Untargeted metabolomic analyses reveal the diversity and plasticity of the specialized metabolome in seeds of different *Camelina sativa* varieties

Stephanie Boutet[1§], Léa Barreda[1§], François Perreau[1], Grégory Mouille[1], Etienne Delannoy[2,3], Marie-Laure Magniette[2,3,4], Andrea Monti[5], Loïc Lepiniec[1], Federica Zanetti[5], Massimiliano Corso[1*]

[1] Institut Jean-Pierre Bourgin, Université Paris-Saclay, INRAE, AgroParisTech, 78000, Versailles, France

[2] Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France.

[3] Université de Paris, CNRS, INRAE, Institute of Plant Sciences Paris Saclay (IPS2) 91405 Orsay

[4] UMR MIA-Paris, AgroParisTech, INRAE, Université Paris-Saclay, 75005, Paris, France

[5] Department of Agricultural and Food Sciences, Alma Mater Studiorum – Università di Bologna, Viale G. Fanin 44, 40127 Bologna, Italy

[§] Equal contribution

* Correspondence to: Massimiliano Corso massimiliano.corso@inrae.fr

E-mails:
SB stephanie.boutet@inrae.fr
LB lea.barreda@agroparistech.fr
FP francois.perreau@inrae.fr
GM gregory.mouille@inrae.fr
ED etienne.delannoy@inrae.fr
MLM Marie-Laure.Magniette@inrae.fr
AM a.monti@unibo.it
LL loic.lepiniec@inrae.fr
FZ federica.zanetti5@unibo.it
MC massimiliano.corso@inrae.fr

Running title: Camelina seed specialized metabolites plasticity

Keywords: *Brassicaceae*, metabolic signature, metabolite decorations, environmental changes, molecular networks, untargeted metabolomics.

Summary

Seeds produce a myriad of Specialized Metabolites (SMs). Nevertheless, despite the essential role of SMs in the interaction of plants with their environment, studying the ability of crop seeds to produce these protective compounds has been neglected. Camelina is an oilseed crop, whose seeds are characterized by high oil content and a unique composition, including broad SM diversity. We characterized SM landscapes in the seeds of six camelina cultivars grown in the open field and harvested during five consecutive growing seasons. We achieved a comprehensive annotation of camelina seed SMs combining molecular and correlation network analyses, which cluster SMs based on their chemical structures and co-accumulation patterns, respectively. Thus, we were able to evaluate the impact of the genotype and environment on the accumulation of these metabolites. Our data showed surprisingly high and unexplored effects of the environment on the stimulation of the seed-specialized metabolome. Moreover, they emphasize that seed SMs display a much higher environmental plasticity than storage compounds (*e.g.* oil and proteins). Last, we identified flavonols as the most plastic metabolic class, revealing highly variable accumulation according to the environmental conditions and/or the genotype. This work highlights the predominant effect of the environment on the regulation of the seed-specialized metabolome, with a potential impact on the seed quality of camelina and other crop species.

Significance statement

Seeds produce a myriad of Specialized Metabolites (SMs) with an essential role in the interaction of plants with their environment. We characterized SM landscapes in the seeds of six camelina cultivars grown in the open field and harvested during five consecutive growing seasons. We highlight the predominant effect of the environment on the regulation of the seed-specialized metabolome, with a potential impact on seed quality of camelina and other crop plant species.

INTRODUCTION

Plants produce large amounts of metabolites with different functions ranging from environmental adaptation to developmental controls (Erb and Kliebenstein, 2020; Fang *et al.*, 2019). Some of these metabolites, namely specialized metabolites (SMs), are usually synthetized in small concentrations, and have a restrained distribution to specific plant families, genera or species (Tissier *et al.*, 2015; Bhatla, 2018; Pichersky and Lewinsohn, 2011; Arimura and Maffei, 2016). They are commonly classified in three major groups according to their chemical structure: terpenoids, nitrogen containing molecules, and phenolic compounds (or phenylpropanoids) (Jacobowitz and Weng, 2020). Each class of metabolites comprises a huge diversity of structures and therefore a wide range of biological activities in different environmental conditions (Wang *et al.*, 2019; Kosmacz *et al.*, 2020).

Terpenoids are the largest group, with more than 25000 compounds identified that derived from the assembly of 5-C isoprene skeletons (terpenes) (Tissier *et al.*, 2015). A second group of SMs, referred as nitrogen-containing compounds, includes alkaloids, glucosinolates, cyanogenic glycosides and some non-protein amino acids (Wang *et al.*, 2019; Ziegler and Facchini, 2008; Halkier and Gershenzon, 2006). Among them, the alkaloids are the most represented, with more than 15000 compounds annotated (Bhatla, 2018). Finally, phenolic compounds (or phenylpropanoids) are chemically characterized as molecules containing one or more hydroxylated aromatic rings, and are synthetized via the malonate and shikimate pathways (Vogt, 2009; Tohge *et al.*, 2017; Corso, Perreau, Mouille, *et al.*, 2020). Plants produce more than 10000 phenolic compounds, which range from simple single aromatic rings such as benzoic acid or cinnamic acid to more complex molecules as flavan-3-ol polymers (also termed proanthocyanidins, PAs) (Corso, Perreau, Mouille, *et al.*, 2020; Lepiniec *et al.*, 2006). Flavonoids are certainly the largest group of phenylpropanoids, with more than 9 000 representatives (Tohge *et al.*, 2017; Corso, Perreau, Mouille, *et al.*, 2020). They have a C6-C3-C6 skeleton, characterized by two benzene rings fused by a heterocyclic pyran ring. Flavonoid diversity, biological activity and transport is strongly affected by the degree of glycosylation, hydroxylation and methylation patterns of the benzene rings (Tohge *et al.*, 2020; Sudheeran *et al.*, 2020; Tohge *et al.*, 2017; Wang *et al.*, 2019).

Specialized metabolites play crucial physiological and ecological roles throughout the plant life cycle. They can protect plants against abiotic stresses such as drought, high or low temperatures, toxic metals and UV, and can constitute effective defences against biotic stresses such as pathogens, herbivores or plant competitors (Delgoda and Murray, 2017; Corso, Perreau, Mouille, *et al.*, 2020; Lacchini and Goossens, 2020; Kosmacz *et al.*, 2020; Lepiniec *et al.*, 2006; Li *et al.*, 2020; Tohge *et al.*, 2016). For example, SMs can act as phytoalexins in response to pathogens, antifeedants against herbivory or allelopathic compounds to stunt the germination/growth of other plants (Corso, Perreau, Mouille, *et al.*, 2020; Corso, Perreau, Rajjou, *et al.*, 2020). These compounds are also involved in plant nutrition and/or reproduction, and they played a role in plant metal homeostasis (Corso and de la Torre, 2020; Corso, Perreau, Mouille, *et al.*, 2020; Croteau, R., Kutchan, T.M. and Lewis, 2000; Maeda, 2019; Arimura and Maffei, 2016).

Seeds present a wide range of specialized metabolites that are important for their physiological, agronomic, nutritional and industrial qualities (Corso, Perreau, Mouille, *et al.*, 2020; Lepiniec *et al.*, 2006; Corso, Perreau, Rajjou, *et al.*, 2020). They show remarkable levels and wide distribution of phenylpropanoids, flavonoids in particular (Routaboul *et al.*, 2006; Routaboul *et al.*, 2012; Alseekh, Ofner, *et al.*, 2020; Auger *et al.*, 2010). Seeds constitute (directly or indirectly) the major component of the human diet, and provide proteins, oil, and starch for both human and animal nutrition. A large diversity of specialized metabolites has been characterized in seeds of several plant species, including Brassicaceae (Routaboul *et al.*, 2006; Quéro *et al.*, 2016; Corso, Perreau, Mouille, *et al.*, 2020; Lepiniec *et al.*, 2006). Nevertheless, we are far from an exhaustive characterization of seed-specialized metabolite diversity and its regulation by the environment.

Camelina [*Camelina sativa* (L.) Crantz] is a Brassicaceae species that has gained a lot of interest in the past few years. This crop is highly resistant to biotic and abiotic stresses. Camelina seeds are characterized by unique oil composition (e.g. high level of polyunsaturated fatty acids, PUFA, and in particular n-3 fatty acids), and high oil content, it reaches adequate yields under low-input management and in limiting environments (Zanetti *et al.*, 2017; Faure and Tepfer, 2016; Yuan and Li, 2020). Nevertheless, further studies are needed to better valorise camelina seed co-products after oil extraction, including specialized metabolites such as polyphenols and carotenoids present in high amounts (Quéro *et al.*, 2016). Only a few data are available on camelina SM diversity and no information is available on their plasticity in response to environmental conditions.

In this work, we characterized the diversity of specialized metabolites in the seeds of six camelina cultivars (MIDAS, OMEGA, WUR, 787-08 789-02 and 887) grown in the open field and harvested during five consecutive growing seasons (2015-2019). Using molecular network analyses, which gather metabolites based on their structures, combined with metabolite co-accumulation networks, we extended the characterization and pathway information of the main metabolic classes in camelina seeds. Moreover, we characterized the impact of the genotype and environment on the plasticity of these metabolites in seeds. Among the main seed specialized metabolite classes, we identified the most plastic ones, which are characterized by highly variable accumulation according to the environmental conditions (i.e. different growing seasons) and/or the genotype.

RESULTS

Plant development and meteorological data

In the multi-year experiment (2015-2019) conducted on six camelina accessions grown in Italy (Bologna), the plants were able to reach maturity in about 100 days (Figure 1a). The amount of precipitation and the temperature patterns were recorded during the entire duration of the experiment (Figure 1b). The average minimum (T_{min}) and maximum (T_{max}) temperatures measured from 50% flowering to seed maturity (Figure 1c), a particularly important phase for seed quality, were quite variable in the years studied (Figure 1b). This particular developmental phase has been reported as the most important in camelina seed development (Walia *et al.*, 2018; Righini *et al.*, 2019), and thus it can be speculated as the most important also for specialized metabolite accumulation. In detail, the average T_{min} ranged from 13.5 °C in 2017 to 16.6 °C in 2019, while the average T_{max} ranged from 25.6 °C in 2016 to 29.3 °C in 2019. The cumulative precipitation in the same growth phase (from 50% flowering to maturity) showed high variability among the five studied years (Figure 1b). Specifically, 2017 had the lowest level of precipitation (52.2 mm), whereas 2016 showed around two- to three-fold higher cumulative precipitation (161.6 mm) compared to the other years (Figure 1b). Besides, in 2016, a hailstorm occurred before harvest (Zanetti *et al.*, 2017). Considering these variations, we will refer to the multiple factors underlying differences among the growing seasons considered in this study as the 'environment'.

Plant phenology and seed quality parameters

Environmental conditions had a significant influence on camelina development across the years considered in this study (Figure 1c). In particular, sowing took place between mid-March to early April, and consequently flowering occurred between early and late May, while seed maturity was reached between mid-June and early-July. Total oil (Figure 1d) and protein (Figure 1e) contents were determined in all genotypes for five consecutive harvest seasons. Seed oil content (%) showed moderate but significant variations among genotypes and years (Tukey's range test; $P \leq 0.05$). WUR and OMEGA were the genotype with the lowest seed oil content (36.7 and 37.4 % respectively), and 787-08 and 789-08 reached the highest values compared with 887, WUR and OMEGA, otherwise MIDAS showed an intermediate content. Concerning years, 2017 and 2019 reported the lowest seed oil content compared with the other growing seasons. Seed protein content (%) was also significantly influenced by years and genotypes. In particular, among genotypes, 787-08 had the lowest protein content, and 887 and WUR the highest, with all the other genotypes showing intermediate values. Concerning years, 2017 was the one reporting the highest seed protein content, but interestingly 2019 showed intermediate value, while 2015 and 2016 had lowest protein content. It is worth mentioning that the coefficient of variation for seed oil content and seed protein content was only 4.2% and 6.7% respectively, highlighting the limited variability for these qualitative traits in camelina seeds (Figure S1a). In addition, a linear regression between oil and protein content (%) highlighted a negative correlation and a major impact of the environment in the relationship between the two variables (Figure S1b).

The metabolomic landscape of camelina seeds

Untargeted profiling of specialized metabolites revealed a large diversity and led to the identification of an unprecedented number of metabolic classes in camelina seeds (Figure 2; Table S1). Untargeted metabolomic data in ElectroSpray Ionization in positive (ESI+) and negative (ESI-) modes allowed the detection of 2202 unique peaks of putative known and unknown metabolites. LC-MS/MS data were used to build a molecular network with MetGem software (Figure 2a; Table S1). This method allows clustering of compounds with high degree of chemical similarity according to their MS^2 spectra, and therefore identifying clusters of unknown metabolites using the MS/MS information for annotated compounds (Olivon *et al.*, 2018). The analysis of camelina seed metabolites using public databases for annotation and internal standard, coupled to molecular network, allowed the annotation of 33% of the detected molecules. Conversely, 15% of the metabolites were included in clusters that could not be associated to a metabolic class, and 52 % of the metabolites were not comprised in a cluster and not annotated (Figure 2b). Overall, seventeen metabolic categories have been identified in camelina seeds: phenolic acids, flavonoids (flavonols, flavan-3-ols, flavones, flavanones and isoflavones), phenolic glycosides, aminoacids, cinnamic acids, alkaloids, nucleosides, terpenes, sulphonamides, sugar-carbohydrates, chalcones, GSHs, vitamins, lignans, glucosinolates, benzoic acids and carboxylic acids (Figure 2c). Flavonols were the most represented metabolic category, with more than twice as many representatives (116) as the aminoacids (53), cinnamic acids (53) or sugars and carbohydrates (46). A relatively lower number of metabolites were annotated as alkaloids (14), flavan-3-ols (14), glucosinolates (14) terpenes (14), and benzoic acids (13). Finally, GSHs, vitamins, lignans, and carboxylic acids presented less than 10 metabolites.

We calculated the average accumulation of specialized metabolites among all samples (Table S2) and identified the top-10 most accumulated metabolites (Figure S2a). According to the molecular network analyses, among the 10 most accumulated metabolites three were flavonols (Rutin, Quercetin-3-(Glucoside [Glu] – Glu-Glu), and Isorhamnetin-3-(Glu - Rhamnoside [Rha])), one cinnamic acid (Sinapine), one aminoacid (γ-glutamyl leucine), and one flavan-3-ol (Catechin-3-Glu). In addition, two compounds that belong to the glucosinolate cluster (n8, n14), and two from non-annotated clusters (p29, p117) of the molecular network constructed with ESI+ data were present among the most highly accumulated metabolites (Figure S2b). The putative structure of the highly accumulated metabolites with no annotation was determined using SIRIUS 4.0 (Dührkop *et al.*, 2019), that propose molecular structures using LC-MS/MS data (see methods for further information). The proposed structures of n8 and n14 suggest their belonging to the glucosinolate class, more precisely n8 was annotated as Glucocamelinin and n14 as Glucoarabin. p29 was putatively annotated as a sugar (most likely trehalose) while metabolite p117 was annotated as an amino acid.

Genetic and environmental regulations of camelina seed specialized metabolome

The ESI+ and ESI- metabolite accumulation was modelled with a linear model to evaluate the genotype and year effects as well as their interactions. After filtering to remove low abundance metabolites, statistical analyses were performed with 1624 out of the 2202 metabolite peaks identified. More precisely, pairwise comparisons were carried out between the different years (*e.g.* 2015 versus 2016 on average for all genotypes), genotypes (*e.g.* OMEGA vs. MIDAS on average for all years) and genotype X environment (*e.g.* [OMEGA-2015 vs. OMEGA-2016] compared with [MIDAS-2015 vs. MIDAS-2016]). Statistical analyses highlighted 1170 and 1548 Differentially Accumulated Metabolites (DAMs) according to the genotype and the environment (*i.e.* the year), respectively (Figure 3a). Furthermore, 1604 metabolites were differentially accumulated because of the interaction between genotype and environment. Thus, the environment had greater influence on metabolite accumulation and affected 95% of the analysed seed-specialized metabolome. In addition, the Venn diagram pinpointed that no metabolites were differentially accumulated according to the sole effect of the genotype (Figure 3a). To measure the specific impact of each year and genotype, DAMs were identified for each comparison (Figure 3b). Overall, the year 2016 showed the highest induction of specialized metabolites compared to the other years, together with the highest number of DAMs, with 2979 induced and 734 reduced metabolites among all comparisons (Figure 3b). For the genotypes, OMEGA showed the highest number of DAMs, with 1681 and 616 induced and reduced accumulations, respectively, compared to the other genotypes (Figure 3b). Taken together, these data highlight that environment has about twice greater effect on metabolite accumulation than genotype (Figure 3b; Figure S3a). Depending on the year or genotype considered, the ratio between induced and repressed metabolites changed considerably (Figure S3b). In the case of genotypes, both OMEGA ($\log_2$ induced/repressed [ind/rep] = 1.4) and 787-08 ($\log_2$ ind/rep = 0.9) showed a higher number of induced than repressed metabolites, while the opposite was true for MIDAS, WUR, 789-02 and 887 (Figure S3b). As concerns years, 2016 was characterized by four-fold higher induced metabolites than repressed ones, and 2017 ($\log_2$ ind/rep = 0.4) showed a slightly higher number of induced than repressed metabolites (Figure S3b). The year 2015 did not show major differences between the numbers of induced and repressed metabolites ($\log_2$ ind/rep = 0.1), while both 2018 and 2019 years had more repressed than induced metabolites. Finally, we identified the DAMs induced solely in a specific year or genotype compared to all the others. The results highlighted a high specificity in the induction of specialized metabolites for the genotype OMEGA (245 specific DAMs) and for the years 2017 (154 DAMs) and, to a lesser extent, 2016 (81 DAMs) (Figure S4a). The contrasting behaviours observed between years and genotypes were further highlighted by the hierarchical cluster analysis of the metabolite accumulation patterns that grouped together years 2016 with 2017 and 2018 with 2019, and the genotypes OMEGA with 787-08, MIDAS with 789-02, and WUR with 887 (Figure 3c).

Identification and accumulation profiles of metabolites affected by environmental and genotype variations

The statistical analyses highlighted a strong effect of the environment on the specialized metabolome, and a majority of DAMs that were in common between the environment and the genotype (G-E). In this light, the G-E DAMs (1134 metabolites) were considered for further analyses (Figure 3a). Using this subset of DAMs, we performed an enrichment analysis using the “known metabolic” categories and ”not-annotated metabolic” categories highlighted by the molecular network analysis (Table S1, column ‘Category Enrichment’). In particular, not-annotated metabolites that clustered together in the molecular network analysis (Figure 2a), which therefore have similar chemical structures, were considered as a distinct metabolic category (see methods for more information). Only metabolic categories with more than five metabolites were taken into account for the enrichment analyses, leading to 14 known and 34 unknown categories (Table S1). Among the DAMs common to genotypes and environments, flavonols (58 DAMs), aminoacids (37 DAMs), cinnamic acids (37 DAMs), and glucosinolates (12 DAMs) were found to be enriched (Figure 4a). In addition, three unknown categories found with the ESI + mode, namely p_2_a (21 DAMs), p_2_c (11 DAMs) and p_16 (10 DAMs) showed enrichment among the G-E DAMs (Figure 4a). The heatmaps constructed using the average accumulation of metabolites for each category showed that the environmental conditions of 2016 induced higher levels of metabolite accumulation compared to the other years for the most enriched metabolic categories, except for flavonols (Figure 4b, c). Conversely, 2018 and 2019 were the years in which the most enriched metabolic categories displayed the lowest accumulation, except for flavonols, which also in this case showed a different accumulation profile (Figure 4b, c). In addition, the genotypes OMEGA and 787-08 showed higher amounts of metabolites for most of the enriched categories. These results confirmed the strong effect of the environment, particularly in 2016, on the accumulation of SMs. Interestingly, flavonols presented an accumulation pattern that significantly differed from the other known and unknown metabolite categories. In particular, flavonol (G-E DAMs) accumulation was not markedly induced in one specific year or genotype (Figure 4b, 4c). Heatmaps, constructed to have an overview of genotype plasticity for the metabolome across years, further confirmed that flavonol accumulation was differently regulated in different years (i.e. by varying environmental conditions) and in different genotypes (Figure 4c).

To characterize the metabolite signature specific to each factor considered, we performed an enrichment analysis using the metabolites uniquely induced in one year or genotype (Figure S4b). In this case, only the annotated/known metabolic categories were taken into account. The results highlighted an enrichment of flavonols and terpenes in MIDAS (65 specific DAMs), and of gluthathione (GSH), alkaloids and phenolic acids in OMEGA (245 specific DAMs). As concern the specific effect of the environment, cinnamic acids, nucleosides and flavonols were enriched in 2016 (81 specific DAMs), while terpenes and sulfonamides showed an enrichment in 2017 (154 specific DAMs) (Figure S4b).

Identification of important metabolite hubs and their environmental plasticity

The results obtained so far highlight the large plasticity of some metabolic categories, flavonols in particular. Nevertheless, it remained unclear how the genotypes or the environmental conditions affected the accumulation of specific metabolites (e.g. quercetin or kaempferol) or their modifications (i.e. glycosylation, acylation and methylation). To establish a link between metabolite structures (and therefore their putative/potential biological activity) with environmental variations or genotypes, a co-accumulation network analysis, based on correlation among metabolite accumulation patterns (Figure 5; $r > 0.8$), was performed using metabolomic data extracted for the enriched known (Figures 6 and S5) and unknown (Figures S6) metabolic classes. Co-accumulated metabolites were visualised in heatmaps created for each cluster and obtained from the metabolite correlation network. In comparison with the other enriched categories, flavonols were split into a higher number of clusters (9 in total) highlighting more diversified accumulation profiles in response to year and/or genotype, compared to the other metabolic classes (Figure 6). As for flavonols, several metabolites were annotated according to public and in-house databases, and a clear metabolic signature was identified for 4 out of 9 clusters identified by co-accumulation analyses. Specifically, quercetin, rutin (Q)-3-(Glu-Rha), and Iso-3-(Glu-Rha) (Figure 7), and kaempferol (K)-3-(Glu-Rha) (Figure 8), belong to a co-accumulation cluster that showed moderately higher accumulation in 2017 and lower accumulation in 789-02 compared with the other genotypes (cluster ‘a’). Acetylated and malonylated Q-3-G belong to cluster ‘b’, characterized by metabolites showing markedly high accumulation in 2016, and in 789-02 and OMEGA genotypes. Di- and tri-glycosylated forms of quercetin were both characterized by metabolites with high accumulation in 2018 and in 789-02 for Q-diglucoside or MIDAS-787-08 for Q-triglucoside (Figure 7).

The other annotated enriched metabolic categories, aminoacid and cinnamic acid metabolites, were split into two and four clusters according to their accumulation patterns, respectively, while the glucosinolates were all associated in one co-accumulation cluster (Figure S5). In particular, the aminoacids with a glutamyl group, such as γ-glutamyl-leucine, γ-glutamyl-tyrosine, glutamyl-phenylalanine, and kyurenic acid, belong to cluster ‘a’, which is characterized by higher accumulation in 2016, while two indole-derivative aminoacids (e.g. 3-formylindole and tryptophan) belong to cluster ‘b’, showing slightly but significantly higher accumulation in 2017 (Figure S7). Twenty-nine cinnamic-acid derivatives were incorporated in the co-accumulation network and divided into four clusters, among which cluster ‘a’ included the vast majority of this class of metabolites (20 out of 29; Figure S5b). Most of the cinnamic acids shown in the network derived from sinapic acids and/or sinapine. Cluster ‘a’ is characterized by metabolites with higher accumulation in 2016, and comprises sinapoyl glucose, benzoylglycerol glucuronic acid ether, feruloylcholine (and some derivatives), and the sinapine-4-glucoside. Other annotated cinnamic acids were 4-coumaroylcholine and 4-hydroxybenzoylcholine (cluster ‘b’), characterized by similar accumulation levels among the years considered, chlorogenic acid (cluster ‘c’), which was highly accumulated in 2016, and the sinapoyl di-glucose (cluster ‘d’) showing higher accumulation in 2018 (Figure S8). Finally, glucosinolates (methylsulfinyloctyl glucosinolate, glucocamelinin, glucoarabin) were markedly more accumulated in 2016 and genotypes 787-08 and 887 (Figure S9).

The unknown (not annotated) metabolic categories that were found using molecular network analyses, were separated into two co-accumulated clusters (‘a’ and ‘b’) in the case of p_2_a, and one cluster in the case of p_2_c and p_16 enriched categories. As already described for the top-10 most accumulated metabolites (Figure S2), the putative structures of the metabolites belonging to the enriched metabolic categories with no annotation were proposed using SIRIUS 4.0 (Dührkop *et al.*, 2019). The metabolites of p_2_a, p_2_c and p_16 categories were annotated as aminoacids, phenolic acids and alkaloids, respectively. All these enriched metabolic categories were characterized by higher accumulation in 2015 and 2016, except for the metabolites belonging to cluster ‘b’ of p_2_a category, which showed higher accumulation in 2015, 2016, and 2017 (Figure S10).

DISCUSSION

Phenotypic plasticity usually refers to the adaptability of an individual (or a genotype) to cope with environmental changes by modulating one or multiple phenotypic traits (Arnold *et al.*, 2019; Fusco and Minelli, 2010). The genetic and environmental impacts on the regulation of gene expression or physiological traits were extensively studied in several plant species and specific tissues or organs (Gruber *et al.*, 2013; dal Santo *et al.*, 2016; Dal Santo *et al.*, 2013; Janková Drdová *et al.*, 2019; Jong *et al.*, 2019; Ibañez *et al.*, 2017). Conversely, little if any information is available on the impact of the environment on the induction or repression of the specialized metabolome, especially in seeds. Studying these processes is particularly interesting in Brassicaceae species such as camelina, a rustic species of which seeds constitute the final products used for food, feed and biobased industrial applications. Although it is well known that camelina seeds contain numerous specialized metabolite classes, most available studies focused on a limited number of compounds, mainly oil composition, fatty acids or anti-nutritional compounds, such as glucosinolates (Amyot *et al.*, 2019).

In the present study, using a molecular network approach (Olivon *et al.*, 2018; Elie *et al.*, 2019) coupled with manual confirmation of metabolite annotation using molecular structure identification (Dührkop *et al.*, 2019), we found a large diversity of specialized metabolite classes (17 in total) in camelina seeds (Figure 2). Our results highlighted that only 33% of the metabolites identified by UPLC-MS/MS are annotated or belong to a cluster with annotated metabolites. About 15% of the specialized metabolome was associated to clusters with no annotation (but similar chemical structure) and the majority (*i.e.* 52%) of metabolites were neither annotated nor associated to a specific molecular network cluster (Figure 2). This is not surprising, since the metabolites present in public databases used for annotation represent only a small proportion of the plant metabolome (7-10%) (Chaleckis *et al.*, 2019; Szymański *et al.*, 2020). It is worth mentioning that the percentage of annotated molecules increased significantly if only the highly accumulated metabolites are considered (Table S1).

Despite this large fraction of non-annotate metabolites, using a molecular network analysis we strongly increased the percentage of metabolites that were assigned to a metabolic class from 12% (public databases) to 33% (public databases and Molecular Network). In addition, we significantly expanded the identification of metabolic classes compared to the published untargeted metabolomic analyses on camelina, which mostly focused their attention on annotated metabolites such as sinapine, a few flavonols and flavan-3-ols, and/or glucosinolates (Quéro *et al.*, 2016; Berhow *et al.*, 2014).

Statistical and data mining analyses allowed measurement of the impact of environment and genotype on metabolite accumulation, and the identification of enriched metabolic categories presenting variable degrees of plasticity relatively to these two factors (Figure 3). Given the five years studied, characterized by major meteorological variability, and the six genotypes, derived from different breeding programs, considered in this study, the results and the trends observed allowed us to identify or speculate on important factors affecting the specialized metabolome landscape in camelina seeds. In particular, the environmental conditions in 2016 caused the highest induction of the metabolome, whereas 2018 and 2019 presented the lowest. These contrasting behaviours can be linked to different weather conditions occurring in each year. Indeed, 2016 was characterized by lower temperatures and high amounts of precipitation during the seed filling phase, while 2018 and 2019 showed higher temperatures (both T_{min} and T_{max}) and much lower precipitation during the seed filling period (<70 mm). It is worth remarking that camelina is able to cope quite well with drought stress (Enjalbert *et al.*, 2013), but its ability to cope with high temperatures (Carmo-Silva and Salvucci, 2012) during the seed filling phase (i.e. 2018 and 2019) is particularly limited, and this is in strong agreement with the present results, which highlighted the lowest metabolome induction under those conditions. In addition, a hailstorm occurred in 2016 before harvest, which might also have affected the accumulation of some classes of seed-specialized metabolites. Indeed, these meteorological conditions might be considered as abiotic stress for the plant, which can severely affect seed physiology and quality. The accumulation of several specialized metabolites, many of them being well-known antioxidants, can also be seen as the plant response to adverse and stressful environmental conditions. It is known that camelina seed yield and quality (e.g. protein and oil content, oil composition, dormancy or longevity) are affected by meteorological conditions (Zanetti *et al.*, 2017; Krzyżaniak *et al.*, 2019). Nonetheless, our results showed that the seed metabolome landscape is also largely affected by environmental factors, and in particular specialized metabolites showed much higher environmental plasticity than seed oil or protein contents (Figure 1d).

Among the enriched metabolic categories modulated by environment and genotype, flavonols were the most represented class among the metabolites annotated or associated to a molecular network cluster (e.g. aminoacids, cinnamic acids and glucosinolates) (Figure 4). Differently from the other categories, flavonols were not markedly accumulated in 2016, but more accumulated in years 2017-2019. In general, the flavonol metabolome was induced in several genotypes and years (i.e. by various environmental conditions), as shown by the high number of co-accumulation clusters (nine, Figure 6) identified for this category. Hence, since the number of clusters can be directly related to the plasticity of a metabolic class, our data suggest that flavonols are the most plastic category among the specialized metabolites observed in camelina seeds. The large accumulation of flavonols in response to different years could be linked to the ability of plants to respond to different environmental stresses re-directing some biosynthetic pathways. Flavonol plasticity could be linked to its two aromatic rings and multiple potential decorations (Figures 7 and 8), while cinnamic acids (Figure S8) and amino acids (Figure S9) showed a lower number of aromatic rings and modifications in their structure.

The accumulation of flavonols in response to abiotic factors has already been documented in different plants, although mainly in vegetative tissues (Corso, Perreau, Mouille, *et al.*, 2020; Nakabayashi *et al.*, 2014; Tohge *et al.*, 2016; Pourcel *et al.*, 2007; Muhlemann *et al.*, 2018; Emiliani *et al.*, 2013). The antioxidant and protective activities of the flavonols would fit their higher accumulation in 2018 and 2019, which presented the highest temperatures during the seed filling phase, and in 2017, characterized by limited amounts of precipitation (Figure 1b; Figure 4). Metabolite co-accumulation analysis (Figures 5 and 6) coupled with their annotation by molecular network allowed the identification of specific patterns of flavonols with diverse structures and decorations (Figures 7 and 8). In relation to the effects of the environment, the glucoside-rhamnoside of Quercetin, Kaempferol and Isorhamnetin were more accumulated in 2017, which is characterized by the lowest level of precipitations. Q-diglucoside and Q-triglycoside were markedly more accumulated in 2018, showing high temperatures, while acylated and malonylated Q-glycosides were more accumulated in 2016, characterized by high precipitations rate and low temperatures (mix of temperature- and humidity-relates stresses). Hence, our analyses suggest that flavonols with different decorations (hydroxylation, glycosylation, methylation and acylation) are differentially accumulated according to environmental conditions (Alseekh, Perez de Souza, *et al.*, 2020; Chen *et al.*, 2020), especially in response to temperature and precipitation. Glycosylation, acylation and methylation play crucial roles in stability, solubility and bioavailability of the flavonoids. For instance, de-glycosylation is a well-characterized fine-tuned mechanism regulating flavonoid localization, availability, and biological activity (Wang *et al.*, 2019). Identifying the contribution of abiotic (e.g. heat or drought stress) and/or biotic stress factors on the accumulation of specific flavonoids and related decorations will be essential for understanding the environmental impact on seed quality.

The plasticity level was lower for aminoacids, cinnamic acids and glucosinolates compared with flavonols, as suggested by the reduced number of co-accumulation clusters (Figure S5). In the case of the aminoacids, one cluster contained mainly glutamate derivatives, whereas the other cluster grouped aromatic aminoacids (tryptophan derivatives). Globally, the glutamate derivatives, leucine and tryptophan derivatives appeared more often than other aminoacid derivatives. These observations are in line with previous studies showing that leucine and glutamic acid were present in higher amounts than others aminoacids (Zubr, 2003). Moreover, it could be speculated that other aromatic aminoacids, such as phenylalanine and tyrosine, were not over-accumulated since they are precursors of phenolic compounds. As concerns glucosinolates, only one compound (methylsulfinyloctyl glucosinolate) showed an annotation. Interestingly, differently from glucoarabin (9- methylsulfinylnonyl glucosinolate), glucocamelinin (10-methylsulfinyldecyl glucosinolate), and 11-methylsulfinylundecylglucosinolate, methylsulfinyloctyl glucosinolate was not previously reported as highly accumulated in camelina seeds (Amyot *et al.*, 2019; Berhow *et al.*, 2013).

Regarding the differences in metabolite accumulation among studied genotypes, both OMEGA and 787-08 showed the highest metabolome induction (Figure 3). Inversely, genotype 789-02 showed the lowest metabolite accumulation for most of the studied metabolic categories. Interestingly, 787-08 has been previously identified as a cultivar with large seeds, providing good and stable oil yields across different environments (Zanetti *et al.*, 2017). Moreover, cultivars 789-02 and 887 are characterized by different seed oil composition (i.e. increased ὠ3/ ὠ6 ratio) particularly suitable to different feed, food and industrial applications. The present results suggest that the agronomic performance of cultivar 787-08 could be related to an improved genetic combination allowing the diversity and abundance of metabolites to provide additional defences to the plant against abiotic stresses. The accordance between metabolite composition and the other traits found in genotypes 787-08 and OMEGA may represent a promising strategy for future improvement of camelina.

In this work we used and combined two different approaches to cluster and annotate specialized metabolites, one based on their chemical structure (allowing to establish “molecular networks”), and the other based on the clustering of co-accumulated metabolites (i.e. having similar patterns of accumulations). This approach allowed new annotations for several flavonols, aminoacids and cinnamic acids. The results obtained showed that, in most cases, the accumulation profile of a metabolite class is linked to the chemical structure (core and decorations). Hence, this approach could be used successfully to annotate new metabolites by linking their chemical structures to the accumulation patterns.

EXPERIMENTAL PROCEDURES

Genetic material, field experimental protocol and meteorological parameters

Six different cultivars of camelina were grown at the experimental farm of Bologna University (Italy) located in Cadriano (Bologna, Italy, 44°30′N, 11°23′E, 32 m a.s.l.) during five consecutive years (2015-2019). The soil at Cadriano is characterized as clay-loam (Udic Ustochrepts, mesic). The six tested cultivars were: MIDAS (AAFC, Saskatoon, Canada), OMEGA (University of Poznan, Poland), WUR (Wageningen University and Research, The Netherlands), and 787-08, 789-02 and 887 (Smart Earth Camelina, Saskatoon, Canada). These genotypes were chosen to represent the highest diversity among breeding programs and qualitative seed traits, in particular OMEGA and WUR belong to the breeding programs carried out at Poznan (Poland) and Wageningen (The Netherlands) Universities, respectively. MIDAS has been selected by AAFC (Saskatoon, Canada) to be highly tolerant to disease (i.e,. downey mildew) and with stable productive performance. Cultivars 787-08, 887 and 789-08 derive from the breeding program of Smart Earth Camelina (Canada), and they are improved for: increased seed weight, and stable and high seed yield and oil content (787-08); improved fatty acid composition (887), seed weight and improved fatty acid composition (789-02). Weather parameters (daily minimum, T_{min}, maximum, T_{max}, and precipitation) were measured and recorded by automated weather stations located at the experimental farm. Camelina was sown each year in late winter/early spring accordingly to the specific meteorological conditions. The same agronomic management for all the trials was adopted, consisting in typical tillage system (ploughing + harrowing), a seeding rate of 500 seeds m^{-2}, no irrigation, a top-dressing fertilization with N (50 kg N ha^{-1}, as urea at bolting stage), with no pest, disease nor weed chemical control. The experimental design was a randomized complete block with 3 or 4 biological replicates depending on the year. The plants were monitored weekly for their phenological phases, which were surveyed according to the BBCH scale for camelina (Martinelli and Galasso, 2011). At full maturity (residual seed moisture <12%), the central portion of each plot (~6 m^2) was cut manually at the soil level, and then threshed using a plot combine. Representative seed samples from individual plots were cleaned at the “Ricerca e Analisi Sementi” laboratory (LaRAS) of DISTAL department (Università di Bologna, IT), and then stored before shipping to IJPB institute (INRAE-Versailles, FR).

Protein and Oil extraction and analysis

Oil and protein determinations in all samples have been carried out by AAFC (Saskatoon, Canada), by near-infrared spectroscopy (NIRS) as reported by Zanetti *et al.* (2017). Oil and protein contents are reported as a percentage on a whole seed dry matter (zero moisture) basis.

Specialized metabolite extraction and UPLC-MS/MS analysis

Metabolites were extracted from 60 mg of camelina dry mature seeds (105 samples in total). The metabolite extraction protocol was adapted from Corso *et al.* (2018) and Giavalisco *et al.* (2011). Briefly, 1 mL of MeOH: Methyl-tert-butyl: H2O (1:3:1), conserved at 4°C, and 200 ng of Apigenin (used as internal standard) were added to each sample, which was then homogenized in 2ml tubes using a FastPrep instrument (1 min, 14,000 rpm). The mixtures were then shaken for 30 min at 4ºC using a ThermoMixer™ C (Eppendorf), placed in an ice cooled ultrasonication bath for 15 min and centrifuged for 1 min at the maximum speed to remove debris. The extracted samples were then transferred in new tubes containing 650 mL of MeOH: water (1:3), previously placed at -20ºC. The mixtures were centrifuged for 1 min at 14,000 rpm. The addition of MeOH: water (1:3) and the centrifugation led to a phase separation, providing an upper organic phase, containing the lipids, a lower aqueous phase, containing the polar and semi-polar metabolites, and a pellet of starch and proteins. The phase containing the polar and semi-polar metabolites was dried down in a SpeedVac vacuum concentrator (o/n) and resuspended in 200 µL of ULC/MS grade water (Biosolve).

Metabolomics data were acquired using a UHPLC system (Ultimate 3000 Thermo) coupled to quadrupole time of flight mass spectrometer (Q-Tof Impact II Bruker Daltonics, Bremen, Germany). A Nucleoshell RP 18 plus reversed-phase column (2 x 100 mm, 2.7 µm; Macherey-Nagel) was used for chromatographic separation. The mobile phases used for the chromatographic separation were (A) 0.1% formic acid in H_2O and (B) 0.1% formic acid in acetonitrile. The flow rate was of 400 µL/min and the following gradient was used: 95% of A for 1-min, followed by a linear gradient from 95% A to 80% A from 1 to 3-min, then a linear gradient from 80% A to 75% A from 3 to 8-min, a linear gradient from 75% A to 40% A from 8 to 20-min. 0% of A was hold until 24-min, followed by a linear gradient from 0% A to 95% A from 24 to 27-min. Finally, the column was washed by 30% A at for 3.5min then re-equilibrated for 3.5-min (35-min total run time). Data-dependent acquisition (DDA) methods were used for mass spectrometer data in positive and negative ESI modes using the following parameters: capillary voltage, 4.5kV; nebuliser gas flow, 2.1 bar; dry gas flow, 6 L/min; drying gas in the heated electrospray source temperature, 140ºC. Samples were analysed at 8Hz with a mass range of 100 to 1500 m/z. Stepping acquisition parameters were created to improve the fragmentation profile with a collision RF from 200 to 700 Vpp, a transfer time from 20 to 70 µs and collision energy from 20 to 40 eV. Each cycle included a MS fullscan and 5 MS/MS CID on the 5 main ions of the previous MS spectrum.

Metabolomic data processing

The .d data files (Bruker Daltonics, Bremen, Germany) were converted to .mzXML format using the MSConvert software (ProteoWizard package 3.0, Chambers *et al.*, 2012). mzXML data processing, mass detection, chromatogram building, deconvolution, samples alignment, and data export, were performed using MZmine 2.52 software (http://mzmine.github.io/) for both positive and negative data files. The ADAP chromatogram builder (Myers *et al.*, 2017) method was used with a minimum group size of scan 3, a group intensity threshold of 1000, a minimum highest intensity of 1500 and m/z tolerance of 10 ppm. Deconvolution was performed with the ADAP wavelets algorithm using the following setting: S/N threshold 8, peak duration range = 0.01-1 min RT wavelet range 0.02-0.2 min, MS^2 scan were paired using a m/z tolerance range of 0.01 Da and RT tolerance of 0.4 min. Then, isotopic peak grouper algorithm was used with a m/z tolerance of 10 ppm and RT tolerance of 0.4. All the peaks were filtered using feature list row filter keeping only peaks with MS2 scan. The alignment of samples was performed using the join aligner with an m/z tolerance of 10 ppm, a weight for m/z and RT at 1, a retention time tolerance of 0.3 min. Metabolites accumulation was normalized according to the internal standard (Apigenin) and weight of seeds used for the extraction.

A first research in library with Mzmine was done with identification module and "custom database search" to begin the annotation with our library, currently containing 93 annotations (RT and *m/z*) in positive mode and 61 in negative mode, with RT tolerance of 0.3 min and *m/z* tolerance of 0.0025 Da or 6 ppm.

Molecular networking

Molecular networks were generated with MetGem software (Olivon *et al.*, 2018; https://metgem.github.io) using the .mgf and .csv files obtained with MZmine2 analysis. The molecular network was optimized for the ESI+ and ESI- datasets and different cosines cosine similarity score thresholds were tested. ESI- and ESI+ molecular networks were generated using cosine score thresholds of 0.75 and 0.8, respectively. Optimal parameters that allowed a good cluster separation among metabolites were set up. Then, molecular networks were exported to Cytoscape software (Shannon *et al.*, 2003; https://cytoscape.org/) to format the metabolic categories.

Metabolites annotation

Metabolite annotation was performed in four consecutive steps. First, the obtained RT and *m/z* data of each feature were compared with our home-made library containing more than 150 standards or experimental common features (RT, *m/z*). Second, the ESI- and ESI+ metabolomic data used for molecular network analyses were searched against the available MS^2 spectral libraries (Massbank NA, GNPS Public Spectral Library, NIST14 Tandem, NIH Natural Product and MS-Dial), with absolute m/z tolerance of 0.02, 4 minimum matched peaks and minimal cosine score of 0.8. Third, not-annotated metabolites that belong to molecular network clusters containing annotated metabolites from step 1 and 2 were assigned to the same chemical family. Finally, for metabolites that had no or unclear annotation, Sirius software (https://bio.informatik.uni-jena.de/software/sirius/) was used. Sirius is based on machine learning techniques that used available chemical structures and MS/MS data from chemical databanks to propose structures of unknown compounds.

Chemical structures drawing

Specialized metabolites structures were drawn using ChemDraw 19.0 with default parameters

Statistical analyses on untargeted metabolomic data

Raw data were normalised on the internal standard (Apigenin) and weight of seeds used for the extraction. Normalised metabolite accumulations were expressed as Relative Units (RU). Only metabolites having values higher than 0.05 RU in at least half of the samples were considered. For these 1624 metabolites, accumulation was modeled with a linear model to decompose the averaged accumulation by a genotype effect plus a year effect and an interaction between the genotype and the year. Several questions were addressed with this model (i) the differences of accumulation between two genotypes averaged on years or given a year (ii) the differences of accumulation between two years averaged on genotypes or given a genotype (iii) the interactions between genotypes and years by testing if the difference of accumulation between two years in a given genotype is the same for a second genotype.

All these questions were written as contrasts in the linear model. P-values were adjusted across the metabolites for each question to control the false discovery rate (FDR), and a difference of accumulation for a metabolite was declared significant if its adjusted p-value was lower than 0.05. A Venn diagram, with the number of DAM depending on the year, genotype and genotype X year interaction, was then made (http://jvenn.toulouse.inra.fr/app/example.html). Enrichment analysis was done by using an hypergeometric test and the FDR was controlled across all the terms at level 0.05. The reference set was defined from the annotated and unknown metabolic categories (Table S1, column 'Category Enrichment'). The analyses was performed with the software R. The contrasts and the enrichment tests were written according to the functions of the tool DiCoExpress (Lambert *et al.*, 2020).

Heatmap, hierarchical clustering and metabolites co-accumulation analyses

The heatmaps and hierarchical clustering analyses of metabolomic data were carried out using the 'heatmap.2' function of GPLOTS R package (https://cran.r-project.org/bin/windows/base/). The metabolite value used for the heatmaps were calculated as percentage related to the sample showing the highest accumulation value for each metabolite (100% and 40% for yellow and blue colours, respectively).

Correlation was calculated among metabolites from enriched metabolic classes or with the average accumulation patterns of all annotated metabolic classes using 'ExpressionCorrelation' of Cytoscape software (Shannon *et al.*, 2003). All 105 samples present in the dataset were used to calculate correlations among metabolites. Metabolite co-accumulation networks were visualised using Cytoscape software.

Statistical analyses on physiological and seed quality parameters

To assess statistical differences in physiological parameters (yield), protein and oil content, and fatty acids composition among the different years (2015, 2016, 2017, 2018, 2019) or *Camelina sativa* accessions (MIDAS, OMEGA, WUR, 787.08, 789.02 and 887), an ANOVA with Tukey's range test ($P \leq 0.05$) post-hoc correction was performed using the AGRICOLAE R package.

ACKNOWLEDGEMENTS

This work was supported by the "Département de Biologie et Amélioration des Plantes" (BAP) of INRAE (grant "Appel à projets scientifiques BAP 2020" to M.C.) and by a grant ("Seed MetSpe") of the Labex Saclay Plant Sciences-SPS (AN-10-LABX-0040-SPS to L.L. and G.M.). IJPB benefits from the support of Labex Saclay Plant Sciences-SPS (ANR-17-EUR-0007). This work has benefited from the support of IJPB's Chemistry and Metabolomic platform. Authors would like to thanks Angela Vecchi for the technical supervision of the field trials in Bologna, and the LaRAS personnel for the technical supervision of the seed sample preparation. Authors thank also the seed providers of all the tested genotypes, and AAFC (Saskatoon, Canada) for running the oil and protein content determinations on camelina seeds.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY

The LC-MSMS untageted metabolomic data and metadata have been deposited at the INRAE data repository portal (data INRAE, doi.org/10.14758/9T8G-WJ20) with the identifier https://doi.org/10.15454/ATTENN.

The R script used for the statistical analyses has been deposited at the INRAE data repository portal with the identifier https://doi.org/10.15454/A3QKZI.

AUTHOR CONTRIBUTIONS

MC and FZ designed the research. MC, FZ, SB, LB and FP performed the experiments. FZ and AM coordinate field experiments and furnished the seed samples for the analyses. SB, FP and GM generated the LC-MSMS data, performed the initial quality analyses on the metabolomic data and the metabolites annotation with SIRIUS. MC, SB and LB performed the molecular network analyses. MC, ED, MLM and LL performed the statistical analyses on metabolomic data. MC and LB generated all Fig.s and supplemental material. MC wrote the paper, which was edited by FZ, AM and LL. All the authors commented on and approved the manuscript.

SUPPORTING INFORMATION

Figure S1. Oil and protein accumulation behaviours

(a) Heat maps showing the average accumulation separately for oil and protein. The accumulation values were calculated as a percentage related to the sample showing the highest accumulation value for each metabolite (100% and 40% for yellow and blue colours, respectively) within years or genotypes.

(b) Linear regression between oil and protein accumulation values (%). The linear regression was calculated separately for each year (environment) and genotype considered in the analysis. The different years or genotypes are highlighted with different colours in the two graphs.

Figure S2. Top-10 most accumulated specialized metabolites in camelina seeds.

(a) Box plot showing the relative accumulation of the top-10 most accumulated specialized metabolites. Centre lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; data points are plotted as open circles. n = 30 sample points. The metabolite levels are represented as relative units normalized according to the internal standard (Apigenin) and relative to the seed weight.

(b) Structures of the most accumulated metabolites that were annotated using public databases.

Figure S3. Induced and repressed metabolites identified in all comparisons for genotype and environment factors.

(a) Differentially accumulated metabolites in all comparisons for the genotype and the environment factors. Induced and repressed metabolites were calculated for each comparison: e.g. for the 2015 vs 2016 comparison, 111 metabolites were more accumulated (induced) in 2015 than in 2016, and 587 less accumulated (repressed) in 2015 than in 2016.

(b) $\log_2$ ratio representing the proportion of induced/repressed metabolites for each genotype and year considered.

Figure S4. Induced metabolites and enriched metabolic categories specific to each year and genotype.

(a) Induced metabolites specific to each year and genotype were identified: e.g. 29 metabolites were uniquely induced in 2015 in at least one comparison. The number of metabolites induced in each year or genotype is indicated.

(b) Histograms showing the annotated enriched metabolic categories using DAM uniquely induced in one year or genotype. The number of metabolites that belong to each category is indicated.

Figure S5. Amino acid, cinnamic acid, and glucosinolate co-accumulation networks.

(a, b, c) Clusters and accumulation patterns are shown for the amino acid (a), cinnamic acid (b) and glucosinolate (c) metabolites present in the co-accumulation network. The heat maps show the average accumulation for each metabolite in all years and genotypes. The accumulation values were calculated as a percentage related to the sample showing the highest accumulation value for each metabolite (100% and 40% for yellow and blue colours, respectively) within years or genotypes. The m/z and annotations are also shown for each metabolite.

Figure S6. Co-accumulation networks constructed with the unknown metabolic categories

(a, b, c) Clusters and accumulation profiles are shown for the metabolites that belong to the p_2_a (a), p_2_c (b) and p_16 (c) unknown categories and present in the co-accumulation networks. The heat maps show the average accumulation for each metabolite in all years and genotypes. The accumulation values were calculated as a percentage related to the sample showing the highest accumulation value for each metabolite (100% and 40% for yellow and blue colours, respectively) within years or genotypes. The m/z and annotations are also shown for each metabolite.

Figure S7. Amino acid-related pathways and clusters

Biological pathway representing the amino acids. The histograms represent year and genotype average accumulation for each co-accumulation cluster. Different letters indicate statistically significant differences ($P \leq 0.05$) by ANOVA with Tukey's range test.

Figure S8. Cinnamic acids-related pathways and clusters

Biological pathway representing the cinnamic acid-related metabolites. The histograms represent year and genotype average accumulation for each co-accumulation cluster. Different letters indicate statistically significant differences ($P \leq 0.05$) by ANOVA with Tukey's range test.

Figure S9. Glucosinolate structures and clusters

Chemical structure representing the annotated glucosinolate metabolite. The histograms represent year and genotype average accumulation for each co-accumulation cluster. Different letters indicate statistically significant differences ($P \leq 0.05$) by ANOVA with Tukey's range test.

Figure S10. Cluster accumulation behaviours, putative structures and annotations for the enriched but unknown metabolic classes.

Putative chemical structures representing the unknown metabolic classes that showed an enrichment. The histograms represent year and genotype average accumulation for each co-accumulation cluster. Different letters indicate statistically significant differences ($P \leq 0.05$) by ANOVA with Tukey’s range test.

REFERENCES

Alseekh, S., Ofner, I., Liu, Z., Osorio, S., Vallarino, J., Last, R.L., Zamir, D., Tohge, T. and Fernie, A.R. (2020) Quantitative trait loci analysis of seed-specialized metabolites reveals seed-specific flavonols and differential regulation of glycoalkaloid content in tomato. *Plant J.*, **103**, 2007–2024.

Alseekh, S., Perez de Souza, L., Benina, M. and Fernie, A.R. (2020) The style and substance of plant flavonoid decoration; towards defining both structure and function. *Phytochemistry*, **174**, 112347.

Amyot, L., McDowell, T., Martin, S.L., Renaud, J., Gruber, M.Y. and Hannoufa, A. (2019) Assessment of Antinutritional Compounds and Chemotaxonomic Relationships between Camelina sativa and Its Wild Relatives. *J. Agric. Food Chem.*, **67**, 796–806.

Arimura, G. and Maffei, M. (2016) *Plant Specialized Metabolism: Genomics, Biochemistry, and Biological Functions*,.

Arnold, P.A., Kruuk, L.E.B. and Nicotra, A.B. (2019) How to analyse plant phenotypic plasticity in response to a changing climate. *New Phytol.*, **222**, 1235–1241.

Auger, B., Marnet, N., Gautier, V., Maia-Grondard, A., Leprince, F., Renard, M., Guyot, S., Nesi, N. and Routaboul, J.M. (2010) A detailed survey of seed coat flavonoids in developing seeds of brassica napus L. *J. Agric. Food Chem.*, **58**, 6246–6256.

Berhow, M.A., Polat, U., Glinski, J.A., Glensk, M., Vaughn, S.F., Isbell, T., Ayala-Diaz, I., Marek, L. and Gardner, C. (2013) Optimized analysis and quantification of glucosinolates from Camelina sativa seeds by reverse-phase liquid chromatography. *Ind. Crops Prod.*, **43**, 119–125.

Berhow, M.A., Vaughn, S.F., Moser, B.R., Belenli, D. and Polat, U. (2014) Evaluating the Phytochemical Potential of Camelina: An Emerging New Crop of Old World Origin. In *Phytochemicals – Biosynthesis, Function and Application*. pp. 129–148.

Bhatla, S.C. (2018) Secondary Metabolites. In *Plant Physiology, Development and Metabolism*. Springer Singapore, pp. 1099–1166.

Carmo-Silva, A.E. and Salvucci, M.E. (2012) The temperature response of CO 2 assimilation, photochemical activities and Rubisco activation in Camelina sativa, a potential bioenergy crop with limited capacity for acclimation to heat stress. *Planta*, **236**, 1433–1445.

Chaleckis, R., Meister, I., Zhang, P. and Wheelock, C.E. (2019) Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics Metabolomics and sample complexity. *Curr. Opin. Biotechnol.*, **55**, 44–50.

Chambers, M.C., MacLean, B., Burke, R., et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.*, **30**, 918–920.

Chen, J., Hu, X., Shi, T., et al. (2020) Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnol. J.*, **18**, 1722–1735.

Corso, M. and la Torre, V.S.G. de (2020) Biomolecular approaches to understanding metal tolerance and hyperaccumulation in plants. *Metallomics*, **12**, 840–859.

Corso, M., Perreau, F., Mouille, G. and Lepiniec, L. (2020) Specialized Phenolic compounds in seeds: structures, functions, and regulations. *Plant Sci.*, **296**, 110471.

Corso, M., Perreau, F., Rajjou, L., Malek, R. Ben, Lepiniec, L. and Mouille, G. (2020)

Specialized metabolites in seeds. *Adv. Bot. Res.*, doi: 10.1016/bs.abr.2020.11.001.

Corso, M., Schvartzman, M.S., Guzzo, F., Souard, F., Malkowski, E., Hanikenne, M. and Verbruggen, N. (2018) Contrasting cadmium resistance strategies in two metallicolous populations of Arabidopsis halleri. *New Phytol.*, **218**, 283–297.

Croteau, R., Kutchan, T.M. and Lewis, N.G. (2000) Natural Products (Secondary Metabolites). In *Biochemistry and Molecular Biology of Plants*. pp. 1250–1319.

Dal Santo, S., Battista Tornielli, G., Zenoni, S., Fasoli, M., Farina, L., Anesi, A., Guzzo, F., Delledonne, M. and Pezzotti, M. (2013) The plasticity of the grapevine berry transcriptome. *Genome Biol.*, **14**, R54.

dal Santo, S., Fasoli, M., Negri, S., D'Incà, E., Vicenzi, N., Guzzo, F., Tornielli, G.B., Pezzotti, M. and Zenoni, S. (2016) Plasticity of the Berry Ripening Program in a White Grape Variety. *Front. Plant Sci*, **7**, 970.

Delgoda, R. and Murray, J.E. (2017) Evolutionary Perspectives on the Role of Plant Secondary Metabolites. In *Pharmacognosy: Fundamentals, Applications and Strategy*. Elsevier Inc., pp. 93–100.

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A. V., Meusel, M., Dorrestein, P.C., Rousu, J. and Böcker, S. (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods*, **16**, 299–302.

Elie, N., Santerre, C. and Touboul, D. (2019) Generation of molecular network from electron ionization mass spectrometry data by combining MzMine2 and MetGem software. *Anal. Chem.*, **91**, 11489–11492.

Emiliani, J., Grotewold, E., Falcone Ferreyra, L. and Casati, P. (2013) Flavonols Protect Arabidopsis Plants against UV-B Deleterious Effects. *Mol. Plant*, **6**, 1376–1379.

Enjalbert, J.N., Zheng, S., Johnson, J.J., Mullen, J.L., Byrne, P.F. and McKay, J.K. (2013) Brassicaceae germplasm diversity for agronomic and seed quality traits under drought stress. *Ind. Crops Prod.*, **47**, 176–185.

Erb, M. and Kliebenstein, D. (2020) Plant secondary metabolites as defenses, regulators and primary metabolites- The blurred functional trichotomy. *Plant Physiol.*, **184**, pp.00433.2020.

Fang, C., Fernie, A.R. and Luo, J. (2019) Exploring the Diversity of Plant Metabolism. *Trends Plant Sci.*, **24**, 83–98.

Faure, J.-D. and Tepfer, M. (2016) Camelina, a Swiss knife for plant lipid biotechnology. *OCL*, **23**, D503.

Fusco, G. and Minelli, A. (2010) Phenotypic plasticity in development and evolution: facts and concepts. *Phil. Trans. R. Soc. B*, **365**, 547–556.

Giavalisco, P., Li, Y., Matthes, A., et al. (2011) Elemental formula annotation of polar and lipophilic metabolites using 13 C, 15 N and 34 S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.*, **68**, 364–376.

Gruber, B.D., Giehl, R.F.H., Friedel, S. and Wirén, N. Von (2013) Plasticity of the Arabidopsis root system under nutrient deficiencies. *Plant Physiol.*, **163**, 161–179.

Halkier, B.A. and Gershenzon, J. (2006) Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.*, **57**, 303–333.

Ibañez, C., Poeschl, Y., Peterson, T., Bellstädt, J., Denk, K., Gogol-Döring, A., Quint, M. and Delker, C. (2017) Ambient temperature and genotype differentially affect developmental and phenotypic plasticity in Arabidopsis thaliana. *BMC Plant Biol.*, **17**.

Jacobowitz, J.R. and Weng, J. (2020) Exploring Uncharted Territories of Plant Specialized Metabolism in the Postgenomic Era. *Annu. Rev. Plant Biol.*, 1–28.

Janková Drdová, E., Klejchová, M., Janko, K., Hála, M., Soukupová, H., Cvrčková, F. and Žárský, V. (2019) Developmental plasticity of Arabidopsis hypocotyl is dependent on exocyst complex function. *J. Exp. Bot.*, **70**, 1255–1265.

Jong, M. de, Tavares, H., Pasam, R.K., et al. (2019) Natural variation in Arabidopsis shoot branching plasticity in response to nitrate supply affects fitness. *PLoS Genet.*, **15**, e1008366.

Kosmacz, M., Sokołowska, E.M., Bouzaa, S. and Skirycz, A. (2020) Towards a functional understanding of the plant metabolome. *Curr. Opin. Plant Biol.*, **55**, 47–51.

Krzyżaniak, M., Stolarski, M.J., Tworkowski, J., Puttick, D., Eynck, C., Załuski, D. and Kwiatkowski, J. (2019) Yield and seed composition of 10 spring camelina genotypes cultivated in the temperate climate of Central Europe. *Ind. Crops Prod.*, **138**, 111433.

Lacchini, E. and Goossens, A. (2020) Combinatorial Control of Plant Specialized Metabolism: Mechanisms, Functions, and Consequences. *Annu. Rev. ofCell Dev. Biol.*, **36**, 1–23.

Lambert, I., Paysant-Le Roux, C., Colella, S. and Martin-Magniette, M.L. (2020) DiCoExpress: A tool to process multifactorial RNAseq experiments from quality controls to co-expression analysis through differential analysis based on contrasts inside GLM models. *Plant Methods*, **16**, 68.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N. and Caboche, M. (2006) Genetics and Biochemistry of Seed Flavonoids. *Annu. Rev. Plant Biol.*, **57**, 405–430.

Li, D., Halitschke, R., Baldwin, I.T. and Gaquerel, E. (2020) Information theory tests critical predictions of plant defense theory for specialized metabolism. *Sci. Adv.*, **6**, eaaz0381.

Maeda, H.A. (2019) Evolutionary Diversification of Primary Metabolism and Its Contribution to Plant Chemical Diversity. *Front. Plant Sci.*, **10**, 881.

Martinelli, T. and Galasso, I. (2011) Phenological growth stages of Camelina sativa according to the extended BBCH scale. *Ann. Appl. Biol.*, **158**, 87–94.

Muhlemann, J.K., Younts, T.L.B. and Muday, G.K. (2018) Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. *Proc. Natl. Acad. Sci. U. S. A.*, 10.1073/pnas.1811492115.

Myers, O.D., Sumner, S.J., Li, S., Barnes, S. and Du, X. (2017) One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. *Anal. Chem.*, **89**, 8696–8703.

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., et al. (2014) Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *Plant J.*, **77**, 367–379.

Olivon, F., Elie, N., Grelier, G., Roussi, F., Litaudon, M. and Touboul, D. (2018) MetGem Software for the Generation of Molecular Networks Based on the t-SNE Algorithm. *Anal. Chem.*, **90**, 13900–13908.

Pichersky, E. and Lewinsohn, E. (2011) Convergent Evolution in Plant Specialized Metabolism. *Annu. Rev. Plant Biol.*, **62**, 549–66.

Pourcel, L., Routaboul, J.-M., Cheynier, V., Lepiniec, L. and Debeaujon, I. (2007) Flavonoid oxidation in plants: from biochemical properties to physiological functions. *Trends Plant Sci.*, **12**, 29–36.

Quéro, A., Molinié, R., Mathiron, D., et al. (2016) Metabolite profiling of developing Camelina sativa seeds. *Metabolomics*, **12**, 1–14.

Righini, D., Zanetti, F., Martínez-Force, E., Mandrioli, M., Toschi, T.G. and Monti, A. (2019) Shifting sowing of camelina from spring to autumn enhances the oil quality for bio-based applications in response to temperature and seed carbon stock. *Ind. Crops Prod.*, **137**, 66–73.

Routaboul, J.-M., Dubos, C., Beck, G., Marquis, C., Bidzinski, P., Loudet, O. and Lepiniec, L. (2012) Metabolite profiling and quantitative genetics of natural variation for flavonoids in Arabidopsis. *J. Exp. Bot.*, **63**, 3749–3764.

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

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–504.

Sudheeran, P.K., Ovadia, R., Galsarker, O., Maoz, I., Sela, N., Maurer, D., Feygenberg, O., Oren Shamir, M. and Alkan, N. (2020) Glycosylated flavonoids: fruit's concealed antifungal arsenal. *New Phytol.*, **225**, 1788–1798.

Szymański, J., Bocobza, S., Panda, S., Sonawane, P. and Cárdenas, P.D. (2020) Analysis of Wild Tomato Introgression Lines Elucidates the Genetic Basis of Transcriptome and Metabolome Variation Underlying Fruit Traits and Pathogen Response. *Nat. Genet.*, **52**, 1111–1121.

Tissier, A., Ziegler, J. and Vogt, T. (2015) Specialized Plant Metabolites: Diversity and Biosynthesis. In *Ecological Biochemistry: Environmental and Interspecies Interactions*. wiley, pp. 14–37.

Tohge, T., Scossa, F., Wendenburg, R., et al. (2020) Exploiting the natural variation in tomato to define pathway structure and metabolic regulation of fruit polyphenolics in the lycopersicum complex. *Mol. Plant*, 1–20.

Tohge, T., Souza, L.P. de and Fernie, A.R. (2017) Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J. Exp. Bot.*, **68**, 4013–4028.

Tohge, T., Wendenburg, R., Ishihara, H., et al. (2016) Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat. Commun.*, **7**, 12399.

Vogt, T. (2009) Phenylpropanoid Biosynthesis. *Mol. Plant*, **3**, 2–20.

Walia, M.K., Wells, M.S., Cubins, J., Wyse, D., Gardner, R.D., Forcella, F. and Gesch, R. (2018) Winter camelina seed yield and quality responses to harvest time. *Ind. Crops Prod.*, **124**, 765–775.

Wang, S., Alseekh, S., Fernie, A.R. and Luo, J. (2019) The Structure and Function of Major Plant Metabolite Modifications. *Mol. Plant*, **12**, 899–919.

Yuan, L. and Li, R. (2020) Metabolic Engineering a Model Oilseed Camelina sativa for the Sustainable Production of High-Value Designed Oils. *Front. Plant Sci.*, **11**, 11.

Zanetti, F., Eynck, C., Christou, M., et al. (2017) Agronomic performance and seed quality attributes of Camelina (Camelina sativa L. crantz) in multi-environment trials across Europe and Canada. *Ind. Crops Prod.*, **107**, 602–608.

Ziegler, J. and Facchini, P.J. (2008) Alkaloid Biosynthesis: Metabolism and Trafficking. *Annu. Rev. Plant Biol.*, **59**, 735–769.

Zubr, J. (2003) Dietary fatty acids and amino acids of camelina sativa seed. *J. Food Qual.*, **26**, 451–462.

FIGURE LEGENDS

Figure 1. Growth cycle of camelina, meteorological data and seed quality parameters.

(a) Developmental cycle of camelina plants in Bologna (Italy). The average phenological stages and days after sowing (DAS) are indicated.

(b) Meteorological parameters obtained from 50% flowering to seed maturity. Minimum (Tmin), average (Tavg) and maximum (Tmax) temperatures (°C), and amount of precipitation (mm) are shown.

(c) Sowing, flowering and harvest dates are indicated for each year studied (2015 to 2019).

(d, e) Total seed oil (d) and protein (e) contents are shown for each year and genotype. Vertical bars = standard deviations. Different letters indicate statistically significant differences ($P \leq 0.05$, Tukey's range test).

Figure 2. Molecular Networks and annotation of the seed-specialized metabolites

(a) The molecular networks are shown for metabolite analyses carried out in positive (ESI+) and negative (ESI-) ionization modes. Different colours correspond to different metabolic classes. Metabolites are grouped based on their chemical structures. Cosine similarity scores of 0.8 and 0.75 were used for ESI+ and ESI-, respectively.

(b) Percentage of specialized metabolites present in a molecular network with or without an annotation.

(c) Histogram showing the annotated/known metabolic categories identified in camelina seeds. The number of metabolites that belong to each category is indicated in the figure.

Figure 3. Genetic and environmental impact on the accumulation of seed-specialized metabolites.

(a) Venn diagram and histogram indicating the number of Differentially Accumulated Metabolites (DAMs) according to the Environment, Genotype and Genotype x Environment (GxE) factors.

(b) Number of induced (yellow) and repressed (blue) DAMs for each year and genotype. The DAM numbers are cumulative and were calculated considering all the statistical comparisons for each year and/or genotype.

(c) Hierarchical clustering analysis using the amounts of metabolites. The results are shown separately for years and genotypes.

Figure 4. Enriched metabolic categories

(a) Histograms showing the annotated and not-annotated enriched metabolic categories using DAM affected by the genotype (G) and the environment (E). The number of metabolites that belong to each category is indicated.

(b, c) Heat maps showing the average metabolite accumulation separately for years and genotypes (b), and the metabolite accumulation of all genotypes during the five years considered (c) for each enriched metabolic category. The accumulation values were calculated as a percentage related to the sample showing the highest accumulation value for each metabolite (100% and 40% for yellow and blue colours, respectively) within years or genotypes.

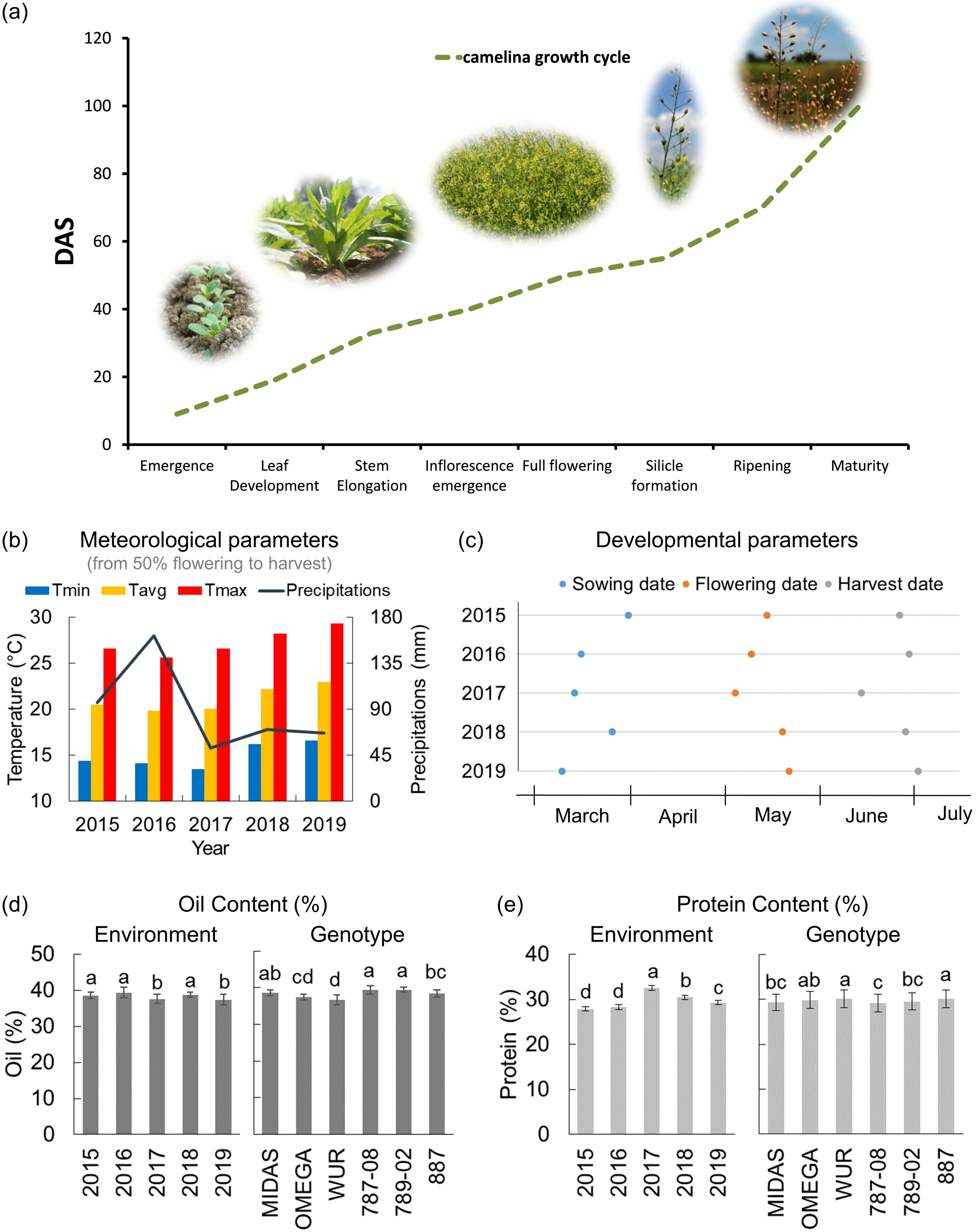
Figure 5. Specialized metabolites co-accumulation networks.

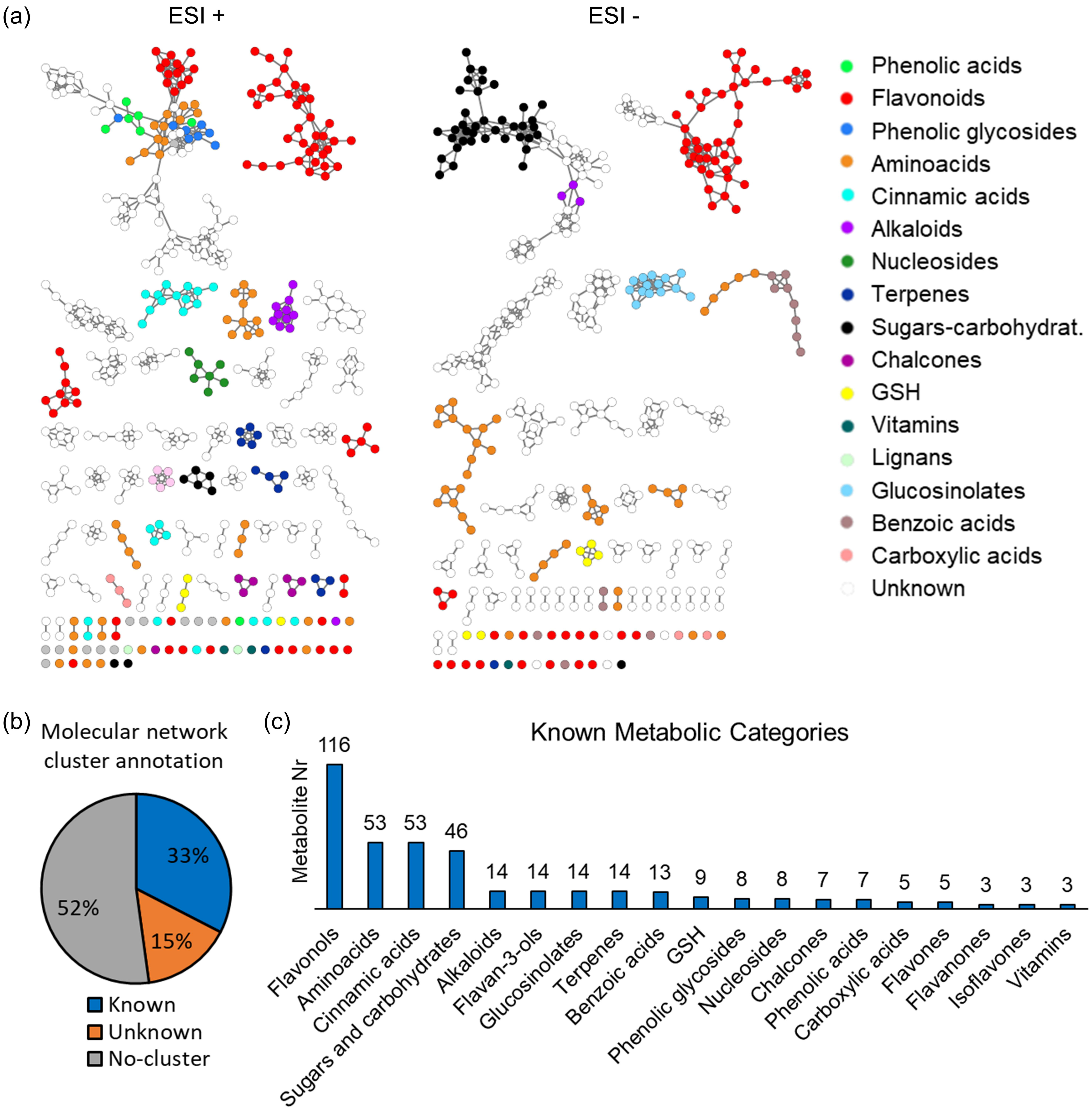
The accumulation patterns were used to calculate correlations ($r > 0.8$) among metabolites and define the clusters. Only the metabolites belonging to enriched categories were considered in this analysis. The accumulation profiles of the years (Y) and genotypes (G) are shown for each co-accumulation cluster. The accumulation values were calculated as a percentage related to the sample showing the highest accumulation value for each metabolite (100% and 40% for yellow and blue colours, respectively) within years or genotypes. The metabolites from each enriched category are highlighted with different colours.

Figure 6. Flavonol co-accumulation network.
Clusters and accumulation patterns are shown for the flavonols present in co-accumulation networks. The heat maps shows the average accumulation for each metabolite in all years and genotypes. The accumulation values were calculated as a percentage relative to the sample showing the highest accumulation value for each metabolite (100% and 40% for yellow and blue colours, respectively) within years or genotypes. The *m/z* and available annotations are also shown for each metabolite.

Figure 7. Quercetin-related pathways and clusters
Biological pathway representing the quercetin-related metabolites. The histograms represent average accumulations for the years and genotypes for each co-accumulation cluster. Different letters indicate statistically significant differences ($P \leq 0.05$) by ANOVA with Tukey's range test.

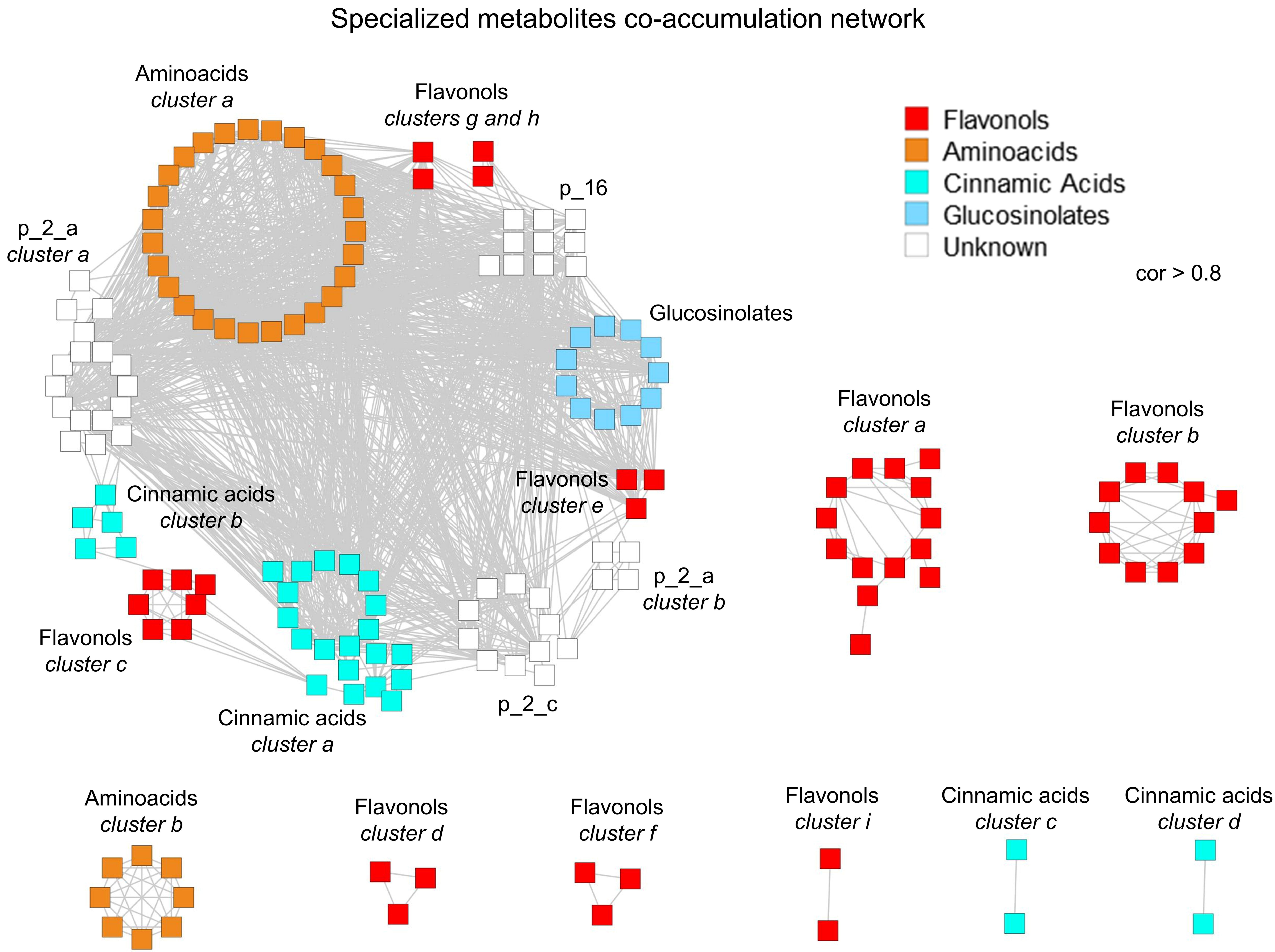
Figure 8. Kaempferol-related pathway and cluster
Biological pathway representing the kaempferol-related metabolites. The histograms represent average accumulations for the years and genotypes, for each co-accumulation cluster. Different letters indicate statistically significant differences ($P \leq 0.05$) by ANOVA with Tukey's range test.

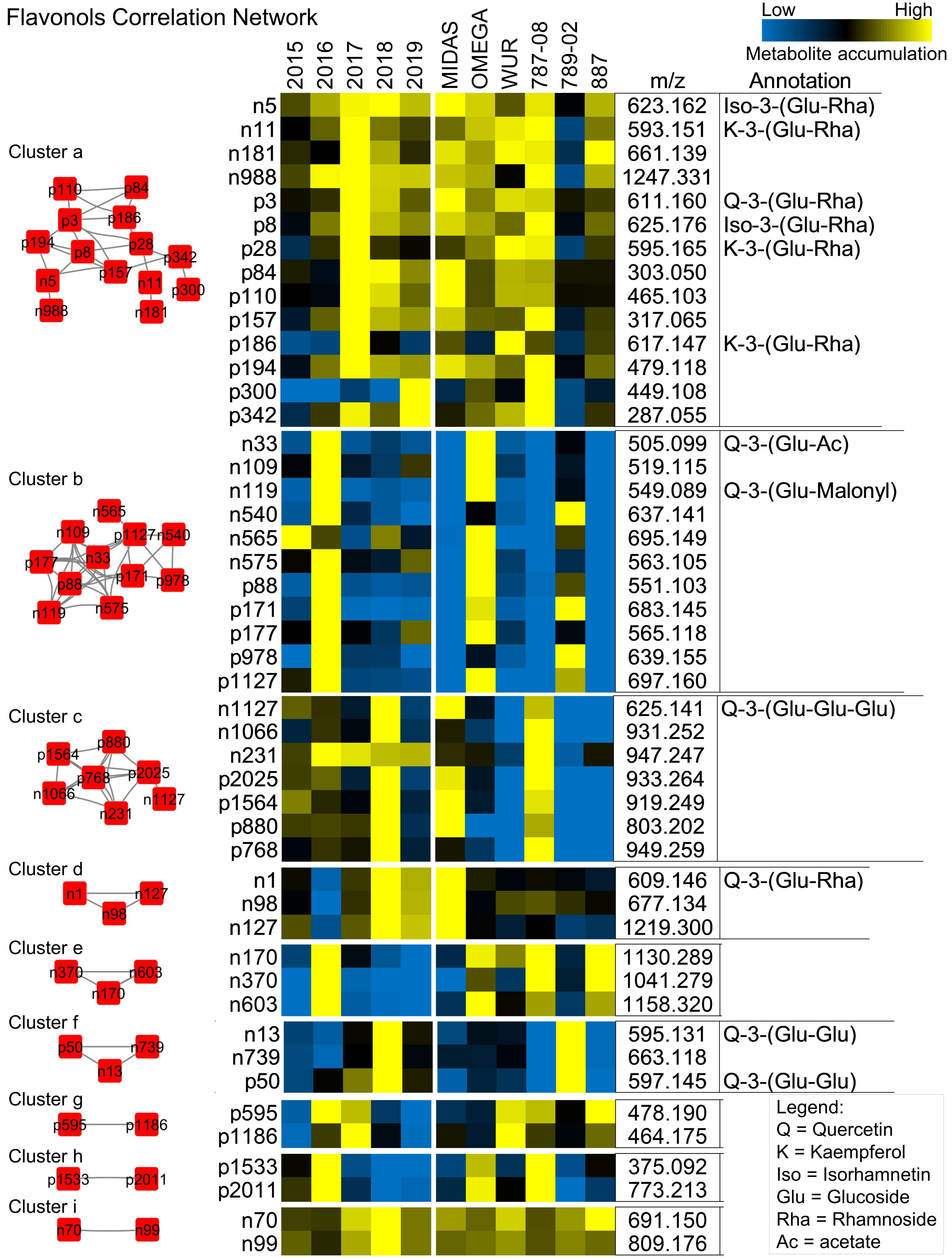

(a)
camelina growth cycle
DAS
0
20
40
60
80
100
120
Emergence
Leaf Development
Stem Elongation
Inflorescence emergence
Full flowering
Silicle formation
Ripening
Maturity
(b)
Meteorological parameters
(from 50% flowering to harvest)
Tmin
Tavg
Tmax
Precipitations
Temperature (°C)
10
15
20
25
30
Precipitations (mm)
0
45
90
135
180
2015
2016
2017
2018
2019
Year
(c)
Developmental parameters
Sowing date
Flowering date
Harvest date
2015
2016
2017
2018
2019
March
April
May
June
July
(d)
Oil Content (%)
Environment
Genotype
Oil (%)
0
10
20
30
40
50
a
a
b
a
b
2015
2016
2017
2018
2019
ab
cd
d
a
a
bc
MIDAS
OMEGA
WUR
787-08
789-02
887
(e)
Protein Content (%)
Environment
Genotype
Protein (%)
0
10
20
30
40
d
d
a
b
c
2015
2016
2017
2018
2019
bc
ab
a
c
bc
a
MIDAS
OMEGA
WUR
787-08
789-02
887

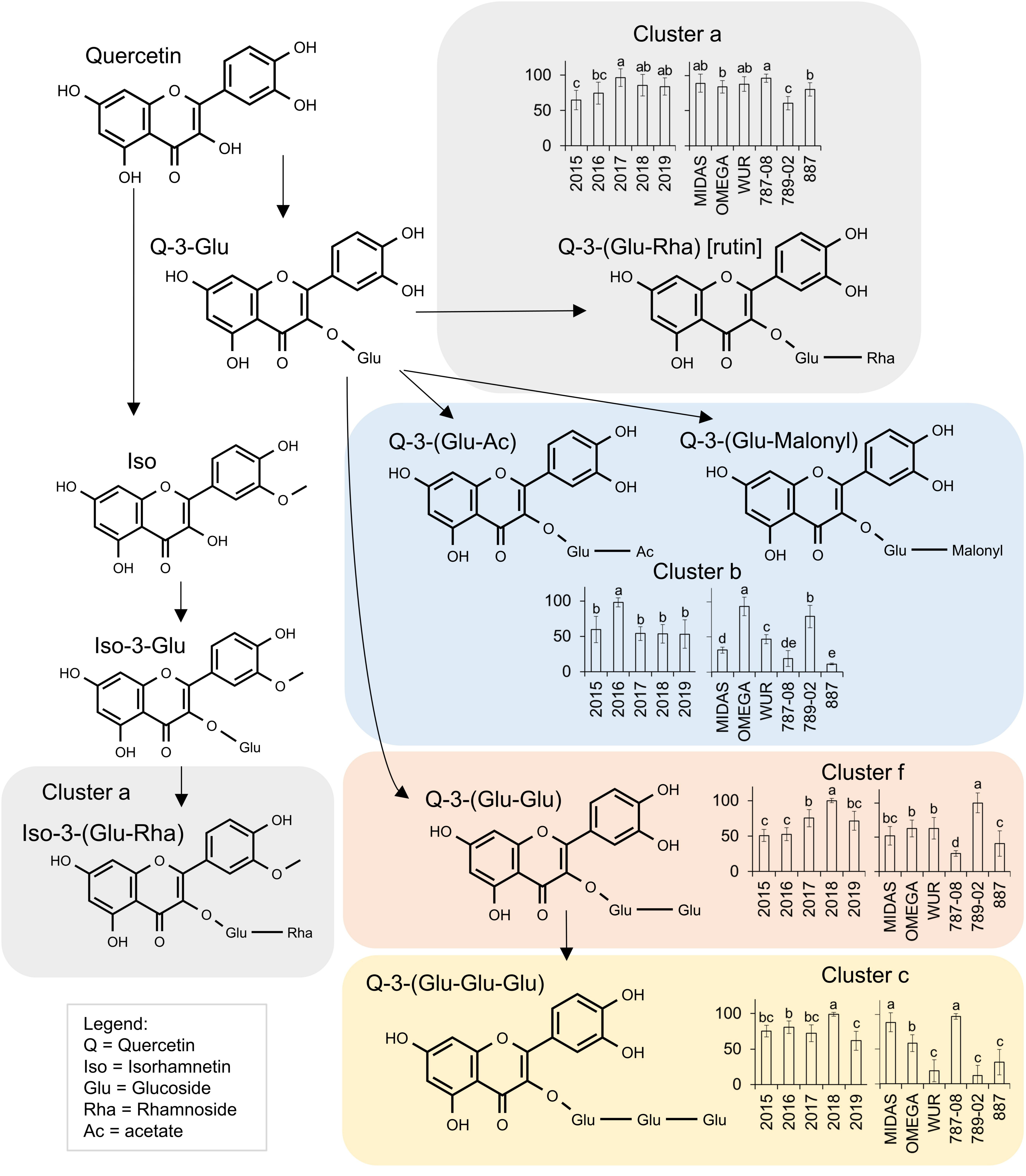

(a)
ESI +
ESI -
Phenolic acids
Flavonoids
Phenolic glycosides
Aminoacids
Cinnamic acids
Alkaloids
Nucleosides
Terpenes
Sugars-carbohydrat.
Chalcones
GSH
Vitamins
Lignans
Glucosinolates
Benzoic acids
Carboxylic acids
Unknown
(b)
Molecular network cluster annotation
33%
15%
52%
Known
Unknown
No-cluster
(c)
Known Metabolic Categories
Metabolite Nr
Flavonols 116
Aminoacids 53
Cinnamic acids 53
Sugars and carbohydrates 46
Alkaloids 14
Flavan-3-ols 14
Glucosinolates 14
Terpenes 14
Benzoic acids 13
GSH 9
Phenolic glycosides 8
Nucleosides 8
Chalcones 7
Phenolic acids 7
Carboxylic acids 5
Flavones 5
Flavanones 3
Isoflavones 3
Vitamins 3

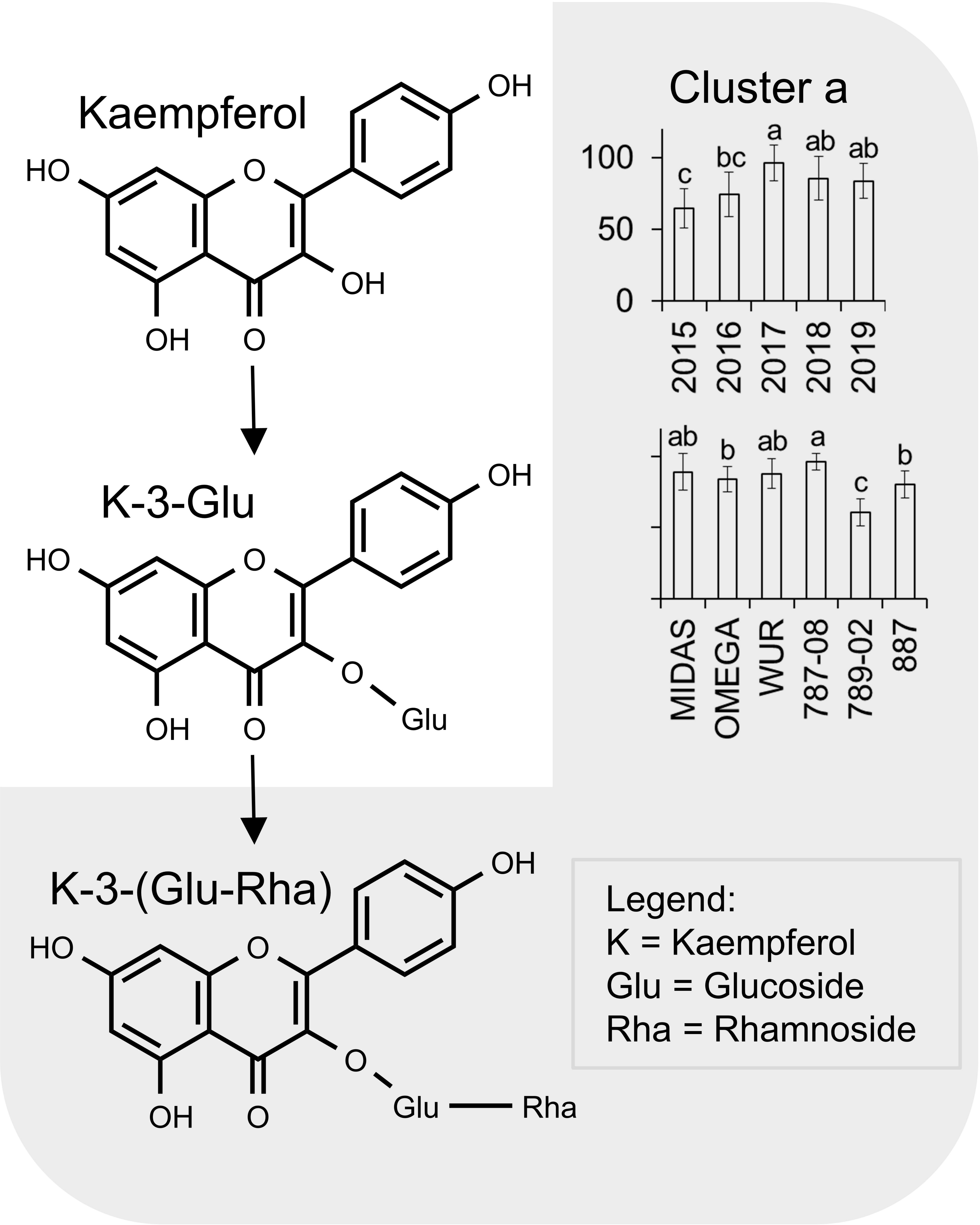
(a) Differentially accumulated Metabolites (DAM)
Genotype
Environment
GxE
0
8
12
1126
36
402
40
Group size
Nr Metabolites
1170
1548
1604
Genotype
Environment
GxE
(b)
Genotype
Nr DAM
688
946
1681
616
481
1043
1188
646
319
1414
695
725
MIDAS
OMEGA
WUR
787-08
789-02
887
Environment
Nr DAM
1643
1535
2979
734
1750
1356
1080
1882
480
2425
2015
2016
2017
2018
2019
🟨 Induced metabolites
🟦 Repressed metabolites
(c)
Genotype
Height
50
70
90
OMEGA
787-08
MIDAS
789-02
WUR
887
Environment
Height
50100
200
2015
2018
2019
2016
2017

(a)
Enriched categories G-E
Nr Metabolites
0
20
40
60
58
37
37
12
21
11
10
Flavonols
Aminoacids
Cinnamic acids
Glucosinolates
p_2_a
p_2_c
p_16
(b)
Flavonols
Aminoacids
Cinnamic acids
Glucosinolates
Pos_2_a
Pos_2_c
Pos_16
789-02
MIDAS
WUR
OMEGA
887
787-08
2019
2018
2016
2017
2015
(c)
Flavonols
Aminoacids
Cinnamic acids
Glucosinolates
MIDAS
OMEGA
WUR
787-08
789-02
887
2015
2016
2017
2018
2019
p_2_a
p_2_c
p_16
MIDAS
OMEGA
WUR
787-08
789-02
887
2015
2016
2017
2018
2019
Lower accumulation
Higher accumulation
40%
100%


Specialized metabolites co-accumulation network
Aminoacids
cluster a
Flavonols
clusters g and h
p_16
p_2_a
cluster a
Glucosinolates
Flavonols
cluster a
Flavonols
cluster b
Flavonols
cluster e
Cinnamic acids
cluster b
p_2_a
cluster b
Flavonols
cluster c
p_2_c
Cinnamic acids
cluster a
Aminoacids
cluster b
Flavonols
cluster d
Flavonols
cluster f
Flavonols
cluster i
Cinnamic acids
cluster c
Cinnamic acids
cluster d
🟥 Flavonols
🟧 Aminoacids
🟦 Cinnamic Acids
🟦 Glucosinolates
☐ Unknown
cor > 0.8


Flavonols Correlation Network
Low
High
Metabolite accumulation
2015
2016
2017
2018
2019
MIDAS
OMEGA
WUR
787-08
789-02
887
m/z
Annotation
Cluster a
p110
p84
p3
p186
p194
p8
p28
p342
p157
n5
n11
p300
n988
n181
n5 623.162 Iso-3-(Glu-Rha)
n11 593.151 K-3-(Glu-Rha)
n181 661.139
n988 1247.331
p3 611.160 Q-3-(Glu-Rha)
p8 625.176 Iso-3-(Glu-Rha)
p28 595.165 K-3-(Glu-Rha)
p84 303.050
p110 465.103
p157 317.065
p186 617.147 K-3-(Glu-Rha)
p194 479.118
p300 449.108
p342 287.055
Cluster b
n565
n109
p1127
n540
p177
n33
p171
p978
p88
n119
n575
n33 505.099 Q-3-(Glu-Ac)
n109 519.115
n119 549.089 Q-3-(Glu-Malonyl)
n540 637.141
n565 695.149
n575 563.105
p88 551.103
p171 683.145
p177 565.118
p978 639.155
p1127 697.160
Cluster c
p880
p1564
p768
p2025
n1066
n1127
n231
n1127 625.141 Q-3-(Glu-Glu-Glu)
n1066 931.252
n231 947.247
p2025 933.264
p1564 919.249
p880 803.202
p768 949.259
Cluster d
n1
n127
n98
n1 609.146 Q-3-(Glu-Rha)
n98 677.134
n127 1219.300
Cluster e
n370
n603
n170
n170 1130.289
n370 1041.279
n603 1158.320
Cluster f
p50
n739
n13
n13 595.131 Q-3-(Glu-Glu)
n739 663.118
p50 597.145 Q-3-(Glu-Glu)
Cluster g
p595
p1186
p595 478.190
p1186 464.175
Cluster h
p1533
p2011
p1533 375.092
p2011 773.213
Cluster i
n70
n99
n70 691.150
n99 809.176
Legend:
Q = Quercetin
K = Kaempferol
Iso = Isorhamnetin
Glu = Glucoside
Rha = Rhamnoside
Ac = acetate


Quercetin
Q-3-Glu
Iso
Iso-3-Glu
Cluster a
Iso-3-(Glu-Rha)
Cluster a
100
50
0
c bc a ab ab
2015 2016 2017 2018 2019
ab b ab a c b
MIDAS OMEGA WUR 787-08 789-02 887
Q-3-(Glu-Rha) [rutin]
Q-3-(Glu-Ac)
Q-3-(Glu-Malonyl)
Cluster b
100
50
0
b a b b b
2015 2016 2017 2018 2019
d a c de b e
MIDAS OMEGA WUR 787-08 789-02 887
Q-3-(Glu-Glu)
Cluster f
100
50
0
c c b a bc
2015 2016 2017 2018 2019
bc b b d a c
MIDAS OMEGA WUR 787-08 789-02 887
Q-3-(Glu-Glu-Glu)
Cluster c
100
50
0
bc b bc a c
2015 2016 2017 2018 2019
a b c a c c
MIDAS OMEGA WUR 787-08 789-02 887
HO
OH
O
Glu
Rha
Ac
Malonyl
Legend:
Q = Quercetin
Iso = Isorhamnetin
Glu = Glucoside
Rha = Rhamnoside
Ac = acetate


Kaempferol
K-3-Glu
K-3-(Glu-Rha)
HO
OH
O
Glu
Rha
Cluster a
100
50
0
c
bc
a
ab
ab
2015
2016
2017
2018
2019
ab
b
ab
a
c
b
MIDAS
OMEGA
WUR
787-08
789-02
887
Legend:
K = Kaempferol
Glu = Glucoside
Rha = Rhamnoside