significant SNP per mQTL, the position of the mQTL, and the peak SNP on the chromosome of the DarmorV10 reference sequence are indicated as well as the confidence interval of the mQTL, the p-value, the minor allele effect, and the R² of the peak SNP,

				Chro	nb		Desition of					
Panel	Organ	Pheno_trait	Nom_QTL	moso	SNP	Peak SNP	POSICIONO	QTL_start	QTL_end	Pvalue	SINP_Weig	R²
				me	S*		реак змр				IIL	
All	Leaves	L_TOT_PPN	mQA08_2.4	A08	16	Bn.A08.p2897827	2 871 709	1 588 102	9 998 472	1.34E-08	2.481	0.116
All	Leaves	L_HAGE	mQA08_2.4	A08	26	Bn.A08.p2897827	2 871 709	1 588 102	11 101 588	4.30E-08	426.479	0.142
All	Leaves	L_HQA	mQA07_8.6	A07		Bn.A07.p3729992	8 641 107	8 537 207	8 630 300	1.41E-06	79.735	0.111
All	Leaves	L_HQA	mQC04_63.2	C04	5	Bn.C4.p51194144	63 195 304	62 322 712	63 603 792	7.63E-07	92.279	0.116
All	Leaves	L_HQA	mQC07_41.8	C07	1	Bn.Scaffold01327.p4	41 818 429	41 783 926	42 066 234	8.16E-07	104.330	0.112
All	Leaves	L_ISOR	mQC07_38.5	C07	3	Bn.C7.p32304272	38 494 702	37 542 265	38 623 201	9.12E-08	-147.637	0.131
All	Leaves	L_ISOR	mQC08_33.0	C08	1	Bn.C8.p34052524	33 536 019	33 503 465	33 680 821	4.03E-06	137.840	0.108
All	Leaves	L_QUER	mQA06_41.2	C07	1	Bn.A06.p22806976	41 252 557	41 105 243	41 332 296	4.30E-06	-236.475	0.110
All	Leaves	L_QUER	mQC03_7.8	C03	9	Bn.C3.p7566009	8 225 531	8 118 430	8 800 845	4.42E-07	229.454	0.123
All	Leaves	L_PL01	mQA04_22.4	A04	1	Bn.C4.p53155836	22 487 493	22 384 755	22 732 061	1.40E-06	-63.638	0.106
All	Leaves	L_PL01	mQC04_63.2	C04	3	Bn.C4.p51194144	63 195 304	63 167 007	63 603 792	3.98E-09	84.765	0.148
All	Leaves	L_PL02	mQA06_2.9	A07	1	Bn.A06.p2348428	2 897 295	2 853 938	2 915 040	3.10E-06	51.189	0.102
All	Leaves	L_PL02	mQC02_4.6	C02	1	Bn.C2.p4066068	4 678 159	4 662 633	4 680 910	2.61E-06	-67.370	0.106
All	Leaves	L_PL03	mQA08_2.4	A08	27	Bn.A08.p8426380	11 420 680	2 864 626	12 840 615	7.21E-07	108.074	0.113
All	Leaves	L_PL04	mQA08_2.4	A08	10	Bn.A08.p2543882	2 463 565	2 354 646	9 998 472	1.20E-06	138.591	0.114
All	Leaves	L_PL05	mQC03_7.8	C03	6	Bn.C3.p7115054	7 690 464	7 570 337	7 825 754	3.26E-06	21.829	0.058
All	Leaves	L_PL05	mQC07_50.7	C07	2	Bn.C7.p43655170	50 687 714	50 685 410	50 779 734	1.67E-06	23.857	0.101
All	Leaves	L_PL06	mQC08_33.0	C08	14	Bn.C8.p33520874	33 008 931	31 822 119	33 964 436	1.82E-11	22.399	0.170
All	Leaves	L_PL10	mQC08_33.0	C08	27	Bn.C8.p33520874	33 008 931	31 818 983	33 964 436	2.25E-14	174.670	0.224
All	Leaves	L_PL11	mQC08_33.0	C08	6	Bn.C8.p33520874	33 008 931	32 926 325	33 852 566	9.26E-07	9.830	0.099
All	Leaves	L_PL12	mQA01_28.6	A01	2	Bn.C1.p40541584	28 666 611	28 665 028	28 800 699	1.16E-09	25.431	0.146
All	Leaves	L_PL13	mQA01_28.6	A01	2	Bn.C1.p40541584	28 666 611	28 665 028	28 800 699	2.52E-17	28.658	0.252
All	Leaves	L_PL13	mQC01_47.5	C01	7	Bn.C1.p39438146	47 485 457	46 305 054	47 941 447	3.78E-10	-24.892	0.148
All	Leaves	L_PL13	mQC09_50.3	C09	1	Bn.C1.p39423755	50 341 575	49 893 806	51 132 226	3.24E-08	25.306	0.127
All	Leaves	L_PL14	mQA03_15.0	A03	1	Bn.A03.p14385006	15 044 467	15 030 277	15 046 651	9.90E-07	56.042	0.111
All	Leaves	L_PL14	mQA03_17.0	A03	1	Bn.A03.p16342771	17 056 851	17 019 835	17 084 268	1.59E-06	52.442	0.104
All	Leaves	L_PL14	mQA07_21.8	A07	1	Bn.A07.p15622870	21 858 070	21 848 070	21 885 454	1.41E-06	-79.691	0.111
All	Leaves	L_PL14	mQA09_48.5	A09	1	Bn.A09.p32344408	48 517 803	48 451 623	48 694 471	6.00E-08	-75.958	0.129
All	Leaves	L_PL14	mQC03_38.0	C03	1	Bn.A06.p15966650	38 070 801	37 912 954	38 347 211	3.96E-06	-59.694	0.102
All	Leaves	L_PL14	mQC04_2.1	C04	1	Bn.C4.p1267536	2 084 425	1 919 303	2 176 213	5.83E-08	-86.603	0.130
All	Leaves	L_PL15	mQA02_7.3	A02	1	Bn.A02.p7788840	7 288 610	7 285 951	7 434 310	3.15E-06	-27.554	0.097
All	Leaves	L_PL15	mQA10_15.7	A10	1	Bn.A08.p4876592	15 755 072	15 734 484	15 806 724	3.04E-06	20.071	0.094
All	Leaves	L_PL16	mQA03_26.4	A03	1	Bn.A03.p25473832	26 435 455	26 374 939	26 458 423	2.08E-06	-371.784	0.118
All	Leaves	L_PL16	mQA08_2.4	A08	188	Bn.A08.p2543882	2 463 565	1 588 102	12 926 354	5.89E-09	-402.172	0.152
All	Leaves	L_PL16	mQA10_17.6	A10	1	Bn.A10.p14547767	17 603 723	17 600 997	17 699 274	1.40E-06	350.103	0.120
All	Leaves	L_PL17	mQA10_17.6	A10	2	Bn.A10.p14636670	17 699 274	16 766 311	17 709 209	3.97E-08	38.260	0.128
All	Leaves	L_PL17	mQC07_38.5	C07	3	Bn.C7.p32098446	38 294 955	37 542 265	38 623 201	1.15E-06	-48.308	0.108
All	Leaves	L_PL19	mQA03_2.5	A03	1	Bn.A03.p2469223	2 488 544	2 477 315	2 554 138	4.51E-06	13.603	0.091
All	Leaves	L_PL19	mQA03_11.7	A03	1	Bn.A03.p11202719	11 727 815	11 702 147	11 758 697	6.17E-07	-18.647	0.106
All	Leaves	L_PL19	mQA09_48.5	A09	1	Bn.A09.p32344408	48 517 803	48 451 623	48 694 471	6.82E-07	-18.112	0.105
All	Leaves	L_PL19	mQC04_2.1	C04	1	Bn.C4.p1267536	2 084 425	1 919 303	2 176 213	4.76E-07	-21.219	0.109
All	Leaves	L_PL20	mQA07_20.2	A07	1	Bn.A07.p13971525	20 187 022	20 179 474	20 215 502	7.42E-08	30.323	0.121
All	Leaves	L_PL20	mQA10_20.2	A10	1	Bn.A10.p17089148	20 253 766	20 079 951	20 304 113	4.15E-06	-18.733	0.094
All	Leaves	L_PL24	mQA09_10.3	A09	1	Bn.A09.p8635608	10 359 219	10 358 583	10 359 388	9.02E-07	-84.278	0.110
All	Leaves	L_PL28	mQA08_2.4	A08	59	Bn.C8.p9243618	11 259 004	2 553 365	11 877 121	3.60E-09	-112.802	0.150
All	Leaves	L_PL28	mQC03_5.7	C03	1	Bn.C3.p5237605	5 762 597	5 749 646	5 815 616	3.76E-06	84.046	0.105
All	Leaves	L_PL29	mQA10_2.4	A10	1	Bn.A10.p1582783	2 381 524	2 371 823	2 597 222	4.20E-06	77.439	0.104
All	Leaves	L_PL29	mQC03_7.8	C03	24	Bn.C3.p7248169	7 839 472	5 749 646	9 304 028	6.23E-11	107.927	0.175
All	Leaves	L_PL30	mQA10_10.0	A10	2	Bn.A10.p6120400	10 025 910	9 927 327	10 033 507	1.31E-10	185.727	0.173
All	Leaves	L_PL32	mQC07_38.5	C07	3	Bn.C7.p32304272	38 494 702	37 542 265	38 623 201	9.12E-08	-147.637	0.131

Supplementary Table S4: QTL identified for total PHL, subcategories in the leaves, and individual leaf PHL compounds.

The number of significant SNP per mQTL, the position of the mQTL, and the peak SNP on the chromosome of the DarmorV10 reference sequence are indicated as well as the confidence interval of the mQTL, the p-value, the minor allele effect, and the R² of the peak SNP,

Compo	Organ	Pheno_trait	Nom_QTL	Chrom	nb SND S*	Peak SNP	Position of	QTL_start	QTL_end	Pvalue	SNP Weight	R ²
GIS	leaves	ali GIS GIS	mOA03 22.0	Δ03	1	Bn 403 n21294836	22 009 685	22 001 204	22 064 566	3 50E-06	-5.816	0.087
GIS	leaves		mQA09_22.0	A03	1	Bn A09 n2733282	4 101 935	3 305 420	4 564 748	6.37E-16	-10.658	0.007
GLS	leaves		mQC09_6.8	00	4	Bn C9 n6344887	6 804 840	6 488 132	6 992 962	9 38F-07	-5 754	0.105
GLS	leaves	ind GIS GIS	mQA03_21_1	A03	1	Bn A03 n20417630	21 162 949	21 154 400	21 242 020	3.89E-06	3 155	0.079
GLS	leaves	ind GIS GIS	mQA09_4_1	A09	3	Bn A09 n2733282	4 101 935	3 305 420	4 564 748	3 54F-10	4 645	0 113
GLS	leaves	alkyl GLS GLS	mQA01_2.9	A01	1	Bn.A01.p27988223	29 764 089	29 420 126	31 229 077	3.58E-08	1.791	0.108
GLS	leaves	C3 tot ali	mQA08_22.9	A08	1	Bn.A08.p18750116	23 050 234	22 909 753	23 267 389	4.63E-06	0.110	0.056
GLS	leaves	C3 tot ali	mQC07 13.0	C07	3	Bn.C7.p10181642	13 721 335	9 984 488	16 394 951	2.26E-06	0.128	0.061
GLS	leaves	C4 tot ali	mQC01 7.3	C01	24	Bn.C1.p6868720	7 360 772	7 061 851	7 469 021	3.72E-11	-7.162	0.161
GLS	leaves	C5 tot ali	mQA07 11.5	A07	1	Bn.A07.p6541352	11 528 722	11 404 433	11 537 868	3.13E-06	-3.912	0.085
GLS	leaves	C5 tot ali	mQC01 7.3	C01	24	Bn.C1.p6854343	7 334 923	7 061 851	7 469 021	1.48E-08	7.256	0.123
GLS	leaves	C6 tot ali	mQA07 13.3	A07	1	Bn.A07.p8334245	13 356 426	13 341 127	13 383 332	4.19E-06	0.089	0.055
GLS	leaves	C6 tot ali	mQC04 7.0	C04	1	Bn.C4.p5393354	7 084 250	6 870 129	7 148 513	4.54E-06	0.095	0.048
GLS	leaves	C6 tot ali	mQC07 25.0	C07	1	Bn.A07.p8096658	24 959 521	24 866 759	25 033 026	9.62E-07	0.095	0.061
GLS	leaves	C6 tot ali	mQC09 61.6	C09	1	Bn.C9.p51326764	61 689 514	61 555 243	61 871 488	1.03E-06	0.098	0.063
GLS	leaves	hydroxyalkenyl_tot_ali	mQA03_8.3	A03	1	Bn.A03.p7793076	8 126 678	7 986 653	8 714 811	1.05E-07	4.118	0.106
GLS	roots	ali GLS .GLS	mQA06 42.5	A06	1	Bn.A06.p23776151	42579204	42 569 236	42 606 575	2.78E-06	-3.132	0.084
GLS	roots	ali_GLSGLS	mQA09_4.1	A09	1	Bn.A09.p1765907	3543971	3 517 509	3 559 703	3.52E-06	-3.492	0.084
GLS	roots	ind GLS .GLS	mQA01 32.1	A01	1	Bn.A01.p28462071	32166670	32 131 188	32 682 417	2.88E-07	2.855	0.098
GLS	roots	ind_GLSGLS	mQA03_1.4	A03	3	Bn.A03.p1352996	1387520	1 057 128	1 518 204	9.28E-08	2.984	0.105
GLS	roots	ind_GLSGLS	mQA04_0.3	A04	2	Bn.A04.p320972	360 959	267 447	575 558	5.06E-07	-3.284	0.095
GLS	roots	ind_GLSGLS	mQA04_10.8	A04	1	Bn.A04.p7797300	10 811 589	10 806 448	10 828 098	3.38E-06	-3.019	0.083
GLS	roots	ind_GLSGLS	mQA05_39.1	A05	2	Bn.A05.p23143046	39163555	39 137 056	39 326 383	1.49E-06	-2.625	0.087
GLS	roots	ind_GLSGLS	mQC04_28.6	C04	1	Bn.C4.p22318210	28653565	21 574 015	29 192 858	3.42E-07	4.086	0.100
GLS	roots	alkyl_GLS_GLS	mQA02_10.7	A02	1	Bn.A02.p11086855	10724754	10 715 108	10 734 166	1.31E-06	-0.263	0.072
GLS	roots	alkyl_GLS_GLS	mQA09_41.4	A09	1	Bn.A09.p25747213	41371883	41 352 833	41 371 975	5.65E-07	-0.371	0.080
GLS	roots	alkyl_GLS_GLS	mQC02_44.6	C02	22	Bn.C2.p38175317	44611642	24 358 921	54 239 899	4.26E-09	-0.266	0.059
GLS	roots	C4_tot_ali	mQC01_7.3	C01	25	Bn.C1.p6854343	7334923	7 061 851	7 469 021	5.76E-08	-6.405	0.114
GLS	roots	C5_tot_ali	mQC01_7.3	C01	25	Bn.C1.p6854343	7334923	7 061 851	7 469 021	2.46E-07	6.111	0.104
GLS	roots	C6_tot_ali	mQA04_0.3	A04	1	Bn.A04.p324144	364191	267 447	575 558	4.61E-06	-1.075	0.073
GLS	roots	C6_tot_ali	mQA05_35.6	A05	4	Bn.A05.p20319622	35757678	35 752 192	36 087 858	1.50E-07	0.812	0.092
GLS	roots	C6_tot_ali	mQA08_17.9	A08	1	Bn.A08.p13785675	17904116	17 825 223	17 952 491	4.83E-06	1.003	0.072
GLS	roots	C6_tot_ali	mQC01_1.4	C01	1	Bn.C1.p1471732	1410757	1 406 503	1 438 140	1.30E-06	0.784	0.079
GLS	roots	C6_tot_ali	mQC02_57.2	C02	3	Bn.C2.p48143863	57182578	57 100 829	57 635 562	1.80E-06	0.894	0.072
GLS	roots	C6_tot_ali	mQC07_13.0	C07	3	Bn.C7.p9955727	13038379	9 984 488	16 394 951	2.08E-08	1.240	0.106
GLS	roots	methylsulfinyl_tot_ali	mQA03_22.0	A03	1	Bn.A03.p21294836	22009685	22 001 204	22 064 566	2.37E-07	3.201	0.100
GLS	roots	methylsulfinyl_tot_ali	mQA09_4.1	A09	14	Bn.A09.p1989556	3320549	2 168 403	4 564 748	1.59E-11	4.266	0.160
GLS	roots	methylsulfinyl_tot_ali	mQC03_13.3	C03	1	Bn.Scaffold01023.p	13353482	13 351 056	13 479 153	2.94E-08	-2.834	0.109
GLS	roots	alkenyl_tot_ali	mQA03_2.0	A03	1	Bn.A03.p2005352	2084799	2 045 223	2 110 267	2.19E-06	1.111	0.078
GLS	roots	alkenyl_tot_ali	mQA09_4.1	A09	1	Bn.A09.p1989556	3320549	3 305 420	3 451 098	9.90E-07	-1.427	0.085
GLS	roots	hydroxyalkenyl_tot_ali	mQA03_14.6	A03	5	Bn.A03.p13943992	14 576 370	14 531 159	14 619 769	8.65E-08	-2.673	0.105
GLS	roots	hydroxyalkenyl_tot_ali	mQA09_4.1	A09	4	Bn.A09.p2733282	4101935	2 757 086	4 564 748	8.23E-08	-3.902	0.099
GLS	roots	neoglucobrassicin_tot_ind	mQC02_21.6	C02	44	Bn.C2.p19939091	21680680	21 460 578	22 609 267	6.99E-07	-3.207	0.093
GLS	roots	GLS22_tot_ind	mQC02_57.2	C02	1	Bn.C2.p48462140	57507685	57 100 829	57 635 562	3.97E-06	0.022	0.046
PHL	leaves	TOT_PPN_PHL	mQA08_2.4	A08	197	Bn.A08.p2897827	2871709	1 588 102	12 927 320	1.34E-08	2.481	0.116
PHL	leaves	TOT_FLA_PHL	mQA08_2.4	A08	197	Bn.A08.p3237433	3 231 749	1 588 102	12 927 320	7.47E-09	-2.526	0.120
PHL	leaves	HQA_PPN	mQA06_6.7	A06	1	Bn.A06.p6182593	6 785 465	6 764 993	6 936 249	5.01E-06	1.278	0.073
PHL	leaves	HQA_PPN	mQC04_63.2	C04	5	Bn.C4.p51315477	63 297 495	62 322 712	63 603 792	9.21E-08	1.521	0.100
PHL	leaves	HQA_PPN	mQC07_36.3	C07	4	Bn.C7.p29949146	36 350 384	35 603 274	36 634 276	4.43E-07	1.827	0.092
PHL	leaves	HAGE_PPN	mQA06_6.7	A06	1	Bn.A06.p6182593	6 785 465	6 764 993	6 936 249	5.01E-06	-1.278	0.073
PHL	leaves	HAGE_PPN	mQC04_63.2	C04	5	Bn.C4.p51315477	63 297 495	62 322 /12	63 603 792	9.21E-08	-1.521	0.100
PHL	leaves	HAGE_PPN	mQC07_36.3	C07	4	Bn.C7.p29949146	36 350 384	35 603 274	36 634 276	4.43E-07	-1.827	0.092
PHL	leaves	PPN_monoA_PPN	mQC08_33.0	C08	5	Bn.C8.p33518339	33 006 364	32 926 325	33 033 719	9.52E-08	-2.385	0.104
PHL	leaves	PPN_diAc_PPN	mQA09_46.1	A09	2	Bn.A09.p30270747	46 131 663	45 983 582	46 140 360	4.73E-06	-1.798	0.076
PHL	leaves	PPN_GIAC_PPN	mQC01_33.0	LU8	20	Bn.C8.p33518339	33 006 364	32 863 135	33 033 /19	2.23E-08	2.521	0.114
PHL	leaves		mQC01_28.7	AUI	2	Bn.C1.p40541584	28 666 611	28 665 028	28 800 699	2.24E-12	0.881	0.162
PHL	leaves	FLA_MONOAC_FLA	mQA03_2.4	A08	322	Bn.AU8.p/260514	10 487 005	1 588 102	15 397 234	6.70E-15	3.5/8	0.205
PHL	leaves	KAEIVI_FLA	mQA10_10.7	AU2	2	BII.AUZ.p10869120	10 48/ 005	10 388 981	10 851 309	2.04E-U/	2.440	0.098
PHL	leaves		mQA10_10.0	A10	10	вп.А10.р5/2//80	903/435	8 /20 536	9 082 307	1.00E-07	-1.838	0.152
PHL	leaves	QUER_FLA	mQA10_10.0	A10	10	BII.A10.pb120400	10 025 910	δ 627 463	10 3/4 523	1.38E-11	1.95/	0.152
	loaves		mQC02 7.9	C03	ر وت	Bn C2 n7791512	9 420 600	/ 223 143	0 000 045	1.23E-U/	1.4/1	0.070
	leaves	ISOR FLA	m0C07 29 5	C05	2 A	Bn (7 n32204272	28 101 707	36 050 414	28 672 201	1 72E 00	-1.0/2	0.076
ICT IL	reaves	IJON_FLA	111000/_30.3	00/	4	DII.C/.p32304272	30 494 702	30 939 414	30 023 201	T. / 3E-00	-1.203	0.110



Supplementary Figure S1: Correlations between contents of GLSs in leaves among the *B. napus* accessions



Supplementary Figure S2: Correlations between contents of GLSs in roots among the *B. napus* accessions

Ph D. manuscript | Anani Amegan MISSINOU



Supplementary Figure S3: Correlation matrix of GLS and PHL contents in leaves among *B. napus* accessions



Supplementary Figure S4: Correlation between contents of leaf phenolics among the *B. napus* accessions

References

Abedini D, Jaupitre S, Bouwmeester H, Dong L (2021) Metabolic interactions in beneficial microbe recruitment by plants. Curr Opin Biotechnol 70: 241–247

An H, Qi X, Gaynor ML, Hao Y, Gebken SC, Mabry ME, McAlvay AC, Teakle GR, Conant GC, Barker MS, et al (2019) Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. Nat Commun. doi: 10.1038/s41467-019-10757-1

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using Ime4. J Stat Softw. doi: 10.18637/jss.v067.i01

Baumert A, Milkowski C, Schmidt J, Nimtz M, Wray V, Strack D (2005) Formation of a complex pattern of sinapate esters in *Brassica napus* seeds, catalyzed by enzymes of a serine carboxypeptidase-like acyltransferase family? Phytochemistry 66: 1334–1345

Bhattacharya A, Sood P, Citovsky V (2010) The roles of plant phenolics in defence and communication during *Agrobacterium* and *Rhizobium* infection. Mol Plant Pathol 11: 705–719

Bus A, Körber N, Snowdon RJ, Stich B (2011) Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theor Appl Genet 123: 1413–1423

Chalhoub B et al (2014) Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science (80-) 345: 950–953

Francisco M, Joseph B, Caligagan H, Li B, Corwin JA, Lin C, Kerwin RE, Burow M, Kliebenstein DJ (2016) Genome wide association mapping in *Arabidopsis thaliana* identifies novel genes involved in linking allyl glucosinolate to altered biomass and defense. Front Plant Sci 7: 1–13

Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol 32: 361–369

Giamoustaris A, Mithen R (1995) The effect of modifying the glucosinolate content of leaves of oilseed rape (*Brassica napus* ssp. oleifera) on its interaction with specialist and generalist pests. 347–363

Hu D, Jing J, Snowdon RJ, Mason AS, Shen J, Meng J, Zou J (2021) Exploring the gene pool of *Brassica napus* by genomics-based approaches. Plant Biotechnol J 19: 1693–1712

Kittipol V, He Z, Wang L, Doheny-Adams T, Langer S, Bancroft I (2019) Genetic architecture of glucosinolate variation in *Brassica napus*. J Plant Physiol 240: 152988

Kondra ZP, Stefansson BR (1970) Inheritance of the major glucosinolates of rapeseed (*Brassica napus*) Meal. J Geol 97: 129–147

Lê S, Josse J, Husson F (2008) FactoMineR: An R package for multivariate analysis. J Stat Softw 25: 1–18

Li F, Chen B, Xu K, Wu J, Song W, Bancroft I, Harper AL, Trick M, Liu S, Gao G, et al (2014) Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). DNA Res 21: 355–367

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D (2011) FaST linear mixed models for genome-wide association studies. Nat Methods 8: 833–835

Liu S, Huang H, Yi X, Zhang Y, Yang Q, Zhang C, Fan C, Zhou Y (2020) Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome-wide association study. Plant Biotechnol J 18: 1472–1484

Liu T, Zhang Y, Agerbirk N, Wang H, Wei X, Song J, He H, Zhao X, Zhang X, Li X (2019) A highdensity genetic map and QTL mapping of leaf traits and glucosinolates in *Barbarea vulgaris*. BMC Genomics 20: 371 Mandal SM, Chakraborty D, Dey S, Mandal SM, Chakraborty D, Dey S (2017) Phenolic acids act as signaling molecules in plant- microbe symbioses Phenolic acids act as signaling molecules in plant-microbe symbioses. Plant Signal Behav 2324: 359–368

Milkowski C, Baumert A, Schmidt D, Nehlin L, Strack D (2004) Molecular regulation of sinapate ester metabolism in *Brassica napus*: Expression of genes, properties of the encoded proteins and correlation of enzyme activities with metabolite accumulation. Plant J 38: 80–92

Missinou AA, Ferreira de Carvalho J, Marnet N, Delhaye T, Hamzaoui O, Abdel Sayed D, Guitton Y, Lebreton L, Langrume C, Laperche A, et al (2022) Identification and Quantification of Glucosinolates and Phenolics in a Large Panel of *Brassica napus* Highlight Valuable Genetic Resources for Chemical Ecology and Breeding . J Agric Food Chem. doi: 10.1021/acs.jafc.1c08118

Mithen R (1992) Leaf glucosinolate profiles and their relationship to pest and disease resistance in oilseed rape. Euphytica 63: 71–83

Negro SS, Millet EJ, Madur D, Bauland C, Combes V, Welcker C, Tardieu F, Charcosset A, Nicolas SD (2019) Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. BMC Plant Biol 19: 1–22

Obermeier C, Hossain MA, Snowdon R, Knüfer J, von Tiedemann A, Friedt W (2013) Genetic analysis of phenylpropanoid metabolites associated with resistance against *Verticillium longisporum* in *Brassica napus*. Mol Breed 31: 347–361

Rosa EAS, Heaney RK, Fenwick GR, Portas CAM (2010) Glucosinolates in Crop Plants. Hortic Rev (Am Soc Hortic Sci). doi: 10.1002/9780470650622.ch3

Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE, Malvar RA, Moreno-Gonzalez J, Nicolas S, Madur D, Combes V, et al (2014) Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genetics 197: 375–387

Rousseau-Gueutin M, Belser C, Silva C Da, Richard G, Istace B, Cruaud C, Falentin C, Boideau F, Boutte J, Delourme R, et al (2020) Long-read assembly of the *Brassica napus* reference genome 'Darmor-*bzh*'. Gigascience 9: 1–16

Spinelli F, Cellini A, Marchetti L, Mudigere K, Piovene C (2011) Emission and Function of Volatile Organic Compounds in Response to Abiotic Stress. Abiotic Stress Plants - Mech Adapt. doi: 10.5772/24155

Van Raden PM (2008) Efficient Methods to Compute Genomic Predictions. J Dairy Sci 91: 4414–4423

Venables WN, Ripley BD (2002) Modern Applied Statistics with S, Fourth edition. doi: 10.1007/978-1-4757-2719-7_14

Wang B, Wu Z, Li Z, Zhang Q, Hu J, Xiao Y, Cai D, Wu J, King GJ, Li H, et al (2018) Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. Plant Biotechnol J 16: 1336–1348

Wei T SV (2021). R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.90). https://github.com/taiyun/corrplot

Wu J, Zhao Q, Yang Q, Liu H, Li Q, Yi X, Cheng Y, Guo L, Fan C, Zhou Y (2016) Comparative transcriptomic analysis uncovers the complex genetic network for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. Sci Rep 6: 1–16

Yang J, Guo H, Jiang NJ, Tang R, Li GC, Huang LQ, Van Loon JJA, Wang CZ (2021) Identification of a gustatory receptor tuned to sinigrin in the cabbage butterfly *Pieris rapae*. PLoS Genet 17: 1–28

Zhang Y, Huai D, Yang Q, Cheng Y, Ma M, Kliebenstein DJ, Zhou Y (2015) Overexpression of three glucosinolate biosynthesis genes in *Brassica napus* identifies enhanced resistance to *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS One 10: 1–17

Paper III: Genetical phytochemistry of *Brassica napus* roots revealed QTLs involved in the control of two pteridine-like aminoacid conjugates

Genetical phytochemistry of *Brassica napus* roots revealed QTLs involved in the control of two pteridine-like aminoacid conjugates

Anani Amegan Missinou ^{#,a,}, Clément Orione ^c, Julie Ferreira de Carvalho ^a, Youcef Haddad ^a, Nathalie Marnet ^b, Philippe Jéhan ^c, Anne Laperche ^a, Régine Delourme ^a, Alain Bouchereau ^a, Maria J. Manzanares-Dauleux ^a, Antoine Gravot ^{*,a}

^a Univ Rennes, Institut Agro, INRAE, IGEPP, 35653, Le Rheu, France

^b INRAE, BIA, 35653, Le Rheu, France

^c Univ Rennes, CRMPO, 35700 Rennes, France

First author and * corresponding author, Antoine.Gravot@univ-rennes1.fr

Abstract

Infra-specific diversity of plant-specialized metabolic compounds is assumed to play major ecological roles but much remains to be more deeply explored in the root compartment as an interface for biotic and abiotic environments. In this work, we present a detailed quantitative profiling of 18 previously unreported root-specific UV-absorbing compounds among 281 accessions of *Brassica napus*. The resulting dataset allowed identifying three root chemotypes, and GWAS analysis allowed the identification of QTL involved in the control of 11 of those compounds. After purification, the structural elucidation of three of those compounds was obtained from spectroscopic data, including 1D, and 2D nuclear magnetic resonance, and high-resolution mass spectrometry. This analysis revealed one sulfur-containing compound derived from ferulic acid (m/z 493.115 [M-H]⁻). It also revealed two pteridine-like polynitrogenated aromatic heterocycles conjugated to leucine (m/z 320.099 [M-H]⁻)

Ph D. manuscript | Anani Amegan MISSINOU

and homophenylalanine (368.099 [M-H]⁻) which were found at higher concentrations in the roots of spring oilseed rape varieties. One QTL on chromosome A04 was found to control these two chemically-related molecules, suggesting that this QTL might be involved in the biosynthesis of the core pteridine-like moiety. Each of the two molecules was also controlled by additional distinct QTLs, which might be each involved in the conjugation with each amino acid. The genomic regions involved in the control of every 11 compounds encompassed a series of genes coding for enzyme families associated with amino acid, carbohydrate, methionine, and phenylpropanoid metabolism. Beyond the canonic roles played by glucosinolates and phenolics in *Brassicaceae* species, this exploratory analysis highlights the importance of additional categories of specialized metabolites, which can also contribute to the phytochemical diversity in *B. napus*.

Keywords: oilseed rape, specialized metabolism, mGWAS, mass spectrometry, NMR, mQTL.

1. Introduction

A key challenge in plant chemical ecology is to extend the current knowledge of specialized metabolite (SM) diversity and identify genomic factors controlling their biosynthesis. This should increase our understanding of the roles played by these SMs in plant growth and development (Nakabayashi and Saito, 2015; Chattha et al., 2016), biotic interactions (Gresshoff, 2003; Shroff et al., 2008; Erktan et al., 2018), and abiotic challenges (Stringlis et al., 2018; Huang et al., 2019). Besides typical chemical approaches for isolation and identification, the research on SM genomics in plants has benefited from collaborations between phytochemists, geneticists, physiologists, and

molecular biologists. For instance, in *Brassicaceae*, integrative studies were able to highlight the main genes controlling glucosinolate (Kroymann et al., 2003; Sun-Ju Kim, et al., 2008; Kittipol et al., 2019; Liu et al., 2020), phenylpropanoids (Francisco et al., 2016), and flavonoids (Lee et al., 2014) contents. Such studies, aiming to identify associations between genetic variants and chemical traits using large genetic diversity are so-called metabolic genome-wide association studies (mGWAS).

Brassica napus is an allopolyploid species derived from the post-neolithic interspecific hybridization between B. rapa and B. oleracea and includes oilseed rape, swede, and fodder crops (Nagaharu U, 1935; Chalhoub et al. 2014). This species is economically important for seed oil and protein meal production across different continents, including temperate and subtropical regions in Europe, Canada, South Asia, China, and Australia (FAOSTAT, 2022). Despite its significance, there are still many uncharacterized SM possibly involved in plant production performance and product quality. Although roots play a determinant role in the growth and development of plants, little attention has been paid to root phytochemicals, especially those with aromatic structures. Based on the benzene-type nucleus, aromatic compounds can be divided into three groups: (1) alkyl, aryl-, and alicyclic- nitrogen-sulfur- substituted benzene derivatives, (2) di- and polyphenyl, and (3) polycyclic compounds formed from the fusion of two or more benzene ring systems (Baxter and Warshawsky, 2012). These compounds contain extended-electron π-systems and display almost as much light absorption in the UVA (315-400 nm) as in the UVB (280-315 nm) ranges and null absorption of visible radiation (400 to 800 nm). Phenolic and aromatic compounds are poorly described in the roots of Brassicaceae. For instance, coumarins are aromatics lactone-derived from the multi-branched phenylpropanoid pathway. The exudation of
coumarins was shown to act as an allelopathic plant growth regulator (Chattha et al., 2016) or as an iron uptake stimulator from substrates with low iron availability (Schmid et al., 2014). Coumarins may shape root microbiota assembly to promote plant health (Stringlis et al., 2018), or be involved in biotic and abiotic stress responses (Nakabayashi and Saito, 2015).

Brassica species typically accumulate very large amounts of flavonols and phenylpropanoids in leaves, but the few available studies on the root compartment reported only very low concentrations of phenolics (Fernandes et al., 2007). In previous work, we documented large-scale metabolic profiling of glucosinolates and phenolic compounds in the roots and leaves of a genetic panel of 304 accessions of *B. napus*, B. rapa, and B oleracea (Missinou et al., 2022). This work confirmed that flavonol and phenylpropanoid derivatives are accumulated at very low concentrations (under quantification ranges) in Brassica roots. Also, unpublished chromatographic results obtained during these investigations highlighted a series of atypical UV-absorbing root compounds. For most of those compounds, mass spectra did not match with any available phenolics from public databases. Two of those unknown compounds (m/z 320.099 and m/z 368.099) have been previously reported in our studies (Wagner et al. 2019) to be genetically related to clubroot resistance in *B. napus*. Thus, because of the potential roles that this series of unknown compounds may potentially play in biotic interactions, we focused our efforts in the present work on the investigation of the chemical and genetic characteristics of this series of unknown aromatic structures.

We first investigated the concentrations of these compounds in the roots of the abovementioned genetic panel of 304 *Brassica* accessions, using a UV-detection semiquantification approach. We then used a mGWAS approach to investigate the genomic regions involved in the control of their concentrations in roots. Three of these compounds were further purified and their chemical structures were investigated by a combination of Q-Exactive Orbitrap High-Resolution Mass Spectrometry (HRMS) and NMR (1D and 2D).

2. Materials and Methods

2.1. Plant material and sampling

Seeds from 304 *Brassica* accessions (281 *B. napus*, 12 *B. rapa*, and 11 *B. oleracea*) were germinated, cultivated, sampled, and stocked as described in Missinou et al. 2022. The subset of spp. oleifera from *B. napus* was constituted by 130 accessions of winter oilseed rape (WOSR) and 120 spring oilseed rape (SOSR). Eight plants from each accession were cultivated in two spatially randomized blocks of 4 plants according to cultivation conditions described in Missinou et al. 2022. Three independent biological replicates were performed at 6-week intervals. Four-week-old plants per accession were harvested, and roots were rapidly washed with tap water, immediately frozen in liquid nitrogen, then stored at -80°C before lyophilization. The lyophilized materials were powdered, homogenized, and stored in the dark until further phytochemical analysis.

2.2. Chemicals

HPLC-grade acetonitrile, methanol, and formic acid were from Carlo Erba. Authentic phenolic standards were provided by Sigma-Aldrich. For NMR experiments, deuterated solvents (MeOD-d₃, MeOD-d₄, and CD₃OD) were obtained from Eurisotop (France).

2.3. Metabolite extraction procedure

Thirty milligrams (mg) aliquots of freeze-dried powders were extracted with 1mL of methanol/formic acid (99/1, v/v). The suspensions were vortexed and then sonicated in a water bath at 25°C for 10 min. After centrifugation at 7000 rpm for 10 min, the supernatants were collected and filtered using a 0.22 µm PTFE membrane. Extracts were stored at 4°C and analyzed by LC-MS less than 24h after extraction. This extraction procedure was conducted on each sample for quantitative analyses. In addition, for "bulk" analysis, similar extractions were performed on pools from several root powders sampled from different accessions (five randomly chosen accessions for each of the following groups: *B. napus* WOSR, SOSR, fodder, swede, *B. oleracea*, and *B. rapa*). After that, these powder bulks were used for HPLC-Orbitrap LTQ-XL high-resolution mass spectrometry analysis. As some compounds were accumulated more specifically in specific genotypes, additional powder bulks were also prepared by combining typical accessions to gain more intense signals and allow better chemical fragmentation.

2.4. Chromatography, semi-quantitative profiling

Sample extracts were processed through UPLC-UV-TQD as previously described (Missinou et al. 2022). UV-absorbing compounds content was simultaneously measured at λ 330mn, and peak deconvolution of UV-chromatograms was preprocessed using Empower (version 2). Alphanumeric IDs were assigned to the UVabsorbing compounds as numbered peaks noted from R01 to R18 (according to the chromatographic elution order). These numbers were followed by an underscore and by the monoisotopic m/z, except for three peaks R12, R14, and R17 harboring multiple m/z possibly corresponding to multiple coeluted compounds. Those three compounds were then annotated as MNA for "Mass Not Available". The measure was based on determining arbitrary area units under the absorbance curve at λ 330 nm using the baseline method and fitting with 1/x weighting.

2.5. MS/MS fragmentation

The experiments of MS2 fragmentation from "bulk" extracts were executed on LTQ-Orbitrap- LTQ-XL mass spectrometer as previously described (Missinou et al. 2022). Succinctly, mass spectrometry experiments were performed using the LTQ-Orbitrap XL (Thermo Fisher Scientific) equipped with an ESI source operating in negative mode. The MS1 spectra were acquired as a profile in full scan mode and used for recording accurate mass ranging from 250-1500 m/z, with 500 ms of scan time (with a resolution of 30.000 at m/z 200). The MS/MS scanning mode was set as a data-dependent ms2 scan (dd-ms2). The high energy collision dissociation was 30eV, and the cone voltage at 35V. Several diagnostic neutral losses were searched to get initial structural information and define the family of each compound.

2.6. Multivariate statistical analysis

Analyses were performed using the software R and the interface RStudio (R Core Team, 2021). Variables with undetermined values representing more than 5% of the dataset were removed from the following analyses. Normalized datasets were mean-centered using the least-squares method and a linear model including biological replicates using the "emmeans" R package (Lenth et al., 2021). Mean-centered values

of each compound were visualized on boxplots, and comparisons between species were evaluated using Tukey's tests. Multivariate statistical analyses were performed to describe phytochemical profile diversity. PCA models were made with "gridextra" and "ade4" R packages (Dray and Dufour, 2007). Heatmaps were performed with "complexheatmaps" R package, converting the data to z-scores (centered samples divided by standard deviation) (Gu et al., 2016). PLS-DA models were performed using the mixOmics R package (Rohart et al., 2017). All visualization plots were realized using the "ggplot2" R package (Wickham, 2016).

2.7. Metabolic genome-wide association analysis

Adjusted quantitative values using the least-squares method were further processed through GWA analyses for the *B. napus* diversity panel. These genetic analyses were conducted using the FastLMM algorithm (Lippert et al., 2011) with a kinship calculated using the Van Raden algorithm (VanRaden, 2008) for each chromosome as described by Rincent and coworkers (2014). In addition to the kinship, the three main axes of the PCoA explaining respectively 13.0, 9.1, and 6.0% of the variance were also added in the FastLMM analysis added as covariables. To run these different steps the R script developed by Negro and coauthors (2019) was adapted. The significance threshold was set based on the Bonferroni threshold of 5% calculated on a corrected SNP population as proposed by the simple M method developed by (Gao et al., 2008). This method is based on the composite linkage disequilibrium (CLD) correlation between SNPs. The CLD was used to calculate the effective number of independent tests (M eff). Here, Meff was calculated and fixed to 5,544 SNPs. The threshold -log10(p) was thus defined at 5.296 for the *B. napus* panel. Then, significative SNPs were fused to

identify mQTLs as long as their confidence intervals (using LD decay, r²=0.25) overlapped.

2.8. Identification of candidate genes

First, the location (start and end) of genes in the *B. napus* genome v05 (Chalhoub et al., 2014) was searched into the genetic intervals of the mQTLs using a homemade R script. Then, the functional annotations of oilseed rape genes inferred from Arabidopsis were retrieved from *Brassica* genome v05 (Chalhoub et al., 2014) to identify the candidate genes for the mQTLs detected. Attention has been given to genes encoding for transcription factors, major enzyme families known to be involved in specialized metabolism, and their presumed cousins in primary metabolism, considering chemical functions and structural information of purified compounds.

2.9. Metabolite purification

Reversed-phase SPE fractionation procedure

The purification of specific compounds implied the production of large amounts of root material with a selection of the genotypes harboring the highest concentrations (details in the support table S1). Plants were cultivated as described above. For each selected genotype qualified as super-accumulators, 60 g freeze-dried roots were extracted with 400 mL of methanol/formic acid (99/1, v /v). These extracts were concentrated to dryness using a rotary evaporator (40°C, 150 rpm), then re-dissolved in 5 mL of a methanol/water (50/50, v/v) mixture, and sonicated for 2 min. This extract was loaded on a series of vacuum solid-phase extraction systems (SPE columns, Chromabond, C18, 45 µm, 70 mL/10000 mg, 60 Å, Macherey-Nagel, Germany) (Krishnan and

Ibraham, 1994; Zhuang et al. 1997) to pre-fractionate root methanolic extracts removing the contaminant and interfering substances. The SPE columns were preconditioned with two column volumes (CV) of methanol and equilibrated with two CVs of water. Subsequently, 1 mL of the concentrated aqueous phase was loaded onto every single column, and rapidly washed polar compounds were removed, including GLS. Cartridges were eluted with 2 CVs of pure methanol, which disrupted the interaction (non-polar and hydrophobic) of the remaining analytes adsorbed by the stationary phase. The eluted fraction was then concentrated by a rotary evaporator (40°C, 150 rpm) to dryness, re-dissolved in 5 mL of methanol/water (50/50, v/v) mixture, sonicated for 2 min, and centrifuged at 7000 rpm for 10 min before loading on the semi-preparative HPLC system.

Semi-preparative HPLC

HPLC chromatographic runs were performed using a Dionex Ultimate 3000 HPLC system, coupled to a UV/vis absorbance detector set at λ 330 nm and piloted by the Chromeleon software and using a Macherey-Nagel semi-preparative HPLC column (Nucleodur, C18, 250 mm × 10 mm internal diameter, five µm, 110 Å) maintained at 36 °C. The mobile phase consisted of a mixture of solvent A (H₂O with 0.1% formic acid, v/v) and solvent B (acetonitrile with 0.1% formic acid, v/v) at a flow rate of 3.5 mL/min. The gradient elution schedule was as follows: initially 2% B, 2-7% B from 0-3 min, 7-17% B from 3-12 min,17-20% B from 12-13 min, 20-32% B from 13-24 min, 32% B maintained from 24-34 min, 32-90% B between 34-36 min, held at 90% B between 36-39 min, then 90-2% B from 39-40 min, and the remaining 11 minutes of the gradient allowed for re-equilibrating the column for the next run. 250 µL were injected into the

column for each run. Fractions of 1.75 mL were collected each 30 sec. Fractions corresponding to UV- absorbing peaks were thereafter analyzed by UPLC-UV-TQD-MS to get additional spectrometric information on the compounds of interest. Fractions obtained from iterated (about 20) injections were pooled, then concentrated by a rotary evaporator to dryness, and were re-dissolved in 5 mL 50% MeOH in water and centrifuged at 7000 rpm for 10 min. Aliquots of 250 µL were then re-loaded on the same HPLC system, using another gradient. This elution gradient was adapted around the eluting % solvent B for each targeted compound and optionally maintained under isocratic conditions. The products collected were concentrated to dryness under nitrogen flux.

2.10. Structural analysis of purified molecules

Q-Exactive Orbitrap MS/MS experiments

The three purified compounds were investigated by HRMS2 and HRMS3 by infusion on a Q-Exactive Orbitrap. For each compound, some basic rules of mass spectrometry (C, H, N, O, S) were applied to identify fragmentation patterns (neutral losses, degree of unsaturation, and formula of small neutral fragments) and the empirical molecular formula could be designed (Bright and Chen, 1983; Pellegrin, 1983). These constraints allowed us to remove more than 95% of false candidate formulas for compounds.

NMR experiments

NMR spectra were recorded in MeOD-d3, MeOD-d4, and DMSO-d6 at 300 K on a Bruker Avance III 500 spectrometer equipped with a BBFO+ probe and a Bruker Avance NEO 500 MHz spectrometer with CryoProbe, both driven by the TopSpin NMR software (v3.6.1). The solvent was used as an internal standard to determine chemical shifts. Spectra were recorded employing standard Bruker pulse sequences except for the band-selective experiments. The 1D 1H, 13C-zgpg, and 13C-JMOD and the 2D 1H-1H COSY (correlated spectroscopy), 1H-13C HSQC (heteronuclear singlequantum correlation), 1H-13C HMBC (heteronuclear multiple bond correlation) experiments were performed, respectively with cosygpqf, hsacetapsisp2.2, hmbcctetgpl2nd. During the acquisition, all the 1H-13C 2D experiments were recorded with the NUS algorithm, with the NUS amount of 25% for the HSQC and 50% for the HMBC. Band-selective experiments were also performed, with shsqcetqpsisp2.2 & shmbcctetgpl2nd, on 50 ppm wide windows in f1 to increase the precision of the assignations for all the close signals. To separate the pure spectra of compounds within a mixture and check simple purity, diffusion-ordered spectroscopy (DOSY) with stebpgp1s sequence was performed. The DOSY spectra were processed with the bayesian algorithm with 32 points in diffusion dimension. NMR data processing and plotting were performed using MestreNova software (v14.2. 3) (Willcott, 2009). Combining NMR and MS allowed a simultaneous assignment and structural elucidation.

3. Results

3.1. LC-MS/MS fragmentation of eighteen unidentified UV-absorbing compounds extracted from *Brassica roots*

The complete analysis of the 304 accessions revealed a newly described series of UVabsorbing compounds (at λ 330 nm). Eighteen peaks that were well resolved among a series of chromatographic injections were selected for further analyses (Figure 1).



Figure 1: UPLC chromatogram showing UV absorbance at λ330nm of root extract (QC) from *Brassica.* R01-R18 represented root-specific UV-absorbing compounds (with their mass-to-charge ratio in negative mode) observed in pooled *Brassica napus*.

In Table 1 are compiled their retention times (after either HPLC or UHPLC separation, see material and methods), accurate parent masses (<10 ppm), and daughter anions. The fragmentation patterns (daughter anions and neutral losses) and parent anions of these compounds did not share any feature with any of the phenolic compounds that had been identified in the leaf samples in Missinou et al. (2022) or with any known Phe/Tyr-derived phenylalkyl-GLSs, or Trp-derived indolic-GLSs (which also display UV-absorbance at λ 330 nm), except for **R04_385** and **R09_385**. These two isobaric anions at m/z 385.113 (HPLC-RT 12.03 min and 16.84 min, respectively) displayed daughter anions at m/z 339 [M-H]-, 317 [M-H]-, and 249 [M-H]-, common neutral loss of 46 Da (C₂H₆O, ethanol) and C₅H₈ (68 Da, possibly pentadiene), and loss of 136 Da

that also suggest the structural proximity between R04 and R09 (Table 1). Although the parent masses of those two compounds may suggest sinapoyl-glucose, both of those two compounds harbored different chromatographic elution times compared to the compound PL04 which was annotated as sinapoyl-glucose in Missinou et al. 2022. This could suggest the accumulation of different stereoisomers in roots and leaves for this family of compounds.

We searched the mass features in the databases (COCONUT, KNAPSAcK, KEGG, BioCyc, Natural Products, PubChem, PlantCyc) and found that accurate masses of UR01_M315 and UR08_681 agreed with dihydroxybenzoylhexose and dihexosyl derivative of pinoresinol. For every other peak, to the best of our efforts, none of the other mass features matched with any compound from public databases. After fragmentation, neutral losses of 18 Da, 46 Da, 68 Da, and 136 Da, shared by R02_M431, R04_M385, R07_M551, R10_M565, R11_M523, R15_M380, and R16_M493 features, suggested that they may have quite close structures (Table 1). Furthermore, three pairs of features (R03_M434/R05_M434; R04_M385/R09_M385; R13 M320/ R18 M368) seem very close from a chemical point of view based on their fragmentation patterns. Features R03_M434 and R05_M434 yielded the deprotonated quasi-molecular ions at m/z 434.207 [M-H]⁻ (HPLC-RT 11.43 min and 12.41 min, respectively). They both generated the daughter anions at 391 [M-H], 390 [M-H], 388 [M-H]⁻, 366 [M-H]⁻ illustrating the putative losses of C₂H₅N (43 Da), CO₂ (44 Da), C₂H₆O (46 Da), respectively (Table 1). These fragmentation patterns suggest the structural isomerization of a single structure.

Features **R13_M320** and **R18_M368** were relevant to compounds m/z 320.099 [M-H]⁻ (at HPLC-RT 13.48 min) and m/z 368.099 [M-H]⁻ (at HPLC-RT 17.45 min), that was previously identified in roots of *B. napus* as genetically related to a QTL conferring resistance to the root pathogen *P. brassicae* (Wagner et al., 2019). These compounds exhibited common fragments at m/z 233 [M-H]⁻ and m/z 163 [M-H]⁻ and putative similar neutral losses of 44 Da (CO₂, carboxylic acid group) generating two main fragments at m/z 276 [M-H]⁻ and m/z 324 [M-H]⁻, from R13 and R18 respectively. Ions 233 [M-H]⁻ and 281 [M-H]⁻ putatively resulted from the secondary fragmentation of in-source generated m/z fragments 276 [M-H]⁻ and m/z 324 [M-H]⁻ and m/z 324 [M-H]⁻, respectively. They both suggested the loss of the 43 Da typical of amide function HOCN (Table 1). The remaining features that displayed specific fragments suggest a structural singularity (Table 1).

3.2. Targeted metabolite profiling of UV-absorbing at λ 330 nm in *Brassica* roots highlights wide intraspecific quantitative variations

Based on the integration of $UV_{\lambda 330}$ peaks, relative quantification of the eighteen eluted compounds was acquired from the methanolic root extracts of the 304 *Brassica* accessions.

Table 1: Mass fragmentation (m/z) patterns of unknown root compounds absorbing at λ330nm in the panel of 304 accessions of *Brassica*. The IDs from R01_M315 to R18_M368 represent the metabolites numbered according to the order of elution. RT min HPLC and RT min UPLC determine retention time for qualitative and semi-quantitative analysis. [M] stands for the neutral structure relative intensity of the fragment to the base peak is shown in parentheses. "Exp" strands for experimental mass (m/z [M-H]⁻) and MS2 (daughter anions) were acquired on HPLC-Orbitrap LTQ-XL. MNA: for "Mass Not Available").

			Diagnostic ions [M-H] ⁻ MS2 (m/z) (% relative intensity)							
[M-H] ⁻ MS1 (m/z)				Neutral losses						
				M-H-H₂O	M-H-CO ₂	M-H-?	M-H-?	M-H-CO ₂ - HCN	M-H-?	Other ions
ID	RT min (HPLC)	RT min (UPLC)	Exp.	[M-H-18] ⁻	[M-H-44] ⁻	[M-H-46] ⁻	[M-H-68] ⁻	[M-H-87] ⁻	[M-H-136] ⁻	
R01_M315	8.81	4.55	315.071							
R02_M431	10.99	5.86	431.117	413(1)		385(1)	363(1)		295(1)	137(100)[M-H-294] ⁻ , 386(1)[M-H-45] ⁻
R03_M434	11.43	6.16	434.207		390(2)	388(2)	366(3)			237(100)[M-H-197] ⁻ , 175(6)[M-H-259] ⁻ , 238(5)[M-H-196] ⁻ , 391(2.6)[M-H-43] ⁻ , 305(2)[M-H-129]-
R04_M385	12.03	6.61	385.113			339(6)	317(5)		249(3)	177(100)[M-H-208] ⁻ , 162(7)[M-H-223] ⁻ , 207(1)[M-H-178] ⁻
R05_M434	12.41	7.14	434.207	416(10)	390(10)	388(15)	366(20)			391(100)[M-H-43] ⁻ , 392(48)[M-H-42] ⁻ , 175(30)[M-H-259] ⁻ , 406(12)[M-H-28] ⁻
R06_M479	13.3	7.63	479.041							301(100)[M-H-178] ⁻ , 283(82)[M-H-196] ⁻ , 259(17)[M-H-220] ⁻ , 399(12)[M-H- 80] ⁻ , 176(12)[M-H-303] ⁻ , 174(12)[M-H-305] ⁻ , 237(9)[M-H-242] ⁻
R07_M551	15.41	9.1	551.175	533(<1)		505(<1)	483(5)		415(1)	389(100)[M-H-162] ⁻ , 193(2)[M-H-358] ⁻ , 341(1)[M-H-210] ⁻
R08_M681	15.91	10.11	681.237	NA	NA	NA	NA	NA	NA	NA
R09_M385	16.84	10.42	385.113	249(9)	341(<1)	339(1)	317(2)			267(100)[M-H-118] ⁻ , 268(<1)[M-H-117] ⁻ , 181(*)[M-H-204] ⁻
R10_M565	18.29	11.22	565.191	547(9)			497(3)		429(3)	339(100)[M-H-226] ⁻ , 327(43)[M-H-238] ⁻ , 357(38)[M-H-208] ⁻ , 518(28)[M-H- 47] ⁻ , 498(5)[M-H-67] ⁻
R11_M523	18.71	12.72	523.144	505(3)		477(15)	455(11)		387(3)	343(100)[M-H-180] ⁻ , 456(7)[M-H-67] ⁻ , 434(2)[M-H-89] ⁻ , 494(2)[M-H-29] ⁻ , 433(2)[M-H-90] ⁻
R12_MNA		12.97	NA	NA	NA	NA	NA	NA	NA	NA
R13_M320	19.54	13.48	320.098	302(2)	276(100)			233(5)		163(9)[M-H-157] ⁻ , 206(2)[M-H-114] ⁻ , 205(1)[M-H-115] ⁻ , 120(*)[M-H-200] ⁻
R14_MNA	20.46	14.09	NA	NA	NA	NA	NA	NA	NA	NA
R15_M380	21.93	15.38	380.079	362(19)	336(5)		312(12)		244(2)	176(100)[M-H-204] ⁻ , 333(4)[M-H-47] ⁻ , 243(3)[M-H-137] ⁻ , 320(3)[M-H-60] ⁻ , 311(2)[M-H-69] ⁻
R16_M493	22.49	15.71	493.115	475(5)		447(18)	425(53)		357(14)	299(100)[M-H-194] ⁻ , 211(6)[M-H-282] ⁻ , 426(5)[M-H-67] ⁻
R17_MNA	23.87	16.89	NA	NA	NA	NA	NA	NA	NA	NA
R18_M368	24.4	17.45	368.099	350(1)	324(100)		300(2)	281(4)		163(4)[M-H-205] ⁻ , 233(2)[M-H-135] ⁻ , 325(2)[M-H-43] ⁻

In the absence of any authentic standard, no calibration could be applied. Thus, relative root contents were expressed in an arbitrary unit of UV absorbance at λ 330nm per dry weight (AU_{λ 330}.g⁻¹ DW). As the molecular extinction coefficients can vary greatly according to the nature of the structures concerned, the semi-quantitative comparisons proposed below only consider the variations observed between accessions for the same compound and not the variations compared between compounds. The supporting Table S1 provides an overview of metabolite contents detected in the whole *Brassica* panel. Sheet 1b, 1c, and 1d summarizes the means, percent, and 5th to 95th percentile values, including minimum, and maximum values of contents of each compound in all accessions (Supporting Table S1, coming soon in the published article).

The #087NSL accession of *B. napus ssp. oleifera* and #282R50 accession of *B. rapa ssp. trilocularis* showed high levels (1594 and 2090 AU_{λ 330}.g⁻¹ DW, respectively) of **R03_M434**. The accessions #027CET of *B. napus ssp oleifera* and #302BAL of *B. oleracea ssp. capitata* exhibited huge amounts of **R07_M551**, which reached up to 1303 and 1562 AU_{λ 330}.g⁻¹ DW on average, respectively. For **R14_MNA**, the top mean contents were pointed out in accessions #286TER of *B. rapa ssp. rapa* (5840 AU_{λ 330}.g⁻¹ DW), #074MAR of *B. napus ssp. oleifera* (892 AU_{λ 330}.g⁻¹ DW), and #300VER of *B. oleracera ssp. capitata* (687 AU_{λ 330}.g⁻¹ DW) (Table 2, supporting Tables S1b and S1c, coming soon in the published article).

Quantitative differences were also observed at the level of subgroups (WOSR and SOSR) and subspecies (Table 2, supporting Tables S1b and S1c, coming soon in the published article). Strikingly, seven compounds (**R02_M431**, **R03_M434**, **R06_M479**, **R09_M385**, **R12_MNA**, **R14_MNA**, and **R16_M493**) showed estimated contents that vary from tens to 100-fold in whole accessions of *B. napus*, *B. rapa* and *B. oleracea*.

Table 2: Descriptive statistics (mean, percent, minimum, maximum) related to the contents of 18 $UV_{\lambda_{330}}$ -absorbing unidentified compounds in *Brassica* roots. Details for each accession and compound are given in Supporting Table S1, coming soon in the published article). AU: absorbance unit, and DW: dry weight.

Unknown root compounds absorbing at λ330nm (AU.g-1 DW)									
		B. na	pus	B. ra	apa	B. oleracea			
[M-H]-	MS1 (m/z)								
ID	Exp.	mean (%)	min - max	mean (%)	min - max	mean (%)	min - max		
R01_M315	315.071	124.9 (2.7)	0 - 268	95.1 (1.8)	0 - 264	122.6 (2.9)	67 - 173		
R02_M431	431.117	366.2 (7.9)	80 - 1396	1096.8 (20.4)	126 - 3857	171.6 (4.1)	97 - 397		
R03_M434	434.207	623.4 (13.5)	136 - 1594	808.5 (15)	152 - 2090	155.7 (3.7)	72 - 286		
R04_M385	385.113	258.6 (5.6)	0 - 976	159.4 (3)	0 - 341	60.7 (1.4)	0 - 136		
R05_M434	434.207	155.1 (3.4)	0 - 668	54.3 (1)	0 - 85	349.2 (8.3)	113 - 525		
R06_M479	479.041	303 (6.5)	120 - 721	473.2 (8.8)	83 - 727	209.4 (5)	144 - 410		
R07_M551	551.175	543.2 (11.7)	83 - 1303	162.6 (3)	0 - 546	923.8 (21.9)	679 - 1562		
R08_M681	681.237	164.1 (3.5)	0 - 489	106.8 (2)	0 - 191	105.2 (2.5)	0 - 191		
R09_M385	385.113	185.9 (4)	74 - 419	204.2 (3.8)	110 - 313	225.3 (5.3)	62 - 539		
R10_M565	565.191	105.7 (2.3)	0 - 372	138.9 (2.6)	88 - 256	91.2 (2.2)	0-214		
R11_M523	523.144	100.8 (2.2)	0 - 319	135.6 (2.5)	24 - 304	57.2 (1.4)	0 - 127		
R12_MNA	NA	364 (7.9)	181 - 855	337.6 (6.3)	208 - 574	380.4 (9)	279 - 491		
R13_M320	320.098	222.9 (4.8)	0 - 652	215.8 (4)	65 - 469	223.2 (5.3)	85 - 439		
R14_MNA	NA	525 (11.3)	141 - 892	864.7 (16.1)	275 - 5840	532.7 (12.6)	395 - 687		
R15_M380	380.079	83.1 (1.8)	0 - 232	74.2 (1.4)	0 - 157	98 (2.3)	76 - 135		
R16_M493	493.115	173.9 (3.8)	59 - 630	161.4 (3)	93 - 310	233.9 (5.5)	97 - 460		
R17_MNA	NA	118.5 (2.6)	0 - 323	116.1 (2.2)	75 - 165	85.1 (2)	56 - 112		
R18_M368	368.099	208 (4.5)	0 - 1192	169.9 (3.2)	70 - 304	200.8 (4.8)	0 - 673		

3.3. Chemotypes pointed out through multivariate analyses

Figure 2 presents the heatmap built with the $UV_{\lambda 330}$ -absorbance values of the 18 root compounds highlighted in the *Brassica* panel. In Supporting Table S1c (coming soon in the published article), the distribution of the mean absorbance values is displayed for each compound, and 5% lowest or highest mean values are shown. Both cultivar-

specific patterns, as well as cultivar-common patterns ("hot points" and "cold points" represented by regions of high-dark red- or low-dark blue- respectively), were observed (Figure 2A). Moreover, the heatmap highlighted three chemotypes and revealed moderate intra-specific variability in *Brassica* (Figure 2A).



Figure 2: Heatmap showing absorbance at λ 330 nm based-quantitative differences among 304 accessions of *Brassica* for eighteen unidentified root compounds. (A) Each row and column represent one accession and one new compound detected in the study. Each point in the heatmap stands for the averaged data from three biological replicates (n = 3 of each accession). On the left, the long and short branches of the tree indicate complete similarity and maximum dissimilarity. The scale bar represents the normalized mean absorbance of features (from R01_M315 to R18_M368). Hierarchical clustering was performed using the z-score values of each accession. The color gradient highlights the relative compound contents across accessions. The Z-score transformation ranged from red for positive Z-score values to blue for negative values. (B) Annotations of accession groups per species and cultivars. (C) Boxplots display the variation level of absorbance (AU_{λ 330}.g⁻¹ DW) of each

Ph D. manuscript | Anani Amegan MISSINOU

unknown compound across accessions. Outlier accessions were removed only for plot improvement. Horizontal lines within the boxplots indicate the median values of the compound content. Each box represents the interquartile range (5-95th percentile). Raw and detailed data of the absorbance of unknown root compounds absorbing at λ 330nm for the heatmap are listed in supplemental Table S1 (coming soon in the published article).

Briefly, chemotype 1 (cluster 1 on top) mostly contained spring cultivars of *B. napus* (SOSR) and one accession of *B. rapa* (see the caption of species and groups in Figure 2B). This cluster was characterized by the relative abundance of R01_M315, R04 M385, R08 M681, R11 M523, R10 M565, R13 M320, R14 MNA, R15 M380, **R17_MNA**, and **R18_M368**. Chemotype 2 (cluster 2 in the middle) comprised *B. napus*, B. oleracea, and B. rapa, accessions characterized mainly by "cold spots" for all compounds except R16_M493, R09_M385, R05_M434, R07_M551, R04_M385, R10_M565, and R11_M523. Chemotype 3 was mainly described by the relative of R02_M431, R03_M434, R06_M479, R12_MNA, R08_M681, abundance **R10 M565**, and **R11 M523**. This cluster mostly comprised winter cultivars accessions (see species and groups highlighted in Figure 2. B). Likewise, hierarchical clustering of compounds in all accessions revealed that pairs of putative isomers (R03_M434/R05_M434 and R04_M385/R09_M385) were clustered separately (see "cluster b" and "cluster c" in Figure 2C). At the same time, R13_M320 and R18_M368 were strongly correlated and grouped in the same cluster (see cluster a in Figure 2C). Unsupervised and supervised analyses were performed with the mean content values measured on 250 accessions belonging to four cultigroups of *B. napus* (spring and winter). As a first step, principal component analysis (PCA) provided an overview of

variable levels (loadings) and accessions (scores) distribution but did not identify variables driving clear separation among cultivars (Figure 3A).



Figure 3: Multivariate analysis of unidentified root compounds in 250 *B. napus* accessions. (A) Biplot of the unsupervised clustering using PCA model based on $UV_{\lambda 330}$ -absorbing root compounds in *Brassica* accessions. Each point represents the mean of the three biological replicates. The quality of compounds represented was estimated by cos2 values of variables and illustrated by the increased size of the circle. Compounds strongly contributing to the projection of PC1, PC2, and PC3 are highlighted

Ph D. manuscript | Anani Amegan MISSINOU

by the orange gradient. The explained variances are shown in brackets on each axis. **(B)** Biplot of the supervised clustering using the PLS-DA model based on unknown root compounds of *B. napus varieties* which compared the co-variance of absorbance profiles among cultivars of *B. napus*. The variance explained is shown in parentheses on each model's axis. Both models were generated using mean-centered and scaled compound contents data from UPLC-UV-TQD spectroscopy. For other component contributions, please see supporting Figures S2 and 3. Biplot overlays the score plot and the loading plot of absorbance variation among accessions.

The three first components were kept, which would be enough to explain a moderate proportion of the variance of data, according to supporting Figure S1. The first three PCs explained about 37.1% of the total variance, with about 15.4%, 12.3%, and 9.4% of the total variation explained by the first, second, and third principal components, respectively (Supporting Figure S1, coming soon in the published article).

Thus, these PCA biplots displayed scattered distributions of accessions, with the highest contribution of three compounds (**R03_M434**, **R13_M320**, and **R18_M368**) (Figure 3A). It is noteworthy that the variation of **R03_M434** was not correlated with those of **R13_M320** and **R18_M368**. The remaining compounds mainly contribute to the other PCA components that explained a small proportion of the variance. Nevertheless, PCA biplots allowed us to distinguish between spring and winter cultivar accessions of *B. napus* with slight overlapping (Figure 3A). Some of the 18 compounds exhibited positively correlated pairwise relationships (highlighted by background color) and tended to have similar loadings: **R13_M320** with **R18_M368**, **R17_MNA** with **R4_385**, and **R1_315** with **R14_MNA**.

Further supervised component analysis based on partial least squares discriminant analysis (PLS-DA) regression was performed to better reveal the different chemical profiles corresponding to the four groups of *B. napus* accessions. The definition of latent components was based on the loading weights of regression coefficients. The biplot of PC1 and PC2 highlights the significant loadings exhibited by essential compounds in the PLS-DA model (on the left of Figure 3B). Although some overlap persisted, this approach allowed a better separation of winter and spring accessions of *B. napus*.

3.4. mGWAS identified genomic regions (mQTLs) associated with novel chemical traits in the roots of *Brassica*

The root contents (based on UV_{A330}-absorbance levels) of the 18 compounds were used as phenotypic variables for quantitative trait locus (QTL) analysis. All metabolic features showed a continuous frequency distribution (Supporting Figure S2, coming soon in the published article). Features **R01_M315**, **R12_MNA**, and **R17_MNA** showed slight variations and a normal distribution. In contrast, the remaining compounds were distributed along with skewed density distributions. Therefore, QTL analysis was performed using single-nucleotide variant (SNP) markers (JF de Carvalho & Missinou, et al, in preparation).

mGWAS thus highlighted several variants associated with UV_{λ330}-absorbing compound content traits. In total, 208 significant SNPs were detected above the Gao threshold set to ^{log10}pval = 5.296 (Figure S3). Seventeen UV_{λ330}-absorbing compound content mQTLs were associated with the 11 metabolite features **R01_M315**, **R02_M431**, **R03_M434**, **R05_M434**, **R07_M551**, **R08_M681**, **R11_M523**, **R12_MNA**, **R13_M320**, **R16_M493**, and **R18_M368** (Table 3). These mQTLs were distributed on ten chromosomes A01, A03, A04, A06, A07, A09, C01, C03, C07, and C05 representing 18 distinct genomic regions on the genetic map (Table 3, Supporting Figure S3, Supporting Table S2a, coming soon in the published article)).

Ph D. manuscript | Anani Amegan MISSINOU

mQTL name	Metabolic Trait	Chrs	SNP Marker	R ² (%)
mQA06_40.0	R01_M315	A06	Bn.A06.p21800812	0.12
mQC01_34.4	R02_M431	C01	Bn.C1.p29015489	0.20
mQA03_28.6	R03_M434	A03	Bn.A03.p27377892	0.21
mQA06_37.7	R03_M434	A06	Bn.A06.p17175388	0.10
mQC07_52.2	R03_M434	C07	Bn.C7.p45129145	0.21
mQC07_52.2	R05_M434	C07	Bn.C7.p45128577	0.25
mQC03_13.6	R07_M551	C03	Bn.C3.p18789565	0.11
mQC05_39.2	R07_M551	C05	Bn.C5.p44425675	0.11
mQC03_22.2	R07_M551	C03	Bn.C3.p20525188	0.31
mQA09_15.9	R08_M681	A09	Bn.A09.p14783306	0.12
mQA07_23.3	R11_M523	A07	Bn.A07.p17017277	0.11
mQC07_37.3	R12_MNA	C07	Bn.C7.p31107417	0.10
mQA04_4.7	R13_M320	A04	Bn.A04.p4325390	0.10
mQA04_9.36	R13_M320	A04	Bn.A04.p6588935	0.24
mQA04_9.9	R13_M320	A04	Bn.A04.p7066595	0.14
mQC04_46.3	R13_M320	A04	Bn.A04.p7044446	0.11
mQA09_15.9	R16_M493	A09	Bn.A09.p14751809	0.14
mQC05_32.6	R16_M493	C05	Bn.C5.p22057936	0.10
mQA01_12.6	R18_M368	A01	Bn.A01.p12544054	0.11
mQA04_9.36	R18_M368	A04	Bn.A04.p6588935	0.15

Table 3: Genome-wide association study of $UV_{\lambda 330}$ -absorbing compound contents in *B. napus* roots.

The explained phenotypic variation of the detected QTLs ranged from 10% (mQA06_37.7, mQC07_37.3, mQA04_4.7, and mQC05_32.6) to 31% (mQC03_22.2). The content of five compounds was associated with more than one mQTL, i.e., three mQTLs for **R03_M434** and **R07_M551**, two mQTLs for **R16_M493** and **R18_M368**, and four mQTLs for **R13_M320** (Table 3, Supporting Table S3a, Supporting Figure S3, coming soon in the published article). In two genomic regions, at least two mQTLs co-localized on chromosome A04 (**R13_M320** and **R18_M368**) and chromosome A09 (**R08_M681** and **R16_M493**) (Supporting Table S3a, coming soon in the published article). Interestingly, **R13_M320**/**R18_M368** and **R08_M681**/**R16_M493** were already shown to display common structural and profile properties, such as fragmentation

patterns (Table 1), correlated level profiles, and best discriminant loadings of the spring accessions between the panel (PLS-DA, Figure 3. B).



Figure 4: Network-dissected mQTL regions illustrating associations between phenolics, and root UV-absorbing compounds content traits. K-triHx = Kaempferol-triHexoside, I-triHx = IsorhamnetintriHexoside. The genetic marker of QTL C07_37.3 was in the region of the mQTL C07_37.5 controlling two leaf phenolic compounds, as previously shown³³. Additional details are given in Supporting Table S3c, coming soon in the published article.

The network representation of the genetic architecture controlling the root concentration of 11 compounds (Figure 4), highlighted that most compounds were independently controlled by distinct and relatively small QTL networks. Only one of those QTLs, C07_37.3 (marker Bn.C07.p31107417) controlling the compound

³³ Cf. chapitre II, article 2

Ph D. manuscript | Anani Amegan MISSINOU

R12_MNA was located in the same genomic regions that one QTL that has been previously identified as associated with the control of two flavonol trihexosides (**PL32**=Isorhamnetin trihexoside and **PL17**=Kaempferol trihexoside) (Chapter II, article 2). The two structurally related and co-regulated compounds **R13_320** and **R18_368** were found to share one QTL A04_9.36, and two additional QTL were specifically associated with the control of **R13_320** and one QTL to the control of R18_368. Similar patterns of shared mQTL were also observed for two additional couples of metabolites, the QTL C07_52.2 controlling the two isomers **R03_M434** and **R05_M434**, and the QTL A09_15.9 controlling both **R08_M681** and **R16_M493**.

3.5. Analysis of potential candidate genes

Overall, 2444 genes were identified in the confidence interval of the 17 mQTL involved in the quantitative variation of the 11 compounds. About 87.7% of these genes have functional annotations inferred from *Arabidopsis thaliana*. In the confidence interval of each mQTL, at least one candidate gene was identified (Figure 5).



Figure 5: Number of genes identified in each mQTL region. The numbers between the parentheses represent the number of annotated genes. The research of candidate genes was performed on the reference genome of *B. napus* accession 'Darmor-*bzh*' version 5 (Chalhoub et al., 2014).

In the mQA09_15.9 (R08_M681 and R16_M493) confidence interval, 169 genes (with 140 annotated genes) were identified. Several genes were involved in the metabolism of sulfur-containing amino acids, carbohydrates, and phenylpropanoids. Of these, an ortholog of A. thaliana acyltransferase (BnaA09g23800D and BnaA09g23800D), (BnaA09g23290D, BnaA09g23510D, hexosyltransferase BnaA09g23800D, BnaA09g24120D, BnaA09g23290D, BnaA09g24120D), glutaredoxin and (BnaA09g24210D) were found. For the very close QTL mQTLA04_9.9 and mQTLA04 9.36 (controlling R13 M320 and R18 M368), very few genes were identified. This list included MYB family transcription factors (BnaA04g10750D and BnaA04q10750D).

3.6 Combining NMR and MS allowed a simultaneous assignment and structural elucidation of two novel root compounds in *Brassica*

R13_M320 and R18_M368 correspond to the two metabolic features reported by Wagner et al. 2019 putatively involved in the genetic resistance of *B. napus* to the clubroot isolate eH. In the present work, the confidence intervals of the mQTL controlling those two molecules encompass a very small number of candidate genes. We thus focused our efforts on the further structural characterization of those two compounds. Purification steps of R13_M320 and R18_M368 led to the isolation of 8 mg, 6 mg, and 4 mg products, respectively. When redissolved at high concentration (in methanol or DMSO) these three compounds harbored a gold-yellow coloration. Q-Exactive Orbitrap MS/MS carried out a forward analysis of the molecular formula. The elemental composition was limited to C, H, N, and O, with the error cutoff set to <3ppm. The molecular formula, C₁₃H₁₅N₅O₅, and C₁₇H₁₅N₅O₅ were respectively assigned to m/z 320.0998 [M-H]⁻ and 368.0998 [M-H]⁻ pseudo-molecular ion peaks of **R13_M320**, R18_M368, respectively. MS and MS/MS/ fragmentation allowed us to uniquely assign the molecular formula of the small anionic fragments and the loss of neutral features. The daughter ion m/z 163 is shared by the two compounds and corresponds to the molecular mass of C₆H₃N₄O₂. R13_M320 and R18_M368 both displayed a nitrogencontaining aromatic moiety and were both carboxylated. Additional structural investigations were then performed on these three compounds using a series of NMR techniques, as follows:



Figure 6: Comparison of 1H NMR and 13C NMR 500 MHz data from experimental chemical shifts (δ_{exp}) for R13_M320 and R18_M368 compounds in deuterated methanol (CD₃OD) (δ in ppm, J in Hz).

R13_M320 (C₁₃H₁₅N₅O₅): This compound was analyzed in MeOD-d4 solvent at a temperature of 300K. A leucine fragment was fully characterized in ¹H and ¹³C for all the signals by HSQC / HMBC / COSY / NOESY spectra except for the COOH function (suggested by mass spectra). A connection to a poly-nitrogen aromatic structure has been identified through the NH of the leucine fragment. The structure of this compound has been partially solved (figure 6. A-C) as a leucine fragment (C₇H₁₂NO₃) connected

Ph D. manuscript | Anani Amegan MISSINOU

to a pteridine-like poly-nitrogenated aromatic heterocycle ($C_6H_3N_4O_2$), for which the exact position of the four nitrogen atoms is still unresolved (Figure 7 and Figure 9).



Figure 7: NMR-based characterization of the substructure of R13_M320. A/ Characterized fragment **B**/ Observed HMBC correlations **C**/ observed NOESY correlations of R13_M320 characterized fragment. The wavy line represents a still not fully resolved poly-nitrogenated aromatic heterocycle (C₆H₃N₄O₂).

R18_M368 (C₁₇H₁₅N₅O₅): This compound has been analyzed in DMSO-d6 solvent at a temperature of 300K. The fragment has been fully characterized in ¹H and ¹³C for all the signals, by HSQC / HMBC / COSY / NOESY spectra. The connection to the polynitrogen aromatic cycles has been identified through the NH of the amino acid fragment. The structure of this compound was thus partially solved (figure 7) as an amino acid-like with a chain R = $-CH_2-CH_2-Ph$, connected to a pteridine-like polynitrogenated aromatic heterocycle (C₆H₃N₄O₂) for which the exact position of the four nitrogen atoms is still unresolved (Figure 8 and Figure 9).



Figure 8: NMR-based characterization of the substructure of R18_M368. A/ Characterized fragment **B**/ Observed HMBC correlations **C**/ Observed NOESY correlations of R18_M368 characterized fragment. The wavy line represents a still not fully resolved poly-nitrogenated aromatic heterocycle (C₆H₃N₄O₂).



Monoisotopic Mass: 321.10731Formula: $C_{13}H_{15}N_5O_5$ Charged anion: m/z 320.10004 [M-H]⁻ Charged cation: m/z 322.11459 [M+H]⁺

Monoisotopic Mass: 369.10731 Formula: C₁₇H₁₅N₅O₅ Charged anion: m/z 368.10004 [M-H]⁻ Charged cation: m/z 370.11459 [M+H]⁺

Figure 9: Hypothetical structures of the two compounds R13 (A) and R18 (B). The proposition of a pteridine structure for the poly-nitrogenated aromatic heterocycle is detailed in the discussion.

4. Discussion

The *Brassica* root repertoire of specialized metabolites is largely dominated by glucosinolates. However, we need to further investigate the still unexplored root chemical diversity that could also play important roles in plant interactions with its soil environment. This work was originally focused on the search for root-specific phenolic compounds. However, the difficulties that we encountered for the annotation of $UV_{\lambda330}$ -chromatographic profiles from *Brassica* root extracts, reflecting the absence of consistent mass-spectrum data in public databases, suggested that these compounds could be worthy of additional investigations.

UV_{A330}-based semi-quantitative data highlighted large ranges of concentrations for every 18 compounds among the accessions of *B. napus*. Multivariate approaches on these unknown compounds clearly distinguish between chemotypes for SOSR and WOSR accessions. It also highlighted, among others, the large contribution of **R13_M320** and **R18_M368** accumulation to the SOSR chemotype, thus bringing an interesting complementary view on the root phytochemical diversity of oilseed rape (following our previous work on glucosinolates and phenolicsf, Missinou et al., 2022³⁴). This also opens new perspectives about the possible biological roles of those compounds and the possible agronomic significance of their variations between oilseed rape cultigroups.

Eleven compounds were controlled by independent mQTL networks, which means that the chemotypical contrasts resulted from combinations of allele variations at multiple mQTL. The absence of colocalization with mQTL previously identified in this panel of

³⁴ Chapitre II, articles 1 et 2

accessions, suggests that the phytochemical profiles defined by these eleven compounds are controlled by evolutionary or breeding processes which are independent of those involved in the control of glucosinolates and phenolics (Missinou et al., 2022).

Experimental data support that the poly-nitrogenated aromatic heterocycle (C₆H₃N₄O₂) from **R13_M320** and **R18_M368** is based on a pteridine-like structure, although the exact position of nitrogen atoms on the bicyclic structure was not fully resolved by the NMR approaches used in the present work (figure 9). To the best of our knowledge however, pteridine basic structure would be the only proposition that has been reported in natural products, and biosynthetic pathways of pteridine-based compounds are reported in plants including flavin and folic acid (M. Pahor, Manini, 2008). The occurrence in the roots of *B. napus* of two amino-acid conjugates of this basic structure is reminiscent of the aminoacid conjugation of a series of plant signaling compounds like jasmonic acid, indole-3-acetic acid, or salicylic acid, by GH3 enzymes belonging to a firefly luciferase-like superfamily.

Bioinformatic molecular network approaches use the similarities between MS/MS fragmentation patterns to screen for sub-groups of compounds that share common chemical sub-structures in large metabolomic datasets. In the present work, the lack of high-resolution MS2 fragments impaired the use of classical bioinformatic approaches (such as GNPS). However, the fact that pairs of compounds displayed both common MS2 fragmentation patterns and common mQTL not only suggested that they could display similar chemical structures, but also suggested that they could share common biosynthetic or regulation pathways. One QTL on chromosome A04

controlled both R13_M320 and R18_M368, i.e. two pteridine-like polynitrogenated aromatic heterocycles conjugated to leucine (m/z 320.099 [M-H]) and homophenylalanine (368.099 [M-H]). This suggests that this QTL might be involved in the biosynthesis of the core pteridine-like moiety. Each of these two molecules was also controlled by additional distinct QTLs, that might be involved in the conjugation with each amino acid. Additional work is now required not only to finalize the full annotation of the pteridine-like structure, which will require enhanced amounts of purified compounds. The confidence intervals of the mQTL encompass a reasonably small number of candidate genes, thus encouraging further work toward the validation of their possible regulatory roles in the biosynthesis of these compounds. This will include 1/ a complementary analysis of the gene content in the mQTL confidence intervals in genome sequences from other B. napus accessions 2/ the analysis of available root transcriptomes to identify the subset of expressed genes, 3/ an analysis of the possible presence of metabolic gene clusters, and 4/ the use of Cas9 strategies for the validation of hypotheses.

The work of Wagner et al. 2019 (based on a linkage genetic analysis on a progeny derived from a 'Darmor-*bzh*' x 'Yudal' cross), had previously reported that one major mQTL on chromosome C09 is involved in the control of the constitutive root content of **R13_M320/R18_M368** and the resistance to the isolate eH of *P. brassicae*. We however did not detect this mQTL in the present work, which might suggest that the underlying allele variation might be rare in the whole genetic panel, thus impairing the detection of this QTL by the GWAS approach in the present work.

5. Conclusion

We provided new insights into root metabolism and identified new compounds and the key genomic regions underlying the variations of their concentrations in roots. The structural draft of two of these root chemical traits suggests that these compounds belong to a poorly studied subfamily of pteridine-like molecules, which brings an added value to the catalog of phytochemical diversity of *Brassica* in the springboard for a better understanding of root chemical ecology. The next challenge will be to characterize the candidate genes through cloning or transgene validation to interpret these associations in a physiological and genomic context. Therefore, metabolic GWAS-based dissection of genetic bases underlying the diversity of UV-absorbing compounds content combined with mass spectrometry and nuclear magnetic resonance is complementary to uncovering genes that regulate their accumulation in plants and offers discoveries and perspectives into the research of specialized plant metabolism.

Author contributions

Writing-review and editing: AAM, AG, MMD, AB, CO, PJ. Contributed data: NM, AAM YH, AND CO. Figures and tables: AAM, JFC, and CO. Sample analysis: AAM, YH, NM, and CO. Statistical analysis and software development: AAM, JFC, YH, PJ, and CO. Protocol and workflow design: AAM, NM, YG, and AG. Funding acquisition: MMD, AB, and AG.

Funding sources

The salaries of A.A.M. and J.F.C. were provided by INRAE and Région Bretagne. Operating funding was supported by INRAE, Institut Agro, Université Rennes 1, and PROMOSOL. Salaries of O.H. and D.A.S. were provided by PROMOSOL (project DESCRIBE).

Associated content

Supporting Figures



Supporting Figure S1: Barplot of the eigenvalues explaining the percent of variance absorbance of unknown root compounds on each principal component. Absorbance was measured at λ 330nm in the panel of 304 accessions of *Brassica*.



Ph D. manuscript | Anani Amegan MISSINOU

Supporting Figure S2: Frequency distribution histograms of eighteen unknown root chemical traits UVabsorbing (λ 330nm) in 250 accessions of *Brassica napus*. Each bar represents the mean of absorbance for one variety, from 3 independent biological replicates. The units of absorbance are absorbance units per gram of dry weight (AU.g⁻¹ DW).





Supporting Figure S3: Manhattan plot of genome-wide association analysis of root UV-absorbing compounds content. Homoeologous chromosomes of the A and C subgenomes of the allopolyploid *Brassica napus* are represented in blue and red blocks, respectively. The horizontal lines represent the

Ph D. manuscript | Anani Amegan MISSINOU
Gao threshold (Gao, Starmer, & Martin, 2008) calculated using the Bonferroni multitest at 5% and 10%, respectively, to adjust the threshold of the genome-wide significance p-value.

List of supporting Figures

Supporting Figure S1: Root UV-absorbing compounds content in a panel of 304 *Brassica* accessions. **S1b**: UV-absorbance measurement of unknown compounds in each biological replicate AU.g⁻¹ DW (arbitrary unit of UV-absorbing compounds content fitting with 1/x weighting). **S1c**: Mean of UV-absorbing root unknown compounds, min, max, percentile 5, and percentile 95. **S1d**: Selection of accessions belonging to percentile 5 and percentile 95 for each UV-absorbing root unknown compound (arbitrary unit peak area per DW).

Supporting Figure S2: Frequency distribution histograms of eighteen unknown root chemical traits UV-absorbing (λ 330nm) in 250 accessions of Brassica napus. Each bar represents the mean of absorbance for one variety, from 3 independent biological replicates. The units of absorbance are absorbance units per gram of dry weight (AU.g⁻¹ DW).

Supporting Figure S3: Manhattan plot of

genome-wide association analysis of root UV-absorbing compounds content. Homoeologous chromosomes of the A and C subgenomes of the allopolyploid Brassica napus are represented in blue and red blocks, respectively. The horizontal lines represent the Gao threshold (Gao, Starmer, & Martin, 2008) calculated using the Bonferroni multitest at 5% and 10%, respectively, to adjust the threshold of the genome-wide significance p-value.

List of tables

Table 1: Unknown root compounds absorbing at λ 330nm in the panel of 304 accessions of *Brassica*. [M] stands for the neutralized structure and (*) stands for the very percentage of relative intensity.

Table 2: Descriptive statistics (mean, percent, minimum, maximum) of unknown *Brassica* root compounds absorbing at λ 330nm. Details for each accession and compound are given in Supporting Table S1). AU: UV-absorbing compounds content unit, and DW: dry weight.

Table 3: Genome-wide association study of $UV_{\lambda 330}$ -absorbing compound contents in *B. napus* roots.

List of Figures

Figure 1: UPLC-UV chromatogram of a leaf extract (QC).

Figure 2: Heatmap showing differences of absorption among 304 accessions of *Brassica* for eighteen unknown root compounds absorbing at λ 330nm. (A) Each row and column respectively represent one accession and one unknown compound detected in the study. (B) Annotations of accession groups per species and cultivars. (C) Boxplots display the variation level of absorbance (AU.g⁻¹ DW) of each unknown compound across accessions.

Figure 3: Multivariate analysis of unknown root compounds of *B. napus* cultivars accessions. (A) Biplot of the unsupervised clustering using PCA model based on unknown root compounds of *Brassica* varieties (B) Biplot of the supervised clustering using the PLS-DA model based on unknown root compounds of *B. napus* varieties.

Figure 4: Phenotypic variation and Pearson correlation coefficients of unknown *Brassica* root compounds absorbing at λ 330nm. (A) An upper triangular matrix showing Pearson's correlation for 18 UV-absorbing in 304 accessions of *Brassica*. (B) QTL numbers associated with UV-absorbing compounds content from genetic linkage study.

Figure 5: Number of genes identified in each mQTL region. The numbers between the parentheses represent the number of genes. The research of candidate genes was performed on the reference genome of *B. napus* version 5 (Chalhoub et al., 2014).

Figure 6: Comparison of ¹H NMR and ¹³C NMR 500 MHz data from experimental chemical shifts (δexp) for R13_M320 and R18_M368 compounds in deuterated methanol (CD₃OD) (δ in ppm, J in Hz).

Figure 7: NMR-based characterization of the substructure of R13_M320. A/ Characterized fragment B/ Observed HMBC correlations C/ observed NOESY correlations of R13_M320 characterized fragment. The wavy line represents a still not fully resolved poly-nitrogenated aromatic heterocycle ($C_6H_3N_4O_2$).

Figure 8: NMR-based characterization of the substructure of R18_M368. A/ Characterized fragment B/ Observed HMBC correlations C/ Observed NOESY correlations of R18_M368 characterized fragment Wavy line represents a still not fully resolved poly-nitrogenated aromatic heterocycle ($C_6H_3N_4O_2$).

Figure 9: Hypothetical structures of the two compounds R13 (A) and R18 (B). The proposition of a pteridine structure for the poly-nitrogenated aromatic heterocycle is detailed in the discussion.

Ph D. manuscript | Anani Amegan MISSINOU

References

Baxter CS, Warshawsky D. 2012. Polycyclic Aromatic Hydrocarbons and Azaaromatic Compounds. Patty's Toxicology 5, 371–428.

Bergmeier SC. 1999. Principles of Chemical Nomenclature. A Guide to IUPAC Recommendations.

Bright JW, Chen ECM. 1983. Mass spectral interpretation using the 'rule of "13". Journal of Chemical Education 60, 557–558.

Chalhoub B et al. 2014. Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science 345, 950–953.

Chattha FA, Nisa M, Munawer MA, Kousar S. 2016. Coumarin-Based Heteroaromatics as Plant Growth Regulators. Plant Growth.

Dray S, Dufour AB. 2007. The ade4 package: Implementing the duality diagram for ecologists. Journal of Statistical Software 22, 1–20.

Erktan A, McCormack ML, Roumet C. 2018. Frontiers in root ecology: recent advances and future challenges. Plant and Soil 424, 1–9.

FAOSTAT. 2022. Production of rapeseed by main producing countries 2020/2021. Crops and Livestock Products (Production). Available online.

Fernandes F, Valentão P, Sousa C, Pereira JA, Seabra RM, Andrade PB. 2007. Chemical and antioxidative assessment of dietary turnip (*Brassica* rapa var. rapa L.). Food Chemistry 105, 1003–1010.

Francisco M, Ali M, Ferreres F, Moreno DA, Velasco P, Soengas P. 2016. Organ-specific quantitative genetics and candidate genes of phenylpropanoid metabolism in *Brassica* oleracea. Frontiers in Plant Science 6, 1–14.

Gao X, Starmer J, Martin ER. 2008. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genetic Epidemiology 32, 361–369.

Gresshoff PM. 2003. Post-genomic insights into plant nodulation symbioses. Genome Biology 4.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849.

Huang AC, Jiang T, Liu YX, Bai YC, Reed J, Qu B, Goossens A, Nützmann HW, Bai Y, Osbourn A. 2019. A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. Science 364.

Kittipol V, He Z, Wang L, Doheny-Adams T, Langer S, Bancroft I. 2019. Genetic architecture of glucosinolate variation in *Brassica napus*. Journal of Plant Physiology 240, 152988.

Knill T, Schuster J, Reichelt M, Gershenzon J, Binder S. 2008. Arabidopsis branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. Plant Physiology 146, 1028–1039.

Krishnan TR, Ibraham I. 1994. Solid-phase extraction technique for the analysis of biological samples. Journal of Pharmaceutical and Biomedical Analysis 12, 287–294.

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T. 2003. Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proceedings of the National Academy of Sciences of the United States of America 100, 14587–14592.

Lee RWH, Malchev IT, Rajcan I, Kott LS. 2014. Identification of putative quantitative trait loci associated with a flavonoid related to resistance to cabbage seedpod weevil (*Ceutorhynchus obstrictus*) in canola derived from an intergeneric cross, Sinapis alba × *Brassica napus*. Theoretical and Applied Genetics 127, 419–428.

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. 2011. FaST linear mixed models for genome-wide association studies. Nature Methods 8, 833–835.

Liu S, Huang H, Yi X, Zhang Y, Yang Q, Zhang C, Fan C, Zhou Y. 2020. Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome-wide association study. Plant Biotechnology Journal 18, 1472–1484.

M. Pahor, Manini and MC. 2008. Pterin chemistry and its relationship to the molybdenum cofactor. Bone 23, 1–7.

Nagaharu U. 1935. Genome-analysis in *Brassica* with specila reference to the experimental formation of *B. napus* and particular mode of fertilization.

Nakabayashi R, Saito K. 2015. Integrated metabolomics for abiotic stress responses in plants. Current Opinion in Plant Biology 24, 10–16.

Negro SS, Millet EJ, Madur D, Bauland C, Combes V, Welcker C, Tardieu F, Charcosset A, Nicolas SD. 2019. Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. BMC Plant Biology 19, 1–22.

Pellegrin V. 1983. Molecular formulas of organic compounds the nitrogen rule and degree of unsaturation. Journal of Chemical Education 60, 626–633.

Rincent R, Moreau L, Monod H, et al. 2014. Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genetics 197, 375–387.

Rohart F, Gautier B, Singh A, Cao K-AL. 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. bioRxiv, 108597.

Russell V. Lenth, Paul Buerkner, Maxime Herve, Jonathon Love, Hannes Riebl HS. 2021. emmeans: Estimated Marginal Means, aka Least-Squares Means [R package emmeans version 1.7.0]. American Statistician 34, 216–221.

Schmid NB, Giehl RFH, Döll S, Mock HP, Strehmel N, Scheel D, Kong X, Hider RC, von Wirén N. 2014. Feruloyl-CoA 6'-Hydroxylase1-dependent coumarins mediate iron acquisition from alkaline substrates in Arabidopsis. Plant Physiology 164, 160–172.

Shroff R, Vergara F, Muck A, Svatos A, Gershenzon J. 2008. Nonuniform distribution of glucosinolates in *Arabidopsis thaliana* leaves has important consequences for plant defense. Proceedings of the National Academy of Sciences of the United States of America 105, 6196–201.

Stringlis IA, Yu K, Feussner K, De Jonge R, Van Bentum S, Van Verk MC, Berendsen RL, Bakker PAHM, Feussner I, Pieterse CMJ. 2018. MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. Proceedings of the National Academy of Sciences of the United States of America 115, E5213–E5222.

Sun-Ju Kim, Kouei Fujii1, Zaidul Islam Sarker Mohamed, Hyun-Woong Kim, Hiroaki Yamauchi GI. 2008. Identification and Quantitative Determination of Glucosinolates in *Brassica napus* cv. Hanakkori. 17, 1097–1101.

VanRaden PM. 2008. Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science 91, 4414–4423.

Wagner G, Laperche A, Lariagon C, Marnet N, Renault D, Guitton Y, Bouchereau A, Delourme R, Manzanares-Dauleux MJ, Gravot A. 2019. Resolution of quantitative resistance to clubroot into QTL-specific metabolic modules. Journal of Experimental Botany.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis.

Willcott MR. 2009. MestRe Nova. Journal of the American Chemical Society 131, 13180–13180.

Zhuang Q, Gan J, Kong L, Wang H ZY. 1997. Design and application of vacuum manifold for solid phase extraction. Chinese Journal of Chromatography.

Ph D. manuscript | Anani Amegan MISSINOU

Chapter III: Metabolic responses of Brassica napus

after Leptosphaeria maculans infection

Introduction du chapitre III

L'objectif de ce troisième chapitre est d'identifier les défenses chimiques mobilisées par le colza lorsqu'il perçoit, au niveau de sa tige, l'infection par *L. maculans*. L'analyse des données transcriptomiques générées par l'équipe d'accueil dans le contexte de l'interaction *B. napus-L. maculans* (Delourme et al, en préparation), a mis en lumière l'expression de gènes jouant un rôle prépondérant dans cette interaction et notamment ceux impliqués dans les voies de biosynthèse des GLSs aliphatiques et indoliques. L'objectif de mon travail était de valider au niveau métabolique ces résultats. En utilisant la méthode de profilage ciblé développée dans le deuxième chapitre, nous avons suivi les modifications métaboliques induites par l'infection au niveau du contenu en GLSs aliphatiques dérivés de la méthionine et en GLSs indoliques dérivés du tryptophane. Cette analyse a tout d'abord révélé une induction des teneurs en GLSs indoliques dans la biosynthèse de ces composés (Delourme et al., article en préparation).

Dans une deuxième partie, je me suis attaché à déterminer la localisation de ces composés chimiques de défense (constitutifs et induits) mobilisés par *B. napus* au stade tige suite à l'infection par *L. maculans*. Le postulat de départ est que certains métabolites s'accumulent au site d'infection et agissent directement sur l'agent pathogène pour limiter sa progression centripète au sein des tissus internes de la tige, en particulier chez le génotype résistant. Avec Anne Levrel, nous avons développé, en collaboration avec Mélanie Lagarrigue-Reboutier et Régis Lavigne de la plateforme PROTIM (<u>PRO</u>Téomes et <u>IM</u>ages) de Rennes, une approche de métabolomique spatiale appliquée à des lamelles (coupes transversales) de tiges infectées et non infectées de colza. Elle nous a permis de caractériser des réponses métaboliques spécifiques aux zones entourant la nécrose induite par l'infection. L'analyse de ces

réponses chez un génotype sensible et chez un génotype partiellement résistant favorise une meilleure exploration du potentiel rôle de ces réponses métaboliques dans la défense contre *L. maculans*.

1. Introduction

The production of oilseed rape in numerous countries is affected by the fungal pathogen *Leptosphaeria maculans* (commonly named phoma), which causes blackleg disease also known as stem canker or phoma stem canker (Fitt et al., 2006; Dukhnytskyi, 2019). The phoma-infected oilseed rape displays necrotic lesions at the stem, provoking symptoms of water and mineral impairment, reducing growth, and leading to plant mortality (McGee and Emmett, 1977; Howlett et al., 2001). A high incidence of disease results in production losses and severe damage to the rapeseed oil industry (Wang et al., 2020). *L. maculans* is a hemibiotroph whose complex life cycle includes alternating phases of saprophytism on stem residues, a short phase of the asymptomatic biotrophic stage during which it colonizes the apoplast, subsequently switching to a necrotrophic lifestyle in leaf lesions after ascopores landing on plant leaves, a long asymptomatic endophytic stage in plant tissue, and finally the second phase of symptomatic necrotrophy at the base of the stem, which affects the central tissues and kills the plant (West et al., 2001; Rouxel and Balesdent, 2005; Bousset et al., 2018).

Genetic factors involved in quantitative resistance to Blackleg mainly exert their effects at the stem stage of infection leading to the dampening of the extent of necrosis caused by *L. maculans* in partially resistant accessions (Delourme et al., 2014; Pilet-Nayel et al., 2017). One hypothesis is that specific metabolites accumulate at the site of infection and act directly on the pathogen to limit its centripetal progression within tissues, particularly in resistant genotypes.

Since its introduction by Caprioli et al., 1997, mass spectrometry imaging technology has allowed the molecular mapping of diverse tissue cross-sections, allowing visualization of molecular details in their tissue context (Qin et al., 2018). This technology is based on the irradiation of very thin tissue with a sub-micron laser spot, previously covered with an optimal matrix solution that allows the desorption and provides a suitable environment for the ionization of the molecules (Dong et al., 2016). At thousands of different spatial locations, MALDI-MSI records each sampled position (pixel), mass spectrum, and intensity of each feature (mass-to-charge ratio), which represents the abundance of local analytes (metabolites, lipids, peptides, and proteins). MALDI-MSI provides multivariate two-spatial dimensional bioimages containing mass signals per pixel at the end of data collection. Although widely used in proteomics, MALDI-MSI for screenings of plant micromolecules is not very common, especially in the field of biotic interactions. Some works have shown that MSI is useful in the study of symbiosis (Ye et al., 2013), endophytism (Kusari et al., 2014; Wang et al., 2015), herbivory (Shroff et al., 2015; Shroff et al., 2008), and plant-pathogen interactions (Klein et al., 2015; Santos et al., 2015; Seneviratne et al., 2015; Barbosa et al., 2018; Elian ST et al., 2021).

Since little is known about plant cellular responses and partial resistance mechanisms at the stem stage of infection, we focused the present work on this stem stage, using the phytopathological test developed by (Levrel et al., 2022). In the first approach, we followed the regulation of GLSs in infected stems, using the UPLC-UV-TQD method described in chapter 2 (Missinou et al., 2022). Then, we developed a spatial metabolomics approach to investigate the hypothesis that the tissular spatialization of *B. napus* responses to phoma infection could help identify localized biochemical regulations that could be induced at the interface between plant tissues and pathogen progression's front. The metabolic responses of two genotypes (the partially resistant

'Darmor' and the moderately susceptible 'Bristol') of oilseed rape were analyzed and compared by MALD-MSI.

2. Materials and Method

2.1. Chemicals

HPLC-grade acetonitrile, methanol, and formic acid were from Carlo Erba. Authentic glucosinolate standards (Supporting Table S2) were acquired from PHYTOPLAN Diehm & und Neuberger GmbH. Chemical structures were drawn using MarvinSketch 20.21 (ChemAxon Ltd.).

2.2. Plant growth

Two accessions of *B. napus ssp. oleifera* (winter oilseed rape (WOSR)), the partially resistant 'Darmor' and the moderately susceptible 'Bristol' were cultivated as described in Missinou et al., 2022 and analyzed in three biological replicates, each with eight plants per accession. Seeds were germinated in a climate-controlled greenhouse in pots using a mixture of soil and sand (ratio 2:1). Plants were grown in a growth chamber under a daily cycle of 16h/8h in light/dark at 20°C and 18°C, respectively. They were fertilized twice with a half-dosed Hoagland solution.

2.3. Inoculum preparation and petiole inoculations

Inoculum consisted of a suspension at 10⁷ conidia per mL. Spore concentrations were standardized to 10⁷ spores per mL after counting aliquots with a Malassez cell under a magnifying lens, aliquoted, and stored at -20°C until the plant inoculation experiment. To reduce a long asymptomatic period, plantlets at the 3-leaves stage were inoculated with 10µL of the spore suspension at 10⁷ spores/mL of JN2 isolate of *L. maculans* on

cut petioles (Ermel et al., 2021). Plants treated with distilled water instead of spore suspension were used as controls. Both groups of plants (control and inoculated) were maintained under controlled conditions and covered with opaque plastic to create a 100% relative humidity atmosphere for 48 h, in the dark for the first 24 h. In the absence of light, increased humidity is favorable to spore germination.

2.4. Metabolomic analysis

2.4.1. Targeted metabolomic profiling

All steps of targeted metabolomic profiling, including extraction, chromatographic conditions (gradient, column, parameters), and quantification, were performed as previously described (Missinou et al., 2022).

2.4.2. Untargeted metabolomic profiling procedures: Sampling, GLSs extraction,

and quantification

Fourteen days after inoculation, stems were sampled by a cross-section 1-5 mm around the necrotic area, wrapped in aluminum foil, and immediately immersed in liquid nitrogen. The samples were stored at -80 °C until they were lyophilized to dryness. The dried tissues were disintegrated by shaking for 60 s with a frequency of 30 Hz using the TissueLyser with grinding jars (10 ml) and stainless-steel grinding balls (Qiagen, Hilden, Germany). GLSs were quantified using the chromatographic procedures detailed in Missinou et al. 2022³⁵.

³⁵ Chapitre II, article 1

2.4.3. Mass spectrometry imaging (MSI) by MALDI (Matrix-Assisted Laser Desorption-Ionisation) procedures

2.4.3.1. Preparation of stem cross-sections

At fourteen days post-incubation (14 dpi), the infected stems (including the necrotic cavity and the proximal area) were sampled and directly embedded in a carboxymethylcellulose (homogenized CMC solution at 2.5% prepared at 60°C) using Peel-A-Way embedding molds under dry ice. To preserve tissue shape and protect biological components from degradation, these procedures were executed in 30-60 seconds. The samples were kept at -80 °C until sectioning. The frozen tissue block was removed and was cross-sectioned in a cryostat microtome (Leica Microsystems GmbH, Wetzlar, Germany) with the microtome chamber chilled at -22°C and the specimen holder at -12°C. Cross-sections 20-µm thick were thaw mounted onto ITO-coated microscopic slides (Bruker Daltonik GmbH) adapted for MALDI mass spectrometry. Samples were sprayed before mass spectrometry analysis with a matrix of DHB or 9AA for the mass-spectrometry analyses with positive and negative modes, respectively, dissolved at 50 mg/mL in 5:5 (v/v) MeOH/1% TFA in water.

2.4.3.2. High-resolution MALDI FT-ICR mass spectrometry imaging measurement

The analytes of the stems diffuse onto the matrix droplets, and MALDI mass spectrometry technology analyzes them every 50 μ m, ionizes them, and measures the mass-to-charge ratio (m/z), generating one spectrum per spot, optical images of the stem, and a feature distribution map. MALDI-MSI was performed using Solarix XRTM 7 Tesla FT-ICR (Bruker Daltonics, Bremen, Germany) mass spectrometer. The spatial resolution was set to 50 μ m, and laser shots per spectrum were used. The mass spectra were acquired in the positive ion mode with an m/z range of 100 up to 1400.

2.4.2.3. Mass spectrometry imaging data processing

The ions images were visualized using SCiLS Lab software 2016 (v4.01.8781, SCiLS Lab GmbH, Bremen, Germany), containing some customizable algorithms in the analysis pipeline. Unsupervised multivariate analysis was performed to generate a spectral similarity map. The clusters of similar spectra were coded by the same color using the bisecting K-means method, spectra differences were measured by correlation distance, and cluster centers were randomly initialized. Probabilistic Latent Semantic Analysis (pLSA) was used to assign tissular components to mass spectra (Hofmann and Thomas, 1999; Hanselmann et al., 2008). Discriminative m/z values in different regions were obtained by ROC (receiver operating characteristics) analysis. It is a supervised univariate that identifies features (m/z values) that most strongly differentiate two predefined groups by the measure of the area under the curve (AUC) of 0.75 ³⁶. All data processing was performed using weak denoising, root means square normalization methods, three ppm as minimal interval width, and maximal value on interval was chosen as interval processing mode.

3. Results and Discussion

3.1. Regulation of GLS content in oilseed rape stems infected by *L. maculans*

The analysis of necrosis symptoms revealed a greater progression of centripetal necrosis in the susceptible accession 'Bristol' compared to partially resistant accession

³⁶ An AUC of 0.75 represents a probability (75% chance) that the model correctly separates them into two different classes.

'Darmor' (as illustrated in **Figure 1A**). A detailed analysis of the partial resistance of 'Darmor' and the near isogenic dwarf line 'Darmor-*bzh*' is reported in Levrel et al. 2022, revealing that the level of partial resistance was similar in 'Darmor' and its dwarf nearisogenic line 'Darmor-*bzh*' (harboring short stems). This suggested that the observed phenotypic contrasts between 'Bristol' and 'Darmor', recorded by the length of necrosis, depend on intrinsic defense factors and not on plant developmental aspects.

Next, to examine plant metabolite patterns regulated by *L. maculans*, metabolites were extracted from the inoculated and non-inoculated stems of three accessions at 14- and 28-days post-inoculation (three or two biological replicates, each replicate was made from pooled samples from 6 individual plants). Based on the analytic procedure described in Missinou et al., 2022, thirty-six metabolites, including thirteen Met-derived aliphatic-GLSs, five Trp-derived indolic-GLSs, five aromatic-GLSs, and thirteen branched and linear alkyl-GLSs, were quantitatively profiled using liquid chromatography-mass spectrometry (UPLC).

The most remarkable metabolic consequence of fungal infection was a large increase of the concentration of Trp-derived indolic-GLSs, at the two kinetic times, in both partially resistant genotypes 'Darmor' and 'Darmor-*bzh*' and the susceptible genotype 'Bristol' (**Figure 1B**). Additional regulations of individual GLSs were also observed, notably an induction of G32 (hexyl-GLS, isomer 1), a branched alkyl-GLS. This induction was weak, with only few nmol.g⁻¹ DW in infected stems, but this response was sharply contrasted compared to the undetectable concentrations in non-inoculated stems.



Figure 1: Regulation of GLSs induced by *Leptosphaeria maculans* in the stems of the partially susceptible accession 'Bristol' and the partially resistant accession 'Darmor'. (A) Illustration of the necrotic cavity at stem level at 14dpi in 'Bristol' and 'Darmor' (Image by Anne Levrel). (B) GLS contents in stems (NI=non-inoculated, JN2=infected by *L. maculans* isolate JN2), at 14- and 28-days post-inoculation (dpi). Yellow = Met-derived aliphatic GLSs. Green = Phe/Try-derived phenylalkyl GLSs (aromatic-). Light salmon = Trp-derived indolic-GLSs. Dark blue= branched and linear alkyl-GLSs. n=3 biological replicates (except 'Darmor' NI and JN2 with n=2). Each replicate is a pool of 6 plants. Stars indicate statistical differences following the student test (<0.05). G32=hexyl-GLS (one of the branched alkyl-GLSs) was not detected in the stems of non-inoculated plants.

Robin et al., (2017) previously reported an increase in aliphatic glucosinolates and indolic glucosinolates 15 days after *L. maculans* infection correlated with the simultaneous changes in levels of expression of GLS biosynthetic genes in the leaves of resistant and susceptible lines of cabbage. Similar regulations have been proposed to be associated with the response to white mold induced by *Sclerotinia sclerotiorum* fungus with a simultaneous increase in the levels of glucoiberverin (aliphatic-GLS) and

glucobrassicin (indolic- GLS) (Abuyusuf et al., 2018). In the present work, however, the observed fungal-triggered induction of indolic-GLS content in stems was observed at similar magnitudes in susceptible and partially resistant accessions. Thus, if those responses might play a role in the general defense against the fungus, they do not seem to explain the partial resistance.

The induction of indolic-GLS is consistent with transcriptomic data that were investigated in a previous project (Leptolife, coordinated by Thierry Rouxel, BIOGER, INRAE-AgroParisTech), using the same series of biological samples. These data are not presented here but this work highlighted the induction of genes involved in Trp-derived indolic-GLSs as one of the key elements of the plant responses to the fungal infection in stems.

The transcriptomic work conducted in the project LEPTOLIFE also highlighted the induction of genes identified by (Klein and Sattely, 2017), as involved in the biosynthesis of a series of indolic phytoalexins derived from indolic-GLSs. Thus, it can be hypothesized that the induction of indolic-GLSs observed in our chromatographic analyses might serve in furnishing the metabolic precursors for the biosynthesis of multiple phytoalexins involved in the defense response. To further explore this hypothesis, a complementary untargeted LC-HRMS approach was started during my Ph.D. to elucidate the impact of fungal infection on a large spectrum of indolic-derived specialized compounds, including those reported by Pedras (Pedras et Yaya, 2010) and additional putative other sulfur-containing phytoalexins (Figure 2). However, due to a long series of technical misadventures, a high-quality metabolomic raw dataset from this work has been acquired only recently. A detailed analysis of this dataset

cannot be presented here. Soon, the combination of these untargeted metabolomic and transcriptomic datasets will enable the analysis of metabolite/gene co-regulations, thus offering interesting opportunities for the investigation of the still unknown metabolic pathways involved in the biosynthesis of phytoalexins and branched and linear alkyl-GLSs in *Brassica*.



Figure 2: Untargeted full-scan LC-HRMS analyses of the metabolic responses to *Leptosapheria* infection in oilseed rape stems in 'Darmor' and 'Bristol'. The two major indolic phytoalexins spirobrassinin and cyclobrassinin are indicated. This preliminary work highlighted the induction of several previously unknown sulfur-containing and pathogen-induced metabolic features (not shown in this figure). Additional work (including further biological replicates) is now required to further document this phenomenon.

3.2. Mass spectrometry imaging reveals the spatial distribution of metabolic responses in stems of *B. napus* after *L. maculans* infection

3.2.1. Metabolic responses to *L. maculans* infection in the two genotypes of *B. napus* 'Darmor' and 'Bristol'

The spatial metabolomic approach was performed on inoculated and non-inoculated

stem samples at 14 dpi. The MALDI-FT-ICR-MSI analysis of stem cross-sections at

the spatial resolution of 50 µM led to the acquisition of more than 20.000 mass-spectra

per sample with the positive mode with the DHB matrix (23.448, 22.421, 24.879 and 21.380 mass-spectra for 'Darmor' inoculated, 'Darmor' control, 'Bristol' inoculated and Bristol control, respectively). Benefiting from the strong resolving power (140.000) of the FT-ICR-MSI, each mass spectrum from this dataset contained a huge amount of mass signals. A series of supervised and unsupervised statistical analyses of this MALDI-FT-ICR MSI dataset were performed using the SCiLS (Figure 3).

First, the spatial segmentation (unsupervised multivariate analysis) generated a spectral similarity map (Figure 3. A), allowing the comparison of the global metabolic responses to phoma in susceptible and resistant genotypes. This highlighted that 1/ metabolic spectra from non-inoculated stems were strongly different from those of inoculated samples, for both plant genotypes, and 2/ different metabolic responses were observed depending on spatial sub-zones in the infected stems.

Then, a pLSA analysis of the dataset allowed us to distinguish 5 groups (components 1 to 5) of metabolomic features, each of those five components being associated with different patterns of regulations (Figure 3. B). First, components 1 and 2 both regroup metabolomic features that were induced by the infection. Then, components 3 and 4 gather the metabolomic features that were measured in the non-inoculated but depleted by the disease. Finally, metabolomic features that are constitutively accumulated in stem tissues, and displaying weak regulations by the infection, are represented by component 5.

Finally, an analysis of the Receiver Operating Characteristic (ROC) identified 384 metabolomic features (m/z) that most strongly differentiated the infected group and the control group. 197 m/z were found at higher levels in the infected samples compared

Ph D. manuscript | Anani Amegan MISSINOU

to the controls. 187 m/z were found at lower levels in the infected samples compared to the control group. These selected metabolomic features were then analyzed by the METASPACE plate- form (Palmer et al., 2016) dedicated to the automated annotation of imaging mass-spectrometry data, using a false discovery rate (FDR) below 10%. This analysis was based on the SWISSLIPID, LIPIDMAP, ChEBI public databases, and on our in-house *Brassica*-focused chemical database, generated from the scientific literature. Based on the excellent mass resolution (error close to 3 ppm, or 0.001 m/z), and taking account of possible adducts, this approach allowed the putative annotation for 18 m/z among the 197 chemical features induced by the infection (with multiple possible stereoisomers for each) and 18 m/z were putatively annotated from the 187 m/z features that were depleted by the infection. In the example given in Figure 2. C, the metabolomic feature m/z 471.089, which was induced by the fungus, can be proposed as a sodium adduct of Kaempferol hexose.



Figure 3: The common response of resistant and susceptible *Brassica napus* genotypes to phoma infection. (A) The three steps of the MALDI-MSI data analysis pipeline are implemented in the

SCiLS tool. **(B)** The metabolic response was common to the two genotypes (R: resistant, S: susceptible) 14 days after infection with the JN2 strain of *L. maculans*. The arrows represent the point of infection/inoculation. Spatial segmentation reveals areas of interest in the stem that respond to infection. pLSA reveals a pattern of metabolic responses to infection and depletion of metabolites) clearly distinguishable from constitutive metabolism. **(C)** Analysis of the ROC curve identified the metabolic features (m/z) that differentiated the infected group from the control using an AUC (area under the curve) threshold of 0.75.

3.2.2. Genotype-specific metabolic responses to *L. maculans*

To determine the differences between metabolic responses of susceptible and resistant genotypes to *L. maculans* infection, ROC analyses of the metabolic features induced or depleted by the infection were performed separately for 'Darmor' and 'Bristol'. The four feature lists obtained were compared to each other. Most of the depleted chemical traits were common to 'Darmor' and 'Bristol'. Many induced metabolic features were also common to the two genotypes (31 accumulated features in 'Bristol' are included in the 224 accumulated features in 'Darmor'). However, some accumulated metabolic features were specific to each genotype (Figure 4). Unfortunately, we have not yet been able to annotate these features. In our future work, MS/MS fragmentation may help to explore the putative chemical structure of those compounds.



Figure 4: The genotype-specific metabolic response to *Leptosphaeria maculans.* The images come from a spatial segmentation between the resistant genotype (A) and the susceptible genotype (B). The arrows indicate the inoculation point. The figures represent the number of chemical characteristics resulting from the different comparisons of the metabolic profiles by analyzing the ROC curve with an AUC threshold of 0.75. The top 10 features in each comparison are displayed.



Figure 5: Molecular maps of glucosinolates in healthy stems of *B. napus* (Darmor). Met-derived aliphatic-GLSs: gluconapin, glucobrassicanapin, (epi)progoitrin, gluconasturtiin, glucoalyssin, Trp-derived indolic-GLSs: glucobrassicin.



Figure 6: Segmentation was carried out individually on each biological replicate per experimental condition. Tissular zones with spectral similarity are coded in the same arbitrary colors within the same section, but not between sections. (A) Susceptible ('Bristol') and (B) resistant ('Darmor'), control at the top and 14dpi after JN2 inoculation at the bottom. Positive mode and negative mode MSI analysis are displayed.

4. Conclusion and perspectives

MALDI-MSI technology has given us access to hundreds of biomolecules, allowing us to better understand the metabolic responses to blackleg disease. We offer a view of the commonalities and differences in metabolic responses of the genotypes.

At the beginning of this project, many methodological aspects had to be solved, including the preparation of samples and post-analytic data treatment. For this reason, the initial project included only one biological replicate. Based on the results described

above, we decided to engage the whole experiment with three biological replicates. The resulting samples have been recently analyzed in both positive and negative modes. The analysis of this new dataset is still in progress, but a few aspects of the results can be given here:

First, the use of the negative mode allowed to locate of a series of glucosinolates, namely gluconapin, glucobrassicanapin, (epi)progoitrin, gluconasturtiin, glucoalyssin (aliphatic GLS derived from methionine), glucobrassicin (indolic GLS derived from tryptophan) (Figure 5). Interestingly, the spatial metabolomics approach highlighted that these GLSs were mostly accumulated in the vascular parts of the stem (Figure 6).

Finally, contrasting with our previous metabolomic works for which every independent biological replicate was made from pooled samples from 6 to 8 individual plants, in the present work our biological replicates consist of three unique individual plants. Then, we will have to cope with inevitable biological variations between these three samples, and then it will soon be necessary to work on the definition of adequate statistical analyses to cope with this possible problem.

References

Abuyusuf M, Robin AHK, Kim HT, Islam MR, Park JI, Nou IS (2018) Altered glucosinolate profiles and expression of glucosinolate biosynthesis genes in ringspot-resistant and susceptible cabbage lines. Int J Mol Sci. doi: 10.3390/ijms19092833

Barbosa EA, Bonfim MF, Bloch C, Engler G, Rocha T, de Almeida Engler J (2018) Imaging Mass Spectrometry of Endogenous Polypeptides and Secondary Metabolites from Galls Induced by Root-Knot Nematodes in Tomato Roots. Mol Plant-Microbe Interact 31: 1048–1059

Bhandari DR, Wang Q, Li B, Friedt W, Römpp A, Spengler B, Gottwald S (2018) Histology-guided high-resolution AP-SMALDI mass spectrometry imaging of wheat-*Fusarium graminearum* interaction at the root–shoot junction. Plant Methods 14: 103

Bousset L, Ermel M, Lebreton L (2018) The full life cycle of *Leptosphaeria maculans* completed on inoculated oilseed rape incubated under controlled conditions. Plant Pathol 67: 1321–1328

Caprioli RM, Farmer TB, Gile J (1997) Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. Anal Chem 69: 4751–4760

Delourme R, Bousset L, Ermel M, Duffé P, Besnard AL, Marquer B, Fudal I, Linglin J, Chadœuf J, Brun H (2014) Quantitative resistance affects the speed of frequency increase but not the diversity of the virulence alleles overcoming a major resistance gene to *Leptosphaeria maculans* in oilseed rape. Infect Genet Evol 27: 490–499

Dong Y, Li B, Malitsky S, Rogachev I, Aharoni A, Kaftan F, Svatoš A, Franceschi P (2016) Sample Preparation for Mass Spectrometry Imaging of Plant Tissues: A Review. Front Plant Sci 7: 60

Dukhnytskyi B (2019) World agricultural production. Ekon APK 59–65

Magali Ermel M, Brun H, Delourme R, Bousset L, (2021) Caractériser l'agressivité des souches de *Leptosphaeria maculans* par un test sur pétiole en conditions contrôlées. HAL ld : hal-03288140, version 1

Fitt BDL, Brun H, Barbetti MJ, Rimmer SR (2006) World-wide importance of phoma stem canker (*Leptosphaeria maculans* and *L. biglobosa*) on oilseed Rape (*Brassica napus*). Eur J Plant Pathol 114: 3–15

Hanselmann M, Kirchner M, Renard BY, Amstalden ER, Glunde K, Heeren RMA, Hamprecht FA (2008) Concise representation of mass spectrometry images by probabilistic latent semantic analysis. Anal Chem 80: 9649–58

Hofmann T, Thomas (1999) Probabilistic latent semantic indexing. Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '99. ACM Press, New York, New York, USA, pp 50–57

Howlett BJ, Idnurm A, Pedras MSC (2001) *Leptosphaeria maculans*, the causal agent of blackleg disease of *Brassicas*. Fungal Genet Biol 33: 1–14

Klein AP, Sattely ES (2017) Biosynthesis of cabbage phytoalexins from indole glucosinolate. Proc Natl Acad Sci U S A 114: 1910–1915

Klein AT, Yagnik GB, Hohenstein JD, Ji Z, Zi J, Reichert MD, Macintosh GC, Yang B, Peters RJ, Vela J, et al (2015) Investigation of the chemical interface in the soybean-aphid and rice-bacteria interactions using maldi-mass spectrometry imaging. Anal Chem 87: 5294–5301

Kusari S, Lamshöft M, Kusari P, Gottfried S, Zühlke S, Louven K, Hentschel U, Kayser O, Spiteller M (2014) Endophytes Are Hidden Producers of Maytansine in *Putterlickia* Roots. J Nat Prod 77: 2577–2584

Levrel A, Lemoine J, Ermel M, Abu-Ahmad Y, Gravot A, Delourme R (2022) Differential growth of *Leptosphaeria maculans* in the stem of susceptible and partially resistant oilseed rape (*Brassica napus* L.) genotypes. https://doi.org/101080/0706066120222031300. doi: 10.1080/07060661.2022.2031300

Lilian ST C, Daiane G R, Eder A B, Luciano P S, Angela M (2021) MALDI-MSI method for the detection of large biomolecules in plant leaf tissue. J Plant Sci Phytopathol 5: 058–061

McGee DC, Emmett RW (1977) Black leg (*Leptosphaeria maculans* (Desm.) Ces. et de Not.) of rapeseed in Victoria: Crop losses and factors which affect disease severity. Aust J Agric Res 28: 47–51

Missinou AA, Carvalho JF De, Marnet N, Delhaye T, Hamzaoui O, Sayed DA, Guitton Y, Lebreton L, Langrume C, Laperche A, et al (2022) Identification and Quantification of Glucosinolates and Phenolics in a Large Panel of *Brassica napus* Highlight Valuable Genetic Resources for Chemical Ecology and Breeding. J Agric Food Chem. doi: 10.1021/acs.jafc.1c08118

Palmer A, Phapale P, Chernyavsky I, Lavigne R, Fay D, Tarasov A, Kovalev V, Fuchser J, Nikolenko S, Pineau C, et al (2016) FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. Nat Methods 14: 57–60

Pedras MSC, Yaya EE (2010) Phytoalexins from *Brassicaceae*: News from the front. Phytochemistry 71: 1191–1197

Ph D. manuscript | Anani Amegan MISSINOU

Pilet-Nayel M-L, Moury B, Caffier V, Montarry J, Kerlan M-C, Fournet S, Durel C-E, Delourme R (2017) Quantitative Resistance to Plant Pathogens in Pyramiding Strategies for Durable Crop Protection. Front Plant Sci 8: 1838

Qin L, Zhang Y, Liu Y, He H, Han M, Li Y, Zeng M, Wang X (2018) Recent advances in matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI) for in situ analysis of endogenous molecules in plants. Phytochem Anal 29: 351–364

Rouxel T, Balesdent MH (2005) The stem canker (blackleg) fungus, *Leptosphaeria maculans*, enters the genomic era. Mol Plant Pathol 6: 225–241

Santos M, Fernandes D, Rossi M, Fátima M, Batista J, Cezar P, Brentan D, Peporine N, Alves S, Carvalho D, et al (2015) Phytochemistry Quantification and localization of hesperidin and rutin in Citrus sinensis grafted on *C. limonia* after *Xylella fastidiosa* infection by HPLC-UV and MALDI imaging mass spectrometry. Phytochemistry 115: 161–170

Seneviratne HK, Dalisay DS, Kim KW, Moinuddin SGA, Yang H, Hartshorn CM, Davin LB, Lewis NG (2015) Non-host disease resistance response in pea (Pisum sativum) pods: Biochemical function of DRR206 and phytoalexin pathway localization. Phytochemistry 113: 140–148

Shroff R, Schramm K, Jeschke V, Nemes P, Vertes A, Gershenzon J, Svatoš A (2015) Quantification of plant surface metabolites by matrix-assisted laser desorption-ionization mass spectrometry imaging: glucosinolates on *Arabidopsis thaliana* leaves. Plant J 81: 961–972

Shroff R, Vergara F, Muck A, Svatos A, Gershenzon J (2008) Nonuniform distribution of glucosinolates in *Arabidopsis thaliana* leaves has important consequences for plant defense. Proc Natl Acad Sci U S A 105: 6196–201

Wang W, Kusari S, Sezgin S, Lamshöft M, Kusari P, Kayser O, Spiteller M (2015) Hexacyclopeptides secreted by an endophytic fungus Fusarium solani N06 act as crosstalk molecules in *Narcissus tazetta*. 7651–7662

Wang Y, Strelkov SE, Hwang SF (2020) Yield losses in canola in response to blackleg disease. Can J Plant Sci 100: 488–494

West JS, Kharbanda PD, Barbetti MJ, Fitt BDL (2001) Epidemiology and management of *Leptosphaeria maculans* (phoma stem canker) on oilseed rape in Australia, Canada and Europe. Plant Pathol 50: 10–27

Ye H, Gemperline E, Venkateshwaran M, Chen R, Delaux P-M, Howes-Podoll M, Ané J-M, Li L (2013) MALDI mass spectrometry-assisted molecular imaging of metabolites during nitrogen fixation in the *Medicago truncatula* - Sinorhizobium meliloti symbiosis. Plant J 75: 130–145

Chapitre IV : Discussion générale et Perspectives

Comme exposé dans l'introduction bibliographique de ce manuscrit, avant le début de cette thèse en 2018, la recherche portant sur le métabolisme spécialisé de *Brassica napus* recélait un vaste territoire inexploré. Dans un contexte où la résistance génétique aux bio-agresseurs constitue un levier majeur des scénarios de transition agro-écologique, il est nécessaire de consolider notre compréhension des mécanismes biologiques impliqués dans les interactions entre le colza et son environnement biotique. Chez cette espèce, il semble aujourd'hui nécessaire d'approfondir les investigations sur la diversité des défenses chimiques constitutives (phytoanticipines), les déterminants génétiques contrôlant leur biosynthèse, et comment la plante mobilise son arsenal de défense chimique (constitutive et induite) en réponses aux contraintes biotiques. C'est dans ce contexte que s'est inscrit mon projet de thèse.

IV. A. Discussion et conclusion générale

L'objectif de cette thèse était de caractériser le métabolisme spécialisé de *B. napus*, avec un "fil d'Ariane" à deux sous-objectifs : étudier dans un panel d'accessions la diversité (1) et le contrôle génétique des contenus en phytoanticipines dans les racines et les feuilles, et caractériser la régulation (2) des phytoanticipines et des phytoalexines en réponse à l'infection fongique par *L. maculans* dans les tiges de *B. napus*.

Au cours de mes recherches, j'ai participé au développement de plusieurs stratégies pour atteindre cet objectif. Dans la première partie de mon travail, nous avons tiré parti d'outils analytiques (spectrométrie de masse et résonnance magnétique nucléaire)

Ph D. manuscript | Anani Amegan MISSINOU

performants afin de générer les meilleurs jeux de données possibles. Cette étape, bien qu'étant assez contraignante, est essentielle pour la suite.

Certes, la recherche sur le métabolisme spécialisé des plantes est traversée actuellement par une révolution informatique, ce qui offre une grande diversité d'outil permettant, d'extraire automatiquement un grand nombre d'information autour des données métabolomiques. Cependant, en raison d'un certain nombre de contraintes techniques, en particulier la nécessaire exploitation d'un jeu de données pour lequel les données de fragmentation MS2 ont été acquises en basse résolution, et dans un contexte de consolidation de nos capacités à interpréter des données de spectrométrie de masse, nous avons dans un premier temps délibérément privilégié une exploration manuelle des données. Cet effort a permis d'annoter 32 composés phénoliques et 36 glucosinolates, dont les schémas de fragmentation ont été proposé dans les données supplémentaires du premier article décrivant la diversité chimique des composés foliaires et racinaires de *Brassica*.

Brassica napus, une "corne d'abondance" de métabolites spécialisés anciens et nouveaux

Mi-2021, la littérature sur les GLSs répertoriait 139 glucosinolates chez les Brassicales (dont 90 structures caractérisées de façon concluantes et 49 structures partiellement caractérisées) (Blažević et al., 2020; Montaut et al., 2020; Trabelcy et al., 2021). Rares sont les études qui rapportent plus de 20 GLSs chez une même espèce. Chez *B. napus*, pendant longtemps, les études se sont concentrées sur les GLSs contenus

dans les graines qui sont en majorité des aliphatiques dérivés de la méthionine. Cependant, l'article 1 du chapitre 2 de ce document révèle 36 GLSs dont certains ont été rarement rapportés chez B. napus (Missinou et al., 2022). Ceux-ci représentent environ le quart de l'ensemble des GLSs répertoriés dans les synthèses bibliographiques les plus récentes (Blažević et al., 2020; Montaut et al., 2020; Trabelcy et al., 2021). Certaines des structures GLSs que nous avons observées chez le colza ont été habituellement attribuées de manière spécifique aux genres taxonomiques, à savoir la p-hydroxygluconasturtiine chez Arabis soyeri; l'epiglucobarbarine et la glucobarbarine utilisée pour des affiliations taxonomique des types P et G de Barbarea vulgaris; la m-hydroxygluconasturtiine commune au genre Arabis et Barbarea; la 4methoxyglucobrassicine chez Barbarea vulgaris; l'isopropyl-GLS chez la Cardamine diphylla; la glucocochlearine chez Cochlearia officinalis; et 3-methylpentyl-GLS chez Cardamine pratensis. Nous avons montré que B. napus synthétise en plus de ces formes relativement rares de GLSs, un nombre important et stimulant d'isomères d'alkyl-GLSs (2 isomères hydroxy-C4-alkyl-GLSs, 3 isomères de C5- et C6-alkyl-GLSs) de structures inconnues (Missinou et al., 2022).

IV.A.1. Les GLSs de Brassica issues de voies de biosynthèses anciennes

D'un point de vue évolutif, les voies les plus anciennes de biosynthèse des GLS correspondent à celles mises en place chez les espèces basales dans la phylogénie des *Brassicales*. Il s'agit des familles telles que les *Akaniaceae*, *Tropaelaceae*, *Moringaceae*, *Caricaceae*, *Setchellanthaceae*, *et Limnanthaceae* (Edger et al., 2018). Par exemple, les formes branchées de GLS, telles que la 1-methylpropyl-GLS (glucoconringianine) et la 1-methylpropyl-

GLS (glucocochlearine) ont été rapportées chez *Limnanthaceae* (Agerbirk et al., 2022). Ces voies de biosynthèse sont considérées comme antérieures à celle des GLS indoliques, et d'après les résultats de mon travail, elles semblent avoir été conservées chez *B. napus* et chez ancêtres des *Brassica*.

IV.A.2. Les GLSs issues des voies de biosynthèses plus récentes

La diversité des GLSs décrite dans notre travail contient un certain nombre de GLSs issue d'innovation biosynthétique à partir des GLSs ancestraux. En effet, une grande partie des GLSs de *B. napus* sont issus d'élongation de la chaine R primaire des GLSs contenus dans des familles de plantes basales la phylogénie des Brassicales, à savoir l'homométhionine, l'homophénylamine et l'homotyrosine.

Il est admis que la structure de GLS la plus ancienne est la glucotropaeolin et ces dérivés hydroxylés et méthoxylés (Oscar et al., 2006; Edger et al., 2018). Bien que le benzyl-GLS et l'hydroxybenzyl-GLS aient été retrouvés chez *Sinapis alba,* qui est relativement proche du genre *Brassica*, aucune étude ne rapporte la présence de ces composés chez le colza. Cette dernière semble avoir été perdu ces GLSs en faveur de la gluconasturitine et de ses dérivés hydroxylés (homosinalbine, epiglucobarbarine, glucobarbarine, et m-hydroxygluconasturtiine).

Chez *Barbarea vulgaris*, il existe un polymorphisme en stéréochimie de la 2hydroxylation de la gluconasturtiine (l'epiglucobarbarine et la glucobarbarine), déterminé par la variation allélique sur deux homologues de type FMO_{GS-OX} d'*Arabidopsis* (SHO : S-HYDROXYLATION) et RHO : R-HYDROXYLATION) qui déterminerait l'accumulation respective des épimères dans le chémotypes G et P de l'espèce en Europe de l'Est (Liu et al., 2019). Très curieusement, les dérivés hydroxylés de GLSs, à l'exception de p/m-hydroxyglucobarbarine et epi/progoitrine, étaient associés à des QTLs localisés sur des chromosomes A03 ou C02. En considérant que le gène CYP81F2 réalisant l'hydroxylation de la glucobrassicine en 1hydroxyglucobrassicine nécessaire à formation de formation la 4methoxyglucobrassicine est localisé sur C02 chez B. oleracea (Sotelo et al., 2014), d'une façon spéculative, nous pouvons proposer que ces régions sur les chromosomes A03 et C02 contiendraient des loci codant des enzymes de type FMOGSox, un contrôle plus général des oxydations réalisées sur des résidus C-, N-, Sprécurseurs de glucosinolates.

IV.A.3. Redondance structurelle dans la biodiversité GLSs

Nous avons déterminé un grand nombre d'isomères de glucosinolates (hydroxylés, méthoxylés, à chaine R) et des phénoliques (hexosylés et acylés). Ceci suggère, une capacité du colza à co-synthétiser différents composés isomérique hydroxylés, méthoxylés, à chaîne carbonée identique, hexosylés et phénylacylés. Cette co-occurrence de ces formes isomériques peut potentiellement être, d'une part le résultat de rétention des gènes dupliqués impliqués dans la biosynthèse de composés spécialisés chez *B. napus* (Hofberger et al., 2013), mais d'autre part, le fruit de séparation des enzymes homologues dans différent compartiments cellulaire, comme cela a été démontré pour les MAMs³⁷ d'*Arabidopsis*.

³⁷ Les enzymes MAM catalysent la réaction de condensation qui étend la chaîne carbonée dans les précurseurs GSL dérivés d'acides aminés

Ph D. manuscript | Anani Amegan MISSINOU

En conclusion, nous avons avancé dans la caractérisation du métabolisme spécialisé du colza, et fourni un premier jalon sur l'impact de la sélection des GLSs graines sur la diversité génétique aux allèles impliqués dans le contrôle génétique de la teneur en GLSs (feuilles/racines), et la caractérisation des réponses métaboliques à l'infection. L'ensemble de mes travaux ouvrent la voie à d'autres thématiques, d'autres thèses potentielles, dans des domaines liés à l'amélioration des plantes et l'écologie chimique.

IV.B. Perspectives

IV.B.1. Étendre l'inventaire des métabolites spécialisés chez B. napus

En plus des GLSs rarement rapportés que nous avons identifiés chez le colza, nous avons repéré après publication, quatre autres glucosinolates, dont deux ont été provisoirement annotés comme étant des isomères de C5-alkyl-GLSs potentiellement équivalent à ceux rapportés chez *B. rapa* (Taveira et al., 2009), un isomère de dihydroxyphenylalkyl-GLS potentiellement équivalent à l'un des deux isomères signalés chez *Barbarea vulgaris* (Agerbirk et al., 2015), et un dérivé dihydroxylés de l'indol-3-ylméthyl-GLS (dihydroxyglucobrassicine), qui cependant n'a jamais été reporté dans aucune espèce (Missinou, communication personnelle des données non publiés). En outre, nous détectons à un très faible niveau, des masses équivalentes à des dérivés dihydroxybenzyl-GLS, le cinnamoyl-eptyl-GLS et le sinapoyl-hydroxyglucobrassicine déjà signalés chez le chou (Capriotti et al., 2018), dans les racines et feuilles de notre panel.

Certains auteurs ont rapporté 51 GLSs chez *Brassica oleracea L. var. botrytis*, dont six dérivés hydroxylés de l'indol-3-ylméthyl-GSLs, trois méthylsulfinylbutyl-GLSs, deux méthylsulfinylpropyl-GLSs, des dérivés sinapoyl, cinnamoyl-, coumaroyl-GLSs entre autres (Capriotti et al., 2018). Face à ce nombre croissant de GLSs retrouvés chez *B. napus* et au fait que *Brassica oleracea L. var. botrytis* présente d'autres formes inhabituelles de GLS, il se profile que le nombre et la diversité structurelle des GLSs chez *B. napus* excède au moins les 44 GLSs. De futures optimisations analytiques seront nécessaires pour fournir les preuves de la présence/absence de ces GLSs atypiques chez le colza. En outre, près d'une quarantaine de composés phénoliques détectés n'ont pas été publiés, en raison de leur insuffisante résolution chromatographique qui empêche leur quantification par intégration des pics UV. Certaines d'entre eux ont été provisoirement annotés comme étant des dérivés phénoliques complexés à de l'acide malique, polyhexosylés et polyacylés.

De façon très remarquable, certaines études avaient rapporté la présence de la glucotropaeoline chez le chou (*Brassica oleracea L. var. gernrnifera*) (Heaney and Fenwick, 1980). Il serait intéressant de rechercher ce composé dans les variétés anciennes de *B. napus* (WOSR++).

Aussi, il serait avantageux de tester des combinaisons de solvant binaire ou tertiaire, afin d'élargir le prisme du rendement d'extraction et des familles des molécules isolées des matrices cellules, comme cela a été prouvé dans chez d'autre modèles (Silva et al., 2020) et d'optimiser nos conditions analytiques, en testant par exemple les colonnes chromatographiques Kinetex XB-C18 and Kinetex Biphenyl (Capriotti et al., 2018). Enfin le fractionnement bio-guidé à partir des plantes infectées permettrait de
cibler des fractions biologiquement actives au moyen de tests d'inhibition de croissance des agents pathogènes en culture. L'activité des métabolites **R13_M320** et **R18_M368** pourrait être testée de cette façon.

IV.B.2. Profilage métabolomique non ciblé de *B. napus* et annotation des métabolites à l'aide de réseaux moléculaires

La spectrométrie de masse (MS) couplée à des méthodes de séparation (LC et GC) est couramment appliquée pour le profilage différentiel et quantitatif. Il serait donc judicieux d'appliquer ces méthodes d'analyse utilisant plusieurs plateformes analytiques (LC-MS et GC-MS) sur les échantillons du panel de Brassica. Le défi le plus important de ces approches dites de profilage métabolomique non ciblé est l'efficacité des méthodes de traitement et d'extraction des informations pertinentes. Avec l'avènement des méthodes de traitement des données métabolomiques non ciblées, il est possible d'extraire, pour un grand nombre de molécules, les premiers jalons à l'annotation. Il existe plusieurs outils dont les plus célèbres sont XCMS (Mo et al., 2004), MS-DIAL (Tsugawa et al., 2020), mzMine2 (Pluskal et al., 2010). A l'heure actuelle, ces outils contiennent des étapes de traitement des données pour les réseaux moléculaires basés sur les fonctionnalités (FBMN) (Nothias et al., 2020) et des réseaux d'identité ionique (IIN). IIN analyse les corrélations de formes de pic LC-MS et corrèle différents ions, y compris les isopotologues, les adduits, les multimères et la fragmentation dans la source (Schmid et al., 2021) et fournissent des sorties à GNPS (Products Social Molecular Networking : http://gnps.ucsd.edu) (Wang et al., 2016). Le serveur GNPS permet de réaliser la déréplication des variables (correspondances

approximatives aux spectres de molécules apparentées), la fusion des modes d'ionisation et d'annoter les spectres de fragmentation dans les réseaux moléculaires grâce à plusieurs bases de données et des outils tel que SIRIUS (Dührkop et al., 2019), ZODIAC (Ludwig et al., 2020), MS2LDA (Rogers et al., 2019) et de fournir une sortie Cytoscape. Au cours de ma thèse, les échantillons de tige de colza, 14 et 28 jours post-inoculation par *L. maculans (JN2)*, ont été analysés selon les deux modes d'ionisation et une fragmentation systématique a été réalisée sur les ions majoritaires. L'approche mzMine2-GNPS, que je me suis appropriée durant ma thèse, a été initiée sur ces données. Cependant par manque de temps, elle n'est pas présentée dans ce manuscrit.

IV.B.3. Affiner la puissance de détection de gènes candidats

Les études d'association à l'échelle du génome (GWAS) ont identifié des associations de plusieurs QTL contenant des polymorphismes mononucléotidiques (SNP) associés aux métabolites spécialisés secondaires de colza. Cependant, la précision de nos analyses GWAS reste à améliorer pour identifier des gènes candidats afin de valider leur implication par l'étude de variants fonctionnels et de caractériser les mécanismes sous-jacents aux QTLs métaboliques. Les intervalles de confiance des QTL correspondent à des régions génomiques incluant de nombreux gènes. Dans nos analyses mGWAS, les pourcentages de variation phénotypique expliquée par les QTLs individuels varient entre 3 et 37%, ce qui correspond plus ou moins forte de la variation génétique de chaque trait. De plus, certains variants SNP peuvent ne pas avoir été capturés dans nos analyses mGWAS, du fait de l'existence d'allèles rares dans le panel colza utilisé. Par exemple, nous n'avons pas détecté de QTL qui soit spécifiquement associé à l'épimérisation 2-hydroxy-3-butenyl-glucosinolate

(progoitrine et epiprogoitrine) ni à celle des hydroxyphenylethyl-glucosinolates (epiglucobrabarine et glucobarbarine). Face à ces limites d'identification des associations génétiques, il existe différentes options pour améliorer la puissance et précision de détection des gènes candidats.

Premièrement, pour exploiter les variants rares spécifiques à certains cultivars, difficilement détectable par GWAS, il serait intéressant d'analyser la génétique des traits métaboliques par des approches de génétique de liaison dans des populations biparentales ou de génétique d'association dans des populations issues de croisements multiparentaux, comme des populations MAGIC (multiparent advanced-generation intercross population) (Cavanagh et al. 2008, Holland 2015) en ciblant les parents selon les contrastes visés.

Deuxièmement, les données de séquençage du génome entier (WGS) peuvent être utilisées pour densifier en marqueurs certaines régions et mieux capturer les variants rares et les variations structurales. L'obtention des séquences génomiques des accessions du panel *Brassica* serait sans doute possible dans les années qui viennent, en raison de la chute vertigineuse du coût du séquençage ADN qui a baissé d'un facteur de 1.000.000 ces 15 dernières années.

IV.B.4. Validation fonctionnelle des gènes candidats

Le développement du système CRISPR-Cas9 pour l'édition du génome dans les cellules de plantes, a révolutionné le domaine de la recherche en génomique fonctionnelle, permettant une validation relativement rapide des gènes candidats suspects de rôle dans la biosynthèse de métabolites spécialisées (Qian et al., 2017; Fan et al., 2020). Il serait aussi intéressant de réaliser des mutations par

CRISPR/cas9, technologie qui peut être mise en œuvre chez le colza mais qui est parfois difficile en raison de la nature dupliquée du génome du colza. Par ailleurs, les protocoles de transformation génétique actuels ne fonctionnent que pour un nombre restreint de génotypes.

IV.B.5. Investiguer les clusters de gènes biosynthétiques (BGCs) et leur évolution chez les *Brassicas*

La découverte de métabolites spécialisées connait une révolution avec l'avènement dédiés aux outils informatiques permettant de prédire les BGCs de façon automatique dans des ressources génomiques de plantes. Il existe plusieurs outils bioinformatiques, dont PhytoClust (Töpfer et al., 2017), PlantClusterFinder (Schläpfer et al., 2017). Lors de ma thèse, j'ai exploré cette piste dans le but de déterminé les BGCs de B. napus. Nous avons prédit 104 BGCs dans le génome de B. napus (v10) à l'aide de l'outil plantiSMASH (Kautsar et al., 2017; Kautsar and Medema, 2018). Malheureusement, l'analyse de leur expression avec les données d'expression transcriptomique du projet Metaphor a révélé qu'un faible nombre était régulé par l'infection par le phoma, à l'exception d'un cluster d'une taille de 80,39 kb. Ce cluster contenait trois gènes biosynthétiques non-homologues, A10p30340.1_BnaDAR (oxydoréductase), A10p30430.1_BnaDAR (CYP71B20), et A10p30490.1_BnaDAR (UDP-glucosyltransférase), régulé à la hausse par l'infection des souches JN2 et NZT4 dans les génotypes 'Darmor-bzh', 'Darmor' et 'Bristol'. Certes ces résultats ne répondent pas à la question de la résistance de 'Darmor/Darmor-bzh' par rapport à 'Bristol', mais fournissent potentiellement des gènes candidats impliqués dans la synthèse de phytoanticipines potentiellement inconnus. Par manque de temps, je n'ai pas investigué les métabolites possiblement produits par ce cluster. Non seulement il

serait intéressant de mieux caractériser les variations génomiques sur ces clusters en exploitant les séquences génomiques exploitables, mais aussi, de les analyser avec d'autres jeux d'expression en condition de stress biotique et abiotique.

La liaison des informations entre les ensembles de données (génomique, génétique, et métabolomique) est utile, car elle permet l'annotation structurelle et fonctionnelle. L'esprit à l'heure actuelle est d'explorer comment évolue les BGCs dans différentes population (Medema et al., 2015; Navarro-Muñoz et al., 2020) et de les connecter aux données de fragmentation de la spectrométrie de masse non-ciblée (van der Hooft et al., 2020; Leao et al., 2021).

References

- Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Altaf-Ul-Amin M, Darusman LK, et al (2012) KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol 53: 1–12
- Agerbirk N, Olsen CE, Heimes C, Christensen S, Bak S, Hauser TP (2015) Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer *Barbarea vulgaris*. Phytochemistry **115**: 130–142
- Agerbirk N, Pattison DI, Mandáková T, Lysak MA, Montaut S, Staerk D (2022) Ancient Biosyntheses in an Oil Crop: Glucosinolate Profiles in Limnanthes alba and Its Relatives (Limnanthaceae, Brassicales). J Agric Food Chem. doi: 10.1021/acs.jafc.1c07299
- Ahmed AS, Sánchez CP, Candela ME (2000) Evaluation of induction of systemic resistance in pepper plants (*Capsicum annuum*) to *Phytophthora capsici* using *Trichoderma harzianum* and its relation with capsidiol accumulation. Eur J Plant Pathol **106**: 817–824
- Ahuja I, Kissen R, Bones AM (2012) Phytoalexins in defense against pathogens. Trends Plant Sci 17: 73–90
- An H, Qi X, Gaynor ML, Hao Y, Gebken SC, Mabry ME, McAlvay AC, Teakle GR, Conant GC, Barker MS, et al (2019) Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. Nat Commun. doi: 10.1038/s41467-019-10757-1
- Anarat-Cappillino G, Sattely ES (2014) The chemical logic of plant natural product biosynthesis. Curr Opin Plant Biol 19: 51–58
- Aubertot JN, West JS, Bousset-Vaslin L, Salam MU, Barbetti MJ, Diggle AJ (2006) Improved resistance management for durable disease control: A case study of phoma stem canker of oilseed rape (*Brassica napus*). Eur J Plant Pathol **114**: 91–106
- Babst BA, Harding SA, Tsai CJ (2010) Biosynthesis of phenolic glycosides from phenylpropanoid and benzenoid precursors in populus. J Chem Ecol **36**: 286–297
- Balalić I, Marjanović-Jeromela A, Crnobarac J, Terzić S, Radić V, Miklič V, Jovičić D (2017) Variabilty of oil and protein content in rapeseed cultivars affected by seeding date. Emirates J Food Agric 29: 404–410
- Barz W, Mackenbrock U (1994) Constitutive and elicitation induced metabolism of isoflavones and pterocarpans in chickpea (*Cicer arietinum*) cell suspension cultures. Plant Cell Tissue Organ Cult **38**: 199–211
- Barz W, Welle Roland (1992) Biosynthesis and Metabolism of Isoflavones and Pterocarpan Phytoalexins in Chickpea, Soybean and Phytopathogenic Fungi. Phenolic Metabolism in Plants. doi: 10.1007/978-1-4615-3430-3_5
- Bayer PE, Hurgobin B, Golicz AA, Chan CKK, Yuan Y, Lee HT, Renton M, Meng J, Li R, Long Y, et al (2017) Assembly and comparison of two closely related *Brassica napus* genomes. Plant Biotechnol J **15**: 1602–1610
- Bednarek P, Piślewska-Bednarek M, Svatoš A, Schneider B, Doubský J, Mansurova M, Humphry M, Consonni
 C, Panstruga R, Sanchez-Vallet A, et al (2009) A Glucosinolate Metabolism Pathway in Living Plant Cells
 Mediates Broad-Spectrum Antifungal Defense. Science (80-) 323: 101–106
- Belser C, Istace B, Denis E, Dubarry M, Baurens FC, Falentin C, Genete M, Berrabah W, Chèvre AM, Delourme
 R, et al (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat Plants 4: 879–887

Bentley R (1999) Secondary metabolite biosynthesis: the first century. Crit Rev Biotechnol 19:1-40

Beran F, Köllner TG, Gershenzon J, Tholl D (2019) Chemical convergence between plants and insects:

biosynthetic origins and functions of common secondary metabolites. New Phytol 223: 52-67

- Blažević I, Montaut S, Burčul F, Olsen CE, Burow M, Rollin P, Agerbirk N (2020) Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. Phytochemistry 169: 112100
- Bolton MD (2009) Primary metabolism and plant defense-fuel for the fire. 22: 487-497
- Bora KS, Sharma A (2011) The genus Artemisia: A comprehensive review. Pharm Biol 49: 101-109
- Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12: 2383–2393
- Böttcher C, Chapman A, Fellermeier F, Choudhary M, Scheel D, Glawischnig E (2014) The Biosynthetic Pathway of Indole-3-Carbaldehyde and Indole-3-Carboxylic Acid Derivatives in Arabidopsis. Plant Physiol **165**: 841–853
- **Bousset L, Ermel M, Lebreton L** (2018) The full life cycle of *Leptosphaeria maculans* completed on inoculated oilseed rape incubated under controlled conditions. Plant Pathol **67**: 1321–1328
- **Boyd RS** (2012) Plant defense using toxic inorganic ions: Conceptual models of the defensive enhancement and joint effects hypotheses. Plant Sci **195**: 88–95
- Brockington SF, Walker RH, Glover BJ, Soltis PS, Soltis DE (2011) Complex pigment evolution in the Caryophyllales. New Phytol 190: 854–864
- de Bruijn WJC, Gruppen H, Vincken JP (2018) Structure and biosynthesis of benzoxazinoids: Plant defence metabolites with potential as antimicrobial scaffolds. Phytochemistry 155: 233–243
- De Bruijn WJC, Vincken JP, Duran K, Gruppen H (2016) Mass Spectrometric Characterization of Benzoxazinoid Glycosides from Rhizopus-Elicited Wheat (*Triticum aestivum*) Seedlings. J Agric Food Chem **64**: 6267–6276
- Brun H, Chèvre AM, Fitt BD, Powers S, Besnard AL, Ermel M, Huteau V, Marquer B, Eber F, Renard M, et al (2010) Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. New Phytol **185**: 285–299
- Brun H, Levivier S, Somda I, Ruer D, Renard M, Chevre AM (2000) A field method for evaluating the potential durability of new resistance sources: Application to the *Leptosphaeria maculans-Brassica napus* pathosystem. Phytopathology **90**: 961–966
- **Bu'lock JD** (1973) Comparative and Functional Aspects of the Isoprene Pathway in Fungi. Pure Appl Chem **34**: 435–462
- **Cambier V, Hance T, De Hoffmann E** (1999) Non-injured maize contains several 1,4-benzoxazin-3-one related compounds but only as glucoconjugates. Phytochem Anal **10**: 119–126
- Cammue BPA, Thevissen K, Hendriks M, Eggermont K, Goderis IJ, Proost P, Van Damme J, Osborn RW, Guerbette F, Kader JC, et al (1995) A potent antimicrobial protein from onion seeds showing sequence homology to plant lipid transfer proteins. Plant Physiol 109: 445–455
- Campos ML, De Souza CM, De Oliveira KBS, Dias SC, Franco OL (2018) The role of antimicrobial peptides in plant immunity. J Exp Bot 69: 4997–5011
- Capriotti AL, Cavaliere C, La Barbera G, Montone CM, Piovesana S, Zenezini Chiozzi R, Laganà A (2018) Chromatographic column evaluation for the untargeted profiling of glucosinolates in cauliflower by means of ultra-high performance liquid chromatography coupled to high resolution mass spectrometry. Talanta 179: 792–802
- Cardinal-McTeague WM, Sytsma KJ, Hall JC (2016) Biogeography and diversification of Brassicales: A 103 million year tale. Mol Phylogenet Evol 99: 204–224

Carré P, Pouzet A (2014) Rapeseed market, worldwide and in Europe. OCL - Oilseeds fats 21: 1-12

- **Castellarin SD, Di Gaspero G, Marconi R, Nonis A, Peterlunger E, Paillard S, Adam-Blondon AF, Testolin R** (2006) Colour variation in red grapevines (Vitis vinifera L.): Genomic organisation, expression of flavonoid 3'hydroxylase, flavonoid 3',5'-hydroxylase genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. BMC Genomics **7**: 1–17
- **Cavanagh C, Morell M, Mackay I, Powell W** (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215-221
- Chalhoub B et al (2014) Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science (80-) **345**: 950–953
- Charbonnier E, Fugeray-Scarbel A, Lemarié S (2019) Rapeseed: how to value varieties with higher nitrogen use efficiency in France. Ocl 26: 26
- Chen YC, Holmes EC, Rajniak J, Kim JG, Tang S, Fischer CR, Mudgett MB, Sattely ES (2018) N-hydroxy-pipecolic acid is a mobile metabolite that induces systemic disease resistance in *Arabidopsis*. Proc Natl Acad Sci U S A **115**: E4920–E4929
- Chen Z, Agnew JL, Cohen JD, He P, Shan L, Sheen J, Kunkel BN (2007) *Pseudomonas syringae* type III effector AvrRpt2 alters *Arabidopsis thaliana* auxin physiology. Proc Natl Acad Sci U S A **104**: 20131–20136
- Christianson DW (2017) Structural and Chemical Biology of Terpenoid Cyclases. Chem Rev 117: 11570–11648
- Clay NK, Adio AM, Denoux C, Jander G, Ausubel FM (2009) Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response. Science (80-) 323: 95–101
- Cohen Y, Eyal H, Goldschmidt Z, Sklarz B (1983) A preformed chemical inhibitor of tobacco powdery mildew on leaves of *Nicotiana glutinosa*. Physiol Plant Pathol 22: 143-IN2
- **Cooke J, Leishman MR** (2012) Tradeoffs between foliar silicon and carbon-based defences: Evidence from vegetation communities of contrasting soil types. Oikos **121**: 2052–2060
- **Corwin JA, Kliebenstein DJ** (2017) Quantitative resistance: More than just perception of a pathogen. Plant Cell **29**: 655–665
- Croteau R, Kutchan TM, Norman G. Lewis (2019) Natural Products (Secondary Metabolites). Compr Biotechnol 131–143
- Cui H, Tsuda K, Parker JE (2015) Effector-triggered immunity: From pathogen perception to robust defense. Annu Rev Plant Biol 66: 487–511
- Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411: 826-833
- Davies WKD (2015) The Plant Hormones: Their Nature, Occurrence, and Functions. Geoj Libr 112: 1–16
- **Debieu M, Huard-Chauveau C, Genissel A, Roux F, Roby D** (2016) Quantitative disease resistance to the bacterial pathogen *Xanthomonas campestris* involves an *Arabidopsis* immune receptor pair and a gene of unknown function. Mol Plant Pathol **17**: 510–520
- Delazar A, Byres M, Gibbons S, Kumarasamy Y, Modarresi M, Nahar L, Shoeb M, Satyajit D. Sarker (2004) Iridoid Glycosides from *Eremostachys baissunensis*. J Nat Prod **51**: 991–992
- **Delgoda R, Murray JE** (2017) Evolutionary Perspectives on the Role of Plant Secondary Metabolites. Pharmacogn Fundam Appl Strateg. doi: 10.1016/B978-0-12-802104-0.00007-X
- Delourme R, Bouchereau A, Hubert N, Renard M, Landry BS (1994) Identification of RAPD markers linked to a fertility restorer gene for the *Ogura radish* cytoplasmic male sterility of rapeseed (*Brassica napus L*.). Theor Appl Genet 88: 741–748

- Delourme R, Bousset L, Ermel M, Duffé P, Besnard AL, Marquer B, Fudal I, Linglin J, Chadœuf J, Brun H (2014) Quantitative resistance affects the speed of frequency increase but not the diversity of the virulence alleles overcoming a major resistance gene to *Leptosphaeria maculans* in oilseed rape. Infect Genet Evol **27**: 490– 499
- Delourme R, Chèvre AM, Brun H, Rouxel T, Balesdent MH, Dias JS, Salisbury P, Renard M, Rimmer SR (2006) Major gene and polygenic resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). Eur J Plant Pathol **114**: 41–52
- Denis M, Delourme R, Gourret JP, Mariani C, Renard M (1993) Expression of engineered nuclear male sterility in *Brassica napus*: Genetics, morphology, cytology, and sensitivity to temperature. Plant Physiol **101**: 1295– 1304
- Diener AC, Ausubel FM (2005) Resistance to Fusarium oxysporum 1, a dominant *Arabidopsis* disease-resistance gene, is not race specific. Genetics **171**: 305–321
- Dinkova-Kostova, A.T.; Kostov, R.V (2012). Glucosinolates and isothiocyanates in health and disease. *Trends* Mol. Med., 18, 337–347.
- Dixon RA, Strack D (2003) Phytochemistry meets genome analysis, and beyond. Phytochemistry 62: 815-816
- Drew SW, Demain AL (1977) Effect of primary metabolites on secondary metabolism. Annu Rev Microbiol **31**: 343–356
- Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik A V., Meusel M, Dorrestein PC, Rousu J, Böcker S (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods 16: 299–302
- Dukhnytskyi B (2019) World agricultural production. Ekon APK 59-65
- Durán-Medina Y, Ruiz-Cortés BE, Guerrero-Largo H, Marsch-Martínez N (2021) Specialized metabolism and development: An unexpected friendship. Curr Opin Plant Biol 64: 102142
- Edger PP, Hall JC, Harkess A, Tang M, Coombs J, Mohammadin S, Schranz ME, Xiong Z, Leebens-Mack J, Meyers BC, et al (2018) Brassicales phylogeny inferred from 72 plastid genes: A reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. Am J Bot **105**: 463–469
- **Erb M, Kliebenstein DJ** (2020) Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. Plant Physiol **184**: 39–52
- Eunice Karinho-Betancourt AAA, Halitschke R, Nunez-Farfan J (2015) Phylogenetic correlations among chemical and physical plant defenses change with ontogeny. New Phytol. doi: 10.1111/nph.13300
- Evans N, Baierl A, Semenov MA, Gladders P, Fitt BDL (2008) Range and severity of a plant disease increased by global warming. J R Soc Interface 5: 525–531
- Fabro G, Kovács I, Pavet V, Szabados L, Alvarez ME (2004) Proline accumulation and AtP5CS2 gene activation are induced by plant-pathogen incompatible interactions in *Arabidopsis*. Mol Plant-Microbe Interact 17: 343–350
- Faccio, G. (2020) Plant complexity and cosmetic innovation. *IScience 23*, 101358.
- Fahey JW, Zalcmann AT, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry 56: 5–51
- Fan P, Wang P, Lou YR, Leong BJ, Moore BM, Schenck CA, Combs R, Cao P, Brandizzi F, Shiu SH, et al (2020) Evolution of a plant gene cluster in solanaceae and emergence of metabolic diversity. Elife 9: 1–26
- **Farag MA, Sharaf Eldin MG, Kassem H, Abou El Fetouh M** (2013) Metabolome classification of *Brassica napus L*. organs via UPLC-QTOF-PDA-MS and their anti-oxidant potential. Phytochem Anal **24**: 277–287

- Ferreres F, Valentão P, Llorach R, Pinheiro C, Cardoso L, Pereira JA, Sousa C, Seabra RM, Andrade PB (2005) Phenolic compounds in external leaves of tronchuda cabbage (*Brassica oleracea L. var. costata* DC). J Agric Food Chem 53: 2901–2907
- Van Der Fits L, Memelink J (2000) ORCA3, a Jasmonate- Responsive Transcriptional Regulator of Plant Primary and Secondary Metabolism. Science (80-) 289: 295–297
- Fitt BDL, Brun H, Barbetti MJ, Rimmer SR (2006) World-wide importance of phoma stem canker (*Leptosphaeria maculans* and *L. biglobosa*) on oilseed Rape (*Brassica napus*). Eur J Plant Pathol **114**: 3–15
- Francisco IA, Pinotti MHP (2000) Cyanogenic glycosides in plants. Brazilian Arch Biol Technol 43: 487–492
- Francisco M, Ali M, Ferreres F, Moreno DA, Velasco P, Soengas P (2016) Organ-specific quantitative genetics and candidate genes of phenylpropanoid metabolism in *Brassica oleracea*. Front Plant Sci 6: 1–14
- Francisco M, Moreno DA, Cartea ME, Ferreres F, García-Viguera C, Velasco P (2009) Simultaneous identification of glucosinolates and phenolic compounds in a representative collection of vegetable *Brassica rapa*. J Chromatogr A **1216**: 6611–6619
- Freezman NJ (1997) Analysis of a Chemical Plant Defense Mechanism in Grasses. 277:
- Funk C, Croteau R (1994) Diterpenoid Resin Acid Biosynthesis in Conifers: Characterization of Two Cytochrome P450-Dependent Monooxygenases and an Aldehyde Dehydrogenase Involved in Abietic Acid Biosynthesis. Arch Biochem Biophys 308: 258–266
- Gay EJ, Soyer JL, Lapalu N, Linglin J, Fudal I, Da Silva C, Wincker P, Aury JM, Cruaud C, Levrel A, et al (2021) Large-scale transcriptomics to dissect 2 years of the life of a fungal phytopathogen interacting with its host plant. BMC Biol **19**: 1–27
- Gervais J, Plissonneau C, Linglin J, Meyer M, Labadie K, Cruaud C, Fudal I, Rouxel T, Balesdent MH (2017) Different waves of effector genes with contrasted genomic location are expressed by *Leptosphaeria maculans* during cotyledon and stem colonization of oilseed rape. Mol Plant Pathol **18**: 1113–1126
- **Glazebrook J** (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annu Rev Phytopathol **43**: 205–227
- **Gleadow RM, Møller BL** (2014) Cyanogenic glycosides: Synthesis, physiology, and phenotypic plasticity. Annu Rev Plant Biol **65**: 155–185
- González-Coloma A, López-Balboa C, Santana O, Reina M, Fraga BM (2011) Triterpene-based plant defenses. Phytochem Rev 10: 245–260
- **Goodger JQD, Heskes AM, Woodrow IE** (2013) Contrasting ontogenetic trajectories for phenolic and terpenoid defences in Eucalyptus froggattii. Ann Bot **112**: 651–659
- **Goodger JQD, Senaratne SL, Nicolle D, Woodrow IE** (2018) Differential metabolic specialization of foliar oil glands in *Eucalyptus brevistylis* Brooker (Myrtaceae). Tree Physiol **38**: 1451–1460
- Gout L, Fudal I, Kuhn ML, Blaise F, Eckert M, Cattolico L, Balesdent MH, Rouxel T (2006) Lost in the middle of nowhere: The AvrLm1 avirulence gene of the Dothideomycete *Leptosphaeria maculans*. Mol Microbiol 60: 67–80
- Grant MR, Jones JDG (2009) Hormone (Dis)harmony Moulds Plant Health and Disease. Science (80-) 324: 750– 753
- **Gravot A** (2019) Génomique fonctionnelle des processus métaboliques impliqués dans les résistances quantitatives. Université de Rennes 1. L'archive ouverte pluridisciplinaire HAL, https://hal.archivesouvertes.fr/tel-03531005

Hanschen FS, Schreiner M (2017) Isothiocyanates, Nitriles, and Epithionitriles from Glucosinolates Are Affected

by Genotype and Developmental Stage in *Brassica oleracea* Varieties. Front Plant Sci. doi: 10.3389/fpls.2017.01095

- Harbaum B, Hubbermann EM, Wolff C, Herges R, Zhu Z, Schwarz K (2007) Identification of flavonoids and hydroxycinnamic acids in pak choi varieties (*Brassica campestris L. ssp. chinensis var. communis*) by HPLC-ESI-MSn and NMR and their quantification by HPLC-DAD. J Agric Food Chem **55**: 8251–8260
- Hartmann M, Zeier J (2019) N-hydroxypipecolic acid and salicylic acid: a metabolic duo for systemic acquired resistance. Curr Opin Plant Biol 50: 44–57
- Hartmann M, Zeier J (2018) L-lysine metabolism to N-hydroxypipecolic acid: an integral immune-activating pathway in plants. Plant J. doi: 10.1111/tpj.14037
- Hartmann M, Zeier T, Bernsdorff F, Reichel-Deland V, Kim D, Hohmann M, Scholten N, Schuck S, Bräutigam A, Hölzel T, et al (2018) Flavin Monooxygenase-Generated N-Hydroxypipecolic Acid Is a Critical Element of Plant Systemic Immunity. Cell 173: 456-469.e16
- Hartmann T (2008) The lost origin of chemical ecology in the late 19th century. Proc Natl Acad Sci U S A 105: 4541–4546
- Hartmann T (2007) From waste products to ecochemicals: Fifty years research of plant secondary metabolism. Phytochemistry **68**: 2831–2846
- Heaney RK, Fenwick GR (1980) Glucosinolates in brassica vegetables. Analysis of 22 varieties of brussels sprout (*Brassica oleracea var. gemmifera*). J Sci Food Agric **31**: 785–793
- Helliwell CA, Chandler PM, Poole A, Dennis ES, Peacock WJ (2001) The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. Proc Natl Acad Sci U S A 98: 2065–2070
- Hofberger JA, Lyons E, Edger PP, Chris Pires J, Eric Schranz M (2013) Whole Genome and Tandem Duplicate Retention Facilitated Glucosinolate Pathway Diversification in the Mustard Family. Genome Biol Evol 5: 2155–2173
- Holland J, (2015) MAGIC maize: a new resource for plant genetics. Genome Biol. 16: 163.
- van der Hooft JJJ, Mohimani H, Bauermeister A, Dorrestein PC, Duncan KR, Medema MH (2020) Linking genomics and metabolomics to chart specialized metabolic diversity. Chem Soc Rev 49: 3297–3314
- https://ourworldindata.org/grapher/rapeseed-production Hannah Ritchie and Max Roser Agricultural Production. https://ourworldindata.org/grapher/rapeseed-production?time=2017
- Huang R, O'Donnell AJ, Barboline JJ, Barkman TJ (2016) Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. Proc Natl Acad Sci U S A **113**: 10613–10618
- Huang YJ, Pirie EJ, Evans N, Delourme R, King GJ, Fitt BDL (2009) Quantitative resistance to symptomless growth of *Leptosphaeria maculans* (phoma stem canker) in *Brassica napus* (oilseed rape). Plant Pathol **58**: 314–323
- Huang YJ, Qi A, King GJ, Fitt BDL (2014) Assessing Quantitative resistance against *Leptosphaeria maculans* (phoma stem canker) in *Brassica napus* (oilseed rape) in young plants. PLoS One **9**: 1–10
- Huard-Chauveau C, Perchepied L, Debieu M, Rivas S, Kroj T, Kars I, Bergelson J, Roux F, Roby D (2013) An Atypical Kinase under Balancing Selection Confers Broad-Spectrum Disease Resistance in *Arabidopsis*. PLoS Genet. doi: 10.1371/journal.pgen.1003766
- Hurni S, Scheuermann D, Krattinger SG, Kessel B, Wicker T, Herren G, Fitze MN, Breen J, Presterl T, Ouzunova M, et al (2015) The maize disease resistance gene Htn1 against northern corn leaf blight encodes a wallassociated receptor-like kinase. Proc Natl Acad Sci U S A 112: 8780–8785

- Ishihara H, Tohge T, Viehöver P, Fernie AR, Weisshaar B, Stracke R (2016) Natural variation in flavonol accumulation in *Arabidopsis* is determined by the flavonol glucosyltransferase BGLU6. J Exp Bot 67: 1505–1517
- Iven T, König S, Singh S, Braus-Stromeyer SA, Bischoff M, Tietze LF, Braus GH, Lipka V, Feussner I, Dröge-Laser
 W (2012) Transcriptional Activation and Production of Tryptophan-Derived Secondary Metabolites in Arabidopsis Roots Contributes to the Defense against the Fungal Vascular Pathogen Verticillium longisporum. Mol Plant 5: 1389–1402
- Jan R, Asaf S, Numan M, Lubna, Kim KM (2021) Plant secondary metabolite biosynthesis and transcriptional regulation in response to biotic and abiotic stress conditions. Agronomy **11**: 1–31
- Kautsar SA, Medema MH (2018) Genomic Identification and Analysis of Specialized Metabolite Biosynthetic Gene Clusters in Plants Using PlantiSMASH. 1795: 127–141
- Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH (2017) PlantiSMASH: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters. Nucleic Acids Res 45: W55–W63
- Katz E, Li JJ, Jaegle B, Ashkenazy H, Abrahams SR, Bagaza C, Holden, Pires C. J, Angelovici R, Kliebenstein D. J (2021) Genetic variation, environment and demography intersect to shape *Arabidopsis* defense metabolite variation across Europe. Elife;10:e67784
- Keukens EAJ, de Vrije T, van den Boom C, de Waard P, Plasman HH, Thiel F, Chupin V, Jongen WMF, de Kruijff B (1995) Molecular basis of glycoalkaloid induced membrane disruption. BBA - Biomembr **1240**: 216–228
- Kittipol V, He Z, Wang L, Doheny-Adams T, Langer S, Bancroft I (2019) Genetic architecture of glucosinolate variation in *Brassica napus*. J Plant Physiol **240**: 152988
- Klayman DL (1985) Qinghaosu (artemisinin): An antimalarial drug from China. Science (80-) 228: 1049–1055
- Klein AP, Sattely ES (2017) Biosynthesis of cabbage phytoalexins from indole glucosinolate. Proc Natl Acad Sci U S A 114: 1910–1915
- Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001a) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. Genetics 159: 359– 370
- Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T (2001b) Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. Plant Physiol **126**: 811–825
- KNApSAcK database . A Compr. Species-Metabolite Relatsh. Database (Date accessed 2022-03-22), http://www.knapsackfamily.com/KNApSAcK/
- Kono Y, Kojima A, Nagai R, Watanabe M, Kawashima T, Onizawa T, Teraoka T, Watanab M, Koshino H, Uzawa J, et al (2004) Antibacterial diterpenes and their fatty acid conjugates from rice leaves. Phytochemistry 65: 1291–1298
- Kourelis J, van der Hoorn RAL (2018) Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. Plant Cell **30**: 285–299
- Kushalappa AC, Yogendra KN, Karre S (2016) Plant Innate Immune Response: Qualitative and Quantitative Resistance. CRC Crit Rev Plant Sci 35: 38–55
- Kusumi T, Nishino T, Nakayama T, Yonekura-Sakakibara K, Sato T, Kikuchi S, Fukui Y, Fukuchi-Mizutani M, Ueda
 T, Nakao M, et al (2000) Aureusidin synthase: a polyphenol oxidase homolog responsible for flower coloration. Science (80-) 10: 1163–1166
- Kuzina V, Ekstrøm CT, Andersen SB, Nielsen JK, Olsen CE, Bak S (2009) Identification of defense compounds in Barbarea vulgaris against the herbivore Phyllotreta nemorum by an ecometabolomic approach. Plant Physiol 151: 1977–1990

- Labandeira CC, Tremblay SL, Bartowski KE, Vanaller Hernick L (2014) Middle Devonian liverwort herbivory and antiherbivore defence. New Phytol **202**: 247–258
- Larkan NJ, Lydiate DJ, Parkin IAP, Nelson MN, Epp DJ, Cowling WA, Rimmer SR, Borhan MH (2012) The *Brassica napus* blackleg resistance gene LepR3 encodes a receptor-like protein. New Phytol.
- Larkan NJ, Ma L, Borhan MH (2014) The *Brassica napus* receptor-like protein RLM2 is encoded by a second allele. plant
- Larkan NJ, Ma L, Haddadi P, Buchwaldt M, Parkin IAP, Djavaheri M, Borhan MH (2020) The *Brassica napus* wallassociated kinase-like WAKL gene Rlm9 provides race-specific. Plant J.
- Leao TF, Wang M, Silva R, Hooft JJJ Van Der, Bauermeister A, Brejnrod A, Glukhov E, Gerwick L, Gerwick WH, Bandeira N, et al (2021) A supervised fingerprint-based strategy to connect natural product mass spectrometry fragmentation data to their biosynthetic gene clusters.
- Lee HT, Chawla HS, Obermeier C, Dreyer F, Abbadi A, Snowdon R (2020) Chromosome-Scale Assembly of Winter Oilseed Rape *Brassica napus*. Front Plant Sci **11**: 1–13
- Lehmann T, Janowitz T, Sánchez-Parra B, Alonso M-MP, Trompetter I, Piotrowski M, Pollmann S (2017) Arabidopsis NITRILASE 1 Contributes to the Regulation of Root Growth and Development through Modulation of Auxin Biosynthesis in Seedlings. Front Plant Sci. doi: 10.3389/fpls.2017.00036
- Leong BJ, Last RL (2017) Promiscuity, impersonation and accommodation: evolution of plant specialized metabolism. Curr Opin Struct Biol 47: 105–112
- Li B, Förster C, Robert CAM, Züst T, Hu L, Machado RAR, Berset JD, Handrick V, Knauer T, Hensel G, et al (2018) Convergent evolution of a metabolic switch between aphid and caterpillar resistance in cereals. Sci Adv 4: 1–15
- Li N, Euring D, Cha JY, Lin Z, Lu M, Huang LJ, Kim WY (2021) Plant Hormone-Mediated Regulation of Heat Tolerance in Response to Global Climate Change. Front Plant Sci **11**: 1–11
- Li Y, Wu YL (2010) A golden phoenix arising from the herbal nest A review and reflection on the study of antimalarial drug Qinghaosu. Front Chem China 5: 357–422
- Liu S, Huang H, Yi X, Zhang Y, Yang Q, Zhang C, Fan C, Zhou Y (2020) Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome-wide association study. Plant Biotechnol J 18: 1472–1484
- Liu T, Zhang Y, Agerbirk N, Wang H, Wei X, Song J, He H, Zhao X, Zhang X, Li X (2019) A high-density genetic map and QTL mapping of leaf traits and glucosinolates in *Barbarea vulgaris*. BMC Genomics **20**: 371
- Liu Z, Carpenter SB, Bourgeois WJ, Yu Y, Constantin RJ, Falcon MJ, Adams JC (1998) Variations in the secondary metabolite camptothecin in relation to tissue age and season in *Camptotheca acuminata*. Tree Physiol **18**: 265–270
- Llorach R, Gil-Izquierdo A, Ferreres F, Tomás-Barberán FA (2003) HPLC-DAD-MS/MS ESI characterization of unusual highly glycosylated acylated flavonoids from cauliflower (*Brassica oleracea L. var. botrytis*) agroindustrial byproducts. J Agric Food Chem **51**: 3895–3899
- Lô-Pelzer E, Aubertot J-N, David O, Jeuffroy M-H, Bousset-Vaslin L (2009) Relationship between severity of blackleg *Leptosphaeria maculans L biglobosa* species. Plant Pathol. doi: 10.1111/j.1365-3059.2008.01931.x
- Ludwig M, Nothias LF, Dührkop K, Koester I, Fleischauer M, Hoffmann MA, Petras D, Vargas F, Morsy M, Aluwihare L, et al (2020) Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. Nat Mach Intell 2: 629–641
- Lv H, Wang Y, Han F, Ji J, Fang Z, Zhuang M, Li Z, Zhang Y, Yang L (2020) A high-quality reference genome for cabbage obtained with SMRT reveals novel genomic features and evolutionary characteristics. Sci Rep 10:

1–9

- **Machingura M, Salomon E, Jez JM, Stephen D. Ebbs** (2016) The β-cyanoalanine synthase pathway: beyond cyanide detoxification. Plant Cell Environ.
- Macías FA, Marín D, Oliveros-Bastidas A, Molinillo JMG (2009) Rediscovering the bioactivity and ecological role of 1,4-benzoxazinones. Nat Prod Rep 26: 478–489
- Marcroft SJ, Sprague SJ, Pymer SJ, Salisbury PA, Howlett BJ (2004) Crop isolation, not extended rotation length, reduces blackleg (*Leptosphaeria maculans*) severity of canola (*Brassica napus*) in south-eastern Australia. Aust J Exp Agric 44: 601–606
- Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, De Bruijn I, Chooi YH, Claesen J, Coates RC, et al (2015) Minimum Information about a Biosynthetic Gene cluster. Nat Chem Biol **11**: 625–631
- Meraj TA, Fu J, Raza MA, Zhu C, Shen Q, Xu D, Wang Q (2020) Transcriptional Factors Regulate Plant Stress Responses Through Mediating Secondary Metabolism. World J Microbiol Biotechnol 8: 90–91
- Milo R, Last RL (2012) Achieving diversity in the face of constraints: lessons from metabolism Science, 336, pp. 1663-1667
- Missinou AA, Carvalho JF De, Marnet N, Delhaye T, Hamzaoui O, Sayed DA, Guitton Y, Lebreton L, Langrume C, Laperche A, et al (2022) Identification and Quantification of Glucosinolates and Phenolics in a Large Panel of *Brassica napus* Highlight Valuable Genetic Resources for Chemical Ecology and Breeding. J Agric Food Chem. doi: 10.1021/acs.jafc.1c08118
- Mittaine JF, Mielke T (2012) The globalization of international oilseeds trade. OCL-Oleagineux Corps Gras Lipides 19: 249–260
- Mo XJ, Gao EQ, He Z, Li WJ, Yan CH (2004) XCMS2: Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. Inorg Chem Commun 7: 353–355
- Montaut S, Read S, Blažević I, Nuzillard JM, Roje M, Harakat D, Rollin P (2020) Investigation of the glucosinolates in *Hesperis matronalis* L. and *Hesperis laciniata* All.: Unveiling 4'-O-β-D-apiofuranosylglucomatronalin. Carbohydr Res **488**: 107898
- Moore BM, Wang P, Fan P, Leong B, Schenck CA, Lloyd JP, Lehti-Shiu MD, Last RL, Pichersky E, Shiu S-H (2019) Robust predictions of specialized metabolism genes through machine learning. Proc Natl Acad Sci U S A 116: 2344–2353
- Morice J (1979) Avant-propos : la sélection du colza sans acide érucique. Ann Biol Anim Biochim Biophys 19: 471– 477
- Morkunas I, Ratajczak L (2014) The role of sugar signaling in plant defense responses against fungal pathogens. Acta Physiol Plant **36**: 1607–1619
- Morrissey JP, Osbourn AE (1999) Fungal Resistance to Plant Antibiotics as a Mechanism of Pathogenesis. Microbiol Mol Biol Rev 63: 708–724
- Murata J, Ono E, Yoroizuka S, Toyonaga H, Shiraishi A, Mori S, Tera M, Azuma T, Nagano AJ, Nakayasu M, et al (2017) Oxidative rearrangement of (+)-sesamin by CYP92B14 co-generates twin dietary lignans in sesame. Nat Commun 8: 1–10
- Nagaharu U (1935) Genome-analysis in *Brassica* with specila reference to the experimental formation of B. napus and particular mode of fertilization.
- Naseri B, Davidson JA, Scott ES (2009) Maturation of pseudothecia and discharge of ascospores of *Leptosphaeria* maculans on oilseed rape stubble. Eur J Plant Pathol **125**: 523–531

Návarová H, Bernsdorff F, Döring AC, Zeier J (2013) Pipecolic acid, an endogenous mediator of defense

amplification and priming, is a critical regulator of inducible plant immunity. Plant Cell 24: 5123–5141

- Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, et al (2020) A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16: 60–68
- Neilson EH, Goodger JQD, Motawia MS, Bjarnholt N, Frisch T, Olsen CE, Møller BL, Woodrow IE (2011) Phenylalanine derived cyanogenic diglucosides from *Eucalyptus camphora* and their abundances in relation to ontogeny and tissue type. Phytochemistry **72**: 2325–2334
- Nelson R, Wiesner-Hanks T, Wisser R, Balint-Kurti P (2018) Navigating complexity to breed disease-resistant crops. Nat Rev Genet 19: 21–33
- 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
- Nielsen JK, Nørbæk R, Olsen CE (1998) Kaempferol tetraglucosides from cabbage leaves. Phytochemistry 49: 2171–2176
- Nielsen LJ, Stuart P, Pičmanová M, Rasmussen S, Olsen CE, Harholt J, Møller BL, Bjarnholt N (2016) Dhurrin metabolism in the developing grain of *Sorghum bicolor* (L.) Moench investigated by metabolite profiling and novel clustering analyses of time-resolved transcriptomic data. BMC Genomics **17**: 1–24
- Niemeyer HM (2009) Hydroxamic acids derived from 2-hydroxy-2h-1,4-benzoxazin-3(4h)-one: Key defense chemicals of cereals. J Agric Food Chem 57: 1677–1695
- Nothias LF, Daniel Petras RS, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H (2020) Feature-Based Molecular Networking in the GNPS Analysis Environment. Physiol Behav **176**: 139–148
- Nowacki EK, Waller R (1975) Use of the metabolic grid to explain the metabolism of quinolizidine alkaloids in Leguminosae. 14: 165–171
- Ochoa-Lopez S, Damian X, Rebollo R, Fornoni J, Dominguez C, A., Boege1 K (2019) Ontogenetic changes in the targets of natural selection in three plant defenses. New Phytol.
- Olsen H, Aaby K, Borge GIA (2009) Characterization and quantification of flavonoids and hydroxycinnamic acids in curly kale (*Brassica oleracea L. convar. acephala var. sabellica*) by HPLC-DAD-ESI-MSn. J Agric Food Chem 57: 2816–2825
- **Osbourn AE** (1996) Preformed antimicrobial compounds and plant defense against fungal attack. Plant Cell **8**: 1821–1831
- Van Overbeek J (1966) Plant hormones and regulators. Science (80-) 152: 721-731
- Palaniswamy UR, McAvoy RJ, Bible BB, Stuart JD (2003) Ontogenic variations of ascorbic acid and phenethyl isothiocyanate concentrations in watercress (*Nasturtium officinale R.Br.*) leaves. J Agric Food Chem 51: 5504–5509
- Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome Biol **15**: 1–18
- Pedras MSC, Yaya EE (2010) Phytoalexins from *Brassicaceae*: News from the front. Phytochemistry **71**: 1191–1197
- Pedras S, Yaya EE (2015) Plant chemical defenses: Are all constitutive antimicrobial metabolites phytoanticipins? Nat Prod Commun 10: 209–218
- Pei Y, Zhu Y, Jia Y, Ge X, Li X, Li F, Hou Y (2020) Molecular evidence for the involvement of cotton GhGLP2, in enhanced resistance to Verticillium and Fusarium Wilts and oxidative stress. Sci Rep 10: 1–15

- Pellan-Delourme R, Renard M (1988) Cytoplasmic male sterility in rapeseed (*Brassica napus L*.): female fertility of restored rapeseed with "Ogura" and cybrids cytoplasms . Genome **30**: 234–238
- Pfalz M, Mukhaimar M, Perreau F, Kirk J, Hansen CIC, Olsen CE, Agerbirk N, Kroymann J (2016) Methyl Transfer in Glucosinolate Biosynthesis Mediated by Indole Glucosinolate O-Methyltransferase 5. Plant Physiol 172: 2190–2203
- Pichersky E, Lewinsohn E (2011) Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol 62: 549–566
- **Pihlava JM, Kurtelius T** (2016) Determination of benzoxazinoids in wheat and rye beers by HPLC-DAD and UPLC-QTOF MS. Food Chem **204**: 400–408
- Pilet-Nayel M-L, Moury B, Caffier V, Montarry J, Kerlan M-C, Fournet S, Durel C-E, Delourme R (2017) Quantitative Resistance to Plant Pathogens in Pyramiding Strategies for Durable Crop Protection. Front Plant Sci 8: 1838
- Pinochet X, Renard M (2012) Progrès génétique en colza et perspectives. 19: 147–154
- Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. doi: 10.1186/1471-2105-11-395
- Poloni A, Schirawski J (2014) Red card for pathogens: Phytoalexins in sorghum and maize. Molecules 19: 9114– 9133
- Ponikvar-Svet M, Zeiger DN, Liebman JF (2021) Alkaloids and selected topics in their thermochemistry. Molecules. doi: 10.3390/molecules26216715
- Poschenrieder C, Tolrà R, Barceló J (2006) Can metals defend plants against biotic stress? Trends Plant Sci 11: 288–295
- Poulton JE, Li CP (1994) Tissue level compartmentation of (R)-amygdalin and amygdalin hydrolase prevents largescale cyanogenesis in undamaged Prunus seeds. Plant Physiol **104**: 29–35
- **Preston JC** (2016) Evo-Devo: Regulatory and Protein-Coding Evolution in Plant Diversification. Encycl Evol Biol **2**: 13–18
- Qian W, Wu C, Fu Y, Hu G, He Z, Liu W (2017) Novel rice mutants overexpressing the brassinosteroid catabolic gene CYP734A4. Plant Mol Biol 93: 197–208
- Rai A, Saito K, Yamazaki M (2017) Integrated omics analysis of specialized metabolism in medicinal plants. Plant J 90: 764–787
- Rajniak J, Barco B, Clay NK, Sattely Elizabeth S (2017) A new cyanogenic metabolite in Arabidopsis required for inducible pathogen defense. Physiol Behav 176: 139–148
- Rajniak J, Barco B, Clay NK, Sattely ES (2015) A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defence. Nature **525**: 376–9
- Roddick JG (1977) Subcellular localization of steroidal glycoalkaloids in vegetative organs of *Lycopersicon* esculentum and Solanum tuberosum. Phytochemistry **16**: 805–807
- Rogers S, Ong CW, Wandy J, Ernst M, Ridder L, Van Der Hooft JJJ (2019) Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. Faraday Discuss 218: 284–302
- **Rojas CM, Senthil-Kumar M, Tzin V, Mysore KS** (2014) Regulation of primary plant metabolism during plantpathogen interactions and its contribution to plant defense. Front Plant Sci **5**: 1–12

- Rousseau-Gueutin M, Belser C, Silva C Da, Richard G, Istace B, Cruaud C, Falentin C, Boideau F, Boutte J, Delourme R, et al (2020) Long-read assembly of the *Brassica napus* reference genome *Darmor-bzh*. Gigascience 9: 1–16
- Routaboul J-M, Dubos C, Beck G, Marquis C, Bidzinski P, Loudet O, Lepiniec L (2012) Metabolite profiling and quantitative genetics of natural variation for flavonoids in *Arabidopsis*. J Exp Bot **63**: 695–709
- Roux F, Voisin D, Badet T, Balagué C, Barlet X, Huard-Chauveau C, Roby D, Raffaele S (2014) Resistance to phytopathogens e tutti quanti: Placing plant quantitative disease resistance on the map. Mol Plant Pathol 15: 427–432
- Rouxel T, Balesdent MH (2005) The stem canker (blackleg) fungus, *Leptosphaeria maculans*, enters the genomic era. Mol Plant Pathol 6: 225–241
- Rouxel T, Penaud A, Pinochet X, Brun H, Gout L, Delourme R, Schmit J, Balesdent MH (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
- Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. Plant Cell 20: 1199–1216
- Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D (2018) The genome biology of effector gene evolution in filamentous plant pathogens. Annu Rev Phytopathol 56: 21–40
- Schläpfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, Dreher K, Chavali AK, Nilo-Poyanco R, Bernard T, et al (2017) Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. Plant Physiol **173**: 2041–2059
- Schmelz EA, Kaplan F, Huffaker A, Dafoe NJ, Vaughan MM, Ni X, Rocca JR, Alborn HT, Teal PE (2011) Identity, regulation, and activity of inducible diterpenoid phytoalexins in maize. Proc Natl Acad Sci U S A **108**: 5455–5460
- Schmid R, Petras D, Nothias LF, Wang M, Aron AT, Jagels A, Tsugawa H, Rainer J, Garcia-Aloy M, Dührkop K, et al (2021) Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. Nat Commun. doi: 10.1038/s41467-021-23953-9
- Schmidt S, Zietz M, Schreiner M, Rohn S, Kroh LW, Krumbein A (2010) Identification of complex, naturally occurring flavonoid glycosides in kale (*Brassica oleracea var. sabellica*) by high-performance liquid chromatography diode-array detection/electrospray ionization multi-stage mass spectrometry. Rapid Commun Mass Spectrom 24: 2009–2022
- Schreiner M, Beyene B, Krumbein A, Stützel H (2009) Ontogenetic changes of 2-propenyl and 3-Lndolylmethyl glucosinolates in *Brassica carinata* leaves as affected by water supply. J Agric Food Chem **57**: 7259–7263
- Shao Y, Jiang J, Ran L, Lu C, Wei C, Wang Y (2014) Analysis of flavonoids and hydroxycinnamic acid derivatives in rapeseeds (*Brassica napus L. var. napus*) by HPLC-PDA-ESI(-)-MSn/HRMS. J Agric Food Chem 62: 2935–2945
- Silva E, da Graça JP, Porto C, Martin do Prado R, Hoffmann-Campo CB, Meyer MC, de Oliveira Nunes E, Pilau EJ (2020) Unraveling Asian Soybean Rust metabolomics using mass spectrometry and Molecular Networking approach. Sci Rep **10**: 1–11
- So G, Ericsson S, Fox GA, Schaffer WM, Rosko D, Kaitala V, Lundberg P, Levin SA, Rohani P, Dabholkar S, et al (2006) A Plant miRNA Contributes to Antibacterial Resistance by Repressing Auxin Signaling. Science (80-) 436–439
- Song JM, Guan Z, Hu J, Guo C, Yang Z, Wang S, Liu D, Wang B, Lu S, Zhou R, et al (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nat Plants. doi: 10.1038/s41477-019-0577-7

- Sotelo T, Soengas P, Velasco P, Rodriguez VM, Cartea ME (2014) Identification of metabolic QTLs and candidate genes for glucosinolate synthesis in *Brassica oleracea* leaves, seeds and flower buds. PLoS One. doi: 10.1371/journal.pone.0091428
- Spoel SH, Dong X (2012) How do plants achieve immunity? Defence without specialized immune cells. Nat Rev Immunol 12: 89–100
- Sprague SJ, Balesdent MH, Brun H, Hayden HL, Marcroft SJ, Pinochet X, Rouxel T, Howlett BJ (2006a) Major gene resistance in *Brassica napus* (oilseed rape) is overcome by changes in virulence of populations of *Leptosphaeria maculans* in France and Australia. Eur J Plant Pathol **114**: 33–40
- Sprague SJ, Marcroft SJ, Hayden HL, Howlett BJ (2006b) Major gene resistance to blackleg in *Brassica napus* overcome within three years of commercial production in Southeastern Australia. Plant Dis **90**: 190–198
- Tanaka Y, Brugliera F (2013) Flower colour and cytochromes P450. Philos Trans R Soc B Biol Sci. doi: 10.1098/rstb.2012.0432
- Tanwir F, Fredholm M, Gregersen PL, Fomsgaard IS (2013) Comparison of the levels of bioactive benzoxazinoids in different wheat and rye fractions and the transformation of these compounds in homemade foods. Food Chem 141: 444–450
- Taveira M, Fernandes F, Guedes de Pinho P, Andrade PB, Pereira JA, Valentão P (2009) Evolution of *Brassica* rapa var. rapa L. volatile composition by HS-SPME and GC/IT-MS. Microchem J **93**: 140–146
- Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A (2014) Triterpene biosynthesis in plants. Annu Rev Plant Biol 65: 225–257
- Threlfall DR, Whitehead IM (1988) Co-ordinated inhibition of squalene synthetase and induction of enzymes of sesquiterpenoid phytoalexin biosynthesis in cultures of *Nicotiana tabacum*. Phytochemistry 27: 2567–2580
- Traka, M.; Mithen, R. (2009) Glucosinolates, isothiocyanates and human health. Phytochem. Rev. 2009, 8, 269–282
- Tissier A, Ziegler J, Vogt T (2015) Specialized Plant Metabolites: Diversity and Biosynthesis. Ecol Biochem Environ Interspecies Interact 14–37
- Tohge T, Wendenburg R, Ishihara H, Nakabayashi R, Watanabe M, Sulpice R, Hoefgen R, Takayama H, Saito K, Stitt M, et al (2016) Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in *Brassicaceae*. Nat Commun. doi: 10.1038/ncomms12399
- Töpfer N, Fuchs LM, Aharoni A (2017) The PhytoClust tool for metabolic gene clusters discovery in plant genomes. Nucleic Acids Res 45: 7049–7063
- Trabelcy B, Chinkov N, Samuni-Blank M, Merav M, Izhaki I, Carmeli S, Gerchman Y (2021) Investigation of glucosinolates in the desert plant *Ochradenus baccatus* (Brassicales: Resedaceae). Unveiling glucoochradenin, a new arabinosylated glucosinolate. Phytochemistry **187**: 112760
- Tsugawa H, Ikeda K, Takahashi M, Satoh A, Mori Y, Uchino H, Okahashi N, Yamada Y, Tada I, Bonini P, et al (2020) MS-DIAL 4: accelerating lipidomics using an MS/MS, CCS, and retention time atlas.
- Valente Pereira FM, Rosa E, Fahey JW, Stephenson KK, Carvalho R, Aires A (2002) Influence of temperature and ontogeny on the levels of glucosinolates in broccoli (*Brassica oleracea var. italica*) sprouts and their effect on the induction of mammalian phase 2 enzymes. J Agric Food Chem **50**: 6239–6244
- VanEtten HD, Mansfield JW, Bailey JA, Farmer EE (1994) Two Classes of Plant Antibiotics: Phytoalexins versus "Phytoanticipins." Plant Cell 6: 1191
- Velasco P, Francisco M, Moreno DA, Ferreres F, García-Viguera C, Cartea ME (2011) Phytochemical fingerprinting of vegetable *Brassica oleracea* and *Brassica napus* by simultaneous identification of glucosinolates and phenolics. Phytochem Anal 22: 144–152

- Vleminckx J, Salazar D, Fortunel C, Mesones I, Dávila N, Lokvam J, Beckley K, Baraloto C, Fine P V A (2018) Divergent secondary metabolites and habitat filtering both contribute to tree species coexistence in the peruvian amazon. Front. Plant Sci., 9, p. 836
- Villagrasa M, Guillamón M, Labandeira A, Taberner A, Eljarrat E, Barceló D (2006) Benzoxazinoid allelochemicals in wheat: Distribution among foliage, roots, and seeds. J Agric Food Chem 54: 1009–1015
- Wagner G, Laperche A, Lariagon C, Marnet N, Renault D, Guitton Y, Bouchereau A, Delourme R, Manzanares-Dauleux MJ, Gravot A (2019) Resolution of quantitative resistance to clubroot into QTL-specific metabolic modules. J Exp Bot. doi: 10.1093/jxb/erz265
- Waminal NE, Perumal S, Lee J, Kim HH, Yang T-J (2016) Repeat Evolution in *Brassica rapa* (AA), *B. oleracea* (CC), and *B. napus* (AACC) Genomes. Plant Breed Biotechnol **4**: 107–122
- Wang-Zilong, Wang S, Xu Z, Li M, Chen K, Zhang Y, Hu Z, Zhang M, Zhang Z, Qiao X, et al (2019) Highly Promiscuous Flavonoid 3- O-Glycosyltransferase from *Scutellaria baicalensis*. Org Lett **21**: 2241–2245
- Wang B, Wu Z, Li Z, Zhang Q, Hu J, Xiao Y, Cai D, Wu J, King GJ, Li H, et al (2018) Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. Plant Biotechnol J 16: 1336–1348
- Wang D, Pajerowska-Mukhtar K, Culler AH, Dong X (2007) Salicylic Acid Inhibits Pathogen Growth in Plants through Repression of the Auxin Signaling Pathway. Curr Biol **17**: 1784–1790
- Wang M, Carver JJ, Phelan V V, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Clifford A Kapono, Tal Luzzatto-Knaan3 CP (2016) Sharing and community curation of mass spectrometry data with GNPS. Nat Biotechnol 34: 828–837
- Wang S, Alseekh S, Fernie AR, Luo J (2019) The Structure and Function of Major Plant Metabolite Modifications. Mol Plant 12: 899–919
- Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet **43**: 1035–1040
- Weng JK (2014) The evolutionary paths towards complexity: A metabolic perspective. New Phytol **201**: 1141–1149
- Weng JK, Lynch JH, Matos JO, Dudareva N (2021), Nat. Chem. Biol., 17, 1037–1045
- West JS, Kharbanda PD, Barbetti MJ, Fitt BDL (2001) Epidemiology and management of *Leptosphaeria maculans* (phoma stem canker) on oilseed rape in Australia, Canada and Europe. Plant Pathol **50**: 10–27
- Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis, second edi. J R Stat Soc Ser A (Statistics Soc. doi: 10.1111/j.1467-985x.2010.00676_9.x
- Williams DJ, Pun S, Chaliha M, Scheelings P, O'Hare T (2013) An unusual combination in papaya (*Carica papaya*): The good (glucosinolates) and the bad (cyanogenic glycosides). J Food Compos Anal **29**: 82–86
- **Wink, M.** (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64, 3–19.
- Wink M (2010) Introduction: Biochemistry, Physiology and Ecological Functions of Secondary Metabolites. Biochem Plant Second Metab Second Ed **40**: 1–19
- Winter M, Koopmann B (2016) Race spectra of *Leptosphaeria maculans*, the causal agent of blackleg disease of oilseed rape, in different geographic regions in northern Germany. Eur J Plant Pathol **145**: 629–641
- Woldemariam MG, Dinh ST, Oh Y, Gaquerel E, Baldwin IT, Galis I (2013) NaMYC2 transcription factor regulates a subset of plant defense responses in *Nicotiana attenuata*. BMC Plant Biol. doi: 10.1186/1471-2229-13-73

- Won C, Shen X, Mashiguchi K, Zheng Z, Dai X, Cheng Y, Kasahara H, Kamiya Y, Chory J, Zhao Y (2011) Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in *Arabidopsis*. Proc Natl Acad Sci U S A **108**: 18518–23
- van de Wouw AP, Cozijnsen AJ, Hane JK, Brunner PC, McDonald BA, Oliver RP, Howlett BJ (2010) Evolution of linked avirulence effectors in *Leptosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. PLoS Pathog. doi: 10.1371/journal.ppat.1001180
- Van de Wouw AP, Sheedy EM, Ware AH, Marcroft SJ, Idnurm A (2022) Independent breakdown events of the *Brassica napus* Rlm7 resistance gene including via the off-target impact of a dual-specificity avirulence interaction. Mol Plant Pathol 1–14
- Yamane H, Konno K, Sabelis M, Takabayashi J, Sassa T, Oikawa H (2010) Chemical defence and toxins of plants. Compr Nat Prod II Chem Biol 4: 339–385
- Yang J, Guo H, Jiang NJ, Tang R, Li GC, Huang LQ, van Loon JJA., Wang CZ (2021) Identification of a gustatory receptor tuned to sinigrin in the cabbage butterfly *Pieris rapae*. PLoS Genet; 17(7):e1009527.
- Yildiz I, Mantz M, Hartmann M, Zeier T, Kessel J, Thurow C, Gatz C, Petzsch P, Köhrer K, Zeier J (2021) The mobile SAR signal N-hydroxypipecolic acid induces NPR1-dependent transcriptional reprogramming and immune priming. Plant Physiol **186**: 1679–1705
- Oscar R . Ortega , Daniel J . Kliebenstein , Carlos Arbizu , Ramiro Ortega and Carlos F . Quiros (2006) Glucosinolate Survey of Cultivated and Feral Mashua (*Tropaeolum tuberosum* Ruíz & Pavón) in the Cuzco Region of Peru. Economic Botany, 60(3), 2006, pp. 254-264. The New York Botanical Garden Press, Bronx, NY 10458-5126 U.S.A
- Zachariah TJ, Leela NK (2018) Spices: secondary metabolites and medicinal properties. In: Sharangi A (ed) Indian Spices. Springer International Publishing, Cham, pp 277–316
- Zhang X, Peng G, Kutcher HR, Balesdent MH, Delourme R, Fernando WGD (2016) Breakdown of Rlm3 resistance in the *Brassica napus-Leptosphaeria maculans* pathosystem in western Canada. Eur J Plant Pathol **145**: 659–674
- Zhao F, Chen YP, Salmaki Y, Drew BT, Wilson TC, Scheen AC, Celep F, Bräuchler C, Bendiksby M, Wang Q, et al (2021) An updated tribal classification of Lamiaceae based on plastome phylogenomics. BMC Biol **19**: 1–27
- Zhao N, Ferrer JL, Moon HS, Kapteyn J, Zhuang X, Hasebe M, Neal Stewart C, Gang DR, Chen F (2012) A SABATH Methyltransferase from the moss *Physcomitrella patens* catalyzes S-methylation of thiols and has a role in detoxification. Phytochemistry 81: 31–41
- Zheng X, Koopmann B, Ulber B, von Tiedemann A (2020) A Global Survey on Diseases and Pests in Oilseed Rape-Current Challenges and Innovative Strategies of Control. Front Agron 2: 1–15
- Zipfel C (2014) Plant pattern-recognition receptors. Trends Immunol 35: 345–351
- Zou J, Mao L, Qiu2 J, Wang M, Jia L, Wu D, He Z, Chen M, Shen Y, Shen E, et al (2019) Genome-wide selection footprints and deleterious variations in young Asian. Plant Biotechnol. J.

DOCTORAT/ECOLOGIE BRETAGNE\GEOSCIENCES LOIRE/AGRONOMIE ALIMENTATION

Titre : Métabolisme spécialisé chez *Brassica napus* : Caractérisation de la diversité phytochimique, de ses déterminants génétiques et de sa régulation par l'infection d'un agent pathogène

Mots clés : colza, métabolites spécialisées, diversité, mGWAS, agent pathogène, métabolomique

Le métabolisme spécialisé joue un rôle essentiel dans les interactions plante-environnement. Sa caractérisation chez les plantes cultivées représente un enjeu scientifique important. Au cours de ma thèse, j'ai travaillé à la caractérisation du métabolisme spécialisé et de ses déterminants génétiques chez Brassica napus, et j'ai étudié sa régulation en réponse à l'infection par Leptosphaeria maculans, agent causal du phoma. J'ai contribué au développement d'une méthode de profilage métabolique ciblé et identifié 36 glucosinolates (GLSs) foliaires/racinaires, 32 composés phénoliques (PHLs) foliaires et 18 composés racinaires non documentés jusqu'à présent. La quantification des ces SMs dans 304 accessions de Brassica a permis de révéler des contrastes phytochimiques importants au sein du panel. L'analyse GWAS de ces variations phytochimiques a identifié 104 locus (QTLs). L'architecture génétique ainsi mise en évidence suggère un controle indépendant des souscatégories de GLSs, PHLs et des nouveaux composés racinaires. Ces travaux fournissent une ressource utile pour l'écologie chimique et la sélection des Brassicas. En utilisant la même méthode de profilage ciblé, nous avons montré une induction des GLSs indoliques 14 jours après inoculation de la tige de B. napus par L. maculans. L'imagerie par spectrométrie de masse de la section transversale de tige de colza infectée a permis de révéler une spatialisation des réponses métaboliques suite à l'infection par L. maculans.

Title: Specialized metabolism in *Brassica napus*: Characterization of phytochemical diversity, genetic determinants, and its regulation by a pathogen infection.

Keywords: oilseed rape, specialized metabolites, diversity, mGWAS, pathogen, metabolomic

Specialized metabolism plays an essential role plant-environment interactions. in lts characterization in cultivated plants represents a significant scientific challenge. During my thesis, I worked on the characterization of the specialized metabolism and its genetic determinants in Brassica napus (oilseed rape), and I studied its regulation in response to infection by Leptosphaeria maculans (causal agent of phoma). I contributed to developing a targeted metabolic profiling method and identified 36 foliar/root glucosinolates (GLSs), 32 foliar phenolic compounds (PHLs), and 18 previously undocumented root compounds. The quantification of these SMs in 304 Brassica accessions revealed important phytochemical contrasts within the panel. GWA analysis of these phytochemical variations identified 104 loci (QTLs). The genetic architecture thus highlighted an independent control of the subcategories of GLSs, PHLs, and new root compounds. This work provides a useful resource for Brassica chemical ecology and breeding. Using the same targeted profiling method, we showed induction of indole GLSs 14 days after inoculation of the B. napus stems with L. maculans. Mass spectrometric imaging of the cross-section of infected rapeseed stem spatialization of metabolic revealed the responses to *L. maculans* infection.