The impact of genetics and environment on the polar fraction metabolome of commercial *Brassica napus* seeds: a multi-site study

*Author for correspondence:
Jean-Charles Martin, Email: jean-charles.martin@univ-amu.fr

Djawed Bennouna[1] (ORCID), Jean-Christophe Avice[2], Clément Rosique[1], Ljubica Svilar[1,6], Célia Pontet[3], Jacques Trouverie[2], Frédéric Fine[3], Xavier Pinochet[3], Karl Fraser[4,5] and Jean-Charles Martin[1,]*

[1]Aix Marseille Université, INSERM, INRA, C2VN, BioMeT, Marseille, France; [2]Normandie Université, UNICAEN, INRA, SFR Normandie Végétal (FED4277), UMR 950 Ecophysiologie Végétale et Agronomie, F-14032 Caen, France; [3]Terres Inovia, Paris, France; [4]Food Nutrition & Health Team, Food & Bio-Based Products Group, AgResearch Grasslands Research Centre, Palmerston North, New Zealand; [5]Riddet Institute, Massey University, Palmerston North 4442, New Zealand and [6]CriBioM, Criblage Biologique Marseille, Faculté de Médecine de la Timone, Marseille, France

Abstract

This study was designed to elucidate the biological variation in expression of many metabolites due to environment, genotype, or both, and to investigate the potential utility of metabolomics to supplement compositional analysis for the design of a new resilient cultivar of *Brassica napus* that can be steady in phytochemicals in different regions in France. Eight rapeseed varieties, grown in eight regions of France, were compared using a non-targeted metabolomics approach. The statistical analysis highlighted the distance and closeness between the samples in terms of both genotypes and geographical regions. A major environmental impact was observed on the polar metabolome, with different trends, depending on the varieties. Some varieties were very sensitive to the environment, while others were quite resilient. The identified secondary metabolites were mapped into the KEGG pathway database to reveal the most sensitive target proteins susceptible to environmental influences. A glucosyl-transferase encoded by the *UGT84A1* gene involved in the biosynthesis of phenylpropanoid was identified. This protein could be rate limiting/promoting in this pathway depending on environmental conditions. The metabolomics approach used in this study demonstrated its efficiency to characterize the environmental influence on various cultivars of *Brassica napus* seeds and may help identify targets for crop improvement.

Introduction

Rapeseed or canola (*Brassica napus*) is an annual plant that belongs to the Brassicaceae family formerly known as the cruciferae (Miller-Cebert *et al.*, 2009; Yang *et al.*, 2015). This important family consists of 3200 species which correspond essentially to herbaceous plants cultivated mainly for the production of oil, for human and animal food, or as ornamental plants (Scott *et al.*, 1999). Among the Brassicaceae species are cabbage, mustard, turnip and also thale cress (*Arabidopsis thaliana*). *Brassica napus* is an allopolyploid cross between its parents *Brassica rapa* (AA, 2n = 20) and *Brassica oleracea* (CC, 2n = 18) and is considered one of the world's most important oleoproteaginous crops, sought after primarily for vegetable oil, animal feed and biofuel (Misra, 2016). Oil extracted from *B. napus* seeds is regarded as a rich source of polyunsaturated fatty acids, with about 7–10% α-linolenic and 17–21% linoleic acids (Baux *et al.*, 2008). The end-product of the oil seed extraction process corresponds to the rapeseed meal and represents a good source of a high-quality protein that displays a well-balanced amino acid composition with high levels of essential sulphur-containing amino acids. Thus, rapeseed meal is regarded as a valuable by-product for the food industry, both as animal feed and for human nutrition (Farag *et al.*, 2013). Moreover, several studies have intensively investigated the beneficial effects of certain phytochemicals found in rapeseed meal such as phenolic compounds, which play an essential role in the prevention of cardiovascular diseases and cancer (Shao *et al.*, 2014). However, this rapeseed meal is also rich in anti-nutritive metabolites, which decreases its market value (Vermorel *et al.*, 1988). Currently, the cultivation of high-quality rapeseed is one of the main objectives of most oilseed breeders and industries. Gains by both have been made, such as the improvement of rapeseed quality by selective breeding of yellow-seeded *B. napus* for thinner coat and higher protein content (He *et al.*, 2008), or the optimization of industrial processes to reduce anti-nutritional metabolites such as glucosinolates (Fenwick *et al.*, 1986). Nevertheless, using an upstream approach by combining crop cultivars with environmental conditions and cultural management could be a catalyst to further improve the nutritional quality of the plants and to reach a constant quality. From this perspective, the use

CAMBRIDGE UNIVERSITY PRESS

of a metabolomics approach seems to be the appropriate option to explore the production of metabolites of interest among varieties and across different locations. Metabolomics can be described as a global analysis of small molecules in a living system which are produced or modified as a result of an adaptation to the environment (Nicholson *et al.*, 1999; Fiehn, 2001). This approach is a powerful tool in the exploration of various aspects of plant physiology and biology, and broadens our knowledge of the metabolic and molecular regulatory mechanisms that condition stress responses such as environment or culturing management (Hong *et al.*, 2016). In this study, we investigated the influence of both genetics and environment on the expression of the polar fraction metabolome of commercial *B. napus* seeds by using liquid chromatography–mass spectrometry (LC-MS), and how this could be useful to identify new targets of crop improvement, such as nutritional quality.

Materials and methods

Plant material

Eight different varieties of *Brassica napus* were harvested in eight different regions of France by Terres Inovia, the technical centre for the research and development of oilseed crops, grain legumes and industrial hemp in France. The oilseed varieties chosen for this study were Pamela, Bonanza, Dalton, Stefano-KWS, DK-Exceptions, DK-Extenciel, DK-Extorm and DK-Expertise. These genotypes were grown in eight locations contrasted for their climate and soil characteristics: Lot-et-Garonne, Meurthe-et-Moselle, Allier, Loiret, Vendée, Ille-et-Vilaine, Marne and Somme (see Tables S1 and S2 and Fig. S1 in the Supplementary Material). All samples were planted during 2014 and harvested in 2015. Seeds were collected when fully ripe. Due to insufficient harvests, some of the varieties no longer conformed to the parameters of the study and had to be excluded, leaving 59 samples instead of the desired 64.

Sample preparation

Ten grams of seeds were processed to obtain the metabolic profile for every sample. This quantity corresponds to approximately 3000 seeds, which should be representative of every field trial of rapeseed that has to be studied. Seeds were stored at –80°C and then were flash-frozen in liquid nitrogen and finely ground into flour using an automatic mortar (Pulverisette 2, Fritsch Idar-Oberstein, Germany). All samples were freeze-dried over 24 hours by a high vacuum line (Christ bioblock scientific, Rungis, France) before the extraction.

Chemical reagents

Methanol (MeOH), acetonitrile (ACN) (LC-MS grade) and ethyl acetate (high performance liquid chromatography isocratic grade) were purchased from Carlo Erba Reagents (France). Formic acid was obtained from Sigma Aldrich (St Louis, MO, USA). Acetic acid was purchased from Merck KGaA (Darmstadt, Germany). MilliQ water was used for ultra performance liquid chromatography (UPLC) analysis.

Extraction protocol

Triplicate samples were extracted in two steps using the protocol of Jervis *et al.* (2015) with minor modifications. Two hundred milligrams of the lyophilized flour was used for every replicate, then homogenized with 3.5 ml of ethyl acetate. The latter was used as an alternative to hexane to remove the non-polar fraction (Lohani *et al.*, 2015). All the samples were vortexed then extracted in an ultrasonicated bath at room temperature for 20 min. The supernatant was collected after centrifugation (1700 g for 15 min at 10°C). This extraction procedure was repeated twice. The supernatants were combined, dried under the gentle nitrogen stream and then stored at –80°C until further analysis. The remaining ethyl acetate was removed from the defatted rapeseed powder under the nitrogen stream, then stored at –80°C. The polar extraction was performed by homogenizing 30 mg of the dried defatted flour with MeOH: 0.1% aqueous acetic acid (HOAc) (0.5ml, 9:1, v/v). The extraction procedure was repeated twice with a sequence of vortexing, sonication for 20 min, and centrifugation (1700 g for 15 min at 4°C). The supernatants were pooled, dried under the gentle nitrogen stream and stored at –80°C.

UPLC-HRMS (high resolution mass spectrometry) analysis

Before the analysis, all the dried polar extracts were reconstituted with 0.1% aqueous formic acid:MeOH (9:1, v/v, 240 μl). All samples were vortexed for 5 min, followed by centrifugation (13000 g for 10 min at 4°C). An aliquot (20 μl) was transferred to an LC-MS-grade vial and diluted with 80 μl of 0.1% aqueous formic acid:ACN (9:1, v/v). Sample separation was performed by UPLC ultimate 3000 (Thermo Scientific), coupled with a HRMS, a hybrid quadrupole-orbitrap mass spectrometer, Q-Exactive Plus (Thermo Scientific) equipped with a heated-electrospray ionization source (H-ESI II). The chromatographic separation was performed on a binary solvent system at a flow rate of 0.4 ml min^{-1} using a reverse phase C18 column (Hypersil Gold, Thermo Scientific, 100 mm × 2.1mm, 1.9 μm) at 40°C. The mobile phase consisted of a combination of solvent A (0.1% FA in water, v/v) and solvent B (0.1% FA in acetonitrile, v/v). The injection volume was 5 μl. The following gradient conditions were used: 0 to 1 min, isocratic 100% A; 1 to 11 min, linear from 0 to 100% B; 11 to 13 min, isocratic 100% B; 13 to 14 min, linear from 100 to 0% B; 14 to 16 min, isocratic 100% A. The separated molecules were analysed in both positive and negative ionization modes in the same run. The mass spectra were collected using resolving power 35,000 Full width at half maximum (FWHM) for the theoretical mass to charge ratio (m/z) 200. Full scan mass spectra were acquired in the 80−1000 m/z range. The ionization source parameters for positive and negative ion mode were as follows: capillary temperature 320°C, spray voltage 3.5 kV, sheath gas 30 (arbitrary units), auxiliary gas 8 (arbitrary units), probe heater temperature 310°C and S-Lens RF level was set at 55 v. except MS/MS experiments were performed using higher energy collision induced dissociation (HCD) and the normalized collision energy (NCE) applied was ramped from 10 to 40%. To ensure a good repeatability of the analysis, a quality control sample (QC) was formed by pooling a small aliquot of each biological sample. The QC sample was analysed intermittently (1 out of every 5 samples) for the duration of the analytical study to assess the variance observed in the data throughout the sample preparation, data acquisition and data pre-processing steps. In order to preliminarily evaluate the analytical variability of UPLC-HRMS analyses, the variation of the chromatographic pressure during the run, along with variability of the retention time and intensity of a randomly selected ion were calculated between the first and last analysis. The evaluation of these three parameters was carried out using

all the QC analyses. In order to verify the compliance of these three parameters, they were compared with compliant values set by the laboratory. The biological variability was measured during the data processing for each variable and evaluated during the statistical analysis.

Data processing and analysis

All the raw data generated by the LC-MS were converted to mzXML by ProteoWizard (version 2.0), then processed by XCMS, an open source package written in R for high throughput omics data analysis (Smith *et al.*, 2006). The data processing was performed in different steps. The first step was peak picking and required the application of the 'centWave' algorithm, the method of choice for processing centroided data acquired by HRMS. During this first step, the algorithm identifies regions of interest (ROIs) by combining consecutive centroids within a tolerated *m/z* deviation, defined by the parameter 'ppm', then the chromatographic peaks are built within the ROIs by applying other parameters such as peak width. The centwave method is fully described in Tautenhahn *et al.* (2008). The peak picking step is followed by retention time correction performed by the obiwarp algorithm and grouping. After XCMS, the data were further processed to eliminate artifacts (manual inspection of peak shape) and false peaks (peaks that do not appear with Gaussian shape). Indeed, the metabolites present in the blanks with an intensity ratio greater than 2 compared with those present in the samples were eliminated from the list. Additional filtrations were then carried out such as the elimination of the features without a Gaussian shape or the elimination of the features that represent 30% of the deviation from the median by applying the relative standard deviation filtration (a coefficient of variation less than 30% on QC samples peaks was applied to all samples). Analytical drift was corrected according to the linear correction algorithm developed by Van der Kloet *et al.* (2009). Simca-P software (version 14, Sartorius Stedim Biotech, Aubagne, France) was used to perform principal component analysis (PCA), hierarchical ascendant classification (HAC), and partial least squares-discriminant analysis (PLS-DA). The statistical significance of the PLS-DA classification model was assessed using a permutation test and cross-validation ANOVA (CV-ANOVA). Permutation testing consists of changing randomly the order of the rows in the data set so the class labels are assigned randomly to the measurements. The classification model is then recalculated using this permuted data set. A permutation test verifies the null hypothesis, namely that a given classification model is not significant and describes noise. If the null hypothesis is true, there should be no difference in the value of the quality-of-fit criteria between the original data set and the permuted one (Westerhuis *et al.*, 2008). The predictive capacity of the model is evaluated by the factor Q2. PermutMatrix (Caraux and Pinloche, 2005) was used to generate heatmaps, a graphical representation of data where the samples are clustered according to the proximity of their metabolome (intensity metabolites). MetaboAnalyst (Xia *et al.*, 2009) was used for the determination of *P*-values, and Cytoscape (Shannon *et al.*, 2003) was used to generate the metabolic network. The identification of the discriminating variables is performed in three steps. During the first step, the accurate mass of the compound is searched in metabolite databases such as Riken, knapsack, or literature. The metabolites suggested by the databases must be filtered by considering the physicochemical characteristics of the molecules as well as those

belonging to the same studied species. The level of identification at this stage corresponds to level 3 according to the metabolomics identification task group (Sumner *et al.*, 2014). The second identification step consists of fragmenting the candidate ion and comparing its fragmentation profile with that of the molecules suggested by the databases. If the profiles are the same then the identification level corresponds to level 2 (Sumner *et al.*, 2014). Level 1 identification is achieved once the searched metabolites occur at the same retention time as a model molecule corresponding to a purified metabolite which was suggested by the databases (Dunn *et al.*, 2013).

We have validated the above LC-MS analytical workflow and spectral data processing elsewhere (Martin *et al.*, 2015)

Results and Discussion

Multivariate analysis of the data

Untargeted metabolomics

The chromatographic conditions used in this study allowed the observation of chromatographic peaks with a good resolution (Fig. S2 in Supplementary Material). The PCA performed before and after filtrations based on the false peaks are presented in Fig. S3. The PCA performed after all filtrations steps are presented in Fig. S4.

We thus first used PCA to determine and visualize the impact of the environment and the genotype on the expression of the metabolome. From the 59 samples, a total of 963 analyses in both positive and negative mode were retained after the XCMS deconvolution processing and post-processing steps. PCA analysis allowed explanation of 74.2% of the total variance in 12 principal components. The first three components explained 36.5% of the total variance. The visualization of the data was performed only in two-dimensional component space using a score plot. Different clusters can be observed according to the component selected. When selecting the first and second component several clusters can be observed (Fig. S5a in Supplementary Material) and correspond to a group of locations such as Ille-et-Vilaine, Meurthe-et-Moselle, Loiret and Somme, that gathers different varieties. This shows a significant environmental effect. Another cluster was observed when selecting the second and the third component (Fig.e S5b). This cluster gathers the Pamela cultivar across five locations out of six and was distinctly separated from all the other samples, which shows a strong influence of the genotype for this particular cultivar and a resilience to the environment. Other clusters were also observed when selecting the sixth and seventh component (Fig. S5c). In this case, the clusters include the Bonanza and DK-Extorm cultivars and showed the same genotype influence as observed for the Pamela cultivar. Nevertheless, the visualization of the data in two-dimensional component space makes the information scattered and it is quite challenging to have an exhaustive view of the distribution of the samples without considering the twelve components. To overcome this, a hierarchical clustering analysis (HCA) was applied from the PCA scores. This classification method thus takes into account the contribution of all the components (12 principal components) of the PCA and returns it into a single graph. The results obtained are shown in Fig. 1, in which eight main clusters can be observed. Among them, three consist of grouped cultivars across different locations: the first cluster observed gathers the Pamela cultivar across four locations: Vendée, Ille-et-Vilaine, Somme and Marne (cluster E). A second
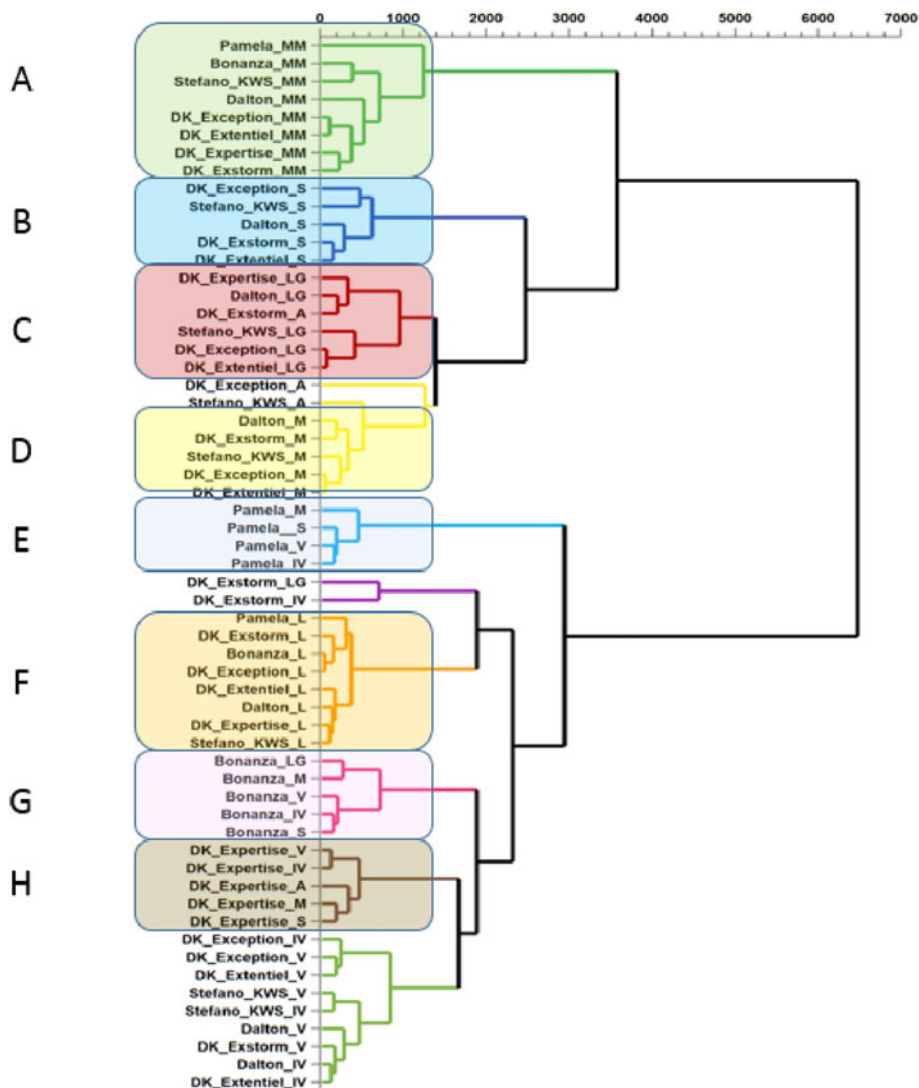
Fig. 1. Hierarchical Ascendant Classification (HAC) calculated from the 12 principal components of the PCA model.

cluster gathered the Bonanza cultivar across five locations: Lot-et-Garonne, Marne, Vendée, Ille-et-Vilaine, Somme (cluster G). Another cluster was also observed and corresponds to the DK-Expertise cultivar which is gathered across five out of the eight locations (cluster H). Together with the Bonanza cultivar, the DK-Expertise cultivar seemed to be the least affected by the environment. On the other hand, other cultivars seem to be very sensitive to the cultivation location. For instance, the Dalton cultivar was present in all clusters formed mainly from geographical locations, indicating a great sensitivity of that cultivar to the environment and a slight genotype influence. Interestingly, some cultivation locations exerted a dramatic environmental influence on the metabolome of most of the cultivars. In fact, two major clusters are distinctly observed in which all the varieties are gathered across the same locations. These two major locations correspond to Meurthe-et-Moselle and Loiret (clusters A and F). Other clusters can also be observed where the environment has a major impact such as Lot-et-Garonne and Somme (clusters C and B). These observations pinpoint the major role of some environments over the genetic background on the expression of the seed metabolome, and thereby seed quality.

Correlation network analysis was performed as an additional exploratory tool to assess the heterogeneity between different varieties and their interaction with the environment. This analysis is in agreement with the HCA results and it also shows different clusters corresponding to all the varieties gathered according to the same locations they were cultivated in, and therefore they also revealed a slight genetic influence and a very strong environmental effect (Fig. S6 in Supplementary Material). Such representation allowed us to better emphasize and visualize the impact of these two factors on the seed metabolome.

Although the rapeseed cultivars may display a different resilient effect with regard to the cultivation location, our main results indicated that the environmental impact is the determining factor in the expression of the metabolome of the studied varieties. In addition, two distinct locations have largely contributed to offset the genetic difference between the varieties. We therefore focused on metabolites that drove the environmental differences among the two main metabolome impacting regions (Meurthe-et-Moselle and Loiret). This comparison was carried out using a PLS-DA. Indeed, PLS-DA is a discriminating statistical analysis that optimizes the intergroup differences, by rotating PCA components in order to maximize the separation among
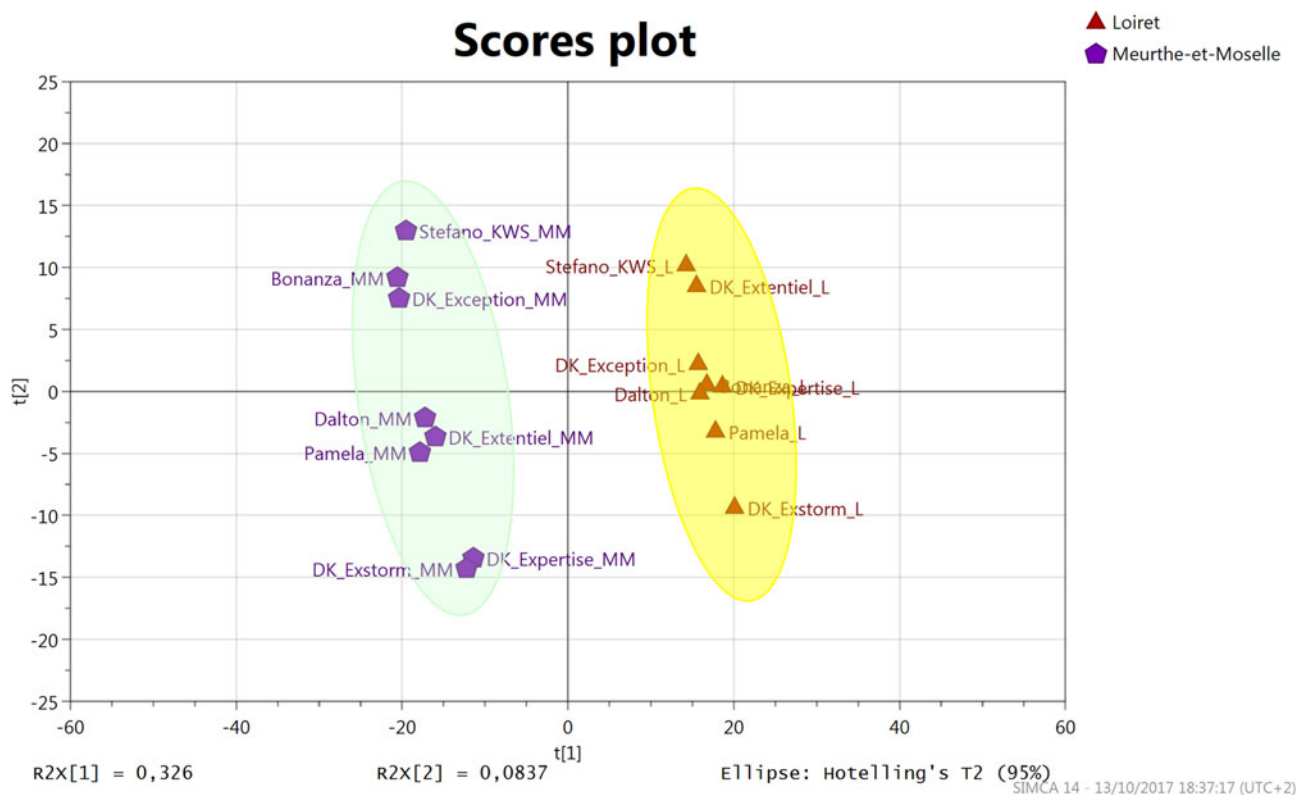
Fig. 2. Partial Least Squares Discriminant Analysis (PLS-DA) of the cultivar seeds metabolome grown in either Meurthe-et-Moselle or Loiret regions. Both regions are used as a class setting in the supervised model.

classes, and sheds light on variables carrying the class-separating information. The statistical model obtained from the PLS-DA allowed a very significant class separation between the two locations (Meurthe-et-Moselle and Loiret) (prediction factor Q2 of 0.96, a P-value of 1.94×10^{-7} after CV-ANOVA, and Q2 values of -0.11 after permutation) (Fig. 2). The identification of the metabolites was performed initially depending on the importance of the PLS partial correlation coefficient. In order to represent the significance of all the impacted molecules between the Loiret region and Meurthe-et-Moselle as described below, an ANOVA followed by Wilcoxon *post-hoc* test, a non-parametric statistical test, was also performed. All the P-values of the most significantly different metabolites along with their fold change between the two regions are listed in Table 1. Of note is that most of metabolites affected by the locations are secondary metabolites, which are produced in response to environmental adaptation.

Identification of the discriminating metabolites
All the metabolites identified in this study are of level 2 according to Sumner *et al.* (2014). The first discriminant metabolite identified corresponds to glucobrassicanapin. This glucosinolate is an aliphatic glucosinolate and was characterized by a $[M-H]^-$ at m/z 386.0582 with a fragment ion at m/z 96.95 which corresponds to the HO_4S^{-1}, a typical ion fragment which exhibits the common fragmentation behaviour of glucosinolates. The product ions at m/z 275, 259, 208, 144 (Fig. 3A) have already been described in Lelario *et al.* (2012). It seems that the Loiret region is more favourable for the production of this glucosinolate than the Meurthe-et-Moselle region. For both regions, this environmental impact is expressed similarly for all the varieties as

shown in Fig. 3B. The second metabolite identified was also an aliphatic glucosinolate and corresponds to gluconapin which is characterized by a $[M-H]^-$ at m/z 372.0425. The fragmentation pattern obtained from this glucosinolate has also been described in Cataldi *et al.* (2010) (Fig. 3C). This glucosinolate is more expressed in the Loiret region than in Meurthe-et-Moselle and this observation is reported similarly for all the varieties (Fig. 3D). Another major glucosinolate also identified among the discriminant variables was progoitrin. In fact, progoitrin is an aliphatic glucosinolate and is considered along with gluconapin as the most prevalent glucosinolate in *B. napus* (Fang *et al.*, 2012) as well as in other Brassicacea. Progoitrin was characterized by a $[M-H]^-$ at m/z 388.0375. This glucosinolate was also more expressed in Loiret than it was in Meurthe-et-Moselle. The results also showed that the Loiret region is favourable for the accumulation of other aliphatic glucosinolates, such as glucoalyssin, gluconapoleiferin, glucocochlearin and sinigrin, but also indole glucosinolates such as 4-hydroxyglucobrassicin (Table 1 and Fig. S7). The results described above seems to be in agreement with other studies where the authors reported how the accumulation of glucosinolates in seeds was modified by environmental factors. For instance, Jensen *et al.* (1996) Jensen *et al.*, 1996 reported that the level of glucosinolates in seeds of *Brassica napus* L. was dependent on the soil properties especially in response to drought. Indeed, by controlling irrigation of *B. napus* L. in field trials, these authors showed that (i) the glucosinolate content in seeds was 2-fold higher in plants exposed to drought during the vegetative, flowering or pod filling stages, and (ii) the accumulation of glucosinolates in seeds was more important for water-stressed plants grown in loamy soil than in

Table 1. Discriminant metabolites when comparing Meurthe-et-Moselle (MM) and Loiret region (L)

Metabolites	[M-H]⁻	[M + H]⁺	Fragments	P-value (Wilcoxon rank-sum test)	FDR	Log2 FC (L/MM)	identification level
1-O-sinapoyl-β-D-glucose	385.1139	—	223; 205	0.0001554	0.00053872	−0.5748	2
4-Hydroxyglucobrassicin	463.0486	—	285; 267; 259; 221; 97	0.0001554	0.00053872	0.4925	2
Caffeic acid glucoside	341.0876	—	179; 161; 135	0.0001554	0.00053872	0.6702	2
Caffeyl alcohol 3-4-O-hexoside	327.1085	—	165; 147; 129	0.0001554	0.00053872	1.0300	2
Glucobrassicanapin	386.0582	—	275; 259; 208; 195; 97	0.0001554	0.00053872	1.1246	2
Gluconapin	372.0425	—	259; 195; 130; 97	0.0001554	0.00053872	0.8499	2
L-Glutathione	306.0767	—	288; 272; 254; 210; 143	0.0001554	0.00053872	2.0198	2
Malic acid	133.0141	—	115	0.0001554	0.00053872	−0.5592	2
Sinapoyl ester	577.1562	—	353; 223	0.0001554	0.00053872	0.7876	2
Sinapoyl malate	339.0717	—	223; 164; 149; 115; 71	0.0001554	0.00053872	−0.5701	2
Adenine	—	136.0618	119; 97; 94	0.0001554	0.00053872	−1.6063	2
Adenosine	—	268.104	136	0.0001554	0.00053872	−1.5663	2
p-Aminobenzoic acid	—	138.0551	120; 94; 77	0.0001554	0.00053872	−0.3779	2
Gluconapoleiferin	402.0531	—	275; 259; 224; 195; 97	0.0003108	0.00085061	0.8829	2
Progoitrin	388.0375	—	332; 308; 275; 259; 97	0.0003108	0.00085061	0.4586	2
Tryptophane	—	205.0973	188; 170; 159; 146; 144	0.0003108	0.00085061	−0.5666	2
Phenylalanine	—	166.0864	120; 103	0.0006216	0.0016162	−0.7506	2
Sinigrin	358.0274	—	259; 195; 181; 97	0.0010878	0.0026936	0.4869	2
Hydroxycinnamic acid derivative	723.2147	—	499; 259; 223	0.006993	0.015152	0.6784	2
Vanilloylcholine 4-O-hexoside	—	416.1915	357; 254; 195	0.006993	0.015152	0.4799	2
1,2-Disinapoylgentiobioside	753.2246	—	223; 205	0.014763	0.029526	0.3795	2
glucoalyssin	450.057	—	386; 97	0.014763	0.029526	0.9673	2
1-O-(4-coumaroyl)-β-D-glucose	325.0928	—	163; 145	0.020668	0.039805	−1.0007	2

sandy soil. Recent work has also demonstrated that quantitative and qualitative composition of glucosinolates in seeds of *B. napus* L. were significantly modified by the mineral composition of soil. For instance, an increase of sulfur (Jankowski *et al.*, 2008) or boron (Jankowski *et al.*, 2016) fertilization led to increased content of progoitrin in seeds of *B. napus* L. Interestingly, in our study, the sites of Loiret and Meurthe-et-Moselle differ mainly by their soil characteristics (Table S1; Loiret: flinty loam; Meurthe-et-Moselle: calcacerous clay). Thus, regarding previous work and data presented in our study, it appears that it would be relevant to consider environmental factors such as soil properties (including the mineral composition) in relationship with abiotic factors (temperature, rainfall) to influence the glucosinolate content in seeds of *B. napus* L.

Among the discriminating metabolites, several polyphenols were also identified. Hydroxycinnamic acids were the major

phenolic compounds with the most discriminating power. The hydroxycinnamic acids are a class of non-flavonoid phenols characterized by the C6-C3 structure. These compounds are abundant in plants and are used for both structural and chemical defence purposes. It is also known that in Brassica vegetables the most common are sinapic, *p*-coumaric and ferulic acids, often found in conjunction with sugar or other hydroxycinnamic acids (Cartea *et al.*, 2011). Based on the statistical analysis, 1-O-sinapoyl-beta-D-glucose is one of the most discriminating hydroxycinnamic derivatives. This phenolic compound was characterized by a [M-H]⁻ at *m/z* 385.1139. The product ion at *m/z* 223.0606 corresponds to sinapic acid and is a typical ion fragment which exhibits the common fragmentation behaviour of the sinapic acid derivatives (Fig. 4A). It is very important to note that 1-O-sinapoyl-beta-D-glucose is one of the major phenolic compounds found in *B. napus* besides sinapine and sinapic acid
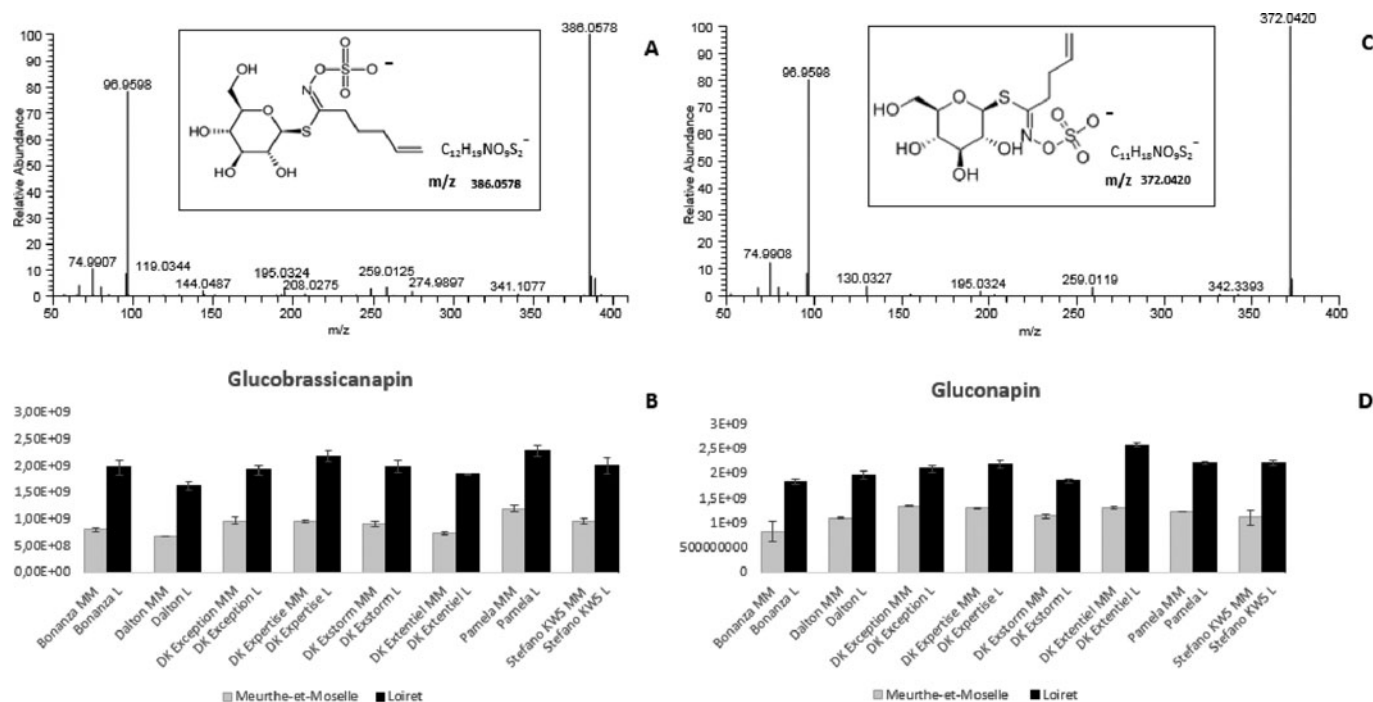
Fig. 3. MS/MS spectrum of glucobrassicanapin and gluconapin (A,C). Relative abundance of glucobrassicanapin and gluconapin in the Loiret and Meurthe-et-Moselle regions (B,D).

(Khattab *et al.*, 2010). PLS-DA results show that 1-O-sinapoyl-beta-D-glucose was more expressed in the region of Meurthe-et-Moselle than in the Loiret region and this observation is reported similarly for all the varieties (Fig. 4B). Another sinapic acid derivative identified among the discriminating molecules was sinapoyl malate (Fig. 4C). This phenolic compound is characterized by [M-H]⁻ at *m/z* 339.0717 and a product ion at *m/z* 223.0606 which corresponds to sinapic acid. Similar to 1-O-sinapoyl-beta-D-glucose, the PLS-DA results show that sinapoyl malate was also more expressed in Meurthe-et-Moselle than in Loiret (Fig. 4D). Another discriminating metabolite identified was 1-O-(4-coumaroyl)-β-D-glucose. This coumaric acid derivative was characterized by [M-H]⁻ at *m/z* 325.0928 and a product ion at *m/z* 163 which corresponds to coumaric acid. The results show that Meurthe-et-Moselle is also more favourable than Loiret for the expression of 1-O-(4-coumaroyl)-β-D-glucose (Fig. S7, in Supplementary Material).

From an agronomic point of view, these results are very interesting for the following reasons. Not only was the Meurthe-et-Moselle region favourable for the expression of some of the major phenolic compounds, but it was also unfavourable for the production of the predominant glucosinolates considered to be undesirable molecules for livestock as their derived compounds interfere with the synthesis of thyroid hormones, leading eventually to hypothyroidism and enlargement of the thyroid gland (goitre) (Fenwick and Heaney, 1983). They are also considered anti-nutritional by decreasing protein absorption in cattle (Mawson and Heaney, 1994). Nevertheless, the PLS-DA results show that the Loiret region could also be favourable to the expression of some minor hydroxycinnamic acid derivatives such as 1,2-Disinapoylgentiobioside or feruloyl sinapoyl hexose which were characterized by [M-H]⁻ at *m/z* 753.2246 and 561.1615, respectively (Fig. S7). The results showed also that there were no flavonoids among the discriminating molecules,

such as sinapine which is the most predominant phenolic compound in *B. napus* (Fang *et al.*, 2012). The main difference was seen for secondary metabolites, in which the ratio of phenolic compounds/glucosinolates varied reversely between the two locations. We thus demonstrated that some cultivation conditions can be overwhelming over genetic influence for seeds quality in *B. napus*, and that metabolomics is well suited to highlight such differences.

Semi-targeted metabolomics

All the results thus far discussed demonstrated how non-targeted metabolomics facilitates investigation into the impact of both genotype and environment on the metabolome of *B. napus* without any prior knowledge of the identity of the molecules. Also, of great interest is the fact that the major impacted molecules are secondary metabolites, among which we found mostly the phenolic compounds and glucosinolates. Nevertheless, neither all the phenolic compounds nor the minor glucosinolates were involved in the discriminating results when comparing the most impacting environment. If the objective is to focus on animal or human health, then a semi-targeted analysis, focusing only on health-promoting compounds such as secondary metabolites, would be more appropriate to determine the most stable or favourable candidate (variety) for the expression of these metabolites (Dunn *et al.*, 2013). In order to achieve this second purpose, i.e. the determination of the most stable or favourable candidate for health-promoting molecules, a heatmap was generated by comparing all the samples according to secondary metabolites where the phenolic compounds and glucosinolates represented the main molecules. The heatmap is a commonly used visualization tool for metabolomic data where the relative abundance of ions detected in each sample is represented with colour intensity (Ivanisevic *et al.*, 2015) and provides a global perspective of metabolite changes in response to genetic background and
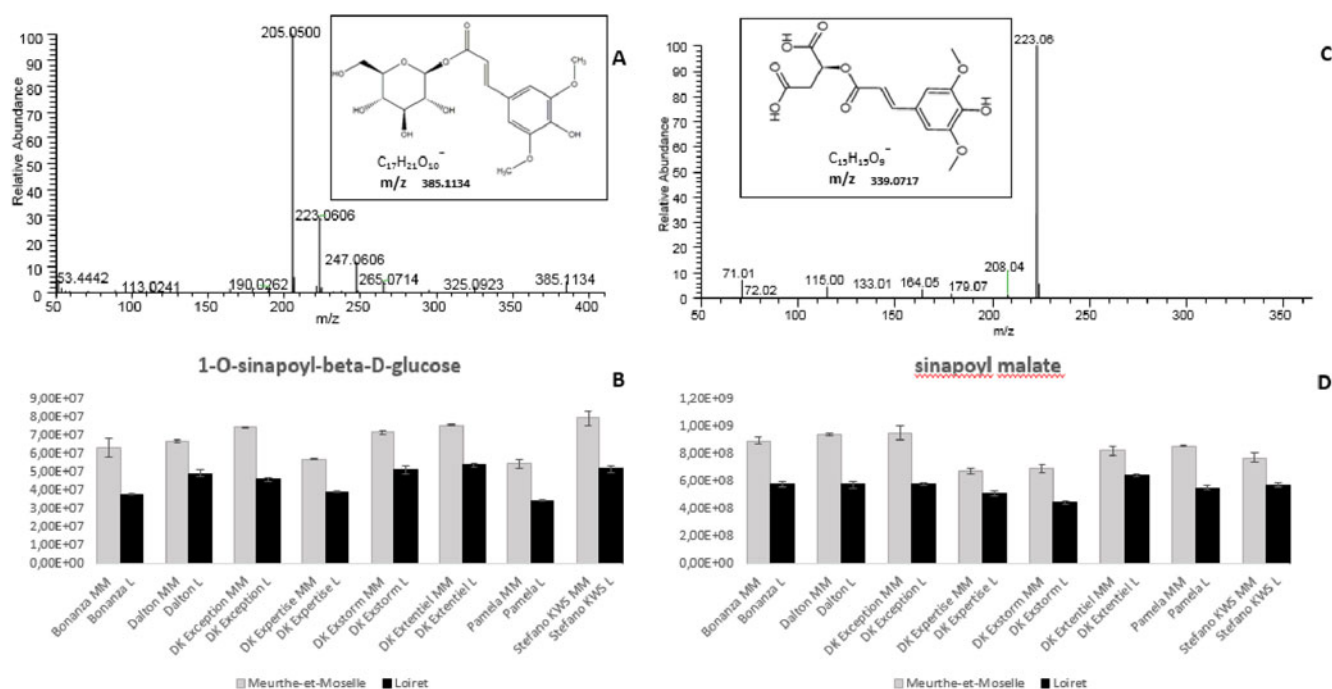
Fig. 4. MS/MS spectrum of 1-O-sinapoyl-beta-D-glucose and sinapoyl malate (A,C). Relative abundance of 1-O-sinapoyl-beta-D-glucose and sinapoyl malate in the Loiret and Meurthe-et-Moselle regions (B,D).

environment. As shown in Fig. 5, the samples are gathered according to the similar patterns of the metabolites. Interestingly, some varieties behave similarly across different locations and are grouped in a distinct cluster, meaning that their level of glucosinolates and phenolic compounds remain similar even when cultivated in different environments. These clusters are Dalton, Pamela, Bonanza and DK-Expertise (Fig. 5a–d). The observation described below shows that the above-mentioned cultivars consistently tended to express these kinds of major health-impacting molecules. This stability is more or less relevant depending on the variety. In fact, among the above-mentioned varieties, the Dalton cultivar seems to be the most environment-resistant cultivar regarding the expression of glucosinolates and phenolic compounds. The heatmap shows also that there is a slight environmental effect when considering only the major secondary metabolites. In fact, the Loiret region seems to be the major impacting environment followed to a lesser extent by Meurthe-et-Moselle and Lot-et-Garonne (Fig. 5e–g). Overall, our graph analysis provided an overview of how metabolite levels changed due to genetic background or environment. Furthermore, not only does it provide an overview of similarities between the samples under analysis, but the metabolite pattern throughout the data set can also be obtained, thereby suggesting possible metabolic relationships (Widodo et al., 2009). Almost all these observations agree with the HCA results described earlier with few exceptions, such as the Dalton cultivar. The semi-targeted metabolomics focused on secondary metabolites and displayed in the heatmap, revealing that the Dalton cultivar behaved consistently and is irrespective of environmental influences about major secondary metabolites, while the non-targeted metabolomics revealed a sensitivity to the environment for this same cultivar. When combining these two approaches, the environment could affect pathways other than secondary metabolism, which may explain why the Dalton cultivar was spread across the

hierarchical ascendant classification when performing the non-targeted approach. Therefore, the semi-targeted and non-targeted approaches are complementary approaches in providing information on which scale of the metabolism (overall, or primary or secondary) is mainly associated to environmental or genetic influences.

Glucosinolates, phenolic compounds and metabolic pathway

The results described above showed how both the environment and the genetic background could influence the expression of the secondary metabolites. From a biological point of view, these observations agree with previous studies that describe the secondary metabolites as a natural defence system which is induced in response to biotic or abiotic stress, in connection to the genetic background. Among these secondary metabolites, the glucosinolates and the phenolic compounds are the major metabolites involved in the defence system of the brassica family (Jahangir et al., 2009). The glucosinolates are synthesized from a variety of typical amino acids (methionine, tryptophan and phenylalanine) and are responsible for diverse physiological effects such as inhibitors of microbial growth, attractants for particular insects, deterrents of different herbivores and as flavouring compounds (Petersen et al., 2002). This biological effect is predominantly related to the glucosinolate-derived compounds which correspond to breakdown products released after the hydrolysis of the intact form of the glucosinolates by the myrosinase upon tissue damage. Although some of these derived products have toxic effects on farm animals, it is now very well established that some of them can be beneficial to health such as supporting anti-cancer activity (Keum et al., 2004). As described earlier, the statistical analysis performed on our samples revealed a significant environmental impact on the accumulation of the major glucosinolate in B. napus which corresponded to the
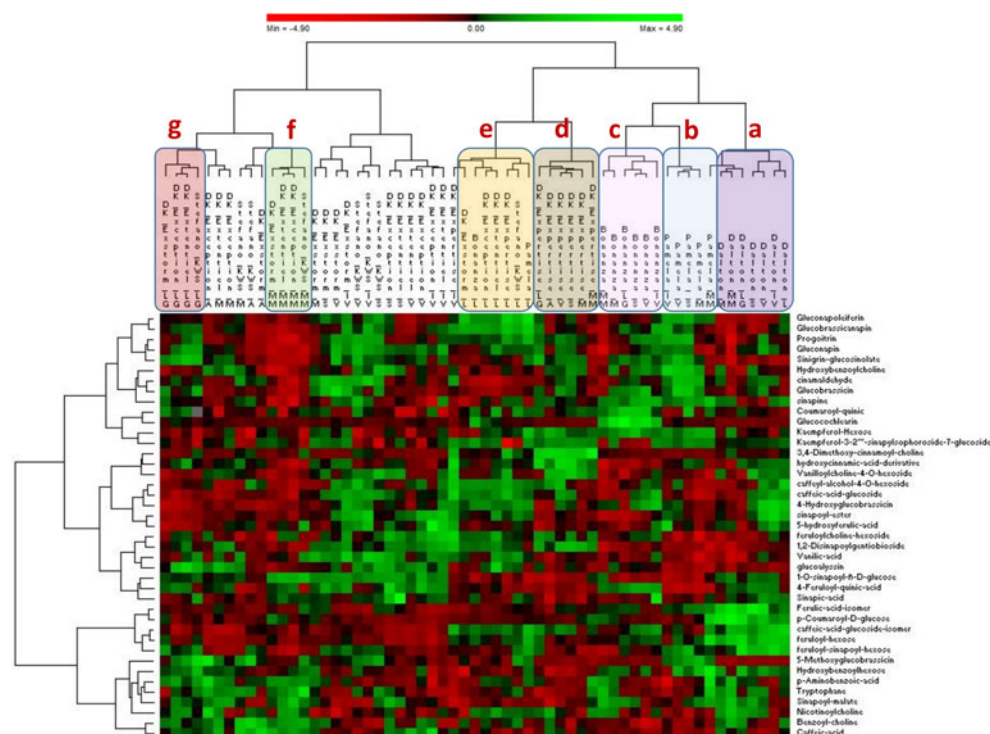
Fig. 5. The heatmap is a graphical representation of data where the samples are clustered according to the proximity of their metabolome (intensity metabolites). The green areas correspond to over-expressed metabolites, while the red areas are the under-expressed metabolites.

gluconapine and progoitrin. This observation is in agreement with other studies performed on *B. napus* (Jensen *et al.*, 1996; Jankowski *et al.*, 2008, 2016). Furthermore, these two glucosinolates have been correlated to bitterness in brassicacea (Kim *et al.*, 2010); however, unlike other aliphatic glucosinolates, nothing is known about the anti-cancer properties of these glucosinolates. Therefore, an attractive objective would be to reveal the possible health beneficial effects of the above-mentioned glucosinolates and then use them in functional food products or pharmaceuticals which may provide additional interesting routes of rapeseed by-products valorization.

The phenolic compounds represent the second largest group among the secondary metabolites in the brassica family. They are derived from the aromatic amino acid phenylalanine which is produced via the shikimic pathway. In the last two decades, there has been a strong interest in the biological effects of phenolic compounds in dietary plants. Besides the multiple roles in plants such as attracting insects for seed dispersion and pollination or hormone controllers, phenolic compounds are also part of the natural defence system and are responsible for diverse physiological effects against insects, bacteria, viruses and fungi (Lattanzio *et al.*, 2006). Moreover, the phenolic compounds have been intensively explored in recent years because of their potential health-promoting effects. Thus, several beneficial properties for human health have been reported such as antimicrobial, anti-allergic, anti-inflammatory, vascular protection and cytotoxic anti-tumor activity and most importantly an antioxidant activity (Pandey and Rizv, 2009; Huang *et al.*, 2016). This information is very useful for developing new cultivars with an appropriate phenolic compound/glucosinolates profile, from which high-quality added value products can be produced. However, the manipulation of the content of phenolic compounds and glucosinolates in *B. napus* by the environment or cultural practices is very challenging and requires a deep understanding of the activation and the regulation of the defence mechanisms. One way to achieve this purpose is to map the detected metabolites into metabolic pathways by using informatics tools such as the Kegg mapper Atlas (Kanehisa *et al.*, 2017). This tool can represent the entire metabolism of *B. napus* and allow not only the visualization of all the discriminating compounds by highlighting them according to their expression level, but also show all the enzymes and genes responsible for these observed modifications. As most of the identified molecules are among the hydroxycinnamic derivatives, we have focused only on the phenylpropanoid pathway. As shown in Fig. 6, the non-discriminant metabolites we identified are highlighted in green, whereas the discriminant ones are highlighted in red. When looking at the metabolic map we noticed that the impacted point starts from 1-O-sinapoyl-beta-D-glucose as all the upstream molecules seem to be unmodified (coloured in green). Therefore, this observation could be explained either by an increase of its precursors which correspond to the sinapate and glucose or by a modification in the expression of the gene coding for the sinapate 1-glucosyltransferase. This enzyme belongs to the UGT84A subfamily of plant glycosyltransferase family 1 and is designated UGT84A2 (Sinlapadech *et al.*, 2007). The results obtained from the statistical analyses show that the two precursors could not be responsible for the modification of the expression of 1-O-sinapoyl-beta-D-glucose as these metabolites were not impacted (*P*-value greater than 0.05). This could be explained by a modification of the activity of the sinapate glucosyl transferase gene upon environmental influences. This is in agreement with other studies where the authors described how a stress-responsive glucosyltransferase belonging to the subfamily of
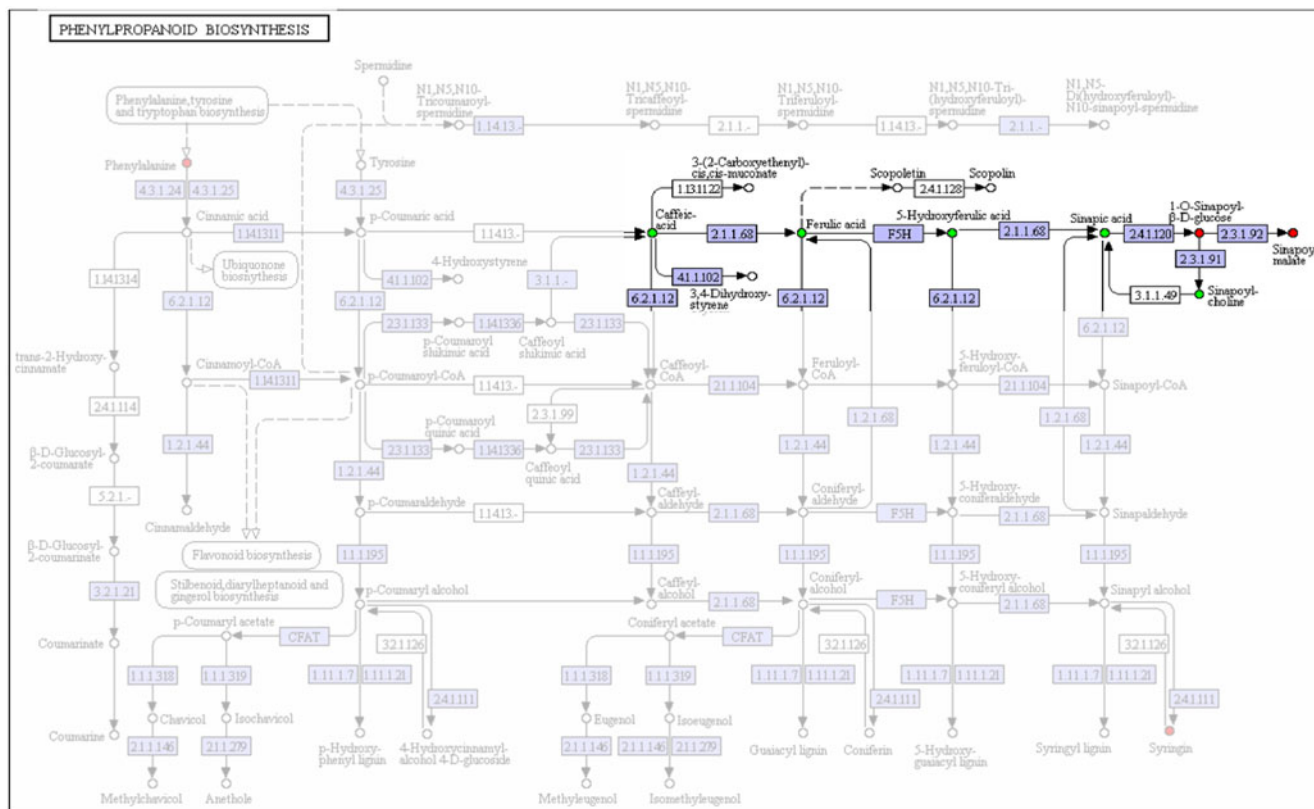
Fig. 6. Metabolic pathways (Kegg Mapper), green: normally expressed; red: impacted by the environment.

UGT84A played a crucial role in modulating phenylpropanoid synthesis, modification, bioactivity, and/or stability in response to environmental cues (Babst *et al.*, 2014). Moreover, the metabolic map shows also that the other impacted metabolite is the sinapoyl malate (coloured in red). This metabolite corresponds to the end product of the phenylpropanoid pathway and is generated from the conversion of the 1-O-sinapoyl-beta-D-glucose under the action of sinapoylglucose:malate sinapoyltransferase. This enzyme belongs to the SNG subfamily and is designated as SNG1 (Lehfeldt *et al.*, 2000). This seems to be the preferred route of metabolism over the production of sinapoyl choline which remained unchanged. The observation described below seems to be in agreement with the modification of the 1-O-sinapoyl-beta-D-glucose suggesting that this modification of the sinapoyl malate could be the result of an accumulation of its precursor sinapoylglucose or the accumulation of the malic acid. The results obtained from the statistical analyses show that the malic acid appears among the discriminating molecules ($P = 1.55 \times 10^{-4}$). This observation is in agreement with other studies describing how the expression of the malate during the Krebs cycle is highly modified under the stress conditions (Lehmann *et al.*, 2015). This observation could be the origin of a chain reaction that led to the accumulation of the sinapoyl malate. Nevertheless, the modification of the sinapoyl malate by the accumulation of the 1-O-sinapoyl-beta-D-glucose seems to be the most plausible explanation. All these observations show how metabolomics is capable of generating a new regulating hypothesis at the enzyme and/or gene expression level, and could help selecting cultivar and/or cultivation conditions to promote one pathway over another to control the content of some

anti-nutritive compounds or health-promoting molecules in the *B. napus* seeds.

In conclusion, we have applied the metabolomic approach to reveal the impact of gene × environment interactions on the quality of rapeseed. We were able to determine that some varieties are very sensitive to environmental conditions such as soil characteristics, while others are more resilient for certain traits. We also demonstrated in this study that the semi-targeted metabolomics is a great approach to allow an overview of how the glucosinolates and the phenolic compounds are regulated in all the varieties across the different locations in one analysis. This helps to the identification of varieties with high and stable glucosinolate or phenolic compound content. It allows identification of target genes and helps designing/selecting new varieties with desired characteristics, either useful for human health or livestock feeding. We believe that this is an innovative paradigm as it reverses the downward vision of the gene to the prevailing phenotype in favour of a top-down approach starting from the phenotype to identify the target genes. Reversely, it can be also used to seek and control the main contributing environmental factors to obtain the desired seed quality.

Although metabolomics has been used in other plants with a similar design (Asiago *et al.*, 2012), this is the first study of that kind performed on *Brassica napus* seeds. It was thus necessary to first get an overview of the impact of the main 'raw' factors on the seeds metabolome, e.g. the region and cultivar (genetics). This allowed us to draw reliable conclusions about the variation or stability of the polar fraction of the different varieties over several regions gathering high contrast in terms of climate and nature of soil. From that first step, it is now essential to design new studies

addressing the quality traits over multiple growth and several harvest cycles and to refine intra-regional variability. Careful experimental designs using well-controlled environments (drought, light, heat, soil and minerals) addressing plant-to-plant variations can also provide meaningful information and give complementary mechanistic explanations with regard to quality traits.

References

Asiago VM, Hazebroek J, Harp T and Zhong C (2012) Effects of genetics and environment on the metabolome of commercial maize hybrids: a multisite study. *Journal of Agricultural and Food Chemistry* **60**, 11498–11508.

Babst BA, Chen HY, Wang HQ, Payyavula RS, Thomas TP, Harding SA and Tsai CJ (2014) Stress-responsive hydroxycinnamate glycosyltransferase modulates phenylpropanoid metabolism in *Populus*. *Journal of Experimental Botany* **65**, 4191–4200.

Baux A, Hebeisen T and Pellet D (2008) Effects of minimal temperatures on low-linolenic rapeseed oil fatty-acid composition. *European Journal of Agronomy* **29**, 102–107.

Caraux G and Pinloche S (2005) PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* **21**, 1280–1281.

Cartea ME, Francisco M, Soengas P and Velasco P (2011) Phenolic compounds in Brassica vegetables. *Molecules* **16**, 251–280.

Cataldi TRI, Lelario F, Orlando D and Bufo SA (2010) Collision-induced dissociation of the A + 2 isotope ion facilitates glucosinolates structure elucidation by electrospray ionization-tandem mass spectrometry with a linear quadrupole ion trap. *Analytical Chemistry* **82**, 5686–5696.

Dunn WB, Erban A, Weber RJM, Creek DJ, Brown M, Breitling R, Hankemeier T, Goodacre R, Neumann S, Kopka J and Viant MR (2013) Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* **9**, 44–66.

Fang J, Reichelt M, Hidalgo W, Agnolet S and Schneider B (2012) Tissue-specific distribution of secondary metabolites in rapeseed (*Brassica napus* L.). *PLoS ONE* **7**, 1–8.

Farag MA, Sharaf Eldin MG, Kassem H and Abou El Fetouh M (2013) Metabolome classification of *Brassica napus* L. organs via UPLC-QTOF-PDA-MS and their anti-oxidant potential. *Phytochemical Analysis* **24**, 277–287.

Fenwick GR and Heaney RK (1983) Glucosinolates and their breakdown products in cruciferous crops, foods and feedingstuffs. *Food Chemistry* **11**, 249–271.

Fenwick GR, Spinks AE, Wilkinson PA, Heaney KR and Legoy AM (1986) Effect of processing on the antinutrient content of rapeseed. *Journal of the Science of Food and Agriculture* **37**, 735–741.

Fiehn O (2001) Combining geonomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics* **2**, 155–168.

He F, Pan QH, Shi Y and Duan CQ (2008) Biosynthesis and genetic regulation of proanthocyanidins in plants. *Molecules* **13**, 2674–2703.

Hong J, Yang L, Zhang D and Shi J (2016) Plant metabolomics: an indispensable system biology tool for plant science. *International Journal of Molecular Sciences* **17**, E767.

Huang W, Zhang X and Chen W (2016) Role of oxidative stress in Alzheimer's disease (review). *Biomedical Reports* **4**, 519–522.

Ivanisevic J, Benton HP, Rinehart D, Epstein A, Kurczy ME, Boska MD, Gendelman HE and Siuzdak G (2015) An interactive cluster heat map to visualize and explore multidimensional metabolomic data. *Metabolomics* **11**, 1029–1034.

Jahangir M, Abdel-Farid IB, Kim HK, Choi YH and Verpoorte R (2009) Healthy and unhealthy plants: the effect of stress on the metabolism of Brassicaceae. *Environmental and Experimental Botany* **67**, 23–33.

Jankowski K, Budzynski W and Szymanowski A (2008) Effect of sulfur on the quality of winter rape seeds. *Journal of Elementology* **13**, 521–534.

Jankowski KJ, Sokólski M, Dubis B, Krzebietke S, Żarczyński P, Hulanicki P and Hulanicki PS (2016) Yield and quality of winter oilseed rape (*Brassica napus* L.) seeds in response to foliar application of boron. *Agricultural and Food Science* **25**, 164–176.

Jensen CR, Mogensen VO, Mortensen G, Fieldsend JK, Milford GFJ, Andersen MN and Thage JH (1996) Seed glucosinolate, oil and protein contents of field-grown rape (*Brassica napus* L.) affected by soil drying and evaporative demand. *Field Crops Research* **47**, 93–105.

Jervis J, Kastl C, Hildreth SB, Biyashev R, Grabau EA, Saghai-Maroof MA and Helm RF (2015) Metabolite profiling of soybean seed extracts from near-isogenic low and normal phytate lines using orthogonal separation strategies. *Journal of Agricultural and Food Chemistry* **63**, 9879–9887.

Kanehisa M, Furumichi M, Tanabe M, Sato Y and Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**, D353–361.

Keum YS, Jeong WS and Kong AN (2004) Chemoprevention by isothiocyanates and their underlying molecular signaling mechanisms. Mutation Research – Fundamental and Molecular Mechanisms of Mutagenesis **555**, 191–202.

Khattab R, Eskin M, Aliani M and Thiyam U (2010) Determination of sinapic acid derivatives in Canola extracts using high-performance liquid chromatography. *Journal of the American Oil Chemists' Society* **87**, 147–155.

Kim JK, Chu SM, Kim SJ, Lee DJ, Lee SY, Lim SH, Ha SH, Kweon SJ and Cho HS (2010) Variation of glucosinolates in vegetable crops of *Brassica rapa* L. ssp. pekinensis. *Food Chemistry* **119**, 423–428.

Lattanzio V, Lattanzio VMT, Cardinali A and Amendola V (2006) Role of phenolics in the resistance mechanisms of plants against fungal pathogens and insects. Available at: https://www.researchgate.net/publication/303270594_Role_of_phenolics_in_the_resistance_mechanisms_of_plants_against_fungal_pathogens_and_insects

Lehfeldt C, Shirley AM, Meyer K, Ruegger MO, Cusumano JC, Viitanen PV, Strack D and Chapple C (2000) Cloning of the SNG1 gene of Arabidopsis reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. *The Plant Cell* **12**, 1295.

Lehmann MM, Rinne KT, Blessing C, Siegwolf RTW, Buchmann N and Werner RA (2015) Malate as a key carbon source of leaf dark-respired CO2 across different environmental conditions in potato plants. *Journal of Experimental Botany* **66**, 5769–5781.

Lelario F, Bianco G, Bufo SA and Cataldi TRI (2012) Establishing the occurrence of major and minor glucosinolates in Brassicaceae by LC-ESI-hybrid linear ion-trap and Fourier-transform ion cyclotron resonance mass spectrometry. *Phytochemistry* **73**, 74–83.

Lohani UC, Fallahi P and Muthukumarappan K (2015) Comparison of ethyl acetate with hexane for oil extraction from various oilseeds. *Journal of the American Oil Chemists' Society* **92**, 743–754.

Martin JC, Maillot M, Mazerolles G, Verdu A, Lyan B, Migné C, Defoort C, Canlet C, Junot C, Guillou C, Manach C, Jabob D, Bouveresse DJR, Paris E, Pujos-Guillot E, Jourdan F, Giacomoni F, Courant F, Favé G, Le Gall G, Chassaigne H, Tabet JC, Martin JF, Antignac JP, Shintu L, Defernez M, Philo M, Alexandre-Gouaubau MC, Amiot-Carlin MJ, Bossis M, Triba MN, Stojilkovic N, Banzet N, Molinié R, Bott R, Goulitquer S, Caldarelli S and Rutledge DN (2015) Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study. *Metabolomics* **11**, 807–821.

Mawson R and Heaney RK (1994) Rapeseed meal-glucosinolates and their antinutritional effects. *Part 5. Animal reproduction.* Die Nahrung **38**, 588–598.

Miller-Cebert RL, Sistani NA and Cebert E (2009) Comparative mineral composition among canola cultivars and other cruciferous leafy greens. *Journal of Food Composition and Analysis* **22**, 112–116.

Misra BB (2016) Cataloging the *Brassica napus* seed metabolome. *Cogent Food & Agriculture* **2**, 1–16.

Nicholson JK, Holmes E and Lindon JC (1999) 'Metabonomics': understanding the metabolic responses of living systems to path physiological stimul via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189.

Pandey K and Rizv S (2009) Plant polyphenols as dietary antioxidants in human health and disease. *Oxidative Medicine and Cellular Longevity* **2**, 270–278.

Petersen BL, Chen S, Hansen CH, Olsen CE and Halkier BA (2002) Composition and content of glucosinolates in developing *Arabidopsis thaliana*. *Planta* **214**, 562–571.

Scott JK, Vitou J and Jourdan M (1999) Review of the potential for biological control of wild radish (*Raphanus raphanistrum*) based on surveys in the Mediterranean region. 377–380.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 2498–2504.

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

Sinlapadech T, Stout J, Ruegger MO, Deak M and Chapple C (2007) The hyper-fluorescent trichome phenotype of the brt1 mutant of Arabidopsis is the result of a defect in a sinapic acid:UDPG glucosyltransferase. *Plant Journal* **49**, 655–668.

Smith CA, Want EJ, O'Maille G, Abagyan R and Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* **78**, 779–787.

Sumner LW, Lei Z, Nikolau BJ, Saito K, Roessner U and Trengove R (2014) Proposed quantitative and alphanumeric metabolite identification metrics. *Metabolomics* **10**, 1047–1049.

Tautenhahn R, Bottcher C and Neumann S (2008) Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**, 1–16.

Van Der Kloet FM, Bobeldijk I, Verheij ER and Jellema RH (2009) Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping research articles. *Journal of Proteome Research* **8**, 5132–5141.

Vermorel M, Heaney RK and Fenwick GR (1988) Antinutritional effects of the rapeseed meals, Darmor and Jet Neuf, and progoitrin together with myrosinase in the growing rat. *Journal of the Science of Food and Agriculture* **44**, 321–334.

Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, Velzen EJJ, Duijnhoven JPM and Dorsten FA (2008) Assessment of PLSDA cross validation. *Metabolomics* **4**, 81–89.

Widodo, Patterson JH, Newbigin E, Tester M, Bacic A and Roessner U (2009) Metabolic responses to salt stress of barley (*Hordeum vulgare* L.) cultivars, Sahara and Clipper, which differ in salinity tolerance. *Journal of Experimental Botany* **60**, 4089–4103.

Xia J, Psychogios N, Young N and Wishart D S (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research* **37**, 652–660.

Yang S, Arasu MV, Chun J, Jang Y, Lee Y, Kim IH, Lee K, Hong S and Kim S (2015) Identification and determination of phenolic compounds in rapeseed meals (*Brassica napus* L.). *Journal of Agricultural Chemistry and Environment* **4**, 14–23.