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

Stéphanie Boutet[1§], Léa Barreda[1§], François Perreau[1], Jean-Chrisologue Totozafy[1], Caroline Mauve[2,3], Bertrand Gakière[2,3], Etienne Delannoy[2,3], Marie-Laure Martin-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] Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris-Saclay, CNRS, INRAE, University of Evry, 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@inrae.fr
FP francois.perreau@inrae.fr
JCT jean-chrisologue.totozafy@inrae.fr
CM caroline.mauve@universite-paris-saclay.fr
BK bertrand.gakiere@ips2.universite-paris-saclay.fr
ED etienne.delannoy@inrae.fr
MLMM marie_laure.martin-magniette@agroparistech.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 metabolite diversity and plasticity

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

Summary

Despite the essential role of Specialized Metabolites (SMs) in the interaction of plants with the environment, studying the ability of crop seeds to produce these protective compounds has been neglected. Furthermore, seeds produce a myriad of SMs providing an interesting model to investigate their diversity and plasticity. *Camelina sativa* gained a lot of interest in the past few years as rustic oil seed crop. A characterization of seed SM landscapes in six camelina genotypes grown in the field and harvested during five growing seasons has been undertaken in this work. This allowed a comprehensive annotation of seed SMs combining analyses that cluster SMs based on their chemical structures and co-accumulation patterns. These data showed broad effects of the environment on the stimulation of the seed-specialized metabolome. Among well annotated compounds, flavonols were identified as the metabolic class characterised by high plasticity, revealing significant variable accumulation according to the year and/or the genotype. Lastly, a deeper characterisation of primary metabolites and lipids in two selected genotypes has been performed. We showed that, in addition to flavonols, alkaloids and glucosinolates displayed a higher phenotypic plasticity with respect to most of the primary metabolites, including some sugars and major storage compounds such as fatty acids, proteins and most lipid classes (e.g. DAG, TAG), but similar plasticity compared to free aminoacids and carboxylic acids. This work highlighted major and unexplored effects of the environment on the seed specialized metabolome demonstrating that seeds exhibit a dynamic and plastic metabolism, with an impact on seed quality.

Significance statement

Seeds produce a myriad of Specialized Metabolites (SMs) with an essential role in the interaction of plants with the environment. We characterized SM landscapes, primary metabolites and lipid composition in the seeds of camelina genotypes grown in the open field in five consecutive growing seasons. Our results showed the predominant effect of the environment on the regulation of the seed - specialized metabolome, with a potential impact on seed quality of camelina that may also occur in other oilseed crops.

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 (or secondary) metabolites (SMs), are usually synthetized in small concentrations, and have a limited 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). In addition, many lipids are also characterized by huge diversity among plant species and many of them can be considered as SMs (Matsuzawa *et al.*, 2021; Tsugawa *et al.*, 2020). Hence, each class of metabolites comprises a huge diversity of structures leading to a wide range of biological activities (Wang *et al.*, 2019; Kosmacz *et al.*, 2020).

In plants, terpenoids are the most structurally divers group of annotated SMs, 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, alkaloids are the most represented, with more than 15000 compounds annotated (Bhatla, 2018). Finally, phenolic compounds (or phenylpropanoids) are molecules containing one or more hydroxylated aromatic rings, and are synthetized via the malonate and shikimate pathways (Vogt, 2009; Tohge *et al.*, 2017; Corso *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 *et al.*, 2020; Lepiniec *et al.*, 2006). Flavonoids are certainly the largest group of phenylpropanoids, with more than 9000 representatives (Tohge *et al.*, 2017; Corso *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 are 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 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 *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 *et al.*, 2020; Corso *et al.*, 2021). These compounds are also involved in plant nutrition and/or reproduction, they play a role in plant metal homeostasis and in plant-microbes interaction, as in the case of root flavones affecting composition rhizosphere microbial community to improve maize performance upon nitrogen deprivation (Corso and de la Torre, 2020; Corso *et al.*, 2020; Croteau, R., Kutchan, T.M. and Lewis, 2000; Maeda, 2019; Arimura and Maffei, 2016; Korenblum *et al.*, 2020; Yu *et al.*, 2021). Given the key role of SM in plant-environment interaction, metabolomics is considered the omics technique that is closest to the phenotype (Guijas *et al.*, 2018). In fact, SMs give a substantial contribution to plant phenotypic plasticity, defined as the ability of a genotype to express different phenotypes (in this case the metabolites) according to environmental conditions (Fox *et al.*, 2019; Bonamour *et al.*, 2018; Diouf *et al.*, 2020; Bradshaw, 1965; Grenier *et al.*, 2016).

Seeds constitute (directly or indirectly) a major component of the human diet, providing proteins, oil, and starch for both human and animal nutrition. They present a wide range of SMs that are important for their physiological, agronomic, nutritional and industrial qualities (Corso *et al.*, 2020; Lepiniec *et al.*, 2006; Corso *et al.*, 2021). A large diversity of SMs has been characterized in seeds of several plant species, including *Brassicaceae* (Routaboul *et al.*, 2006; Quéro *et al.*, 2016; Corso *et al.*, 2020; Lepiniec *et al.*, 2006). 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). Nevertheless, we are far from an exhaustive characterization of seed-specialized metabolite diversity and of the impact of the environment on this diversity.

Camelina [*Camelina sativa* (L.) Crantz] is a *Brassicaceae* species that has gained a lot of interest in the past few years (Zanetti *et al.*, 2021). This crop is highly tolerant 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 18:3 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 SMs 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 and induction in response to environmental conditions.

In this work, we characterized the diversity of specialized metabolites in the seeds of six camelina genotypes (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. Among the main seed SM classes, we identified those influenced by the environment and/or the genotype. Moreover, we characterized the impact of the environment on the phenotypic plasticity of primary and specialized metabolites in camelina seeds.

RESULTS

Plant development and meteorological data

In the multi-year experiment (2015-2019) conducted on six camelina genotypes grown in Italy (Bologna), the plants were able to reach maturity in about 100 days (Figure 1a). The average minimum (T_{min}) and maximum (T_{max}) temperatures (Figure 1b) measured from 50% flowering to seed maturity (Figure 1c), a particularly important phase for seed quality, were quite variable in the years studied. This particular developmental phase has been reported as the most important in camelina seed development and filling (Walia *et al.*, 2018; Righini *et al.*, 2019). 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 even higher variability among the five years studied (Figure 1b). Specifically, 2017 had the lowest precipitation (52.2 mm), whereas 2016 showed around two- to three-fold higher cumulative precipitation (161.6 mm) compared with 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 growing seasons as the 'environment'.

Plant phenology and seed quality parameters

Environmental conditions had a significant influence on camelina development across years (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. Seed oil content (%) showed low variations among genotypes and years (Tukey's range test; $P \leq 0.05$). WUR was the genotype with the lowest average seed oil content (36.7 %), while 787-08 and 789-02 (both 39.4 %) reached the highest values. Concerning years, 2016 (39.3 %) and 2019 (37.4 %) reported the highest and the lowest seed oil content, respectively, compared with the other growing seasons. Seed protein content (%) was also significantly, albeit modestly, influenced by years and genotypes. Among genotypes, 787-08 (29.2 %) had the lowest protein content, and 887 and WUR the highest (30.1 %). Concerning years, 2017 was the one reporting the highest seed protein content, 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 (Figures 1d, e).

The metabolomic landscape of camelina seeds

Untargeted SM profiling revealed a large diversity in camelina seeds (Figure 2; Table S1). Untargeted metabolomic data in ElectroSpray Ionization in positive (ESI+) and negative (ESI-) modes allowed the detection of 1595 peaks of putative known and unknown metabolites (i.e. metabolic features).

The average intensity of SMs among all samples (Table S2) was calculated and identified the top-10 with higher average intensity (Figure S1a). Among them four were flavonoids (Rutin, Quercetin-3-(Glucose [G] - Rhamnose [R] - Pentose [P]), Isorhamnetin-3-(G - R) and Catechin-3-G), one cinnamic acid (Sinapine), and one aminoacid (γ-glutamyl leucine). In addition, four compounds had no annotation in homemade or public libraries (n8, n14, p29, p117; Figure S1b). The putative structure of these 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). The proposed structures of n8 and n14 belong to the glucosinolate class, more precisely n8 was annotated as Glucocamelinin and n14 as Glucoarabin. p29 was putatively annotated as a sugar while metabolite p117 was annotated as an aminoacid.

LC-MS/MS data were used to build a molecular network with MetGem software (Figure 2a, b; Table S1). This method allows clustering of compounds with high degree of structural 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 assignment of 30% of the detected molecules to a metabolic category. Conversely, 18% 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 2a). Overall, SMs identified in camelina seeds have been assigned to sixteen putative metabolic categories: flavonoids (flavonols, flavan-3-ols, flavones, flavanones and isoflavones), aminoacids and derivatives, cinnamic acids, carbohydrates, alkaloids, benzoic acids, glucosinolates, terpenes, nucleosides, chalcones, phenolic glycosides, phenolic acids, glutathione (GSH) and lignans (Figure 2c). Flavonols were the most represented metabolic category (68), followed by aminoacids and derivatives (43), cinnamic acids (40) and sugars and carbohydrates (36). A relatively lower number of metabolites were annotated as alkaloids (13), flavan-3-ols (13), benzoic acids (13) and glucosinolates (10). The other metabolic categories presented less than 10 metabolites.

Genetic and environmental regulations of camelina seed specialized metabolome

The ESI+ and ESI- relative metabolite intensities data were analysed by a linear model to evaluate the genotype and year effects as well as their interactions. After filtering, statistical analyses were performed with 1481 out of the 1595 metabolic features detected (see experimental procedures). Pairwise comparisons were carried out between the different years (*e.g.* 2015 vs. 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]). From this point on Differentially Accumulated Metabolic features (DAMf) will be used to indicate the number of metabolic features showing significantly different levels (induced or repressed) between two samples or group of samples/factors. Statistical analyses highlighted 1111 and 1378 DAMf in response to the genotype and the environment (*i.e.* the year), respectively (Figure 3a). Furthermore, 1452 DAMf were affected by genotype X environment interaction. Thus, statistical analyses revealed that the environment was the most important factor influencing metabolite accumulation and affected 93% of the analysed camelina seed-specialized metabolome. To measure the specific impact of each year and genotype, DAMf were identified for each comparison (Figure 3b, Figure S2). Overall, the year 2016 showed the largest number of cumulated DAMf, with 2494 induced and 660 reduced metabolites compared to the other years among all comparisons (Figure 3b, Figure S2). Concerning genotypes, OMEGA showed the largest number of cumulated DAMf, with 1690 and 402 induced and reduced accumulation of SMs, respectively, compared with the others (Figure 3b, Figure S2).

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

The analyses highlighted a strong and broad effect of the environment on SM diversity and accumulation, and the majority of DAMf that were in common between the environment and the genotype (G-E). In this light, the G-E DAMf (1051 metabolic features) were considered for further analyses (Figure 3a). Using this subset of DAMf, we performed an enrichment analysis using the metabolic categories highlighted by the molecular network analysis (Table S1, column 'Metabolic category'). Not-annotated metabolites clustering together in the molecular network analysis (Figure 2b) and therefore having similar chemical structures, were considered as a distinct metabolic category (see methods for more information). Among the DAMf common to genotypes and environments, flavonols (47 DAMf), cinnamic acids (35 DAMf), and glucosinolates (10 DAMf) were found to be enriched compared with the metabolite reference set library (see methods for further information; Figure 4a). In addition, two unknown categories, namely p_2_a (20 DAMf) and p_2_c (9 DAMf) were also enriched (Figure 4a). The heatmaps built using the average metabolite intensity for each category showed that the environmental conditions of 2016 induced higher average metabolite accumulation compared with the other years for the enriched metabolic categories, except for flavonols and, to a lesser extent, to not annotated category p_2_c (Figure 4b). Conversely, 2018 and 2019 were the years during which cinnamic acids, glucosinolates, p_2_a and p_2_c metabolic categories displayed the lowest accumulation levels. In addition, the genotypes OMEGA and 787-08 showed higher amounts of metabolites for most of the 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 metabolite categories (Figure 4b).

Identification of important metabolite hubs and their environmental induction

The results obtained so far highlighted large variations according to the environment and/or the genotype of several metabolic categories. Nevertheless, it remained unclear how genotypes or the environmental conditions affected the accumulation of specific metabolites or their decorations (e.g. glycosylated / acylated / methylated quercetin or kaempferol). To establish a link between metabolite structures (and therefore their putative/potential biological activity) and environmental conditions or genotypes, a co-accumulation network analysis was performed based on correlation among metabolite intensity (Figure 5; $r > 0.7$) using metabolomic data extracted for the enriched metabolic classes (Figures 6 and S4). Co-accumulated metabolites obtained from the metabolite correlation network were visualised in boxplots for each cluster. In comparison with the other enriched categories, which were grouped in only one cluster, flavonols were splitted into several clusters, highlighting more diversified accumulation profiles in response to year and/or genotype (Figure 6). A metabolic signature was identified for several flavonol clusters identified by co-accumulation analyses, and specific metabolite decorations were associated to each year (i.e. to different environmental conditions). Specifically, rutin (Quercetin [Q])-3-(G-R)), Kaempferol [K]-3-(G-R), Isorhamnetin [Iso]-3-(G-R), K-3-(G-R), Q-3-(G-R-P), Iso-3-(G-R-P) and Q-3-(G-R-R) (Figure 7), belong to a co-accumulation cluster that showed slightly lower accumulation in 2015 (cluster 'a'). Other flavonol clusters showed a much clearer metabolite accumulation behaviour. In particular, acetylated and malonylated mono- or di-glycosylated quercetin and isorhamnetin belong to cluster 'b', characterized by metabolites showing markedly high accumulation in 2016, and in 789-02 and OMEGA genotypes. Cluster 'c', which is characterized by 12-times higher accumulation of metabolites in 2019 compared to other years, includes K-3-(G-Malonyl), K-(OH-3-MethylGlutaryl)-G and Q-3-(G-Galloyl). Q-3-(R-G-G) (cluster 'd') and diglycosylated quercetin bounded to feruloyl (cluster 'f') were markedly more accumulated in 2016, but their accumulation patterns differ according to the genotypes. Finally, metabolites belonging to cluster 'e', which are characterized by higher accumulation in 2018 and in MIDAS and 787-08, were tri-glycosylated quercetin and flavonols with coumaroyl, feruloyl or sinapoyl alcohol groups (Figure 7).

The other annotated enriched metabolic categories, cinnamic acid and glucosinolates, were merged into only one cluster, showing higher accumulation in 2016 (Figure S3). Twenty-nine cinnamic-acid derivatives were incorporated in the network and clustered together (Figure S3a). Most of the cinnamic acids shown in the network derived from sinapic acids and/or sinapine, including different glycosylated forms (Figure S5). Finally, annotated glucosinolates were 9-methylthiononyl glucosinolate, 8-methylsulfinyloctyl glucosinolate and its chain-elongated homologues glucoarabin and glucocamelinin (Figure S6).

Both p_2_a and p_2_c enriched unknown (not annotated) metabolic categories, found using molecular network analyses, were merged into one co-accumulated clusters. As already described for the top-10 most accumulated metabolites (Figure S1), 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 and p_2_c categories were putatively annotated as aminoacids and phenolic acids, respectively. These enriched metabolic categories were characterized by higher accumulation in 2015 and 2016, especially for p_2_c category (Figure S7).

Comparison of primary metabolites, oil composition and specialized metabolite plasticity in

camelina seeds

To gain deeper knowledge on the metabolic composition and the plasticity of seed specialized (e.g. flavan-3-ols, flavonols and glucosinolates) and primary (free aminoacids, sugars, carboxylic acids, fatty acids, lipids) metabolites, additional analyses were performed on selected genotypes. The analysis of primary metabolites (92 metabolites, Table S3), and fatty acids (17 metabolites, Table S4) as well as lipids (1476 metabolic features [Mf], Table S5) were obtained by GC-MS and LC-MS/MS, respectively. Metabolite analyses were carried out on seeds of OMEGA and 789-02 genotypes harvested from 2015 to 2019. These genotypes were chosen based on their contrasted profiles for SMs (Figure 3b, 4b), OMEGA and 789-02 being the ones showing the highest and lowest SM accumulation, respectively. A co-accumulation network analysis (similarly to what has been done previously for SMs; Figure 8a) was performed to calculate the phenotypic plasticity, using a widely used index (Relative Distance Plasticity Index, RDPI; Valladares *et al.*, 2006), for the main primary and specialized metabolic categories (Figure 8b; Table S6). The RDPI index allows the quantification of phenotypic plasticity based on phenotypic distances between different individuals of a given species (e.g. camelina plants) exposed to different environments (e.g. 2015 to 2019). While this index has been used to compare plasticity among different species, we applied the method to compare the plasticity of different phenotypes (i.e. the metabolic classes) from the same species and tissue. Co-accumulated (i.e. highly correlated: $r > 0.8$) metabolites were visualised in heatmaps build for each cluster and obtained from the metabolite correlation network (Figure 8a), and the phenotypic plasticity was calculated for each cluster and category (Figure 8b), using the average metabolite intensity (see method for further information). The correlation cut-off has been increased compared to the previous correlation network because of the lower number of genotypes considered and, therefore, of samples used to calculate the correlation.

Alkaloids, flavonols, and glucosinolates SMs, together with free aminoacids and carboxylic acids primary metabolites, were the categories showing the largest plasticity level. Conversely, seed oil and protein contents, major fatty acids ($18{:}1^{\Delta 9}$, 18:2 and 18:3), most lipid categories (Diacylglycerols, DAG; Triacylglycerols, TAG; Phospholipids) and protein content were characterized by markedly lower plasticity than that just reported for SMs. These categories showed indeed a quite similar accumulation behaviour and no clear metabolic signature across the five years considered in this study. Lower plasticity was also observed for cinnamic acids, flavan-3-ols, and one sugar cluster (which included sucrose, the most abundant sugar). According to metabolite accumulation profiles, 2016, 2017 and 2019 were the years inducing the highest number of specialized and primary metabolites (Figure 8a), including alkaloids (2016), flavonols (2016 and/or 2019), glucosinolates (2016 and 2017), free aminoacids (2016 and 2019), sugars (2016 and, 2017 or 2019), carboxylic acids (2019).

Last, carbon and nitrogen content in seeds of the two genotypes was also analysed. The results for C/N did not reveal a significant impact of genotype nor environment in camelina seeds (Figure S8).

The environmental regulation of oil composition in camelina seeds

Untargeted lipidomic analyses allowed the annotation of 199 metabolic features (out of 1476), mainly classified into 7 metabolic categories: TAG (85 Mf), Phospholipids (43 Mf), DAG (33 Mf), Fatty acid methyl esters (5 Mf), Sterol lipids (5 Mf), Monoradylglycerols (3 Mf) and Glycerolipids (1 Mf) (Table S5). We also detected (+)-.gamma.-Tocopherol, and minor classes corresponding to Phosphatidylmethanols (8 Mf) and Fatty acid methyl esters, respectively (5 Mf). These last two metabolic classes could be artefacts derived from the extraction protocol that was used. Statistical analyses reveal that 596 (40 %) lipid DAMf were affected by both genotype and environment, belonging to the DAG, TAG, phospholipids and sterol lipids. The heatmaps constructed using the accumulation of metabolites for each category or separately for all metabolites showed that the environmental conditions of 2019 induced lower levels of metabolite accumulation compared to the other years (Figure 9b, Figure S9). In addition, 2017 and 2018 were characterized also by moderately lower level of many DAG and sterol lipids. As already shown by the analysis of phenotypic plasticity, fatty acids (FAs) profiles were stable among the different years, despite the significant but moderate impact of the environment on the accumulation of some major FAs was moderately but significantly affected by the environment ($C18:1^{\Delta 9}$, C18:2, C18:3) (Figure 9a).

DISCUSSION

Phenotypic plasticity refers to the ability 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 overall impact of the environment on the induction or repression of the specialized metabolome, especially in seeds. Studying these processes is particularly interesting in crops, including *Brassicaceae* species such as camelina, a cultivated oilseed crop, whose seeds constitute a valuable feedstock for food, feed and biobased industrial applications. Although it is well known that camelina seeds contain numerous SM classes, most available studies focused on a limited number of compounds, mainly 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 SM classes (16 in total) in camelina seeds (Figure 2). The present results highlighted that only 30% of the identified metabolites are annotated or belong to a cluster with annotated metabolites. This is not totally unexpected, since the metabolites present in public databases used for annotation represent only a small portion 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 when only the highly accumulated metabolites are considered (Table S1). Despite the large fraction of non-annotate metabolites, using Sirius and molecular network analyses allowed us to increase the percentage of metabolites that were assigned to a metabolic class from 12% (public databases) to 30%. In addition, we significantly expanded the identification of metabolic classes compared with 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). Major challenges in the selection of metabolite datasets and their annotation include the duplication of the signal for metabolic features and in-source fragments causing SM miss annotation (Xu *et al.*, 2015; Domingo-Almenara *et al.*, 2018; Godzien *et al.*, 2015). Thus, the impact of in-source ion fragmentation was limited by filtering the metabolic features showing high correlation ($r > 0.95$) and the same retention time.

Given the five years studied characterized by high meteorological variability and the six genotypes, the trends observed allowed to hypothesize on the major factors impacting the specialized metabolome landscape in camelina seeds. Although the selected genotypes were obtained from different breeding programs and showed major differences in morphology, seed yield, protein and oil content (Righini *et* *al.*, 2019; Zanetti *et al.*, 2017; Krzyżaniak *et al.*, 2019), metabolite accumulation and diversity can be mainly linked to different weather conditions occurring in each growing season. Indeed, 2016 was characterized by lower temperatures and strikingly higher amounts of precipitation during the seed filling phase, while 2018 and 2019 showed higher temperatures and lower precipitation during the seed filling period. It is worth remarking that while camelina is able to cope quite well with drought stress (Enjalbert *et al.*, 2013), its ability to handle high temperatures (Carmo-Silva and Salvucci, 2012) during the seed filling phase is quite limited, and this is in strong agreement with the present results, which highlighted the lowest metabolome induction under those conditions(Figure 3b; 2018 and 2019 growing seasons). In addition, a hailstorm occurred in 2016 prior to harvest, which might have also affected the accumulation of some classes of seed-specialized metabolites. Many of them being well-known antioxidants, which can also act as natural plant defence to adverse and stressful environmental conditions.

Among the annotated metabolites modulated by environment and genotype, flavonols were the most represented class (Figure 4). 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 (six, Figure 6) identified for this category. Hence, since the number of clusters can be directly related to the diversity of a metabolic class, the present data suggest that flavonols are a very plastic category among the SMs observed in camelina seeds. The large accumulation of flavonols in different years could be linked to the ability of plants to respond to different environmental stresses re-directing some biosynthetic pathways. This diversity could be linked to flavonol structure with two aromatic rings and multiple potential decoration sites (glycosylated, methylated and/or acylated; Figures 6 and 7) compared with cinnamic acids (Figure S5) and glucosinolates (Figure S6). The importance of decorations in metabolite activity and plant-interaction with abiotic and biotic factors has been recently highlighted by Behr *et al.* (2020). This is also the case of N-hydroxy-pipecolic acid, whose glycosylation affects plant immunity responses and plant growth (Cai *et al.*, 2021; Mohnike *et al.*, 2021).
The accumulation of flavonols in response to abiotic stresses has been extensively documented for vegetative tissues (Corso *et al.*, 2020; Nakabayashi *et al.*, 2014; Tohge *et al.*, 2016; Pourcel *et al.*, 2007; Muhlemann *et al.*, 2018; Emiliani *et al.*, 2013), while very few information are available for seeds (Macgregor *et al.*, 2015; Corso *et al.*, 2020) The antioxidant and protective activities of some flavonols would explain their higher accumulation in 2018 and 2019, which presented the highest temperatures during the seed filling phase, and in 2017, characterized by very limited amounts of precipitation (Figure 1b; Figure 4).

The present 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 fluctuating temperatures and high precipitation. In relation to the effects of the environment, the Q -diglycoside – malonyl or feruloyl, and Q-triglycoside, together with acylated and malonylated Q-glycosides were markedly more accumulated in 2016, characterized by high precipitation rate and low temperatures (mix of the two stresses/conditions). Instead, Q and/or K –monoglycoside and -triglycoside (+ malonyl, galloyl or feruloyl) were significantly more accumulated in 2018 and 2019, characterized by higher temperatures (heat stress). Glycosylation, acylation and methylation play crucial roles in stability, solubility and bioavailability of flavonoids. For instance, de-glycosylation is a well-characterized fine-tuned mechanism regulating flavonoid localization, availability, and biological activity (Wang *et al.*, 2019). Characterizing the effects of abiotic (e.g. heat or drought stress) and/or biotic stresses on the accumulation of specific flavonoids and related decorations will be essential for understanding the environmental impact on seed specialized metabolism and seed quality.

It has been reported 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, the present study showed that the environment has a limited impact on major seed oil primary metabolites (lipids and fatty acids) and protein contents (Figures 8 and 9), which were characterized by low phenotypic plasticity. Nevertheless, DAG and sterol lipids showed slightly lower accumulation in 2018 and/or 2019 compared with the other years (Figure 9). The impact of abiotic (Hou *et al.*, 2016; Shiva *et al.*, 2020) and biotic (Jeon *et al.*, 2020; Kazaz *et al.*, 2021) stresses on lipid accumulation has been already described for vegetative tissues, but the data on seed specialized metabolome are scarce. In particular, osmotic, drought and high temperature stresses enhance the accumulation of phospholipids, including phosphatidylcholine (PC) and phosphatidylethanolamine (PE). The low variability of seed lipids and fatty acid content and composition might be related to camelina domestication history. This species, together with other domesticated *Brassicaceae* (e.g. *Brassica napus*), has been indeed selected for oil production, and showed a higher and less variable seed oil content compared with wild *Brassicaceae* (Barker *et al.*, 2007; Sharafi *et al.*, 2015).

Conversely, in addition to many SM classes (alkaloids, flavonols, or glucosinolates), a surprisingly high plasticity was observed for free aminoacids, carboxylic acids and some sugars (primary metabolites) in camelina seeds. This suggests that, in addition to the seed specialized metabolome, some metabolic classes belonging to the seed free primary metabolome can be shaped by the environment (Maeda, 2019). This could be due to the fact that many primary metabolites (e.g. free aminoacids) are precursors of major SMs, as in the case of L-phenylalanine for flavonoids (Maeda, 2019; Maeda and Fernie, 2021). A positive induction on the accumulation of seed free primary metabolites has been also observed in Arabidopsis plants subjected to ABA-deficiency, which has been linked to a premature metabolism resumption state (Chauffour *et al.*, 2019). Besides free primary metabolites, protein content and turnover might also play a major role in triggering the events leading to successful seed germination and seedling establishment (Galland *et al.*, 2014), which could explain the higher stability of camelina seed total protein content highlighted in the current work.

Regarding the differences in metabolite accumulation among the genotypes, both OMEGA and 787-08 showed the highest metabolome induction (Figure 3). Inversely, the genotype 789-02 showed the lowest metabolite accumulation level for most of the studied metabolic categories. Conversely, 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 the cultivar 787-08 could be related to an improved genetic combination allowing the diversity and abundance of metabolites. The accordance between metabolite composition and the other traits found in genotypes 787-08 and OMEGA may represent a promising strategy for future breeding in camelina.

In this work were combined two different approaches to annotate and cluster specialized metabolites based on their structure and co-accumulation, respectively. A more precise characterization of new metabolites was achieved by linking their chemical structures to their accumulation patterns. In addition, using fatty acid profiles and untargeted lipidomics it was possible to dissect the impact of the environmental conditions on seed oil composition, important characteristics for a seed oil crop as camelina.

To conclude, major and unexplored effects of the environment on the regulation of the seed specialized metabolome were highlighted, suggesting that seeds might not be exclusively storage and "metabolically static" organs, but they show a very plastic specialized metabolism.

EXPERIMENTAL PROCEDURES

Genetic material, field experimental protocol and meteorological parameters

Six different genotypes 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 genotypes 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 at 4 °C in the dark 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.

Extraction of polar/semipolar specialized metabolites, primary metabolites and lipid fraction

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 / semi-polar metabolites and lipids were dried down in a SpeedVac vacuum concentrator (o/n) and resuspended in 200 µL of ULC/MS grade water (Biosolve).

UPLC-MS/MS analysis of polar/semipolar specialized metabolites and lipid fractions

Untargeted metabolomic 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). For untargeted polar and semipolar metabolomic analysis, 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 3.5 min then re-equilibrated for 3.5-min (35-min total run time).

For untargeted lipidomic analysis, an EC 100/2 Nucleoshell Phenyl-Hexyl 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) H_20 + 1% ammonium formate in H_2O + 0.1% formic acid and (B) Acetonitrile : isopropanol (7 : 3) + 1% of 10mM ammonium formate in H_2O + 0.1% formic acid. The flow rate was of 400 µL/min and the following gradient was used: 45% of A for 1-min, followed by a linear gradient from 45% A to 30% A from 1 to 2-min, then a linear gradient from 30% A to 15% A from 2 to 7-min, a linear gradient from 15% A to 10% A from 7 to 15-min, a linear gradient from 10% A to 6% A from 15 to 19-min, a linear gradient from 6% A to 2% A from 19 to 26-min. 0% of A was hold until 27-min, followed by a linear gradient from 0% A to 45% A from 27 to 35-min (35-min total run time).

For both untargeted metabolomics and lipidomics, 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.

UPLC-MS/MS 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 of untargeted metabolomic data

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 cosine similarity score thresholds were tested. ESI- and ESI+ molecular networks were generated using cosine score thresholds of 0.75 and 0.8, respectively. Molecular networks were exported to Cytoscape software (Shannon *et al.*, 2003; https://cytoscape.org/) to format the metabolic categories.

Metabolite annotation of untargeted metabolomic data

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. Raw data were normalised on the internal standard (Apigenin) and weight of seeds used for the extraction.

Metabolite annotation of untargeted lipidomic data

Lipid metabolites annotation was performed in different consecutive steps. First, 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 and NIH Natural Product), with absolute *m/z* tolerance of 0.02; 4 minimum matched peaks and minimal cosine score of 0.65. Second, in the different clusters of the molecular network, the result of the database search was validated using the different specific fragments and neutral loss for the different lipid class with their MS^2 spectrum (*Lipid Class Specific Fragments - Lipidomics-Standards-Initiative (LSI)*). If the database search result was valid, annotation of other features was performed by stepwise comparison from the valid lipid metabolite. Finally, for the cluster of molecular network that had no data base search result, Sirius software (https://bio.informatik.uni-jena.de/software/sirius/) was used.

Chemical structures drawing

Specialized metabolites structures were drawn using ChemDraw 19.0 with default parameters

Fatty Acids analysis

Fatty acid composition of the seed oil was determined using the same protocol described in Zanetti *et al.* (2017)

Primary Metabolites: GC-MS data analysis and processing

The lower aqueous phase that was obtained after phase separation of the extracted samples was used to analyse primary metabolites via GC-MS.

GC-MS analyses and data processing were performed according to Fiehn (2006), Fiehn *et al.* (2008) and Ponnaiah *et al.* (2019). Briefly, 150µl of the extraction solution were taken and dry in a Speed-Vac evaporator for 2 h at 30°C before adding 10 µL of 20 mg.mL-1 methoxyamine in pyridine to the samples. The reaction was performed for 90 min at 30°C under continuous shaking in an Eppendorf thermomixer. 90 µL N-methyl-N-trimethylsilyl- trifluoroacetamide (MSTFA) (Regis Technologies, Morton Grove, IL, USA) were then added and the reaction continued for 30 min at 37°C. After cooling, all the samples were transferred to an Agilent vial for injection. At 4 h after derivatization, 1 µL of sample was injected in splitless mode on an Agilent 7890B gas chromatograph coupled to an Agilent 5977A mass spectrometer. The column was an Rxi-5SilMS from Restek (30 m with 10 m Integra-Guard column). An injection in split mode with a ratio of 1:30 was systematically performed for saturated compounds quantification. Oven temperature ramp was 60°C for 1 min then 10°C min-1 to 325°C for 10 min. Helium constant flow was 1.1 mL.min-1. Temperatures were the following: injector: 250°C, transfer line: 290°C, source: 230°C and quadrupole 150°C. The quadrupole mass spectrometer was switched on after a 5.90 min solvent delay time, scanning from 50 to 600 *m/z*. Absolute retention times were locked to the internal standard d27-myristic acid using the RTL system provided in Agilent's Masshunter software. Retention time locking reduces run-to-run retention time variation. Samples were randomized. A fatty acid methyl esters mix (C8, C9, C10, C12, C14, C16, C18, C20, C22, C24, C26, C28, C30) was injected at the beginning of analysis for external RI calibration. The Agilent Fiehn GC/MS Metabolomics RTL Library (version June 2008) was employed for metabolite identifications. Peak areas determined with the Masshunter Quantitative Analysis (Agilent Technologies, Santa Clara, CA, USA) in splitless and split 30 modes. Resulting areas were compiled into one single MS Excel file for comparison. Peak areas were normalized to Ribitol and Dry Weight. A total of 92 unique metabolites were identified are expressed in arbitrary units (semi-quantitative determination).

Statistical and enrichment analyses on polar and semi/polar specialized metabolomic data

Normalised metabolite accumulations were expressed as Relative Units (RU). Only metabolites having values higher than 0.05 RU in at least 1/6 of the samples (corresponding to the number of genotypes) were considered. For these 1481 metabolic features, 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 a hypergeometric test and the FDR was controlled across all the terms at level 0.05. The metabolite reference set was defined from the annotated and unknown metabolic categories (Table S1, column 'Metabolic category'). 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).

Statistical analyses on lipids and primary metabolite data

Similarly to polar/semipolar specialized metabolites, lipids and primary metabolites data were normalized to an internal standard. Differentially Accumulated Metabolites (DAM) or metabolic features (DAMf) were then identified following the same statistical pipeline used for polar/semipolar metabolites. DAM and/or DAMf affected by the environment and/or the genotype were then used for further computational analyses and data interpretation.

Total carbon and nitrogen contents

Three milligram of seeds were placed in tin capsules and burned in an elemental analyser (Pyrocube, Elementar, Lyon, France). Separation of resulting N_2 and CO_2 gas was performed on two gas-selective columns with Thermal Conductivity Detector (TCD). They can be quantified against standards (ammonium sulfate 21.2% N, benzoic acid 68.85 % C, glutamic acid 9.51% N / 40.82% C and glutamine 19.17% N / 41.09% C). The elemental carbon or nitrogen content was given in % (mass fraction).

Heatmap, hierarchical clustering, metabolites co-accumulation analysis and calculation of phenotypic plasticity

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 (maximum and minimum accumulation for yellow and blue colours, respectively; specific values are reporter for each heatmap).

Correlation was calculated among metabolites from enriched metabolic classes or with the average accumulation patterns of all annotated metabolic classes using ‘ExpressionCorrelation’ app 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.

To measure phenotypic plasticity, the Relative Distance Plasticity Index (RDPI; Valladares *et al.*, 2006) was calculated for specialized and primary metabolic categories using the *Plasticity* R package (https://github.com/ameztegui/Plasticity).

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 and IPS2 benefit from the support of Labex Saclay Plant Sciences-SPS (ANR-17-EUR-0007). This work has benefited from the support of IJPB’s Plant Observatory technological platforms and IPS2’s metabolomic platform. Authors would like to thank 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. We would like to thank as well Dr. Sébastien Baud, Dr Grégory Mouille and Prof. Loïc Rajjou (IJPB, INRAE) for the constructive discussion on seed fatty acids, primary and specialized metabolites.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY

The metabolomic data and metadata, and script used for the statistical analyses have been deposited at the INRAE data repository portal (data INRAE) with the following identifiers:

1. LC-MS/MS untargeted metabolomics, https://doi.org/10.15454/ATTENN.
2. LC-MS/MS untargeted lipidomics, https://doi.org/10.15454/ODKCCS.
3. GC-MS metabolomics, https://doi.org/10.15454/XYNDVB.
4. R script used for the statistical analyses, https://doi.org/10.15454/A3QKZI.

AUTHOR CONTRIBUTIONS

MC and FZ designed the research. MC, FZ, AM and LL directed the research. FZ and AM coordinated field experiments and furnished the seed samples for the analyses. SB and FP generated the LC-MS/MS data, performed the initial quality analyses on the metabolomic data and the metabolites annotation with SIRIUS. FZ and AM performed the analysis of fatty acids. JCT performed the untargeted lipidomic analyses (LC-MS/MS). CM and BK performed the analyses on primary metabolites (GC-MS). MC, SB and LB performed the molecular network analyses. MC, LB, ED and MLM performed the computational analyses on metabolomic data. MC and LB generated all Fig.s and supplemental material. MC and LB wrote the paper, which was edited by FZ, AM and LL. All the authors commented on and approved the manuscript.

SUPPORTING INFORMATION

Table S1. Annotation and information on the Camelina sativa specialized metabolome.

Table S2. Relative accumulation of camelina seed specialized metabolites (LC-MS/MS).

Table S3. Diversity and accumulation of seed primary metabolites (GC-MS) for OMEGA and 789-02 camelina genotypes.

Table S4. Diversity and accumulation of seed fatty acids (GC-MS) for OMEGA and 789-02 camelina genotypes.

Table S5. Diversity and accumulation of seed lipids (LC-MS/MS) for OMEGA and 789-02 camelina genotypes.

Table S6. Co-accumulation networks of primary and specialized metabolites for OMEGA and 789-02 camelina genotypes.

Figure S1. Top-10 most accumulated specialized metabolites in camelina seeds.

Figure S2. Induced and repressed metabolites identified in all comparisons for genotype and environment factors.

Figure S3. Cinnamic acid and glucosinolate co-accumulation networks.

Figure S4. Co-accumulation networks constructed with the unknown metabolic categories

Figure S5. Cinnamic acids-related pathways and clusters

Figure S6. Glucosinolate structures and clusters

Figure S7. Cluster accumulation behaviours, putative structures and annotations for the enriched but unknown metabolic classes.

Figure S8. Elemental carbon and nitrogen content in 789-02 and OMEGA seeds.

Figure S9. Heatmap representing lipids influenced by the environment and the genotype.

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* Taylor and. G. Arimura and M. Maffei, eds., New York, NY: Taylor and Francis.

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.

Barker, G.C., Larson, T.R., Graham, I.A., Lynn, J.R. and King, G.J. (2007) Novel Insights into Seed Fatty Acid Synthesis and Modification Pathways from Genetic Diversity and Quantitative Trait Loci Analysis of the Brassica C Genome. *Plant Physiol.*, **144**, 1827–1842.

Behr, M., Neutelings, G., Jaziri, M. El and Baucher, M. (2020) You Want it Sweeter: How Glycosylation Affects Plant Response to Oxidative Stress. *Front. Plant Sci.*, **11**, 1–14.

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.

Bonamour, S., Chevin, L.-M., Charmantier, A. and Teplitsky, C. (2018) Phenotypic plasticity in response to climate change: the importance of cue variation. *Philos. Trans. R. Soc. B*, **374**, 20180178.

Bradshaw, A.D. (1965) Evolutionary Significance of Phenotypic Plasticity in Plants. *Adv. Genet.*, **13**, 115–155.

Cai, J., Jozwiak, A., Holoidovsky, L., Meijler, M.M., Meir, S., Rogachev, I. and Aharoni, A. (2021) Glycosylation of N-hydroxy-pipecolic acid equilibrates between systemic acquired resistance response and plant growth. *Mol. Plant*, **14**, 440–455.

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.

Chauffour, F., Bailly, M., Perreau, F., et al. (2019) Multi-omics analysis reveals sequential roles for aba during seed maturation. *Plant Physiol.*, **180**, 1198–1218.

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. (2021) Specialized metabolites in seeds. In *Advances in Botanical Research*. pp. 35–70.

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.

Diouf, I., Derivot, L., Koussevitzky, S., Carretero, Y., Bitton, F., Moreau, L. and Causse, M. (2020) Genetic basis of phenotypic plasticity and genotype × environment interactions in a multi-parental tomato population. *J. Exp. Bot.*, **71**, 5365–5376.

Domingo-Almenara, X., Montenegro-Burke, J.R., Paul Benton, H. and Siuzdak, G. (2018) Annotation: a computational solution for streamlining metabolomics analysis. *Anal. Chem*, **90**, 480–489.

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**, 39–52.

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.

Fiehn, O. (2006) Metabolite profiling in Arabidopsis. *Methods Mol. Biol.*, **323**, 439–447.

Fiehn, O., Wohlgemuth, G., Scholz, M., Kind, T., Lee, D.Y., Lu, Y., Moon, S. and Nikolau, B. (2008) Quality control for plant metabolomics: Reporting MSI-compliant studies. *Plant J.*, **53**, 691–704.

Fox, R.J., Donelson, J.M., Schunter, C., Ravasi, T. and Gaitán-Espitia, J.D. (2019) Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. *Philos. Trans. R. Soc. B*, **374**, 20180174.

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

Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D. and Rajjou, L. (2014) Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. *Mol. Cell. Proteomics*, **13**, 252–268.

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.

Godzien, J., Armitage, E.G., Angulo, S., Martinez-Alcazar, M.P., Alonso-Herranz, V., Otero, A., Lopez-Gonzalvez, A. and Barbas, C. (2015) In-source fragmentation and correlation analysis as tools for metabolite identification exemplified with CE-TOF untargeted metabolomics. *Electrophoresis*, **36**, 2188–2195.

Grenier, S., Barre, P. and Litrico, I. (2016) Phenotypic Plasticity and Selection: Nonexclusive Mechanisms of Adaptation. *Scientifica (Cairo).*, **2016**, 7021701.

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.

Guijas, C., Montenegro-Burke, J.R., Warth, B., Spilker, M.E. and Siuzdak, G. (2018) Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat. Biotechnol.*, **36**, 316–320.

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

Hou, Q., Ufer, G. and Bartels, D. (2016) Lipid signalling in plant responses to abiotic stress. *Plant Cell Environ.*, **39**, 1029–1048.

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**, 114.

Jacobowitz, J.R. and Weng, J. (2020) Exploring Uncharted Territories of Plant Specialized Metabolism in the Postgenomic Era. *Annu. Rev. Plant Biol.*, **71**, 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.

Jeon, J.E., Kim, J.G., Fischer, C.R., Mehta, N., Dufour-Schroif, C., Wemmer, K., Mudgett, M.B. and Sattely, E. (2020) A Pathogen-Responsive Gene Cluster for Highly Modified Fatty Acids in Tomato. *Cell*, **180**, 176-187.e19.

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.

Kazaz, S., Miray, R. and Baud, S. (2021) Acyl-Acyl Carrier Protein Desaturases and Plant Biotic Interactions. *Cells*, **10**, 674.

Korenblum, E., Dong, Y., Szymanski, J., Panda, S., Jozwiak, A., Massalha, H., Meir, S., Rogachev, I. and Aharoni, A. (2020) Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 3874–3883.

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. Cell 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.

Macgregor, D.R., Kendall, S.L., Florance, H., Fedi, F., Moore, K., Paszkiewicz, K., Smirnoff, N. and Penfield, S. (2015) Seed production temperature regulation of primary dormancy occurs through control of seed coat phenylpropanoid metabolism. *New Phytol.*, **205**, 642–652.

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

Maeda, H.A. and Fernie, A.R. (2021) Evolutionary History of Plant Metabolism. *Annu. Rev. Plant Biol.*, **72**, 185–216.

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

Matsuzawa, Y., Higashi, Y., Takano, K., Takahashi, M., Yamada, Y., Okazaki, Y., Nakabayashi, R., Saito, K. and Tsugawa, H. (2021) Food Lipidomics for 155 Agricultural Plant Products. *J. Agric. Food Chem.*

Mohnike, L., Rekhter, D., Huang, W., Feussner, K., Tian, H., Herrfurth, C., Zhang, Y. and Feussner, I. (2021) The glycosyltransferase UGT76B1 modulates N-hydroxy-pipecolic acid homeostasis and plant immunity. *Plant Cell*, **33**, 735–749.

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.

Ponnaiah, M., Gilard, F., Gakière, B., El-Maarouf-Bouteau, H. and Bailly, C. (2019) Regulatory actors and alternative routes for Arabidopsis seed germination are revealed using a pathway-based analysis of transcriptomic datasets. *Plant J.*, **99**, 163–175.

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.

Sharafi, Y., Majidi, M.M., Amir, S., Goli, H. and Rashidi, F. (2015) Oil Content and Fatty Acids Composition in Brassica Species. *Int. J. Food Prop.*, **18**, 2145–2154.

Shiva, S., Samarakoon, T., Lowe, K.A., et al. (2020) Leaf lipid alterations in response to heat stress of Arabidopsis thaliana. *Plants*, **9**, 1–22.

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., et al. (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.

Tsugawa, H., Ikeda, K., Takahashi, M., et al. (2020) A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol.*

Valladares, F., Sanchez-Gomez, D. and Zavala, M.A. (2006) Quantitative estimation of phenotypic plasticity: Bridging the gap between the evolutionary concept and its ecological applications. *J. Ecol.*, **94**, 1103–1116.

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.

Xu, Y.-F., Lu, W. and Rabinowitz, J.D. (2015) Avoiding Misannotation of In-Source Fragmentation Products as Cellular Metabolites in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. *Anal. Chem*, **84**, 2273–2281.

Yu, P., He, X., Baer, M., et al. (2021) Plant flavones enrich rhizosphere Oxalobacteraceae to improve maize performance under nitrogen deprivation. *Nat. Plants*, 1–19.

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., Alberghini, B., Marjanovi, A., Grahovac, N., Rajkovi, D., Kiprovski, B. and Monti, A. (2021) Camelina, an ancient oilseed crop actively contributing to the rural renaissance in Europe. A review. *Agron. Sustain. Dev.*

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.

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 (DAMf) according to the Environment, Genotype and Genotype x Environment (GxE) factors.
(b) Number of induced (yellow) and repressed (blue) DAMf 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) Heat maps showing the average metabolite accumulation separately for years and genotypes 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 ($r > 0.7$) among metabolites and define the clusters. Only the metabolites belonging to enriched categories were considered in this analysis. 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 boxplots represent the accumulation patterns for each metabolite in all years and genotypes. The

m/z, retention time, MetGem cluster and annotations are also shown for each metabolite. The accumulation values (boxplots) were normalized and calculated as a percentage related to the sample showing the highest accumulation value for each metabolite (100% for the highest accumulation value) within years or genotypes.

Figure 7. Flavonol-related pathways and clusters

Biological pathway representing flavonol-related metabolites. The boxplots represent the accumulation patterns for each metabolite in all years. The accumulation values (boxplots) were normalized and calculated as a percentage related to the sample showing the highest accumulation value for each metabolite (100% for the highest accumulation value) within years or genotypes.

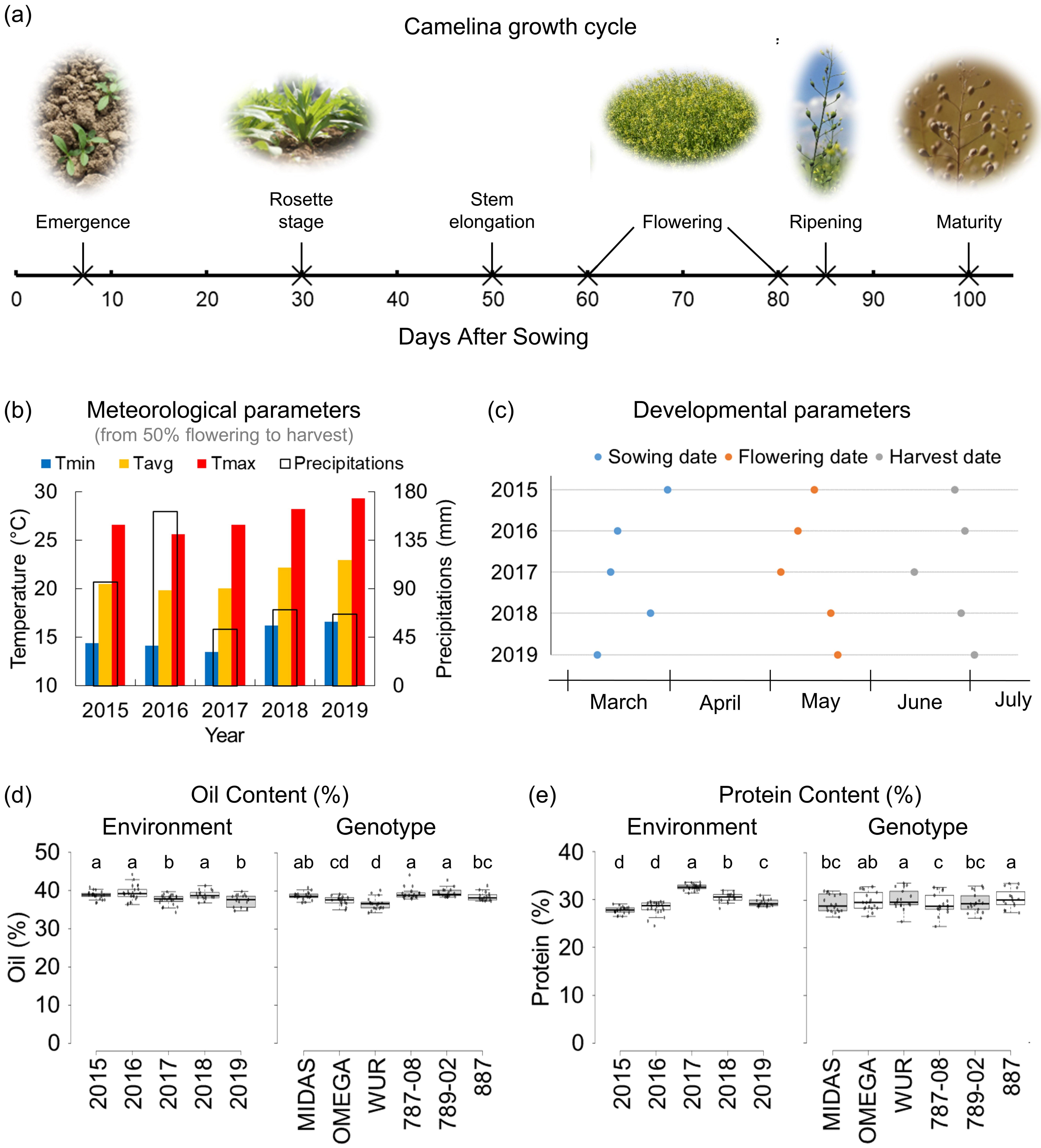
Figure 8. Comparison and phenotypic plasticity of seed primary and specialized metabolites, and oil composition in OMEGA and 789-02 varieties.

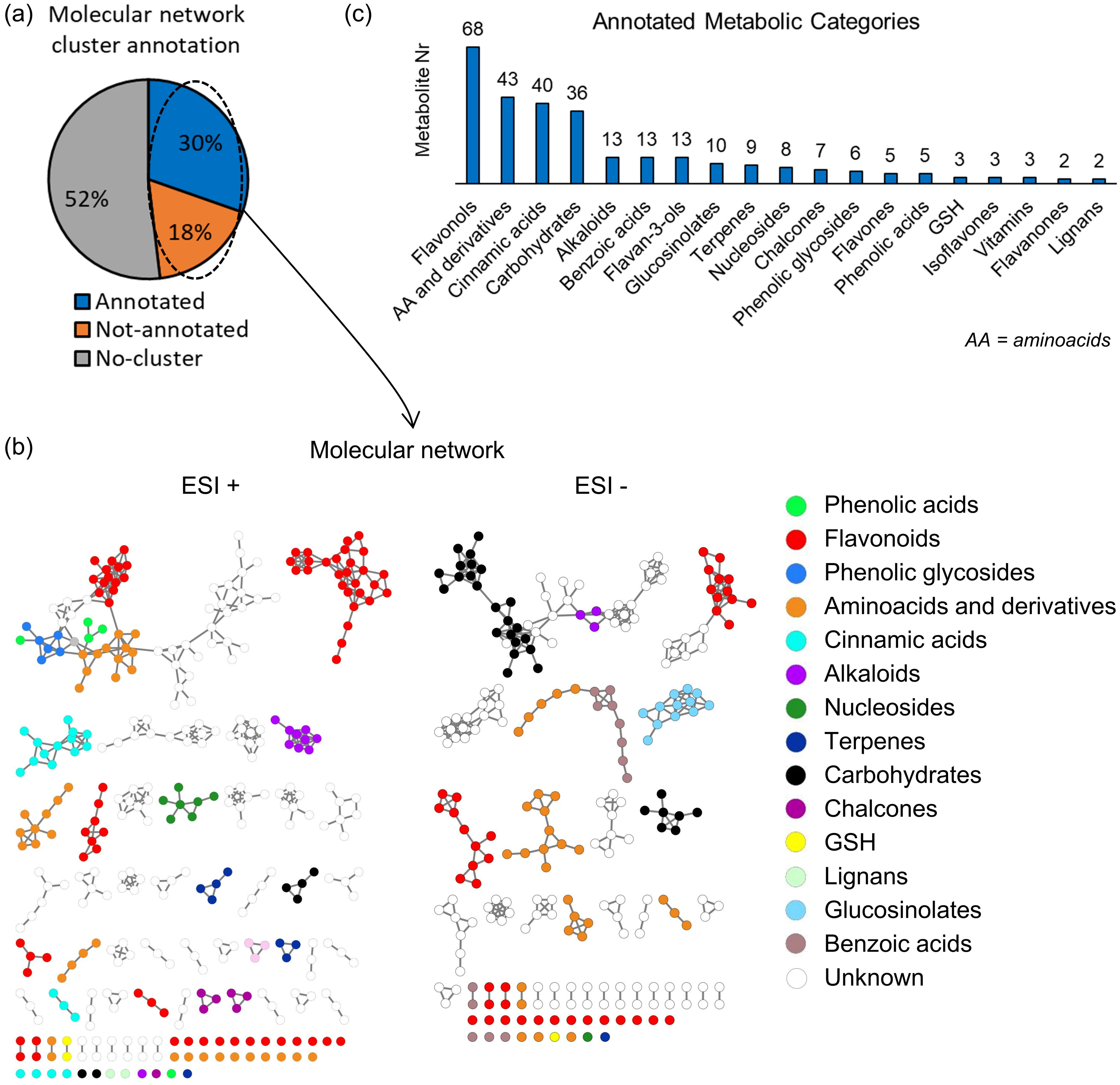
(a) Heat maps showing the average metabolite accumulation, from 2015 to 2019, for differentially accumulated metabolites belonging to specialized and primary metabolic categories. The accumulation patterns were used to calculate correlations ($r > 0.8$) among metabolites and assign a metabolic class to a specific cluster. The accumulation values were calculated as a percentage related to the sample showing the highest accumulation value for each metabolite (100% and 0% for yellow and blue colours, respectively) within years or genotypes.

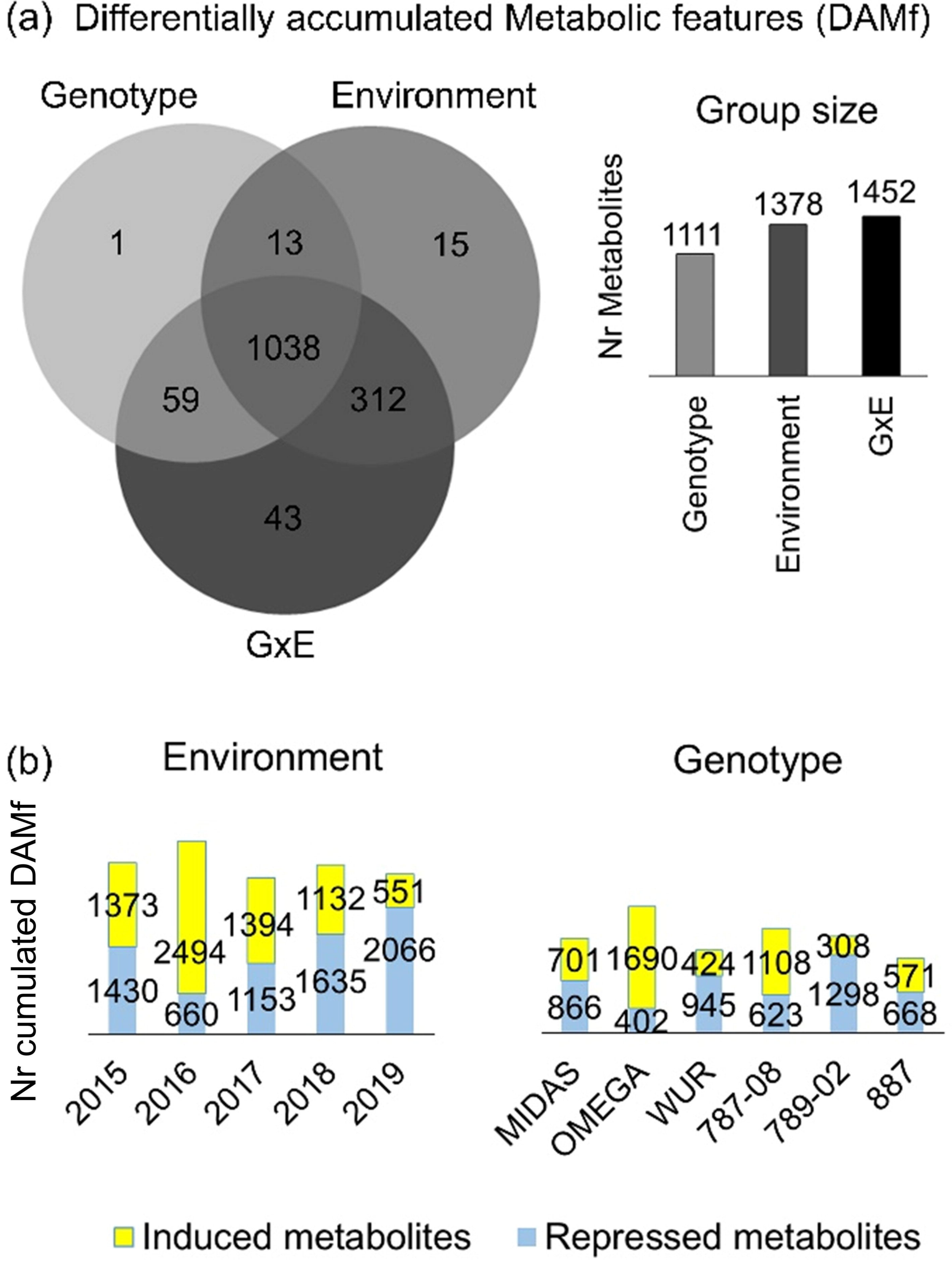
(b) Boxplots showing the Relative Distance Plasticity Index (RDPI), representing the phenotypic plasticity for the main specialized and primary (sugars, free aminoacids, carboxylic acids, lipids and proteins) metabolic categories. RDPI index is comprised between 0 (low/no plasticity) and 1 (very high plasticity).

Figure 9. Seed oil composition in OMEGA and 789-02 varieties.

(a, b) Fatty acid (a) and lipid (b) profiles and accumulation patterns are shown. Heat maps showing the average metabolite accumulation according to the different years are shown. 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.


(a)
Camelina growth cycle
Emergence
Rosette stage
Stem elongation
Flowering
Ripening
Maturity
0
10
20
30
40
50
60
70
80
90
100
Days After Sowing
(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
ab cd d a a bc
2015
2016
2017
2018
2019
MIDAS
OMEGA
WUR
787-08
789-02
887
(e)
Protein Content (%)
Environment
Genotype
Protein (%)
0
10
20
30
40
d d a b c
bc ab a c bc a
2015
2016
2017
2018
2019
MIDAS
OMEGA
WUR
787-08
789-02
887


(a) Molecular network cluster annotation
30%
18%
52%
Annotated
Not-annotated
No-cluster
(c)
Annotated Metabolic Categories
Metabolite Nr
68
43
40
36
13
13
13
10
9
8
7
6
5
5
3
3
3
2
2
Flavonols
AA and derivatives
Cinnamic acids
Carbohydrates
Alkaloids
Benzoic acids
Flavan-3-ols
Glucosinolates
Terpenes
Nucleosides
Chalcones
Phenolic glycosides
Flavones
Phenolic acids
GSH
Isoflavones
Vitamins
Flavanones
Lignans
AA = aminoacids
(b)
Molecular network
ESI +
ESI -
Phenolic acids
Flavonoids
Phenolic glycosides
Aminoacids and derivatives
Cinnamic acids
Alkaloids
Nucleosides
Terpenes
Carbohydrates
Chalcones
GSH
Lignans
Glucosinolates
Benzoic acids
Unknown



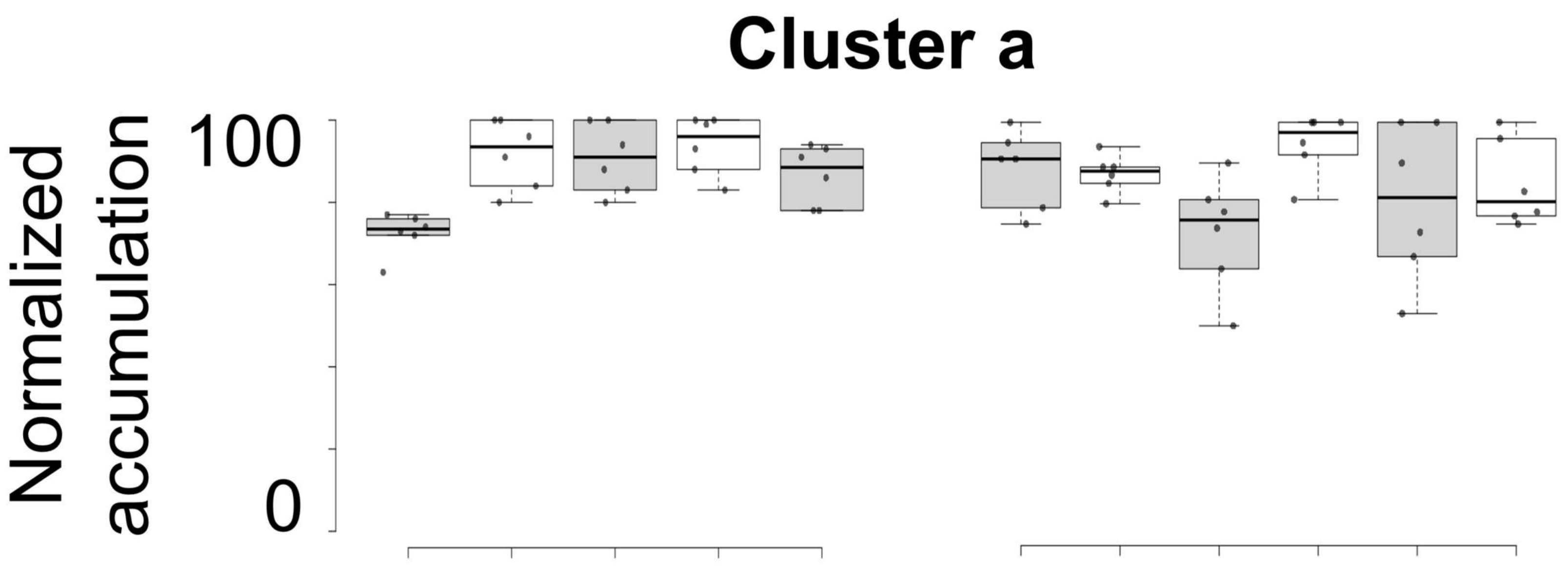
Cumulated DAMf:
DAMf present in at least one comparison, a DAMf can be present in multiple comparisons

(a)
Enriched categories Genotype-Environment
Nr Metabolites
0
20
40
60
47
35
10
20
9
Flavonols
Cinnamic acids
Glucosinolates
p_2_a
p_2_c
(b)
Flavonols
Cinnamic acids
Glucosinolates
Pos_2_a
Pos_2_c
2019
2018
2017
2015
2016
OMEGA
787-08
789-02
MIDAS
WUR
887
Lower accumulation
Higher accumulation
40%
100%

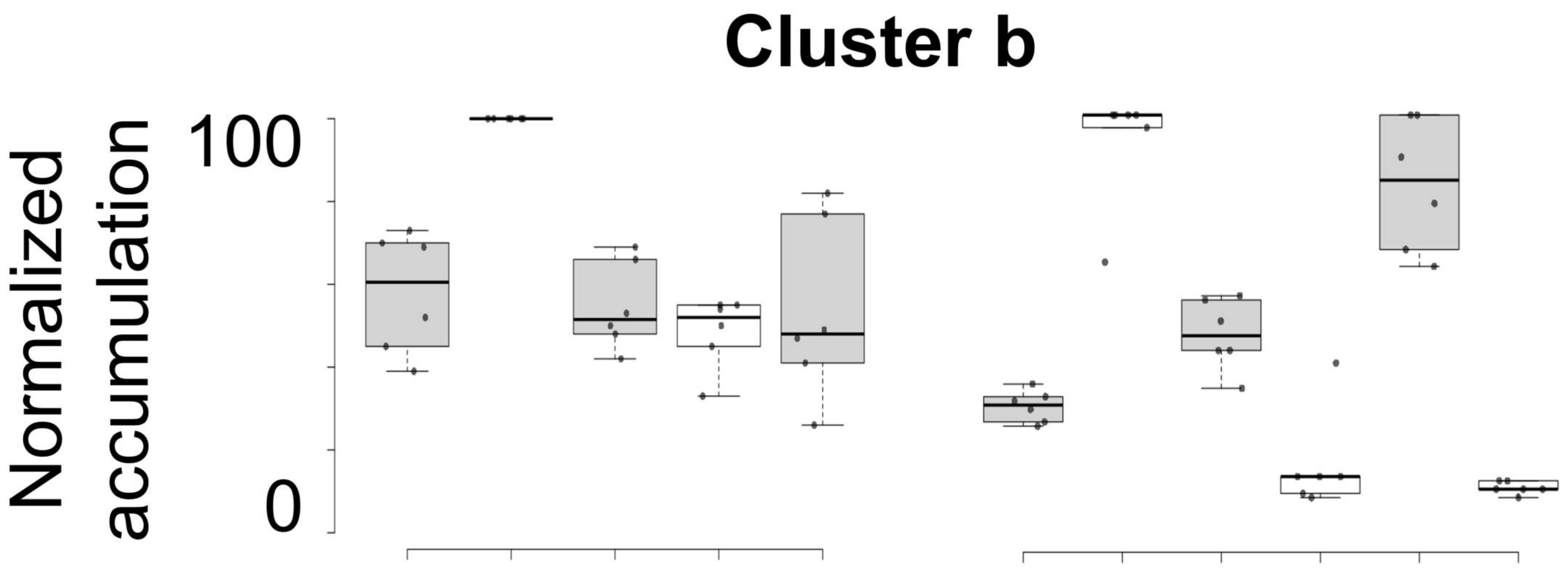
Specialized metabolites co-accumulation network

Cinnamic acids
Flavonols
cluster e
🟥 Flavonols
🟦 Cinnamic Acids
🟦 Glucosinolates
☐ Unknown
cor > 0.7
Flavonols
cluster d
p_2_a
p_2_c
Glucosinolates
Flavonols
cluster a
Flavonols
cluster b
Flavonols
cluster c
Flavonols
cluster f

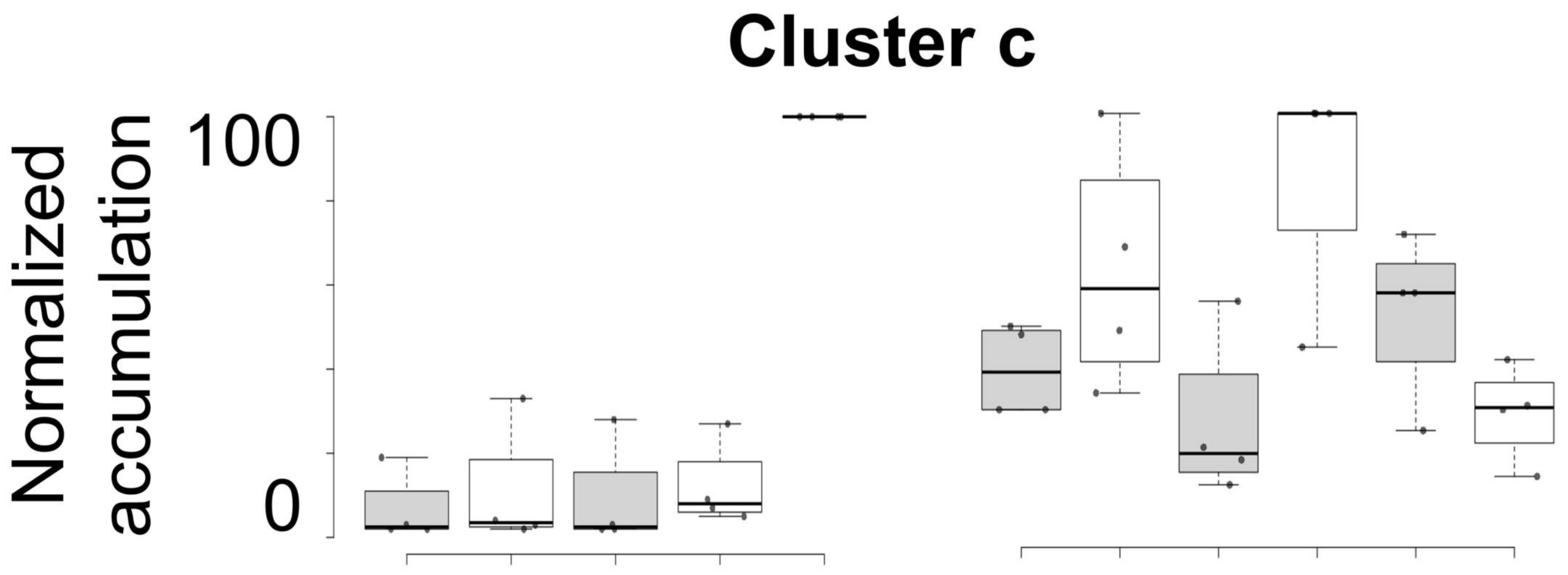
Flavonols Co-accumulation Network



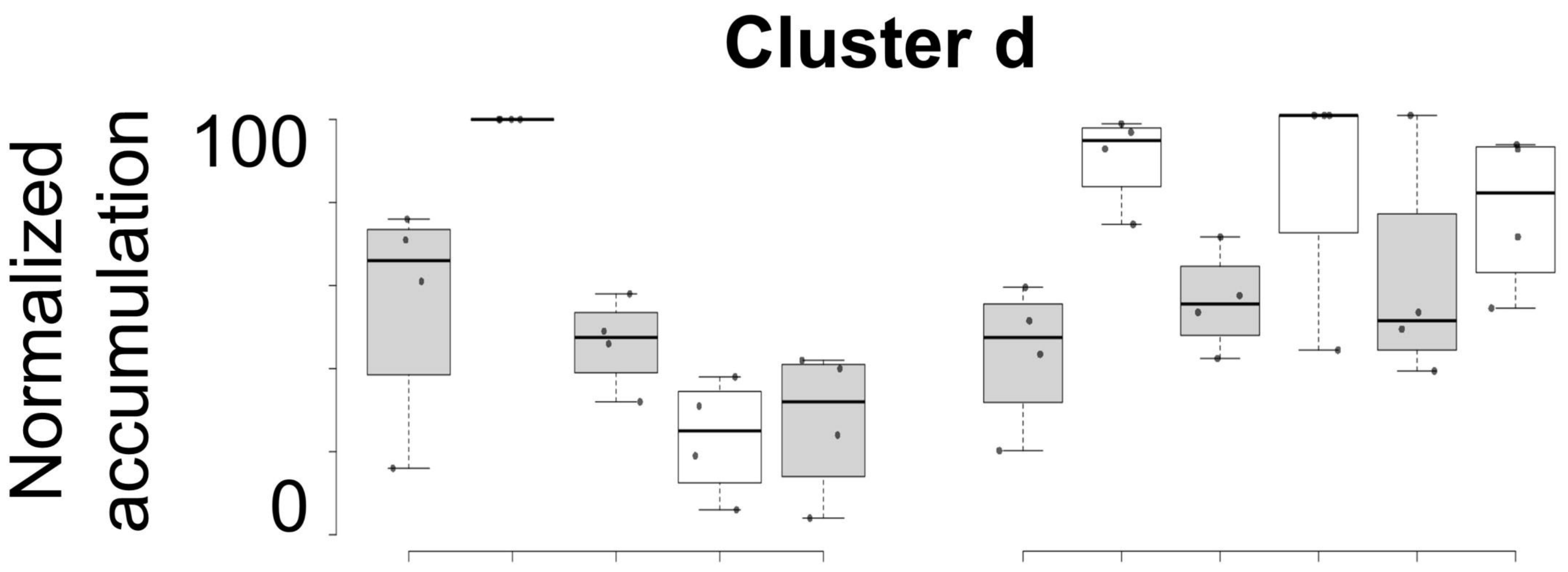
ID	mz	RT	Cluster MetGem	Annotation
p28	595.17	5.00	p_3	K-3-(G-R)
p3	611.16	4.50	p_3	Q-3-(G-R)
p8	625.18	5.20	p_3	Iso-3-(G-R)
p6	743.20	4.30	p_3	Q-3-(G-R-P)
p78	757.22	4.60	p_3	Iso-3-(G-R-P)
p403	757.22	4.10	p_3	Q-3-(G-R-R)



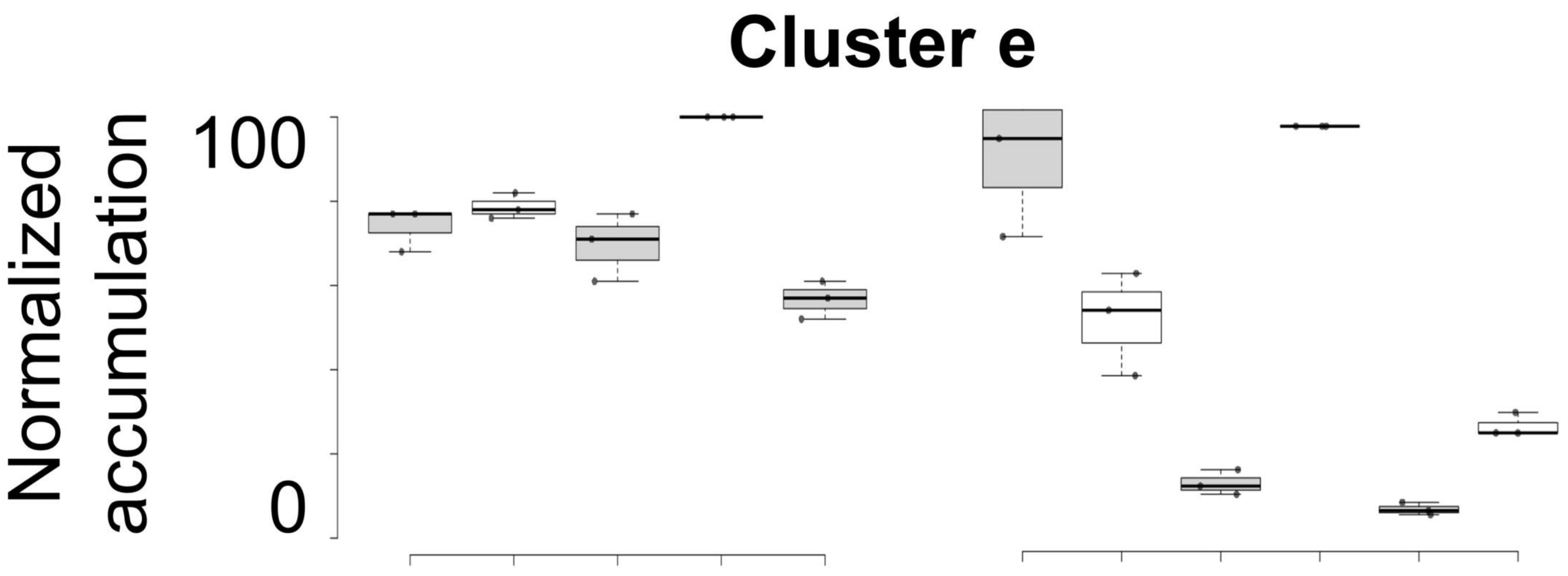
ID	mz	RT	Cluster MetGem	Annotation
n109	519.11	6.20	n_1_a	Iso-3-(G-Acetyl)
p88	551.10	5.00	p_3	Q-3-(G-Malonyl)
p177	565.12	6.30	p_3	Iso-3-(G-Malonyl)
p978	639.16	5.20	p_3	Q-3-(G-Acetyl-P)
p171	683.15	4.70	p_3	Q-3-(G-Malonyl-P)
p1127	697.16	5.50	p_3	Iso-3-(G-Malonyl-P)



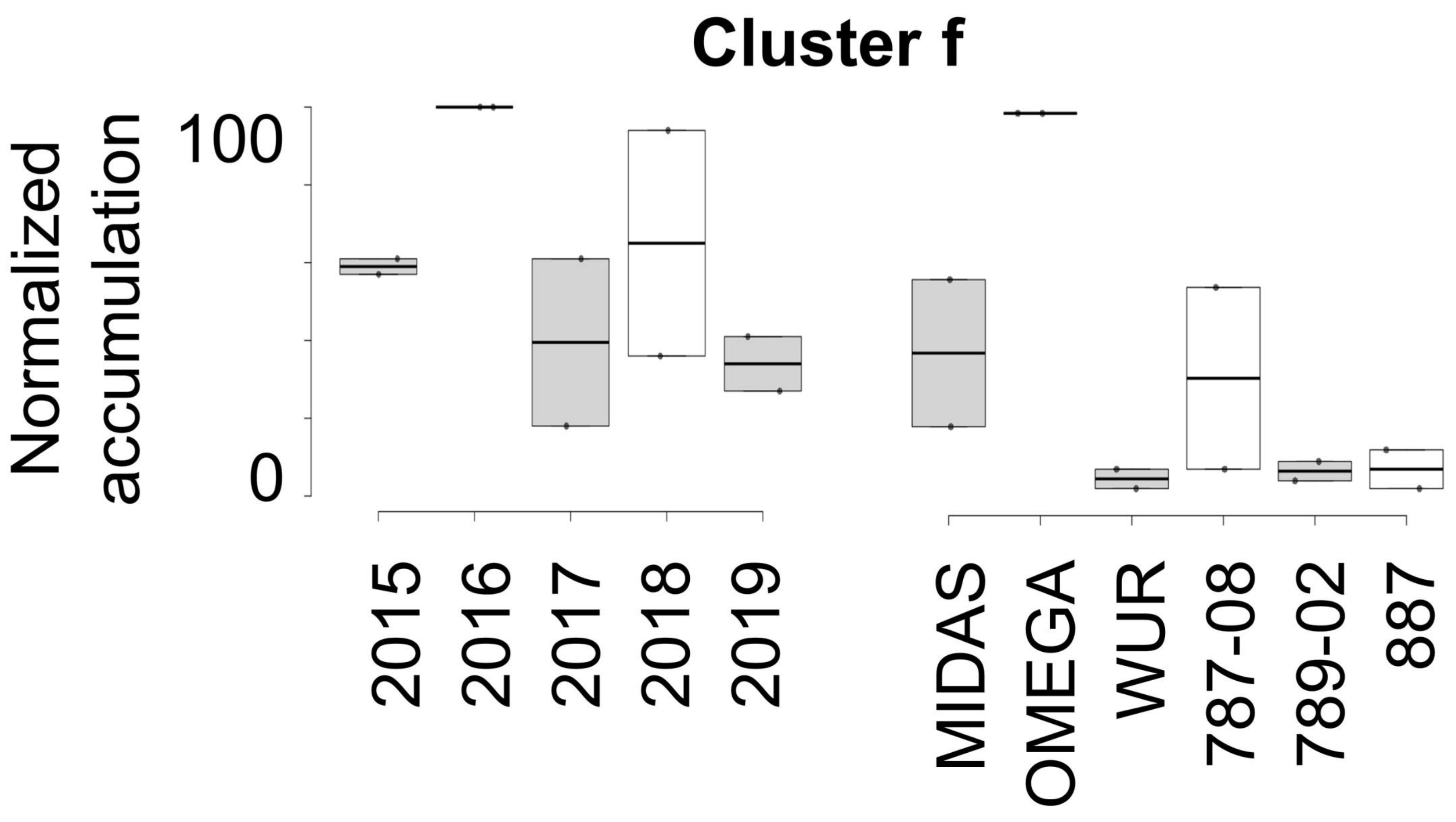
ID	mz	RT	Cluster MetGem	Annotation
n108	447.09	5.10	n_1_a	
p245	535.11	5.90	p_3	K-3-(G-Malonyl)
p290	593.15	5.90	p_3	K-(OH-3-MethylGlutaryl)-G
p1487	617.11	4.40	p_2	Q-3-G-Galloyl



ID	mz	RT	Cluster MetGem	Annotation
p1533	375.09	1.70	p_3	Q-3-(R-G-G)
p2011	773.21	3.50	p_3	
n544	897.26	4.60	n_1_a	
n554	1116.3	4.50	n_1_a	



ID	mz	RT	Cluster MetGem	Annotation
p1564	919.25	5.90	p_39	Q-3-(G-R-Coumaroyl-G)
p2025	933.26	6.20	p_39	K-3-(G-R-Feruloyl-G)
p768	949.26	5.60	p_39	Q-3-(Sinapoyl Alcohol-G-R-R)

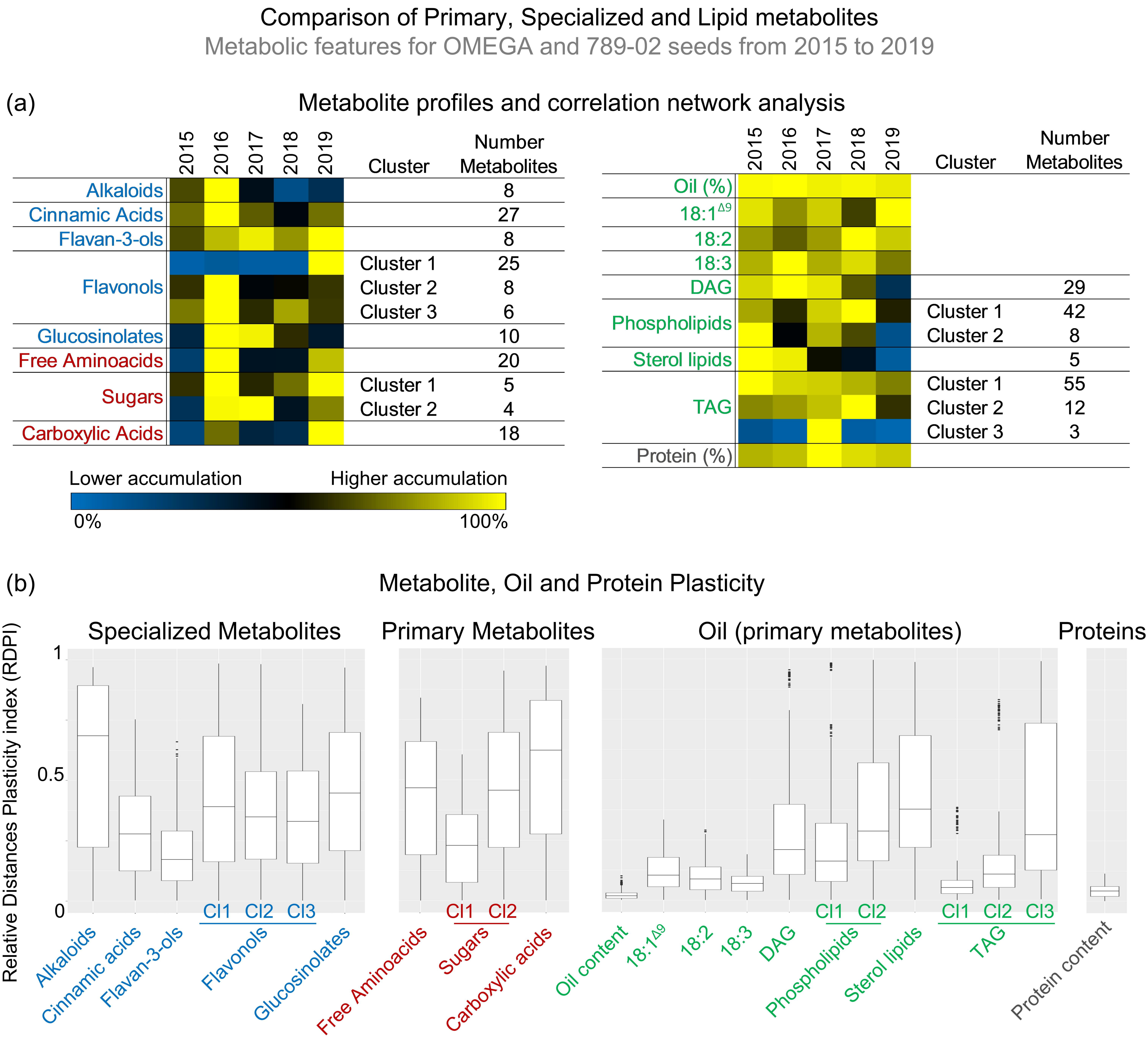


ID	mz	RT	Cluster MetGem	Annotation
n439	801.19	6.00	n_1_a	Q-3-(G-G-Feruloyl)
p836	889.20	6.60	p_39	Q-3-(Malonyl-G -Feruloyl-G)

Legend:

Q = Quercetin	G = Glucose or another hexose
K = Kaempferol	R = Rhamnose
Iso = Isorhamnetin	P = Pentose

Kaempferol
Quercetin
Isorhamnetin
Cluster a (6 SMs)
K-3-(G-R)
Q-3-(G-R) [rutin]
Iso-3-(G-R)
Q-3-(G-R-R)
Q-3-(G-R-P)
Iso-3-(G-R-P)
G - R - R
G - R - P
Norm. Acc
100
0
2015
2016
2017
2018
2019
Cluster b (6 SMs)
Q-3-(G-Malonyl)
Q-3-(G-Malonyl-P)
Q-3-(G-Acetyl-P)
Iso-3-(G-Malonyl)
Iso-3-(G-Malonyl-P)
Iso-3-(G-Acetyl)
G - Malonyl
G - Malonyl - P
G - Acetyl - P
G - Acetyl
Cluster c (4 SMs)
K-3-(G-Malonyl)
Q-3-(G-Galloyl)
G - Galloyl
Cluster f (2 SMs)
Q-3-(G-R-Feruloyl)
Feruloyl
Cluster d (4 SMs)
Q-3-(R-G-G)
R - G - G
Cluster e (3 SMs)
Q-3-(G-R- G-Coumaroyl)
G - R – G-Coumaroyl
HO
OH
O
Legend:
Q = Quercetin
K = Kaempferol
Iso = Isorhamnetin
G = Glucose or another hexose
R = Rhamnose
P = Pentose
SMs = specialized metabolites
Norm. Acc. = Normalized accumulation
Tentative structures are proposed
Flavonol Co-accumulation network



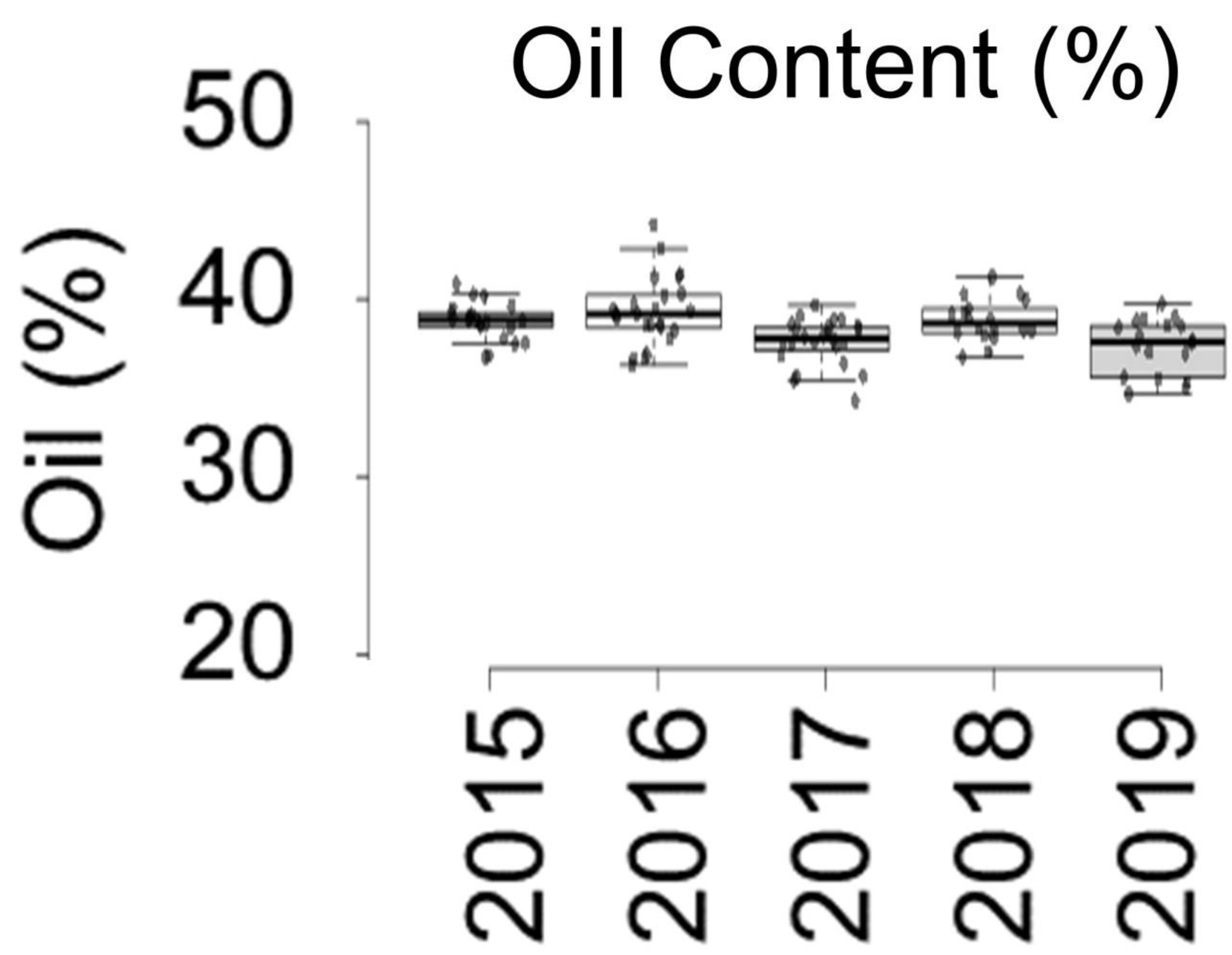
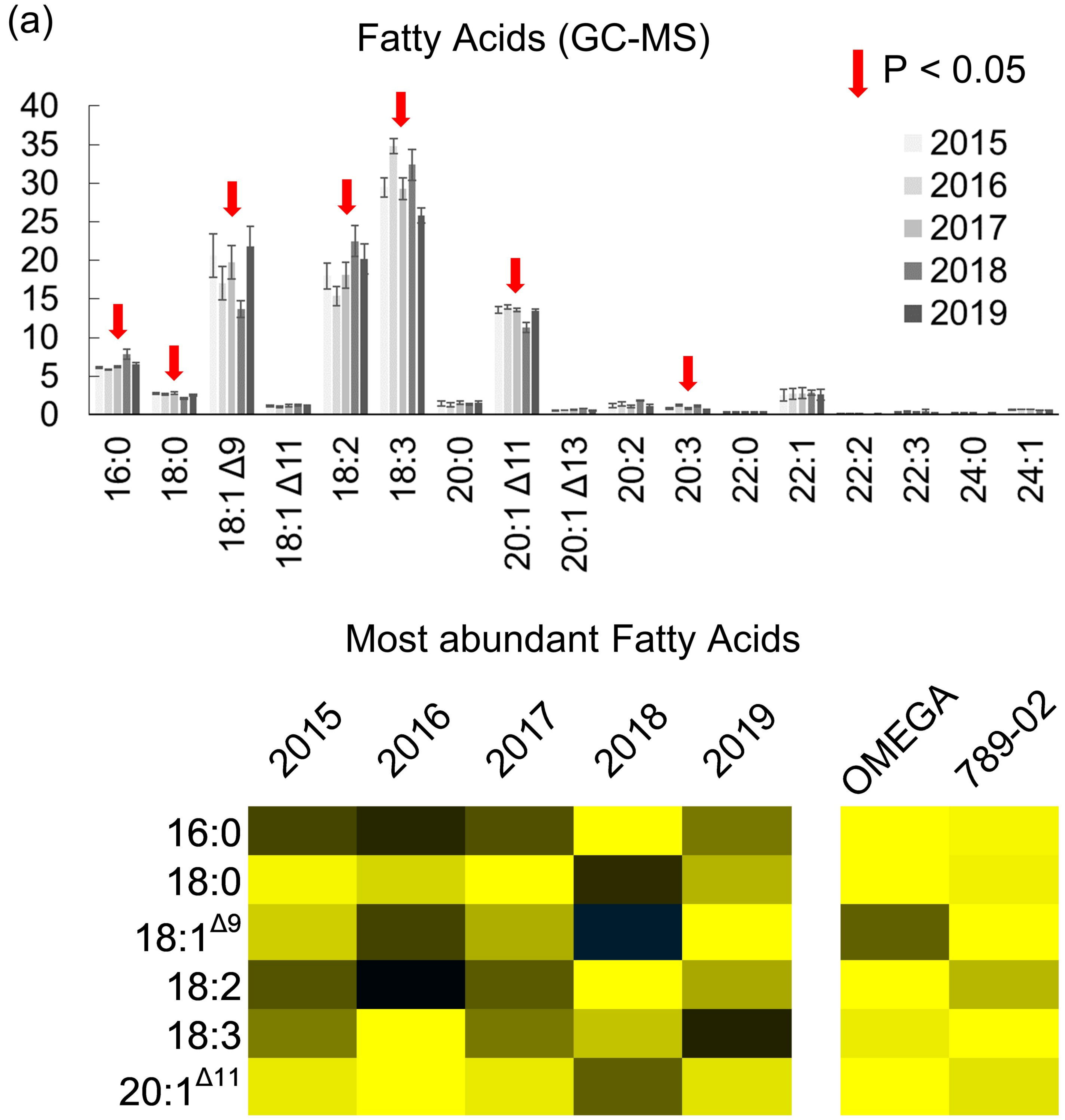
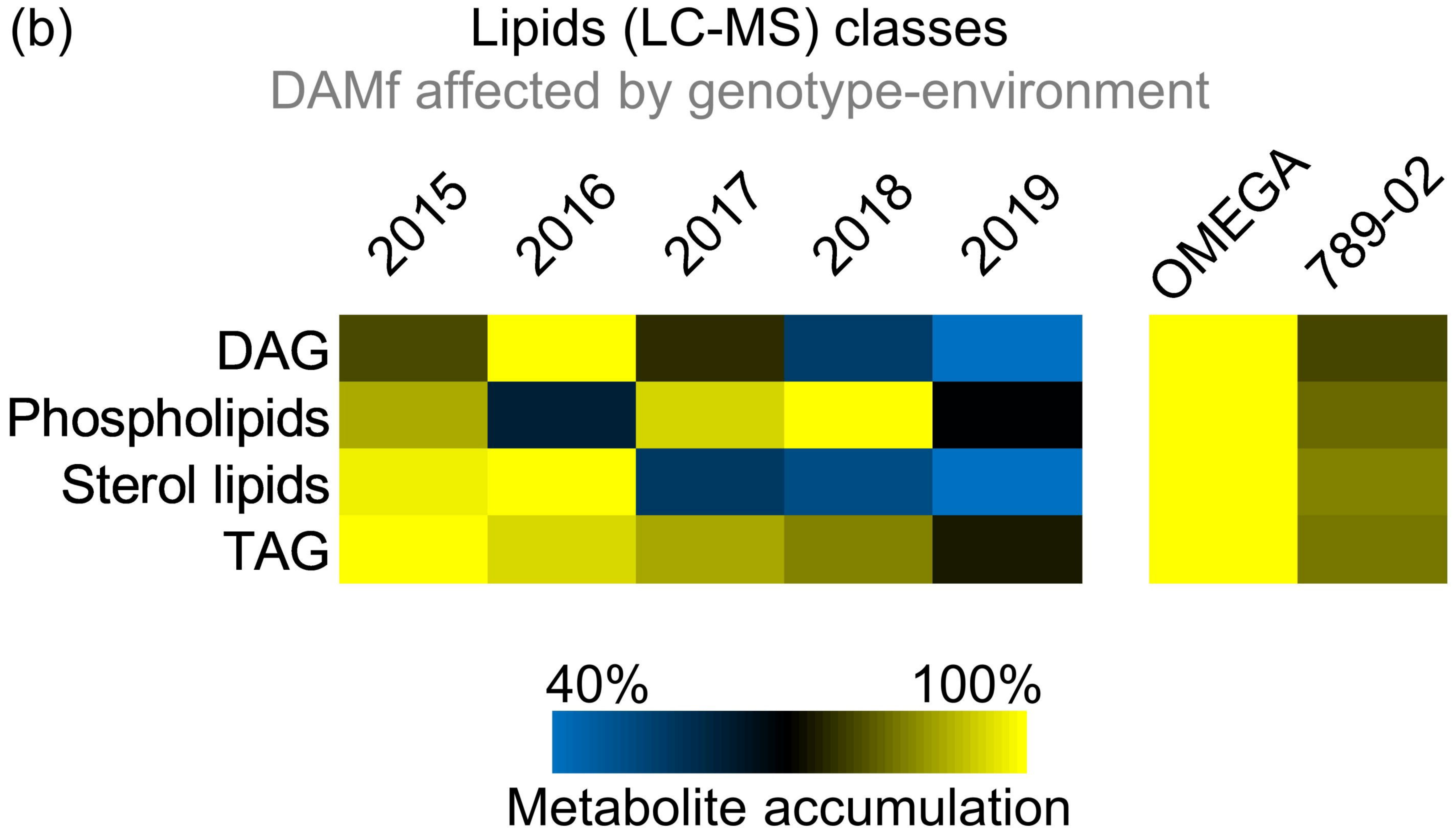
Comparison of Primary, Specialized and Lipid metabolites

Metabolic features for OMEGA and 789-02 seeds from 2015 to 2019

(a) Metabolite profiles and correlation network analysis

(b) Metabolite, Oil and Protein Plasticity

Oil composition

OMEGA and 789-02
2015 to 2019